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

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

Atmospheric, Climate, and Earth Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA

Correspondence: Lingcheng Li (lingcheng.li@pnnl.gov) and Gautam Bisht (gautam.bisht@pnnl.gov)

9   **Abstract**

10   Earth system models (ESMs) are progressively advancing towards the kilometer scale (k-scale).

11   However, the surface parameters for Land Surface Models (LSMs) within ESMs running at the k-

12   scale are typically derived from coarse resolution and outdated datasets. This study aims to develop

13   a new set of global land surface parameters with a resolution of 1 km for multiple years from 2001

14   to 2020, utilizing the latest and most accurate available datasets. Specifically, the datasets consist

15   of parameters related to land use and land cover, vegetation, soil, and topography. Differences

16   between the newly developed 1k land surface parameters and conventional parameters emphasize

17   their potential for higher accuracy due to the incorporation of the most advanced and latest data

18   sources. To demonstrate the capability of these new parameters, we conducted 1 km resolution

19   simulations using the E3SM Land Model version 2 (ELM2) over the contiguous United States.

20   Our results demonstrate that land surface parameters contribute to significant spatial heterogeneity

21   in ELM2 simulations of soil moisture, latent heat, emitted longwave radiation, and absorbed

22   shortwave radiation. On average, about 31% to 54% of spatial information is lost by upscaling the

23   1 km ELM2 simulations to a 12 km resolution. Using eXplainable Machine Learning (XML)

24   methods, the influential factors driving the spatial variability and spatial information loss of ELM2

25   simulations were identified, highlighting the substantial impact of the spatial variability and

26   information loss of various land surface parameters, as well as the mean climate conditions. The

27   comparison against four benchmark datasets indicates that ELM generally performs well in

28   simulating soil moisture and surface energy fluxes. The new land surface parameters are tailored

29   to meet the emerging needs of k-scale LSMs and ESMs modeling with significant implications for

30   advancing our understanding of water, carbon, and energy cycles under global change. The 1 km

31   land surface parameters are publicly available at https://zenodo.org/records/10523833 (Li et al.,

32   2024).

Deleted: https://doi.org/10.25584/PNNLDH/1986308 (Li et al., 2023)…

2

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 and energy cycles (Giorgi and Avissar, 1997; Chaney et al., 2018; Xu et al., 2023), biogeochemical cycles (Chaney et al., 2018), as well as land–

3

58   atmosphere coupling (Liu et al., 2017; Zhou et al., 2019; Bou-Zeid et al., 2020). For example,

59   Singh et al. (2015) demonstrated that increasingly capturing topography and soil texture

60   heterogeneity at finer resolutions (e.g., 1 km) improves land surface modeling of water and energy

61   variables. Li et al. (2022) have shown that the spatial heterogeneities of land surface parameters

62   (including land use and land cover (LULC) and topography) are essential for modeling the spatial

63   variability of land surface energy and water partitioning. Hao et al. (2022) found that 1 km

64   simulations with sub-grid topographic configurations can better capture the topographic effects on

65   surface fluxes.

66

67   The parameters for LSMs within ESMs being run at the k-scale are typically derived from coarse

68   resolution datasets or outdated datasets. Consequently, k-scale modeling may not accurately

69   represent fine-scale land surface heterogeneity unless high-resolution land surface parameters at

70   the kilometer or finer scales are utilized. Publicly available land surface parameters are primarily

71   provided at coarse resolutions and based on outdated datasets (see details in Table 1). For example,

72   the Community Land Model version 5 (CLM5; Lawrence et al., 2019) typically relies on land

73   surface parameters with spatial resolutions ranging from 1km to 0.5º based on source datasets that

74   were processed more than 10 years ago (see Table 1 for details). Although LULC-related

75   parameters are available at a relatively high resolution of 0.05º, they are temporally static and were

76   derived from a combination of data from different years spanning 1993 to 2012 (Table 1). Leaf

77   area index (LAI) was derived from the now outdated products of Moderate Resolution Imaging

78   Spectroradiometer (MODIS) collection 4 (Myneni et al., 2002). The canopy height for tree Plant

79   Functional Types (PFTs) is based on forest canopy height data derived from the Geoscience Laser

80   Altimeter System (GLAS) aboard ICESat, collected in 2005 (Simard et al., 2011). Canopy height

4

81   for short vegetation is represented by PFT-specific values that remain invariant in space (Bonan et

82   al., 2002). Soil sand and clay content were obtained from the International Geosphere-Biosphere

83   Programme (IGBP) soil dataset (Global Soil Data Task 2000) consisting of 4931 soil mapping

84   units (IGBP, 2000). These CLM5 land surface parameters have been widely utilized in the LSMs

85   and ESMs communities, despite being developed over a decade ago. Subsequently, Ke et al. (2012;

86   hereafter referred to as K2012) developed an updated set of LULC and vegetation-related land

87   surface parameters for CLM4 at a resolution of 0.05º. These parameters were developed based on

88   MODIS collection 5 products or datasets derived from MODIS collection 5 products, including

89   PFTs and non-vegetation land cover, LAI, and Stem Area Index (SAI). K2012 has also been widely

90   used by LSMs, including CLM (e.g., Leng et al., 2013; Ke et al., 2013; Singh et al., 2015; Xia et

91   al., 2017) and the Energy Exascale Earth System Model (E3SM) Land Model (ELM) (e.g.,

92   Caldwell et al., 2019; Leung et al., 2020; Li et al., 2022). However, the CLM5 and K2012 datasets,

93   with their relatively coarse resolution and reliance on outdated data from over a decade ago, may

94   not fully meet the requirements for k-scale modeling. Additionally, these datasets include LULC,

95   LAI, and SAI that are year invariant. Consequently, they are inappropriate for studies involving

96   LULC changes, such as urbanization. In addition, some recently developed land surface processes

97   and their associated parameters are not included in previous datasets. For instance, Hao et al. (2021)

98   introduced a sub-grid topographic parameterization of solar radiation with five associated

99   topographic factors in ELM, which have been found to significantly affect the surface energy

100   budget. the surface energy budget.

101

102   High-resolution and up-to-date datasets at kilometer or finer resolutions are now widely available

103   and can be utilized to derive more accurate land surface parameters for k-scale LSM simulations.

104   For example, the MODIS Land Cover Type Collection 6 (MCD12Q1 C6) data product provides

105   global land cover types yearly from 2001 to the present (Friedl et al., 2019; Sulla-Menashe et al.,

106   2019) at 500-meter resolution. Compared to the MODIS Collection 4 (used in CLM5 land surface

107   parameters) and Collection 5 products (used in K2012 land surface parameters), the C6 data

108   represents a significant advancement in algorithm improvements and the quality of land cover

109   information. Despite the availability of high-resolution MODIS LAI products, such as the 500 m

110   MCD15A2H (Myneni et al., 2021), they suffer from noise and gaps with spatially and temporally

111   inconsistent values due to clouds, seasonal snow cover, instrument issues, and uncertainties in

112   retrieval algorithms (Yuan et al., 2011). To address these limitations, Yuan et al. (2011)

113   reprocessed MODIS LAI products and generated a more accurate and spatiotemporally continuous

114   and consistent LAI dataset that is available continuously to the present period. Additional high-

115   resolution and up-to-date datasets are available for preparing land surface parameters, such as soil

116   texture and soil organic matter at 250-meter resolution (Poggio et al., 2021) and vegetation height

117   at 10-m resolution (Lang et al., 2023).

118

119   This study aims to develop a new set of global land surface parameters with a resolution of 1 km

120   for multiple years, utilizing the latest and most accurate available datasets. These parameters will

121   be tailored to meet the needs of k-scale Earth system modeling. The newly developed land surface

122   parameters include four categories: (1) LULC-related parameters, such as the spatial distributions

123   of PFTs, lakes, wetlands, urban areas, and glaciers; (2) vegetation-related parameters, including

124   PFTs' LAI and SAI for multiple years ranging from 2001 to 2021, and the canopy top and bottom

125   height; (3) soil-related parameters, such as soil textures and soil organic matter; and (4)

126   topography-related parameters, such as elevation, slope, aspect, and sub-grid topographic factors.

127   We conducted a comparison of the new 1k parameters against the K2012 and ELM2/CLM5 default

128   parameters. Utilizing ELM version 2 (ELM2) as a testbed, we demonstrated the modeling

129   capability enabled by the new high-resolution parameters through a 5-year simulation at 1 km

130   resolution over the CONUS. We performed a spatial scaling analysis on four ELM2 simulated

131   variables, which included soil moisture, latent heat, emitted longwave radiation, and absorbed

132   shortwave radiation, to underscore the significance of high-resolution land surface parameters on

133   ELM2 simulations. We employed eXplainable Machine Learning (XML) methods to evaluate the

134   most important factors of land surface parameters and climate conditions (e.g., mean temperature

135   and precipitation) in driving the spatial variability and spatial information loss of ELM2

136   simulations.

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). All data sources utilized in this study have been rigorously validated in their respective original publications. The detailed methods for deriving these parameters are described below.

146 Table 1 Comparison between new and previous land surface parameters

Category	Land surface parameters	This study	ELM2 / CLM5 *	K2012
LULC	PFTs, Lake, Glacier, Urban	• Resolution: 1 km, yearly, 2001-2020 • Data source: 500 m, yearly, MODIS collection 6 (Friedl et al., 2019)	• Resolution: 0.05°, temporally static, processed based on data from mixed years • 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) • Lake data source: 3 km (90 arc seconds) lake data (Kourzeneva 2009, 2010) • Glacier data source: glacier and ice sheet vector data (Arendt et al. 2012; Rastner et al. 2012) • Urban data source: 1 km urban data (Jackson et al., 2010)	• Resolution: 0.05°, year 2005 • Data source: 500 m, yearly, MODIS collection 5 (Friedl et al., 2010)
Vegetation	LAI, SAI	• Resolution: 1 km, monthly, 2001-2020 • Data source: 450 m, 8-day, reprocessed MODIS collection 6 LAI (Yuan et al., 2011; Friedl et al., 2019)	• Resolution: 0.5°, 12 months • Data source: 1 km, 8-day, MODIS collection 4 LAI (Myneni et al., 2002)	• Resolution: 0.05°, year 2005 • Data source: 450 m, 8-day, reprocessed MODIS collection 5 LAI (Yuan et al., 2011; Friedl et al., 2010)
	Canopy top height, Canopy bottom height	• Resolution: 1 km, temporally static • Data source: 10 m, vegetation canopy height (Lang et al., 2023)	• Resolution: 0.5° or PFT specified value, temporally static • Tree PFT data source: 1 km, forest canopy height derived using 2005 GLAS aboard ICESat data (Simard et al., 2011); • Short vegetation data source: PFT specific values (Bonan et al., 2002)	--
Soil	Percent sand, Percent clay	• Resolution: 1 km, temporally static • Data source: 250 m, Soilgrid v2 (Poggio et al., 2021)	• Resolution: 10 km (0.083°), temporally static • Data source: IGBP soil data of 4931 mapping units (IGBP, 2000)	--
	Soil organic matter			
Topography	Elevation, Slope, Standard deviation of elevation	• Resolution: 1 km, temporally static • Data source: 90 m, MERIT Hydro elevation (Yamazaki et al., 2019)	• Resolution: merge of 1 km and 10 arc minutes, temporally static • Data source: global most regions are based on USGS HYDRO1k (Verdin and Greenlee 1996); but 10 arc minute data is used over Greenland and Antarctica.	--
	Aspect, Sky view factor, Terrain configuration factor	• Resolution: 1 km, temporally static • Data source: 90 m, MERIT Hydro elevation (Yamazaki et al., 2019)	--	

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

149
150

9

2.1 LULC-related parameters

In this study, the MODIS MCD12Q1 version 6 (Friedl et al., 2022) was employed to ascertain the Plant Functional Types (PFT) as well as other non-vegetative land categories at a spatial resolution of 1 km spanning the years 2001 to 2020. The integrity of the MODIS land cover product has been established through a 10-fold cross-validation accuracy assessment using the Terrestrial Ecosystem Parameterization database (Sulla-Menashe et al., 2019). This land cover product offers richer and more flexible land cover data with higher accuracy and substantially less year-to-year stochastic variation in classification results (Sulla-Menashe et al., 2019). Being the sole operational global land cover product available with annual intervals, it addresses a significant gap in the realm of global change research.

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, needleleaf deciduous trees, broadleaf evergreen trees, broadleaf deciduous trees, shrub, grass, and crop. 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).

173 • Grass was reclassified as C3 and C4 grass using the approach presented by Still et al. (2003),

174   with the assistance of monthly LAI (processed in section 2.2.1) and meteorological

175   variables from WorldClim V1.

176 • The "non-vegetated land" was classified as barren soil class.

177 • The "permanent snow and ice" was assigned as the glacier land unit.

178 • Global lakes were identified based on the classification of "water bodies" over the global

179   land, constrained using the global land mask obtained from Natural Earth

180   (https://www.naturalearthdata.com/).

181 • The urban land unit was determined based on the MODIS "urban and built-up"

182   classification. These urban grids were further classified into three urban classes, namely,

183   tall building district (TBD), high density (HD), and medium density (MD), based on

184   Jackson et al. (2010; hereinafter referred to as J2010). J2010 generated global urban extent

185   maps for the TBD, HD, and MD classes at a spatial resolution of 1 km, based on rules of

186   building   height   and   vegetation   coverage   fraction

187   (https://gdex.ucar.edu/dataset/188a_oleson/file.html). However, the J2010 dataset is

188   temporally static and cannot reflect changes in urban boundaries over time. Therefore, we

189   reclassified the yearly MODIS urban land class as TBD, HD, and MD based on the J2010

190   dataset using the nearest neighbor sampling method for each year.

191 After determining the distribution of 15 PFTs, bare soil, lake, glacier, and urban land, any

192 remaining 1 km grids were assigned as ocean (Table S1). It should be noted that the wetland land

193 unit was not explicitly classified in this study. This is because, instead of treating wetlands as an

194 individual land unit, many LSMs (e.g., ELM2 and CLM5) integrate wetland functioning processes

195   prognostically within other land units where a surface water storage component is implemented to

196   represent wetland functioning.

197

198   **2.2 Vegetation-related parameters**

199   **2.2.1 Monthly LAI and SAI**

200   The monthly LAI parameters were obtained from Beijing Normal University (BNU_LAI; Yuan et

201   al., 2011). BNU_LAI, an enhanced version of the MODIS LAI product, has been subjected to

202   thorough quality control, incorporating multiple algorithms for improved accuracy (Yuan et al.,

203   2011). Its validation involved an extensive array of LAI reference maps and employed the bottom-

204   up approach advocated by the CEOS Land Product Validation sub-group (Morisette et al., 2006).

205   Compared to the original MODIS LAI, the BNU_LAI dataset exhibits superior performance, along

206   with enhanced spatiotemporal continuity and consistency. The 8-day BNU_LAI product at a

207   resolution of 15 seconds (~450 m) over 2001–2020 was downloaded from

208   http://globalchange.bnu.edu.cn/research/laiv061. Subsequently, the data were resampled to a

209   resolution of 1 km using an area-weighted average method and averaged temporally for each

210   month. The processed monthly LAI at 1 km resolution was subsequently assigned to each of the

211   15 PFTs described above at each grid. The monthly SAI was then calculated based on the

212   processed monthly LAI using the methods and PFT parameters described in Zeng et al. (2002).

213

214   **2.2.2 Vegetation canopy height**

215   We leveraged a global vegetation canopy height dataset sourced from Lang et al. (2023). This

216   dataset, derived using a probabilistic deep learning model, fuses Sentinel-2 images with the Global

217   Ecosystem Dynamics Investigation (GEDI) to retrieve canopy height. It stands out as the inaugural

218   global canopy height dataset offering consistent, wall-to-wall coverage at a 10 m spatial resolution

219   across all vegetation types. Assessments using hold-out GEDI reference data and comparisons

220   with independent airborne LiDAR data demonstrate that the approach outlined by Lang et al. (2023)

221   produces a meticulously quality-controlled, state-of-the-art global map product, accompanied by

222   quantitative uncertainty estimates. The canopy height served as the canopy top height parameter.

223   Canopy bottom height was calculated by multiplying PFT-based ratios derived from the ratio of

224   ELM2's (same as CLM5) canopy top and bottom heights for different PFTs (Table S2).

225

226   **2.3 Soil-related parameters**

227   We obtained the Soilgrid v2 data with an original resolution of 250 m (Poggio et al., 2021) to

228   prepare soil properties. Soilgrid is generated using machine learning based on multiple data

229   sources of soil profiles and remote sensing data (Hengl et al., 2017). The soil product underwent

230   rigorous quantitative evaluation using a cross-validation method, which ensures alignment with

231   established pedo-landscape features and provides spatial uncertainty to guide product users

232   (Poggio et al., 2021). Soilgrid v2 provides percent clay, percent sand, and soil organic matter for

233   six standard soil layers: 0–5 cm, 5–15 cm, 15–30 cm, 30–60 cm, 60–100 cm, and 100–200 cm.

234   The original SoilGrid version 2 data obtained from GEE were processed at 1 km resolution with

235   multiple layers using an area-weighted average method. To facilitate the demonstration, we

236   restructured the six soil layers vertically into ELM2's ten effective soil layers (0–1.8 cm, 1.8–4.5

237   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–

238   229.6 cm, and 229.6–380.2 cm) using the nearest neighboring method. It should be noted that the

239   lake module in ELM2 and CLM5 requires soil properties, but the Soilgrid v2 data may not provide

240   coverage over water surfaces. To address this, we utilized the nearest neighbor sampling method

241   to map the 1 km soil properties onto the terrestrial water surface.

242

243   **2.4 Topography-related parameters**

244   We employed the digital elevation from the Multi-Error-Removed Improved-Terrain DEM

245   (MERIT DEM, Yamazaki et al., 2019) to obtain topography-related parameters. The MERIT DEM

246   provides globally consistent elevation data at 90 m resolution, distinguished by its exceptional

247   vertical accuracy. This accuracy was rigorously validated against ICESat's lowest elevations in

248   both forested and non-forested regions and was further benchmarked using the UK's premium

249   airborne LiDAR DEM (Yamazaki et al., 2019). We first acquired the 1km elevation and standard

250   deviation of elevation using GEE based on the original 90 m elevation. Further, we calculated the

251   slope, aspect, sky view factor, and terrain configuration factor from the 1km elevation using the

252   parallel computing tool developed by Dozier (2022). The sky view factor represents the proportion

253   of visible sky limited by adjacent terrain, and the terrain configuration factor describes the

254   proportion of adjacent terrain which is visible to the ground target. Finally, to drive the

255   parameterization of sub-grid topographical effects on solar radiation (Hao et al., 2022) in ELM2,

256   we calculated the $\sin(slope) \cdot \sin(aspect)$ and $\sin(slope) \cdot \cos(aspect)$ for calculating the

257   local solar incident angle, and two normalized angle-related factors, the sky view factor, and terrain

258   configuration factor by $\cos(slope)$. It is important to note that the standard deviation of elevation

259   calculated in this study is specific to the 1 km resolution simulation. For applications requiring

260   coarser resolutions (e.g., 0.5 degree), the standard deviation should be recalculated directly from

261   the 1 km elevation, rather than averaging from the 1k standard deviation of elevation.

262

2.5 Comparison between new and existing land surface parameters

In this study, since the data sources used to develop the 1k global land surface parameters have already undergone rigorous validation, we do not perform additional evaluations against reference datasets (e.g., observations). Instead, our focus is on comparing the newly developed 1k parameters with those from K2012 and the ELM2/CLM5 default parameters. The K2012 parameters, obtained through personal communication (refer to the data availability section for details). The ELM2/CLM5 default parameters were sourced from the CESM input data repository (https://svn-ccsm-inputdata.cgd.ucar.edu/trunk/inputdata/). Given the different resolutions of these datasets—our new parameters at 1km, K2012 at 0.05 degree, and ELM2/CLM5 defaults with varying resolutions—we adapt our comparison at different resolutions for different variables.

For PFT parameters, we aggregated both the 1k new parameters and the 0.05-degree K2012 data to the 0.5-degree resolution of the ELM2/CLM5 default. For non-vegetated land units (i.e., urban, glacier, and lake), we upscaled the 1k new parameters to a 0.05-degree resolution to align with the ELM2/CLM5 default. It is important to note that the urban parameter in K2012 is only available for the northern hemisphere, due to limitations in data acquisition.

When comparing LAI, we aggregated the 1k new and K2012 LAI to 0.5-degree resolution, matching the ELM2/CLM5 default LAI/SAI resolution. We excluded the comparison of SAI from our analysis due to the limited availability of the global K2012 dataset, from which we only acquired coverage for North America. We have not included a comparison of vegetation canopy height (top and bottom parameters) in our study. This is because the K2012 dataset does not contain these parameters, and the ELM2/CLM5 default parameters in the CESM input data repository provide only tabular values for each PFT, rather than spatially variable canopy heights for tree PFTs.

286   For soil and topography-related parameters, our comparison was limited to the 1k new parameters

287   and the ELM2/CLM5 default, as K2012 does not include these parameters. Specifically, for soil

288   comparisons, we aggregated the new 1k parameters to 0.083° resolution to match the ELM2/CLM5

289   default soil parameters. For topography, given that the ELM2/CLM5 default parameters is a

290   combination of 1k and 10 arc-minute data sources, we simplify the comparison by aggregating

291   both the new 1k parameters and ELM2/CLM5 default to 0.5-degree resolution, including elevation

292   and slope.

293

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 0.5° to drive ELM. The spatial homogeneity of atmospheric forcings within 0.5° grid cell guarantees that the spatial variability of ELM simulated variables (e.g., latent heat) within 0.5° 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 et al. (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, λ_{scale} of 1/60°, 1/40°, 1/30°, 1/24°, 1/20°, and 1/10°, and calculated the spatial standard deviation (σ_{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 σ_{scale} to $\sigma_{1\,km}$. Third, we fitted a $\log\left(\frac{\sigma_{scale}}{\sigma_{1\,km}}\right) \propto \beta \times \log\left(\frac{\lambda_{scale}}{\lambda_{1\,km}}\right)$ relationship, where β is an indicator to quantify data spatial variability persistence

17

317 across scales (Hu et al., 1997). A more negative β indicates a larger dependency of data spatial

318 variability on spatial scales, resulting in a higher information loss, denoted as $\gamma_{scale} =$

319 $(1 - \sigma_{scale}/\sigma_{1\,km}) \times 100\%$. In this study, we focus on information loss at a 12 km scale, denoted

320 as $\gamma_{12\,km}$. For simplicity in subsequent discussion, $\gamma_{12\,km}$ will be referred to as γ in the results

321 section. Given the possibility that β may not demonstrate significant temporal variation (Mälicke

322 et al., 2020), and considering that our scaling analysis is intended for demonstration purposes, our

323 spatial scaling analysis is based on the annual mean of ELM2 simulations.

324 It is crucial to clarify that the upscaled 1 km simulation results in the spatial scaling analysis are

325 not equivalent to the results obtained from a coarse resolution ELM conducted using upscaled

326 parameters. The spatial scaling analysis is intended to emphasize the value of high-resolution

327 modeling in capturing fine-scale spatial variabilities, and to highlight the contributions of high-

328 resolution land surface parameters on the simulated variables.

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

$\lambda_{scale}/\lambda_{1\,km}$	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

330

331 **3.3 Attribution analysis utilizing XML methods**

332 We conducted additional analysis to determine the primary land surface parameters that influence

333 the spatial scaling of ELM simulations. We employed XML methods, specifically the eXtreme

334 Gradient Boosting (XGBoost; Chen and Guestrin, 2016) machine learning algorithm and the game

335 theoretic approach SHapley Additive exPlanations (SHAP; Lundberg and Lee, 2017; Lundberg et

336 al., 2018, 2020). XML methods were utilized to assess the influence of land surface parameters on

337 the spatial variability and information loss of ELM2 simulations across the CONUS. Taking spatial

18

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

3.4 Reference datasets for evaluating ELM simulation

We also performed a comparison of all four ELM-simulated variables against reference datasets. It is important to note that we used the default model parameters and did not perform any calibration (see discussions for details). For reference datasets, soil moisture was obtained from the Global Land Evaporation Amsterdam Model (GLEAM; Martens et al., 2017), latent heat flux data was from the MODIS product (Running et al., 2021), and both ELR and ASR data were processed from the land component of the fifth generation of European ReAnalysis (ERA5_Land; Muñoz-Sabater et al., 2021). For the soil moisture evaluation, we compared the surface layer soil moisture from GLEAM (10 cm depth) with the weighted average of the first four-layer soil moisture from ELM (about 11 cm depth). To ensure comparability, we unified the spatial resolution of both reference datasets and ELM simulations to a 0.5-degree resolution and focused our analysis on the annual mean data for 2014.

4. Results

### 4.1 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.

(a) LAI (m²/m²)

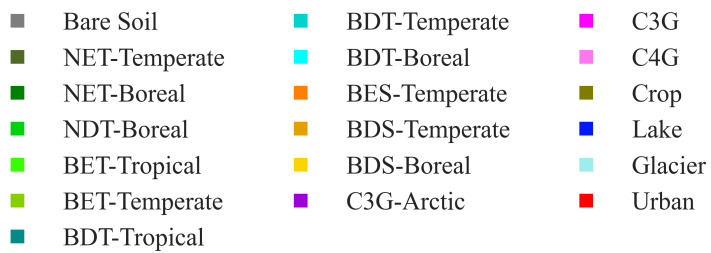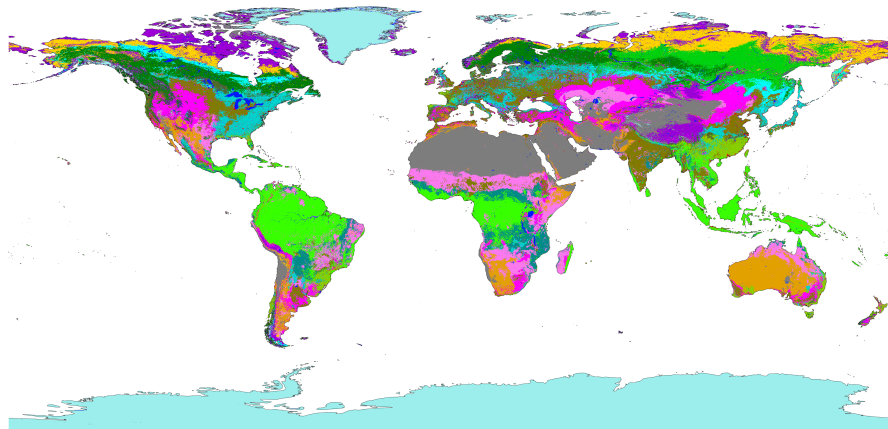(b) R1 (c) R2 (d) R3 (e) R4
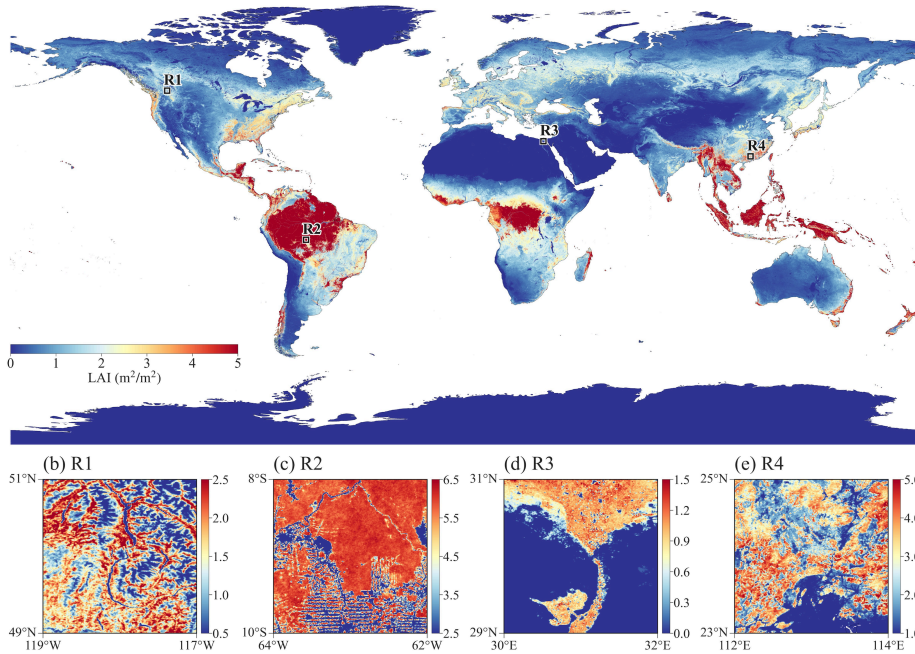
Figure 1. The spatial pattern of LAI (annual mean in 2010) over (a) global land and (b)~(e) four

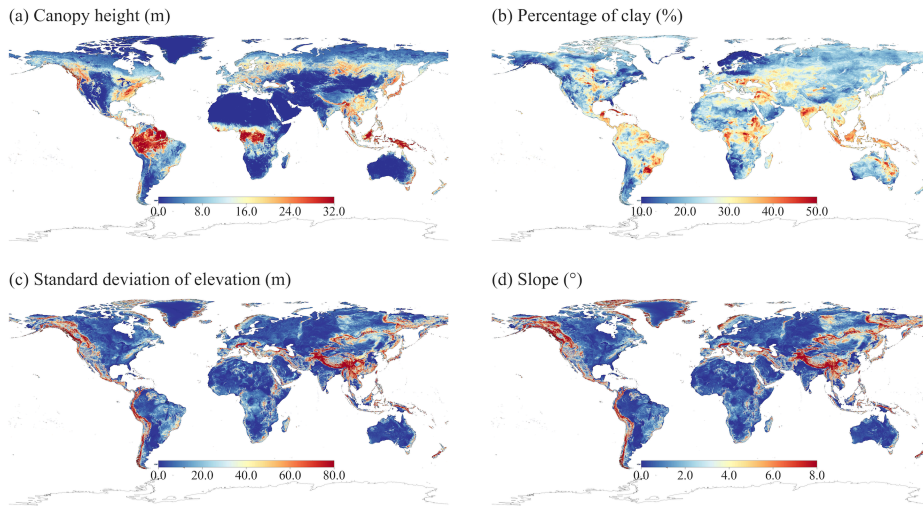subregions R1~R4 within 2-degree boxes marked in (a). Subregions R1~R4 represent

topography, deforestation, irrigations, and urbanization effects on LAI.

■ Bare Soil	■ BDT-Temperate	■ C3G
■ NET-Temperate	■ BDT-Boreal	■ C4G
■ NET-Boreal	■ BES-Temperate	■ Crop
■ NDT-Boreal	■ BDS-Temperate	■ Lake
■ BET-Tropical	■ BDS-Boreal	■ Glacier
■ BET-Temperate	■ C3G-Arctic	■ Urban
■ BDT-Tropical		

387

388   Figure 2. Global LULC distribution in year 2010. PFT abbreviations include: Bare Soil, Needleleaf

389   Evergreen Trees in temperate (NET-Temperate) and boreal (NET-Boreal) regions, Needleleaf

390   Deciduous Trees in boreal regions (NDT-Boreal), Broadleaf Evergreen Trees in tropical (BET-

391   Tropical) and temperate (BET-Temperate) regions, Broadleaf Deciduous Trees in tropical (BDT-

392   Tropical), temperate (BDT-Temperate), and boreal (BDT-Boreal) regions, Broadleaf Evergreen

393   Shrubs in temperate regions (BES-Temperate), Deciduous Shrubs in temperate (BDS-Temperate)

394   and boreal (BDS-Boreal) regions, C3 Grass in arctic (C3G-Arctic) and general (C3G) varieties,

395   C4 Grass (C4G), Crop, Lake, Glacier, and Urban.

(a) Canopy height (m)

(b) Percentage of clay (%)

0.0   8.0   16.0   24.0   32.0

10.0   20.0   30.0   40.0   50.0

(c) Standard deviation of elevation (m)

(d) Slope (°)

0.0   20.0   40.0   60.0   80.0

0.0   2.0   4.0   6.0   8.0

396

397   Figure 3. Demonstration of global 1km datasets (a) Canopy top height, (b) percent clay, (c)

398   standard deviation of elevation, and (d) slope.

399

400   **4.2 Comparison between new and existing land surface parameters**

401   The global distributions of different PFTs show varying degrees of difference when comparing the

402   new parameters with the K2012 and ELM2/CLM5 default parameters (Figure 4 and

403   Supplementary Figures S1 to S16). Predominant types such as bare soil, BET-Tropical tree, C3

404   and C4 grass, and crop are found consistently across all datasets. Notable differences include less

405   bare soil in the new parameters and K2012 compared to ELM2/CLM5 default, especially in high-

406   latitude North America, western US, South Africa, Central Asia, and Central Australia (Figure S1).

407   While the new NDT PFT shows larger coverage in Siberia than K2012 and ELM2/CLM5 (Figure

408   S4), BET-Tropical PFT is more prevalent in the new parameters across Central and South America

409   (Figure S5). BET-Temperate PFT has greater area coverage in southern China in the new

410   parameters (Figure S6). For BDT-Tropical, BDT-Temperate, and BDT-Boreal PFTs, both the new

23

412   and ELM2/CLM5 default parameters surpass K2012 data in coverage (Figures S7 to S9). The

413   coverage of new BDS-Temperate PFT is smaller than K2012 but larger than ELM2/CLM5 default

414   (Figure S11), and the new BDS-Boreal PFT is less extensive in the boreal northern hemisphere

415   compared to both K2012 and ELM2/CLM5 defaults (Figure S12). The C3-Arctic PFT shows

416   larger areas in the new parameters, particularly in northern Canada, with the new C4 grass PFT

417   being similar to that of K2012 and larger than ELM2/CLM5 C4 grass. Crop PFT is less extensive

418   in the new parameters, particularly in Southeastern China, Europe, South America, Africa, and

419   Australia.
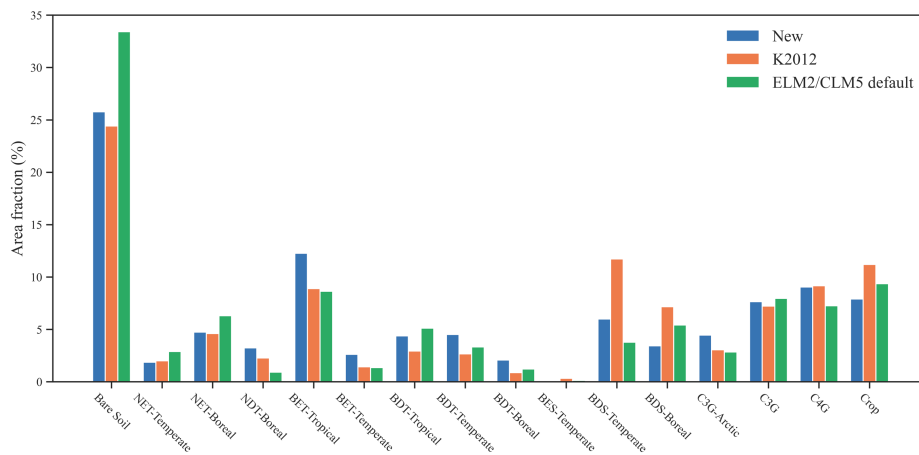
420

421   Figure 4. The global average area fractions of PFTs for three land surface parameter datasets. PFT

422   abbreviations used on the X-axis are displayed in Figure 2.

423

424   The global distributions of non-vegetated land covers of lake, glacier and urban areas vary among

425   the datasets (Figure S17–S19). The new dataset shows slightly less lake coverage than K2012, but

426   both are smaller than ELM2/CLM5 default, particularly in high-latitude North America (Figure

427   S17). Glacier coverage in the new parameter is around 0.7% smaller than K2012, with noticeable

24

428   differences in the Arctic North America, while ELM2/CLM5 default shows more extensive glacier

429   coverage in Antarctica (Figure S18). Regarding urban areas, K2012 has the smallest urban

430   coverage in the Northern Hemisphere compared to both the new dataset and ELM2/CLM5 default

431   (Figure S19). Meanwhile, ELM2/CLM5 default exhibits more expansive urban areas in India and
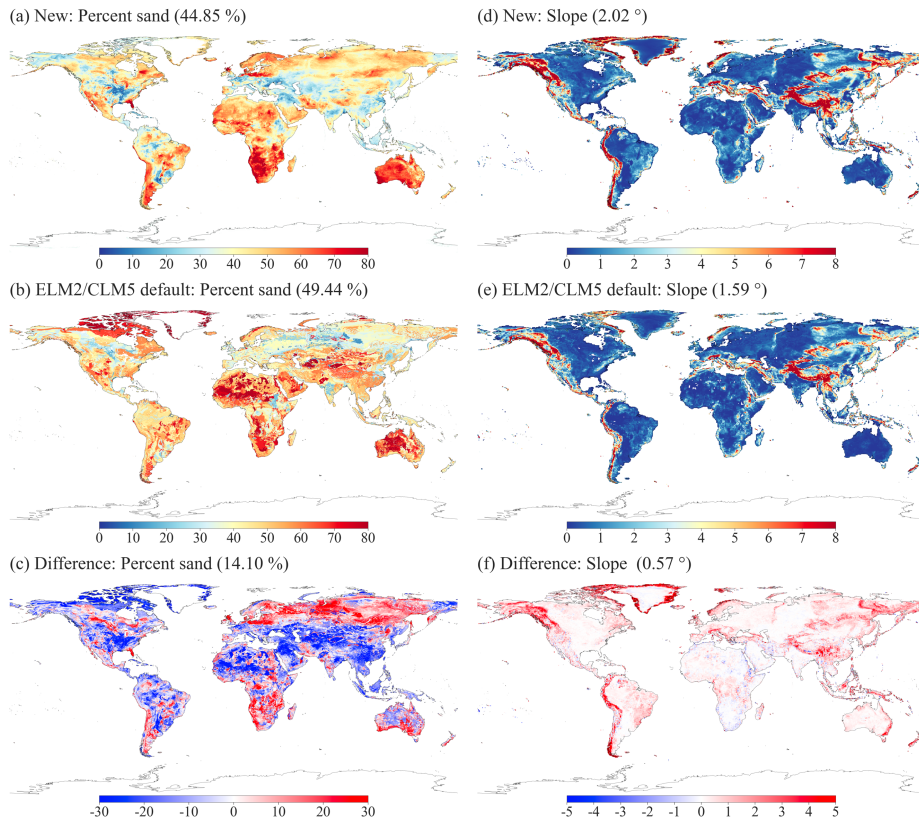
432   China than the new dataset and K2012.

433

434   The global annual mean LAI exhibits similar spatial patterns among the new parameter, K2012,

435   and ELM2/CLM5 (Figure 5). The overall global mean LAI for the new parameter (1.28 m²/m²) is

436   slightly higher than that of K2012 (1.14 m²/m²) and the ELM2/CLM5 default data (1.24 m²/m²).

437   In terms of spatial pattern, the new LAI, relative to K2012 (Figure S20a), shows lower values in

438   the NET-Boreal PFT over the northern hemisphere, but higher values in the BET-Tropical PFT

439   over the tropics. Similarly, compared with the ELM2/CLM5 default LAI (Figure S20b), the new

440   LAI also presents smaller values in both the NET-Boreal and NDT PFTs over the northern

441   hemisphere, but larger values in the BET-Tropical PFT regions.

(a) New: LAI (1.28 m²/m²)

(b) K2012: LAI (1.14 m²/m²)

(c) ELM2/CLM5 default: LAI (1.24 m²/m²)

442
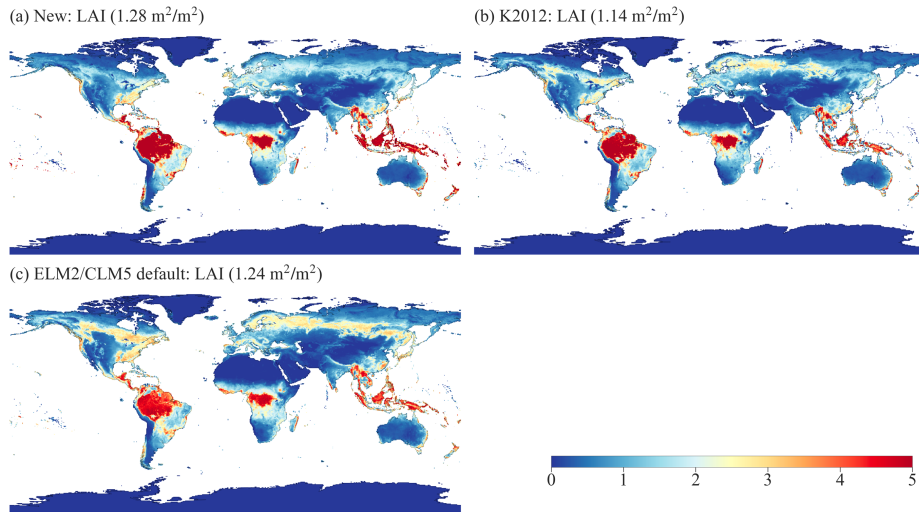
Figure 5. Comparison of global annual mean LAI for (a) new, (b) K2012, and (c) ELM2/CLM5

default parameters. The global average is indicated in the subplot title.

Soil parameters exhibit significant differences between the new and ELM2/CLM5 default datasets

(Figures 6a-bc, S21, and S22). The global mean absolute differences between the new and

ELM2/CLM5 default for percent sand, percent clay, and organic matter are 14.1%, 8.1%, and 30.5

kg/m³, respectively. Generally, the new soil parameters are spatially distributed more smoothly

than those from ELM2/CLM5 with more patchy patterns (Figure 6a vs. 6b). Specifically, the new

percent sand is higher in regions like Europe, Siberia, South Africa, and Southern Australia, but

lower in areas such as the Lower Mississippi River Basin, North Africa, and Central and

Southeastern Asia (Figure 6c). The new percent clay shows larger values in the Western US, North

Africa, Central Asia, and Australia, but smaller values in Alaska and Eastern Europe (Figure S21).

455 For organic matter, the new parameter indicates smaller values in the Northern Hemisphere but

456 larger values in other global regions compared to the ELM2/CLM5 default (Figure S22).

457 Topography-related parameters exhibit broadly similar spatial patterns but with notable

458 differences between the new and ELM2/CLM5 default parameters, as seen in Figures 6d-6f and

459 S23. The new slope parameter generally shows a larger slope relative to the ELM2/CLM5 default,

460 particularly in mountainous regions (Figure 6f). This could be attributed to the new 1 km slope

461 being calculated from a finer 90 m resolution elevation. Differences in elevation between the new

462 and ELM2/CLM5 parameters are more pronounced in areas such as various mountainous regions,

463 Greenland, the Amazon Basin, the Tibetan Plateau, and Australia (Figure S23).

(a) New: Percent sand (44.85 %)

(d) New: Slope (2.02 °)

(b) ELM2/CLM5 default: Percent sand (49.44 %)

(e) ELM2/CLM5 default: Slope (1.59 °)

(c) Difference: Percent sand (14.10 %)

(f) Difference: Slope (0.57 °)

Figure 6. Comparisons of percent sand and slope. (a) new and (b) ELM2/CLM5 default percent sand, along with (c) their difference (new – ELM2/CLM5 default) for percent sand; (d) new, (e) ELM2/CLM5 default, and (f) their difference for slope. The global average is shown in the subplot titles, with the global average of the absolute difference provided for (c) and (f).

4.3 Demonstration 1km simulation over CONUS

ELM simulations at a 1 km resolution display significant spatial heterogeneity over CONUS (Figure 7). The values of SM, LH, ELR, and ASR across CONUS follow approximately normal

28

474   distributions, with averages of 0.3 m³/m³, 39.0 W/m², 371.7 W/m², 156.7 W/m², respectively (as

475   shown in the histogram plots in Figure 7). SM shows drier conditions over the West and Southwest

476   and wetter conditions over the Midwest, Corn Belt, Mississippi River basin, and Northeast (Figure

477   7a). LH shows high values over the central and southeast, and lower values over the west and

478   southwest (Figure 7b). The ELR generally shows higher values over regions with high surface

479   temperature in the south (Figure 7c). The ASR shows higher values over the southwestern regions

480   determined by incoming solar radiation and albedo (Figure 7d). Despite the high-resolution

481   heterogeneity shown at 1 km resolution, we can still see the spatial patterns distinguished at coarse

482   resolution, i.e., 0.5º × 0.5º. These coarser footprints are from the GSWP3 atmospheric forcing with

483   0.5º resolution. As concluded by Li et al. (2022), atmospheric forcing is one primary heterogeneity

484   source for land surface modeling. Therefore, k-scale atmospheric forcing needs to be developed to

485   further advance k-scale offline land surface modeling.
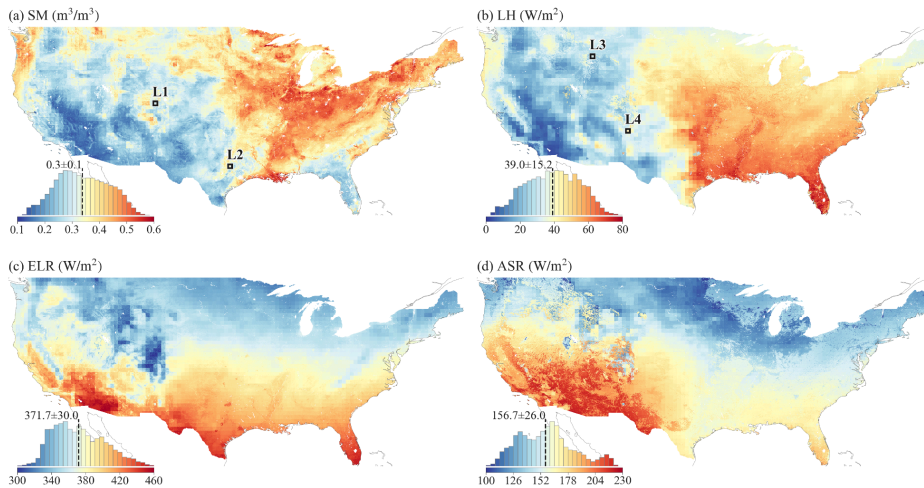
486

487   Figure 7. The annual mean of 1 km simulations of (a) SM, (b) LH, (c) ELR, and (d) ASR over

488   CONUS. The 0.5° × 0.5° boxes marked as L1, L2, L3, and L4 in (a) and (b) are selected to

490 demonstrate the spatial scaling analysis. The inserted histogram plot illustrates the distribution of

491 ELM2 simulations.

492

493 **4.4 Demonstration of spatial scaling across scales**

494 We next demonstrate the relationships between spatial variabilities and spatial scales for SM and

495 LH. Four locations (in Figures 4a and 4b) are specifically chosen to showcase varying levels of

496 spatial information loss: L1 and L3 demonstrate a relatively large loss for SM and LH, respectively,

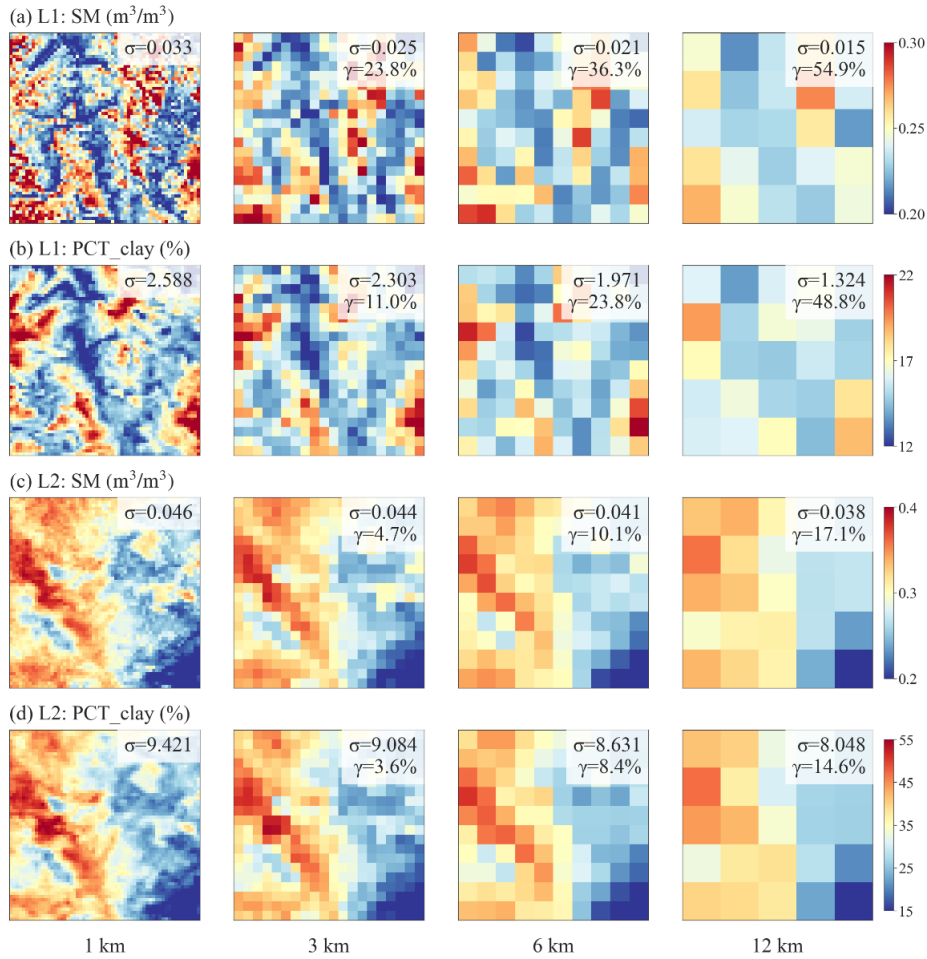497 while L2 and L4 represent a relatively small loss for SM and LH, respectively.

498 At location L1 (Figure 8a), when the 1 km simulation is upscaled to coarser resolutions (i.e., larger

499 spatial scale ratios), the spatial variability of SM decreases, resulting in a negative slope of β. As

500 shown in Figure 9a, compared to the original 1 km resolution, the information loss γ reaches up to

501 54.9% at the 12 km spatial scale. The spatial pattern of SM is consistent with the spatial pattern of

502 percent clay (Figures 6a vs. 6b and 6c vs. 6d), indicating that soil texture contributes significantly

503 to the spatial variability of SM. However, SM has a more negative β than the percent clay ($\beta = -$

504 0.28 vs. –0.19 at L1, as shown in Figure 8a), suggesting that SM variability is amplified likely by

505 other processes that are also influenced by soil texture. In contrast to location L1, location L2

506 exhibits less negative β values for both SM and percent clay, suggesting that their spatial

507 variabilities exhibit less scale dependence (Figures 5a, 6c, and 6d). Both SM and percent clay at

508 location L2 approximately maintain their spatial patterns of high values in the west and low values

509 in the east across spatial scales (Figures 6c and 6d).

510 For LH, there is a more negative β value at location L3 than at location L4 ($\beta = -0.27$ at L3 vs. –

511 0.08 at L4, as shown in Figure 8b), which indicates a larger decrease of spatial variability across

512 spatial scales and lower variability persistence at location L3 than location L4 (Figure 10). The

spatial pattern of LH is consistent with the spatial pattern of LAI (Figures 7a vs. 7b and 7c vs. 7d) at different spatial scales, suggesting that vegetation plays a significant role in the spatial variability of LH. Similar to comparison between SM and soil texture, LH has a more negative β than LAI (Figure 8b).
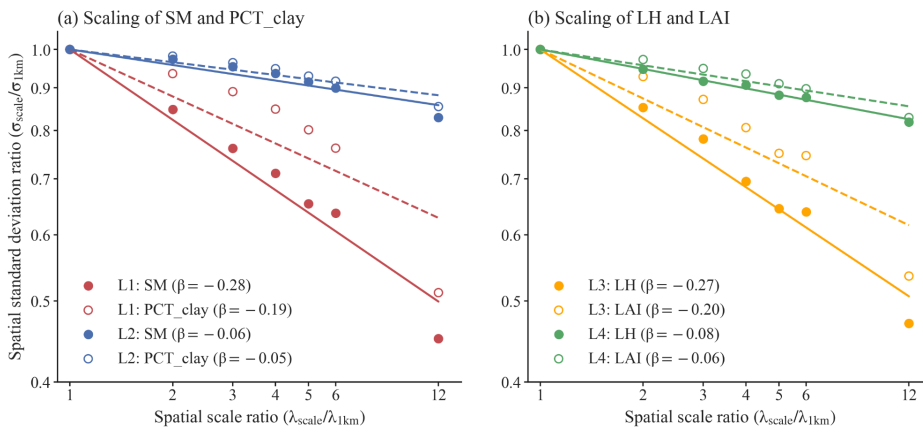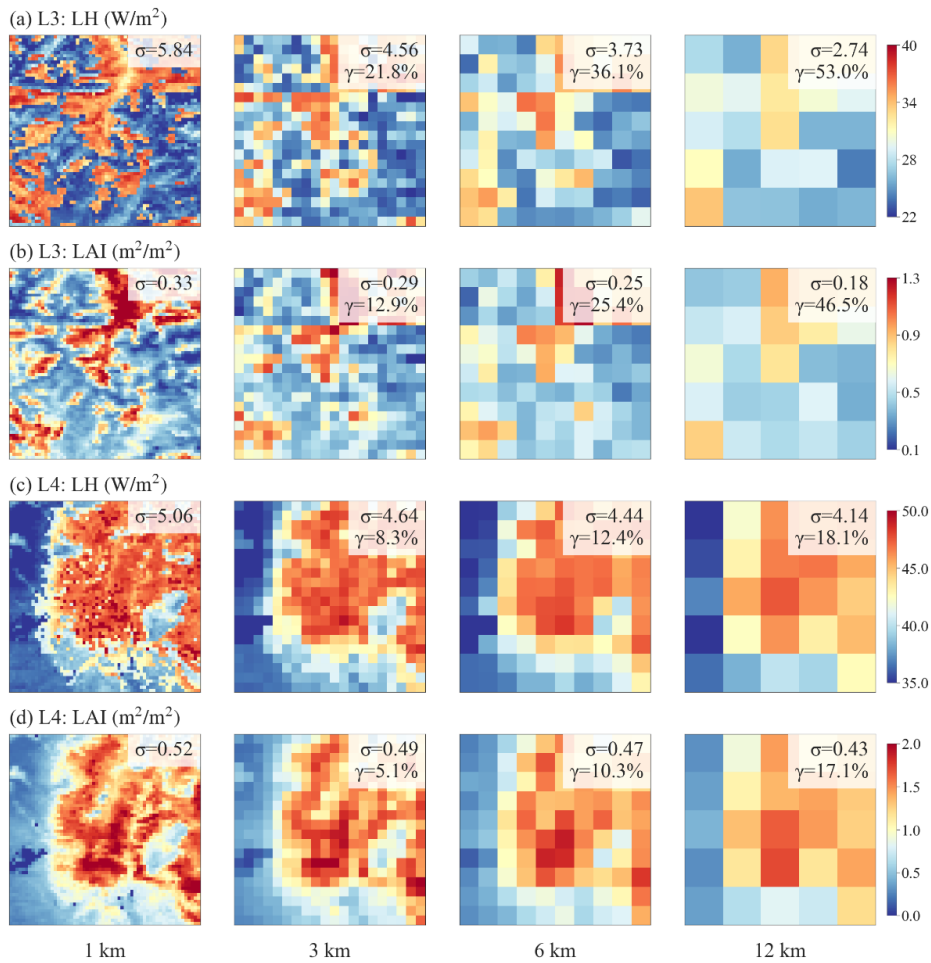

Figure 8. The scaling of spatial variabilities for (a) SM and percent clay, and (b) LH and LAI. Both the x-axis and y-axis are in logarithmic scale. The slope of the linear regression line, β, quantifies the strength of the negative relationship between spatial scale and spatial variability. A more negative β value indicates a higher spatial-scale dependency and increased information loss at coarser spatial scales. Four $0.5° \times 0.5°$ boxes (displayed in Figure 7), 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).

(a) L1: SM (m³/m³)

σ=0.033   σ=0.025 γ=23.8%   σ=0.021 γ=36.3%   σ=0.015 γ=54.9%

(b) L1: PCT_clay (%)

σ=2.588   σ=2.303 γ=11.0%   σ=1.971 γ=23.8%   σ=1.324 γ=48.8%

(c) L2: SM (m³/m³)

σ=0.046   σ=0.044 γ=4.7%   σ=0.041 γ=10.1%   σ=0.038 γ=17.1%

(d) L2: PCT_clay (%)

σ=9.421   σ=9.084 γ=3.6%   σ=8.631 γ=8.4%   σ=8.048 γ=14.6%

1 km   3 km   6 km   12 km

526

527   Figure 9. Comparison of SM and percent clay across spatial scales at locations L1 and L2

528   highlighted in Figure 7. Each subplot displays the spatial patterns of SM or percent clay within a

529   0.5° × 0.5° box, with the σ and γ presented in the legend.

(a) L3: LH (W/m²)

σ=5.84   σ=4.56 γ=21.8%   σ=3.73 γ=36.1%   σ=2.74 γ=53.0%

(b) L3: LAI (m²/m²)

σ=0.33   σ=0.29 γ=12.9%   σ=0.25 γ=25.4%   σ=0.18 γ=46.5%

(c) L4: LH (W/m²)

σ=5.06   σ=4.64 γ=8.3%   σ=4.44 γ=12.4%   σ=4.14 γ=18.1%

(d) L4: LAI (m²/m²)

σ=0.52   σ=0.49 γ=5.1%   σ=0.47 γ=10.3%   σ=0.43 γ=17.1%

1 km   3 km   6 km   12 km

530

531   Figure 10. Similar to Figure 9, but for LH and LAI at locations L3 and L4.

33

532 **4.5 The spatial variability of water and energy simulations and their drivers**

533 We quantified the spatial variability simulated at 1 km resolution using σ within each 0.5º × 0.5º

534 box across CONUS. Four ML models were built to explore the spatial relationships between σ and

535 its potential drivers including σ of the land surface parameters and the temperature and

536 precipitation averaged over the grid box. Overall, the ML models performed well in predicting the

537 σ of the simulated variables, with small root mean square error (RMSE) and large R^2 (see Figure

538 S24). SM shows larger spatial variability in the US Southern Coastal Plain, lower Mississippi

539 River, Northeast, Southeast, and regions around the Great Lake (Figure 11a), which is roughly

540 consistent with the spatial heterogeneity of the high-resolution SM simulation in Vergopolan et al.

541 (2022). Based on the SHAP method, the spatial variability of SM across CONUS is driven by

542 various factors, mainly including the spatial variabilities of percent sand and percent clay, mean

543 precipitation, the σ and μ of soil organic matter, the σ of canopy height, and mean temperature

544 (Figure 11b). Mean precipitation and temperature reflect climate conditions (Figure S26), which

545 are related to the water supply and water demand of soil water content. The spatial heterogeneity

546 of soil properties, such as texture and organic matter content, affects soil hydraulic properties and

547 generate more spatially variable soil water content. Vegetation characteristics, such as canopy

548 height and LAI, could influence SM spatial variability through their effect on roughness length

549 and rooting depth.

550 The spatial variability of LH is large in the southeastern, central, and western mountainous regions

551 of the US (Figure 11c). Vegetation properties and climate conditions mainly drive the variability

552 of LH (Figure 11d). The μ and σ of LAI can affect transpiration and soil evaporation, while canopy

553 height can influence surface roughness length and, in turn, evapotranspiration. Mean precipitation

34

555   and temperature reflect the overall climate conditions related to the water and energy available for

556   latent heat.

557   ELR and ASR exhibit large spatial variability mainly over the western US, with ASR additionally

558   showing significant spatial variability across the Northern US (Figures 8e and 8g). This variability

559   is primarily driven by climate conditions such as mean precipitation and temperature, topographic

560   features such as standard deviation of elevation and slope, and vegetation properties including LAI

561   and canopy height (Figures 8f and 8h). These factors are related to the radiation input and surface

562   properties, such as albedo and roughness length, which impact the energy cycles and availability

563   of ELR and ASR.

(a) Spatial pattern of σ_{SM} (m^3/m^3)

(b) Drivers of σ_{SM}

σ_{PCT_sand} 17.0
σ_{PCT_clay} 17.0
$\mu_{precipitation}$ 16.0
$\mu_{organic\_matter}$ 12.1
σ_{canopy_height} 6.1
$\mu_{temperature}$ 5.2
$\sigma_{organic\_matter}$ 4.5
σ_{LAI} 2.3

Positive correlation
Negative correlation

Relative importance (%)

(c) Spatial pattern of σ_{LH} (W/m^2)

(d) Drivers of σ_{LH}

μ_{LAI} 19.5
σ_{LAI} 18.3
$\mu_{precipitation}$ 12.6
$\mu_{temperature}$ 10.0
μ_{canopy_height} 8.3
σ_{canopy_height} 5.8
$\mu_{STD\_elevation}$ 2.7
$\mu_{organic\_matter}$ 2.4

Relative importance (%)

(e) Spatial pattern of σ_{ELR} (W/m^2)

(f) Drivers of σ_{ELR}

$\mu_{precipitation}$ 13.9
μ_{canopy_height} 10.6
σ_{canopy_height} 9.4
$\mu_{organic\_matter}$ 9.1
$\sigma_{STD\_elevation}$ 8.4
σ_{LAI} 8.2
μ_{LAI} 6.4
σ_{slope} 5.9

Relative importance (%)

(g) Spatial pattern of σ_{ASR} (W/m^2)

(h) Drivers of σ_{ASR}

σ_{canopy_height} 15.3
$\sigma_{STD\_elevation}$ 14.6
$\mu_{temperature}$ 11.9
μ_{LAI} 9.9
$\mu_{organic\_matter}$ 6.4
σ_{LAI} 5.6
μ_{canopy_height} 4.6
$\mu_{precipitation}$ 4.2

Relative importance (%)

564

Figure 11. The spatial variability over each 0.5º × 0.5º 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

36

567   histogram plot illustrates the probability distribution of the spatial variability across CONUS. The

568   relative importance of each variable in determining the spatial variability is calculated as the ratio

569   of the mean |SHAP value| of the variable to the sum of the mean |SHAP value| of all variables.

570   Therefore, the sum of the relative importance of all variables is 100%.

571

572   **4.6 The information loss of water and energy simulations and their drivers**

573   We also evaluated the information loss in simulations when upscaling from 1 km to 12 km

574   resolution and analyzed the drivers of their spatial patterns over CONUS. Four ML models were

575   built to explore the relationships between the γ of the simulations and its drivers including the γ of

576   the land surface parameters and the mean temperature and precipitation averaged over the 0.5º ×

577   0.5º box. These ML models performed well in predicting the simulations' γ, with small RMSE and

578   large R^2 (Figure S25).

579   Significant information loss ranging from 31% to 54% with maximum values exceeding 90% is

580   observed for SM, LH, ELR, and ASR simulations (Figure 12). Their spatial patterns and drivers

581   show distinct variations. γ_{SM} is primarily driven by the information loss of percent clay and sand,

582   mean soil organic matter, and mean temperature, which affects the soil hydraulic properties and

583   soil water balance (Figures 9a and 9b). γ_{LH} displays high values in the eastern US and low values

584   in the western US (Figure 12c). It is primarily contributed by the information loss of vegetation

585   properties such as LAI and canopy height, and mean LAI, which influences the partitioning of LH

586   and sensible heat, and the partitioning of transpiration and evaporation (Figure 12d). γ_{ELR} exhibits

587   high values in the central and eastern US, particularly in the northeastern US, while γ_{ASR} has high

588   values almost all over the US, especially in the eastern regions (Figures 9e and 9g). γ_{ELR} and γ_{ASR}

589   are largely driven by vegetation properties such as LAI and canopy height, which are associated

Deleted: 3

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

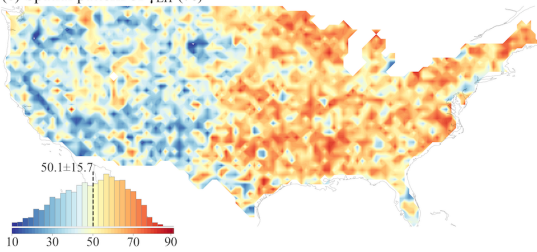592   standard deviation of elevation and slope also slightly contribute to γ_{ASR}.
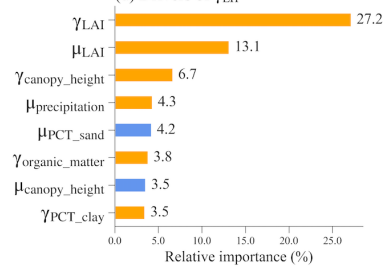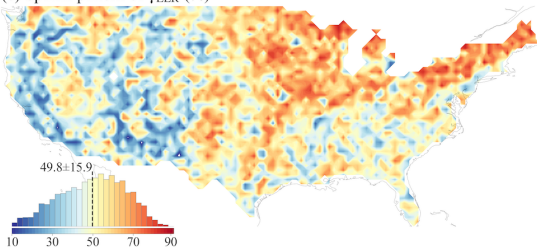
(a) Spatial pattern of γ_{SM} (%)

31.3±13.4

(b) Drivers of γ_{SM}

- γ_{PCT_clay} — 26.3
- γ_{PCT_sand} — 16.8
- $\mu_{temperature}$ — 10.5
- $\mu_{organic\_matter}$ — 6.1
- μ_{PCT_sand} — 4.3
- μ_{LAI} — 3.8
- γ_{canopy_height} — 3.2
- $\mu_{precipitation}$ — 2.8

Positive correlation
Negative correlation

Relative importance (%)

(c) Spatial pattern of γ_{LH} (%)

50.1±15.7

(d) Drivers of γ_{LH}

- γ_{LAI} — 27.2
- μ_{LAI} — 13.1
- γ_{canopy_height} — 6.7
- $\mu_{precipitation}$ — 4.3
- μ_{PCT_sand} — 4.2
- $\gamma_{organic\_matter}$ — 3.8
- μ_{canopy_height} — 3.5
- γ_{PCT_clay} — 3.5

Relative importance (%)

(e) Spatial pattern of γ_{ELR} (%)

49.8±15.9

(f) Drivers of γ_{ELR}

- γ_{LAI} — 28.3
- γ_{canopy_height} — 17.6
- $\mu_{temperature}$ — 11.7
- μ_{LAI} — 5.1
- $\mu_{organic\_matter}$ — 4.5
- μ_{canopy_height} — 4.1
- $\mu_{precipitation}$ — 3.6
- μ_{PCT_sand} — 2.8

Relative importance (%)

(g) Spatial pattern of γ_{ASR} (%)

54.1±14.8

(h) Drivers of γ_{ASR}

- μ_{LAI} — 21.5
- γ_{LAI} — 17.5
- γ_{canopy_height} — 8.8
- $\mu_{temperature}$ — 4.7
- μ_{canopy_height} — 4.4
- $\mu_{STD\_elevation}$ — 4.1
- $\mu_{precipitation}$ — 3.5
- γ_{slope} — 3.0

Relative importance (%)

593

594   Figure 12. Same to Figure 11 but for information loss.

4.7 Comparison of ELM simulation against reference data

The average spatial biases between ELM and reference datasets across CONUS are relatively small, with SM bias at -0.01 m³/m³, LH bias at 1.8 W/m², ELR bias at -3.8 W/m², and ASR bias at 1.1 W/m² (Figure 13 and Figure S27). The correlation coefficient (R^2) between ELM and reference datasets was relatively high at 0.60 (for SM), 0.70 (for LH), 0.96 (for ELR), and 0.90 (for ASR). However, the spatial distribution of these biases exhibits variability, with some areas showing more pronounced biases than others. Specifically, in comparison with GLEAM SM, ELM tends to underestimate SM in the southeastern Texas and across the eastern and southeastern CONUS, while it overestimates SM in the western, central, and southwestern CONUS, including the central eastern US which are primarily agricultural areas. For LH, ELM simulates higher values than the MODIS LH dataset in the western and central US and Florida, but lower values in regions such as the eastern and northeastern CONUS, the western US coastal areas, and the Pacific Northwest. Regarding radiation variables, ELM generally underestimates ELR across nearly all of CONUS and tends to overestimate ASR, particularly in the southwestern, southern, eastern, northeastern, and northern regions of CONUS.

(a) Difference: SM (m^3/m^3); $R^2 = 0.60$

(b) Difference: LH (W/m^2); $R^2 = 0.70$

(c) Difference: ELR (W/m^2); $R^2 = 0.96$

(d) Difference: ASR (W/m^2); $R^2 = 0.90$
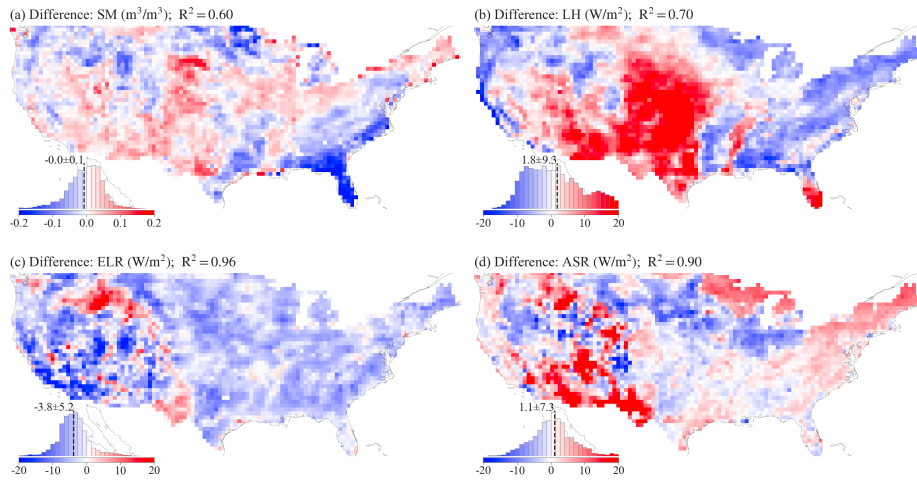
Figure 13. Annual mean bias between ELM-simulated variables and reference datasets over CONUS: (a) SM, (b) LH, (c) ELR, and (d) ASR. The negative values indicate lower ELM values compared to the reference data. The inserted histogram plot illustrates the distribution of grid values. For spatial patterns of the reference datasets, refer to Figure S27. The correlation coefficient (R^2) between the ELM simulation and the reference dataset is calculated and displayed in the title of each subplot."

41

5. Discussion

The development of new 1 km land surface parameter datasets in this study marks a substantial improvement over commonly used land surface parameters such as CLM5 and K2012, leveraging the latest high-resolution data sources with rigorous validation, including MODIS PFTs, enhanced LAI and canopy height, soil properties, and topography factors. When compared with K2012 and ELM2/CLM5 default datasets, the new 1k parameters exhibit notable differences, suggesting potential improvement due to the use of more advanced data sources. Distinct features of the new parameters include a reduction in bare soil compared to ELM2/CLM5, especially in regions like North America and Central Asia, and diverse coverage of specific PFTs such as NDT and BET-Tropical in areas like Siberia and South America. The LAI of the new parameters diverges from K2012 and ELM2/CLM5, showing lower values in NET-Boreal PFT of the northern hemisphere but higher BET-Tropical PFT in the tropics. The soil parameters, particularly in regions like Europe, Central Asia, and the Western US, show significant differences between the new and ELM2/CLM5 defaults. Moreover, the new parameters indicate larger slopes in mountainous regions and more distinct elevation differences in areas such as Greenland and the Tibetan Plateau compared to ELM2/CLM5. These differences potentially highlight enhanced accuracy and sophistication of the new 1k parameters. Their enhanced resolution and rigorous validation suggest a substantial capacity to improve ESMs modeling. Additionally, the richness of multi-year data for LULC, LAI, and SAI in these datasets is especially valuable for examining land use and cover changes, urbanization trends, deforestation impacts, and agricultural transformations.

The new 1 km land surface parameters can improve k-scale offline LSMs modeling by better capturing spatial surface heterogeneity. As evidenced by the 1 km ELM simulation over CONUS,

42

soil properties, vegetation properties, and topographic factors contribute a lot to the spatial heterogeneities of ELM water and energy simulations. Upscaling 1 km to a coarser 12 km resolution, we observe significant spatial information loss, with SM experiencing an average loss of 31%, and LH, ELR, and ASR experiencing around 50% information loss on average (Figure 12). This conclusion is in line with the results of Vergopolan et al. (2022), which showed a substantial loss of spatial information in soil moisture when upscaling from 30 m to 1 km resolution, with an average loss of approximately 48% and up to 80% over the CONUS region. The XML analysis reveals that the spatial variability and information loss of ELM2 simulations are influenced by the spatial variability and information loss of the different variables of land surface parameters, as well as the mean precipitation and temperature (Figures 11 and 12). Our findings highlight the critical role of land surface parameters in contributing to the spatial variability of water and energy in land surface simulations, showcasing the value of the developed high-resolution datasets. Another implementation example where our 1 km land surface parameters can be beneficial is in hillslope-scale simulations, which are fundamental for organizing water, energy, and biogeochemical processes (Fan et al., 2019). Krakauer et al. (2014) have highlighted the significance of between-cell groundwater flow, which becomes comparable in magnitude to recharge at grid spacings smaller than 10 km. Advancements have been made in ESMs to address hillslope-scale processes, including the representation of intra-hillslope lateral subsurface flow within grid cells in CLM5 (Swenson et al., 2019), the development of explicit lateral flow processes between grid cells (Qiu et al., 2023), and the incorporation of topographic radiation effects within and between grid cells (Hao et al., 2021). Another notable example is the integrated hydrology-land surface model ParFlow-CLM, which incorporates three-dimensional groundwater flow, two-dimensional overland flow, and land surface exchange processes (Maxwell, 2013).

663   ParFlow-CLM has demonstrated remarkable reliability in reproducing hydrologic processes, such

664   as its simulations at 3 km resolution for pan-European and 1 km resolution for CONUS (Naz et al.,

665   2023; O'Neill et al., 2021). More recently, Fang et al. (2022) coupled ParFlow with ELM and the

666   Functionally Assembled Terrestrial Ecosystem Simulator (FATES) to simulate carbon-hydrology

667   interactions at hillslope scale. By incorporating our 1 km datasets and leveraging these

668   advancements, we can improve simulations of hillslope-scale processes and enhance our

669   understanding of water and energy dynamics within ESMs.

670

671   Additionally, the new land surface parameters are also a timely resource for supporting the

672   emerging need for k-scale Earth system modeling, particularly in improving land-atmosphere

673   interaction processes. Representing the impact of spatial heterogeneity on land-atmosphere

674   interaction processes is a major challenge in Earth system modeling. Taking E3SM as an example,

675   researchers have proposed three key approaches to enhance spatial heterogeneity representation to

676   address this challenge. In line with these approaches, our newly developed 1 km land surface

677   parameters offer promising opportunities for improving land-atmosphere coupling within ESMs.

678   The first approach to enhance the representation of spatial heterogeneity is to directly conduct

679   simulations at high resolution. For instance, the Simple Cloud-Resolving E3SM Atmosphere

680   Model (SCREAM) has been used to perform global simulations at 3.25 km (Caldwell et al., 2021),

681   although the land surface parameters were based on coarser resolution datasets. By utilizing the

682   new 1 km land surface parameters, we can enhance the representation of land surface heterogeneity

683   within the ELM component of SCREAM, potentially improving modeling of land–atmosphere

684   coupling. The second and third approaches focus on improving the representation of land surface

685   heterogeneity within ESMs run at a coarse resolution while accounting for subgrid heterogeneity

44

686   in two different ways. In the second approach, the Cloud Layers Unified By Binormals (CLUBB)

687   has been implemented in E3SM Atmosphere Model (EAM) version 1 (Rasch et al., 2019;

688   Bogenschutz et al., 2013), to better account for subgrid atmospheric heterogeneity of turbulent

689   mixing, shallow convection, and cloud macrophysics. Recently, Huang et al. (2022) developed a

690   novel land-atmosphere coupling scheme in EAM that enables the communication of subgrid land

691   surface heterogeneity information to the atmosphere model with CLUBB, significantly impacting

692   boundary layer dynamics. The new 1km datasets can provide more accurate land surface

693   representations of the variability of individual patches and the inter-patch variability that were

694   used in Huang et al. (2022). The third approach is the Multiple Atmosphere Multiple Land (MAML)

695   approach used in the multiscale modeling framework (MMF) in which a cloud resolving model

696   (CRM) is embedded within each grid cell of the atmosphere (Baker et al., 2019; Lin et al., 2023;

697   Lee et al., 2023). In the MAML approach, each CRM column within the atmosphere grid is coupled

698   directly with its own independent land surface. This enables a more explicit representation of the

699   impact of spatial heterogeneity on land-atmosphere interactions within each grid and has shown

700   notable impacts on water and energy simulations (Baker et al., 2019; Lin et al., 2023). Lee et al.

701   (2023) highlighted the limitation of the current MAML approach, which utilizes the same land

702   surface characteristics for each land surface model interacting with the CRM column within the

703   same grid, which could lead to a weak representation of land-atmosphere interactions. To address

704   this limitation, incorporating the new 1 km land surface parameters within the MAML approach

705   can provide more detailed information about land surface heterogeneity, enabling a more accurate

706   capture of land-atmosphere interactions.

707

708   Evaluation of k-scale simulations, while essential, faces significant challenges as merely updating

709   the land surface input data to the new 1k parameters for k-scale simulations does not guarantee

710   improved model performance. This is clearly evidenced in our ELM demonstration simulations,

711   where, despite relatively low CONUS averaged biases for water and energy simulations, the spatial

712   variation in these biases cannot be overlooked, with some regions exhibiting notably larger biases.

713   It is important to emphasize that enhancing model performance requires not just updated input

714   data, but also appropriate calibration of model parameters and faithful model structures to

715   represent various processes. First, LSMs and ESMs that have been adapted for simulations at

716   coarser resolutions commensurate with the resolutions of previous land surface data require

717   recalibration for effective high-resolution modeling. This necessity for recalibration is echoed by

718   Ruiz-Vásquez et al., (2023), who noted that updating the ECMWF system with new land surface

719   data did not inherently improve performance, but improvements were seen after recalibrating key

720   soil and vegetation-related parameters. Second, high-resolution modeling requires the

721   incorporation of new physical processes crucial at finer scales. For example, hillslope-scale

722   processes like lateral flow and topography-radiation interactions are key to water and energy fluxes

723   at high resolution (Han et al., 2023; Hao et al., 2021). With increased heterogeneity at higher

724   resolutions, larger differences in land surface properties such as vegetation water use strategies

725   requires more attention to plant hydraulics besides the traditional focus on soil hydraulics for a

726   more accurate depiction of plant water use, as highlighted by Li et al., (2021). Third, the lack of

727   high-resolution benchmarks for large-scale applications, like k-scale atmospheric forcing data,

728   remains a challenge, despite the availability of relative coarse resolution global datasets such as

729   ERA5_Land (Muñoz-Sabater et al., 2021) and MSWX (Beck et al., 2021). Additionally, using soil

730   moisture as an example, multiple high-resolution datasets exhibit significantly different

performance when compared to in-situ measurements (Beck et al., 2021). Lastly, when evaluating

simulations against benchmarks, it is crucial not only to assess absolute differences using metrics

like bias and root mean square error but also to examine other metrics, such as the relationships

between physical variables (e.g., rainfall vs. runoff; soil moisture vs. evapotranspiration),

information loss, and the tail quantiles of the probability distribution functions for simulations (e.g.,

extreme events, Li et al., 2020).

There are certain opportunities for future development of 1k parameters. The urban extension may

vary based on 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),

necessitating careful consideration in specific modeling applications. Additionally, urban

classification in J2010, based on global building height data, is limited by the lack of a consistent

and publicly accessible global dataset, despite available regional data for Europe (Frantz et al.,

2021), the US (Li et al., 2020a), and China (Cao and Huang, 2021; Yang and Zhao, 2022), thus

posing challenges to future urban classification enhancements. Incorporating local climate zones

offers a promising approach for urban classification and modeling. Moreover, the multiple-year

high-resolution PFT maps like the ones 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). Soil color, crucial for soil albedo and surface energy balance, lacks extensive global datasets

for ESMs modeling, but the global soil color map derived by Rizzo et al. (2023) offers potential

for further kilometer-scale ESMs and LSMs modeling.

754   The strategic aggregation of high-resolution parameters to coarser resolutions are crucial to

755   maintain accuracy and effectiveness in modeling applications. For instance, in soil properties, the

756   basic parameters (e.g., percent sand) are often utilized to derive secondary parameters (e.g.,

757   saturated water content). This aggregation procedure, whether performs before or after deriving

758   secondary parameters—known as 'aggregating first' and 'aggregating after'—is influenced by the

759   non-linear relationships between basic and derived parameters, with the latter method generally

760   preferred (Shangguan et al., 2014; Dai et al., 2019). Our study's initial approach in upscaling soil-

761   and topography-related parameters follows the 'aggregate first' approach, aligning with the

762   structure of models like ELM2 and CLM5. Conversely, models such as Common Land Model

763   (CoLM, Dai et al., 2003) and community Noah with multi-parameterization options (Noah-MP,

764   He et al., 2023; Niu et al., 2011; Yang et al., 2011) integrate secondary derived soil related

765   parameters directly as inputs, effectively demonstrating the advantages of the 'aggregating after'

766   approach. By leveraging secondary derived parameters from comprehensive databases such as

767   SoilGrids (Hengl et al., 2017) and GSDE (Shangguan et al., 2014), these models provide a valuable

768   framework for future development of models like ELM2 and CLM5 by directly integrating

769   secondary derived parameters.

770

6. Data availability

771

772 The 1 km land surface parameters are publicly available at Zenodo:

773 https://zenodo.org/records/10523833 (Li et al., 2024) and PNNL Datahub:

774 https://doi.org/10.25584/PNNLDH/1986308 (Li et al., 2023).

775

7. Conclusions

776

777 We developed 1 km global land surface parameters using the latest available datasets covering

778 multiple years from 2001 to 2020. These parameters comprise four categories: LULC of PFTs and

779 non-vegetative land cover, vegetation properties, soil properties, and topographic factors. The new

780 1k parameters, when compared to the K2012 and ELM2/CLM5 default datasets, display

781 significant differences, indicating their potential superiority stemming from the utilization of latest

782 and more advanced data sources. The 1 km resolution ELM simulations conducted over CONUS

783 demonstrate the valuable capabilities of the new datasets in enabling k-scale land surface modeling.

784 Through scaling analysis of the 1 km resolution simulations within 0.5º × 0.5º boxes where spatial

785 heterogeneity of the simulations is induced only by spatial heterogeneity of the land surface

786 parameters, we revealed the significant impact of land surface parameters on the spatial variability

787 of water and energy simulations. The spatial information loss of these simulations over CONUS

788 is significant when upscaling from 1 km to a coarser 12 km resolution, with an average ranging

789 from 31% to 54% and up to more than 90%. The XML analysis reveals that the spatial variability

790 and spatial information loss of ELM2 simulations are primarily impacted by the spatial variability

791 and information loss of soil properties, vegetation properties and topography factors, as well as the

792 mean climate conditions of precipitation and temperature. Furthermore, the spatial variability of

793 water and energy in the 1 km simulations is not dominated by the spatial heterogeneity of any land

796   surface parameters, suggesting the usefulness of the multi-parameter high-resolution land surface

797   parameter dataset. Furthermore, the comparison against four benchmark datasets indicates that

798   ELM generally performs well in simulating soil moisture and surface energy fluxes. The

799   availability of 1 km land surface parameters is a valuable resource that addresses the emerging

800   needs of k-scale LSMs and ESMs modeling. By providing accurate and precise information, these

801   1 km land surface parameters will significantly enhance our understanding of the water, carbon,

802   and energy cycles under global change.

803

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.

827

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.
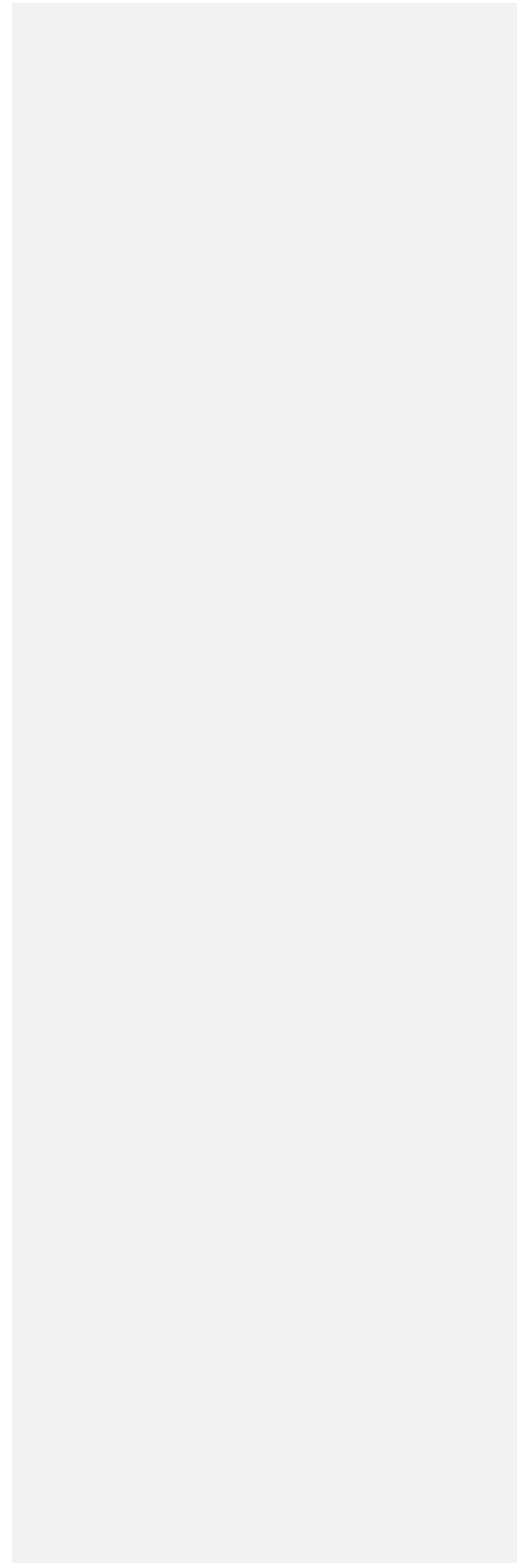
831 **Reference**

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

835 Baker, I. T., Denning, A. S., Dazlich, D. A., Harper, A. B., Branson, M. D., Randall, D. A.,
836 Phillips, M. C., Haynes, K. D., and Gallup, S. M.: Surface-Atmosphere Coupling Scale, the Fate
837 of Water, and Ecophysiological Function in a Brazilian Forest, J Adv Model Earth Sy, 11, 2523–
838 2546, https://doi.org/10.1029/2019ms001650, 2019.

839 Batjes, N.H.: ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid. Report
840 2006/02, available through : http://www.isric.org, 2006.

841 Beck, H. E., Van Dijk, A. I., Larraondo, P. R., McVicar, T. R., Pan, M., Dutra, E., & Miralles, D.
842 G.: MSWX: Global 3-hourly 0.1 bias-corrected meteorological data including near-real-time
843 updates and forecast ensembles, BAMS, 103(3), E710-E732, https://doi.org/10.1175/BAMS-D-
844 21-0145.1, 2022.

845 Beck, H. E., Pan, M., Miralles, D. G., Reichle, R. H., Dorigo, W. A., Hahn, S., Sheffield, J.,
846 Karthikeyan, L., Balsamo, G., Parinussa, R. M., van Dijk, A. I. J. M., Du, J., Kimball, J. S.,
847 Vergopolan, N., and Wood, E. F.: Evaluation of 18 satellite- and model-based soil moisture
848 products using in situ measurements from 826 sensors, Hydrol Earth Syst Sci, 25, 17–40,
849 https://doi.org/10.5194/hess-25-17-2021, 2021.

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

853 Bonan, G. B., Oleson, K. W., Vertenstein, M., Levis, S., Zeng, X., Dai, Y., Dickinson, R. E., and
854 Yang, Z.-L.: The Land Surface Climatology of the Community Land Model Coupled to the
855 NCAR Community Climate Model*, J Climate, 15, 3123–3149, https://doi.org/10.1175/1520-
856 0442(2002)015<;3123:tlscot>2.0.co;2, 2002.

857 Bonan, G. B., Levis, S., Kergoat, L., & Oleson, K. W.: Landscapes as patches of plant functional
858 types: An integrating concept for climate and ecosystem models. Global Biogeochemical Cycles,
859 16(2), 5-1–5-23. https://doi.org/10.1029/2000gb001360, 2002

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

863 Caldwell, P. M., Mametjanov, A., Tang, Q., Roekel, L. P. V., Golaz, J., Lin, W., Bader, D. C.,
864 Keen, N. D., Feng, Y., Jacob, R., Maltrud, M. E., Roberts, A. F., et al.: The DOE E3SM Coupled
865 Model Version 1: Description and Results at High Resolution, J Adv Model Earth Sy, 11, 4095–
866 4146, https://doi.org/10.1029/2019ms001870, 2019.

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

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

873 Chaney, N. W., Huijgevoort, M. H. J. V., Shevliakova, E., Malyshev, S., Milly, P. C. D.,
874 Gauthier, P. P. G., and Sulman, B. N.: Harnessing big data to rethink land heterogeneity in Earth
875 system models, Hydrol Earth Syst Sc, 22, 3311–3330, https://doi.org/10.5194/hess-22-3311-
876 2018, 2018.

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

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

881 Dai, Y., Zeng, X., Dickinson, R. E., Baker, I., Bonan, G. B., Bosilovich, M. G., et al.: The
882 common land model, BAMS, 84(8), 1013-1024, https://doi.org/10.1175/BAMS-84-8-1013,
883 2003.

884 Dai, Y., Shangguan, W., Wei, N., Xin, Q., Yuan, H., Zhang, S., Liu, S., Lu, X., Wang, D., and
885 Yan, F.: A Review of the Global Soil Property Maps for Earth System Models, SOIL, 5, 137-
886 158. https://doi.org/10.5194/soil-5-137-2019, 2019.

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

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

893 Fan, Y., Clark, M., Lawrence, D. M., Swenson, S., Band, L. E., Brantley, S. L., et al.: Hillslope
894 Hydrology in Global Change Research and Earth System Modeling, Water Resour Res, 55,
895 1737–1772, https://doi.org/10.1029/2018wr023903, 2019.

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

899 Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H.,
900 Woodcock, C. E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., and Schaaf, C.:

54

901 Global land cover mapping from MODIS: algorithms and early results, Remote Sens Environ,
902 83, 287–302, https://doi.org/10.1016/s0034-4257(02)00078-0, 2002.

903 Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and
904 Huang, X.: MODIS Collection 5 global land cover: Algorithm refinements and characterization
905 of new datasets, Remote Sens Environ, 114, 168–182, https://doi.org/10.1016/j.rse.2009.08.016,
906 2010.

907 Friedl, M., Sulla-Menashe, D.: MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3
908 Global 500m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC. Accessed
909 2022-11-21 from https://doi.org/10.5067/MODIS/MCD12Q1.006, 2019.

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

913 Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R.: Google Earth
914 Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, 202,
915 18–27. https://doi.org/10.1016/j.rse.2017.06.031, 2017.

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

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

922 Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Carroll, M., Dimiceli, C., and Sohlberg, R.
923 A.: Global Percent Tree Cover at a Spatial Resolution of 500 Meters: First Results of the MODIS
924 Vegetation Continuous Fields Algorithm, Earth Interact, 7, 1–15, https://doi.org/10.1175/1087-
925 3562(2003)007<;0001:gptcaa>2.0.co;2, 2003.

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

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

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

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

940   He, C., Valayamkunnath, P., Barlage, M., Chen, F., Gochis, D., Cabell, R., Schneider, T.,
941   Rasmussen, R., Niu, G.-Y., Yang, Z.-L., Niyogi, D., and Ek, M.: Modernizing the open-source
942   community Noah with multi-parameterization options (Noah-MP) land surface model (version
943   5.0) with enhanced modularity, interoperability, and applicability, Geosci Model Dev, 16, 5131–
944   5151, https://doi.org/10.5194/gmd-16-5131-2023, 2023.

945   Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A.: Very high resolution
946   interpolated climate surfaces for global land areas, International Journal of Climatology, 25,
947   1965–1978, https://doi.org/10.1002/joc.1276, 2005.

948   Hu, Z., Islam, S., and Cheng, Y.: Statistical characterization of remotely sensed soil moisture
949   images, Remote Sens Environ, 61, 310–318, https://doi.org/10.1016/s0034-4257(97)89498-9,
950   1997.

951   Huang, M., Ma, P.-L., Chaney, N. W., Hao, D., Bisht, G., Fowler, M. D., Larson, V. E., and
952   Leung, L. R.: Representing surface heterogeneity in land-atmosphere coupling in E3SMv1
953   single-column model over ARM SGP during summertime, Geoscientific Model Dev Discuss,
954   2022, 1–20, https://doi.org/10.5194/gmd-2021-421, 2022.

955   Huang, F., Jiang, S., Zhan, W., Bechtel, B., Liu, Z., Demuzere, M.: Mapping local climate zones
956   for cities: A large review, Remote Sens. Environ., 292, 113573,
957   https://doi.org/10.1016/j.rse.2023.113573, 2023.

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

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

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

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

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

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

975   Ke, Y., Leung, L. R., Huang, M., Coleman, A. M., Li, H., and Wigmosta, M. S.: Development of
976   high resolution land surface parameters for the Community Land Model, Geosci Model Dev, 5,
977   1341–1362, https://doi.org/10.5194/gmd-5-1341-2012, 2012.

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

981   Kim, H.: Global Soil Wetness Project Phase 3 Atmospheric Boundary Conditions (Experiment 1)
982   [Data set]. Data Integration and Analysis System (DIAS). https://doi.org/10.20783/DIAS.501,
983   2017.

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

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

989   Krakauer, N. Y., Li, H., and Fan, Y.: Groundwater flow across spatial scales: importance for
990   climate modeling, Environ Res Lett, 9, 034003, https://doi.org/10.1088/1748-9326/9/3/034003,
991   2014.

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

996   Lang, N., Jetz, W., Schindler, K., and Wegner, J. D.: A high-resolution canopy height model of
997   the Earth. Nat Ecol Evol, https://doi.org/10.1038/s41559-023-02206-6, 2023.

998   Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., et al.:
999   The Community Land Model Version 5: Description of New Features, Benchmarking, and
1000   Impact of Forcing Uncertainty, J Adv Model Earth Sy, 11, 4245–4287,
1001   https://doi.org/10.1029/2018ms001583, 2019.

1002   Lawrence, D., Fisher, R., Koven, C., Oleson, K., Swenson, S., et al. (2018). Technical
1003   description of version 5.0 of the Community Land Model (CLM). National Center for
1004   Atmospheric Research, University Corporation for Atmospheric Research, Boulder, CO.
1005   https://escomp.github.io/ctsm-docs/versions/release-clm5.0/html/tech_note/index.html

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

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

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

Li, L., Yang, Z., Matheny, A. M., Zheng, H., Swenson, S. C., Lawrence, D. M., Barlage, M., Yan, B., McDowell, N. G., and Leung, L. R.: Representation of Plant Hydraulics in the Noah-MP Land Surface Model: Model Development and Multiscale Evaluation, J Adv Model Earth Sy, 13, https://doi.org/10.1029/2020ms002214, 2021.

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

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

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

Li, L., She, D., Zheng, H., Lin, P., and Yang, Z.-L.: Elucidating Diverse Drought Characteristics from Two Meteorological Drought Indices (SPI and SPEI) in China, J Hydrometeorol, 21, https://doi.org/10.1175/JHM-D-19-0290.1, 2020.

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

Li, L., Bisht, G., Hao, D., Leung, L.R.: Global 1km Land Surface Parameters for Kilometer-Scale Earth System Modeling [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10523833, 2024

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

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

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

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

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

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

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

Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A., and Verhoest, N. E. C.: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, Geosci. Model Dev., 10, 1903–1925, https://doi.org/10.5194/gmd-10-1903-2017, 2017.

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

Morisette, J. T., Baret, F., Privette, J. L., Myneni, R. B., Nickeson, J. E., et al.: Validation of global moderate-resolution LAI products: A framework proposed within the CEOS land product validation subgroup, IEEE Trans Geosci Remote Sens, 44(7), 1804-1817. https://doi.org/10.1109/TGRS.2006.872529, 2006.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, Earth Syst Sci Data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G. R., Lotsch, A., Friedl, M., Morisette, J. T., Votava, P., Nemani, R. R.,

1078   and Running, S. W.: Global products of vegetation leaf area and fraction absorbed PAR from
1079   year one of MODIS data, Remote Sens Environ, 83, 214–231, https://doi.org/10.1016/s0034-
1080   4257(02)00074-3, 2002.

1081   Myneni, R., Knyazikhin, Y., Park, T. (2021). MODIS/Terra+Aqua Leaf Area Index/FPAR 4-Day
1082   L4 Global 500m SIN Grid V061 [Data set]. NASA EOSDIS Land Processes DAAC. Accessed
1083   2022-11-21 from https://doi.org/10.5067/MODIS/MCD15A3H.061

1084   Naz, B. S., Sharples, W., Ma, Y., Goergen, K., and Kollet, S.: Continental-scale evaluation of a
1085   fully distributed coupled land surface and groundwater model, ParFlow-CLM (v3.6.0), over
1086   Europe, Geosci Model Dev, 16, 1617–1639, https://doi.org/10.5194/gmd-16-1617-2023, 2023.

1087   Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al.: The community
1088   Noah land surface model with multiparameterization options (Noah-MP): 1. Model description
1089   and evaluation with local-scale measurements, J Geophys Res Atmos, 116(D12),
1090   https://doi.org/10.1029/2010JD015139, 2011

1091   O'Neill, M. M. F., Tijerina, D. T., Condon, L. E., and Maxwell, R. M.: Assessment of the
1092   ParFlow–CLM CONUS 1.0 integrated hydrologic model: evaluation of hyper-resolution water
1093   balance components across the contiguous United States, Geosci Model Dev, 14, 7223–7254,
1094   https://doi.org/10.5194/gmd-14-7223-2021, 2021.

1095   Poggio, L., Sousa, L. M. de, Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., and
1096   Rossiter, D.: SoilGrids 2.0: producing soil information for the globe with quantified spatial
1097   uncertainty, Soil, 7, 217–240, https://doi.org/10.5194/soil-7-217-2021, 2021.

1098   Qiu, H., Bisht, G., Li, L., Hao, D., and Xu, D.: Development of Inter-Grid Cell Lateral
1099   Unsaturated and Saturated Flow Model in the E3SM Land Model (v2.0), Egusphere, 2023, 1–31,
1100   https://doi.org/10.5194/egusphere-2023-375, 2023.

1101   Rabus, B., Eineder, M., Roth, A., & Bamler, R.: The shuttle radar topography mission—A new
1102   class of digital elevation models acquired by spaceborne radar. ISPRS Journal of
1103   Photogrammetry and Remote Sensing, 57(4), 241–262. https://doi.org/10.1016/s0924-
1104   2716(02)00124-7, 2003.

1105   Ramankutty, N. and Foley, J. A.: Estimating historical changes in global land cover: Croplands
1106   from 1700 to 1992, Global Biogeochem Cy, 13, 997–1027,
1107   https://doi.org/10.1029/1999gb900046, 1999.

1108   Rasch, P. J., Xie, S., Ma, P. -L., Lin, W., Wang, H., Tang, Q., Burrows, S. M., Caldwell, P., et
1109   al.: An Overview of the Atmospheric Component of the Energy Exascale Earth System Model, J
1110   Adv Model Earth Sy, 11, 2377–2411, https://doi.org/10.1029/2019ms001629, 2019.

1111   Rastner, P., Bolch, T., Mölg, N., Machguth, H., Le Bris, R., and Paul, F.: The first complete
1112   inventory of the local glaciers and ice caps on Greenland, The Cryosphere, 6, 1483–1495,
1113   https://doi.org/10.5194/tc-6-1483-2012, 2012.

1114 Rizzo, R., Wadoux, A. M. C., Demattê, J. A., Minasny, B., Barrón, V., Ben-Dor, E., ... &
1115 Salama, E. S. M.: Remote sensing of the Earth's soil color in space and time, Remote Sens
1116 Environ, 299, 113845, https://doi.org/10.1016/j.rse.2023.113845, 2023.

1117 Rouf, T., Maggioni, V., Mei, Y., and Houser, P.: Towards hyper-resolution land-surface
1118 modeling of surface and root zone soil moisture, J Hydrol, 594, 125945,
1119 https://doi.org/10.1016/j.jhydrol.2020.125945, 2021.

1120 Ruiz-Vásquez, M., O, S., Arduini, G., Boussetta, S., Brenning, A., et al: Impact of updating
1121 vegetation information on land surface model performance, J. Geophys. Res. Atmos., 128(21),
1122 e2023JD039076, https://doi.org/10.1029/2023JD039076, 2023.

1123 Running, S., Mu, Q., Zhao, M., Moreno, A. (2021). MODIS/Terra Net Evapotranspiration Gap-
1124 Filled 8-Day L4 Global 500m SIN Grid V061 [Data set]. NASA EOSDIS Land Processes
1125 Distributed Active Archive Center. https://doi.org/10.5067/MODIS/MOD16A2GF.061

1126 Shangguan, W., Dai, Y., Duan, Q., Liu, B., and Yuan, H.: A Global Soil Dataset for Earth
1127 System Modeling, J Adv Model Earth Syst, 6, 249-263, https://doi.org/10.1002/2013MS000293,
1128 2014.

1129 Simard, M., Pinto, N., Fisher, J. B., and Baccini, A.: Mapping forest canopy height globally with
1130 spaceborne lidar, J. Geophys. Res. Biogeosci., 116, G04021,
1131 https://doi.org/10.1029/2011jg001708, 2011.

1132 Singh, R. S., Reager, J. T., Miller, N. L., and Famiglietti, J. S.: Toward hyper-resolution land-
1133 surface modeling: The effects of fine-scale topography and soil texture on CLM4.0 simulations
1134 over the Southwestern U.S., Water Resour Res, 51, 2648–2667,
1135 https://doi.org/10.1002/2014wr015686, 2015.

1136 Slingo, J., Bates, P., Bauer, P., Belcher, S., Palmer, T., Stephens, G., Stevens, B., Stocker, T.,
1137 and Teutsch, G.: Ambitious partnership needed for reliable climate prediction, Nat Clim Change,
1138 12, 499–503, https://doi.org/10.1038/s41558-022-01384-8, 2022.

1139 Still, C. J., Berry, J. A., Collatz, G. J., and DeFries, R. S.: Global distribution of C3 and C4
1140 vegetation: Carbon cycle implications, Global Biogeochem Cy, 17, 6-1-6–14,
1141 https://doi.org/10.1029/2001gb001807, 2003.

1142 Sulla-Menashe, D., Gray, J. M., Abercrombie, S. P., and Friedl, M. A.: Hierarchical mapping of
1143 annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product, Remote
1144 Sens Environ, 222, 183–194, https://doi.org/10.1016/j.rse.2018.12.013, 2019.

1145 Swenson, S. C., Clark, M., Fan, Y., Lawrence, D. M., and Perket, J.: Representing Intrahillslope
1146 Lateral Subsurface Flow in the Community Land Model, J Adv Model Earth Sy, 11, 4044–4065,
1147 https://doi.org/10.1029/2019ms001833, 2019.

1148 Verdin, K. L. and Greenlee, S. K.: Development of continental scale digital elevation models and
1149 extraction of hydrographic features, paper presented at the Third International Workshop on
1150 Integrating GIS and Environmental Modeling, Santa Fe, New Mexico, 21–26 January, Natl.
1151 Cent. for Geogr. Inf. and Anal., Santa Barbara, Calif, 1996.

1152 Vergopolan, N., Chaney, N. W., Beck, H. E., Pan, M., Sheffield, J., Chan, S., and Wood, E. F.:
1153 Combining hyper-resolution land surface modeling with SMAP brightness temperatures to
1154 obtain 30-m soil moisture estimates, Remote Sens Environ, 242, 111740,
1155 https://doi.org/10.1016/j.rse.2020.111740, 2020.

1156 Vergopolan, N., Chaney, N. W., Pan, M., Sheffield, J., Beck, H. E., Ferguson, C. R., Torres-
1157 Rojas, L., Sadri, S., and Wood, E. F.: SMAP-HydroBlocks, a 30-m satellite-based soil moisture
1158 dataset for the conterminous US, Sci Data, 8, 264, https://doi.org/10.1038/s41597-021-01050-2,
1159 2021.

1160 Vergopolan, N., Sheffield, J., Chaney, N. W., Pan, M., Beck, H. E., Ferguson, C. R., Torres-
1161 Rojas, L., Eigenbrod, F., Crow, W., and Wood, E. F.: High-Resolution Soil Moisture Data
1162 Reveal Complex Multi-Scale Spatial Variability Across the United States, Geophys Res Lett, 49,
1163 https://doi.org/10.1029/2022gl098586, 2022.

1164 Wood, E. F., Roundy, J. K., Troy, T. J., Beek, L. P. H. van, Bierkens, M. F. P., et al.:
1165 Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's
1166 terrestrial water, Water Resour Res, 47, https://doi.org/10.1029/2010wr010090, 2011.

1167 Xia, Y., Mocko, D., Huang, M., Li, B., Rodell, M., Mitchell, K. E., Cai, X., and Ek, M. B.:
1168 Comparison and Assessment of Three Advanced Land Surface Models in Simulating Terrestrial
1169 Water Storage Components over the United States, J Hydrometeorol, 18, 625–649,
1170 https://doi.org/10.1175/jhm-d-16-0112.1, 2017.

1171 Xu, C., Torres-Rojas, L., Vergopolan, N., Chaney, N. W.: The Benefits of Using State-Of-The-
1172 Art Digital Soil Properties Maps to Improve the Modeling of Soil Moisture in Land Surface
1173 Models, Water Resour Res, 59(4), e2022WR032336, https://doi.org/10.1029/2022WR032336,
1174 2023.

1175 Yang, Z. L., Niu, G. Y., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al.: The community
1176 Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over
1177 global river basins, J Geophys Res Atmos, 116(D12), https://doi.org/10.1029/2010JD015140,
1178 2011.

1179 Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., and Pavelsky, T. M.: MERIT
1180 Hydro: A High-Resolution Global Hydrography Map Based on Latest Topography Dataset,
1181 Water Resour Res, 55, 5053–5073, https://doi.org/10.1029/2019wr024873, 2019.

1182 Yang, C. and Zhao, S.: A building height dataset across China in 2017 estimated by the spatially-
1183 informed approach, Sci Data, 9, 76, https://doi.org/10.1038/s41597-022-01192-x, 2022.

1184   Yuan, H., Dai, Y., Xiao, Z., Ji, D., and Shangguan, W.: Reprocessing the MODIS Leaf Area
1185   Index products for land surface and climate modelling, Remote Sens Environ, 115, 1171–1187,
1186   https://doi.org/10.1016/j.rse.2011.01.001, 2011.

1187   Yuan, X., Ji, P., Wang, L., Liang, X., Yang, K., Ye, A., Su, Z., and Wen, J.: High-Resolution
1188   Land Surface Modeling of Hydrological Changes Over the Sanjiangyuan Region in the Eastern
1189   Tibetan Plateau: 1. Model Development and Evaluation, J Adv Model Earth Sy, 10, 2806–2828,
1190   https://doi.org/10.1029/2018ms001412, 2018.

1191   Zeng, X., Shaikh, M., Dai, Y., Dickinson, R. E., and Myneni, R.: Coupling of the Common Land
1192   Model to the NCAR Community Climate Model, J Climate, 15, 1832–1854,
1193   https://doi.org/10.1175/1520-0442(2002)015<;1832:cotclm>2.0.co;2, 2002.

1194   Zhao, M., Cheng, C., Zhou, Y., Li, X., Shen, S., and Song, C.: A global dataset of annual urban
1195   extents (1992–2020) from harmonized nighttime lights, Earth Syst Sci Data, 14, 517–534,
1196   https://doi.org/10.5194/essd-14-517-2022, 2022.

1197   Zhou, Y., Li, D., and Li, X.: The Effects of Surface Heterogeneity Scale on the Flux Imbalance
1198   under Free Convection, J Geophys Res Atmospheres, 124, 8424–8448,
1199   https://doi.org/10.1029/2018jd029550, 2019.