Geoscientific
Model Development

Open Access

Enhancing the representation of subgrid land surface characteristics in land surface models

Y. Ke[1,2], L. R. Leung[2], M. Huang[2], and H. Li[2]

[1]Base of the State Key Laboratory of Urban Environment Process and Digital Modelling,
Department of Resource Environment and Tourism, Capital Normal University, 105 Xi San Huan Bei Lu, Beijing,
100048, China
[2]Pacific Northwest National Laboratory, 902 Battelle Boulevard, Richland, WA 99352, USA

Correspondence to: L. R. Leung (ruby.leung@pnnl.gov)

Abstract. Land surface heterogeneity has long been recognized as important to represent in the land surface models. In most existing land surface models, the spatial variability of surface cover is represented as subgrid composition of multiple surface cover types, although subgrid topography also has major controls on surface processes. In this study, we developed a new subgrid classification method (SGC) that accounts for variability of both topography and vegetation cover. Each model grid cell was represented with a variable number of elevation classes and each elevation class was further described by a variable number of vegetation types optimized for each model grid given a predetermined total number of land response units (LRUs). The subgrid structure of the Community Land Model (CLM) was used to illustrate the newly developed method in this study. Although the new method increases the computational burden in the model simulation compared to the CLM subgrid vegetation representation, it greatly reduced the variations of elevation within each subgrid class and is able to explain at least 80 % of the total subgrid plant functional types (PFTs). The new method was also evaluated against two other subgrid methods (SGC1 and SGC2) that assigned fixed numbers of elevation and vegetation classes for each model grid (SGC1: M elevation bands–N PFTs method; SGC2: N PFTs–M elevation bands method). Implemented at five model resolutions (0.1°, 0.25°, 0.5°, 1.0° and 2.0°) with three maximum-allowed total number of LRUs (i.e., N_LRU of 24, 18 and 12) over North America (NA), the new method yielded more computationally efficient subgrid representation compared to SGC1 and SGC2, particularly at coarser model resolutions

and moderate computational intensity ($N\_LRU = 18$). It also explained the most PFTs and elevation variability that is more homogeneously distributed spatially. The SGC method will be implemented in CLM over the NA continent to assess its impacts on simulating land surface processes.

1 Introduction

As the terrestrial component of earth system models, land surface models play important roles in representing the interactions between terrestrial biosphere and atmosphere, which is important for predicting future states of the earth system and assessing anthropogenic impacts on the climate system. Using land surface parameters and meteorological forcing data as input, land surface models simulate key land processes such as photosynthesis, respiration, and evapotranspiration that regulate mass, energy, moisture, and momentum exchange between soil, vegetation and atmosphere. Realistic and high spatial resolution representation of land surface characteristics is important for accurate estimation of surface hydrology, heat fluxes, and surface CO_2 exchanges in climate models for applications across global, regional, and sub-regional scales.

In recent years, numerous efforts have improved the representation of land surface characteristics either by developing higher-resolution land cover datasets (e.g., Bonan et al., 2002a, b; Lawrence and Chase, 2007; Ke et al., 2012) or by representing subgrid spatial heterogeneity of land surface parameters (e.g., Koster and Suarez, 1992; Seth et al., 1994).

Although land surface parameters have been developed with spatial resolution as fine as 5 km globally (Ke et al., 2012), climate models cannot explicitly resolve land surface heterogeneity at fine scales because of computational constraints on the grid resolutions that can be achieved. Therefore, representations of subgrid land surface heterogeneity are still needed and employed in many land surface models. It has been well established that subgrid spatial variability in land surface characteristics such as vegetation cover and topography can significantly affect the estimation of surface evapotranspiration, runoff, soil moisture, surface albedo, snowpack, and other fluxes (Koster and Suarez, 1992; Seth et al., 1994; Giorgi and Avissar, 1997; Ghan et al., 1997; Giorgi et al., 2003; Li and Arora, 2012; Li et al., 2013).

Vegetation plays a key role in land surface water and energy partitioning and carbon cycle. Current land surface models widely adopt the concept of plant functional type (PFT) to describe vegetation distributions (Oleson and Bonan, 2000; Krinner et al., 2005, Ek et al., 2003; Sitch et al., 2003; Niu et al., 2011). For example, the Noah land surface model incorporates 13 PFTs (Ek et al., 2003), CLM has 15 PFTs (Oleson et al., 2010), and ORCHIDEE distinguishes 12 PFTs (Krinner et al., 2012). While some models such as Noah represent a single dominant PFT in each model grid, models such as CLM represent subgrid spatial heterogeneity of vegetation distribution with a composition of multiple PFTs coexisting within each model grid. This representation assumes that all plants of the same type cluster as a "tile" within a model grid and a single atmospheric forcing is assigned to the tiles within a model grid.

In addition to horizontal landscape variability, spatial heterogeneity in topography is also a pronounced land surface characteristic and has been considered in some land surface simulations to help parameterize topographic variability in precipitation, temperature and snow processes (Leung and Ghan, 1995; Nijssen et al., 2001; Giorgi et al., 2003). The Variable Infiltration Capacity (VIC) model is one example of land surface models that divide a model grid cell into multiple subgrid elevation bands with 500 m elevation interval to achieve improved simulations of surface hydrology (Nijssen et al., 2001). Leung and Ghan (1995, 1998) developed a subgrid parameterization to incorporate the influence of topography on precipitation and snow cover and reported improved simulations in a regional climate model that used the subgrid parameterization with an explicit grid resolution of 90 km compared to simulations performed at a finer resolution of 30 km but without the subgrid parameterization, although the latter is much more computationally demanding.

The subgrid parameterization of Leung and Ghan (1998) is one of few that incorporate the joint distribution of both vegetation and topography parameters. Taking advantage of the statistical relationship between topography and vegetation, their method classifies the topography within each model grid into subgrid elevation bands with predetermined elevation interval and parameterizes subgrid vegetation variability by considering vegetation distribution in each elevation band. This subgrid scheme allows a different atmospheric condition to be assigned to each elevation band. For example, elevation bands corresponding to higher elevation have cooler near-surface air temperature and increased precipitation compared to the grid cell mean values. Applying such atmospheric forcing to different vegetation classes within the same elevation class simulates surface fluxes and soil hydrology that reflect the influence of atmospheric forcing at the higher elevation on land surface processes, leading to improved land surface simulations for the specific subgrid elevation/vegetation class as well as the overall grid cell averaged conditions. With only one dominant vegetation type considered for each subgrid elevation class in the study, 67 % of the total vegetation was explained over the study area; the approach resulted in improved surface temperature simulation compared to simulations without considering topography effect.

With the existing subgrid method in CLM4, each subgrid PFT within a model grid is a computational unit, while with the subgrid method presented in Leung and Ghan (1998) each subgrid vegetation–elevation class is a computational unit. Although the subgrid scheme by Leung and Ghan (1998) can be improved by representing each elevation band with distribution of multiple vegetation types, the larger number of subgrid elevation/vegetation classes within each model grid can greatly increase the computational burden because land surface processes are explicitly simulated for each computational unit.

This study generalizes the method of Leung and Ghan (1998) and aims to develop an improved and computationally efficient subgrid scheme based on high-resolution satellite-based land cover and topography products in order to enhance the representation of both vegetation cover and topography. We chose the subgrid structure of CLM as an example and used the PFTs defined in CLM to represent vegetation. The new subgrid PFT scheme assigned a flexible number of elevation bands and PFTs for each model grid and was optimized to explain a maximal amount of elevation and vegetation variations in a computationally efficient manner. The method is applied to North America (NA) at different model resolutions and evaluated by comparing it with other subgrid methods that use a fixed number of subgrid elevation bands and vegetation types.

2   Method

2.1   CLM vegetation representation

CLM is a land model within the Community Earth System Model (CESM). It has been widely applied at continental to global scales to understand the impact of land processes on climate change (Oleson et al., 2010). The spatial heterogeneity of land surface parameters in CLM is represented using
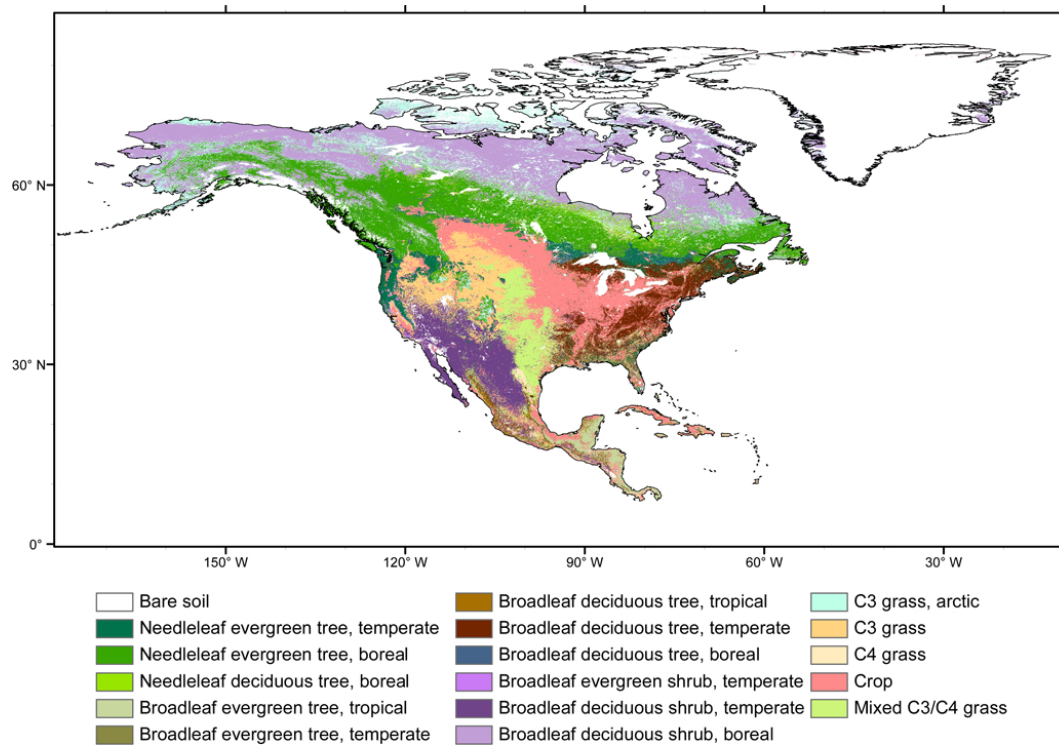
Fig. 1. North America PFT map. *Area with legend "mixed C3/C4 grass" means the fraction of either C3 or C4 grass in each pixel is less than 1.

a nested subgrid hierarchy. Each grid cell is composed of a different number of land units including glacier, lake, wetland, and urban and vegetated surfaces. Vegetated surfaces are represented with a composition of 15 possible PFTs such as temperate needleleaf evergreen trees, temperate broadleaf evergreen trees, etc., plus bare ground. In the current version of CLM (CLM 4.0), the PFT data are available at 0.5° and 0.05° resolutions (Lawrence and Chase, 2007; Lawrence et al., 2011; Ke et al., 2012) and the spatial distribution of each PFT within a model grid is not explicitly represented.

2.2 Plant functional types mapping

The PFT map for North America was generated at 500 m resolution based on the MODIS land cover product and climate data following the method presented in Ke et al. (2012). Briefly, seven PFTs, including needleleaf evergreen trees, needleleaf deciduous trees, broadleaf evergreen trees, broadleaf deciduous trees, shrubs, grasses and crop, were directly determined from the MODIS MCD12Q1 C5 PFT classifications for each 500 m pixel. The WorldClim 5 arc-min (0.0833°) (Hijmans et al., 2005) climatological global monthly surface air temperature and precipitation data were interpolated to the 500 m grids, and the climate rules described by Bonan et al. (2002a) were used to reclassify the 7 PFTs into 15 PFTs in the tropical, temperate and boreal climate groups. Similar to Lawrence and Chase (2007),

the fractions of C3 and C4 grasses were mapped based on the method presented in Still et al. (2003). Pixels with barren land and urban areas were reassigned to the bare soil class. Figure 1 shows the 500 m-resolution PFT map for NA.

2.3 Digital elevation model

The HYDRO1k digital elevation model (DEM) for NA was used to generate elevation data. The HYDRO1k is a comprehensive and consistent geographic database providing global coverage of topographically derived datasets such as elevation, slope, aspect, flow accumulation raster layers, and stream lines vector layers. All raster datasets were generated from the USGS 30 arc-second global digital elevation model (GTOPO30) at 1 km resolution, and covered all global landmasses with the exception of Antarctica and Greenland. (http://eros.usgs.gov/). Compared to other existing global-scale elevation datasets such as 90 m DEM of the Shuttle Radar Topography Mission (SRTM) (http://www2.jpl.nasa.gov/srtm/) and the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) 30 m Global Digital Elevation Model (GDEM) (http://asterweb.jpl.nasa.gov/), HYDRO1k DEM has more complete global coverage and its resolution is closer to the existing global-scale land cover data such as MODIS land cover product. Therefore, it has been widely used for continental and global hydrologic and land surface modeling and was also selected in our study. For
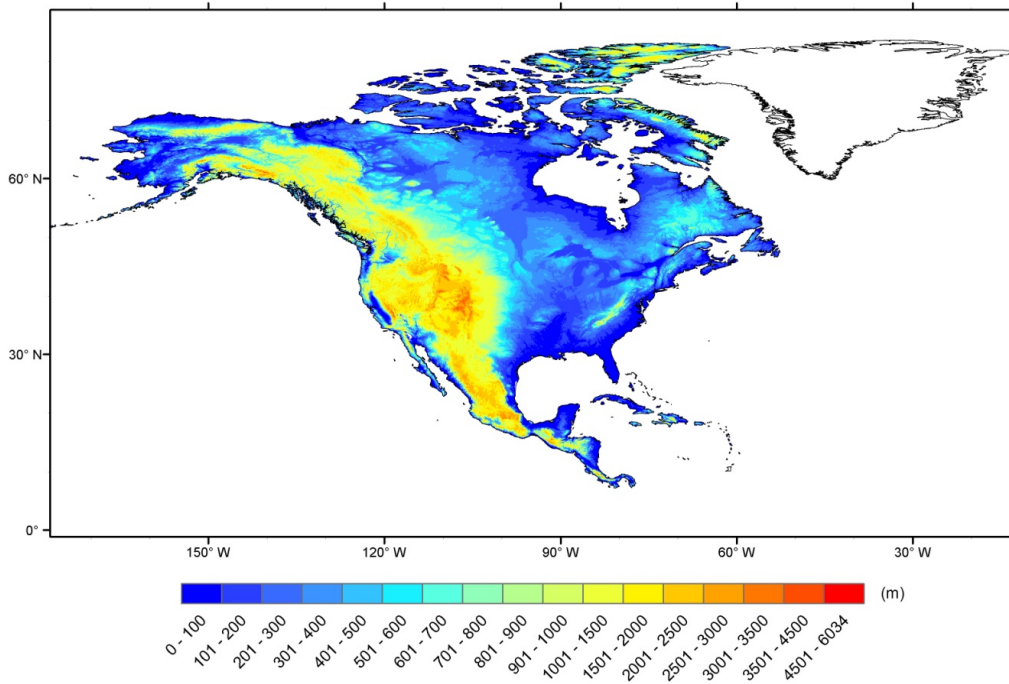
Fig. 2. Elevation distribution in North America.

consistency with the PFT map, the elevation raster layer for North America was bilinearly interpolated to 500 m resolution (Fig. 2). We excluded Greenland from our study because HYDRO1k does not cover this area.

2.4 Optimal subgrid classification (SGC) method of elevation and vegetation

The subgrid classification (SGC) method developed in our study considered the joint distribution of elevation and vegetation. Within each model grid (e.g., at resolution $0.1° \times 0.1°$), the SGC method first classified surface elevation from the 500 m DEM data into a limited number of elevation bands (or classes) of equal elevation range. A minimum area threshold of 1 % was used to limit the area of each elevation band. That is, an elevation band containing less than 1 % of the land area of the model grid is added to the neighboring elevation band so that each elevation band covers at least 1 % of the grid land area. Within each elevation band in each model grid, the area was further classified into a limited number of PFTs. For example, with M elevation bands within a model grid cell and N PFTs within each elevation band, the model grid was represented with a total number of $M \times N$ subgrid classes or land response units (LRUs), with each elevation–PFT LRU treated as a computational unit in the land surface model simulation. The schematic comparison between the new SGC method and CLM subgrid method is shown in Fig. 3.
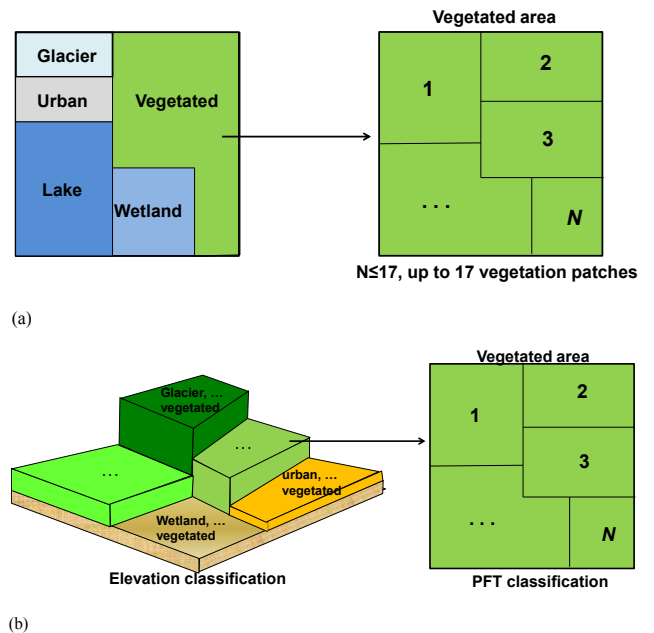

Fig. 3. Schematic diagram of subgrid classification method in **(a)** CLM 4.0 and **(b)** SGC method.

To restrict computational burden, we set the maximum-allowed total number of LRUs to "N_LRU" (e.g., 18 LRUs) for each model grid. The number of elevation bands M and the number of PFTs N for each elevation band are

variable for each model grid, but $M \times N$ should not exceed the maximum-allowed number N_LRU (e.g., 18). Hence the combination of M and N is variable and is chosen to best represent the subgrid variability of both PFT and elevation. For example, for $N\_LRU = 18$, possible combinations include 3 elevation bands and 6 PFTs per elevation band, or 2 elevation bands and 9 PFTs per elevation band, etc, but the optimal combination was selected. Two criteria must be satisfied for the optimal classification: (1) the elevation range of each elevation band is less than and close to 100 m; and (2) total percentage of subgrid PFTs correctly classified by the method is no less than 80 % for each model grid. We prioritized criteria (2) so that if none of the combinations satisfies both conditions, the classification explaining more than 80 % of PFTs and with elevation range greater than but closest to 100 m was selected; if more than one combination satisfy both conditions, the classification that correctly classifies the most subgrid PFTs was selected.

2.5 SGC method evaluation

In CLM 4.0, vegetation was represented as the composition of 15 PFTs plus bare soil. The simplest and least computationally intensive way to incorporate elevation distribution of vegetation is to assign a single elevation band to each PFT within a model grid; that is, the surface elevation in the area covered by each PFT was aggregated to one elevation band. We used this method as a baseline to assess the performance of the SGC at different model resolutions (0.1°, 0.25°, 0.5°, 1.0°, and 2.0°).

The SGC method was also evaluated by comparing it with two other subgrid classification methods based on a fixed number of elevation bands and vegetation types. The first subgrid classification method (SGC1) was the M elevation bands–N PFTs method. Each model grid cell is first divided into M equal-interval elevation bands and each elevation band was further classified into N PFTs. The second subgrid classification method (SGC2) was the N PFTs–M elevation bands method. Each model grid cell was first classified into N PFTs and the area covered by each PFT was further divided into M equal-interval elevation bands. For both methods, we used the minimum area threshold of 1 % to restrict the number of elevation bands in the same way that was used in the SGC method. The SGC1 and SGC2 methods take different perspectives of topography–vegetation distribution in that SGC1 examines the PFT distribution at different elevation bands and SGC2 examines the elevation distribution of each PFT. However, both methods classify the model grids into fixed numbers of elevation bands and PFTs throughout the study area.

We implemented the classification methods SGC, SGC1 and SGC2 in North America at 0.1°, 0.25°, 0.5°, 1.0° and 2.0° resolution with different combinations of number of elevation bands and vegetation types: (1) *Scheme 1:* $N\_LRU = 24$, $M = 6N = 4$, meaning 24 maximum-allowed

classes for the SGC method and the combination of 6 elevation bands and 4 PFTs for the SGC1 and SGC2 methods; (2) *Scheme 2:* $N\_LRU = 18M = 6N = 3$; and (3) *Scheme 3:* $N\_LRU = 12$, $M = 4N = 3$. The baseline subgrid method was also implemented at the five resolutions. The three schemes represent different computational burdens with Scheme 1 being most computationally intensive because it has the largest number of maximum-allowed total subgrid LRUs for each model grid.

The average number of total LRUs across NA was used to evaluate and compare the methods' computational burden. In the baseline method, LRUs correspond to PFTs within each model grid; in the SGC, SGC1 and SGC2 methods, LRUs correspond to elevation–PFT classes within each model grid. The total percentage of PFTs explained within each model grid (% PFT) and the mean standard deviation of elevation averaged over all elevation bands σ_{ep} were used to measure and compare the methods' abilities to characterize subgrid vegetation and topography (Eqs. 1–3).

$$\% PFT = \sum_{i=1}^{N_b} \left[PA_{eb}(i) \times \sum_{j=1}^{N_p} PA_{pft}(i, j) \right], \qquad (1)$$

where $PA_{eb}(i)$ is the percent area of the ith elevation band in that grid, $PA_{pft}(i, j)$ is the percent coverage of the jth dominant PFT in the ith elevation band, N_p is the number of dominant PFTs within each elevation band, and N_b is the number of elevation bands in that grid.

For the SGC and SGC1 methods, σ_{ep} at a given model grid was calculated as

$$\sigma_{ep} = \frac{1}{N_b} \sum_{i=1}^{N_b} \sigma_{eb}(i), \qquad (2)$$

where $\sigma_{eb}(i)$ is the standard deviation of subgrid surface elevation within the ith elevation band, and N_b is the number of elevation bands in that grid.

For the baseline and SGC2 methods, σ_{ep} was calculated as

$$\sigma_{ep} = \frac{1}{N_P} \sum_{j=1}^{N_P} \left[\frac{1}{N_b} \sum_{i=1}^{N_b} \sigma_{eb}(i, j) \right], \qquad (3)$$

where $\sigma_{eb}(i, j)$ is the standard deviation of subgrid surface elevation within the ith elevation band for the jth dominant PFT, and N_p is the number of dominant PFTs in that grid. For the baseline method, N_p can be any number within 15 because all PFTs within each model grid were included, and N_b equals 1 because only one elevation band was assigned to each PFT.

3   Results and discussions

3.1   Comparison of the SGC method and the baseline method

Figure 4 shows the number of PFTs and σ_{ep} from the baseline subgrid classification method over the NA continent at 0.1° and 1.0° resolutions. Both spatial resolutions show similar spatial patterns of subgrid variability of vegetation: more PFT classes (6–8 PFTs at 0.1° resolution, over 8 PFTs at 1.0° resolution) per grid in the coastal areas than in the inland region such as the Great Plains (1–2 PFTs at 0.1° resolution, 3–5 PFTs at 1.0° resolution), where crop dominates the landscape. Increasing model grid size results in greater subgrid-scale variability of PFTs (Fig. 4a and b). When assigning one elevation band to each PFT within the model grid, the spatial distribution of σ_{ep} clearly corresponds with topographic variations (Fig. 4c and d). In western NA with complex topography such as the Coastal Range and Rocky Mountains, σ_{ep} is larger than in flat areas such as the Great Plains and coastal area in southeast NA. The spatial contrast becomes more distinct at coarser resolution as the subgrid topographic variability becomes larger. The average σ_{ep} across the continent rapidly increases from 43.1 m at 0.1° resolution to 62.7 m at 0.25° resolution, 79.9 m at 0.5° resolution, 99.1 m at 1.0° resolution, and 123.3 m at 2.0° resolution, indicating that with the baseline method substantially less subgrid topographic details are represented at coarser resolution.

Figure 5 compares the SGC method and the baseline method in terms of computational burden, % PFT and the elevation variability explained by the methods. Note that although a maximum of 15 PFTs are allowed within each model grid, the computational burden depends on the actual number of PFTs needed to represent the subgrid PFT variability, which can be smaller than 15. It is apparent that SGC is more computationally intensive even with N_LRU (= 12) smaller than the maximum number of PFTs in the baseline method (= 15). In the baseline method, the average number of PFTs increases from 3.5 to 8.4 at resolutions varying from 0.1 to 2.0 degrees, while the average number of elevation–PFT LRUs in the SGC method increases from 5.5 to 10.6 ($N\_LRU = 12$). SGC explained slightly less PFTs than the baseline method that allows all PFTs in each model grid (Fig. 5b). Increasing N_LRU required more computing resources but yielded better PFT representation. However, the small sacrifice in computational burden and % PFT resulted in substantial improvement in elevation variability representation. Figure 5c demonstrates that the standard deviation of elevation within each elevation band is greatly suppressed, with the average σ_{ep} of SGC reduced to almost one third that of the baseline method.

Compared to the baseline method where the number of PFTs and the number of elevation band per PFT (= 1) were predetermined, the optimal SGC method produced a much more spatially variant number of elevation bands (M) and number of PFTs within each elevation band (N) throughout the continent (Figs. 6a, b and 7a, b). In topographically complex areas, the SGC method shows great advantage over the baseline method in representing details of subgrid topography (Fig. 4c vs. Fig. 6d, Fig. 4d vs. Fig. 7d). This advantage is more prominent at coarser resolution as σ_{ep} is generally less than 60 m for the SGC method in the Pacific Northwest compared to over 200 m for the baseline method (Fig. 4d vs. Fig. 7d). However, more detailed elevation information in the SGC method compromises the representation of PFTs compared to the baseline method that considers all PFTs within each model grid. Nevertheless, the SGC method still explained a reasonable amount of PFT variability (over 80 % in Figs. 6c and 7c, over 94 % in Fig. 5b) while greatly improving the elevation variability, which has equally large if not larger impacts on land surface processes than PFT variability (Leung and Ghan, 1998).

3.2   The SGC method with different N_LRU and model resolutions

The SGC method at different model resolutions (Fig. 6a vs. 7a, Fig. 6b vs. 7b) shows that both M and N demonstrate a similar spatial pattern in North America. In the areas with more complex topography such as the Coastal Range, Rocky Mountains, and the Appalachian Mountains (Fig. 2), more elevation bands were generated than in flat areas such as the central and coastal plains (Figs. 6b and 7b). The spatial distribution of N within each elevation band, on the other hand, generally shows an opposite pattern (Figs. 6a and 7a) because the combined number of PFTs and elevation bands was restricted by N_LRU. In the flat areas of central and eastern NA, more than six PFTs were generated that explained over 98 % of the total PFTs in the model grids. In topographically complex areas such as western NA, although smaller N were produced than in other areas, the total PFTs explained by the method was still over 80 % for each model grid (Fig. 6c and 7c) because PFT is more correlated with elevation over complex terrains. This allows the SGC method to optimally represent subgrid variations in both topography and vegetation.

With decreasing model resolutions and increasing number of N_LRU, the average computational burden increases. At the finest resolution of 0.1°, the average number of total LRUs increases only slightly with increasing N_LRU (5.6 for $N\_LRU = 12$, 6.3 for $N\_LRU = 18$, and 6.8 for $N\_LRU = 24$) but increases rapidly at resolution of 2° (10.7, 14.9, and 18.9 for N_LRU equal 12, 18, and 24) (Fig. 5a). This demonstrates that the low subgrid variability within a small grid cell can be well represented by a small number of subgrid LRUs. As the subgrid topography and vegetation variation increase with coarser model resolution, more LRUs are required and the method is more restricted by the maximum-allowed number of LRUs. This shows the method's ability in assigning optimal classes for different model resolutions, and a larger number of maximum-allowed
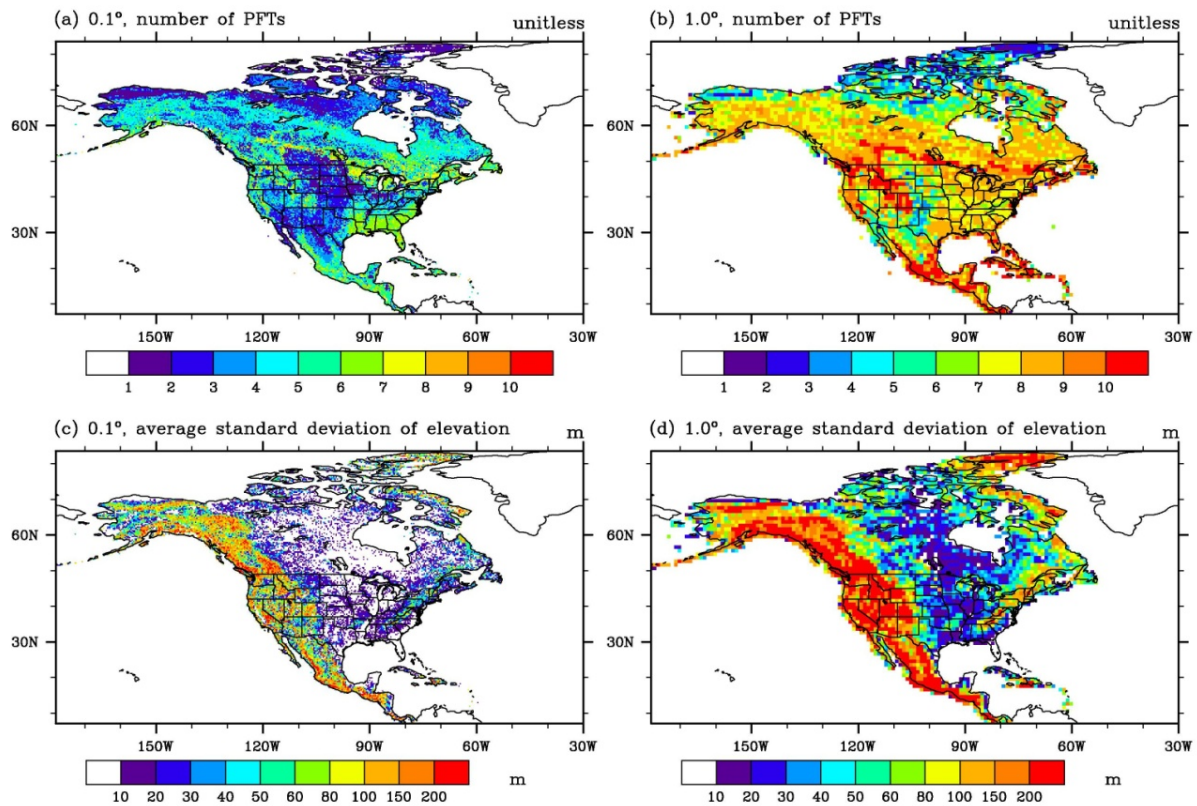
Fig. 4. Baseline subgrid classification method. **(a)** 0.1°, number of PFTs; **(b)** 1.0°, number of PFTs; **(c)** 0.1°, average standard deviation of elevation σ_{ep}; **(d)** 1.0°, average standard deviation of elevation σ_{ep}.

total classes gives more flexibility in assigning the best suitable combination of elevation bands and PFTs.

3.3 Comparison of three subgrid classification methods

The optimal subgrid classification method, SGC, was evaluated against the SGC1 and SGC2 methods with a fixed number of elevation bands and PFTs.

The spatial distribution of the differences in the % PFT (Fig. 8) and σ_{ep} (Fig. 9) illustrate that at both fine and coarse resolutions, the SGC method explained less PFTs (negative values of SGC-SGC1 or SGC-SGC2) and produced distinctly lower σ_{ep} in the mountainous western NA with complex topography. In these areas, the SGC method required more elevation bands to represent reasonable variations of elevation (e.g., over 6 elevation bands were produced in these areas in Fig. 6b), thus sacrificing the total PFTs represented (only 2–3 PFTs were identified in this area in Fig. 6a compared to four PFTs used in SGC1 and SGC2 methods). In the southeast United States such as Florida, Mississippi, South Carolina and Georgia, and in central Canada, the SGC method identified more PFTs than the other two methods because few elevation classes are needed to represent the flat topography (Fig. 8). Although σ_{ep} in flat areas from SGC is slightly greater than that from the other two methods (Fig. 9),

SGC is still able to produce reasonable representation of elevation variability as σ_{ep} is less than 30 m in these areas (Figs. 6d and 7d). This implies that using 6 or 4 elevation bands may be redundant in these areas because topography does not vary much and vegetation distribution has little relation to topography. When fewer PFTs were used in SGC1 and SGC2 and fewer N_LRUs were used in SGC, the areas of positive difference in the % PFT between SGC and the other two methods expanded from central Canada to Alberta and Saskatchewan provinces in western Canada (Fig. 8b and e). In Mexico and Central America with distinct topographic relief, SGC explained a considerably higher percentage of PFT than SGC2. This emphasizes the advantages of SGC in topographically complex and species-rich areas partly because vegetation type correlates with topography.

With increasing model grid size and decreasing M or N, the average % PFT represented by each method decreases (Fig. 10a), while the standard deviation of % PFT increases (Fig. 10b). With $M = 6$, $N = 4$ and $N\_LRU = 24$, the SGC1 method, which first classifies topography into 6 elevation bands and represents each elevation band with the 4 most dominant PFTs, explains the most PFTs, while the optimal classification method produced the lowest percentage of PFTs except at 0.1° resolution. With 18 maximum-allowed
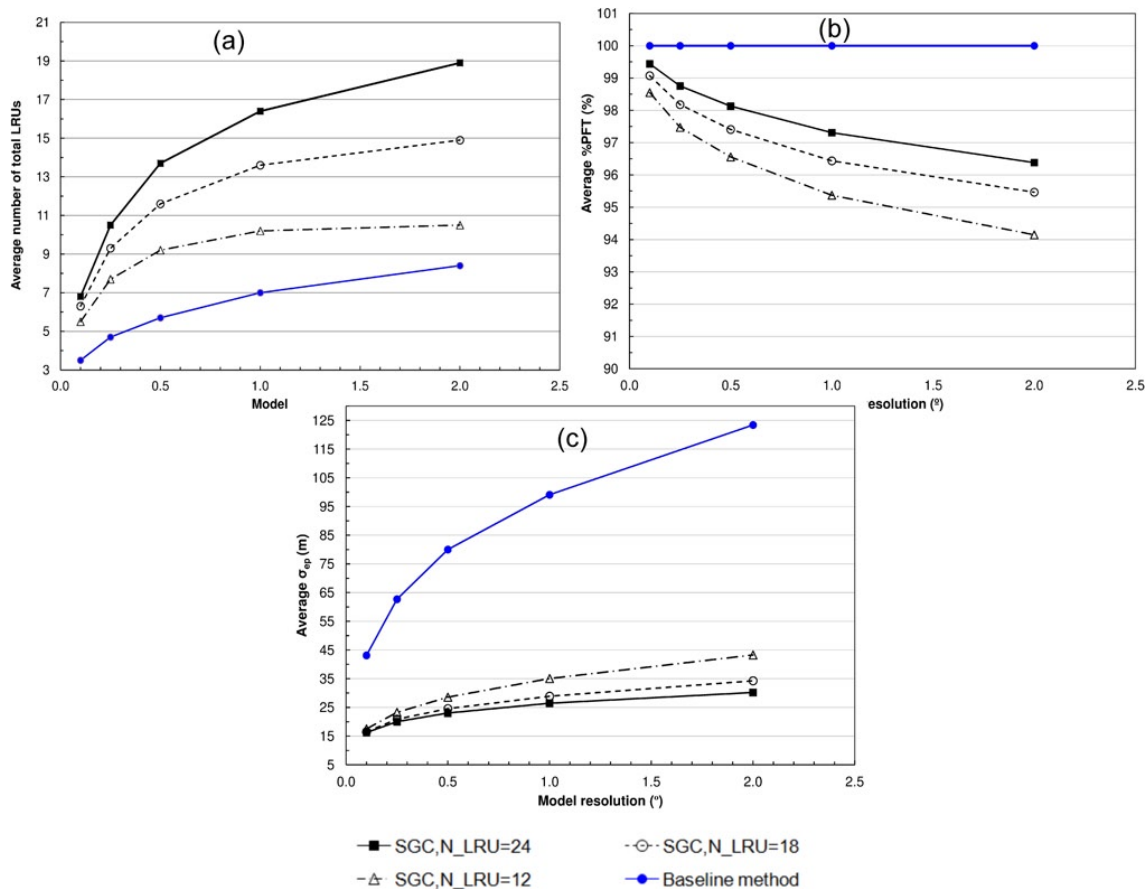
Fig. 5. SGC method compared to baseline method. **(a)** Average number of total LRUs across NA; **(b)** average %PFT; **(c)** average σ_{ep}.

number of total LRUs in SGC, $M = 6$ and $N = 3$, the negative differences in PFTs between SGC and the other two methods were compensated by positive differences (Fig. 8b and e). Hence overall, the average percentage of PFTs explained by SGC is higher than the other two methods across all resolutions (Fig. 10a). As N_LRU decreases to 12, at the finest resolution of 0.1°, SGC explained a slightly greater percentage of PFTs than the other methods. At the resolution of 1° or higher, SGC explained a lower amount of PFTs than the other two methods. As stated above, at coarse resolution SGC is more restricted by the number of maximum-allowed LRUs because larger variations of PFTs and topography exist. Since topography shows more distinct change than PFTs, the balance between the number of elevation bands and PFTs resulted in less PFTs explained by the method. Although the SGC method shows varying performances in terms of the average percent of PFTs explained compared to the other two methods, the PFTs explained by this method is more spatially homogeneous – all model grids in the study area have over 80% of total PFTs explained regardless of N_LRU and model resolution. In contrast, the PFTs explained by SGC1 and SGC2 can be as low as 52% (Fig. 10c), and around

92.7% of the model grids have over 80% of the PFTs explained (Fig. 10d).

The abilities of the three methods in explaining topographic variation are shown in Fig. 11. For all three methods, the average σ_{ep} increase with model grid size, i.e., reduced ability to represent subgrid topographic variability as model resolution decreases. At fine resolutions from 0.1° to 0.5°, the average σ_{ep} from the SGC method with N_LRU of 24 (and 18) is greater than the SGC1 and SGC2 methods with $M = 6$ and $N = 4$ (and $N = 3$) (Fig. 11a), meaning that SGC1 and SGC2 have better overall representation of elevation variability at these scales. When model resolution decreases, the advantage of the SGC method in elevation representation emerges. At both 1° and 2° resolutions, the average σ_{ep} from SGC is lower than that from SGC1 and SGC2. Compared to those from SGC1 and SGC2, the average σ_{ep} from SGC climbs more slowly with increasing model grid size and decreasing number of maximum-allowed classes. Because topography shows much greater variability at coarser scales, the slower change in average σ_{ep} indicates more stable performance of the SGC method across different scales. Although the average σ_{ep} from SGC is higher than that from the other methods at fine resolution (e.g., less than
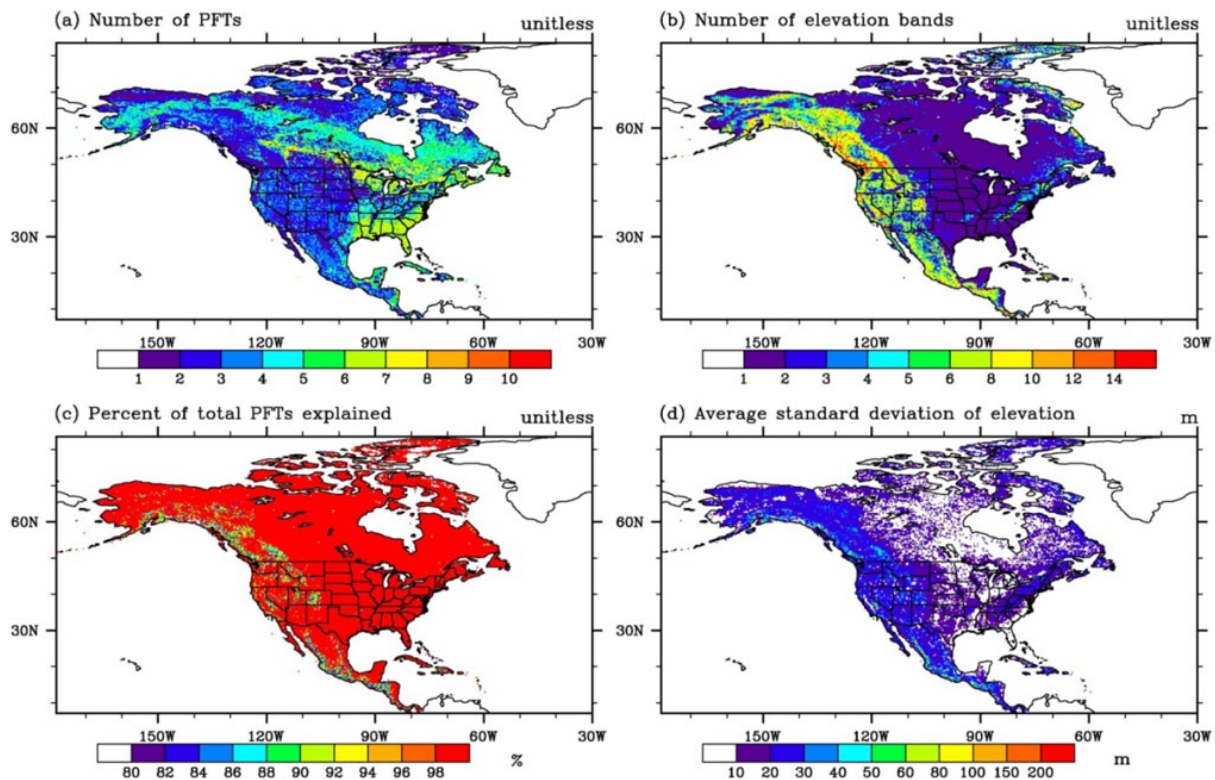
Fig. 6. Optimal classification SGC with maximum-allowed total classes $N\_class = 18$ at $0.1°$ resolution. **(a)** number of dominant PFTs classified for each grid; **(b)** Number of elevation bands classified for each grid; **(c)** total percentage of PFTs explained by the method; **(d)** average standard deviation of elevation in each elevation band in each model grid σ_{ep}.

$1°$), the variation of σ_{ep} across NA is smaller than that from the other two methods (Fig. 11b) at each model resolution and with each scheme of computational burden. This means that the elevation variability explained by the SGC method is more spatially homogeneous than the other two methods. The combined examination of Figs. 8–11 and the statistical analysis in Table 1 show that the SGC method is not necessarily superior to the other two methods in terms of both vegetation and elevation variation explained, when a large number of subgrid classes is allowed ($N\_LRU = 24$). However, when computational burden is moderately alleviated using fewer number of subgrid LRUs ($N\_LRU = 18$, $M = 6$ and $N = 3$), the SGC method begins to demonstrate its advantage in balancing the variability of vegetation and elevation distribution that can be explained. At the coarser resolutions of $1°$ to $2°$, the SGC method becomes clearly superior to both SGC1 and SGC2; i.e., a statistically greater percentage of PFTs was explained and σ_{ep} is smaller (Table 1). With N_LRU of 12, the SGC method is better than SGC2 at the resolutions of $0.5°$ or coarser and better than SGC1 only at the resolution of $0.25°$.

Despite the variable performance of SGC compared to SGC1 and SGC2 in terms of PFT and elevation variability, SGC is more computationally efficient, especially at fine model resolutions (Fig. 12). For example, in Scheme 1 ($N\_LRU = 24$, $M = 6$, $N = 4$), at model resolution of $0.1°$,

the average number of computational units for each model grid for SGC increases quickly from 6.8 to 13.0 and 14.0 for SGC1 and SGC2. With decreasing model resolutions, the differences in computational burden decrease. At coarse resolution such as $2.0°$, the average number of LRUs are similar for all three methods.

4 Conclusions

In this study we presented a new subgrid land surface representation method for LSMs that accounts for the joint distribution of vegetation and topography. Using updated datasets of high-resolution DEM and PFTs, this study provides a systematic analysis and comparison of different ways to classify subgrid surface elevation and vegetation to provide an optimal approach that improves both accuracy and computational efficiency. For each model grid, the new method assigned variable elevation and vegetation classes based on their joint distribution so that the subgrid-scale variability of both can be captured. Using predefined N_LRU, the method ensures that the optimal combination of M classes for elevation and N classes for PFTs is assigned to each model grid under the restriction of a given computational burden. In flat areas with rich vegetation diversity, this method assigns a small M so
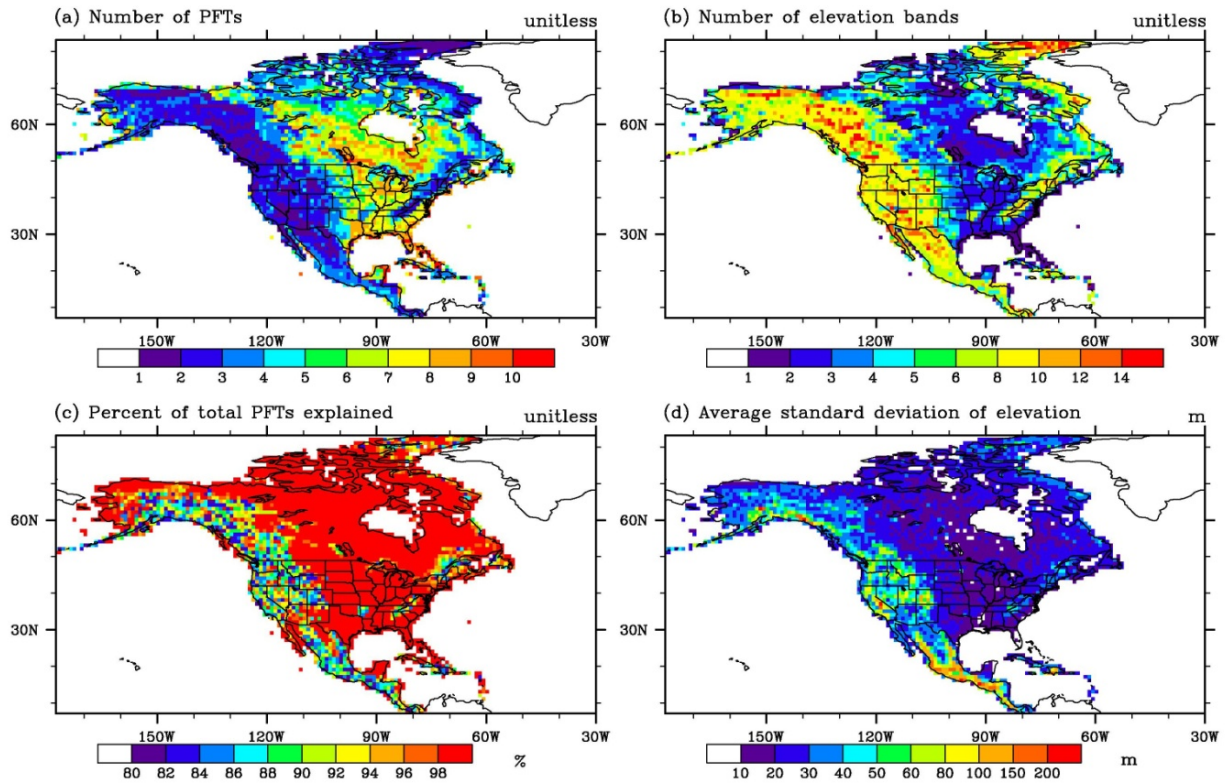
Fig. 7. Optimal classification SGC with maximum-allowed total classes $N\_class = 18$ at $1.0°$ resolution. **(a)** Number of dominant PFTs classified for each grid; **(b)** number of elevation bands classified for each grid; **(c)** total percentage of PFTs explained by the method; **(d)** average standard deviation of elevation in each elevation band in each model grid σ_{ep}.
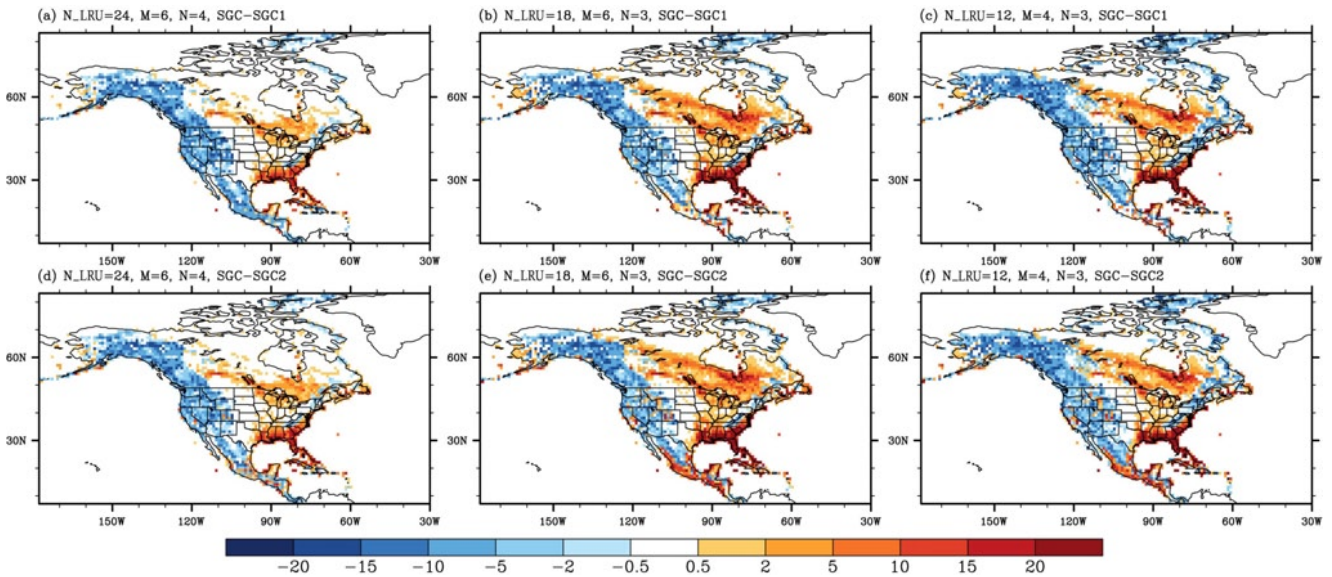
Fig. 8. Difference in percentage of total PFTs explained by SGC compared to SGC1 (top row), and SGC compared to SGC2 (bottom row) at $1.0°$ model resolution.
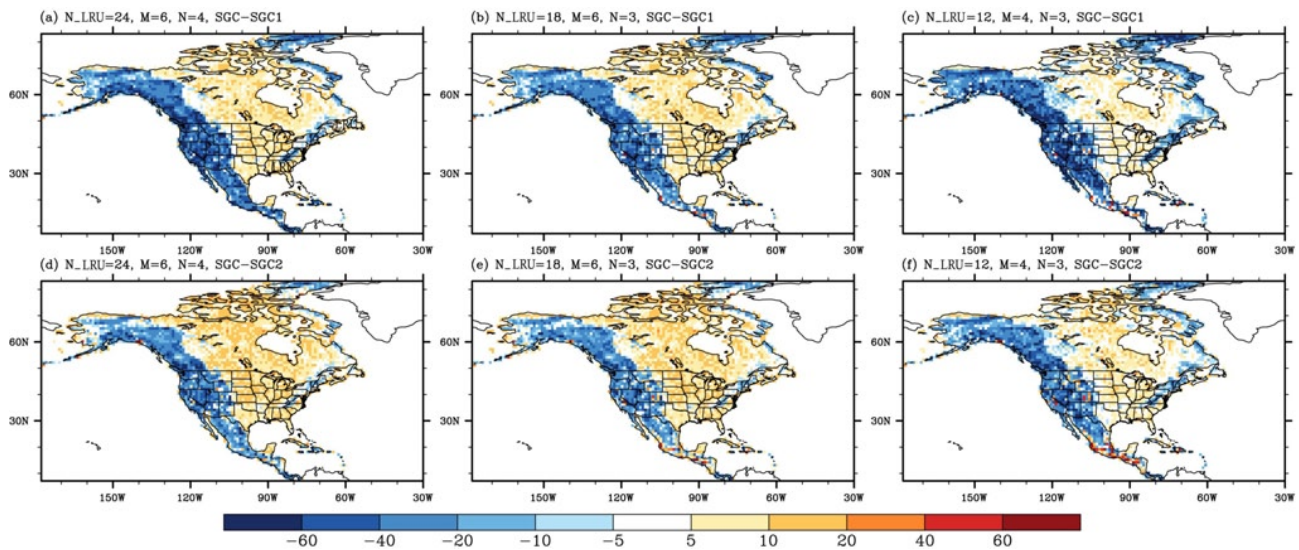
Fig. 9. Difference in average standard deviation of elevation for SGC compared to SGC1 (top row) and SGC compared to SGC2 (bottom row) at 1.0° model resolution.

Table 1. Paired t test statistics for four classification schemes in terms of total PFT explained and standard deviation of elevation σ_{ep}. Values in bold indicate significant difference between the two classifications on a 95 % confidence level. Positive value in PFTs and negative value in elevation standard deviation mean better capability of explaining both PFT and elevation. *: not significant.

Resolution	SGC vs. SGC1		SGC vs. SGC2	
	Percentage of PFTs explained	σ_{ep}	Percentage of PFTs explained	σ_{ep}
N_class = 24, M = 6, N = 4				
0.1°	−0.7*	277.3	47.4	434.8
0.25°	−19.6	24.9	−1.9*	102.1
0.5°	−16.9	−18.2	−7.1	17.7
1.0°	−12.7	−21.5	−7.0	−6.7
2.0°	−8.2	−17.9	−4.5	−11.3
N_class = 18, M = 6, N = 3				
0.1°	106.0	312.4	150.2	464.6
0.25°	27.9	42.5	46.4	111.1
0.5°	**8.6**	**−10.1**	19.4	21.6
1.0°	**1.4**	**−18.1**	**7.6**	**−4.3**
2.0°	0.38*	**−15.8**	4.4	**−9.0**
N_class = 12, M = 4, N = 3				
0.1°	49.0	58.7	88.4	205.8
0.25°	**3.2**	**−36.4**	19.4	23.6
0.5°	−4.9	−40.9	**4.4**	**−10.9**
1.0°	−6.5	−31.4	−1.0*	**−15.7**
2.0°	−4.6	−21.1	−0.9*	**−14.1**

that a large number N of PFTs can be represented while elevation variation is still well represented. However, in topographically complex areas, the method assigns a larger M and a smaller N so that elevation variation can be reasonably explained. Based on our analysis using high-resolution DEM and vegetation data, we find that this classification is feasible because, in topographically complex areas, elevation has a dominant influence on vegetation through its effects on climate, so assigning a small number of PFT classes within each elevation class is able to capture the dominant subgrid
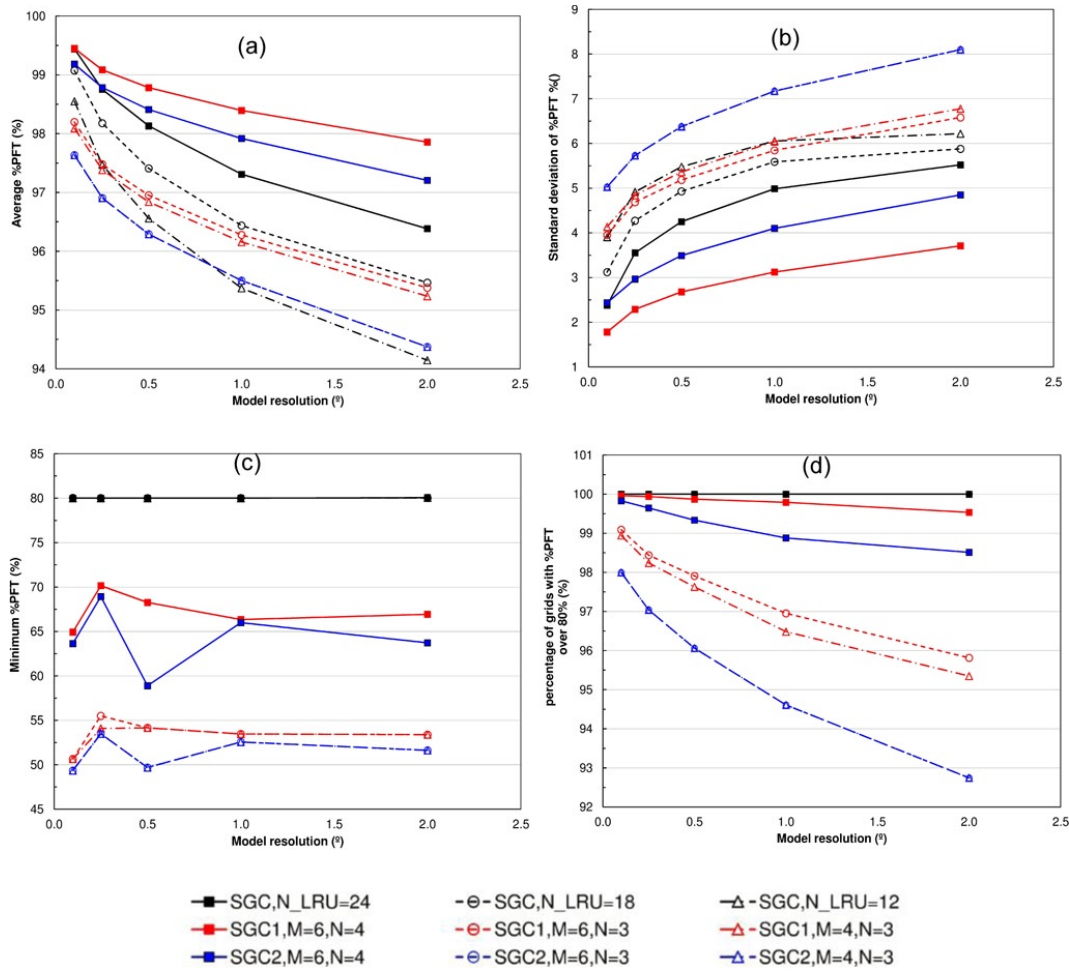
Fig. 10. Total PFTs explained by methods SGC, SGC1 and SGC2 across NA. **(a)** Average % PFT within model grids; **(b)** standard deviation of % PFT within model grids; **(c)** Minimum % PFT within model grids; **(d)** percentage of grids with total % PFT over 80 %. Black lines: SGC; red lines: SGC1; blue lines: SGC2. Square symbol: $N\_LRU = 24$; circular symbol: $N\_LRU = 18$; triangular symbol: $N\_LRU = 12$.
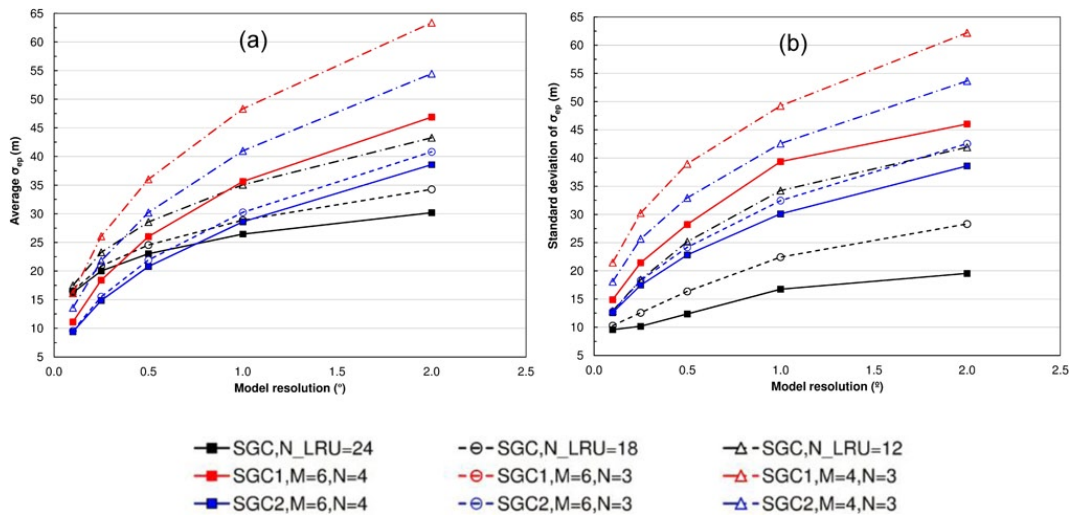
Fig. 11. Elevation variability explained by methods SGC, SGC1 and SGC2. **(a)** Average σ_{ep}; **(b)** standard deviation of σ_{ep}. Black lines: SGC; red lines: SGC1; blue lines: SGC2. Square symbol: $N\_LRU = 24$; circular symbol: $N\_LRU = 18$; triangular symbol: $N\_LRU = 12$.
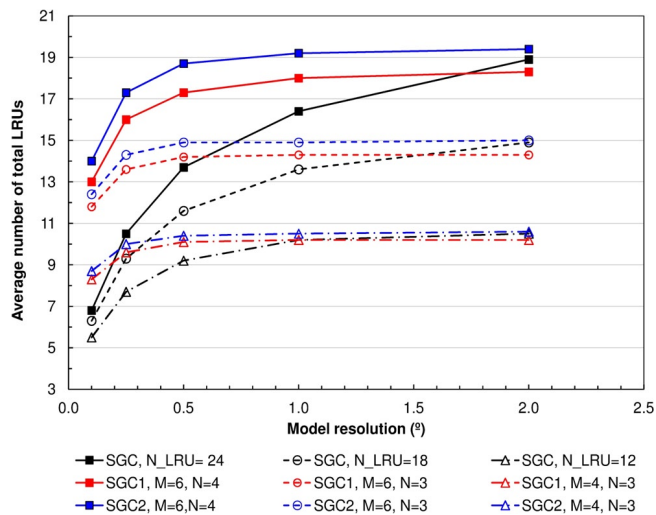
Fig. 12. Average number of total LRUs of method SGC, SGC1 and SGC2 across NA. Black lines: SGC; red lines: SGC1; blue lines: SGC2. Square symbol: $N\_LRU = 24$; circular symbol: $N\_LRU = 18$; triangular symbol: $N\_LRU = 12$.

variations of PFT within the model grid cells. However, we recognized that there is no simple relationship between vegetation distribution and topography since vegetation cover depends not only on elevation but on other environmental factors such as slope/aspect and soil that also play an important role in determining vegetation distribution, so it is important to develop vegetation–elevation distribution for each model grid.

Compared to the baseline method, which assigns a single elevation class to each PFT, the new method provides an obvious advantage in representing topographic variability. Although the SGC method slightly compromised the ability to represent vegetation variability and increased computational burden compared to the baseline method, the new method still explained at least 80 % of the total PFTs in each model grid and produced substantial improvements in topography representation. The effectiveness of the new method in representing subgrid variability in both topography and vegetation is partly related to the correlation between topography and vegetation. However, this effectiveness decreases with decreasing model resolution because the elevation dependence of vegetation is weaker at coarser spatial scales.

Compared to the other subgrid approaches with predetermined number of elevation classes and vegetation types (SGC1 and SGC2), the new method presented in this study balanced the representation of both topography and vegetation variability under the restriction of a maximum-allowed number of total LRUs. With the same maximum-allowed LRUs, it requires less computational burden in LSM simulation than the other two methods. Among the three schemes, the new method shows advantages over the other methods with moderate computation intensity ($N\_LRU = 18$) and at

coarse scales in that both PFTs and topography variability were best explained. Furthermore, the variability of both vegetation and elevation explained by the new method was more spatially homogeneous compared to the SGC1 and SGC2 methods regardless of model resolutions and computational burdens.

With the new subgrid scheme, the fractional area of each elevation band and PFT can be determined and the mean elevation of each elevation band can be defined. When implemented in the model, each surface elevation class can be forced by different atmospheric conditions by disaggregating the atmospheric forcing from each model grid cell to the subgrid elevation class based on temperature and precipitation lapse rate or the subgrid parameterization of orographic precipitation described in Leung and Ghan (1995, 1998) and Ghan et al. (2006). Separate calculations of surface processes can be performed for each subgrid PFT within each subgrid elevation class, and the output fluxes at each class can be then aggregated in an area-weighted manner for each grid cell. This will allow the impacts of subgrid variations of surface topography and PFTs on the partitioning of the energy and water budgets to be represented in land surface models. The impacts of the new subgrid classification on land surface simulations at different model resolutions will be studied in the future.

References

Bonan, G. B., Levis, S., Kergoat, L., and Oleson, K. W.: Landscapes as patches of plant functional types: An integrating concept for climate and ecosystem models, Global Biogeochem. Cy., 16, 5-1–5-23, doi:10.1029/2000GB001360, 2002a.

Bonan, G. B., Oleson, K. W., Vertenstein, M., Levis, S., Zeng, X., Dai, Y., Dickinson, R. E., and Yang, Z. L.: The land surface climatology of the Community Land Model coupled to the NCAR Community Climate Model, J. Climate, 15, 3123–3149, 2002b.

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., and Tarpley, J. D.: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model, J. Geophys. Res., 108, D228851, doi:10.1029/2002JD003296, 2003.

Ghan, S. J., Liljegren, J. C., Shaw, W. J., Hubbe, J. H., and Doran, J. C.: Influence of subgrid variability on surface hydrology, J. Climate, 10, 3157–3166, 1997.

Ghan, S. J., Shippert, T. R., and Fox, J.: Physically based global downscaling: regional evaluation, J. Climate, 19, 429–445, doi:10.1175/JCLI3622.1, 2006.

Giorgi, F. and Avissar, R.: Representation of heterogeneity effects in earth system modeling: Experience from land surface modeling, Rev. Geophys., 35, 413–437, doi:10.1029/97RG01754, 1997.

Giorgi, F., Francisco, R., and Pal, J.: Effects of a subgrid-scale topography and land use scheme on the simulation of surface climate and hydrology, Part I: Effects of temperature and water vapor disaggregation, J. Hydrometerol., 4, 317–333, doi:10.1175/1525-7541(2003)4<317:EOASTA>2.0.CO;2, 2003.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A.: Very high resolution interpolated climate surfaces for global land areas, Int. J. Climatol., 25, 1965–1978, doi:10.1002/joc.1276, 2005.

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

Koster, R. D. and Suarez, M. J.: A comparative analysis of two land surface heterogeneity representations, J. Climate, 5, 1379–1390, doi:10.1175/1520-0442(1992)005<1379:ACAOTL>2.0.CO;2, 1992.

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C.: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, Global Biogeochem. Cy., 19, GB1015, doi:10.1029/2003GB002199, 2005.

Krinner, G., Lezine, A.-M., Braconnot, P., Sepulchre, P., Ramstein, G., Grenier, C., and Gouttevin, I.: A reassessment of lake and wetland feedbacks on the North African Holocene climate, Geophys. Res. Lett., 39, L07701, doi:10.1029/2012GL0500992, 2012.

Lawrence, D., Oleson, K. W., Flanner, M. G., Thorton, P. E., Swenson, S. C., Lawrence, P. J., Zeng, X., Yang, Z. L., Levis, S., and Skaguchi, K.: Parameterization improvements and functional and structural advances in version 4 of the Community Land Model, J. Adv. Model. Earth Syst., 3, M03001, doi:10.1029/2011MS000045, 2011.

Lawrence, P. J. and Chase, T. N.: Representing a new MODIS consistent land surface in the Community Land Model (CLM 3.0), J. Geophys. Res., 112, G01023, doi:10.1029/2006JG000168, 2007.

Leung, L. R. and Ghan, S. J.: A subgrid parameterization of orographic precipitation, Theor. App. Climatol., 52, 95–118, doi:10.1007/BF00865510, 1995.

Leung, L. R. and Ghan, S. J.: Parameterizing subgrid orographic precipitation and surface cover in climate models, Mon. Weather Rev., 126, 3271–3291, doi:10.1175/1520-0493(1998)126<3271:PSOPAS>2.0.CO;2, 1998.

Li, H., Wigmosta, M. S., Wu, H., Huang, M., Ke, Y., Coleman, A. M. and Leung, L. R.: A physically based runoff routing model for land surface and earth system models, J. Hydromet., 14, 808–828, doi:10.1175/JHM-D-12-015.1, 2013.

Li, R. and Arora, V. K.: Effect of mosaic representation of vegetation in land surface schemes on simulated energy and carbon balances, Biogeosciences, 9, 593–605, doi:10.5194/bg-9-593-2012, 2012.

Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models, J. Geophys. Res., 99, 14415–14428, 1994.

Nijssen, B., Schnur, R., and Lettenmaier, D. P.: Global retrospective estimation of soil moisture using the variable infiltration capacity land surface model, 1980–93, J. Climate, 14, 1790–1808, doi:10.1175/1520-0442(2001)014<1790:GREOSM>2.0.CO;2, 2001.

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah©MP): 1.Model description and evaluation with local-scale measurements, J. Geophys. Res., 116, D12109, doi:10.1029/2010JD015139, 2011.

Oleson, K. W. and Bonan, G. B.: The effects of remotely sensed plant functional type and leaf area index on simulations of boreal forest surface fluxes by the NCAR land surface model, J. Hydrometeorol., 1, 431–446, doi:10.1175/1525-7541(2000)001<0431:TEORSP>2.0.CO;2, 2000.

Oleson, K. W., Lawrence, D. M., Bonan, G. B., Flanner, M. G., Kluzek, E., Lawrence, P. J., Levis, S., Swenson, S. C., Thornton, P. E., Dai, A., Decker, M., Dickinson, R., Feddema, J., Heald, C.L., Hoffman, F., Lamarque, J.-F., Mahowald, N., Niu, G.-Y., Qian, T., Randerson, J., Running, S., Sakaguchi, K., Slater, A., Stockli, R., Wang, A., Yang, Z.-L., Zeng, X., and Zeng, X.: Technical Description of version 4.0 of the Community Land Model (CLM). National Center for Atmospheric Research, Boulder, CO, NCAR/TN-478+STR, 2010.

Seth, A., Giorgi, F., and Dickinson, R. E.: Simulating fluxes from heterogeneous land surfaces: Explicit subgrid method employing the biosphere-atmosphere transfer scheme (BATS), J. Geophys. Res., 99, 18651–18667, doi:10.1029/94JD01330, 1994.

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J. O., Levis, S., Lucht, W., Sykes, M. T., Thonicke, K., and Kenevsky, S.: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model, Global. Change Biol., 9, 161–185, 2003.

Still, C., Berry, J. C., Collatz, G. J., and DeFries, R. S.: Global distribution of C3 and C4 vegetation: carbon cycle implications, Global Biogeochem. Cy., 17, 6-1–6-14, doi:10.1029/2001GB001807, 2003.