

Comparison of simulations from a state-of-the-art dynamic global vegetation model (LPJ-GUESS) driven by low- and high-resolution climate data

Dmitry Otryakhin¹, David Martín Belda¹, and Almut Arneth¹

¹Institute of Meteorology and Climate Research Atmospheric Environmental Research IMK-IFU, Karlsruhe Institute of Technology KIT, Kreuzeckbahnstraße 19, 82467 Garmisch-Partenkirchen, Germany

Correspondence: Dmitry Otryakhin (dmitry.otryakhin@pm.me)

Abstract. Simulations of dynamic global vegetation models (DGVMs) are typically conducted at a spatial resolution of 0.5° , while higher-resolution simulations remain uncommon. This coarse resolution eliminates detailed orographic features and hence, associated climate variability, which are especially pronounced in mountainous regions. The impact of disregarding such variability on vegetation dynamics has not been thoroughly examined. In this study, we explore the differences in vegetation outcomes between the DGVM LPJ-GUESS simulations conducted at high and low spatial resolutions. Using the CHELSA algorithm, we create an elevation-informed high-resolution climate dataset for a domain encompassing the European Union. Distinctive features of this algorithm include orographic nature of formation of precipitation, a negative derivative of temperatures with respect to elevation, and also, detailed consideration of shadowing and exposure of the terrain to the Sun in computations of solar radiation. We design a custom experiment protocol and use it to perform LPJ-GUESS simulations on both resolutions. Comparative analysis reveals significant systematic discrepancies between the two resolutions. In mountainous areas, all of the considered output variables show statistically significant differences. In particular, carbon pools are smaller on the high resolution, with the total carbon pool being 37–39% smaller. Furthermore, we quantify the extent to which the under-representation of orographic climate variation affects regional predictions across the European Union. This is expressed as a difference in the total value, which ranges from -3.8% for the net ecosystem productivity to 2.9% for the litter and soil C pools. These values are found to be comparable to differences caused by miss-representation of water bodies and shorelines on the low resolution.

1 Introduction

The rapidly progressing climate change reinforces the urgency with which political and societal measures need to be implemented to reduce greenhouse gas emissions and to mitigate further climate change as much as possible, while also considering appropriate adaptation measures. Due to the complexity of general circulation models, global climate change projections are still available only at very coarse spatial resolutions (> 0.5 degrees; e.g., Masson-Delmotte et al., 2021). These allow to assess very well the basic response of the earth system to climate change. But impacts of climate change on ecosystems and societies are felt locally; weather extremes in particular can happen at scales of few km, rather than tenths or hundreds of km. Many

mitigation and adaptation measures that are being discussed rely on the ecosystem services provided by natural and managed ecosystems (Shin et al., 2022; Smith et al., 2022). Their design and permanence require climate change projections of a spatial resolution much closer to the spatial extents of the ecosystems under study. For instance, using the dynamic global vegetation model LPJ-GUESS, Lagergren et al. (2024) explored how climate change and CO₂ impacts of different vegetation types in Fennoscandia would affect habitats of rare and threatened species and also how reindeer grazing (an important source of income for the local population) would be affected. Another study based on LPJ-GUESS simulated the negative impacts of late-spring frosts on forest productivity, yielding a decline of NPP in frost years of around 50 % compared to non-frost years (Meyer et al., 2024). High-resolution climate for these analyses provided important information on, e.g., seasonal and altitudinal distribution of snowfall Lagergren et al. (2024) and minimum temperatures (Meyer et al., 2024). Similarly, in the Spanish region of Catalonia 1km downscaled climate projections supported simulations of future productivity of a number of species of wild edible mushrooms, which provide both large economic and recreational value (Morera et al., 2024). In this study, too, the capability to resolve climate gradients in mountain areas underpins confidence in the projected patterns.

Downscaling methods can be applied to overcome the mismatch between coarse global climate projections, and the fine-resolution needs of impact models (Karger et al., 2023). At present, terrain-informed downscaling could be executed by either regional climate models for dynamical downscaling, or by topographic downscaling methods. Algorithms of the first class are very precise as they directly model physical and chemical processes in the atmosphere. This comes with the disadvantage of being computationally slow, which makes their application on large scales challenging (Giorgi et al., 2009; Sørland et al., 2021; Schär et al., 2020). Topographic downscaling uses mechanistic relationships to turn low-resolution climatologies into high-resolution ones based on knowledge of terrain. These relationships are quite simple and do not capture atmospheric effects unrelated to topography, so this class of algorithms fails to represent some small-scale effects, such as convective precipitation (Karger et al., 2021). Also, topographic downscaling is characterized by less computational complexity than that of dynamical downscaling. The two best performing and widely known topographic methods are CHELSA (Karger et al., 2017, 2021, 2023) and PRISM (Daly et al., 1994, 1997). For this study we choose CHELSA to perform downscaling for two reasons. First, we need a computationally fast algorithm as we examine a region covering the whole of Europe. Second, out of the two best performing topographic downscaling methods, CHELSA provides the easiest way to interpret the results from the point of view of atmospheric physics.

Here we present a downscaled climate product for the European region at 0.05° for use in climate change impact studies (Sect. 3). The downscaling adopts the approach presented in Karger (2022) and Karger et al. (2023), and uses ISIMIP3b 0.5° climate data (Lange and Büchner, 2021) to obtain their high-resolution counterpart. We used the downscaled data to force LPJ-GUESS simulations and, applying an ensemble approach, tested whether systematic differences in simulated output emerged between fine and coarse resolutions (Sect. 4). Lastly, we evaluated the impact of this bias on the scale of European Union (Sect. 5). The work is part of an ongoing effort to incorporate a simplified downscaling method into LPJ-GUESS, which eventually should allow users to downscale flexibly different climate projections for different regions.

2 Methods

2.1 CHELSA downscaling algorithm

CHELSA (Karger et al., 2017, 2021, 2023) is a family of semi-mechanistic algorithms designed to perform spatial downscaling of near-surface climate data. For this study, we choose CHELSA V2.1 presented in Karger et al. (2023) and its original software implementation (Karger, 2022), that scales ISIMIP3b temperature, precipitation, and downwelling shortwave radiation from an input resolution of 0.5° down to $0.0083(3)^\circ$. The code additionally requires 3D data from the CMIP6 ensemble (Jungclaus et al., 2019), as well as static data such as high-resolution surface elevation.

2.1.1 Temperature

For every low-resolution grid cell, the temperature is projected to the sea level via Eq. (1);

$$T_{lr}^{sea} - T_{lr}^c = \gamma(Z^{sea} - Z_{lr}^c), \quad (1)$$

where γ is the lapse rate, T_{lr}^{sea} is the temperature to compute, Z^{sea} is the sea-level elevation, and T_{lr}^c and Z_{lr}^c are, correspondingly, the temperature and elevation of the cell. The lapse rate γ is obtained by applying linear regression to CMIP6 pressure-level data. The projected values T_{lr}^{sea} are interpolated using B-splines to obtain high-resolution sea-level temperatures (T_{hr}^{sea}). For every high-resolution grid point, the surface temperature is computed using the interpolated values, the surface elevation, the lapse rates and Eq. (2):

$$T_{hr}^c - T_{hr}^{sea} = \gamma(Z_{hr}^c - Z^{sea}). \quad (2)$$

Elevation values are from the Global Multi-resolution Terrain Elevation Data 2010 (Danielson and Gesch, 2011), with the spatial resolution of 30 arcsec. This method is used to downscale mean, maximum, and minimum daily temperatures.

2.1.2 Precipitation

CHELSA considers only orographic precipitation (Karger et al., 2023), which is done by computing the wind effect index H for each high-resolution cell. This index reflects how much moisture gets pushed up towards the top of a mountain as well as rain shadow in its leeward direction, and it is computed using u-wind and v-wind components from CMIP6 data. Those components were interpolated to the high-resolution grid with a B-spline, and then were projected onto a world Mercator projection.

$$H = H_{W,L} \rightarrow d_{LH_i} < 0 \times H_{W,L} \rightarrow d_{LH_i} \geq 0, \quad (3)$$

$$H_W = \frac{\sum_{i=1}^n \frac{1}{d_{WH_i}} \tan^{-1} \left(\frac{d_{WZ_i}}{d_{WH_i}^{0.5}} \right)}{\sum_{i=1}^n \frac{1}{d_{LH_i}}} + \frac{\sum_{i=1}^n \frac{1}{d_{LH_i}} \tan^{-1} \left(\frac{d_{LZ_i}}{d_{LH_i}^{0.5}} \right)}{\sum_{i=1}^n \frac{1}{d_{LH_i}}} \quad (4)$$

$$85 \quad H_L = \frac{\sum_{i=1}^n \frac{1}{\ln(d_{WHi})} \tan^{-1} \left(\frac{d_{LZi}}{d_{WHi}^{0.5}} \right)}{\sum_{i=1}^n \frac{1}{\ln(d_{LHi})}} \quad (5)$$

, where d_{WHi} and d_{LHi} denote the horizontal distances in windward (W) and leeward (L) directions, while d_{WZi} and d_{LZi} are the corresponding vertical distances. The summations in (4) and (5) are performed within a circle with the radius of 75 kilometers.

90 The H index is then corrected according to the atmospheric boundary layer height to account for the contribution of the surface pressure level to the wind effect. Lastly, the low-resolution precipitation p_{lr} is multiplied by the corresponding H indices and normalized to obtain high-resolution precipitations p_{hr} , so that within each low-resolution grid cell the sum of the values p_{hr} remains equal to p_{lr} (see section Methods in Karger et al. (2021)).

2.1.3 Surface downwelling shortwave radiation (RSDS)

The total shortwave radiation, measured in W/m^2 is represented as in Karger et al. (2023), Sect. 2.2.2:

$$95 \quad S_n = S_s + S_h. \quad (6)$$

Here, S_s is direct solar radiation reaching the surface, computed according to the position of the Sun with respect to the high-resolution grid cell. Diffuse solar radiation S_h , which is the energy re-emitted by the atmosphere, takes into account the percentage of the sky observable from a grid cell.

100 Computation of S_s component starts with astronomical equations. For the sun elevation angle θ , sun azimuth φ , latitude λ , the solar declination angle δ , the Julian day number J , hour h , and the hour angle in degrees $\bar{\omega}$, we have the following:

$$\begin{aligned} \sin \theta &= \cos \lambda \cos \delta \cos \bar{\omega} + \sin \lambda \sin \delta \\ \cos \varphi &= \frac{\cos \delta \cos \bar{\omega} - \sin \theta \cos \lambda}{\sin \lambda \cos \theta} \\ \delta &= 23.45 \cdot \sin \left(\frac{360^\circ [284 + J]}{365} \right) \\ \bar{\omega} &= 15^\circ (12 - h). \end{aligned} \quad (7)$$

Using these identities, $\cos \gamma$ is computed as

$$\cos \gamma = \cos \beta \cdot \sin \theta + \sin \beta \cdot \cos \theta \cdot \cos(\varphi - \alpha), \quad (8)$$

105 where γ is the angle between the Sun beam and the normal to the terrain, while α and β are surface slope and aspect. Then, S_s is computed using constants $G_{sc} = 1367 \text{ kW} \cdot m^2$, $\tau = 0.8$, and air optical thickness m defined in formula (13) of Karger et al. (2023):

$$S_s(h) = \varsigma(h) \cdot G_{sc} \cdot \tau^m \cdot \cos \gamma. \quad (9)$$

Diffuse solar radiation is calculated as

$$S_h = (0.271 - 0.294\tau^m)G_{sc}\Psi_s, \quad (10)$$

110 where Ψ_s is the sky view factor computed as

$$\Psi_s = \frac{1}{N} \sum_{i=1}^N [\cos \beta \cos \varphi_i + \sin \beta \cos(\Phi_i - \alpha) \cdot (90 - \varphi_i - \sin \varphi_i \cos \varphi_i)] \quad (11)$$

for $N = 8$ azimuth directions Φ_i and the corresponding horizon angles φ_i .

$$rsds = \bar{S}_n(1 - 0.75 \cdot clt^{3.4}), \quad (12)$$

where $\bar{S}_n$ is an average of S_n over 24 hours, and clt is the cloud cover computed according to formulas (19)–(22) Karger et al.

115 (2023).

To summarize this procedure, we note that the S_s and S_h components are obtained by taking into account shadowing and obstruction of light, the position of the Sun, the slope and the aspect of the terrain, and cloud cover resulting from orographic precipitation formation.

2.2 Bootstrap hypothesis test

120 In Sect. 4, we try to find systematic differences between high and low resolutions by comparing the corresponding regional averages of LPJ-GUESS output variables. We do this by testing if the mean values of the samples of the output variables are equal on both resolutions. Since on the 2 resolutions LPJ-GUESS produces outputs with different distribution variance, we are interested in the mean values only instead of the whole distributions. In order to test whether two random samples come from distributions having equal means, we employ the so-called bootstrap two-sample heterogenic location test (see Dikta and
125 Scheer, 2021, Sect. 4.3) and its implementation in R package boot (Angelo Canty and B. D. Ripley, 2024; Davison and Hinkley, 1997). Assume that there are two samples: $\mathbf{X}_n = \{X_1, \dots, X_n\}$ and $\mathbf{Y}_m = \{Y_1, \dots, Y_m\}$. $\mathbf{X}_n$ is drawn from distribution F , and $\mathbf{Y}_m$ is drawn from G , where both distributions are univariate and have finite (but not necessarily equal) variances σ_x and σ_y as well as mean values μ_x and μ_y . The goal is to test null-hypothesis H_o against the alternative H_a :

$$H_o : \mu_x = \mu_y \quad \text{vs} \quad H_a : \mu_x \neq \mu_y \quad (13)$$

130 Following the standard bootstrap approach,

$$X_{li}^*, \quad i \in [1, \dots, n], \quad \text{and} \quad Y_{lj}^*, \quad j \in [1, \dots, m], \quad l \in [1, \dots, L], \quad (14)$$

are obtained via sampling with replacement from sets $\{X_1, \dots, X_n\}$ and $\{Y_1, \dots, Y_m\}$ respectively, and bootstrap counterparts of the sample means and variances are computed:

$$\begin{aligned} \bar{X}_{l;n}^* &= \frac{1}{n} \sum_{i=1}^n X_{li}^*, \quad s_{l;x}^{*2} = \frac{1}{n-1} \sum_{i=1}^n (X_{li}^* - \bar{X}_{l;n}^*)^2, \\ \bar{Y}_{l;m}^* &= \frac{1}{m} \sum_{j=1}^m Y_{lj}^*, \quad s_{l;y}^{*2} = \frac{1}{m-1} \sum_{j=1}^m (Y_{lj}^* - \bar{Y}_{l;m}^*)^2 \end{aligned} \quad (15)$$

135 Under H_o , the following limiting result holds, which is used by the test:

$$T^* := \frac{(\bar{X}_{l;n}^* - \bar{X}_n) - (\bar{Y}_{l;m}^* - \bar{Y}_m)}{\sqrt{s_{l;x}^{*2}/n + s_{l;y}^{*2}/m}} \xrightarrow{d} \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{s_x^2/n + s_y^2/m}} =: T,$$

when $n \rightarrow \infty$, $n/m \rightarrow \lambda$, $0 < \lambda < \infty$ $L > n + m$. (16)

T is used as a test statistic with the distribution of T^* , providing a p-value for the difference of the two populations. Here, the classical two-sided test is used: if T lies beyond 95% of the generated sample of T^* (no matter to which side), then the null-hypothesis is rejected, otherwise— accepted.

140 We chose this test for two reasons. First, its only restriction on the data is the existence of a finite variance, which is justified by the conservation laws of physics. On the contrary, parametric tests, such as Z-test, require assuming at least a certain distribution family, which is too constraining, as we apply the testing procedure to a number of different quantities. Also, a standard t-test cannot be used due to differences in variations of samples we obtain— the low resolution data have smaller variation than the high-resolution ones (see Sect. 4.2). Second, our data are too small to make use of the central limit theorem; see Sect. 4.1 for the setup of the experiment. In the context of studies of large regions over the historical period 1850–2014, LPJ-GUESS simulations are computationally demanding especially on the high resolution. Because of this, generating samples that contain more than 50–100 observations of averages in the Alpine region is a challenging task on both 0.5° and $0.083(3)^\circ$ resolutions. When the number of observations is larger, for example in regional studies, the central limit theorem ensures convergence in law of the means of the two samples to normal distributions, and it becomes possible to use Welch’s t-test
150 (Welch, 1947) as a good alternative to the bootstrap test.

2.3 LPJ-GUESS

2.3.1 Model description

LPJ-GUESS is a process-based Dynamic Global Vegetation Model (DGVM) that incorporates ecosystem biogeochemistry, water cycling, and tree demography (Smith et al., 2001, 2014). The model is able to simulate several types of land cover and
155 land use change (Lindeskog et al., 2013; Olin et al., 2015; Lindeskog et al., 2021). At any given geographical location (*grid-cell*), the different land cover types constitute separate *stands*, which share the same climate forcings. We restrict our test here to a model configuration that only includes potential natural vegetation. In natural forest stands, vegetation is represented as a mixture of woody Plant Functional Types (PFTs), divided into age classes or *cohorts*, and a grassy understory. Yearly establishment of new cohorts is subject to prescribed bioclimatic limits, which are specific to each PFT. Trees and grasses coexist in
160 the same *patch*, which roughly represents the area of influence of one large, mature tree. Competition for available water, light and nutrients determines the daily rate at which each cohort absorbs atmospheric CO_2 . At the end of every simulation year, the assimilated carbon is allocated to leaves, sapwood, or roots according to a set of PFT-specific allometric constraints. Within the stand, horizontal heterogeneity is represented by simulating a number of replicate patches. Establishment of new cohorts, death of individuals, and vegetation-destroying disturbances are modeled as stochastic processes, giving rise to different suc-

165 cessional histories for each patch. Ecosystem pools and fluxes are estimated by averaging over patches. Wildfires are simulated explicitly with the SIMFIRE-BLAZE submodel (Knorr et al., 2014, 2016; Rabin et al., 2017). The potential burned area for each gridcell is calculated annually as a function of land cover type, meteorological information, and the fraction of absorbed photosynthetically-active radiation (FAPAR) as a proxy for vegetation cover. This is then used to model ignition stochastically, and calculate combustion rates and associated carbon and nitrogen fluxes. A comprehensive description of the fire submodel is
170 available in Molinari et al. (2021).

In this paper we used the ‘European Applications’ branch of LPJ-GUESS. This version differs from the standard (global) version in that the PFTs are parametrized based on observed characteristics of common European species (Hickler et al., 2012; Gregor et al., 2022, 2024).

2.3.2 Model modification

175 Stochastic events in LPJ-GUESS are triggered by the outcomes of a Random Number Generator (RNG). In the unmodified version of the model code, each stand keeps its own random number sequence, which is initialized (seeded) with a hard-coded value when the stand is created at the beginning of the simulation. This implies that all stands in the simulation derive their stochasticity from the same random number sequence. However, ensemble experiments require that observations within the ensemble are statistically independent. In order to emulate statistical independence of stands and gridcells between ensemble
180 realizations, we modified the model code to initialize each stand’s seed as follows:

$$\text{Seed}_{igs} = (1000000 + N_{gc} \cdot i + g) \cdot 1000 + s, \quad (17)$$

where $i = 0, \dots, N - 1$ identifies the simulation within the ensemble, N_{gc} is the total number of gridcells in the simulation domain, $g = 0, \dots, N_{gc} - 1$ runs over gridcells, and $s = 0, \dots, 999$ is a unique stand identifier. This method ensures that each stand in every realization of the ensemble draws random numbers from a different sequence, up to 1000 stands per gridcell.

185 We emphasize, however, that each gridcell contains only one natural vegetation stand in the present study.

3 Climate data downscaling

3.1 Data processing

The CHELSA algorithm was used to downscale ISIMIP3b mean, maximum and minimum daily temperatures, precipitation rate, and downwelling solar radiation in the domain defined by $\lambda \in [26.75^\circ\text{W}, 35.25^\circ\text{E}]$; $\phi \in [34.75^\circ\text{N}, 71.75^\circ\text{N}]$, where λ
190 and ϕ are geographical longitude and latitude, respectively. This domain encompasses the continental European Union plus Norway, Iceland, Switzerland, the United Kingdom, the non-EU Balkan states, Moldova, Belarus, and parts of Ukraine, Russia, Morocco, Algeria, and Tunisia. CHELSA generates one TIFF file per day for each of the input variables at an output resolution of $0.0083(3)^\circ$ (approximately 1 km near the Equator). LPJ-GUESS simulations covering the target domain at this resolution are computationally impractical. We thus first upscaled the files to 0.05° by taking the mean of every 6×6 block of adjacent

0.0083(3)° × 0.0083(3)° gridcells. This upscaled version was stored in NetCDF format. Gaps produced by missing days in CHELSA’s output (~ 0.34% in the historical period, and fewer than 0.14% in the scenarios) were filled with previous day values. The daily NetCDF files were stacked along the time dimension, and we added CF-compliant metadata (Hassell et al., 2017). LPJ-GUESS simulates the whole target period in one gridcell before proceeding to the next location. To optimize data retrieval by the model code, the stacked NetCDF files were *rechunked* along the time dimension. This operation rearranges the internal structure of the file in a way that greatly enhances performance when reading the full time series at a single spatial location. The resulting files underwent quality control, which included checking that there were no missing days, ensuring that all values were non-negative, and manually assessing that annual mean values were reasonable. In addition to the variables downscaled with CHELSA, we remapped ISIMIP3b near-surface wind speed and air relative humidity data to high resolution by applying bilinear interpolation to the original files. These variables are required when running LPJ-GUESS with the SIMFIRE/BLAZE fire model (Knorr et al., 2014, 2016; Rabin et al., 2017) (see Sect. 5). The CHELSA original algorithm depends on a B-spline interpolation for wind, while we adopt here bilinear interpolation. Both techniques derive from the same class-polynomial interpolation, and bi-linear interpolation is expected to capture terrain heterogeneity better. Relative humidity is not included in the original CHELSA approach. The pipeline scripts were implemented in Bash Script and Python, and use the NetCDF Operators (Zender, 2008) and the Climate Data Operators (Schulzweida, 2023).

Table 1 summarizes the properties of the dataset. The data is freely accessible through KIT/IMK-IFU’s *thredds* storage server Otryakhin and Belda (2024), and is made available under the CC BY-SA 4.0 license.

Table 1. Characteristics of the final downscaled dataset. Variable names follow the ISIMIP3b nomenclature: daily average temperature, minimum, and maximum (‘tas’, ‘tasmin’, ‘tasmax’), precipitation (‘prec’), downwelling shortwave radiation (‘rsds’), wind speed (‘sfcwind’), and relative humidity (‘hurs’).

Spatial resolution	0.05°
Temporal resolution	Daily
Format	NetCDF4
Variables (CHELSA)	tas, tasmin, tasmax, pr, rsds
Variables (interp)	sfcwind, hurs
Spatial extent	$\lambda \in [26.75^{\circ}\text{W}, 35.25^{\circ}\text{E}]$ $\phi \in [34.75^{\circ}\text{N}, 71.75^{\circ}\text{N}]$
Temporal extent	1850 – 2100
GCM	MPI-ESM1-2-HR
SSP/RCP scenarios	Historical, SSP1-2.6, SSP3-7.0, SSP5-8.5
Total size	1.3 TB

Table 2. Characteristics of elevations in the study and the control regions computed using values from GMTED2010 (Danielson and Gesch, 2011). All values are in meters.

Region	Min.	1st Qu.	Median	Mean	SD	3rd Qu.	Max.
Study	0	748	1310	1369	761	1941	4505
Control	69	81	88.0	100.4	44.5	105.0	860.0

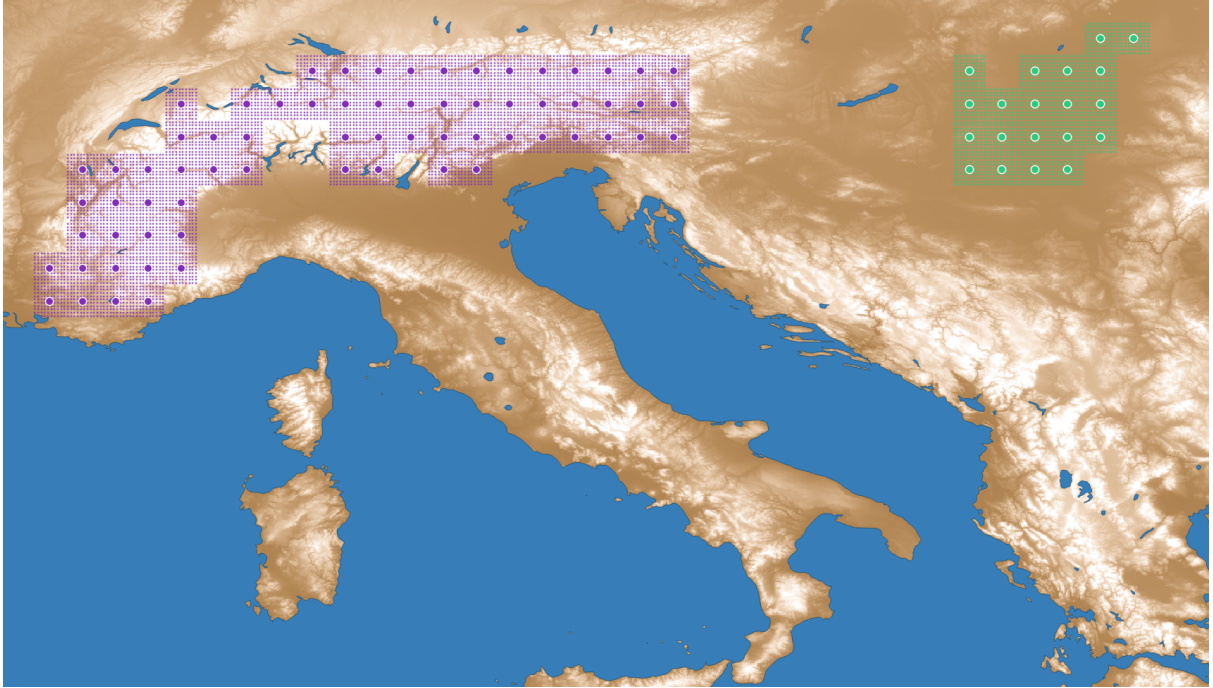


Figure 1. Study (violet) and control (green) regions for the ensemble experiment, overlaid on an elevation map based on values of GMTED2010 (Danielson and Gesch, 2011). The centers of low-resolution gridcells are indicated by solid circles. The smaller dots mark the location of the high-resolution gridcells. Elevation is encoded as a color gradient ranging from dark brown (low elevation) to white (high elevation).

4 Ensemble experiment

4.1 Setup

This experiment aims to find systematic differences in regional predictions between low- and high-resolution LPJ-GUESS simulations, arising from underrepresentation of orography-induced climate variability in the low-resolution forcings. To this end, we ran two sets of ensembles of LPJ-GUESS simulations (high- and low-resolution) in selected *study* and *control* regions. The simulations spanned the historical period 1850–2014. The low-resolution simulations were forced with ISIMIP3b climate,

while the high-resolution simulations were forced with the downscaled dataset. Both simulations use the same soil properties dataset, derived from the Digitized Soil Map of the World (Zobler, 1986), as in Sitch et al. (2003). In order to prevent introducing possible confounding factors, the soil information was not downscaled, and we kept nitrogen deposition at a constant pre-industrial rate of $2\text{kgNha}^{-1}\text{year}^{-1}$. We chose the Alps as our study region for its high variability in surface elevation. The control region, located between the Dinaric Alps and the Carpathian Mountains, was chosen to contain comparatively little mountainous terrain (Table 2), while being in close proximity to the Alpine region and of approximately the same size. The climate between the Pannonian basin and the European alps naturally differs but is still influenced by similar, large-scale circulation patterns that affect the European continent and the choice of the control region intended to prevent significant global differences. Figure 1 shows the study and control regions overlayed on an elevation map. The gridcell centers were chosen such that any low-resolution gridcell contains a block of 10×10 high-resolution gridcells, with no high-resolution gridcells outside these blocks (i.e., a perfect overlap of the high- and low-resolution grids). Prominent water bodies were avoided.

The experimental design is outlined in Fig. 2. We considered the regional averages of each modeled variable (Table 3), evaluated in the last year of the simulation (2014), as random variables. We run 50 simulations per region and resolution, giving rise to 44 ensembles of random variables ($11 \text{ modeled variables/simulation} \times 2 \text{ resolutions} \times 2 \text{ regions}$), each containing 50 observations of its random variable. For a given region and resolution, the simulations were identically set up, but used a different seed for the random number generator (see Sect. 2.3.2). Model runs were assumed to be independent and identically distributed. We note that such an assumption is not imposed on pairs of ecosystem variables within the same model run. The patch number was set to 20 in all simulations.

Let $\mu_{\text{hr}}^{\text{S}}$, $\mu_{\text{lr}}^{\text{S}}$, $\mu_{\text{hr}}^{\text{C}}$, and $\mu_{\text{lr}}^{\text{C}}$ denote the ensemble means of any of the modeled variables, with superscripts ‘S’ and ‘C’ identifying the study and control regions, and subscripts ‘hr’ and ‘lr’ denoting high- and low- resolution. The question of whether there are systematic differences between low- and high-resolution regional predictions in areas with high orographical variability can be cast in terms of two formal hypothesis tests:

$$H_o^{\text{S}} : \mu_{\text{hr}}^{\text{S}} = \mu_{\text{lr}}^{\text{S}} \quad (\text{There are no differences in the study region})$$

vs

$$H_a^{\text{S}} : \mu_{\text{hr}}^{\text{S}} \neq \mu_{\text{lr}}^{\text{S}}, \quad (\text{There are differences in the study region}), \quad (18)$$

and

$$H_o^{\text{C}} : \mu_{\text{hr}}^{\text{C}} = \mu_{\text{lr}}^{\text{C}} \quad (\text{There are no differences in the control region})$$

vs

$$H_a^{\text{C}} : \mu_{\text{hr}}^{\text{C}} \neq \mu_{\text{lr}}^{\text{C}}, \quad (\text{There are differences in the control region}). \quad (19)$$

where the subscript ‘o’ denotes the null hypotheses and ‘a’ identifies the alternatives. We tested these hypotheses for every variable by applying the bootstrap test method described in Sect. 2.2.

CHELSEA conserves the amount of precipitation in the upscaled output by construction (Sect. 2.1.2), but does not impose such constraint on mean temperature and radiation. In order to investigate the possible biases associated with this non-conservative

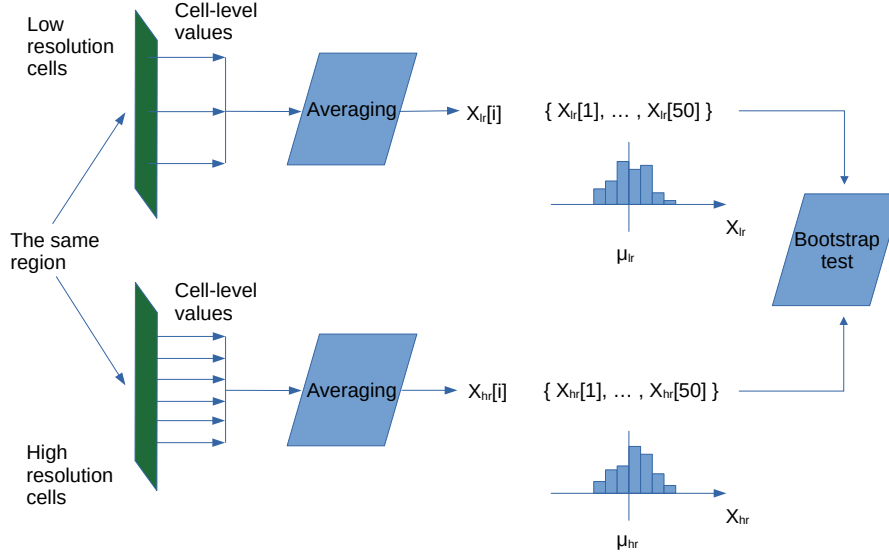


Figure 2. Scheme of computations in the ensemble experiment. Here, X is the average of values at the end of the computation period 1850–2014 in the region, lr and hr are the indicators of the low and high resolution correspondingly, $i = \{1, \dots, 50\}$ is the experiment id, μ 's are the sample mean estimates.

downscaling we repeated the above experiment, this time forcing the low-resolution simulations with CHELSA climate, up-scaled back to 0.5° by spatially averaging the high-resolution data. This ensures that mean temperature, radiation and precipitation are the same on both resolutions. In addition, we wished to investigate the impact of fire disturbance on the results of the above experiments. This was motivated by the fact that, in LPJ-GUESS, fire disturbance depends on the climate forcings through the fire model (Rabin et al., 2017). To explore this effect we rerun the last experiment with fire switched off. The characteristics of the three ensemble experiments are summarized in Table 4.

4.2 Results and analysis

The results of the first experiment (I3b/Fire) are shown in Table 5. We found statistically significant differences between the means of the high- and low-resolution samples ($\delta^S := \mu_{lr}^S - \mu_{hr}^S$) over the study region. The bootstrap test returned a p-value of 0 for all the variables, leading us to reject H_0^S in favor of the alternative H_a^S .

In the control region all differences between ensemble means, $\delta^C := \mu_{lr}^C - \mu_{hr}^C$, are substantially smaller than their study region counterparts, ranging from $|\delta^S/\delta^C| \sim 7$ in the case of NEP and Transpiration to ~ 116 in the case of C_{soil} . For the majority of the variables, the p-values indicate a statistically significant difference between ensemble means δ^C . For NEP and C_{lit} , however, we found no evidence in the data to reject the null hypothesis H_0^C . This indicates that, for these two variables, either there are no significant differences between μ_{lr}^C and μ_{hr}^C , or the ensemble size is too small to detect them with a sufficient

Table 3. List of ecosystem variables modeled by LPJ-GUESS that were included in the experiment. These include carbon fluxes (GPP, NPP, and NEP), carbon pools (C_{lit} , C_{soil} , C_{veg} , and C_{tot}), water cycle variables (Tr. and R_{off}), and vegetation structural variables (FPC and LAI). The units refer to regional aggregates (for all variables except FPC and LAI) and regional averages (for FPC and LAI) of the selected variables.

Variable	Description	Units
GPP	Gross Primary Productivity	$\text{kgC} \cdot \text{yr}^{-1}$
NPP	Net Primary Productivity	$\text{kgC} \cdot \text{yr}^{-1}$
NEP	Net ecosystem productivity	$\text{kgC} \cdot \text{yr}^{-1}$
C_{lit}	Amount of carbon stored in litter	kgC
C_{soil}	Amount of carbon stored in the soil	kgC
C_{veg}	Amount of carbon stored in above-ground vegetation	kgC
C_{tot}	Total amount of stored carbon	kgC
Tr.	Ecosystem transpiration	$\text{mm} \cdot \text{yr}^{-1}$
R_{off}	Runoff	$\text{mm} \cdot \text{yr}^{-1}$
FPC	Foliar Projective Cover	$\text{m}^2 \cdot \text{m}^{-2}$
LAI	Leaf Area Index	$\text{m}^2 \cdot \text{m}^{-2}$

Table 4. Ensemble experiments carried out in Sect. 4. The experiments differ in the climate data used to force the low-resolution simulations and on whether wildfires are allowed.

Experiment	LR forcing	Fire
I3b/Fire	ISIMIP3b	On
uCH/Fire	Upscaled CHELSA	On
uCH/NoFire	Upscaled CHELSA	Off

confidence level. Figure 3 shows δ^S and δ^C , as a fraction of μ_{hr} , for all variables. These values in the study region have the opposite sign to those in the control region, with the exception of NEP and R_{off} .

An additional bootstrap test was run to check whether random variability in the control region could give rise to differences between ensemble means as large as those seen in the study region. With low- and high-resolution values in place of X_i and Y_i , we generated the distribution of T^* from the left side of (16), and found the two-sided p-values using $-|\delta^S|$ and $+|\delta^S|$. These tests returned a p-value $p^X = 0$ for all variables, which strongly suggests that control-region differences as large as those found in the study region could not be explained by mere random variability of the ensemble means.

Table 6 shows the results of the second experiment (uCH/Fire). All bootstrap tests returned a p-value of 0 in the study region, again indicating very strong evidence of a mean difference between high- and low-resolution outputs. In the control region, 5 of

Table 5. Results of the I3b/Fire experiment. The ensemble means and standard deviations are denoted μ and σ , respectively. The superscripts S and C identify the study and the control regions, respectively. The subscripts hr and lr denote high- and low-resolution. The δ is used for the difference between the low- and high-resolution ensemble means for the given region. The p^S and p^C denote the p-values for the statistical tests defined by Eq. 18 and Eq. 19. The p^X are the p-values for the test defined in Sect. 4.2. A bold font face is used to identify tests with a p-value below the significance threshold. The ratios δ^S/μ_{hr}^S , δ^C/μ_{hr}^C , and δ^S/δ^C are expressed as percentages.

	GPP	NPP	NEP	C_{lit}	C_{soil}	C_{veg}	C_{tot}	Tr.	R_{off}	FPC	LAI
μ_{hr}^S	0.96	0.52	-0.05	10.7	13.7	15.8	40.2	328	827	0.70	2.64
σ_{hr}^S	0.0005	0.0003	0.0005	0.014	0.004	0.027	0.02	0.20	0.14	0.0006	0.002
μ_{lr}^S	1.19	0.66	-0.06	15.6	19.2	21.2	56.0	374	789	0.80	3.12
σ_{lr}^S	0.006	0.003	0.005	0.154	0.058	0.257	0.17	2.11	1.60	0.006	0.03
δ^S	-0.22	-0.13	0.01	-4.92	-5.44	-5.41	-15.8	-45.7	38.1	-0.10	-0.48
p^S	0	0	0	0	0	0	0	0	0	0	0
δ^S/μ_{hr}^S	-23.4	-25.5	-15.3	-46.0	-39.6	-34.2	-39.2	-13.9	4.6	-14.8	-18.3
μ_{hr}^C	1.08	0.50	0.08	7.27	8.93	17.4	33.6	402	44.5	0.82	3.14
σ_{hr}^C	0.0007	0.0004	0.0013	0.026	0.005	0.04	0.03	0.35	0.22	0.001	0.005
μ_{lr}^C	1.07	0.49	0.08	7.22	8.88	17.2	33.2	395	42.5	0.81	3.13
σ_{lr}^C	0.0073	0.0037	0.0114	0.216	0.041	0.45	0.33	3.14	1.96	0.013	0.036
δ^C	0.012	0.0055	-0.0014	0.050	0.047	0.23	0.33	6.79	2.00	0.004	0.014
p^C	0	0	0.41	0.10	0	0	0	0	0	0.022	0.003
δ^C/μ_{hr}^C	1.2	1.1	-1.7	0.7	0.5	1.3	1.0	2.0	4.0	0.5	0.5
δ^S/δ^C	-18	-24	-7	-98	-116	-24	-48	-7	19	-25	-34
p^X	0	0	0	0	0	0	0	0	0	0	0

11 tests failed to reject the null hypothesis: NEP, C_{veg} , Tr., FPC, and LAI. Interestingly, the C_{lit} test switched to returning a significant p-value. The relative differences between ensemble means in the study and control regions, δ^S/μ_{hr}^S and δ^C/μ_{hr}^C , are now both negative (Fig. 3) except for the high-resolution R_{off} . Runoff shows the largest relative discrepancy with respect to the previous experiment, but the difference in absolute terms is very small. This sign switch of δ^C with respect to I3b/Fire suggests that CHELSA’s non-conservative properties introduce a bias of sign opposite to the response of the model to the altitude-driven climate differences. The relative importance of this effect is much larger over the control region, where it determines the sign of the overall difference δ^C . The change in magnitude of δ^C with respect to I3b/Fire, $|\Delta\delta^C|$, was between $\sim 110\%$ and $\sim 310\%$, except for NEP ($\sim 57\%$). By contrast, $|\Delta\delta^S|$ is much smaller, ranging between $\sim 0.04\%$ and $\sim 14\%$. This may explain the significance switch in the C_{lit} test over the control region; the non-conservative bias present in the I3b/Fire experiment nearly compensates the altitude-induced differences for this variable. This brings the high- and low-resolution means closer together,

Table 6. Results of the uCH/Fire experiment. Repeated symbols are as in Table 5. $|\Delta\delta^S|$ and $|\Delta\delta^C|$ are the changes in magnitude of δ^S and δ^C with respect to I3b/Fire, expressed as a percentage.

	GPP	NPP	NEP	C_{lit}	C_{soil}	C_{veg}	C_{tot}	Tr.	R_{off}	FPC	LAI
μ_{hr}^S	0.96	0.52	-0.05	11	14	16	40	328	827	0.70	2.64
σ_{hr}^S	0.0005	0.0003	0.0005	0.014	0.004	0.027	0.019	0.20	0.14	0.0006	0.002
μ_{lr}^S	1.18	0.65	-0.06	15.5	18.5	21	55	367	791	0.81	3.12
σ_{lr}^S	0.0055	0.003	0.004	0.15	0.04	0.22	0.15	2.14	1.5	0.006	0.025
δ^S	-0.22	-0.13	0.01	-4.8	-4.8	-5.5	-15	-39	36	-0.11	-0.48
p^S	0	0	0	0	0	0	0	0	0	0	0
δ^S/μ_{hr}^S	-22.8	-24.4	-14.1	-45.2	-35.1	-34.8	-37.7	-12.0	4.4	-15.9	-18.2
μ_{hr}^C	1.08	0.50	0.08	7.3	8.9	17.4	33.6	401	44	0.82	3.14
σ_{hr}^C	0.0007	0.0004	0.0013	0.026	0.005	0.04	0.03	0.35	0.22	0.002	0.005
μ_{lr}^C	1.08	0.50	0.08	7.37	9	17.47	33.8	402	48	0.82	3.15
σ_{lr}^C	0.007	0.003	0.01	0.24	0.04	0.4	0.3	3.7	2	0.01	0.04
δ^C	-0.004	-0.003	-0.0006	-0.1	-0.08	-0.09	-0.26	-0.7	-4.2	-0.002	-0.009
p^C	0	0	0.75	0.0025	0	0.16	0	0.21	0	0.26	0.10
δ^C/μ_{hr}^C	-0.4	-0.6	-0.7	-1.4	-0.9	-0.5	-0.8	-0.2	-9.4	-0.3	-0.3
δ^S/δ^C	56	41	-12	49	63	63	58	56	-9	53	52
p^X	0	0	0	0	0	0	0	0	0	0	0
$ \Delta\delta^S $	0.73	1.3	27	1.8	11	1.8	4	14	5	11	0.04
$ \Delta\delta^C $	133	156	57	299	263	138	180	110	310	154	166

which makes it more difficult to discern them. When this bias is removed in uCH/Fire, the difference between ensemble means increases, and the bootstrap test is able to detect it.

The outcome of the uCH/NoFire experiment is shown in Table 7. Like before, the impact on the control region is comparatively larger than in the study region, as shown by the values of $|\Delta\delta^S|$ and $|\Delta\delta^C|$, and by the fact that some of the tests in the control region returned switched p-values again.

We summarize the analysis described in this section as follows. It was demonstrated that high- and low-resolution simulations produce significantly different average predictions over a study region with high elevation variability. Differences in the control region were also detected, but they are much smaller than in the study region. Climate data downscaled with CHELSA introduces a bias related to its non-conservative treatment of temperature and radiation. This bias is comparable in magnitude to the altitude-related differences in the control region, but small in relation to the magnitude of the variables, and largely inconsequential in the study region. When this bias was removed, average NEP, C_{veg} , Tr., and FPC were indistinguishable

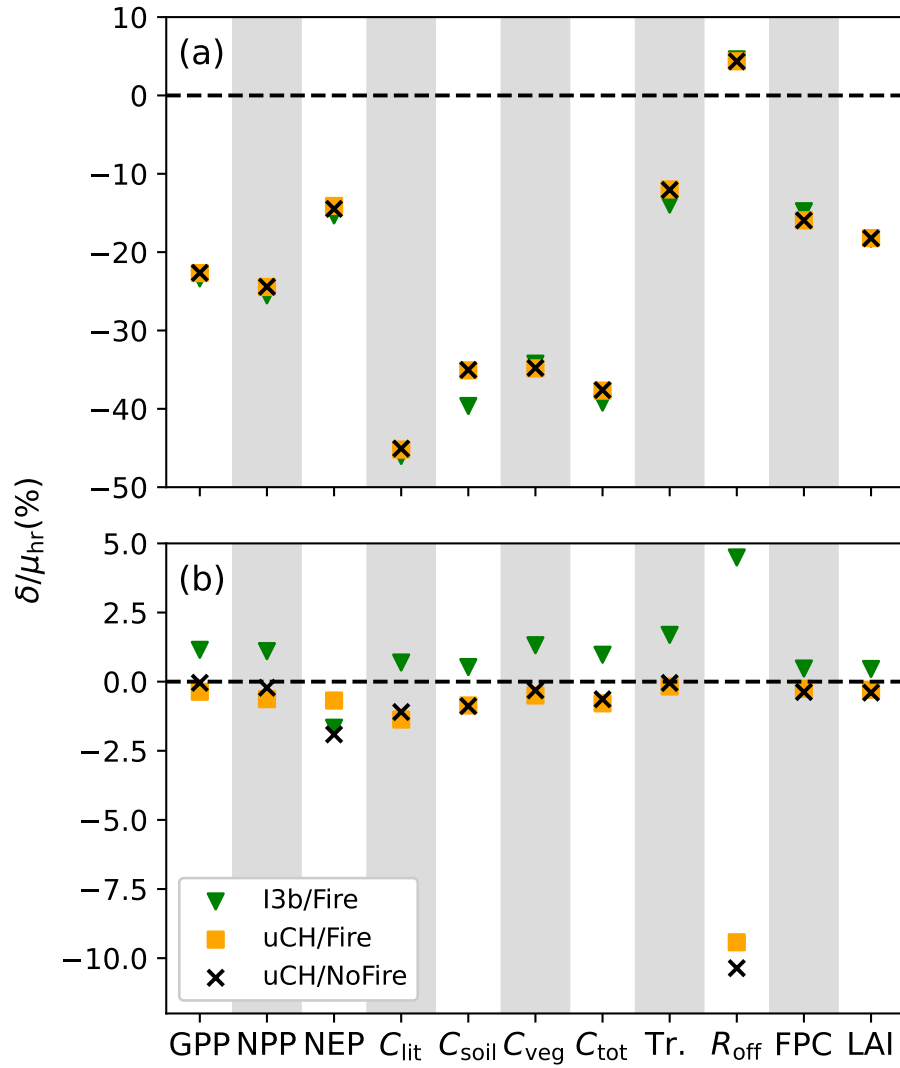


Figure 3. Mean difference between high- and low- resolution ensemble means, as a percentage of the high-resolution ensemble mean, over the study region (a) and the control region (b). The different symbols represent the three experiments; I3b/Fire: low-resolution simulations are forced with ISIMIP3b climate; uCH/Fire: low-resolution simulations are forced with upscaled CHELSA climate; uCH/NoFire: low-resolution simulations are forced with upscaled CHELSA climate and fire is switched off.

in the high- and low-resolution simulations. Fire was found to be a significant contributor to the ensemble mean differences in the control region.

Table 7. Results of the uCH/NoFire experiment. Repeated symbols are as in Table 5. $|\Delta\delta^S|$ and $|\Delta\delta^C|$ are the changes in magnitude of δ^S and δ^C with respect to uCH/Fire, expressed as a percentage.

	GPP	NPP	NEP	C_{lit}	C_{soil}	C_{veg}	C_{tot}	Tr.	R_{off}	FPC	LAI
μ_{hr}^S	0.96	0.53	-0.05	11	14	16	40	328	827	0.7	2.64
σ_{hr}^S	0.0005	0.0003	0.0004	0.013	0.005	0.02	0.01	0.2	0.16	0.0005	0.002
μ_{lr}^S	1.18	0.65	-0.06	16	19	21	55	368	791	0.81	3.12
σ_{lr}^S	0.005	0.003	0.005	0.18	0.041	0.28	0.18	2	1.48	0.006	0.022
δ^S	-0.22	-0.13	0.007	-4.8	-4.8	-5.5	-15	-40	36	-0.11	-0.48
p^S	0	0	0	0	0	0	0	0	0	0	0
δ^S/μ_{hr}^S	-22.6	-24.4	-14.5	-45.1	-35.0	-34.8	-37.6	-12.1	4.3	-15.9	-18.2
μ_{hr}^C	1.08	0.5	0.08	7.28	8.93	17	34	402	45	0.82	3.14
σ_{hr}^C	0.0008	0.0004	0.0011	0.02	0.004	0.042	0.03	0.3	0.22	0.0014	0.004
μ_{lr}^C	1.08	0.5	0.08	7.36	9.01	17	34	402	49	0.82	3.15
σ_{lr}^C	0.007	0.004	0.01	0.215	0.036	0.381	0.3	3.5	1.99	0.013	0.043
δ^C	-0.0004	-0.0011	-0.0016	-0.08	-0.079	-0.055	-0.21	-0.2	-4.6	-0.003	-0.013
p^C	0.63	0.09	0.26	0.0025	0	0.37	0	0.69	0	0.09	0.04
δ^C/μ_{hr}^C	-0.04	-0.22	-1.9	-1.1	-0.9	-0.3	-0.6	-0.05	-10.4	-0.4	-0.4
δ^S/δ^C	545	117	-5	60	61	100	71	215	-8	36	39
p^X	0	0	0	0	0	0	0	0	0	0	0
$ \Delta\delta^S $	0.14	0.08	2.7	0.09	0.04	0.18	0.08	0.53	0.28	0.09	0.23
$ \Delta\delta^C $	90	65	167	20	2.9	37	19	74	10	48	36

CHELSEA-downscaled climate data is closer to observations than the original, coarse resolution data (Karger et al., 2021, 2023).
 295 This motivates us to consider the difference between high- and low-resolution simulations as a systematic bias incurred when
 running LPJ-GUESS at low resolution, arising from the underrepresentation of orographical climate variability.

5 Comparison of Europe-wide simulations

5.1 Setup

In order to assess the impact of systematic biases in low-resolution LPJ-GUESS outputs on a European-regional level, we ran
 300 two simulations, at high and low resolutions, in the European domain from Sect. 3.1 (Table 1). The input to the model is as in
 the ensemble experiment, except now we use historical ISIMIP nitrogen deposition data (Tian et al., 2018). Both simulations
 were fed with the original $0.5^\circ \times 0.5^\circ$ data. To capture coastline features and inland water bodies as accurately as possible for

each resolution, we drop the restriction of one-to-one correspondence between blocks of 10×10 high-resolution gridcells and the low-resolution ones (see Sect. 4.1). The number of patches was set to 100 for both runs, and wildfires were enabled.

305 5.2 Analysis and results

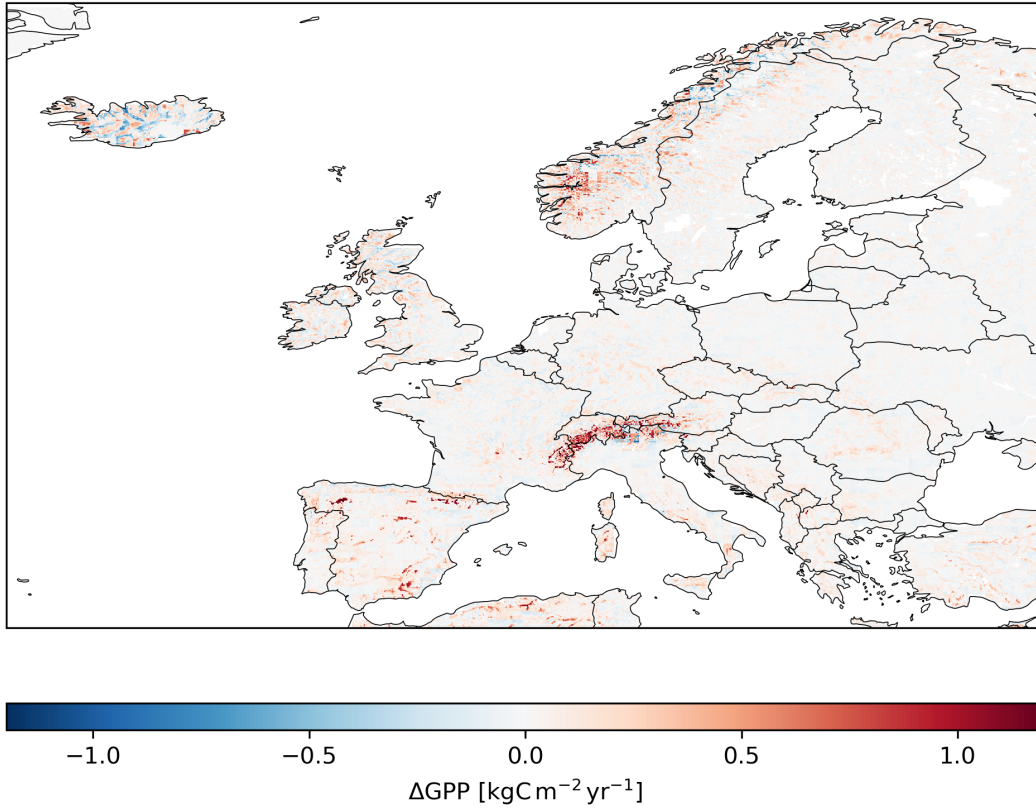


Figure 4. Climate-response bias in low-resolution modeled GPP, calculated as the difference between the low- and high-resolution predictions, averaged over the period 2010-2014. The value from every high-resolution gridcell was subtracted from the value in the corresponding 10×10 low-resolution block. Only fully overlapping low- and high- resolution gridcells are represented. Red indicates a higher GPP value in the low-resolution run than in the high-resolution run.

Forcing LPJ-GUESS with low-resolution climate data introduces a bias in average predictions, related to the underrepresentation of climate spatial variability (Sect. 4). Figure 4 shows this climate-response bias for GPP, averaged over the 2010-2014 period. Similar maps for the rest of the variables can be found in the supplementary materials (Sect. 1: Supplement to Comparison of Europe-wide simulations). The most prominent discrepancies concentrate over highly mountainous regions, such as

310 the Alps, the Spanish mountains and the Scandinavian Mountains. Large differences are also seen in Iceland, where the rapidly changing elevation leads to high spatial variability above and below the low-resolution predictions.

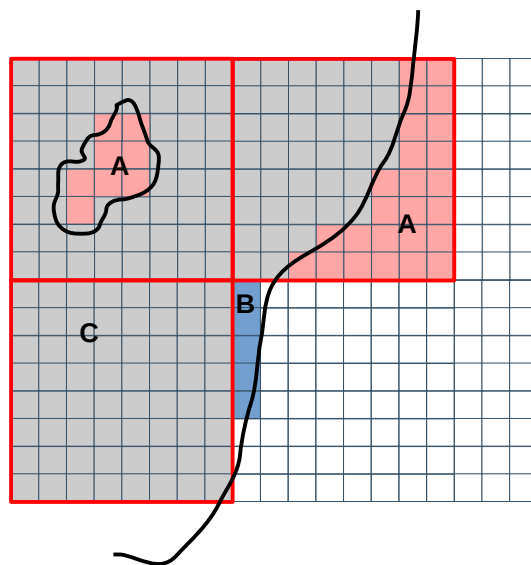


Figure 5. Three low-resolution gridcells (outlined in red) projected onto a high-resolution grid. A small lake and the coastline are represented with black, thick lines. The sea is to the right of the coastline. Red-shaded regions (marked A) indicate areas that are considered land in low-resolution simulations and water in high-resolution simulations. The blue-shaded area (marked B) is accounted for in a high-resolution run, but not in a low-resolution run. Gray areas (marked C) are represented in both high- and low-resolution simulations. White areas are considered water points in both simulations.

Additional bias results from the limitations of the low-resolution grid in representing areas around coastlines and inland water bodies (Fig. 5). In a low resolution simulation, some gridcells protrude outside the coastline, thus covering some sea-surface area (marked A), which is simulated as land. Similarly, the low-resolution grid cannot resolve small lakes, which adds to the overestimation of land-surface area. By contrast, some land-surface areas close to the seashore (marked B) are correctly accounted for in high-resolution simulations, but cannot be captured in low-resolution. In the European domain under consideration, these two counteracting effects amount to a $\sim 3.5\%$ increase in simulated surface area in the low resolution runs. This leads to a *shoreline-representation* bias in regional estimates.

Aggregating GPP across the domain yields an average of 7.01 PgCyr^{-1} over the period 2010-2014 for the high-resolution simulation, and 7.40 PgCyr^{-1} for the low-resolution simulation (a 5.6% increase). The climate-response and shoreline-representation contributions to this increase are $\delta_{\text{cli}} = 2.1\%$ and $\delta_{\text{sho}} = 3.4\%$, respectively. Table 8 shows aggregate high- and low-resolution values for the rest of the selected variables. In this region, the shoreline-representation bias is larger in magnitude than the climate-response bias for all variables. The largest relative discrepancy is seen in C_{lit} ; a 6.6% increase respect to the high-resolution value, with contributions $\delta_{\text{cli}} = 2.9\%$ and $\delta_{\text{sho}} = 3.7\%$ (the bias in NEP is even larger, but NEP is a very small quantity resulting from the difference of two large quantities (GPP and ecosystem respiration), and hence very sensitive

to small variations in either of those terms. In the case of aggregate runoff, the shoreline-representation bias ($\delta_{\text{sho}} = 4.1\%$) and the climate-response bias ($\delta_{\text{cli}} = -1.3\%$) act in opposite directions, adding up to a net total of $\delta_{\text{tot}} = 2.8\%$. The calculation of the climate-response and shoreline-representation contributions to the total bias is detailed in Appendix A.

Table 8. Comparison of domain-wide aggregates of selected ecosystem variables for the high- and low- resolution European simulations (HR and LR, respectively). δ_{tot} is the total bias, δ_{sho} is the shoreline-representation bias, and δ_{cli} is the bias arising from the difference in climate forcings. Percentages are calculated with respect to the high-resolution values.

	HR	LR	δ_{tot}	δ_{sho}	δ_{cli}
GPP	7.01	7.40	0.39 (5.6%)	0.24 (3.4%)	0.15 (2.1%)
NPP	3.68	3.90	0.22 (5.9%)	0.12 (3.3%)	0.09 (2.6%)
NEP	-0.405	-0.434	-0.029 (-7.3%)	-0.014 (-3.5%)	-0.015 (-3.8%)
C_{lit}	64.1	68.3	4.2 (6.6%)	2.3 (3.7%)	1.9 (2.9%)
C_{soil}	72.0	76.7	4.7 (6.5%)	2.6 (3.6%)	2.1 (2.9%)
C_{veg}	99.0	104.4	5.4 (5.5%)	3.0 (3.1%)	2.4 (2.4%)
C_{tot}	235.0	249.4	14.4 (6.1%)	8.0 (3.4%)	6.4 (2.7%)
Tr.	2.322	2.412	0.090 (3.9%)	0.076 (3.3%)	0.014 (0.6%)
R_{off}	2.54	2.61	0.07 (2.8%)	0.10 (4.1%)	-0.03 (-1.3%)
FPC	0.305	0.321	0.016 (5.2%)	0.011 (3.6%)	0.005 (1.6%)
LAI	1.14	1.20	0.06 (5.2%)	0.04 (3.4%)	0.02 (1.8%)

6 Discussion

330 Earlier work by Müller and Lucht (2007) showed little impact on model results when running the LPJ DGVM between 10° and 0.5° , at 0.5° intervals, suggesting that a resolution of 0.5° is still too coarse to account for relevant effects of spatial heterogeneity. Our study suggests that the impacts of resolution on the modeled output, linked to the influence of orography on the input climate, become noticeable at higher resolutions. The relative importance of these effects strongly depends on the focus region. Europe-wide simulations show an impact of resolution on aggregated ecosystem pools and fluxes of $\sim 3\%$,
335 likely smaller than the uncertainty derived from the spread in climate forcings by different GCMs (see, e.g., Schaphoff et al., 2006; Morales et al., 2007; Schurgers et al., 2018). By contrast, these differences increase up to $\sim 46\%$ in an Alpine region. Additional bias may result from poor representation of shorelines and small inland water bodies, but this effect could be mitigated by scaling the model output by the land-cover fraction in the affected gridcells. In areas of low variability in surface elevation, the difference between LPJ-GUESS outputs at different resolutions is much smaller and may be safely ignored

340 in calculations involving regional averages of ecosystem variables. For this type of studies, one could optimize the resource requirements of the simulations by using a coarser resolution in areas with low elevation variability.

The high-resolution simulations were performed on a grid that captures coastlines and water bodies more precisely, and are driven by climate that is generally closer to the observed regional climate (see validation sections in Karger et al., 2021, 2023). This motivates us to interpret these differences as systematic biases incurred when running LPJ-GUESS on a coarse grid. We
345 defer evaluating the simulation results against observational data to future studies. However, we note here that if the model output on low resolution was closer to observations, that would suggest that the model needs recalibration or revision.

Since geographical features and climate effects are not related to intrinsic properties of LPJ-GUESS, we infer that predictions by other DGVMs are likely to be affected in a similar manner. We note, however, that gridcells in LPJ-GUESS are independent of each other (there is no lateral information flow) and completely unaware of gridcell size. By contrast, other models may in-
350 clude processes, such as lateral matter transport, which are sensitive to the coarseness of the grid. This introduces an additional dependence of the output on resolution, on top of the effects discussed in this study.

One possible mechanism underpinning the difference in modeled GPP between high- and low-resolution simulations in areas with high elevation variability is the non-linear relationship between the mean gridcell temperature and the duration of the growing season, which is dynamically calculated by LPJ-GUESS. The linear relationship between temperature and
355 elevation (Eqs. 1 and 2) implies that air temperatures in higher parts of resolved mountainous areas are lower than the average value in the corresponding low-resolution gridcell (Fig. 6), causing a shorter growing season. The lower parts will, in turn, experience a longer growing season. The shorter growing season in high areas leads to reduced productivity and vegetation cover. Because of the non-linear response of the model to climate forcings, this is not fully compensated by the additional productivity in the lower, warmer parts. A similar argument can be made for the photosynthetic rate, which is temperature-
360 dependent. On the other hand, rainfall redistribution in the high-resolution grid may provide a counteracting effect. CHELSA tends to concentrate the total amount of rainfall towards high-elevation areas to account for the influence of orography on precipitation (Fig. 7), which may reduce water availability for plant growth in the lower areas. The radiation downscaling algorithm is more involved, and includes the effects of orographical features as well as those of the position of the Sun (Sect. 2.1.3). Nonetheless, there is a clear pattern in Fig. 8 showing that higher parts of mountains receive less solar radiation per
365 square meter than the corresponding low-resolution value, while the lower parts— more. This suggests that the increased cloud cover resulting from orographic precipitation leads to a decrease in average radiation in areas of higher elevation difference. This effect contributes to the reduction of vegetation in places with high elevations. Another counteracting factor is the excess land simulated in a grid too coarse to resolve small inland water bodies. The interplay between these factors will depend on the specific region being simulated, which emphasizes the complexity of the model's response to orographical and climate drivers.

370 There are many other modeled processes that respond non-linearly to climate forcings. Leaf-level photosynthesis shows a saturating (as opposed to linear) response to absorbed photosynthetically-active radiation when not limited by RuBisCo production (see Haxeltine and Prentice, 1996, for a discussion of the scaling of leaf-level photosynthesis to canopy-level productivity). Soil water transport follows a power law of available water content, which in turn depends on the amount of rainfall (see Gerten et al., 2004). The amount of radiation reaching the forest floor, which determines potential establishment

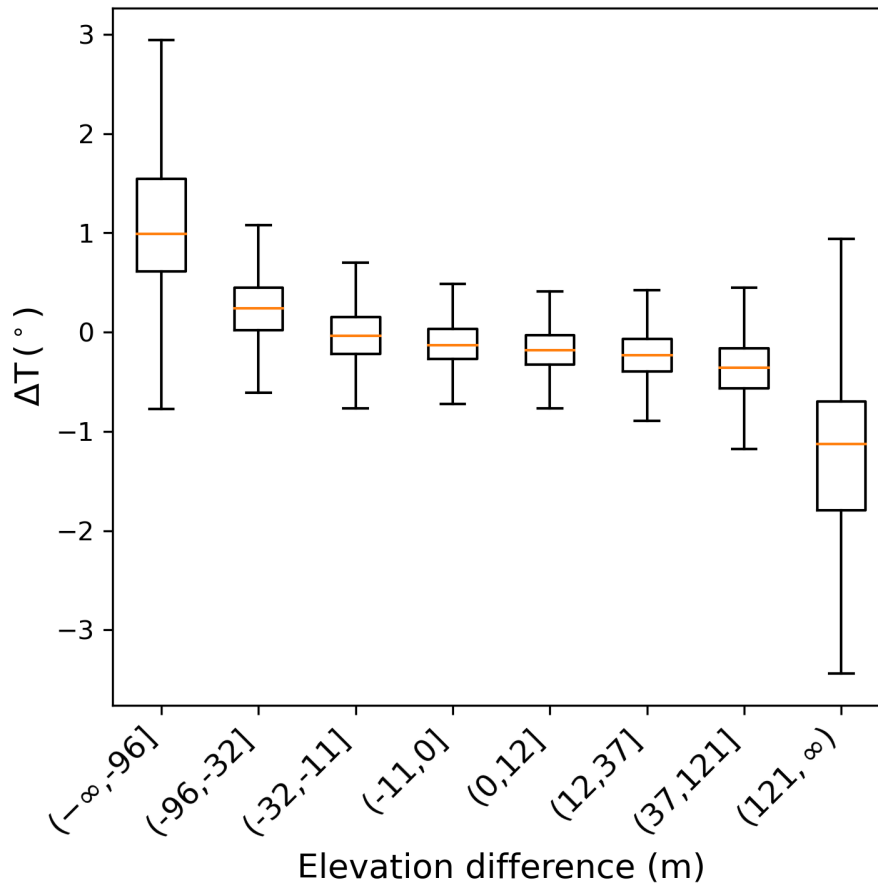


Figure 6. Distribution of the difference in mean air temperature between the downscaled and the original datasets. Positive values indicate a temperature increase after downscaling. The data was binned in octiles of elevation difference, calculated with respect to the average elevation of the low-resolution gridcell that contains the data point. The x -axis labels indicate limits of the bins. Water points are excluded. The distributions were derived from temperature averages over the 2011-2014 period.

375 of new saplings, obeys an exponential law that depends on the forest canopy's LAI (Monsi and Saeki, 1953, 2005). The decay rate of C in the different soil carbon pools is a non-linear function of soil temperature (driven by air temperature in the model) and soil water content (which depends non-linearly on precipitation rate, as mentioned above; see description of the carbon cycle submodel in Smith et al. (2014)).

380 The effect of fire on simulation results was found to be somewhat important, but not as strong as those of non-conservative properties of CHELSA and differences in climate due to orography. The effect includes 2 parts. First, since ignition is stochastic, the presence of the fire module may be able to increase the variation of the simulation results. Comparison of the standard deviations in Tables 6 and 7 shows that this effect does not play a significant role. Second, fire is a rare but destructive event

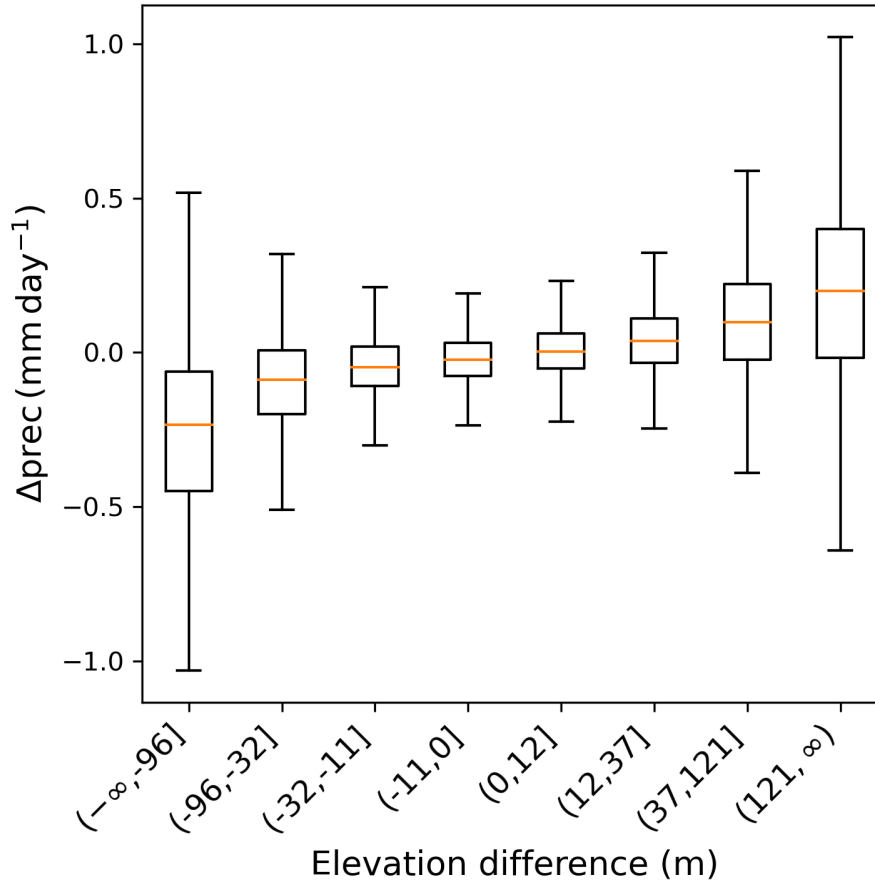


Figure 7. As in Fig. 6, but for average precipitation.

which introduces changes in the potential vegetation structure. This could be one of the reasons why we see more variables with statistically indistinguishable μ_{hr}^C and μ_{lr}^C in the uCH/NoFire experiment than in the uCH/Fire one. In the study region
 385 on the high resolution, ignition is expected to occur more in valleys, which are warmer and drier than mountain tops, thus the effect of reduced vegetation in mountainous areas should be decreased in the uCH/NoFire experiment. However, in Fig. 3 we see that the influence of fire on vegetation in the study region is negligible compared to the influence of orography-induced climate difference.

Systematic biases in model outputs may arise as a consequence of differences in forcings other than resolution. For instance,
 390 high-resolution simulations might be sensitive to the algorithm used to downscale the forcings. In the context of climate change mitigation, correlations between different climate variables might influence relevant modeled variables (Zscheischler et al., 2019). To give an example of mechanisms responsible for these correlations, we notice that at points where light is obstructed, the temperature is lower than at neighboring points with no obstruction. Analogously, a spot with a significant

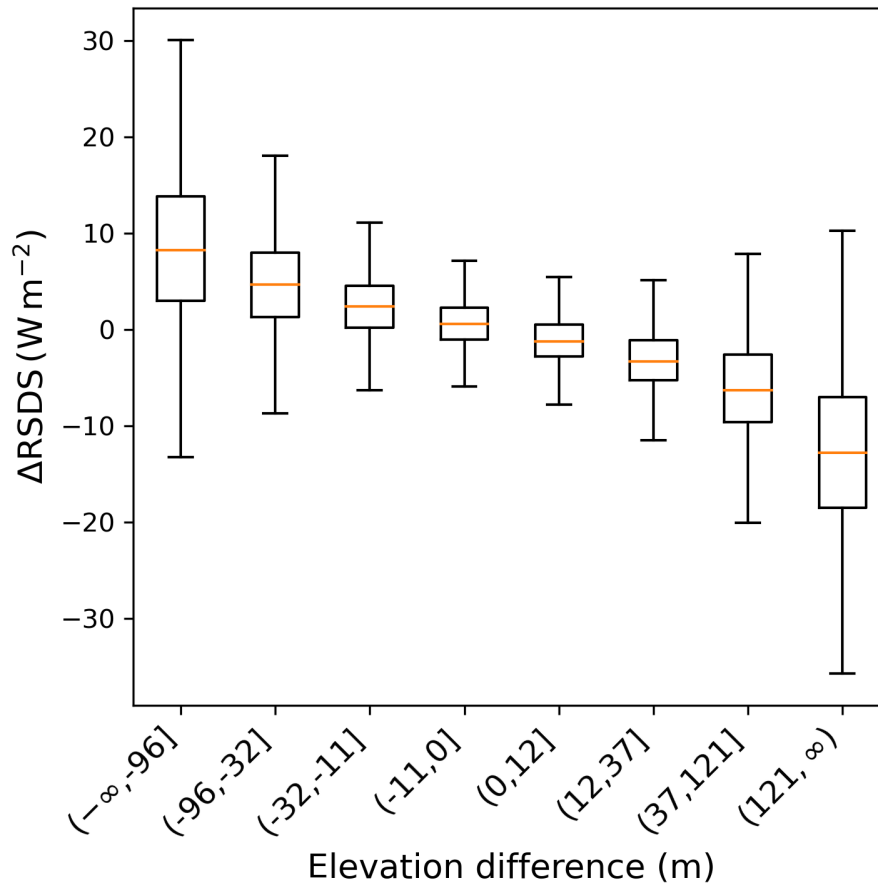


Figure 8. As in Fig. 6, but for average RSDS.

amount of precipitation would be colder and darker than the same spot without precipitation. Such correlations are not built
 395 into univariate methods like CHELSA but can be captured by dynamical or multivariate downscaling methods. These methods
 are, however, generally more complex, and might require intensive use of computational resources. Therefore, it might be of
 interest to find systematic differences between simulations forced by the different methods. This could be done with the help
 of the methodology presented in Sect. 2.2 and 4. A similar setup could also be employed to investigate systematic differences
 originating from alternative model configurations. For example, one could assess whether the modeled impacts of two different
 400 forest managing strategies on regional carbon sinks are significantly different from each other.

7 Summary

In this paper we presented a high-resolution climate dataset for ecosystem modeling applications in Europe. We applied the CHELSA semi-mechanistic algorithm to scale four ISIMIP3b scenarios (historical, SSP1-2.6, SSP3-7.0, and SSP5-8.5) from an original resolution of 0.5° down to 0.05° . Further processing involved quality checks, rechunking to optimize time-series retrieval at a single location, and the addition of CF-compliant metadata. The new dataset is provided in NetCDF format (one file per variable), and is publicly accessible under a CC BY-SA 4.0 license.

We studied systematic differences between high-resolution LPJ-GUESS simulations, forced with the new dataset, and low-resolution simulations. We found that low-resolution simulations are systematically biased. Two main sources of bias were identified: (a) bias associated to the non-linear response of the model to orographical climate variability, and (b) bias associated to the poor representation of coastlines and inland water bodies on a coarse grid. While the latter may be mitigated by rescaling the output by the land cover fraction in the affected gridcells, reducing the climate-response bias requires a finer grid resolution. These sources of bias are independent of the downscaling algorithm, and apply to other DGVMs, insofar as their response to climate forcings is non-linear. Climate-response bias can be very large in mountainous areas; low-resolution simulations overestimated average predictions between $\sim 4\%$ and $\sim 45\%$ in an alpine region, as opposed to a mean bias of $\sim 1.4\%$ in a nearly-flat control region. Biases as large as in the alpine region were shown to be vanishingly unlikely in the control region. On a European scale, climate-response bias led to an overestimation of regional averages of $\sim 3\%$. This suggests that this type of bias is very sensitive to overall changes in elevation, and should be accounted for when the focus region presents high orographical variability.

Code availability. The code base used in this work along with intermediate and final results are available in 10.5281/zenodo.14941305 (Otryakhin and Belda, 2025)

Data availability. The high-resolution climate data described in Sect. 3 is available in IMK-IFU storage https://thredds.imk-ifu.kit.edu/thredds/catalog/catalogues/luc_and_climate_catalog_ext.html (Otryakhin and Belda, 2024).

Appendix A: Bias decomposition

Let X be a modeled variable, S_X the aggregated value of X over the simulated domain, and μ_X the domain-average. In order to calculate the climate-response and shoreline-representation components of the bias, we consider the following quantities, defined in the high resolution grid:

1. X_{ij}^{HR} : Value of the high-resolution output at grid point (i, j) .

2. X_{ij}^{LR} : Value of the low-resolution output at grid point (i, j) . We note that this value will be the same for all (i, j) within the same low-resolution gridcell (see Fig. 5).
3. A_{ij} : Surface area of the gridcell at gridpoint (i, j) .
4. $M_{ij}^{\text{LR,HR}}$: *Overlap mask*. It takes the value 1 at land points where low-resolution values and high-resolution values overlap (gray cells in Fig. 5), and 0 everywhere else.
5. $M_{ij}^{\overline{\text{LR,HR}}}$: Only *high-resolution mask*. It takes the value 1 at land points present in the high-resolution simulation, but not present in the low resolution one (blue cells in Fig. 5) and 0 everywhere else.
6. $M_{ij}^{\text{LR},\overline{\text{HR}}}$: Only *low-resolution mask*. It takes the value 1 at land points present in the low-resolution simulation, but not present in the high resolution one (red cells in Fig. 5) and 0 everywhere else.

A1 Regionally aggregated quantities

For regionally aggregated variables, such as the carbon fluxes and pools, the bias between high- and low- resolution outputs is:

$$\begin{aligned}
 \delta_{\text{tot}} &= S_X^{\text{LR}} - S_X^{\text{HR}} \\
 &= \sum_{i,j} X_{ij}^{\text{LR}} A_{ij} (M_{ij}^{\text{LR,HR}} + M_{ij}^{\overline{\text{LR,HR}}}) \\
 &\quad - \sum_{i,j} X_{ij}^{\text{HR}} A_{ij} (M_{ij}^{\text{LR,HR}} + M_{ij}^{\overline{\text{LR,HR}}}),
 \end{aligned} \tag{A1}$$

where the indices (i, j) cover the whole domain. In this equation, the first sum represents the regional sum of the low resolution values, and the second term is the regional sum of the high-resolution values. Rearranging terms yields:

$$\begin{aligned}
 \delta_{\text{tot}} &= \underbrace{\sum_{i,j} (X_{ij}^{\text{LR}} - X_{ij}^{\text{HR}}) A_{ij} M_{ij}^{\text{LR,HR}}}_{\delta_{\text{cli}}} \\
 &\quad + \underbrace{\sum_{i,j} A_{ij} (X_{ij}^{\text{LR}} M_{ij}^{\overline{\text{LR,HR}}} - X_{ij}^{\text{HR}} M_{ij}^{\overline{\text{LR,HR}}})}_{\delta_{\text{sho}}}.
 \end{aligned} \tag{A2}$$

The first term of the above equation, labeled as δ_{cli} , involves values of X at overlapping gridcells exclusively (shown as gray cells in Fig. 5). Hence this term can be attributed to the difference in climate forcings between the two simulations. The second term, labeled δ_{sho} , involves values of X at non-overlapping gridcells between the high- and low- resolution simulations. These gridcells are the red and blue gridcells from Fig. 5, and are associated with poor shoreline representation at low resolution.

A2 Regionally averaged quantities

The variables FPC and LAI in Sec. 5 are averaged across the domain, rather than aggregated. The bias in this case is calculated as:

$$\delta_{\text{tot}} = \mu_X^{\text{LR}} - \mu_X^{\text{HR}} \quad (\text{A3})$$

$$= \frac{\sum_{i,j} X_{ij}^{\text{LR}} A_{ij} (M_{ij}^{\text{LR,HR}} + M_{ij}^{\text{LR,\overline{HR}}})}{\sum_{i,j} A_{ij} (M_{ij}^{\text{LR,HR}} + M_{ij}^{\text{LR,\overline{HR}}})} - \frac{\sum_{i,j} X_{ij}^{\text{HR}} A_{ij} (M_{ij}^{\text{LR,HR}} + M_{ij}^{\text{LR,\overline{HR}}})}{\sum_{i,j} A_{ij} (M_{ij}^{\text{LR,HR}} + M_{ij}^{\text{LR,\overline{HR}}})}, \quad 455$$

where the first term is the low-resolution regional average, and the second term is the high-resolution regional average. Rearranging terms yields

$$\delta_{\text{tot}} = \delta_{\text{cli}} + \delta_{\text{sho}}, \quad (\text{A4})$$

where

$$\delta_{\text{cli}} = \frac{\sum_{i,j} X_{ij}^{\text{LR}} A_{ij} M_{ij}^{\text{LR,HR}}}{\sum_{i,j} A_{ij} (M_{ij}^{\text{LR,HR}} + M_{ij}^{\text{LR,\overline{HR}}})} - \frac{\sum_{i,j} X_{ij}^{\text{HR}} A_{ij} M_{ij}^{\text{LR,HR}}}{\sum_{i,j} A_{ij} (M_{ij}^{\text{LR,HR}} + M_{ij}^{\text{LR,\overline{HR}}})}, \quad 460 \quad (\text{A5})$$

and

$$\delta_{\text{sho}} = \frac{\sum_{i,j} X_{ij}^{\text{LR}} A_{ij} M_{ij}^{\text{LR,\overline{HR}}}}{\sum_{i,j} A_{ij} (M_{ij}^{\text{LR,HR}} + M_{ij}^{\text{LR,\overline{HR}}})} - \frac{\sum_{i,j} X_{ij}^{\text{HR}} A_{ij} M_{ij}^{\text{LR,\overline{HR}}}}{\sum_{i,j} A_{ij} (M_{ij}^{\text{LR,HR}} + M_{ij}^{\text{LR,\overline{HR}}})}. \quad (\text{A6})$$

465 *Author contributions.* DO: conceptualization, data processing and quality control in Sect. 3 and 4, formal analysis, software in Sect. 3 and 4. DMB: conceptualization, data processing in Sect. 3, 4 and 5, formal analysis, software in Sect. 3, 4 and 5. AA: original idea of using CHELSA with LPJG, conceptualization. All authors helped shape the final form of the manuscript.

Competing interests. The Authors declare no competing interest.

References

- Angelo Canty and B. D. Ripley: boot: Bootstrap R (S-Plus) Functions, R package version 1.3-30, 2024.
- Daly, C., Neilson, R. P., and Phillips, D. L.: A Statistical-Topographic Model for Mapping Climatological Precipitation over Mountainous Terrain, *Journal of Applied Meteorology and Climatology*, 33, 140 – 158, [https://doi.org/10.1175/1520-0450\(1994\)033<0140:ASTMFM>2.0.CO;2](https://doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2), 1994.
- Daly, C., Taylor, G., and Gibson, W.: The PRISM approach to mapping precipitation and temperature, in: *Proc., 10th AMS Conf. on Applied Climatology*, pp. 20–23, 1997.
- Danielson, J. J. and Gesch, D. B.: Global multi-resolution terrain elevation data 2010 (GMTED2010), <https://doi.org/10.3133/ofr20111073>, 2011.
- Davison, A. C. and Hinkley, D. V.: *Bootstrap Methods and Their Applications*, Cambridge University Press, Cambridge, ISBN 0-521-57391-2, <https://doi.org/10.1017/CBO9780511802843>, 1997.
- Dikta, G. and Scheer, M.: *Bootstrap methods: with applications in R*, Springer Nature, <https://doi.org/10.1007/978-3-030-73480-0>, 2021.
- Gerten, D., Schaphoff, S., Haberlandt, U., Lucht, W., and Sitch, S.: Terrestrial vegetation and water balance—hydrological evaluation of a dynamic global vegetation model, *Journal of Hydrology*, 286, 249–270, <https://doi.org/10.1016/j.jhydrol.2003.09.029>, 2004.
- Giorgi, F., Jones, C., Asrar, G. R., et al.: Addressing climate information needs at the regional level: the CORDEX framework, *World Meteorological Organization (WMO) Bulletin*, 58, 175, 2009.
- Gregor, K., Knoke, T., Krause, A., Reyser, C. P. O., Lindeskog, M., Papastefanou, P., Smith, B., Lansø, A.-S., and Rammig, A.: Trade-Offs for Climate-Smart Forestry in Europe Under Uncertain Future Climate, *Earth's Future*, 10, e2022EF002796, <https://doi.org/10.1029/2022EF002796>, 2022.
- Gregor, K., Krause, A., Reyser, C. P. O., Knoke, T., Meyer, B. F., Suvanto, S., and Rammig, A.: Quantifying the impact of key factors on the carbon mitigation potential of managed temperate forests, *Carbon Balance and Management*, 19, 10, <https://doi.org/10.1186/s13021-023-00247-9>, 2024.
- Hassell, D., Gregory, J., Blower, J., Lawrence, B. N., and Taylor, K. E.: A data model of the Climate and Forecast metadata conventions (CF-1.6) with a software implementation (cf-python v2.1), *Geoscientific Model Development*, 10, 4619–4646, <https://doi.org/10.5194/gmd-10-4619-2017>, 2017.
- Haxeltine, A. and Prentice, I. C.: A General Model for the Light-Use Efficiency of Primary Production, *Functional Ecology*, 10, 551–561, <https://doi.org/10.2307/2390165>, 1996.
- Hickler, T., Vohland, K., Feehan, J., Miller, P. A., Smith, B., Costa, L., Giesecke, T., Fronzek, S., Carter, T. R., Cramer, W., Kühn, I., and Sykes, M. T.: Projecting the future distribution of European potential natural vegetation zones with a generalized, tree species-based dynamic vegetation model, *Global Ecology and Biogeography*, 21, 50–63, <https://doi.org/10.1111/j.1466-8238.2010.00613.x>, 2012.
- Jungclaus, J., Bittner, M., Wieners, K.-H., Wachsmann, F., Schupfner, M., Legutke, S., Giorgetta, M., Reick, C., Gayler, V., Haak, H., de Vrese, P., Raddatz, T., Esch, M., Mauritsen, T., von Storch, J.-S., Behrens, J., Brovkin, V., Claussen, M., Crueger, T., Fast, I., Fiedler, S., Hagemann, S., Hohenegger, C., Jahns, T., Kloster, S., Kinne, S., Lasslop, G., Kornbluh, L., Marotzke, J., Matei, D., Meraner, K., Mikolajewicz, U., Modali, K., Müller, W., Nabel, J., Notz, D., Peters-von Gehlen, K., Pincus, R., Pohlmann, H., Pongratz, J., Rast, S., Schmidt, H., Schnur, R., Schulzweida, U., Six, K., Stevens, B., Voigt, A., and Roeckner, E.: MPI-M MPI-ESM1.2-HR model output prepared for CMIP6 CMIP historical, <https://doi.org/10.22033/ESGF/CMIP6.6594>, 2019.

- 505 Karger, D. N.: CHELSA_ISIMIP3b_BA_1km, https://gitlabext.wsl.ch/karger/chelsa_isimip3b_ba_1km, [Online; accessed 06-March-2024], 2022.
- Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., and Kessler, M.: Climatologies at high resolution for the earth's land surface areas, *Scientific Data*, 4, 170 122, <https://doi.org/10.1038/sdata.2017.122>, 2017.
- 510 Karger, D. N., Wilson, A. M., Mahony, C., Zimmermann, N. E., and Jetz, W.: Global daily 1 km land surface precipitation based on cloud cover-informed downscaling, *Scientific Data*, 8, 307, 2021.
- Karger, D. N., Lange, S., Hari, C., Reyer, C. P. O., Conrad, O., Zimmermann, N. E., and Frieler, K.: CHELSA-W5E5: daily 1 km meteorological forcing data for climate impact studies, *Earth System Science Data*, 15, 2445–2464, <https://doi.org/10.5194/essd-15-2445-2023>, 2023.
- 515 Knorr, W., Kaminski, T., Arneth, A., and Weber, U.: Impact of human population density on fire frequency at the global scale, *Biogeosciences*, 11, 1085–1102, <https://doi.org/10.5194/bg-11-1085-2014>, 2014.
- Knorr, W., Jiang, L., and Arneth, A.: Climate, CO₂ and human population impacts on global wildfire emissions, *Biogeosciences*, 13, 267–282, <https://doi.org/10.5194/bg-13-267-2016>, 2016.
- Lagergren, F., Björk, R. G., Andersson, C., Belušić, D., Björkman, M. P., Kjellström, E., Lind, P., Lindstedt, D., Olenius, T., Pleijel, H.,
- 520 Rosqvist, G., and Miller, P. A.: Kilometre-scale simulations over Fennoscandia reveal a large loss of tundra due to climate warming, *Biogeosciences*, 21, 1093–1116, <https://doi.org/10.5194/bg-21-1093-2024>, 2024.
- Lange, S. and Büchner, M.: ISIMIP3b bias-adjusted atmospheric climate input data, <https://doi.org/10.48364/ISIMIP.842396.1>, 2021.
- Lindeskog, M., Arneth, A., Bondeau, A., Waha, K., Seaquist, J., Olin, S., and Smith, B.: Implications of accounting for land use in simulations of ecosystem carbon cycling in Africa, *Earth System Dynamics*, 4, 385–407, <https://doi.org/10.5194/esd-4-385-2013>, 2013.
- 525 Lindeskog, M., Smith, B., Lagergren, F., Sycheva, E., Ficko, A., Pretzsch, H., and Rammig, A.: Accounting for forest management in the estimation of forest carbon balance using the dynamic vegetation model LPJ-GUESS (v4.0, r9710): implementation and evaluation of simulations for Europe, *Geoscientific Model Development*, 14, 6071–6112, <https://doi.org/10.5194/gmd-14-6071-2021>, publisher: Copernicus GmbH, 2021.
- Masson-Delmotte, V., Zhai, P., Pirani, S., Connors, C., Péan, S., Berger, N., Caud, Y., Chen, L., Goldfarb, M., and Scheel Monteiro, P. M.:
- 530 Ipcc, 2021: Summary for policymakers. in: *Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change*, 2021.
- Meyer, B. F., Buras, A., Gregor, K., Layritz, L. S., Principe, A., Kreyling, J., Rammig, A., and Zang, C. S.: Frost matters: incorporating late-spring frost into a dynamic vegetation model regulates regional productivity dynamics in European beech forests, *Biogeosciences*, 21, 1355–1370, <https://doi.org/10.5194/bg-21-1355-2024>, publisher: Copernicus GmbH, 2024.
- 535 Molinari, C., Hantson, S., and Nieradzik, L. P.: Fire Dynamics in Boreal Forests Over the 20th Century: A Data-Model Comparison, *Frontiers in Ecology and Evolution*, 9, <https://doi.org/10.3389/fevo.2021.728958>, publisher: Frontiers, 2021.
- Monsi, M. and Saeki, T.: On the Factor Light in Plant Communities and its Importance for Matter Production, *Japanese Journal of Botany*, 14, 22–52, 1953.
- Monsi, M. and Saeki, T.: On the Factor Light in Plant Communities and its Importance for Matter Production, *Annals of Botany*, 95, 549–567,
- 540 <https://doi.org/10.1093/aob/mci052>, 2005.

Morales, P., Hickler, T., Rowell, D. P., Smith, B., and Sykes, M. T.: Changes in European ecosystem productivity and carbon balance driven by regional climate model output, *Global Change Biology*, 13, 108–122, <https://doi.org/10.1111/j.1365-2486.2006.01289.x>, <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2486.2006.01289.x>, 2007.

Morera, A., LeBlanc, H., Martínez de Aragón, J., Bonet, J. A., and de Miguel, S.: Analysis of climate change impacts on the biogeographical patterns of species-specific productivity of socioeconomically important edible fungi in Mediterranean forest ecosystems, *Ecological Informatics*, 81, 102557, <https://doi.org/10.1016/j.ecoinf.2024.102557>, 2024.

Müller, C. and Lucht, W.: Robustness of terrestrial carbon and water cycle simulations against variations in spatial resolution, *Journal of Geophysical Research: Atmospheres*, 112, <https://doi.org/10.1029/2006JD007875>, publisher: John Wiley & Sons, Ltd, 2007.

Olin, S., Schurgers, G., Lindeskog, M., Wårlind, D., Smith, B., Bodin, P., Holmér, J., and Arneth, A.: Modelling the response of yields and tissue C : N to changes in atmospheric CO₂ and N management in the main wheat regions of western Europe, *Biogeosciences*, 12, 2489–2515, <https://doi.org/10.5194/bg-12-2489-2015>, 2015.

Otryakhin, D. and Belda, D. M.: ISIMP3b-CHELSA climate input data for LPJ-GUESS, https://thredds.imk-ifu.kit.edu/thredds/catalog/catalogues/luc_and_climate_catalog_ext.html, [Online; accessed 28-February-2025], 2024.

Otryakhin, D. and Belda, D. M.: Software for comparison of LPJ-GUESS simulations driven by low- and high-resolution climate data, <https://doi.org/10.5281/zenodo.14941305>, 2025.

Rabin, S. S., Melton, J. R., Lasslop, G., Bachelet, D., Forrest, M., Hantson, S., Kaplan, J. O., Li, F., Mangeon, S., Ward, D. S., Yue, C., Arora, V. K., Hickler, T., Kloster, S., Knorr, W., Nieradzik, L., Spessa, A., Folberth, G. A., Sheehan, T., Voulgarakis, A., Kelley, D. I., Prentice, I. C., Sitch, S., Harrison, S., and Arneth, A.: The Fire Modeling Intercomparison Project (FireMIP), phase 1: experimental and analytical protocols with detailed model descriptions, *Geoscientific Model Development*, 10, 1175–1197, <https://doi.org/10.5194/gmd-10-1175-2017>, publisher: Copernicus GmbH, 2017.

Schaphoff, S., Lucht, W., Gerten, D., Sitch, S., Cramer, W., and Prentice, I. C.: Terrestrial biosphere carbon storage under alternative climate projections, *Climatic Change*, 74, 97–122, <https://doi.org/10.1007/s10584-005-9002-5>, 2006.

Schär, C., Fuhrer, O., Arteaga, A., Ban, N., Charpillot, C., Di Girolamo, S., Hentgen, L., Hoefler, T., Lapillonne, X., Leutwyler, D., et al.: Kilometer-scale climate models: Prospects and challenges, *Bulletin of the American Meteorological Society*, 101, E567–E587, 2020.

Schulzweida, U.: CDO User Guide, <https://doi.org/10.5281/zenodo.10020800>, 2023.

Schurgers, G., Ahlström, A., Arneth, A., Pugh, T. A. M., and Smith, B.: Climate Sensitivity Controls Uncertainty in Future Terrestrial Carbon Sink, *Geophysical Research Letters*, 45, 4329–4336, <https://doi.org/10.1029/2018GL077528>, <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2018GL077528>, 2018.

Shin, Y.-J., Midgley, G. F., Archer, E. R. M., Arneth, A., Barnes, D. K. A., Chan, L., Hashimoto, S., Hoegh-Guldberg, O., Insarov, G., Leadley, P., Levin, L. A., Ngo, H. T., Pandit, R., Pires, A. P. F., Pörtner, H.-O., Rogers, A. D., Scholes, R. J., Settele, J., and Smith, P.: Actions to halt biodiversity loss generally benefit the climate, *Global Change Biology*, 28, 2846–2874, <https://doi.org/10.1111/gcb.16109>, 2022.

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 Venevsky, S.: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model, *Global Change Biology*, 9, 161–185, <https://doi.org/10.1046/j.1365-2486.2003.00569.x>, 2003.

Smith, B., Prentice, I. C., and Sykes, M. T.: Representation of vegetation dynamics in the modelling of terrestrial ecosystems: comparing two contrasting approaches within European climate space, *Global Ecology and Biogeography*, 10, 621–637, <https://doi.org/10.1046/j.1466-822X.2001.t01-1-00256.x>, 2001.

- Smith, B., Wårlind, D., Arneth, A., Hickler, T., Leadley, P., Siltberg, J., and Zaehle, S.: Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model, *Biogeosciences*, 11, 2027–2054, <https://doi.org/https://doi.org/10.5194/bg-11-2027-2014>, 2014.
- Smith, P., Arneth, A., Barnes, D. K. A., Ichii, K., Marquet, P. A., Popp, A., Pörtner, H.-O., Rogers, A. D., Scholes, R. J., Strassburg, B., Wu, J., and Ngo, H.: How do we best synergize climate mitigation actions to co-benefit biodiversity?, *Global Change Biology*, 28, 2555–2577, <https://doi.org/10.1111/gcb.16056>, 2022.
- Sørland, S. L., Brogli, R., Pothapakula, P. K., Russo, E., Van de Walle, J., Ahrens, B., Anders, I., Bucchignani, E., Davin, E. L., Demory, M.-E., et al.: COSMO-CLM regional climate simulations in the Coordinated Regional Climate Downscaling Experiment (CORDEX) framework: a review, *Geoscientific Model Development*, 14, 5125–5154, 2021.
- Tian, H., Yang, J., Lu, C., Xu, R., Canadell, J. G., Jackson, R. B., Arneth, A., Chang, J., Chen, G., Ciais, P., Gerber, S., Ito, A., Huang, Y., Joos, F., Lienert, S., Messina, P., Olin, S., Pan, S., Peng, C., Saikawa, E., Thompson, R. L., Vuichard, N., Winiwarter, W., Zaehle, S., Zhang, B., Zhang, K., and Zhu, Q.: The Global N₂O Model Intercomparison Project, *Bulletin of the American Meteorological Society*, 99, 1231–1251, <https://doi.org/10.1175/BAMS-D-17-0212.1>, publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society, 2018.
- Welch, B. L.: The Generalization of ‘Student’s’ Problem When Several Different Population Variances Are Involved, *Biometrika*, 34, 28–35, <https://doi.org/10.1093/biomet/34.1-2.28>, 1947.
- Zender, C. S.: Analysis of self-describing gridded geoscience data with netCDF Operators (NCO), *Environmental Modelling & Software*, 23, 1338–1342, <https://doi.org/10.1016/j.envsoft.2008.03.004>, 2008.
- Zobler, L.: A world soil file global climate modeling, NASA Technical Memorandum, 32, 1986.
- Zscheischler, J., Fischer, E. M., and Lange, S.: The effect of univariate bias adjustment on multivariate hazard estimates, *Earth System Dynamics*, 10, 31–43, <https://doi.org/10.5194/esd-10-31-2019>, 2019.