Upscaling of spatially explicit and linked time- and space-discrete models simulating vegetation dynamics under climate change

Julia E.M.S. Nabel\(^1,2\) and Heike Lischke\(^1\)

Abstract

Models applied to simulate the impact of climate change on vegetation dynamics generally face the trade-off between computational expenses (computation time and memory) and modelled detail. Models used for simulations of large areas (e.g. continental) often abstract processes entailing spatial linkages, e.g. species migration, and have too coarse resolutions to depict microsite heterogeneity. Regional to local models, on the other hand, are more detailed, but their computational expenses prevent applications on larger scales. For manageable and accurate simulations of vegetation dynamics on large scales, small-scale dynamics need to be integrated with large-scale applications in a balanced way. Several methods have been proposed and applied to expedite the integration of scales. However, each method has different advantages and drawbacks and the applicability of a method also strongly depends on the initial model and the research question.

Here we present a conceptual framework for a further step integrating the scales in simulations with spatially explicit, time- and space-discrete models simulating vegetation dynamics under climate change. In such models, grid cells with similar environmental drivers and species compositions often entail repetitive calculations. Our method strives to reduce this redundancy and aims to disentangle repetitive calculations from processes specific to single cells. The proposed method is based on a dynamic two-layer classification (D2C) concept, in which the majority of processes is simulated in representative cells constituting the coarse layer, and only processes which might lead to changes specific to a single cell are simulated on the original grid, i.e. the fine layer. This new concept is a further step to enable the simulation of more detailed small-scale dynamics on a larger scale. We provide an example applying the D2C concept with the forest-landscape model TreeMig and shortly discuss its advantages and limitations.

1. Introduction

Spatiotemporal vegetation dynamics play a central role in earth system processes, and large-scale changes in vegetation structure and distribution can influence the entire system (Fischlin et al 2007). Changes in the vegetation structure, in turn, are driven by processes on various scales, ranging from photosynthesis on very small scales and competition for light and disturbances on intermediate scales up to large-scale disturbances and species' migration (Neilson et al 2005; Fischlin et al 2007).

Models applied to study vegetation dynamics suffer from limitations when trying to simulate interacting small and large-scale forest dynamics on a large extent. Generally, processes and influences represented in a model result from trade-offs between accuracy on one side, and computational feasibility and efficiency as well as parameterisation requirements on the other side (Huntley et al 2010). Available dynamic vegetation models can broadly be classified into large and small-scale models. Large-scale models – essentially dynamic global vegetation models (DGVMs) – avoid spatially linking processes and use coarse spatial (e.g. 50-200km\(^2\)) or taxonomic resolutions (plant functional types instead of single species). Furthermore, they disregard or strongly simplify processes requiring small-scale information, such as local competition for light (Fisher et al 2010). Small-scale models, on the other hand, often incorporate important small-scale processes and drivers, but are computationally too expensive (time and required memory) to be used

\(^1\) Dynamic Macroecology, Landscape Dynamics, Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland, email: jemsnabel@gmail.com.

\(^2\) Institute of Terrestrial Ecosystems, Swiss Federal Institute of Technology ETH, Universitätsstr. 16, 8092 Zürich, Switzerland.
for large area simulations. Particularly, the naïve approach to enable simulations of larger areas by coarsening the grid cell resolution has been shown to introduce large discretisation errors (e.g. Bocedi et al. 2012). Furthermore, high computational expenses often not only result from the fine resolution, but from the simulation of processes which are usually neglected in large-scale models, for example, spatial linkages between grid cells (Nabel et al. 2013). Typical cell side lengths of small-scale models are 25m to 1km and simulated extents seldom exceed the size of a small country or federal state.

Several methods have been proposed and applied to reach a stronger integration of the scales, either in a top-down manner refining large-scale models (Fisher et al. 2010) or as bottom-up approaches to upscale small-scale models (Urban et al. 1999; Lischke et al. 2007; Auger et al. 2012). The majority of upscaling methods assume that fine-scale components of a modelled system can be represented on a coarse scale through skilful selections and aggregations (Lischke et al. 2007). Such aggregations can be temporal, spatial or functional, i.e. regarding processes or state variables. There is a manifold of methods, ranging from analytical aggregations (Auger et al. 2012) to adaptive grid methods, in which discretisations in time and space are dynamically refined or coarsened according to local gradients (Zumbusch 2003). Other methods use parts of the original model to obtain information for coarser scales. This is for example the case with so-called equation free approaches, which calculate and evaluate selected small-scale experiments to attain the state on the coarse scale at a certain point in time (Kevrekidis/Samaey 2009). A common way to completely change the computation scale is the so-called meta-modelling, involving the development of a new coarse-scale model, parameterised with results of representative small-scale model simulations (Urban et al. 1999).

Most of the listed methods are constrained by certain assumptions, some are not suitable for more complex models and others replace the complex fine-scale model by a simpler coarse-scale model, which then is only valid under specific conditions. In the end, the applicability of a method strongly depends on the initial model and on the investigated research question. Models simulating vegetation dynamics are often implemented as discrete grid-based systems with or without spatial linkage of the single cells. For this kind of models we developed a method which aims to provide a further step in the integration of detailed small-scale dynamics with larger-scale applications. In the following we describe the concept of the method and present a first implementation and test.

2. The dynamic two-layer classification concept

The proposed dynamic two-layer classification (D2C) concept relies on the fact that in many grid-based models simulating sessile organisms, grid cells with similar abiotic drivers are covered by comparable species compositions. Provided the abiotic drivers influencing the cells follow the same temporal pathways, cells with similar species compositions will continue to follow the same successional paths until cell-specific influences, such as immigration or disturbances, cause deviations of the species compositions among single cells. Cells with comparable abiotic drivers can hence lead to repetitive calculations. The D2C concept aims to reduce this redundancy by disentangling repetitive calculations from processes specific to single cells in a dynamic and adaptive way.

When upscaling a model with the D2C concept, processes which might lead to changes specific to a single cell, such as dispersal, recruitment and disturbances, are simulated on the original fine grid constituting the first layer. All other processes (e.g. light competition, growth and seed production) are simulated on the second layer solely consisting of representatives to which the cells of the first layer are assigned. Each element of the second layer represents all first-layer cells with similar environmental influences and similar species compositions, i.e. one type of first-layer cells. In order to classify the first-layer cells into types, thresholds have to be specified determining similarity of environmental drivers and of the model state variables describing species quantities and properties (e.g. size or age) in a cell. The number of these representatives, i.e. the number of types, can be dynamic, since processes simulated on the first layer can
cause splits, and therefore new types. A split would, for example, be necessary when formerly absent species establish in only some of the cells assigned to one type. On the other hand, representatives which are similar enough according to the specified thresholds can be merged. This can, for example, happen when two representatives both reach a similar state, after being simulated separately because a species immigrated at different points in time (e.g. Figure 1).

![Figure 1](image)

**Figure 1**

Simple example for the dynamic two-layer classification concept. (Panel a) Cells with similar species compositions are depicted in the same shade of green. (Panel b) Similarity thresholds the species compositions and on model drivers determine the classification of first-layer cells (grey underlying grid) in different types (numbers) each constituting a second layer representative. The assignment of first-layer cells to their representatives (frames formed by the black lines) can change over time: processes simulated on the first layer can cause splits, for example seed dispersal, leading to new representatives (e.g. 7 – panel c). Second-layer processes change properties of the representatives (panel c – asterisks). Dynamics simulated on both layers can lead to changed associations due to merges and splits (e.g. changed frames panel b, c).

### 2.1 How to apply the D2C concept

In order to apply the D2C concept for a certain model, the following steps have to be conducted:

1. Modelled processes have to be assigned to the two different layers, i.e. the layer representing the fine scale, where processes are calculated for each cell, and the coarse-scale layer, where only representatives are evaluated. Processes which might lead to changes specific to a single cell, such as dispersal, recruitment and cell-specific disturbances need to be simulated on the first layer. All other processes (e.g. light competition, growth and seed production) can be simulated on the second layer.

2. Similarity thresholds have to be specified for the environmental drivers and for model state variables describing species quantities and properties in a cell. These thresholds determine the number of resulting representatives and therefore control large parts of the trade-off between computational expenses (computation time and memory) and accuracy. There are multiple ways to specify the similarity thresholds. The simplest way would be to predefined a maximum number and to equidistantly stratify the range of environmental influences and expected species abundances. Alternatively, heuristics could be developed to determine the thresholds and to control similarity comparisons. One example are heuristics linked to ingrowth: small changes in the ingrowth of established species might not lead to changes in the species composition, as opposed to changes for non- or underrepresented species. Thresholds in this example hence depend on the momentary situation of a considered site.

---

3 Here accuracy is defined as the similarity of simulation results with two layers compared to results when simulating all processes on the original fine layer.
3. Whilst splits are required as soon as any of the thresholds would be violated, merges on the second layer do not necessarily have to be conducted in each iteration. Testing for similarity among all different representatives in each iteration would involve high computational expenses. An efficient application of the D2C concept therefore should also implement heuristics to reduce the overhead involved with these membership decisions. Merging could, for example, only be considered after certain time intervals or for representatives with a specific age. Another applicable heuristic would be to restrict comparisons among representative cells to groups with similar environmental conditions.

The assignment of the first-layer cells to representatives on the second layer at the beginning of a simulation depends on how the state variables of a model run are initialised. Many models simulating vegetation dynamics are initialised by a so-called spin-up from bare ground, i.e. without pre-assigned species' distributions. In this case, a simulation will start with a small number of representatives classified according to the similarity thresholds for the environmental drivers and will increase with simulation time. If the simulation area, on the other hand, is initialised based on species occurrence or abundance data, this data also needs to be classified, which is done according to the similarity thresholds for species compositions.

2.2 When to apply the D2C concept

We expect the D2C concept to be suitable for spatially explicit and linked, time- and space-discrete models in which repetitive calculations for cells with similar environmental drivers and species compositions constitute a big share of the computational expenses. In such cases, the separation into two different layers reduces the memory usage, since not all state variables have to be stored for the fine layer anymore and reduces the computation time for processes now executed on the coarse layer. Hence, the concept is expected to be suitable when the main share of the local processes are realised as deterministic processes (as done in the forest-landscape model TreeMig – see below and Lischke et al 2006) or if stochastic processes are only realised as patch replications within each cell (as in the second generation DGVM LPJ-Guess – Hickler et al 2012). The D2C concept will not be suitable when processes linking first-layer cells predominate the computational expenses or in case of a high variability in species compositions among first-layer cells with similar environmental influences. A high variability in species compositions can especially be expected when the species composition in a cell strongly depends on stochastic processes which are not realised as within cell patch replications (as in the forest-landscape model Landclim – Schumacher 2004).

3. Applying the D2C concept to the forest-landscape model TreeMig

The forest-landscape model TreeMig is a dynamic time- and space-discrete intermediate-complexity model simulating the dynamics and spatial interactions of multiple competing tree species (Lischke et al 2006). Depending on the simulation setup, the TreeMig Fortran implementation used in the following (TreeMig-Netcdf 2.0 – Nabel, submitted) has a computation time of approximately 0.2-0.7 millisecond for each cell and year on a 2.8GHz AMD Opteron CPU and a memory usage of approximately 5kB per cell (see Table 1 for some examples). The computational expenses hence constrain large-scale applications.

TreeMig simulations are driven by annual time series of three bioclimate variables per simulation cell (listed in Figure 2b). These bioclimate variables are derived from monthly average temperatures and precipitation sums in combination with static information on slope and aspect of the terrain and water storage capacity for each cell (Lischke et al 2006). Hence, the number of fluctuating environmental drivers is small and they are, moreover, correlated among each other and in space (Nabel, submitted). These properties imply that TreeMig potentially requires a smaller number of representatives than a model with more or uncorrelated fluctuating environmental drivers.
TreeMig itself is already the result of a previous upscaling. The state variables are real-valued densities of seeds per tree species in the seed bank and of population densities of tree species in a constant number of distinct height classes per grid cell. These state variables represent mean densities determining Poisson distributions of the population density on a given unit area (the ‘patch’ area) and can be regarded as a deterministic representation of the local spatial forest heterogeneity (Lischke et al 1998, 2006). This deterministic representation of the species composition on a constant number of height classes makes TreeMig particularly suitable for the application of the D2C concept because it implies less variability among different cells with similar environmental drivers. The discretisation to a constant number of height classes, furthermore, reduces the number of similarity thresholds required to test if representatives can be merged.

The first step to apply the D2C concept to a model requires the assignment of simulated processes to the two layers. In simulations with TreeMig the main share of the local processes are realised as deterministic processes which can be assigned to the coarse layer. Stochastic influences on the local processes can stem from purely random cell-specific disturbances (Lischke et al 2006) or from random but spatially autocorrelated bioclimatic influences (Nabel, submitted). Therefore, single-cell disturbances and bioclimatic influences have to be simulated on the fine layer. Additionally, TreeMig simulations can be conducted with or without spatial linkage via seed dispersal. In simulations without spatial linkage all seeds stay in the producing cell which is what is typically done in models applied on larger scales. In simulations with spatial linkage seeds produced in a cell are distributed to its neighbourhood following deterministic or stochastic species-specific dispersal kernels. Since seeds are distributed to the neighbours of a cell, dispersal has to be simulated on the fine layer, together with associated seedbank dynamics (see Lischke et al 2006).

3.1 Preclustering the bioclimatic influences to assign representatives

In a first attempt to apply the D2C concept to TreeMig and to assess possible gains of its implementation, we established static assignments to representatives following a clustering of the bioclimatic space. This approach was selected because TreeMig simulations are strongly influenced by its three bioclimatic drivers (listed in Figure 2b), whose interannual fluctuations are highly spatially correlated (Nabel, submitted). For these first tests we selected a nested set of three simulation areas: a small transect embedded in a larger transect, which itself is embedded in Switzerland (Figure 2a). All of these simulation areas are gridded with a cell side length of 200m. We clustered the bioclimatic influences with three sets of bioclimatic classes (E1, E2, E3 – Figure 2b) equidistantly distributed over the bioclimatic space derived for Switzerland from an SRES A1B scenario simulation with RCA3 (Kjellström et al 2005) downscaled to 200m cell side length based on an interpolated grid of MeteoSwiss weather stations.

To derive the representatives for each of the three simulation areas and to assign the single cells to these representatives we clustered the bioclimatic time series of the SRES A1B scenario (1961-2100) following four steps: 1. We defined the average of the first and the last 30 years, as well as the average over the whole time span as sampling points. 2. For all cells we calculated the sampling points for each bioclimatic variable and assigned them to the closest bioclimatic class of the applied sets (Figure 2b). 3. Cells with the same class for all three bioclimatic variables and all sampling points were assigned to the same representative. 4. Finally, we derived bioclimatic time series for the representatives by averaging the values of the associated cells for each year and variable.
Figure 2
(Panel a) Nested simulation areas used for simulations with a D2C TreeMig implementation. The map shows the values of the bioclimate variables $DD_{sum>5.5^\circ C}$ (sum of daily mean temperatures above 5.5°C) derived from the applied climate scenario for the year 2100. (Panel b) Applied sets of bioclimate classes.

Results from TreeMig simulations with two layers based on the assignment to the obtained representatives were compared to results from simulations on the single layer\(^4\). Comparisons of the overall resulting biomass and the biomass resulting for single species were conducted with a similarity coefficient (EQ. 1) previously used in various inter- and intramodel comparisons (e.g. Lischke et al 1998; Nabel et al 2013).

$$SC_y = 1 - \frac{\sum_{i=1}^{\text{ncells}} |p_{\text{um},i,y} - p_{\text{sum},i,y}|}{\sum_{i=1}^{\text{ncells}} S_{\text{sum},i,y}} \quad \text{(EQ. 1)}$$

The similarity coefficient (SC) for a year ($y$) of a run with two layers compared to the run on the single layer is calculated as the sum of the absolute values of the biomass differences ($D_{\text{um}}$) between corresponding cells ($i$) in this year and their sum ($S_{\text{um}}$). When comparing the biomass per species, differences for each species are calculated separately. The SC can range from zero to one, with one resulting for a perfect match and zero for no similarity.

3.1.1 Simulating without spatial linkage

We calculated the similarity coefficient (SC) to compare simulations with two layers and the single layer over time, for simulations without spatial linkage, i.e. without seed dispersal to neighbouring cells and with an additional steady supply of seeds of all species in all cells (Figure 3). The SCs resulting from the comparisons indicate that the simulation results are fairly similar. The nearly identical results in the spin-up diverge in the transient phase of climate warming after 2000 and subsequently stabilise on different levels depending on the resolution of the applied set of representatives (E1-E3). Since these simulations were conducted without spatial linkage and with no other stochastic influences than the bioclimate extrapolation, these comparisons show the error due to the clustering of the bioclimate and could further be reduced using a finer classification or additional sampling points.

Simulations with different sets of representatives involve large differences in computational expenses compared to the simulation on one layer (see Table 1).

\(^4\)All simulations were conducted for 1100 years starting in 1400 with a spin-up from bare ground. Bioclimate for years exceeding the scenario (i.e. 1400-1960 and 2101-2500) was extrapolated by sampling single years for the entire simulation area from a base set derived from the first (1961-1990) and the last 30 years (2071-2100), respectively (see Nabel, submitted).
Similarity coefficients (SCs) comparing results from simulations without spatial linkage conducted with one or two layers. The graphs show SCs for the biomass resulting for all and for the single species over time. Each colour represents SCs comparing results of the simulations with one layer to simulations with one of the applied sets of bioclimate classes (E1-E3 – Figure 2b); lines of the same colour are repetitions with different pseudo-random number streams (PRNS) used to extrapolate bioclimate influences (five for the two transects, two for Switzerland). Comparisons were conducted between runs with the same PRNS.

Table 1 Statistics for simulations with the three applied clustering schemes (E1-E3 – Figure 2b) compared to simulation on one layer. Simulations were conducted without (F) and with spatial linkage (T) via seed dispersal.

<table>
<thead>
<tr>
<th></th>
<th>Small transect</th>
<th>Large transect</th>
<th>Switzerland</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reps. a</td>
<td>#100</td>
<td>#1000</td>
<td>#1000</td>
</tr>
<tr>
<td>(%)</td>
<td>125 (9.6)</td>
<td>1250 (2.8)</td>
<td>1920 (2.9)</td>
</tr>
<tr>
<td>CPU time</td>
<td>min. 63.2</td>
<td>618.1</td>
<td>68.1</td>
</tr>
<tr>
<td>(%)</td>
<td>(10.0)</td>
<td>(3.8)</td>
<td>(47.6)</td>
</tr>
<tr>
<td>Heap mem. c</td>
<td>MB 83.7</td>
<td>679.0</td>
<td>10.2</td>
</tr>
<tr>
<td>(%)</td>
<td>(34.8)</td>
<td>(13.6)</td>
<td>(11.4)</td>
</tr>
</tbody>
</table>

Simulating with spatial linkage

Simulations with the same setup as described above but with spatial linkage via seed dispersal and without additionally supplied seeds were conducted. These simulations, in which ingrowth to a representative was simply averaged from the associated fine layer cells, led to a reduction in the similarities (Figure 4). The memory usage was about the same as for the simulations without spatial linkage, however, computation times increased strongly for both, simulations on one and simulations on two layers (Table 1). This increase in computation time was expected and is due to the higher common expenses due to the simulation of dispersal between fine layer cells. One consequence of this increase in “base load” is that the gain in computation time between simulations on one and on two layers is also decreased.
3.2 Discussion and Outlook

The example application of the D2C concept with TreeMig showed a reasonable gain in computational expenses which, for example, could be used to increase the simulated extent or to refine the resolution of the simulation area. At the same time, the application demonstrated limitations of the concept, in particular the loss of accuracy in simulations with spatial linkage caused by the averaging of the ingrowth from the fine layer. The introduced error will in particular cause problems in simulations where many species migrate in the simulation area, or where the migration of single species should be tracked. The next step to address such situations needs to be the implementation of dynamic partitioning and merging of representatives. However, the overhead introduced with these dynamic decisions will inevitably further reduce the gain in computational expenses, in addition to the reduction already involved with the spatial linkage. For simulations with spatial linkage, the D2C concept will hence only lead to reductions in computational expenses if efficient heuristics for the partitioning and the merging can be identified (see 2.1 for examples). Despite these pitfalls, the D2C concept has several advantages compared to other methods. It particularly allows retaining the fine resolution for important small-scale processes and for the simulation results, as opposed to other methods, such as equation free approaches or meta-modelling, which work with model results on a coarser scale. A further important advantage is that the representatives are not constrained to predefined spatial arrangements in contrast to, for example, adaptive meshing methods, which can only aggregate cells to compact regular shapes.

Acknowledgements

We like to thank Felix Kienast, James Kirchner, Natalie Zurbriggen and David Gutzmman for valuable discussions and comments and Dirk Schmatz for providing the downscaled data. Julia Nabel was partly funded by the Swiss National Science Foundation (SNF) Grant 315230-122434.

Bibliography


