



DATA ARTICLE

A global dataset on subgrid land surface climate (2015–2100) from the Community Earth System Model

Keer Zhang¹ | Xuhui Lee¹ | Natalie M. Schultz¹ | Lei Zhao² | Chunyang He^{3,4} | Qingxu Huang^{3,4} | Zhifeng Liu^{3,4} | Haoran Chu^{5,6} | Jiayu Zhao^{5,6}

¹School of the Environment, Yale University, New Haven, CT, USA

²Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

³Center for Human-Environment System Sustainability (CHESS), State Key Laboratory of Earth Surface Processes and Resource Ecology (ESPRE), Beijing Normal University, Beijing, China

⁴School of Natural Resources, Faculty of Geographical Science, Beijing Normal University, Beijing, China

⁵Yale-NUIST Center on Atmospheric Environment, International Joint Laboratory on Climate and Environment Change (ILCEC), Nanjing University of Information Science and Technology, Nanjing, China

⁶Key Laboratory of Meteorological Disaster, Ministry of Education and Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing, China

Correspondence

Keer Zhang, School of the Environment, Yale University, New Haven, CT, USA.
Email: keer.zhang@yale.edu

Funding information

U.S. National Science Foundation (Grant AGS1933630 to XL and NS); National Key R&D Program of China (Grant 2019YFA0607203 to CH, QH, and ZL); Yale Graduate Fellowship (to KZ)

Abstract

Subgrid data from earth system models are a powerful, yet underutilized, data resource for investigating the climatic impacts of land use and land cover change (LULCC). In this paper, we describe a global dataset on subgrid land surface climate variables produced by the Community Earth System Model in a fully coupled mode. The simulation was conducted at a $0.9^\circ \times 1.25^\circ$ resolution under the Representative Concentration Pathway (RCP) 8.5 scenario from 2015 to 2100. Data are archived for eight subgrid tiles (urban, rural, tree, grass, shrub, bare soil, crop and lake) and include variables on the physical state, surface energy fluxes, runoff and atmospheric forcing conditions. Archival intervals are monthly, daily and hourly. Meta data on land surface parameters are also available. The data files are stored in NetCDF-4 (Network Common Data Form, version 4) format and

Dataset

Identifier: <https://doi.org/10.7910/DVN/HUXAH6>

Creator: Keer Zhang, Xuhui Lee, Natalie M. Schultz, Lei Zhao, Chunyang He, Qingxu Huang, Zhifeng Liu, Haoran Chu, Jiayu Zhao

Dataset correspondence: keer.zhang@yale.edu

Title: A global dataset on subgrid land surface climate (2015–2100) from the Community Earth System Model

Publisher: Harvard Dataverse

Publication year: 2021

Resource type: Dataset

Version: 1.0

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Geoscience Data Journal* published by Royal Meteorological Society and John Wiley & Sons Ltd.

the meta data follow the latest Coupled Model Intercomparison Project phase 6 standards. We anticipate that this dataset will become a useful resource for characterizing local climate changes due to LULCC. This dataset can be downloaded from the Harvard Dataverse (<https://doi.org/10.7910/DVN/HUXAH6>).

KEYWORDS

biophysical impacts, climate model, individual soil column, land use and land cover change, subgrid data

1 | INTRODUCTION

Land use and land cover change (LULCC), such as deforestation and urbanization, alters local and regional climates via biophysical mechanisms (Alkama & Cescatti, 2016; Lawrence et al., 2012; Mahmood et al., 2014; Pitman et al., 2009; Zhao et al., 2014). The biophysical process affects energy and water cycles via changing surface radiative properties, heat storage, evapotranspiration and roughness (Mahmood et al., 2014). It can either amplify or dampen the temperature response to increasing greenhouse gas concentrations (Boysen et al., 2020; Liao et al., 2020). A good understanding of the biophysical impact of LULCC is critical to develop strategies for heat stress mitigation and sustainable land use.

The biophysical impact of LULCC has been investigated via observations and climate modelling (Boysen et al., 2020; Feddema et al., 2005; Lee et al., 2011). Observational studies utilize data from in situ observation and satellite remote sensing to quantify the impacts of LULCC based on a space-for-time analogy (Duveiller et al., 2018; Lee et al., 2011; Schultz et al., 2017; Tang et al., 2018; Zhang et al., 2014). They focus on the physical state differences between two contrasting systems located in the vicinity of each other, such as urban *versus* rural (Scott et al., 2018) and forest *versus* open land (Zhang et al., 2014). It is assumed that they share the same atmospheric background and that differences in their temperatures and energy fluxes are solely attributed to their contrasting biophysical properties. Most modelling studies are based on two sets of simulations with different land cover under the same model configuration (Davin et al., 2020; Findell et al., 2017; Lawrence & Chase, 2010). For instance, the biophysical effect of deforestation can be evaluated by one simulation with intact vegetation cover and another with forest removed. These sensitivity experiments are computationally costly. A less expensive modelling approach takes advantage of the mosaic strategy of many land models, in which the land surface heterogeneity is represented as a mosaic of subgrid land tiles. The impact of LULCC is quantified by the physical state differences

among different land tiles in a grid cell (Malyshev et al., 2015; Oleson et al., 2010). Within one grid cell, all land tiles receive the same atmospheric forcing. The subgrid variabilities allow us to isolate the local climate changes due to LULCC forcing from the changes induced by atmospheric climate variability.

The subgrid strategy has been employed to investigate how the local climate responds to urbanization and deforestation. The temperature response to urbanization, or the urban heat island (UHI) intensity, is the temperature difference between the urban and nonurban land tiles in a grid cell (Oleson et al., 2011). In deforestation studies, the deforestation signal is computed as the temperature difference between the forest and nonforest land tiles (Malyshev et al., 2015). These studies demonstrate that the subgrid information produced by earth system models (ESMs) can provide valuable insights for LULCC studies.

Traditionally, climate model data are archived as grid-mean values. Subgrid data remain underutilized because they do not exist in the Coupled Model Intercomparison Project (CMIP) data depositories. In order to “retrieve” the data, users have to rerun a land surface model with the atmospheric forcing data saved from old coupled simulations (e.g. Zhao et al., 2017). Such retrieval tasks are daunting, requiring supercomputing facilities and skills in climate modelling.

The main objective of this paper is to describe a subgrid dataset produced by the earth system model – CESM (Community Earth System Model) – for the period from 2015 to 2100 under a high-emission climate change scenario. The key biophysical variables (temperature, humidity, energy fluxes, atmospheric forcing and runoff variables) are output at the grid and subgrid levels at monthly, daily and hourly intervals. The subgrid land tiles include urban, rural, tree, grass, shrub, bare soil, crop and lake. Besides, the surface data, such as the fraction of land-use type, lake depth and leaf area index, are also archived to provide key surface information to aid the users. We anticipate that this dataset will become a valuable resource for characterizing local climate changes due to LULCC.

2 | MODEL AND EXPERIMENT

2.1 | Model description

The CESM Version 2 (CESM2) is a fully coupled global climate model (Danabasoglu et al., 2020). The land component of CESM2 is the Community Land Model Version 5 (CLM5; Lawrence et al., 2019). In CLM5, the land surface heterogeneity is represented as a nested subgrid hierarchy (Lawrence et al., 2019). Each grid cell is represented by up to five land units (vegetated, crop, lake, glacier and urban). Each land unit can have different number of columns to capture the state variability in the soil and snow. Each vegetated or crop column consists of at least one plant functional type (PFT) or crop functional type (CFT). All subgrid units in a grid cell receive the same atmospheric forcing, and their physical states are computed separately at the subgrid level and are area-weighted to obtain grid-cell results.

2.2 | Soil column configuration

By default, all PFTs in a grid cell share a single soil column and compete for water and nutrients. There are no variations of soil moisture and nitrogen conditions among the PFTs in a land unit. An alternative configuration is independent soil column whereby each PFT in the vegetated land unit has its own soil column. The individual soil column scheme is preferable for studying the climatic impact of LULCC (Meier et al., 2018; Schultz et al., 2016), and was used in our simulation.

2.3 | Experimental design

We conducted a fully coupled global climate simulation using CESM Version 2.1.2 with a $0.9^\circ \times 1.25^\circ$ grid resolution. The atmosphere, ocean and land components are Community Atmosphere Model version 6, Parallel Ocean Program version 2 and CLM5. The simulation was driven by the future climate forcing under the Representative Concentration Pathway (RCP) 8.5 scenario. The vegetation phenology was prescribed by satellite observations.

The surface data were fixed at present-day level in this simulation. In CLM5, the default present-day surface data use the global urban extent and urban properties at 2000 developed by Jackson et al. (2010) (hereafter J2010). Here, the urban extent in J2010 was replaced with the extent of 2020 produced by He et al. (2021) (hereafter He2021; Supporting Information). He et al. (2021) projected the global urban land from 2020 to 2070 under five Shared Socioeconomic Pathways (SSPs) at a 1-km resolution as binary data (urban or nonurban). We only used the urban land cover in 2020

under the SSP5 scenario from He2021 to generate our new surface data. The reason for using He2021 is that we plan to carry out additional experiments with future urban extent provided by He2021. We wish to use the urban land cover from the same data source for consistency among this simulation and future experiments. The He2021 dataset is in broad agreement, in terms of total urban area and urban area growth rate, with previous urban land cover projections (Figure S2; Chen et al., 2020; Gao & O'Neill, 2020; Huang et al., 2021; Li et al., 2019). The urban properties, including morphological, thermal and radiative characteristics, prescribed for the new surface dataset are the same with those in the J2010 dataset. Besides, all other land use is held constant at 2015 levels.

The key variables of the surface data are archived as meta data (Table S1). Examples include the fraction of land-use type, urban canyon height-to-width ratio, lake depth and leaf area index. Figure 1 displays the fraction of urban, lake, tree (including all tree PFTs) and grass (including all grass PFTs) in the surface dataset.

A 60-year spin-up was conducted offline with the individual soil column configuration in order to achieve an equilibrium land condition. The spin-up was initiated with a start file which has been spun up with the shared column configuration for at least 800 years using the standard CLM spin-up protocol (Lawrence et al., 2019). The land air temperature, vapor pressure and energy fluxes adjusted quickly in the first year after the configuration change (Figure S3). The soil moisture also reached a stable condition after 40 years of simulation. The annual mean ground heat flux of the tree land tile differed by $0.1\text{--}5\text{ W m}^{-2}$ from zero in year 1 of the spin-up and approached near-zero values ($<0.5\text{ W m}^{-2}$ in magnitude) at low and mid latitudes in year 60 (Figure S4). Similarly, the annual mean ground heat flux at low and mid latitudes also converged to zero for grass and other PFTs. In high-latitude regions (north of 44°N and south of 50°S), these subgrid ground heat fluxes were nonzero because of energy consumption by snowmelt.

Our simulation started with the spun-up initial land condition and model integration continued from 2015 to 2100. The key biophysical variables at the grid and subgrid levels were archived at monthly and daily intervals during the entire simulation and at hourly intervals during two 5-year periods (2019–2023 and 2096–2100).

3 | DATA DESCRIPTION

3.1 | Data preparation

Table 1 displays a list of output variables. All variables are archived at the subgrid level except the atmospheric

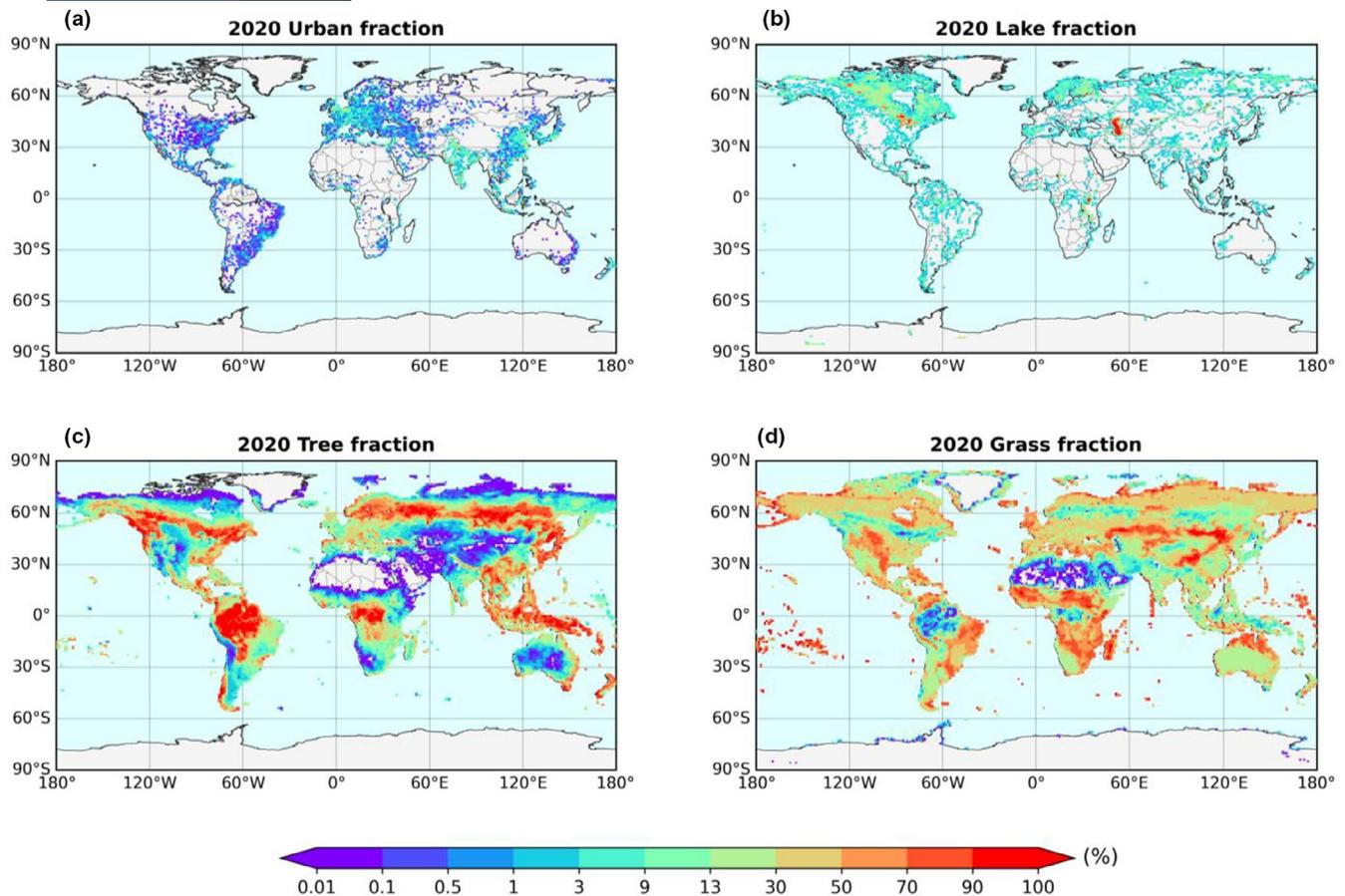


FIGURE 1 Spatial distribution of the percent fraction of (a) urban, (b) lake, (c) tree and (d) grass in the new present-day surface data. Tree and grass fractions include all tree and grass PFTs in a grid cell respectively

forcing variables because all subgrid units within a grid share the same atmospheric forcing. Monthly and daily data are available from 2015 to 2100 except run-off variables which are only available at monthly intervals. Hourly data are available during 2019–2023 and 2096–2100.

The model outputs grid, land unit and PFT level results. We combined these outputs as area-weighted means for eight overlapping subgrid land tiles: urban, rural, tree, grass, shrub, bare soil, crop and lake (Figure 2). The urban tile is the area-weighted result of three urban density classes, and the lake tile is the result of the lake land unit. The rural tile is the area-weighted results of all the natural vegetation PFTs, crop CFTs and the bare soil PFT. Similarly, the tree, grass, shrub and crop tiles are the area-weighted results of their respective PFTs.

The radiative surface temperature (T_s) is not a standard variable in this archive. It can be computed from the surface upwelling longwave radiation flux and surface emissivity (Supporting Information). Other diagnostic information can also be derived from the archived variables. For example, lake ice period can be diagnosed with lake surface albedo which is the ratio of $rsus$ to $rsds$ (Table 1).

An albedo greater than 0.2 usually indicates ice cover (Oke, 1987; Subin et al., 2012).

The data are archived in the NetCDF-4 (Network Common Data Form, version 4) format with one variable stored per file. The meta data follow the CMIP6 standards built on the CF convention. All hourly data are stamped with UTC time. The naming convention for the data files is *b.e21.SSP585UC_IndividualSoil_ + short name + grid/subgrid land tile + output frequency + time period.nc*. As an example, file '*b.e21.SSP585UC_IndividualSoil_tas_Grid_Daily_2015-2019.nc*' contains the grid mean 2-m air temperature daily data from 2015 to 2019.

3.2 | Data and code availability

This dataset can be freely downloaded from the Harvard Dataverse (download page: <https://doi.org/10.7910/DVN/HUXAH6>; Zhang et al., 2021). The total file size is 484 GB. The monthly files are 6.2 GB, daily files 164 GB and hourly files 313 GB in size. The surface data with key variables (Table S1), 386 MB in size, are also available for download. We provide sample python scripts at

TABLE 1 A list of output variables. Symbol × indicates data availability

Variable	Short name	Unit	Grid	Subgrid	Frequency		
					Monthly	Daily	Hourly
2-m air temperature	<i>tas</i>	K	×	×	×	×	×
2-m vapor pressure	<i>huv</i>	Pa	×	×	×	×	×
2-m wet-bulb temperature	<i>wba</i>	°C	×	×	×	×	×
Daily maximum 2-m air temperature	<i>tasmax</i>	K	×	×	×	×	
Daily minimum 2-m air temperature	<i>tasmin</i>	K	×	×	×	×	
Daily maximum 2-m wet-bulb temperature	<i>wbamax</i>	°C	×	×	×	×	
Daily minimum 2-m wet-bulb temperature	<i>wbamin</i>	°C	×	×	×	×	
Latent heat flux	<i>hfls</i>	W m ⁻²	×	×	×	×	×
Sensible heat flux	<i>hfss</i>	W m ⁻²	×	×	×	×	×
Ground heat flux	<i>hfdsl</i>	W m ⁻²	×	×	×	×	×
Waste heat from air conditioning (AC) and space heating	<i>fwaste</i>	W m ⁻²	×	×	×	×	×
Heat removed by AC	<i>fahac</i>	W m ⁻²	×	×	×	×	×
Heat generated by space heating	<i>fahsh</i>	W m ⁻²	×	×	×	×	×
Surface upwelling shortwave radiation	<i>rsus</i>	W m ⁻²	×	×	×	×	×
Surface upwelling longwave radiation	<i>rlus</i>	W m ⁻²	×	×	×	×	×
Surface downwelling shortwave radiation	<i>rsds</i>	W m ⁻²	×		×	×	×
Surface downwelling longwave radiation	<i>rlds</i>	W m ⁻²	×		×	×	×
Precipitation	<i>pr</i>	kg m ⁻² s ⁻¹	×		×	×	×
Atmospheric air pressure at forcing height	<i>fps</i>	Pa	×		×	×	×
Atmospheric air temperature at forcing height	<i>fta</i>	K	×		×	×	×
Atmospheric air potential temperature at forcing height	<i>ftheta</i>	K	×		×	×	×
Atmospheric specific humidity at forcing height	<i>fhus</i>	kg kg ⁻¹	×		×	×	×
10-m wind speed	<i>sfcWind</i>	m s ⁻¹	×		×	×	×
Total runoff	<i>mrro</i>	kg m ⁻² s ⁻¹	×	×	×		
Total surface runoff	<i>sfcmrro</i>	kg m ⁻² s ⁻¹	×	×	×		

Note: Monthly and daily data are available for the whole simulation period and hourly data are available during 2019–2023 and 2096–2100.

the download page for reading the data, generating seasonal and diurnal variation plots, and calculating urban emissivity. The raw outputs of all variables at the PFT, the land unit and the grid level, about 4.4 TB in size, is available upon request.

4 | DATA OVERVIEW

In the subgrid framework, the temporal changes induced by LULCC are approximated by the spatial variations

between a base land tile and a perturbed land tile. For instance, the effects of deforestation can be quantified by differences in the physical states between the grass (perturbed state) and the tree (base state) land tile. To investigate the impact of urbanization, the urban and rural tiles are the perturbed state and base state respectively. Similarly, the lake climatic effect can be examined by comparing the lake tile with the nonlake (rural) tile, using grids that contain both the lake and nonlake tiles.

Here, we examine the seasonal variations of key variables and show the lake effect across latitudes. In the

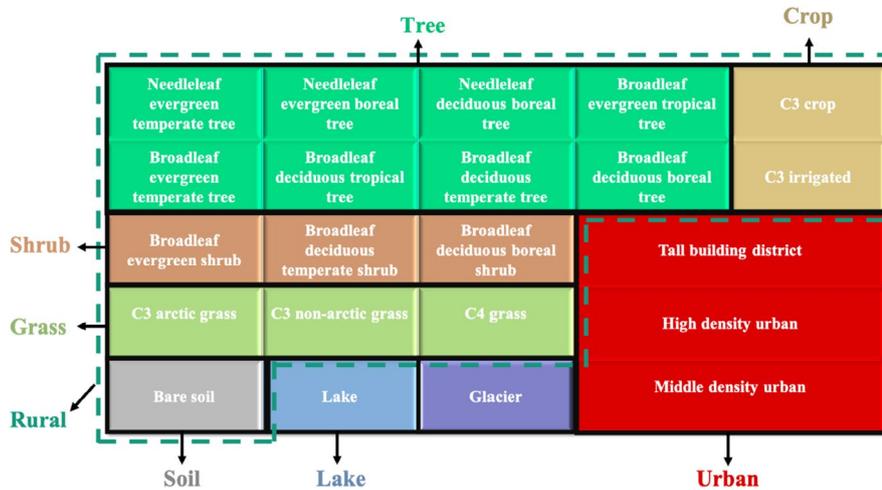


FIGURE 2 Schematic diagram illustrating the compositions of subgrid land tiles from subgrid patches

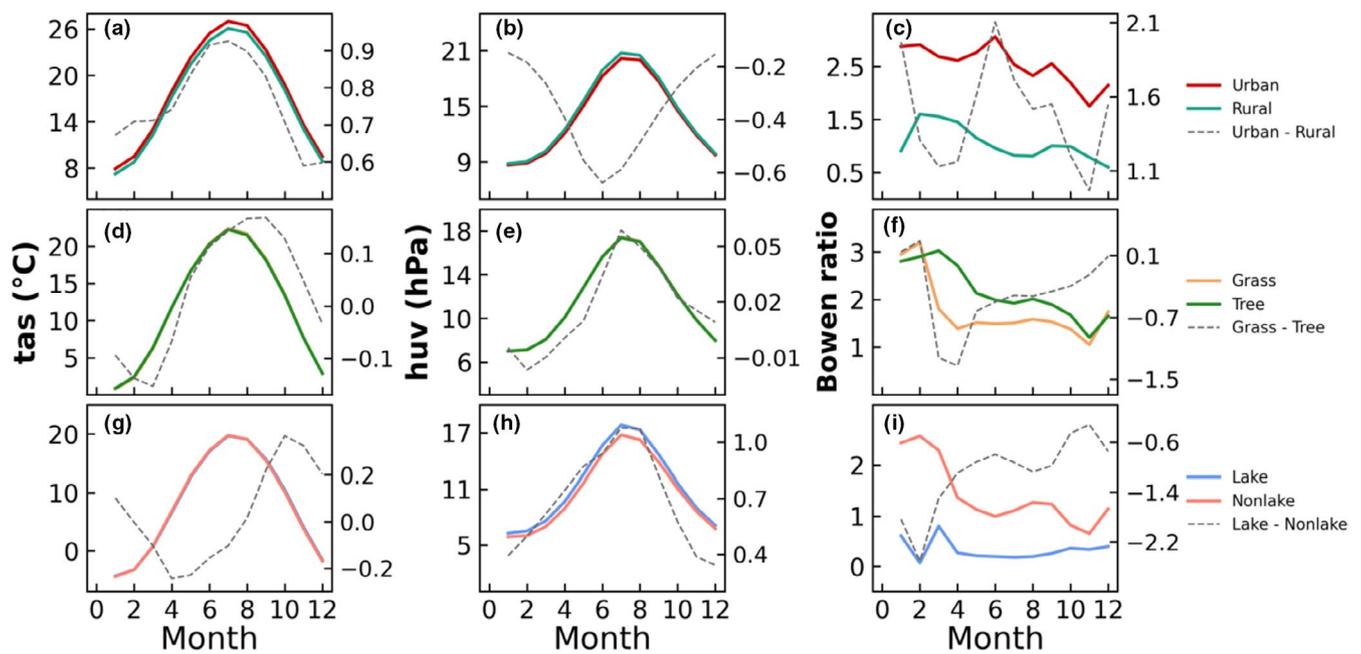


FIGURE 3 Seasonal variations of screen-height air temperature (tas), vapor pressure (huv) and Bowen ratio for urban and rural tiles (top row), grass and tree tiles (middle row) and lake and nonlake (rural) tiles (bottom row) in the Northern Hemisphere during 2015–2044. In each plot, the actual quantities are given by the left y-axis, and the difference between the two tiles is given by the right y-axis

Supporting Information, we present the diurnal cycles of these variables and an example to investigate the impact of urbanization on extreme temperatures by comparing the distributions of air temperature of the urban and rural tiles. All application examples use the data in the summer season, JJA (June–July–August) of the Northern Hemisphere and DJF (December–January–February) of the Southern Hemisphere. These results demonstrate the potential of this dataset to advance LULCC studies. Other potential applications include but not limited to attribution analysis on LULCC signals (Liao et al., 2020; Zhao et al., 2014) and the LULCC effect on runoff.

4.1 | Seasonal variations

Figure 3 shows the seasonal variations of screen-height (2-m) air temperature (tas), vapor pressure (huv) and surface Bowen ratio for three pairs of land tiles in the Northern Hemisphere during 2015–2044. In each plot, only grids that contain both tiles of interest are included. In Figure 3a–c, urban land tiles generally have higher air temperature, lower vapor pressure and higher Bowen ratio than rural land tiles. The annual urban-*versus*-rural temperature difference (0.59–0.92°C) is slightly lower than the global mean value (1.10°C) reported by Oleson et al. (2011), who used an earlier CLM version with overestimated

anthropogenic heat flux. The global temperature and vapor pressure contrasts between tree and grass tiles are minor, their differences being -0.15 to 0.17°C and -0.01 to 0.05 hPa respectively (Figure 3d–e). The higher Bowen ratio at tree tiles than the grass tiles indicates that the tree land cover on average partitions more available energy into sensible heat flux than grass (Figure 3f). These Bowen ratio contrasts are consistent with previous observational studies (Burakowski et al., 2018; Yamazaki et al., 2004). As for the lake effect, Figure 3g indicates that lakes have a slightly higher monthly air temperature than nonlake (rural) tiles from July to February and lower temperature during March to June. Due to higher lake evaporation rates, the vapor pressure is higher and the Bowen ratio is lower at lake tiles than at nonlake tiles (Figure 3h–i).

Figure 4 displays seasonal variations of the energy fluxes of the same three pairs of land tiles in the Northern Hemisphere. Note that the ground heat flux is defined as the heat flux into ground, buildings or lake but excluding the anthropogenic heat flux ($hf_{dsl} - f_{ahac} + f_{ahsh}$; Supporting Information). The net radiation flux ($rs_{ds} + r_{lds} - rs_{us} - rl_{us}$) and the ground heat flux are positive if the energy is going towards the surface, and a positive sensible heat flux (hf_{ss}) or latent heat flux (hf_{ls}) indicates that the flux is going upward towards the atmosphere. The anthropogenic heat flux is the sum of waste heat from AC

and heating appliances (f_{waste}) and the heat generated by space heating (f_{ahsh}). Globally, urbanization increases the sensible heat flux and decreases the latent heat flux, especially during JJA (Figure 4c). Deforestation leads to a reduction of net radiation, sensible and latent heat fluxes (Figure 4f), in broad agreement with previous observational (Duveiller et al., 2018) and modelling studies (Boysen et al., 2020). On average, lake tiles have higher net radiation and latent heat fluxes and lower sensible heat flux than nonlake land tiles (Figure 4i) because of low lake albedo and unlimited water availability. The “ground heat flux” of lakes is higher than that of nonlake tiles from January to August, and the opposite is true for other months when the surface of some lakes is frozen. Here, the heat flux into the lake is the sum of heat conduction and shortwave radiation transmission into the lake (Subin et al., 2012).

4.2 | Lake effects on air temperature

The air temperature difference between lake and nonlake tiles during the daytime and night-time as well as its changes at the end of the 21st century are shown in Figure 5. The daytime period is 8 am to 4 pm local time and the night-time period is 8 pm to 4 am local time. During the day, lake tiles south of 60°N have lower temperature

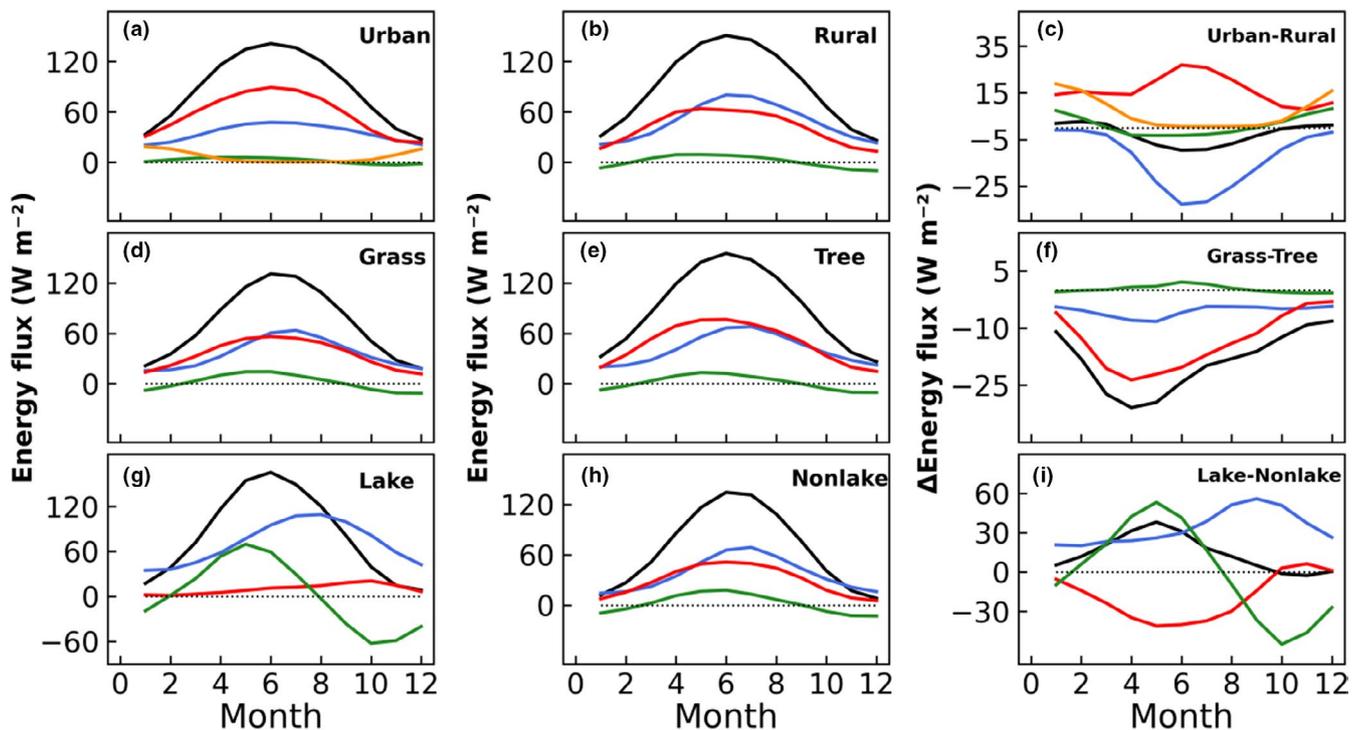


FIGURE 4 Seasonal variations of energy flux variables for perturbed land tiles (a, urban; d, grass; g, lake) and the corresponding base land tiles (b, rural; e, tree; h, nonlake) in the Northern Hemisphere during 2015–2044. Panels (c), (f) and (i) display the differences between the paired tiles. Black lines: net radiation flux; blue lines: latent heat flux; red lines: sensible heat flux; yellow lines: anthropogenic heat flux; green lines: ground heat flux

than nonlake titles in the same grid cell, with a difference of 0 to 2.5°C (Figure 5a). This lake cooling effect is the most prominent (2.1°C to 2.5°C) near 30°N and 30°S. In high-latitude regions north of 60°N, lake tiles have slightly higher temperature, by 0 to 0.35°C, than nonlake tiles. At night, the lake warming effect peaks near the equator and at 50°N (Figure 5b).

The difference in the daytime lake *versus* nonlake temperature contrast between current climate (2019–2023) and the end of the century (2096–2100) is negative in most regions (Figure 5c), indicating a stronger lake cooling effect in the future. One interpretation is that lake tiles warm up more slowly than nonlake tiles over time, similar to the results obtained by Wang et al. (2018) using an earlier version of CLM under the RCP 8.5 scenario. According to Wang et al. (2018), the mechanism underlying the different warming rates of lake and land is that the lake evaporation increases more than land evaporation, due to the changes in surface energy allocation in a warmer climate. In contrast, the night-time lake warming effect generally becomes weaker in the future, as indicated by the negative time change values, except for some lakes located in Southern Canada and Siberia (Figure 5d).

5 | DISCUSSION

5.1 | Comparison with the LUMIP data depository

Recently, the Land Use Model Intercomparison Project (LUMIP), a project endorsed by CMIP6, has included a data request of some key variables for the Tier 1 experiments at (up to) four subgrid land tiles: primary and secondary land, crop, pasture and urban at monthly intervals (Lawrence et al., 2016). LUMIP represents the first attempt to create datasets on subgrid variables. Currently, six modelling groups have deposited subgrid monthly mean data in a public depository (<https://esgf-node.llnl.gov/search/cmip6/>; accessed on 25 June 2021).

Our dataset was generated with a protocol similar to LUMIP, and offers three enhancements. First, the number of land tiles in this dataset is increased to eight (Figure 2), allowing more options than the LUMIP archive for investigation of climate effects of local LULCC.

Second, hourly and daily data are available in addition to monthly means. The daily data are helpful for investigating the impact of LULCC on extreme temperatures (e.g. Figure S7). Hourly data are desirable for heat stress studies. Heat stress indices cannot be accurately calculated from monthly or daily air temperature and humidity because they are nonlinearly related to temperature and humidity. For example, the wet-bulb temperature computed

from monthly or 6-h data is systematically overestimated by 0.5 to 1.5°C (Buzan et al., 2015). Additionally, biophysical processes are highly dynamic over the course of the 24-hr cycle. The hourly data permit a mechanistic examination of how the surface climate responds differently to land perturbation during the day and at night.

Third, a comprehensive set of surface data and atmospheric forcing variables is archived in this dataset. Our surface data provide information on urban (emissivity, fractions of road, roof and wall), the height of canopy top and bottom, soil texture and lake depth, which can help further the investigation. The atmospheric forcing variables are particularly useful in efforts to attribute local climate responses to different biophysical drivers (Lee et al., 2011; Rigden & Li, 2017; Zhao et al., 2014).

Finally, the results archived by LUMIP were produced with historical greenhouse gas forcing. The present study presents the results for a future forcing scenario (RCP8.5). When used together, these datasets allow investigation on how atmospheric CO₂ may influence subgrid variations.

5.2 | Limitations and known issues

We wish to draw the reader's attention to several limitations and known issues of this dataset. First, although CESM2 can capture the climatology, seasonal and inter-annual patterns of most fields (Collier et al., 2018; Fasullo, 2020; Lawrence et al., 2019), biases still exist in surface relative humidity, surface upward longwave radiation, anthropogenic heat flux and the energy partitions between latent and sensible heat fluxes (Chen et al., 2018; Cheng et al., 2021; Oleson & Feddema, 2020). The anthropogenic heat flux scheme in CLM5 omits anthropogenic heat sources such as traffic and industrial production, resulting in underestimation. The equilibrium climate sensitivity, defined as the global surface temperature increase after a doubling of CO₂ concentration, is 5.3 K for CESM2 (Danabasoglu et al., 2020), which is higher than the best estimate of 1.5–4.5 K. Thus, the future climate change projected by CESM2 may be too strong. Previous studies also indicate that the regional bias of CESM2 may vary for different regions of interest (Cheng et al., 2021; Collier et al., 2018). We urge caution when interpreting the regional results from this dataset.

Second, the user may run the risk of double-counting some of the energy terms if a clear distinction is not made between the anthropogenic heat flux, heat storage and the ground heat flux. The surface energy balance equation is expressed as (variable definitions in Table 1):

$$rds + rlds - rsus - rlus + fahac + fwaste = hfsl + hfss + hfdsl \quad (1)$$

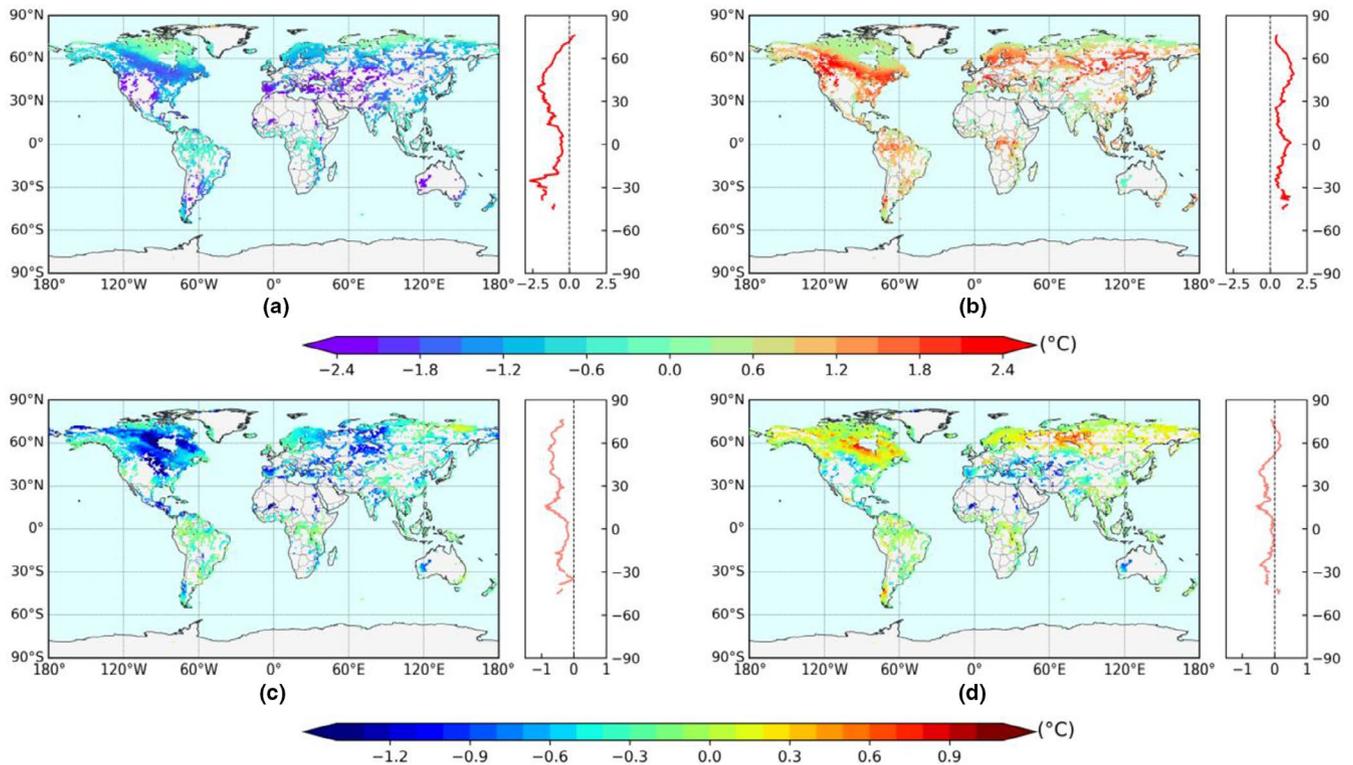


FIGURE 5 The lake minus nonlake air temperature during (a) the daytime and (b) the night-time in 2019–2023. The changes of the lake minus nonlake air temperature between 2096–2100 and 2019–2023 during (c) the daytime and (d) the night-time

For nonurban tiles, $fahac$ and $fwaste$ are zero. For lake tiles, $hfdsl$ includes both heat conduction and shortwave radiation transmission into the water column. In an urban system, $hfdsl$ has incorporated $fahac$ and $fahsh$. To obtain a “natural” ground heat flux excluding AC and space heating contributions (Supporting Information), Equation 1 can be rearranged as:

$$\begin{aligned} rds + rlds - rsus_u - rlus_u + fwaste_u + fahsh_u \\ = hfsl_u + hfss_u + (hfdsl_u - fahac_u + fahsh_u) \end{aligned} \quad (2)$$

The subscript u means they are all on urban land tiles. Here, $fwaste_u + fahsh_u$ is the total anthropogenic heat flux entering the climate system, and $hfdsl_u - fahac_u + fahsh_u$ is the “natural” ground heat flux or the heat flux into ground and buildings but excluding anthropogenic heat flux.

Third, many ESMs (including CESM) have difficulty reproducing the response of the diurnal air temperature cycle to deforestation (Lejeune et al., 2017). The diurnal temperature range (DTR) is smaller in grass tiles than in forest (tree) tiles in this dataset (Figure S5d), but observational studies show larger DTR over grasses than over forests (Lee et al., 2011; Zhang et al., 2014). In other words, the daytime and night-time air temperature difference between grass and forest tiles is negative and positive, respectively, according to CLM5 (Figure S5d),

but the opposite is true according to observations. This inconsistency may be associated with parameterization issues in current land models, such as roughness formulation, the response of partitioning of available energy between latent and sensible heat fluxes to LULCC and the derivation of the screen-height air temperature (Chen & Dirmeyer, 2019; Chen et al., 2018; Lejeune et al., 2017).

Fourth, the subgrid strategy only quantifies the direct biophysical impact of LULCC at the local scale, and cannot capture the indirect impacts associated with LULCC, including the impact caused by changes in atmospheric conditions and water-energy balances at regional and global scales. In addition, the carbon cycle feedback via changing land use and dynamic vegetation phenology, another large-scale impact, is not considered. Previous studies found that the local effects are very similar between simulations with different LULCC scenarios (Winckler et al., 2017) and between fixed land-use simulations and satellite observations (Chen & Dirmeyer, 2020). In other words, LULCC local effects can be investigated independently of the LULCC-induced regional large-scale changes.

Finally, atmospheric CO_2 and transient land cover may affect subgrid variations. Some of the CO_2 effect can be investigated by comparing subgrid results at the beginning and at the end of the RCP 8.5 simulation period and with

the results under the historical scenario in the LUMIP data depository. Simulations of other scenarios (e.g. RCP2.6, RCP4.5, $4 \times \text{CO}_2$) and with transient land cover and dynamic vegetation phenology will be helpful in evaluating scenario uncertainties. It is our future plan to conduct these simulations and publish the subgrid datasets.

6 | SUMMARY

In comparison with the subgrid data in the LUMIP depository, our dataset is enhanced in three respects: (1) We provide data for eight land tiles (urban, rural, tree, grass, shrub, bare soil, crop and lake) instead of four; (2) Biophysical variables are available at high frequencies (hourly and daily) in addition to the monthly interval; (3) A comprehensive list of surface data and atmospheric forcing variables is included. These features allow for wider investigations of the relationships between LULCC and local climate at multiple time scales and across climate zones.

The seasonal variations of key state variables and energy fluxes at subgrid tiles are in broad agreement with those in published observational and modelling studies. Results show that the future daytime lake cooling effect becomes stronger and the night-time warming effect becomes weaker than under the current climate. We anticipate that this dataset will become a useful resource for meteorologists, hydrologists and ecologists to investigate the interplay between surface climate, ecosystem functions and climate change.

ACKNOWLEDGEMENTS

We would like to acknowledge high-performance computing support from Cheyenne (<https://doi.org/10.5065/D6RX99HX>) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation. The Community Earth System Model Version 2 is freely available at <https://www.cesm.ucar.edu/models/cesm2/>. The modified code and raw output of the simulation in this paper are available from authors upon request (keer.zhang@yale.edu).

CONFLICTS OF INTEREST

The authors declare that they have no conflict of interest.

AUTHOR CONTRIBUTIONS

X.L. and N.S. designed this study. K.Z. performed the simulation and data processing. L.Z. contributed ideas to the model simulation. C.H., Q.H. and Z.L. conducted the urban land cover projection. H.C. and J.Z. generated the new surface data for the model simulation. X.L. and K.Z. drafted the manuscript.

OPEN PRACTICES



This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at <https://doi.org/10.7910/DVN/HUXAH6>. Learn more about the Open Practices badges from the Center for OpenScience: <https://osf.io/tvyxz/wiki>.

ORCID

Keer Zhang <https://orcid.org/0000-0001-9966-7566>

REFERENCES

- Alkama, R. & Cescatti, A. (2016) Biophysical climate impacts of recent changes in global forest cover. *Science*, 351, 600–604.
- Boysen, L.R., Brovkin, V., Pongratz, J., Lawrence, D.M., Lawrence, P., Vuichard, N. et al. (2020) Global climate response to idealized deforestation in CMIP6 models. *Biogeosciences*, 17, 5615–5638.
- Burakowski, E., Tawfik, A., Ouimette, A., Lepine, L., Novick, K., Ollinger, S. et al. (2018) The role of surface roughness, albedo, and Bowen ratio on ecosystem energy balance in the Eastern United States. *Agricultural & Forest Meteorology*, 249, 367–376.
- Buzan, J.R., Oleson, K. & Huber, M. (2015) Implementation and comparison of a suite of heat stress metrics within the Community Land Model version 4.5. *Geoscientific Model Development*, 8, 151–170.
- Chen, G., Li, X., Liu, X., Chen, Y., Liang, X., Leng, J. et al. (2020) Global projections of future urban land expansion under shared socioeconomic pathways. *Nature Communications*, 11, 1–12.
- Chen, L. & Dirmeyer, P.A. (2019) Differing responses of the diurnal cycle of land surface and air temperatures to deforestation. *Journal of Climate*, 32, 7067–7079.
- Chen, L. & Dirmeyer, P.A. (2020) Reconciling the disagreement between observed and simulated temperature responses to deforestation. *Nature Communications*, 11, 1–10.
- Chen, L., Dirmeyer, P.A., Guo, Z. & Schultz, N.M. (2018) Pairing FLUXNET sites to validate model representations of land-use/land-cover change. *Hydrology and Earth System Sciences*, 22, 111–125.
- Cheng, Y., Huang, M., Zhu, B., Bisht, G., Zhou, T., Liu, Y. et al. (2021) Validation of the community land model version 5 over the contiguous United States (CONUS) using in situ and remote sensing data sets. *Journal of Geophysical Research: Atmospheres*, 126, 1–27.
- Collier, N., Hoffman, F.M., Lawrence, D.M., Keppel-aleks, G., Koven, C.D., Riley, W.J. et al. (2018) The international land model benchmarking (ILAMB) system: Design, theory, and implementation. *Journal of Advances in Modeling Earth Systems*, 10, 2731–2754.
- Danabasoglu, G., Lamarque, J.F., Bacmeister, J., Bailey, D.A., DuVivier, A.K., Edwards, J. et al. (2020) The community earth system model version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, 12, 1–35.
- Davin, E.L., Rechid, D.I., Breil, M., Cardoso, R.M., Coppola, E., Hoffmann, P. et al. (2020) Biogeophysical impacts of forestation in Europe: First results from the LUCAS (Land Use and Climate across Scales) regional climate model intercomparison. *Earth System Dynamics*, 11, 183–200.

- Duveiller, G., Hooker, J. & Cescatti, A. (2018) The mark of vegetation change on Earth's surface energy balance. *Nature Communications*, 9, 1–12.
- Fasullo, J.T. (2020) Evaluating simulated climate patterns from the CMIP archives using satellite and reanalysis datasets using the Climate Model Assessment Tool (CMATv1). *Geoscientific Model Development*, 13, 3627–3642.
- Feddema, J.J., Oleson, K.W., Bonan, G.B., Mearns, L.O., Buja, L.E., Meehl, G.A. et al. (2005) The importance of land-cover change in simulating future climates. *Science*, 310, 1674–1678.
- Findell, K.L., Berg, A., Gentile, P., Krasting, J.P., Lintner, B.R., Malyshev, S. et al. (2017) The impact of anthropogenic land use and land cover change on regional climate extremes. *Nature Communications*, 8, 1–9.
- Gao, J. & O'Neill, B.C. (2020) Mapping global urban land for the 21st century with data-driven simulations and Shared Socioeconomic Pathways. *Nature Communications*, 11, 1–12.
- He, C., Liu, Z., Wu, J., Pan, X., Fang, Z., Li, J. et al. (2021) Future global urban water scarcity and potential solutions. *Nature Communications*, 12, 1–11.
- Huang, Q., Zhang, H., Vliet, J., Ren, Q., Wang, R.Y., Du, S. et al. (2021) Patterns and distributions of urban expansion in global watersheds. *Earth's Future*, 9, 1–16.
- Jackson, T.L., Feddema, J.J., Oleson, K.W., Bonan, G.B. & Bauer, J.T. (2010) Parameterization of urban characteristics for global climate modeling. *Annals of the Association of American Geographers*, 100, 848–865.
- Lawrence, D.M., Fisher, R.A., Koven, C.D., Oleson, K.W., Swenson, S.C., Bonan, G. et al. (2019) The community land model version 5: description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, 11, 4245–4287.
- Lawrence, D.M., Hurtt, G.C., Arneth, A., Brovkin, V., Calvin, K.V., Jones, A.D. et al. (2016) The Land Use Model Intercomparison Project (LUMIP) contribution to CMIP6: Rationale and experimental design. *Geoscientific Model Development*, 9, 2973–2998.
- Lawrence, P.J. & Chase, T.N. (2010) Investigating the climate impacts of global land cover change in the community climate system model. *International Journal of Climatology*, 30, 2066–2087.
- Lawrence, P.J., Feddema, J.J., Bonan, G.B., Meehl, G.A., O'Neill, B.C., Oleson, K.W. et al. (2012) Simulating the biogeochemical and biogeophysical impacts of transient land cover change and wood harvest in the Community Climate System Model (CCSM4) from 1850 to 2100. *Journal of Climate*, 25, 3071–3095.
- Lee, X., Goulden, M.L., Hollinger, D.Y., Barr, A., Black, T.A., Bohrer, G. et al. (2011) Observed increase in local cooling effect of deforestation at higher latitudes. *Nature*, 479, 384–387.
- Lejeune, Q., Seneviratne, S.I. & Davin, E.L. (2017) Historical land-cover change impacts on climate: Comparative assessment of LUCID and CMIP5 multimodel experiments. *Journal of Climate*, 30, 1439–1459.
- Li, X., Zhou, Y., Eom, J., Yu, S. & Asrar, G.R. (2019) Projecting global urban area growth through 2100 based on historical time series data and future shared socioeconomic pathways. *Earth's Future*, 7, 351–362.
- Liao, W., Liu, X., Burakowski, E., Wang, D., Wang, L. & Li, D. (2020) Sensitivities and responses of land surface temperature to deforestation-induced biophysical changes in two global earth system models. *Journal of Climate*, 33, 8381–8399.
- Mahmood, R., Pielke, R.A., Hubbard, K.G., Niyogi, D., Dirmeyer, P.A., Mcalpine, C. et al. (2014) Land cover changes and their biogeophysical effects on climate. *International Journal of Climatology*, 34, 929–953.
- Malyshev, S., Shevliakova, E., Stouffer, R.J. & Pacala, S.W. (2015) Contrasting local versus regional effects of land-use-change-induced heterogeneity on historical climate: Analysis with the GFDL earth system model. *Journal of Climate*, 28, 5448–5469.
- Meier, R., Davin, E.L., Lejeune, Q., Hauser, M., Li, Y., Martens, B. et al. (2018) Evaluating and improving the Community Land Model's sensitivity to land cover. *Biogeosciences*, 15, 4731–4757.
- Oke, T.R. (1987) *Boundary layer climates*, 2nd edition. New York: Routledge.
- Oleson, K.W., Bonan, G.B. & Feddema, J. (2010) Effects of white roofs on urban temperature in a global climate model. *Geophysical Research Letters*, 37, 1–7.
- Oleson, K.W., Bonan, G.B., Feddema, J. & Jackson, T. (2011) An examination of urban heat island characteristics in a global climate model. *International Journal of Climatology*, 31, 1848–1865.
- Oleson, K.W. & Feddema, J. (2020) Parameterization and surface data improvements and new capabilities for the community land model urban (CLMU). *Journal of Advances in Modeling Earth Systems*, 12, 1–30.
- Pitman, A.J., De Noblet-Ducoudré, N., Cruz, F.T., Davin, E.L., Bonan, G.B., Brovkin, V. et al. (2009) Uncertainties in climate responses to past land cover change: First results from the LUCID inter-comparison study. *Geophysical Research Letters*, 36, 1–6.
- Rigden, A.J. & Li, D. (2017) Attribution of surface temperature anomalies induced by land use and land cover changes. *Geophysical Research Letters*, 44, 6814–6822.
- Schultz, N.M., Lawrence, P.J. & Lee, X. (2017) Global satellite data highlights the diurnal asymmetry of the surface temperature response to deforestation. *Journal of Geophysical Research: Biogeosciences*, 122, 903–917.
- Schultz, N.M., Lee, X., Lawrence, P.J., Lawrence, D.M. & Zhao, L. (2016) Assessing the use of subgrid land model output to study impacts of land cover change. *Journal of Geophysical Research: Atmospheres*, 121, 6133–6147.
- Scott, A.A., Waugh, D.W. & Zaitchik, B.F. (2018) Reduced Urban Heat Island intensity under warmer conditions. *Environmental Research Letters*, 13, 1–9.
- Subin, Z.M., Riley, W.J. & Mironov, D. (2012) An improved lake model for climate simulations: Model structure, evaluation, and sensitivity analyses in CESM1. *Journal of Advances in Modeling Earth Systems*, 4, 1–27.
- Tang, B., Zhao, X. & Zhao, W. (2018) Local effects of forests on temperatures across Europe. *Remote Sensing*, 10, 1–24.
- Wang, W., Lee, X., Xiao, W., Liu, S., Schultz, N., Wang, Y. et al. (2018) Global lake evaporation accelerated by changes in surface energy allocation in a warmer climate. *Nature Geoscience*, 11, 410–414.
- Winckler, J., Reick, C.H. & Pongratz, J. (2017) Robust identification of local biogeophysical effects of land-cover change in a global climate model. *Journal of Climate*, 30, 1159–1176.
- Yamazaki, T., Yabuki, H., Ishii, Y., Ohta, T. & Ohata, T. (2004) Water and energy exchanges at forests and a grassland in eastern

Siberia evaluated using a one-dimensional land surface model. *Journal of Hydrometeorology*, 5, 504–515.

- Zhang, K., Lee, X., Schultz, N.M., Zhao, L., He, C., Huang, Q. et al. (2021) *A global dataset on subgrid land surface climate (2015–2100) from the Community Earth System Model, Harvard Dataverse*. <https://doi.org/10.7910/DVN/HUXAH6>
- Zhang, M., Lee, X., Yu, G., Han, S., Wang, H., Yan, J. et al. (2014) Response of surface air temperature to small-scale land clearing across latitudes. *Environmental Research Letters*, 9(3), 034002.
- Zhao, L., Lee, X. & Schultz, N.M. (2017) A wedge strategy for mitigation of urban warming in future climate scenarios. *Atmospheric Chemistry and Physics*, 17, 9067–9080.
- Zhao, L., Lee, X., Smith, R.B. & Oleson, K. (2014) Strong contributions of local background climate to urban heat islands. *Nature*, 511, 216–219.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Zhang, K., Lee, X., Schultz, N.M., Zhao, L., He, C., Huang, Q., et al (2022) A global dataset on subgrid land surface climate (2015–2100) from the Community Earth System Model. *Geoscience Data Journal*, 00, 1–12. Available from: <https://doi.org/10.1002/gdj3.153>