Geosci. Model Dev. Discuss., 8, C1921–C1935, 2015
www.geosci-model-dev-discuss.net/8/C1921/2015/

Geoscientific Model Development Discussions

Interactive comment on "Upscaling with the dynamic two-layer classification concept (D2C): TreeMig-2L, an efficient implementation of the forest-landscape model TreeMig" *by* J. E. M. S. Nabel

Dr. Nabel

jemsnabel@gmail.com

Received and published: 8 September 2015

Thank you for your review and the suggestions that help to clarify the manuscript. I have duplicated your comments in full below, each followed by a point-by-point response including the modifications that will be adopted in the revised manuscript.

P5536 L25: I think a comment on the nature of DVM's according to Snell 2014 would be useful here.

Response:
I am not sure if I understood your comment correctly. The second sentence comments on the shared view on the nature of DVMs in this manuscript and in the Snell et al. (2014) paper. To make this more clear, I moved the Snell et al. (2014) reference to the end of the second sentence. The sentences now state:

Impact studies of climatic changes on spatio-temporal vegetation dynamics are often conducted with so-called dynamic vegetation models (DVMs). DVMs are mainly implemented as time- and space-discrete models, simulating ecological processes that are key to vegetation dynamics, such as establishment, growth and mortality, usually under consideration of biotic and abiotic influences (see e.g. Snell et al., 2014).

P5537 L27: Could do with references to models that use cohorts. I am not sure that it is very useful to refer so heavily to another paper here.

Response:
References to two example models using cohorts were added and the reference to Snell et al. (2014) removed, the sentences now read:

One example finding broad application in DVMs is the aggregation of individuals with similar properties into cohorts (e.g. Scheller and Mladenoff, 2004; Smith et al. 2014). With cohorts solely one representative calculation, instead of multiple replicate calculations, needs to be conducted.

P5541 L17: I think it is perhaps better to be explicit that you are referring to avalanches.

Response:
The sentence has been adapted according to the suggestion and now states:

Other processes which would require information on the neighbourhood are, for example, spatially connected disturbances such as snow avalanches (e.g. Zurbriggen, 2013).

P5542 L17: 'Supporting period' is not a very intuitive descriptive term here. Could the the need for the definition of the period be described more clearly?
P5542 L20: What happens if the bioclimate bins change between supporting periods? Is there an accommodation for a changing climate?

Response:
I rewrote the paragraph on the pre-structuring of the simulation area to clarify that one average is calculated for each supporting period and that the bioclimate bins can thus change between supporting periods. The paragraph now expresses that the supporting periods have the function to sample the bioclimate for each cell in time. To accommodate for a changing climate more than one supporting period is required. The more supporting periods, on the other hand, the more bioclimate types can theoretically result. More supporting periods will, furthermore, also lead to higher costs in the pre-clustering step.
The first referee also had clarification questions to this section, and I therefore not only adapted the text but also adapted a figure from the supplementary (previous Figure A.2) and moved it to the main text (see Fig. 3). This Figure gives an example for the pre-structuring. It shows the discretisation of the minimum winter temperature with the set E3 of bioclimate bins for one cell and how the bioclimate driver is then calculated from all associated cells. The new Figure includes the information of the

previous Fig. 4, which I therefore removed.

The adapted paragraph now states:

One of the challenges for an efficient organisation of the elements on the non-spatial layer was that the number of elements is not known in advance, because it is an emergent property of the simulation. To allow for an arbitrary number of elements on the non-spatial layer, the elements are stored in linked lists, instead of using an array structure with a predetermined fixed size, as usually done in space discrete DVMs. However, using one large linked list, and comparing all elements with each other, would be very inefficient and would lead to a large organisational overhead. To reduce the organisational overhead required for the comparison of elements during runtime, TreeMig-2L uses several linked lists, and only elements in the same list are compared. To define the number of linked lists, the fact is used that the bioclimate drivers are an input to TreeMig and can thus be used to pre-structure a simulation area (see Fig. 3 for an example, and Supplement Sect. A.2 for additional information). To pre-structure a simulation area, grid cells are classified into "bioclimate types". Each bioclimate driver is thereby discretised according to a set of pre-defined "bioclimate bins" (see e.g. Table 2). To limit the number of possible bioclimate types and to save computation time, the discretisation is only applied for temporal averages of the bioclimate drivers on a pre-defined set of "supporting periods". For each cell, one average per driver and supporting period is calculated. Two cells are classified into the same bioclimate type, if their averages fall into the same bioclimate bin for each driver and supporting period, whereby the bioclimate bins can differ in between supporting periods. For each existing bioclimate type, i.e. each type with which at least one cell is associated, an own linked list is used. The bioclimate drivers for a bioclimate type are calculated as the averages of all associated cells.

Caption of the new figure:

Visualisation of the pre-structuring into bioclimate types on the example of the minimum winter temperature (Min. winter temperature), one of TreeMig's three bioclimate drivers. **(a)** The minimum winter temperature driving cell$_k$ is averaged for each of the supporting periods (here P1: 1901–1930, P2: 1901–2100, P3: 2071–2100). **(b)** The range of the minimum winter temperature ($-14\,^{\circ}$C to $+10\,^{\circ}$C) is discretised into 13 bins with a resolution of $2\,^{\circ}$C (as e.g. done for E3, the set of bioclimate bins with the coarsest resolution – Table 2). The averages of the supporting periods P1–P3 for cell$_k$ are classified according to these bins. **(c)** Cells whose averages fall into the same bin in each of the supporting periods for each of the bioclimate drivers are classified into the same bioclimate type. The bioclimate driver (here the minimum winter temperature) of the bioclimate type is calculated as the average (black line) of its 287 associated cells (grey lines).

P5542 L25-30: Maybe too much detail? Defer to editorial advice on how much description of the actual code is desirable.

Response:

I removed the sentence. For interested readers this information is still contained in the supplementary material.

P5543 L2 : Is this sentence "The communication between the layers is asymmetric." needed? I found it rather inhibited my understanding of this paragraph.

Response:

I rephrased the sentence. The important point here was that the cell starts the communication. The paragraph now starts with:

The information exchange between the layers is induced by the cells.

P5543 L6: Are the light conditions those of the under-storey, above the canopy, or in gaps?

Response:

In TreeMig, each height class has its own frequency distribution of light which is calculated from the tree distribution of all higher height classes (see Lischke et al., 1998; Löffler and Lischke, 2001; Lischke et al., 2006). I added more information to the sentence, which now states:

For regeneration, additionally, the current bioclimatic conditions and the light distribution of the lowest height-class need to be accessed, because they influence the density of newly germinated seeds (see Lischke et al., 1998, 2006).

P5543 L13: Averaged densities of what?

Response:

Added 'of germinated seeds':

Averaged densities of germinated seeds falling below a pre-specified presence threshold are thereby set to zero.

P5544 L4: The concept of 'tracked' and 'untracked' species I find confusing. Can it be elaborated on slightly?

Response:

I added further explanation, the according sentences of the paragraph now state:

In TreeMig-2L this trade-off between accuracy and possible splits is approached by not considering splitting for all species, but only for a set of species previously specified as species to be tracked. I.e. for tracked species differences in the number of germinated seeds among grid cells associated with the same element can lead to splits. For untracked species, in contrast, the number of germinated seeds is not compared among grid cells and deviations can thus not lead to splits. Apart from the splitting step, all species are treated the same, in particular the number of germinated seeds on an element is calculated as the average over all associated cells for each species, no matter if tracked or untracked. As a consequence, untracked species might be represented less accurately, which can feedback on the tracked species via competition. Therefore, species that are expected to change their distribution in the course of the simulation should be defined as tracked species.

P5545 L6: 'influenced', not 'influence'.

Response:
Thank you for spotting this typo, I corrected the sentence and it now states:

The number of bioclimate types will also be influenced by the spatial extent of a simulation area:...

P5545 L 26: Maybe add some brief context here, like 'the scenarios pertain to different regions of Switzerland'.

Response:
The suggested context was added to the sentence, which now states:

TreeMig-2L simulations were conducted for two different application scenarios (A1 and A2), pertaining to different regions of Switzerland.

P5546: Again, perhaps some context might be of use here, in terms of 'scenario A1, in the flatter areas of Northern Switzerland' vs scenario B, a N-S transect across the whole of the Swiss Alps'.

Response:

Thank you, I added the suggested text to the application scenario descriptions, which now state:

Scenario A1, in the flatter areas of Northern Switzerland, stems from a study with a preliminary TreeMig-2L version without dynamic associations between the layers (Nabel and Lischke, 2013). Scenario A2, a north-south transect across the Swiss Alps, stems from a study investigating the influence of interannual bioclimate variability in simulations of the northwards migration of *Ostrya carpinifolia* Scop. (European Hop Hornbeam) with TreeMig-Netcdf 1.0 (Nabel et al., 2013).

P5546 L16: I am not sure what happens to the other species. Are there other species? How many? Do their dynamics not affect the distributions of the 'tracked' species?

Response:
The number of competing species was so far only contained in Table 1. This information has now been added to the main text and further text has been added for clarification. I hope that, together with the explanation to comment P5544 L4 above, it is now clear what happens to the other species and that they indeed compete with the tracked species.

The tracked species in scenario A2 is naturally the investigated migrating species (*O. carpinifolia*). The other 21 competing species simulated in this application scenario were not tracked. For A1 four out of the 31 simulated competing species were selected as tracked species: *Quercus pubescens*, *O. carpinifolia*, *Larix decidua* and *Pinus sylvestris*. These species were selected because they have the highest drought tolerance indices in TreeMig. With increasing drought severity (and increasing temperatures), these species are therefore expected to extend their spatial distributions, whilst the other species might be less effected or have declining distributions.

P5546 L25: This idea of 'stochastic extrapolation' could do with some more explanation. What is the timespan of the simulation compared to the availability of the driving data? Why does the extrapolation need to be stochastic?

Response:
I agree that this is an important topic. However, I suggest that this discussion is not in the scope of the current manuscript and I would like to refer to the publications Nabel et al. (2013, 2014) in which different extrapolation methods are discussed in detail on the example of application scenario A2.

P5547 L1: I am confused about the need to use/define these three sampling periods? Are they 'supporting periods"? Why not use the whole time series?

Response:
Thank you very much for spotting this relict. These are of course the supporting periods and this sentence as well as a sentence in the supplementary have been corrected accordingly. Regarding the question why not using the whole time series: I hope that the text change in response to the comments P5542 L17 and P5542 L20 above clarified that several supporting periods are required to accommodate for a changing climate.

The changed sentence state:

For both scenarios three supporting periods were averaged: the first 30 (A1: 1961–1991; A2: 1901–1931) and the last 30 years (both: 2071–2100), as well as the whole time span (A1: 1961–2100; A2: 1901–2100).

P5547 L7: Suggest adding "Simulations with scenario A2, a much smaller land area required one-tenth..."

Response:
I adapted the sentence which now states:

Simulations with scenario A2, having much less grid cells and requiring only one-tenth of the CPU time of scenario A1, were repeated 100 times.

P5547 L10: Are there not four versions, including the control 1L-ORG?

Response:
The original intention was to refer to three model versions in addition to the original model. However, I agree that this is cumbersome and adapted the sentence which now states:

To disentangle the effects of the pre-structuring into bioclimate types and of dynamic associations between the layers, simulations of four model versions with increasing complexity were compared (Table 3).

L5549 L15: To what extent is the 2L approximation dependent upon how the model is driven? In the nomenclature I am most used to, metrics of bioclimate are not used as drivers, e.g. annual/daily/hourly meteorology is the main driving data set. Longer 'bioclimate' averages are more often used to define the ranges of vegetation, not their driving data. Is there some confusion over the use of the term 'bioclimate' here that requires further clarification?

Response:
Using 'bioclimate' as a driver: TreeMig simulations require time series of three different annual bioclimate drivers, which are derived deterministically from monthly averaged meteorology (temperature and precipitation). Because for TreeMig, the bioclimate is derived in advance in a data pre-processing step, the number of required calculation steps during runtime is reduced, which is especially beneficial for stochastic replicate

simulations. However, TreeMig could also be formulated as a model directly driven by temperature and precipitation. Up to now the information on the meteorology data was only contained in the supplementary and I added this information to Sect. 3.1.1, now stating:

TreeMig simulations require time series of three different annual bioclimate drivers: the minimum winter temperature, the sum of daily mean temperatures above $5.5\,°\mathrm{C}$, and an index denoting the severity of droughts (Lischke et al., 2006; Nabel et al., 2014). These drivers are derived from monthly averaged temperatures and monthly precipitation sums (see e.g. Supplement Sect. B).

Additionally the commented sentences have been slightly rephrased:

Averaging of the bioclimate drivers of associated cells to obtain the drivers for the bioclimate types led to small but visible deviations (e.g. Fig. 3c and Supplement Sect. C). These deviations in the bioclimate driver entail deviations in the simulation results when comparing to results from one layer simulations, i.e. simulations with the original grid cell-based bioclimate driver (see column 1L-PB in Table 5 and Sect. 4.2). Using bioclimate bins with a coarser resolution thereby led to larger deviations in the driving bioclimate and, thus, to larger deviations in the simulation results (Table 5).

Regarding your first question: Following my experiences, the pre-structuring of the simulation area is fundamental for the efficiency of TreeMig-2L (which I do state in Sect. 4.1). Thus, for an efficient 2L approximation of any model, there needs to be some pre-structuring of the simulation area enabling to define a fixed set of linked-lists, among which the elements of the non-spatial layer can be organised. This means that it is e.g. not important, if the drivers of the model are precipitation and temperature or some derived bioclimate variables, but it is important that there is some sort of information which can be used for the pre-structuring. While this will be the case for most DVMs, it would be difficult, for example, to approximate a model with interactive climate, as e.g. in coupled land-atmosphere models. For such a model one could try a

pre-structuring of the land area based on a short time-span simulated with the original model. Since in such a case the driving climate time series is not known in advance, the averaging of the driver in the pre-structuring step would definitively need to be replaced by a dynamic approach in order to accommodate for a changing climate. However, as I also mentioned in Sect. 4.1, such a dynamic approach has not been tested with TreeMig-2L so far and would certainly require more computations and could thus lead to less CPU time reductions.

P5549: L10-30: This discussion is interesting, but slightly breaks the flow of the text. Might it be relegated to the supplementary information?

Response:
I would prefer to keep this discussion in the main text, also with respect to your question above, regarding the dependence of the method on how the model is driven. This discussion underlines the importance of the pre-structuring, explains that the pre-structuring is not necessarily coupled to the 'in advance averaging' of the bioclimate driver, and discusses some of the consequences of replacing the averaging in the pre-structuring step by a dynamic approach at runtime.

P5551 L9: Suggest replacing 'selected set' with 'resolution'.

Response:
Thank you for this suggestion. Several sentences in the manuscript and the supplementary were adapted according to the suggestion. The sentence referred to now states:

The level of the SC was thereby largely determined by the resolution of the applied set of bioclimate bins, the coarser the resolution, the smaller the SCs for all variables (differences in the SCs of up to 0.14 – Table 5).

P5551 L 25-30: Could you add a sentence here about why adding more tracked species reduces the error?

Response:
The trade-off between accuracy and computation costs related to the number of tracked species is described in Sect. 3.1.4. I hope that with the additional information introduced there as a reaction on the comment to P5544 L4 it is now more clear why tracking more species can reduce the error.

P5555 L2-5: Not sure I understand this sentence (Most importantly...). Is there a word missing?

Response:
The sentence has been corrected and now states:

The main determinant for the non-reducible base load is the number and complexity of processes requiring information on the neighbourhood of a grid cell, i.e. spatially linked processes.

P5555 L14: Updated reference for LPJ-GUESS:
Smith,B., Warlind,D., Arneth,A., Hickler,T., Leadley,P., Siltberg,J., & Zaehle,S. (2014). Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model. Biogeosciences, 11, 2027-2054.
ED references should probably include:
Medvigy, D., Moorcroft, P. R. (2012). Predicting ecosystem dynamics at regional scales: an evaluation of a terrestrial biosphere model for the forests of northeastern North America. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1586), 222-235.
and/or

Moorcroft, P. R., Hurtt, G. C., & Pacala, S. W. (2001). A method for scaling vegetation dynamics: the ecosystem demography model (ED). Ecological monographs, 71(4), 557-586.

Response:
Thank you. The references for LPJ-GUESS (Smith et al., 2014) and for ED (Moorcroft et al., 2001) have been added.

References

Lischke, H., Löffler, T. J., and Fischlin, A.: Aggregation of Individual Trees and Patches in Forest Succession Models: Capturing Variability with Height Structured, Random, Spatial Distributions, Theoretical Population Biology, 54, 213–226, 1998.

Lischke, H., Zimmermann, N. E., Bolliger, J., Rickebusch, S., and Löffler, T. J.: TreeMig: A forest-landscape model for simulating spatio-temporal patterns from stand to landscape scale, Ecological Modelling, 199, 409–420, 2006.

Löffler, T. J. and Lischke, H.: Incorporation and influence of variability in an aggregated forest model, Natural Resource Modeling, 14, 103–137, 2001.

Moorcroft, P. R., Hurtt, G. C., and Pacala, S. W.: A Method for Scaling Vegetation Dynamics: The Ecosystem Demography Model (ED), Ecol. Monogr. 71, 557–586, 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, 2014.

Interactive comment on Geosci. Model Dev. Discuss., 8, 5535, 2015.

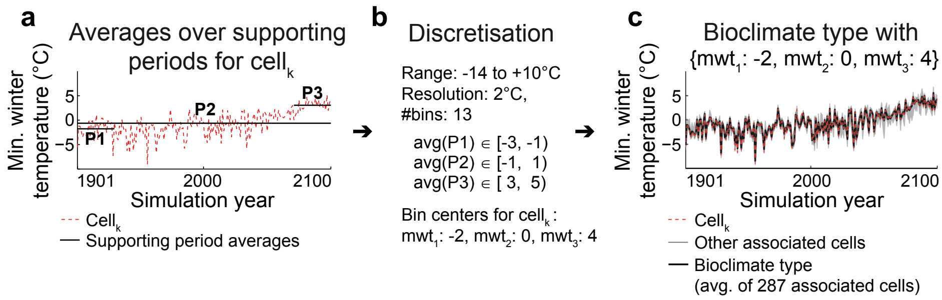

Fig. 3. See response to the comments P5542 L17 and P5542 L20 above for the figure caption