

Authors response to “Simulating migration in dynamic vegetation models efficiently with LPJ-GM 1.0” by Lehsten et al.

We would like to thank all reviewers for their efforts in reviewing our manuscript. Below is a point by point list of the responses and changes in the manuscript.

Anonymous Referee #1

Overall Comments:

The paper describes two approaches, for simulating seed dispersal in global-scale dynamic vegetation models. Vegetation migration in response to climate change (both past and future) is a major area of research, and the ability to simulate dispersal in DGVMs would be a major advance. There is certainly scientific merit in this manuscript, however there are numerous issues that need to be addressed. In general, since this is a paper about model development, more details about the model and justifications for their choices, need to be included.

Major Comments:

Seed production (Section 2.3.1) For the entire section – “each grid cell”, does this refer to the large grid cell or the smaller grid cells within?

Response:

There are no smaller gridcells within gridcells in our model. Whenever we mention gridcell we mean the entire gridcell since we have not subdivided them. We simply use very small (compared to typical LPJ-GUESS uses) gridcells. In each run all gridcells have the same size. However, in some runs we are only simulating a sparse field of gridcells, meaning that while all gridcells are taken into account for the seed dispersal, some of them (often many of them) are not simulated by LPJ-GUESS-GM, hence all seeds landing there will not germinate. We will stress this out in the paper.

L157 – “no specific age of maturity is taken into account”. Maturation age has been shown to be one of the most important factors for determining tree migration rates (e.g., Nathan et al. 2011; Snell 2014). This is especially relevant for trees, as most tree species delayed maturation. Please include a justification for why this was not included.

Nathan R, Horvitz N, He YP, Kuparinen A, Schurr FM, Katul GG. 2011. Spread of North American wind-dispersed trees in future environments. *Ecology Letters*, 14: 211-219.
Snell RS. 2014. Simulating long distance seed dispersal in a dynamic vegetation model. *Global Ecology and Biogeography*, 23: 89-98.

Response:

In our implementation we are not aiming to make a model to fit exactly the real world but rather to see whether our way to implementation leads to comparable results compared to TREEMIG. To be able to compare the results we tried to stick to the way it was implemented in TREEMIG, and there the start of seed production was based on the tree height, hence this is what we implemented as well. Height is also a good indicator of maturity. However we agree with the reviewer that age has certainly its merits and might be a more suitable variable in certain conditions. We have in earlier versions used age as a trigger to start seed production. In the next version the user will be able to switch between tree age and tree height as a trigger for seed production. We added this information to the paper in line 570. The mentioned LAI is used to calculate the amounts of seed similar to Lischke et al. (2006).

Comment:

L159 – please clarify what is meant by “seed bank”, and perhaps use another term. In ecology, seed bank has a very specific meaning (i.e., the dormant seeds in the soil that can germinate in subsequent years).

Response:

This is exactly what we mean. We now use this phrase to define seed bank in the text.

Comment:

How long do seeds stay in this “seed bank”?

Response:

The annual loss of seed germeability hence the decay time for the seeds is taken from the values in the cited TREEMIG publication (for Fagus it is 0.8). It is a species-specific value though currently the values are similar for all since there is little literature comparing the different species (described under 2.3.3. Seed bank dynamics).

Comment:

Does each grid cell have their own seed bank?

Response: *Yes. We mention this now in the text.*

So seeds enter after dispersal has already occurred?

Response: *Yes.*

Comment:

Or is this a central seedbank that all grid cells have access to?

Response: *No. Seed bank dynamics (decay, loss due to germination and new arrival due to seed dispersal). Is calculated independently for each gridcell. We have decided to keep the seed bank description short since we took exactly the same approach as in TREEMIG. We have explained it more exhaustively (at line 221).*

Comment:

L160-160 – please provide a justification for why you chose the LAI approach for seed production, and not the carbon allocation approach already implemented in LPJ-GUESS.

Response:

The currently implemented carbon allocation in LPJ-GUESS allocates a fixed amount of carbon to reproductive tissue which is then added to the litter pool. Here the basic unit is carbon rather than seed number, which is what we work with. There are two main reasons why we did not go the way to change the amount of carbon allocated to reproductive tissue depending on seed weight and seed number. Firstly we did not want to change the carbon dynamics within LPJ-GUESS as this is the result of a lot of fine adjustments and any change in the NPP allocation requires a substantial testing afterwards to assure that no unwanted effects occur. Secondly it would have meant that we also need to take the lateral exchange of carbon between cells into account. Currently there are a number of checks that assure that the carbon cycle is closed. Apart from transferring more data between cells, we would have had to adjust these checks as well. So in essence even though we agree that it would be more reasonable to deduct the carbon for the seeds from the NPP and to transfer it to the adjacent cells (though I guess the amount of carbon actually transported outside a cell is very small compared to the one that stays inside), for the purpose of demonstrating the migration mechanism (focus of this paper) we consider it not necessary but in the next version of LPJ-GM in which we plan to simulate historical tree migration we will implement exactly this. We are commenting on this in the paper under 2.3.3 Seed Bank Dynamics.

Comment.

In addition, please include some more information for how LAI is used to determine the number of seeds?

Response:

We added an example calculation of seed production to the paper (under the section Seed production).

Comment:

What value was chosen for maximum fecundity?

Response:

We used the same value as used in TREEMIG, we have added it to the paper (under Seed production).

Comment:

Is this species specific?

Response: Yes.

Comment:

Seed bank dynamics (Section 2.3.3)

L191-193 – this explanation is not sufficient. What is the difference between yearly loss of germinability and the amount of germinated seeds?

Response:

At the start of each year the amount of seeds that survived the last year is calculated. From this number the amount of germinating seeds is subtracted using a species specific fraction. Then new seeds arrive from the same gridcell and surrounding gridcells that are added to the seed bank the cycle begins again.

We explain this now in more detail in the paper.

Comment:

Is there a single seed bank for

each large grid cell, or each smaller grid cell inside? L194-198 – this is confusing.

Response: As there are no smaller grid cells within gridcells, there is one seed bank for each grid cell, of which all have the same size.

Germination (Section 2.3.4)

Comment:

L202 – 208 – why did you want to add more limitations to establishment? What is the biological justification for this? What does “we fixed this parameter to 0.01 after initial testing” mean? What properties did you evaluate? What does this parameter do? And how does your new limitation interact with the already implemented light limitation (i.e., does this filter happen before or after)?

Response:

This parameter relates to the total area of the gridcell (in which the seeds spread out). In LPJ-GUESS all simulations are done per square meter. If the same number of seeds land in a larger area there will be less per square meter. As we do not change cell size here we fixed this parameter. Basically LPJ-GUESS simulates a certain amount of seedlings to establish each year depending on the amount of light reaching the forest floor. Here, we calculate a probability that the establishment event that LPJ-GUESS simulates happens depending on the amount of seeds available, hence we only decrease LP-GUESS's internal establishment. So the seed limitation filter is applied before the light limitation filter. We made this more clear in the paper at line 235.

Comments:

Corridors (Section 2.5)

This entire section is also very confusing - looking at the figures helped, but the text needs to be clarified. Are these corridors the large grid cells, or the smaller grid cells inside? Or both?

Response:

As there are no small gridcells and large gridcells, but only one size of gridcells, the corridors are those gridcells in which the full LPJ-GM dynamics is calculated. The cells outside the corridors take part in the seed dispersal routine, but no vegetation dynamics is calculated in them. We made this clearer in the text and mention repeatedly that there is only one size of gridcell however only on those on the corridors the vegetation dynamic is calculated by LPJ-GM. To make this clearer to the reader we added a new figure (Fig. 3).

Comment:

L260 – 263 – How is the 1 km scale chosen, appropriate for a species with an average long distance dispersal of 200 m? Only a very, very small proportion of seeds would be able to travel 1 km or more.

Response: Yes it is a small proportion, but it is sufficient to establish the species in the next cell. But we agree, there is a discretization error involved, as in every spatial simulation. We mention this at line 355.

Comment:

The next section (L285), mentions

“parallel and diagonal corridors”. What does this mean? This should be described in this section, with some additional details provided.

Response:

We hope this is now clearer with the additional figure 3.

There are basically two types of gridcells (all with the same size). First there are the cells for which LPJ-GM calculates full vegetation dynamics and seed production (type 1). Secondly there are cells (type 2) for which LPJ-GM assumes a seed production similar to the nearest neighbor for which full vegetation dynamics is calculated. Hence there is a complete matrix of seed production for which one of the two described algorithms (FFTM and SMSM) is applied to calculate seed dispersal. Only in those cells for which LPJ-GM calculates the vegetation dynamics these seeds can cause trees to establish and to produce new seeds. These two types of cells are arranged in a way that the type 1 cells form a corridor surrounded by type 2 cells. Since the diagonal corridors are also parallel we agree that the wording is unfortunate. In the revised version we mention that they are north-south, east-west, northeast-southwest and northwest-southeast corridors and explain this more extensively

Comment:

Results, Explicit seed dispersal (Section 3.1)

It is not clear what results this section is talking about, nor how it relates to the rest of the manuscript. Referring to “pre-studies” is not helpful (i.e., these results are not part of the current manuscript? So why are they included?).

Response:

This pre-study is not part of the manuscript but part of the supplementary material. And it is mentioned in the paragraph. The term pre-study is misleading and we have replaced it and now directly point to the supplement. We mention the Matlab supplement in the Methods where we added a part about the performance of the different algorithms. We also highlight the results in the Results section and discuss this in the Discussion section in the revised version.

Comment:

There are also no values in here at all.

Response: Running times for FFTMS and SMSM are compared to a large detail in the table 1. We have pointed to that in the text as well.

How much faster did the FFTM or SMSM perform compared to the explicit dispersal?

Response:

In the early stadium of the work we did an explicit seed dispersal in LPJ-GM. It became quickly obvious that given the current computation ability this would not allow us to simulate larger areas. We then developed the two methods (FFTM and SMSM). Re-implementation of the explicit dispersal algorithm only to show that it will be much slower would take a considerable time. Therefore we went a different way, by implementing FFTM and SMSM and explicit dispersal mechanism into a Matlab script. This allowed to concentrate on the running time needed for the dispersal mechanism, which is the focus of this paper. A direct comparison of running times required for the seed dispersal for the different algorithms for different area sizes is done in the script and the results are plotted in the pdf. The script also allows the user to simply cut and past the code into Matlab and play around with it.

Comment:

Also, this is the first mention of a Matlab script (perhaps should be mentioned in the methods?).

Response:

We now have a whole section to do so. It was actually just meant as an add on to aid explaining the methods since reading an implementation helps to be implement it in any other model. But since it is also used to evaluate running times we have covered it in the methods and results section in the revised version (Sections 2.7 and 3.1).

Comment:

Since (I assume) the Matlab script doesn't include the additional processes from LPJ-GUESS, how comparable are these results to what you would get in LPJ-GUESS?

Response:

In table 1 we are comparing the running times of the FFTM and the SMSM with the running time where seed production is calculated but no seed dispersal is performed, while in the Matlab script we are comparing only the seed dispersal calculations of the two new mechanisms and the explicit mechanism leaving out the vegetation dynamics.

*Hence though one cannot precisely calculate the difference, one can make a rough estimate. Table 1 lists the percentage of the time used for the dispersal for the FFTM method, Supplement 2 shows the increase in computation time for the dispersal algorithm between the FFTM and the explicit seed dispersal. Hence, by multiplying the time needed for the FFTM in table 1 with the factor from the figure in the Matlab script one can estimate the total difference. As an example: The simulation of 100*100 cells with LPJ-GM uses in total 1800 cpu*h of which ca 200 cpu*h are used for the calculation of the dispersal (11%; table 1). According to the graph in supplement 2 the explicit simulation needs one order of magnitude longer than the FFTM, hence a rough guess would be that instead of 200 cpu*h as used for the FFTM, an explicit seed dispersal would need 2000 cpu*h, which would increase the total required simulation time for 100*100 cells to 4000 cpu*h. One can see in the plot in the supplementary material, that with larger areas, the differences between calculation time of FFTM and supplementary material increase to two orders of magnitude.*

Hence for larger areas the calculation of the seed dispersal would dominate the required calculation time even more. We added that the performance differences in the Matlab script are only to be seen as rough estimates in the Methods section dedicated to the Matlab script.

Minor Comments

L34 – not “at least”, which implies 1 km or greater. But should be “at maximum” implying that 1 km is the greatest size that can be used.

Response: We changed this.

Comment

L 37 – what is “it”?

Response: It stands for ‘simulating the local dynamics’ We changed this.

Comment

L39 – what “both methods” are you referring to here? The comparison of the Fast Fourier transformation vs the iteratively shifting seed matrix, or the comparison of between the simulations with all grid cells, versus the corridors?

Response: we mean FFTMS and SMSM and will name them in the sentence to make this clear.

The corridors are not a method but a way of placing cells.

L39 – what does “reliable” mean?

Response:

It means that for both methods (FFTMS and SMSM) comparable results are gained by calculating either corridors or the whole area. We will rephrase this.

L59 – awkward wording.

Response: We rephrased this.

L59-79 – both of these paragraphs are missing appropriate references. They have none, but include several statements which need to be referenced.

Response:

The first paragraph expressed mainly the viewpoint of the authors, but the second paragraph clearly states facts and we added an appropriate reference.

L95 – although this is explained in more detail in the discussion, it would be helpful to have this information in the introduction. (i.e., what did previous approaches do, and why were they limiting).

Response: We had in fact moved the review of other methods from the introduction to the discussion but will provide a small summary part of the Snell et al.(2014) approach in the introduction section in the revised version.

L108 – 110 – a few more details about LPJ-GUESS? This one sentence is vague and particularly unhelpful for understanding what this model does.

Response:

LPJ-GUESS has been described in 200+ publications (though most of them focus on a small side aspect or a new development that did not become part of the standard version). Finding the right amount of information given to the reader is tricky. We have extended the model description.

L127-128 – if all vegetation is killed, and now seed dispersal is active, you MUST have some vegetation or you won't have any seeds?? Where does the first generation come from?

Response: As described in 225-230, for some cells seed limitation does not apply until a certain point in time. These cells are the refugia of the species which have free establishment. In the example simulations these are the cells in the upper left corner. Hence after the clearing of the vegetation trees can establish freely here, produce seeds that can subsequently disperse to the surrounding cells.

L142-150 – please clarify this how this occurred. So instead of one grid cell with

multiple patches, you simulated one grid cells with multiple grid cells? But these smaller grid cells, had spatial locations and could interact with each other (unlike patches)? This was my interpretation, but this needs to be clearer. A conceptual figure would help.

Response:

We simulated only one patch per grid cell. However in typical LPJ-GUESS simulations, grid cells are as big as the climate data dictate, e.g. 0.5 up to 2.5 degree longitude/latitude.

In our simulation, grid cells are small compared to the standard LPJ-GUESS size. Hence for the area simulated in one standard LPJ-GUESS simulation gridcell (with several patches), LPJ-GM would place many gridcells (with one patch). So there is not one gridcell with multiple gridcells, but just many small grid cells. We thank the reviewer for the suggestion to make a conceptual figure and included it in the revised version (fig.3).

L188-189 – need more details about how these parameters were “roughly estimated” if this approach is to be applied in other models or for different species.

Response:

The term ‘roughly estimated’ indicated that there is a high uncertainty connected to these values. We removed the roughly estimated as we want to indicate that they were not estimated by us, but by Lischke et al. 2006.

L286-289 – this sentence is confusing.

Response: we broke up the sentence and made it clearer.

Figure 5 is not clear – what is causing the white areas? Neither the results nor the methods addresses what the simulation set up was that could cause this pattern. The explanation “no seeds were able to reach them” is not true, as seeds obviously reached all the way around the white circles in the center (i.e., beech arrived in year 2500, but then never migrated in?).

Response:

The main purpose of the simulation displayed in Figure 5 was to demonstrate a how the method performs when a certain seed dispersability is defined caused by a certain terrain.

To demonstrate the effect we created areas in which the permeability for the seeds were set to zero, hence they could not enter the grid cell. This is shown in the methods section in Figure 2 which shows the probability of entering a new grid cell. In the blue areas in figure 2 the cells have a zero permeability hence seeds cannot enter the cell. We have now updated the colorscheme of figure 2 to highlight the areas with zero permeability and added a sentence to the methods section explaining this.

Comment:

Numerous small grammatical errors throughout (this is not a complete list, just a few examples). L25

L51-52, “to have a sufficient amount of seeds”

L72 – unnecessary “of”
L86 – unnecessary “the”
L202 – “depending stochastically depending”
L305 – “Using at a distance of”
L347 – “.we are the first that manage to implement. .

Response:

We have changed these grammatical errors and check the text again for further grammatical issues.

Anonymous Referee #2

Received and published: 21 September 2018

Comment:

The paper presents two methods for simulating tree species migration, newly implemented in the dynamic global vegetation model LPJ-Guess. I find the paper mostly well written and generally an interesting scientific contribution.

Response: Thank you for this summary.

Comment:

What I, based on the presented material, cannot consent to is the reoccurring statement that the model can be used for continental simulations of multiple interacting species, nor that it is suitable for DGVMs beyond special cases (i.e. species simulations in Europe).

Response:

We agree that the performance of the dispersal algorithm as presented by now creates a doubt that a continental simulation is possible. Given that we clearly make this claim should require for us to present an algorithm that would have a suitable performance. The main aim of this paper was to introduce the two algorithms into the DGVM. We aim to perform continental scale simulations in an upcoming work. Hence we did not invest a lot of time into optimization, except for a paragraph in the discussion, for simplicity of the paper. We will rephrase all statements where we state that our implementation can be used for continental applications and write that it has the potential to be used for large areas and also has a lot of potential for performance improvements. Again the intention was to present the two algorithms for seed dispersal, while any continental scale simulation experiments would require to also present a completely new parameterization of some of the trees, as well as many other aspects hence we would like to refrain from performing continental scale simulations for now.

We completely agree with the second statement. Given that we worked a lot in the past with species simulations we completely forgot that the currently most common application of DGVMs use PFTs and there it is not necessary to include seed dispersal.

Comment:

Most DGVMs use plant functional types with mixed dispersal/reproduction traits, particularly when used for large spatial applications. The example application deals with only two species, and only the dominant late successional tree species *Fagus Sylvatica* is tracked, which has a quite narrow dispersal kernel. Furthermore, the application deals with a homogenous landscape. From what the authors show and write I am not convinced that/how a continental simulation with multiple interacting and dispersing species would/can be possible.

Response:

Again we completely agree that only for simulations on the species level, an inclusion of seed dispersal is useful. Currently all dispersal simulation is performed at the master node, if the dispersal simulations for each species are performed at one node per species there should be (theoretically) no reasonable reduction in performance. We also repeat that we have not proven that we can simulate at a continental scale, but only that we are 2-3 orders of magnitude faster in simulating seed dispersal compared to the explicit simulation. Whether this is sufficient for a continental simulation will be shown in successive work. We will rephrase this claim.

Comment:

From the paper I understood that using FFTM with widely spread transects would not be appropriate in heterogenous areas. SMSM with terrain, on the other hand, would not save enough computation time to be applicable on continental area. Is the plan to use FFTM with transects in homogeneous areas and SMSM in heterogeneous areas? But if so, how would these algorithms then communicate with each other in a continental simulation?

Response:

Though not formulated in the paper (for simplicity), yes this is the main idea. Depending on the parameters of the species specific dispersal kernel, there is a maximum distance that the seeds are transported (theoretically there is no such limit, but given the strong decrease of the tail, this assumption has no influence on the final result). One option is to define a certain area as heterogeneous. The seeds produced in this area are dispersed by the SMSM algorithm, while the seeds of the remaining area are dispersed by the FFTM. Though for both the seed production is only taken into account for the assigned areas the seed fall will be calculated for the area plus an edge surrounding the area with the width of half a maximum kernel width. In a last step the dispersed seeds of the two methods are added to a final distribution of seeds. This way there is no complicated communication of the two algorithms, but the edges of the areas that are simulated overlap while the areas of the seed production that go into the algorithm do not. We have presented this reasoning in the Discussion as a last point in 4.2 Comparison of the two dispersal methods.

Comment:

Given that the paper, the presented ideas and the LPJ-GM implementation are already a substantial contribution, I recommend reconsidering the (over?) statements regarding continental applications and DGVMs e.g. in the last sentence of the abstract and particularly the first sentence in the discussion and talk about DVMS with species and spatial extents exceeding applications of a few ha, which is a good and sound contribution.

Response:

We have tuned down our statements and clearly show what level of areal coverage the algorithms are can simulate.

Comment:

Another way could be closing the explanatory gaps, i.e. (1) discussing issues with DGVMs and how DGVMs, which usually use plant functional types (PFTs) for large scale applications, could be parameterized for the algorithms, (2) discuss the costs/difficulties of an application with a realistic number of interacting species, with differing dispersal traits and (3) explaining how a realistic continental simulation could be assembled with the FFTM and/or the SMSM simulation, given spatial fragmentation and spatial heterogeneity.

Response:

We have done this in the Discussion.

General comments

1. In many places in the text the authors state that transect simulations lead to similar/only slightly reduced migration speed. However, from the figures/table it seems to be quite a significant underestimation, and the less transects the worse the underestimation of the migration speed (>20%; i.e. in the 3000y application >600y delay). I recommend stating this clearly and to discuss the consequences.

Response: An underestimation of 20% might look large, but given that most current approaches completely ignore seed dispersal and also that the parameterization of the seed dispersal kernel comes with quite an uncertainty puts the 20% in perspective. The currently reached migration speeds are very likely to be too low as described in the Discussion, which is probably caused by uncertainties in the parameterization of the seed dispersal kernel. However as the aim was to implement the two methods in a DGVM we aimed to keep the kernel similar to the parameterization within TREEMIG to be able to compare results. We did remove the word 'slightly' when mentioning the migration speed reduction.

Comment:

2. The authors claim that they fulfil the stochastic requirements because they have 200 or so 1km² grid-cells when comparing to the usual 0.5 x0.5 grid cell. However, this only holds if the spatial heterogeneity caused by the stochastic disturbances and stochastic mortality does not affect tree species migration. In the example application the authors choose *Fagus Sylvatica*, a dominant late successional tree species, and I can imagine that for this species the stochasticity might indeed play a minor role.

However, what in case of e.g. pioneer, less dominant/more specialised species? These might depend on disturbed areas for establishment – is the transect approach valid for such species? I would find it very helpful to see how the stochasticity and the few available transect cells might affect the spread of such species.

Response:

Pioneer species are typically fast migratory species, hence they would typically be able to colonize the area before the late successional species arrives (if both occupy the same climate space). If on the other hand the late successional species has a larger climate range you are correct that the early successional species will be hindered in its migration into the few spots that are available over a short time. This is true for both a full simulation as well as a simulation along transects. The fact that the seed survival of early successional species is typically higher due to the lower seed mass, should allow the spread still in the small temporal successional gaps. Again the focus here is to introduce the methods. A parameterization of the species that results in a migration speed comparable to observed values is outside the range of this study. However we agree that this is an important point and will discuss it at line #EARLY SUCESSIONAL SPECIES.

Comment:

It is correct that if applied globally DGVMs usually use 0.5 grid-cells, however when applied as DVMS on continents or regionally the resolution is usually much smaller. See e.g. the dispersal experiments by Snell (2014), and the simulation of European potential natural vegetation with LPJ-Guess (Hickler et al. 2012). 4. I would recommend referring a bit more to relevant literature in some parts of the text, since several of the ideas/methods have already been discussed/used elsewhere. (I mentioned some references in the specific comments list below).

Response:

The maximum resolution is dictated by the climate data available. Given that LPJ-GUESS is parameterized using CRU climate data the simulations at coarse scale are typically performed at 0.5 degree or sometimes using the CRU climatology to bias correct the CRU time series it is

performed at 0.25 or 0.1 degree. Even when run with 10 patches and at 0.1 degree would result in 250 patches per 0.5 degree which is comparable to the 200 simulations that we perform at any 50 by 50 km cell. However we now mention that some regional simulations use finer grid cells.

Comment:

What I miss in the current introduction is a bit more on why migration is missing in DGVMs. The authors state that one reason is the '1D' property, i.e. that cells are not interacting and thus the computation costs of making them interacting. But what should also be mentioned is the problem of parameterisation: DGVMs usually use PFTs, often compiled of species with various different traits with respect to migration (dispersal vectors, competitiveness, generation times, ...) (e.g. Snell et al., 2014).

Response: We now mention this at line 103.

Comment :

Furthermore, if I understood it correctly, the example simulation is for 3000y and the tracked species migrates 100km in that time. Several of the criticized studies with DGVMs (e.g. "land use change on vegetation and ecosystem properties") would use well below 3000y; mostly around 100/200y – given the comparable cheapness of 1D simulations and the mentioned constraints due to parameterisation: wouldn't a 'no dispersal' simulation be sufficient for many simulations with large spatial extent and coarse resolution?

Response:

As already mentioned before (and highlighted in the Discussion) the migration speed that we are calculating are way too small compared to measured values. The aim of this paper was to introduce the method and here we choose to measure our success in response to the migration speed simulated by TREEMIG since we used a similar dispersal kernel. Any real world application will require a new parameterization of the kernel to gain a realistic speed. Currently most simulations are not 'no dispersal' simulations but 'extremely fast dispersal' simulations (given that they use free establishment). Otherwise they could not show any response of vegetation on climate. For short term studies where the time horizon is well below the generation time of the species there is a limited use of a dispersal kernel. Especially for Europe where land use is dominating the vast majority of the landscape and at least in the northern part also plants (often alien) species in forests, any kind of simulations assuming semi-natural conditions are questionable. However to understand current tree distribution in those parts which are still semi-natural and especially to understand forest species history taking seed dispersal into account might be important. We now mention this reasoning as a last point in 4.4.3.

Comment:

I would appreciate a more detailed description of the SMSM method. Maybe an illustration? Would this method work with a species with a more pronounced long distance dispersal tail than *Fagus Sylvatica*? What would this mean regarding computation costs? How to parameterise the SMSM? Could a setting like Fig 5 be simulated with transect at all? Looking at the supplementary figure it seems that the matrix shift method with a terrain has a very small computational gain?

Response:

Thank you for suggesting to add a figure that will illustrate the SMSM, we have now added an example sheet which demonstrates the SMSM additionally to the Matlab code which also gives the full details.

With respect to the long tailed distribution: within each SMSM step the maximum dispersal distance increases by one cell. Hence in our example we have a maximum dispersal distance of 10 cells (10 km). A simulation in which you want to allow 20 cells maximum distance require twice as much

computation time for the SMSM. For long tailed species the FFTM is certainly better suited, given that this method has no such limitations.

The parameterization of the SMSM is described in detail in the Supplementary material. While in general any kind of dispersal kernel can be used in the SMSM, this requires to stack several Gaussian kernel on top of each other and hence would of course increase computational demand. If computation time is an issue, the FFTM is a better choice.

Compared to the explicit simulations of seed dispersal, the SMSM is still 3 to 4 times faster. This is much slower than the FFTM which is 1.5 order of magnitudes faster but the improvement is still significant.

Comment:

I would appreciate more discussion of the limitations and a clearer directive how to apply the algorithms for a continental simulation, if possible. When reading the text I got the feeling that the remedy for the FFTM limitations (heterogeneity/fragmentation, wind directions) is to use the SMSM, but that this method, particularly if used with terrain, is not performant enough for continental applications. Some more buzzwords for the limitation section: parameterisation of SMSM; species parameterisation; fragmentation when using the FFTM; what about ecosystems with many species (i.e.. tropics).

Response:

You read correctly that FFTM is not able to handle landscapes resulting in heterogeneous seed dispersal while SMSM has strong performance constrains. As a matter of fact, the FFTM is still applicable if the barrier is larger than the kernel width since it will place seeds there but if the cells are not suitable then the seeds will not germinate (Baltic sea, alps). The SMSM is only required in areas where we have (or rather know) different dispersal tail lengths depending on the terrain. This might be the case in some valleys in the alps, where seed dispersal acts mainly along the valleys, but not the mountain, but given the typical resolution of the output for continental studies this might not be necessary to apply the SMSM at all, while for finer scale studies SMSM might be the best choice. We are discussing this in the Discussion section together with the mentioning that for some simulations the best option might be to combine the FFTM and the SMSM at the end of 4.2.

Comment:

Reduction of migration speed by >20%, i.e. in the 3000y simulation > 600y delay. 8. The editor provided me with the model code. Unfortunately I was not able to understand how the simulations were done. There are no hints on how the simulations were conducted, nor was I able to identify the configuration file (instruction script (ins)?) used for the simulations or to find out how/where the transects were defined. I know that it is cumbersome but in the spirit of "good scientific practice" it might be nice to provide and mark the configurations files?

Response:

The transects are defined in the gridlist. Basically while a typical LPJ-GUESS gridlist contains only columns one for the longitude and one for the latitude, in LPJ-GM there are additional columns in which for each species a time is given in which free establishment is allowed see below for the start of a gridlist.

23 50 TeBS,101.0,IBS,101.0

23.01 50 TeBS,100.0,IBS,10000.0

23.02 50 TeBS,100.0,IBS,10000.0

The first line indicates that at position 23 degree longitude, 50 degree latitude both species TeBS (temperate broadleaved Summer green tree or beech and IBS Intermediate shade tolerant broadleaved Summer green tree; or birch) are allowed free establishment (hence no seed limitation) at the year 101, which is one year after the initialization phase for nitrogen initialization. Hence this cell would form a refugia for beech. The next cell is located at 23.01 degree longitude and 50

degree latitude and beech is only allowed free establishment after the year 10000, hence not within the simulated time of 3000 years, it can only establish at this site if seeds arrive there. Birch is allowed to establish at this site without seed limitation.

The transects are defined in a way that only the cells that are on the transects are listed in the gridlist.txt. This way of defining them might not be the most elegant one, but since the current setup of LPJ-GUESS simply cuts the gridlist into as many pieces and distributes them into different directories in which the simulation is performed, this way I did not had to read in a separate file for the refugia definition, and I am sure that the information is linked to the gridcell.

The configuration script (the ins file) is similar to the one used in Hickler et al. except that it contains these additional entries at the global level.

! migration INSTRUCTION

```
years_total 3000 ! How many years the dispersal simulation is performed
domain 23      50      0.01      0.01 ! which domain is simulated and with what resolution
param "size_lat" (num 100) ! how many cells are in the domain along the latitude
param "size_lon" (num 100) ! how many cells are in the domain along the longitude
dispersal_patchsize 0.99 ! How big a single patch is.
if_dispersal_fft 1 ! whether FFTM dispersal is performed
if_dispersal_float 0 ! whether SMSM dispersal is performed
if_dispersal_ext_fft 0 ! whether another variant of FFTM (not described in the paper) is
performed
stochastic_seed_est_scaler 0.01 ! scaler for the patch size
output_interval 10 ! in years save space since not all years are needed in the output
```

Each species contains the following extra parameters which are taken from TREEMIG (here are the values for beech):

```
max_fecundity 29. ! maximum fecundity
min_height_for_maturity 14.4 ! minimum height for maturity
germination_rate 0.3 ! rate of seeds germinating per year
max_seed_age 3.3 ! maximal survival times for seeds in seed bank
short_range_disp_frac 0.99 ! fraction of seeds that go into short seed dispersal
short_disp_alpha 25 ! parameter for short distance dispersal
long_disp_alpha 200 ! parameter for long distance dispersal
```

I also would prefer to make the whole model code publicly available. However current policies within the modelling consortium only allows to give access to model code after individual contact with the author. I decided that my unit containing the implementation of the code for the actual migration will be made publicly available (as a supplement to this paper), but there are of course some other small bits and technical issues, like for example the MPI related code that is located in other units.

Specific comments:

1.1: Maybe consider to adapt the title, since LPJ-GM does not necessarily lead to a more efficient simulation of migration in dynamic vegetation models per se – e.g.: "Simulating migration in dynamic vegetation models efficiently on the example of LPJ- GM" or maybe better "Simulating migration efficiently in the dynamic vegetation model LPJ-GM"

Response:

Yes we have changed the title to "LPJ-GM 1.0: Simulating migration efficiently in a dynamic vegetation model"

Comment:

1.21: Most DGVMs do not use species but plant functional types (PFTs)

Response:

We mention this in the second paragraph of the introduction.

Comment:

1.31: From the last Figure in the supplementary, SMSM with terrain seems to be much slower than FFTM?

Response: Yes it is especially now that we have optimized the Matlab code (though at the expense of readability) it is. We deleted the word ‘marginally’ in the abstract. However it is still faster than an explicit seed exchange.

Comment:

1.40: “Furthermore, with the transect method both methods”?

Response: We have reformulated the ‘with the transect methods’. We have replaced continents with large regions, since we have not really shown that continents can be simulated with our method. We have made the sentence clearer now.

Comment:

1.49: DGVMs assume that some instance (i.e. species) of the PFT can establish

Response: We added a paragraph stating that:

“Additionally to the reasons mentioned before, most DGVM applications use plant functional types which comprise typically species with very different traits with respect to migration (e.g. dispersal vectors or seed properties), hence introducing migration would require to split up PFTs into smaller groups and to parameterise the additional properties. “

Comment:

1.51: Something seems not correct with the embedded sentence – maybe that instead of the?

Response:

No here that would give a different meaning. We inserted an ‘a’ and hope the sentence is now easier to read. Line:56

Comment:

1.53 & 63: Anyway DGVMs usually do not simulate species but only PFTs

Response:

Since we wrote already before that we are only considering species simulating DGVMs we consider this covered.

Comment:

1.60: When considering ecosystem properties in the future hardly any study would make projections »100y, maybe 200y, but the example in this study uses 3000y. Wouldn’t – based on what is shown in this paper – a “no migration between large grid-cells (0.5 °) assumption” be appropriate for studies with ~ 100-200y?

Response: We mention this now with the following sentences. “For periods of less than 50-100 years ahead, which corresponds to at most a few generations of most tree species, the explicit modelling of seed dispersal might be less important for simulating tree distributions, in particular when taking into account the overwhelming influence of human activities. ”

Comment:

1.76: Another example of a model even accounting for wind speed and direction: LAVESI-WIND (Kruse et al. 2018)

Response:

Thanks for pointing us to the paper, we are citing it now in line 86.

Comment:

1.82: What does the spatial heterogeneity refers to in this context – soil and climate? If I understood the set-up right LPJ-GM also does not account for such heterogeneities within the grid cell, only to such with regards to species interactions and stochasticity?

Response:

Yes: Soil and climate, mountains blocking seed transport as well. This sentence should simply highlight that a simple transfer of migration speeds calculated with models at fine scale into models at coarse scale is challenging.

Comment:

1.94: why every time-step? LPJ-GM only does it once per year?

Response:

Yes of course we mean annually. We change this at line 103.

Comment:

1.100: If I understood it correctly the presented simulations only simulate two species.

Response:

Yes in our example simulation only two species are simulated but the method can simulate more species in a real application case.

Comment:

1.101: What would a simulation with several species look like, does each need one FFT/SMSM? What are the resulting costs?

Response:

In the current simulation time that we present in the table we are actually simulating the seed dispersal of both species independently (though it would of course be faster to only simulate one species). Yes each species needs its own FFTM or SMSM to be performed if the migration of several species is to be evaluated, however they could be potentially performed at separate nodes which would decrease calculation time again. We are mentioning this in the Potential further improvements section.

Comment:

1.110-111: Please list a few key references describing LPJ-GUESS 4.0

Response:

We included Smith 2014 and Lindeskog 2013 which are the main references describing the 4.0 version.

Comment:

1.119: Above and below this node is called master

Response:

We now call it master here as well.

Comment:

.123: “no seed dispersal”-> “no seed limitation”?

Response: Yes we added this as well. Line 146

Comment

1.130: There are species producing seeds throughout the year (see e.g. Owens 1994, Brokaw, 1998)

Response: Yes this is one of the discretization errors that we have to make. Given unlimited computing power and knowledge of weather conditions and plants behavior, we would perform the FFTM or SMSM daily over the time when seeds are produced. However, as a first improvement of the situation in which most models do not consider seed dispersal at all, we suggest to simulate at an annual time step.

Comment:

1.140: Here or generally in LPJ-Guess?

Response:

It is variable but this is the recommended size.

Comment:

1.153-157: How is this similar to Lischke et al., 2006? Lischke et al. (2006) do not mention LAI but state: "The number of seeds S produced per year by each tree depends on its height, species and mast seeding period."?

Response:

We also use the height of maturity, but no mast seeding period. While Lischke et al. scaled the seeds with height we did scale them with LAI, you are correct this is not the same and we have taken that sentence away as it was meant introduce into the chapter but it is not necessary.

Comment:

1.175: For *Fagus Sylvatica*?!?

Response:

Added. Actually we also simulated seed dispersal for Birch but since Birch is set to no seed limitation.

Comment:

1.181-182: But wouldn't the implementation of wind direction lead to anisotropy and therefore make the FFTM not applicable anymore (E.g. Neupane (2015))?

Response:

The FFTM can apply any shape of seed dispersal kernel, it can just not change it with the landscape. Hence certain wind directions are possible like the kernel used for illustration, which is also skewed, for example by wind. Neupane simulates effects of the landscape on fruit dispersing birds. Such an effect would have to be modelled by the SMSM. If the different wind directions in different parts of the domain (e.g. caused by a certain terrain) is to be taken into account, this also needs to be done by SMSM.

1. 185: maybe use θ and add the $\theta = 1$ in the text below?

We changed this equation to make it clearer.

Comment:

1.186: long term -> long distance?

Response: *Changed.*

Comment:

1.197: but how is the number of seeds defined in this case, since in the next para it is stated that the establishment of seedlings depends on the number of available seeds?

Response:

LPJ-GUESS calculates the number of established individuals per species depending on the light reaching the forest floor. LPJ-GM takes this value and sets it to zero depending on the presence (or rather absence) of seeds. In case of establishment free from seed limitation (in our case the birch), this step is not performed. Hence the species can always establish depending only on the light reaching the forest floor.

Comment:

1.202: “depending stochastically depending”

Response:

Thanks for spotting this repetition. It is fixed.

Comment:

1.206: “seed bank per and the germination” remove the per?

Response:

Done

Comment:

1.216: The authors should definitely mention that the method has also already been broadly applied in simulating dispersal. E.g. have a look at Powel (2001) + shortly googling I e.g. found Pueyo et al. (2008) and Prasad et al. (2013) and I assume there are more.

Response:

We are now mentioning that there are a number of applications which already use ffts to simulate dispersal and cite a few of them in the first paragraph of 2.4.

Comment

1.235: “different wind distributions” -> only if they are valid for the whole simulated area, or?

Response:

We added this remark in line 270.

1.242: How is this proportion determined?

Response:

A few lines later we point to a derivation of the parameters (in this case this proportion) in the supplementary material S.1.

Comment:

1.242: 1km² cell?

Response:

Yes in our application all cells have one km² extent.

Comment:

1.249: How often is this done/ needs to be done to account for long distance dispersal?

Response:

Currently this is done 10 times hence we are reaching a maximum of 10 km.

Comment:

What happens with the seeds at the boundary of a simulation area?

Response:

For both the FFT as well as the SMSM simulation we extend the area by one kernel width to avoid / minimize edge effects. Basically all seeds that land of the seed domain are lost.

Comment:

1.255: Figure 2 is not cited in the main text (only in Fig. 5).

Response:

Thanks for spotting this. This sentence must have gotten lost in one of the internal revisions. We now refer to Figure 2 in the description of the simulations in line 320.

Comment:

1. 266: Wouldn't the heterogeneous landscape be much more crucial to test the applicability of the methods?

Response:

No the idea is to only use the corridors in homogenous landscapes and to speed up the simulation there. In heterogeneous landscapes this simplification is not suitable. Hence we only test the corridors in homogenous landscapes.

Comment:

1.274-275: And? But?

Response:

And we do not want to strongly increase the migration speed. We have spent a lot of time trying to come up with a better solution like some kind of distance weighted average, however so far we have not found one. Hence we prefer to have a reduction by 20% hence a conservative estimate rather than a strong increase which also was heterogeneous within the simulated area depending on the arrangement of the corridors. We have not given up the hope to come up with a better solution in a real world application.

Comment:

1.276: Out of interest: how many CPUs were used/ what computing environment?

Response:

We used 200 nodes (with 20 nodes per CPU) at the LUNARC computing facilities.

Comment:

Would it be possible to add a 'no dispersal and no communication' (i.e. a 1D) simulation for comparison?

Response:

Yes have been working on it and planned to add it to the final version of the paper. As it turned out the differences between the version with communication and without are negligible. We also mention this now in the ms.

Comment:

1.303: 1km² grid cells

Response:

We added this.

Comment:

1.306: somewhat? >20%!

Response:

We have removed the somewhat.

Comment:

1.309: Maybe add the numbers for the variability

Response:

What we mean here is visual realization that the distance of the points increases. Since we do not use a mean value to estimate the migration front, it is hard to quantify this variability since we have a different variability above compared to below the line.

Comment:

1.312: Which probably also explains the patterns in the migration front?!

Response:

Yes that is the reason.

Comment:

1.314: When FFT when FFTM?

Response

It should always be FFTM. Thanks for spotting this.

Comment:

1.318-319: How do the simulations compare to a simulation without communication between grid cells, i.e. 1D simulations?

Response:

See above the differences to the version with communication are negligible.

Comment:

1.323: How to specify this parameter when not having a FFTM simulation at hand?

Response:

Here the aim was to parametrize the SMSM in a way that we have a similar migration speed compared to FFTM. In a practical application one would have a certain dispersal kernel and the derivation in Supplement 2 would allow to estimate the parameter to fit a Gaussian kernel. It is also possible to transform the Gaussian kernel to any other shape by adding several Gaussian kernel. If we would have wanted to do this we would have increased the calculation time for the SMSM. Hence we opted for a more practical approach to get comparable results with the two and still keep the kernel and parameterization from TREEMIG.

Comment:

How for SMSM with terrain, does this require a simulation without terrain before?

Response:

Well as stated before, one can mimic the function used in the FFTM or one can use a Gaussian dispersal function to start with and calculate the parameter for the SMSM from the distribution. However in our case we wanted to be comparable to TREEMIG, so we choose their function and parameterization. And to avoid to use several Gaussian to approximate the function used in TREEMIG we simply tested in an homogenous area. We are now mentioning this in the text on in chapter 3.3.

Comment:

What are the cost reductions then?

Response:

When the final dispersal kernel is approximated by stacking several Gaussian dispersal kernel the SMSM has to be performed several times.

Comment:

1.335-337: I would find it valuable to have the simulation times for the terrain simulation in the table, too!

Response:

All SMSM calculations are with terrain, though the terrain is a homogenous grid or 1s. In the Matlab script we have differentiated between simulation of SMSM with terrain (one extra multiplication) and without, however since we are not planning to use any SMSM without terrain and since the LPJ-GM code always does a terrain, we decided to remove the SMSM without terrain from the Matlab script.

Comment:

1.341: This would probably not work with transects?!

Response:

Yes, the transects have to be chosen in a way that they are not disrupted by barriers that are larger than the dispersal kernel. The main idea behind using the transects is to use them only in heterogeneous areas where you would simulate the whole area. However some parts of the typically squared domain might be homogenous so one might choose to use transects there as well.

Comment:

1.358: K is the number of iterations?!

Response:

Yes thanks for spotting that we did not explain this. We now added it to the text at line ...

Comment:

1.361: “a very similar migration pattern” I would delete the “very”

Response: Done.

Comment:

1.362: it is slower by 20%!

Response:

Yes but given the differences in the literature of migration speed within and between measured and simulated migration speed as well as the uncertainty in the parameters of the seed dispersal kernel this is still relatively similar.

Comment:

1.362-363: in l. 310-312 the authors state that its slower because of the migration path?

How do these two different explanations contribute?

Response:

The stochasticity leads to an increase in migration speed if there are surrounding cells right and left that can contribute via diagonal seed exchange to the cells along the transect. We are currently testing the effect of transects being wider than a single cell, but the results of this would make the ms more complex and we will present them in the next application.

Comment:

1.364: how to parameterize “explicit considerations of wind directions”

Response:

Basically one could calculate different Gaussian distributions in different directions and according to the wind distribution in one area using the considerations in Supplement 2.

Comment:

1.376: Something is missing in this sentence

Response:

We added an ‘or the other’ to make it clearer

Comment:

1.379: Maybe in a DVM? But not in ecology and not to simulate dispersal; the authors should mention some applications - as mentioned above: have a look at Powell (2001) + other references such as Pueyo et al. 2008; Prasad et al. 2013 + I imagine there are much more.

Response:

Yes we agree, we meant DGVMs, this is certainly misleading and we are now relating to some other applications in the introduction.

Comment

1.384: “DGVMSs”

Response: Thanks for spotting this.

Comment:

1.410-412: 63-85% instead if 85%? I would maybe remove this quantitative comparison. In my understanding the size of the reduction will be dependent on the model and the set-up of the simulations, i.e. on a variety of factors, such as the number of simulated and dispersing species, the resolution, settings of the applied algorithms, etc., and since its two different models and probably very different simulation set-ups, it seems to me to be comparing apples and oranges?

Response:

Yes we certainly agree that there are a variety of factors influencing this and therefore it might be more suitable to not quantify it here. Instead we write that our method leads to a reduction in a similar range depending on the configuration of the corridors (Line 484).

Comment:

1.413: more pronounced than what or where?

Response:

Thanks for spotting this, the sentence that this was referring to was lost in an internal revision. We are now writing a complete statement at line 486.

Comment:

1.416: maybe 0.5 and 1.0?

Response:

No here we actually mean 0.1 There are some applications at 1 degree and even 2.5 degree, but when vegetation or even species are in the focus the finer scales are more common.

Comment:

1.449: From the last Figure in the supplementary it seems that SMSM simulations with terrain are comparably much slower. Is it possible to speed them up with transects?

Response:

The idea is to have only those parts where the area is very complex or in which we are actually able to define different seed dispersal kernel to be used with the SMSM, all other areas should use the FFTM. The SMSM did also speed up if used with corridors. The values in table 1 are actually calculated with the extra one multiplication required for the SMSM with terrain. We therefore decided to remove the SMSM-without-terrain from the figure in Supp.2.

Comment:

1.462: I would not call 20% slightly

Response:

With respect to the uncertainty both in the parameters available for the seed dispersal kernel as well as the estimates of migration speed in the literature from pollen analysis 20% is still a low uncertainty. However we are removing the word slightly.

Comment:

1.457: “FTTM” -> “FFTMM”

Response:

Thanks for spotting this.

1.465: unfortunate – I think this would be really interesting, especially when simulating

fragmented landscape or non-dominant species

Response:

We absolutely agree and we are already performing test simulations for a further study.

Comment:

481: missing “)

Response

Thanks for spotting this.

Comment:

1.486: What was the tested set-up? I assume FFTM? I.e. no ‘terrain’? Transects with 50km distance? How many competing/migrating species? All grid cells homogeneous? How many years?

Response:

We tested using the FFTM (hence without terrain) using 4000 by 4000 grid cells, running for a few years only.

*Looking at the numbers in table 1 shows that running a full scale simulation with 21000 years over the 3463 ½ degree cells that we typically use for European runs would take a long time: 1800 (CPUh per 100000 cells and 3000 years) /100000(cells in the MS)*3463(half degree cells in Europe) *50*50(rough estimation of how many 1km cells are in a half degree cell) *21000 years in LGM simulation / 3000 years in testsimulation gives us roughly 10 mill CPUh. Given that my current account allows me 45000 CPUh a month that is currently not feasible and that is why we suggest the transect method. (Actually there might be even more time needed given that there are more than 2 species in the final runs).*

Comment:

1.488: “considerable computation costs” – what were they in the tests (e.g. CPU h per simulated y)? Are continental applications possible, or are they not possible?

Response:

See above. If the corridors are clever placed yes they are possible and if a more efficient parallelization of the FFT is implemented. In this ms we are not providing a proof for this (we will do in the next where we aim to perform a European simulation). Therefore we do not refer to continental runs anymore. The statement at this point is meant to say that from a memory requirement there is no problem performing the FFTM over large areas.

Comment:

1.488: plural and singular mix: “a high amounts” + what does “of the FFT as the local simulations” mean?

Response:

amounts : see calculations above.

FFT as local simulation: Currently only one node is taking care of the calculation of the FFT.

One could theoretically perform the FFT at each node and use one master node only for collecting the amount of dispersed seeds and performing the communication. Hence there is still some untapped optimization potential.

Comment:

1.498: what do you mean with “truly mechanistic”? I recommend deleting this statement

Response:

We meant that the migration rates are a result of the dispersal kernel and establishment in a mechanistic way. We agree that the term might be misleading and have removed it.

Comment:

1.503: “related estimates the Conclusion section”?

Response:

Thanks for spotting that there is an ‘in’ missing.

Comment

1.613: Again “10” + something with the formatting

Response:

Thanks for spotting this. It is changed.

Comment:

1.627: Please provide a legend – even if the figure is only schematic

Response:

We added a legend.

Comment:

1.635: When looking at the Figure and reading 2.4 I wondered where the 5×10^{-7} came from and how this parameter is determined? – Finally I found some information in 3.3

Response:

See responses above. It is a fitted parameter to make the two methods result in comparable migration speed.

Comment:

Fig.3 and 4: I would appreciate if the y-axis of the distance plots on the right would have similar scales, this would really help for comparison

Response:

We changed the axes.

Comment:

1.647-648: difficult sentence – maybe: “only taken into account for grid cells ...”?

Response:

Thanks for the suggestion, we reformulated.

Comment:

1.661: Comparing the dark blue spots in Fig . 2 and the white ones in Fig. 5 the Figures seem to be mirrored along the diagonal?

Response:

Thanks for spotting this. We checked now and yes it was mirrored.

Comment : .664: $\text{cpu} \times \text{h} = \text{CPU h}$?

Response:

Changed.

Comment:

1.671: FFTM with 10: shouldn't this be 64% instead of 67%?

Response:

Thanks for spotting this typo.

References

Brokaw, N. V. (1998). *Cecropia schreberiana* in the Luquillo mountains of Puerto Rico. *The Botanical Review*, 64(2), 91-120.

Fisher, R., McDowell, N., Purves, D., Moorcroft, P., Sitch, S., Cox, P., Huntingford, C., Meir, P., and Woodward, F. I. (2010). Assessing uncertainties in a second-generation dynamic vegetation model caused by ecological scale limitations, *New Phytol.*, 187, 666–681.

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. (2012). Projecting the future distribution of European potential natural vegetation zones with a generalized, tree speciesâ

^

^
Rbased dynamic vegetation model. *Global Ecology and Biogeography*, 21: 50-63.

Kruse, S., Gerdes, A., Kath, N. J., and Herzschuh, U. (2018). Implementing spatially explicit seed and pollen dispersal in the individual-based larch simulation model: LAVESI-WIND 1.0, *Geosci. Model Dev.* 11, 4451-4467, <https://doi.org/10.5194/gmd-11-4451-2018>

Neupane, R. C. (2015). Modeling Seed Dispersal and Population Migration Given a Distribution of Seed Handling Times and Variable Dispersal Motility: Case Study for Pinyon and Juniper in Utah.

Owens, J. N. (1995). Constraints to seed production: temperate and tropical forest trees. *Tree Physiology*, 15(7-8), 477-484.

Powell, James (2001). *Spatio-temporal models in ecology; an introduction to integro-difference equations*. Utah State University, Logan (UT).

Prasad, Anantha M., et al. (2013). Exploring tree species colonization potentials using a spatially explicit simulation model: implications for four oaks under climate change. *Global Change Biology* 19.7, 2196-2208.

Pueyo, Y., et al. (2008). Dispersal strategies and spatial organization of vegetation in arid ecosystems. *Oikos* 117.10, 1522-1532.

Snell, R. S. (2014), Simulating long-distance seed dispersal in a dynamic vegetation model, *Glob. Ecol. Biogeogr.*, 23(1), 89–98, doi:10.1111/geb.12106.

Snell, R. S., Huth, A., Nabel, J. E. M. S., Bocedi, G., Travis, J. M. J., Gravel, D., Bugmann, H., Gutiérrez, A. G., Hickler, T., Higgins, S. I., Reineking, B., Scherstjanoi, M., Zurbriggen, N. and Lischke, H. (2014), Using dynamic vegetation models to simulate plant range shifts, *Ecography (Cop.)*, 37(12), 1184–1197, doi:10.1111/ecog.00580.

Interactive comment on *Geosci. Model Dev. Discuss.*, <https://doi.org/10.5194/gmd-2018-161>, 2018.

Anonymous Referee #3

Received and published: 25 September 2018

Lehsten et al. present a nice and timely study focusing on the implementation of migration into dynamic global vegetation models. They show a way to connect established assumptions of seed dispersal based on former studies with two approaches of enhanced seed dispersal based on Fast Fourier transformation and seed matrix shifting. Together with allowing seed dispersal through specified spatially equidistant corridors they show a nice way of how to reduce computation time while losing some accuracy even though there is no real validation presented. The approach has the potential to be applied in the different DGVMs existing today and is an important contribution to their development. My comments mainly concern 1) the reproducibility of their method and 2) the realism of being able to conduct continental scale simulations.

Response:

Thank you for the effort put in the review of our paper. We will respond and react to your comments as listed below. .

Comment:

Computation time: My biggest concern here is that the authors simulated only one species migrating. Therefore, they were able to simulate only one growing patch per 1 km² grid cell. When increasing the species number it is definitively important to also increase the number of growing patches (and probably decrease their size to enable local competition or in other words to avoid an unrealistic growing patch overarching competition which is one of the major selling points of a gap model like in LPJ GUESS typically using 100-1000 m² growing patches) in return increasing the computation time.

Response:

Actually the whole concept of patches as independent replicates of the vegetation succession at one location is somewhat problematic with our simulation set up.

For patches to be independent, there should be no interaction between them. However this is exactly what we want when simulating migration. It seems that the reviewer assumes that we increased the patch size. We did not. The patch size is kept at the standard level of LPJ-GUESS 4.0 of 1000 m². Similar to all other applications however, this patch represents a larger area. In most applications this patch represents an area as large as the climate grid cell (typically 1 (0.5) degree lon/lat). In our simulation the single patch per gridcell represents an area of 1km² which is only important for the dispersal kernel.

The patches in LPJ-GUESS are mainly introduced to take into account the stochastic heterogeneity of vegetation, that means to decouple the small scale successional stages, to be sure that for example an LAI value for a cell is not too low, simply because the patch is just in an early stage of succession, Hence an averaging of many cells (which have a random disturbance event to restart the succession) will assure that this will not happen. In our case we have only one patch per grid cell, however, in a larger scale simulation one would of course average the output over 0.5 degree and in this case over 200 grid cells with a single patch and hence all successional stages should be present in this 0.5 degree with the correct proportion similar to an application in which a large number of patches would be applied.

Comment:

The computation time for a continental scale simulation with many different species still has to be determined and could be topic of a follow up study.

Response:

This point has been raised by the other reviewers as well. To run a truly continental simulation with many species is not yet feasible with the presented method - at the current stage of optimisation. However, the main point of the paper was to present the two methods and not to spend a lot of time on optimisation methods. The main time in large scale simulations will be needed not by the simulation of the seed dispersal but by the simulation of the vegetation dynamics given that many cells will have to be calculated. We expect to be able to present continental simulations over the Holocene soon, but have currently no proof for this. Therefore we refrain from mentioning continental scale simulations and refer to 'larger areas'. Basically our tests indicate that the corridors can be placed rather far from each other over the largest areas of Europe, which would allow the continental scale simulations that we refer to. But again since we are not providing data on this we will not mention it in the paper.

Comment:

Germination rate: It is unclear how sensitive the presented results are in connection to the germination rate used. Obviously the germination rate must influence the speed of migration. The rate of germination directly influences the competitiveness of each species and therefore its dispersal.

Response:

We fully agree with this comment. Here we are using the germination rate used in TREEMIG and compared our results with the results in TREEMIG. In a 'real world' study this should be looked at in detail while here we only want to present the method. We do now mention that this in the Discussion where we changed the title of the section : 'Parameterisation of dispersal kernels ' to 'Parameterisation of dispersal kernels and other plant parameters'. And we discuss this in Line 534.

Comment:

Age of maturity: Even though I am totally ok with not taking into account an age of maturity to keep the findings of this study as simple as possible, it is again very obvious that this variable strongly influences the speed of migration. Therefore, this topic needs an extra space in the discussion or some results in the supplement showing e.g. the influence of assuming a minimum age of maturity.

Response:

We agree, that the time a species needs to start reproduction, i.e. generation time, is one of the most important factors influencing migration speed. Instead of fixed age of maturity we use a height threshold for maturity, which makes generation time dependent on growth and thus on environment and competition. We emphasize it in line 568.

Comments:

Line 132-134: How does the seedbank determine establishment probability and how is environmentally-suitable determined? I would like at least a brief explanation of this crucial aspect.

Response:

While we mention the environmental suitability here, and the probability to establish, we explain the establishment probability in detail in Chapter 2.3.4 Germination.. With respect to the environmental suitability, this is assessed in LPJ-GUESS (and we did not change this part) by using environmental envelopes of some climatological parameters, e.g. minimum temperatures to survive or establish. We now added an inset mentioning this and referring to the LPJ-GUESS publication at line 156- 158.

Comment:

Line 136 – 150:

So what I see in figure 3 and 4 is that for the 50km corridor approach you have 6 corridors per 0.5° grid cell (or do neighboring cells share corridors)? And these corridors need 200 simulation cells each of them 1km² in size? Assuming that a 0.5° grid cell is 50x50km I wonder where the positions of your corridors actually are. At the borders of each grid cell and also diagonal through the middle? It would be helpful to this in Fig. 3 and 4.

Response: We have inserted a figure (Fig. 3) of the sequence of local dynamics on the corridors, interpolation and dispersal on the grid, where we highlight the corridors (line ...) However there seems to be another misconception. We are simulating an area of 100 by 100 km in all simulations. Each time each cell has a size of one by one km. So the cells are completely adjacent. In the simulations with the corridors we are selecting cells (all outside the corridor) for which we do not simulate the vegetation dynamics, but before simulating seed dispersal we interpolate the seed production from neighbouring cells to all cells. Hence in the step of simulating the seed dispersal we have again 100 by 100 cells (adjacent to each other) that produce seeds (though some just have the seed production value from the nearest neighbour). So to come back to your first question, the 100 by 100 km would be roughly a 1 by 1 degree grid cell (in our simulations we need to run LPJ on equal area cells rather than lon lat as usual) and for the 50 km corridors you would have 2 East West, 2 North South, 3 NorthWest – SouthEast and 3 NorthEast –Southwest corridors (lines of cells at which the full vegetation dynamic is calculated).

Comment:

b) You are able to use only 1 patch per cell, because you are only simulating 1 species migrating. It is important to explain in the discussion that you definitely need more (and probably smaller) patches if you consider more species. It actually scales with species number. Therefore, computation time would be much higher as well. This is contradicting potential continental simulations.

Response:

Testruns have shown that we can have multiple species migrating with a single patch. We cannot see the logic why a single patch can only have a single species migrating, and why the required number of patch repetition scales with species number. In fact the calculation in LPJ-GM simulates the seed dispersal for both simulated species (which probably slows the simulations somewhat), though the birch has free establishment.

We agree that more species will require more simulation time. However since this is using LPJ-GUESS and not LPJ it does not scale as strongly with species number since each species can have a multitude of species age cohorts and computation time scales with age cohorts. In a crowded cell there will be less cohorts per species but in general the statement that more species require more computation time holds true. We therefore removed all references to continental simulations in this paper. We hope to present a continental simulation soon in a different paper.

Comment:

Line 155-157: Even though you cite Lischke et al. I would like to see a brief explanation of the “maximum fecundity” method.

Response:

It is not a method but a single value for maximum fecundity. Basically there is a maximum number of seeds that can be produced per tree of a species for which we have an estimate. This one we multiply with the current LAI divided by the maximum LAI for which we also have a value listed. We now give an example to make this clearer at line 181.

Comment:

Line 157: Have you performed tests for age of maturity? I guess setting an age of maturity would lower the speed of migration. I am totally ok with not taking this into account, but it would be good to pick up this issue in the discussion e.g. under 4.4.3.

Response:

As mentioned before we are using height of maturity (see chapters 2.3.1) and now discuss this in the chapter 4.4.3 Parameterisation of dispersal kernels and other plant parameters.

Comment:

Line 163-164: Please provide explanation and reference for mast fruiting effects.

Response:

We provided an explanation on line 190

Comment:

Line 188 – 189: So do you use the values for *Fagus sylvatica*?

Response:

*Since the only species that effectively migrates in this paper is *Fagus*, we do use these values.*

Comment:

Line 191 – 193: Where can I find values of “loss of germinability”? If these are specific values from Lischke et al. I would suggest to list them in a table as well as similar parameter values. This would really help to reproduce the study.

Response:

We agree with this and have added one extra table (in the supplementary material S.4) listing all needed parameters needed to reproduce the study.

Comment:

Line 194-198: I have my problems understanding this whole part. 1. “A year is defined for each species and grid cell before which seed bank constraints are ignored”. I do not understand this sentence. 2. I also do not understand the second sentence. I believe you talk about the initial conditions and refugia. It is probably a very crucial part for migration simulations so please provide a few more sentences of explanation.

Response:

We rephrased the sentences to this (and hope it is now clearer):

For each grid cell and each year we prescribe whether the species requires seeds to establish. By not requiring seeds in some cells for establishment or not requiring seeds for establishment for some species for all cells we define refugia or in the latter we define that the species' seeds are known to be very far dispersed and hence no explicit simulation of establishment by seeds is required for this species. Technically this is implemented by reading in a list for each cell containing a year from which onwards a species' establishment is not limited by the availability of seeds. Explained in lines 225ff.

Comment:

Line 207: Explain “age cohort”. It has not appeared before and is important to understand the approach.

Response:

The term age cohort comes from the general principles of LPJ-GUESS. Since here we only talk about the establishment of a species we decided to remove the term as it might only confuse the reader if we add a 3 sentence explanation.

Comment:

Line 199-208:

Please provide equations and an according explanation instead of an example in

Response:

We have added an equation (eq. 5).

Comment:

b) Moreover, I am quite sure that the germination rate strongly influences results. It is probably important for the speed of migration and definitively for competition and therefore equilibrium

biomass. What do you mean with “initial testing”? I don’t expect a comprehensive full explanation in the text, but I would like to understand why you have chosen certain parameter values.

Response:

See above, we added a section in the chapter:

4.4.3 Parameterisation of dispersal kernels ,and other plant parameters

Comment:

Line 302 – 303: As suggested above. Please provide the numbers of your parametrization.

Response:

All parameters are listed in an extra table in the supplementary material S4.

Comment:

Line 442 -443: Is it possible to give a comparison here? What is the computation time for the same setup with standard LPJ-GUESS? I see your comparisons in table 1, but they all refer to simulations which use a master.

Response: We have added a simulation without the communication and the difference to the simulation with communication is negligible. We mention this in the paper.

Comment:

Line 488 – 490: Have you estimated the CPU hours for this setup? Would be an interesting information.

Response:

Yes we did. We tested using the FFTM (hence without terrain) using 4000 by 4000 grid cells, running for a few years only.

Looking at the numbers in table 1 shows that running a full scale simulation with 21000 years over the 3463 ½ degree cells that we typically use for European runs would take a long time:

*1800 (CPUh per 100000 cells and 3000 years) /100000(cells in the MS)*3463(half degree cells in Europe) *50*50(rough estimation of how many 1km cells are in a half degee cell)) *21000 years in LGM simulation / 3000 years in a test simulation gives us roughly 10 mill CPUh. Given that my current account allows me 45000 CPUh a month, this is currently not feasible and that is why we suggest the transect method. (Actually there might be even more time needed given that there are more than 2 species in the final runs).*

Again we see quite some opportunities for optimization unused and will hopefully soon provide a continental simulation.

Technical comments:

Line 22-24: Indicate that this sentence is about plants in the real world

Response:

We now start the sentence with : Pollen studies have shown that indicating that we are not talking about simulated species anymore.

Comment:

Line 62-64: For me this is one of the major selling points and I would put it in the abstract as well. You decide.

Response:

Thanks for the suggestion. We would love to do this and in fact we just started a project where we look at the spread of alien species, however what we present here is only considering tree migration. For trees the spread by adjusting to climate change is very slow given the long generation time. We will hopefully soon have a model in which we simulate tree pests spread, there

this applies, but in this publication we only consider trees hence the response to current climate change is rather of second order.

Comment:

Line 202: 2 times the word “depending”

Response:

Thanks for spotting this. We removed it.

Comment:

Line 203 -205: Confusing sentence. “The probability that a species establishes is proportional to the seed number in the seed bank multiplied by . . .”. Wrong formulation.

Response:

We now added an equation instead of this sentence.

Comment:

Line 206: The word “year” is missing.

Response:

No, the word ‘per’ was obsolete. Thanks for spotting this. We removed it.

Comment:

Line 242: I would not expect that every reader knows what a Moore neighborhood is?

Response:

We now added” i.e. the surrounding eight cells”.

1
2 **LPJ-GM 1.0: Simulating migration**
3 **efficiently in a dynamic vegetation**
4 **model models efficiently using LPJ-**
5 **GM**

6
7
8
9 Lehsten, Veiko^{a,b}, Mischurow, Michael^b, Lindström, Erik^c, Lehsten, Dörte^b, Lischke, Heike^a

10
11 ^a Dynamic Macroecology/Land change science-Landscape dynamics, Swiss Federal Institute for
12 Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland

13 ^b Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden

14 ^c Centre for Mathematical Sciences, Department for Mathematics Lund University, Lund, Sweden

15
16
17 * Corresponding author: veiko.lehsten@gmail.com

18

19 Abstract

20 Dynamic global vegetation models are a common tool to assess the effect of climate and land use
21 change on vegetation. Though most applications of dynamic global vegetation models use plant
22 functional types, some also simulate species occurrences. While the current development aims to
23 include more processes, e.g. the nitrogen cycle, the models still typically assume an ample seed supply
24 allowing all species to establish once the climate conditions are suitable. Pollen studies have shown
25 that a number of plant species ~~have been shown to~~ lag behind in occupying climatological suitable
26 areas (e.g. after a change in the climate) as they need to arrive at and establish ~~in~~ the newly suitable
27 areas. Previous attempts to implement migration in dynamic vegetation models have allowed
28 simulating either only small areas or have been implemented as post process, not allowing for
29 feedbacks within the vegetation. Here we present two novel methods simulating migrating and
30 interacting tree species which have the potential to be used for ~~continental~~ simulations of large areas.
31 Both distribute seeds between grid cells leading to individual establishment. The first method uses an
32 approach based on Fast Fourier Ttransforms while in the second approach we iteratively shift the seed
33 production matrix and disperse seeds with a given probability. While the former method is
34 computationally ~~marginally~~ faster, it does not allow for modification of the seed dispersal kernel
35 parameters with respect to terrain features, which the latter method allows.

36 We evaluate the increase in computational demand of both methods. Since dispersal acts at a scale no
37 larger than 1 km, all dispersal simulations need to be performed at ~~maximum~~ least at that scale.
38 However, with the current available computational power it is not feasible to simulate the local
39 vegetation dynamics of a ~~large area~~ whole continent at that scale. We present an option to decrease the
40 required computational costs, reducing the number of grid cells where the local dynamics is
41 ~~simulated~~ computed by simulating it only along migration transects. Evaluation of species patterns and
42 migration speeds shows that ~~although~~ the simulation along transects reduces the migration speed, and
43 ~~slightly,~~ both methods applied, on the transects, produce reasonable results, are reliable. Furthermore,
44 using the migration transects, both methods are sufficiently computationally efficient to allow large
45 scale DGVM simulations with migration, ~~on entire continents.~~

46 1. Introduction

47 A large suite of dynamic global vegetation models (DGVMs) is currently used to simulate the effects
48 of climate and / or land use change on vegetation and ecosystem properties. These simulations result
49 in projections (or hind-casts) of species ranges as well as changes in ecosystem properties such as
50 carbon stocks and fluxes. Examples of these DGVMs include ORCHIDEE (Yue et al., 2018)(~~Yue et~~
51 ~~al., 2018~~), LPJ-GUESS (Sitch et al., 2003)(~~Sitch et al., 2003~~), IBIS (Foley et al., 1998)(~~Foley et al.,~~
52 ~~1998~~), (Sato et al., 2007)(~~Sato et al., 2007~~), for a review of DGVM features see (Quillet et al.,
53 2010)(~~Quillet et al., 2010~~).

54 While most DGVM applications use plant functional types (groups of plant species with similar traits
55 and responses to environmental conditions), here we only consider applications which explicitly
56 simulate tree species, e.g. (Hickler et al., 2012). These models typically assume that species can
57 establish at any site once the environmental conditions become suitable. However, in real ecosystems
58 species need not only to establish and replace existing vegetation – the processes gap models describe
59 successfully – but they also need to have a sufficient amount of seeds at a given location to
60 successfully establish. Implicitly, current DGVMs assume that ample amounts of seeds of all species
61 are present in every location.

62 While this approach might seem reasonable in cases where the vegetation can keep up with climate
63 change (i.e. moving sufficiently fast to occupy areas which become suitable), there have been a
64 number of instances reported where a considerable migration lag occurred. For instance *Fagus*
65 *sylvatica* has been shown to have a considerable migration lag and is currently still in the process of
66 occupying its climatological optimum (Bradshaw and Lindbladh, 2005).

67 ~~Not only for the simulation of historical species ranges is the implementation~~ ~~These models typically~~
68 ~~assume that species can establish at any site once the environmental conditions become suitable.~~
69 ~~However, in real ecosystems species need not only to establish and replace existing vegetation – the~~
70 ~~processes gap models describe successfully – but they also need to have sufficient amount of seeds at~~
71 ~~a given location to successfully establish. Implicitly, current DGVMs assume that ample amounts of~~
72 ~~seeds of all species are present in every location.~~

73 ~~While this approach might seem reasonable in cases where the vegetation can keep up with climate~~
74 ~~change (i.e. moving sufficiently fast to occupy areas which become suitable), there have been a~~
75 ~~number of instances reported where a considerable migration lag occurred. For instance *Fagus*~~
76 ~~*sylvatica* has been shown to have a considerable migration lag and is currently still in the process of~~
77 ~~occupying its climatological optimum (Bradshaw and Lindbladh, 2005).~~

78 ~~Inclusion~~ of migration ~~into dynamic vegetation models is not only~~ of interest. ~~Also for to simulate~~
79 ~~species migration in the past. For~~ the projection of ecosystem properties in the future (with projected
80 climate), migration lags might lead to uncertainties in projected ecosystem properties if the wrong
81 species community is predicted to occur at a certain site ~~(Neilson et al., 2005).~~ Especially, given that
82 the speed at which environmental conditions change currently is unprecedented at least over the last
83 centuries, effects of the migration lag of key species should be evaluated when projecting ecosystem
84 properties. ~~This holds in particular for projections over several centuries. For periods of less than 50-~~
85 ~~100 years ahead, which corresponds to at most a few generations of most tree species, the explicit~~
86 ~~modelling of seed dispersal might be less important for simulating tree distributions, in particular~~
87 ~~when taking into account the overwhelming influence of human activities.~~

88 Migration lags can be caused by different factors. Seed transport might only occur over limited
89 distances. But also low seed amounts and in particular long generation times can slow down
90 migration. Seed amount and generation time depend on the competition with other trees: a free
91 standing tree starts earlier to produce seeds and produces more than a tree of the same age in ~~at~~
92 closed forest. The competitors, however, are also migrating, which leads to feedbacks between the
93 species.

94 Thus, for simulations over large areas covering long time spans, species migration – consisting of a)
95 local dynamics influenced by the environment, b) competition between species, and c) ~~of~~ seed
96 dispersal – has to be taken into account simultaneously for several species.

97 Species migration has been implemented successfully in dynamic vegetation models working on
98 smaller extents and finer scales than DGVMs typically use, e.g. forest landscape models (FLMs;
99 review in Shifley et al, 2017), such as TreeMig, (Lischke et al., 2006), Landclim (Schumacher et al.,
100 2004), Landis (Mladenoff, 2004), or Iland (Seidl et al., 2012) or spatially explicit individual based
101 models such as LAVESI (Kruse et al., 2018).

102 ~~Species migration has been implemented successfully in dynamic vegetation models working on~~
103 ~~smaller extents and finer scales than DGVMs typically use, e.g. forest landscape models (FLMs;~~
104 ~~review in Shifley et al, 2017), such as TreeMig, (Lischke et al., 2006), Landclim (Schumacher et al.,~~
105 ~~2004), Landis (Mladenoff, 2004), or Iland (Seidl et al., 2012).~~

106 In these models, seed dispersal is modelled in a straightforward way: seeds are distributed from each
107 producing to each receiving cell with a distance dependent probability. However, transferring these
108 approaches to DGVMs is problematic, due to a number of conceptual and technical difficulties.
109 DGVMs usually operate on a coarse spatial resolution to reduce computational load and input data
110 requirements. This neglects the spatial heterogeneity within the grid cells. Additionally, and even more
111 critical for implementing migration, it leads to discretization errors: if it is assumed that the forest
112 representing the grid cell is located in the centre of the cell, the seeds cannot move far enough to leave
113 the cell (given a typical cell size of 50km by 50km or 10km by 10km). If it is assumed in contrast that
114 the simulated forest is uniformly distributed in the cell, with each time step some ~~of the~~ seeds reach
115 the neighbour cell, leading to a resolution dependent speed up of ~~the~~ migration.

116 ~~Also some~~ Additionally specifics of model implementations might complicate the inclusion of
117 migration in some DGVMs. Many DGVM implementations are done in a way that for each grid cell
118 all years are simulated before the simulation of the next cell is started. This is done to minimize input-
119 output effort since the whole climate data for each cell is read in at once and it also eases
120 parallelisation for multi-core computers, since in this case each node is assigned a number of grid cells
121 which the node calculates independently of the other nodes without ~~with no required~~ communication.

122 However, for simulating seed dispersal, all cells need to be annually evaluated. Additionally to the
123 reasons mentioned before, most DGVM applications use plant functional types which comprise
124 typically species with very different traits with respect to migration (e.g. dispersal vectors or seed
125 properties). Hence introducing migration would require to split up PFTs into smaller groups and to
126 parameterise the additional properties in each time step.

127 There have been a number of attempts to integrate species migration in DGVMs (cf. Snell *et al.*, 2014,
128 and Discussion section). For example, Sato and Ise (2012) developed a DGVM where species could
129 potentially migrate between neighbouring cells with a fixed rate of about 1km/year while Snell *et al.*
130 (2014) simulated migration as an infection process ~~Snell *et al.*, 2014, and Discussion section~~). For
131 ~~example, Sato and Ise (2012) developed a DGVM where species could potentially migrate between~~
132 ~~neighbouring cells with a fixed rate of about 1km/year.~~

133 However, to the knowledge of the authors, there is no implementation into a DGVM which allows
134 simulations with a large extent, which ~~at continental scale~~, takes into account the migration within the
135 grid cell and includes feedbacks between all simulated species.

136 Here we present two methods to fill this gap, i.e. allow simulating species migration of several species
137 simultaneously. The methods are implemented into the LPJ-GUESS DGVM but can potentially also
138 be implemented into other DGVMs. Though they are tested here using a virtual landscape, they can be
139 applied for ~~continental~~ simulations of large areas given current computing resources.

140 **2. Methods**

141 **2.1 The dynamic vegetation model LPJ-GUESS**

142 LPJ-GUESS is a flexible framework for modelling the dynamics of terrestrial ecosystems from
143 landscape to global scales (Sitch *et al.*, 2003; Smith *et al.*, 2001). This DGVM consists of a number of
144 sub-modules containing formulations of subsets of ecosystem processes at defined spatial and
145 temporal scales. Similar to most other DGVMs, it requires time series of climate data (precipitation,
146 air temperature and shortwave radiation), soil conditions and carbon dioxide concentrations as input
147 and explicitly simulates vegetation cover. While it uses plant functional types in most applications,
148 some applications simulate tree species (e.g. Hickler *et al.*, 2012; Lehsten *et al.*, 2015). LPJ-GUESS
149 explicitly simulates canopy conductance, photosynthesis, phenology, and carbon allocation. It uses a
150 detailed individual-based representation of forest stand structure and dynamics. Each species (or PFT)
151 has a specific growth form, leaf phenology, life history and bioclimatic limits, determining its
152 performance and competitive interactions under the forcing conditions and realized ecosystem state of
153 a particular grid cell (Sitch *et al.*, 2003). A large body of publications describes the features of LPJ-
154 GUESS in detail; here we concentrate on the changes that were applied to LPJ-GUESS version 4.0

155 (Lindeskog et al., 2013; Smith et al., 2014). To differentiate between the original version of LPJ-
156 GUESS and our extended version (where we implemented the migration module) we refer to the
157 extended version as LPJ-GM (short for LPJ-GUESS-MIGRATION).

158 **2.2 Technical implementation**

159 Standard LPJ-GUESS simulations are typically performed at a computing cluster with cells running on
160 different nodes of the cluster without any interaction of the nodes. We implemented a distributed
161 simulation using MPI (Clarke et al., 1994) with the grid cells communicating with a master process.

162 Seeds are produced potentially in each grid cell at the end of each migration year. The number of seeds
163 produced is sent to the node computing the dispersal while all nodes wait for this master LPJ-GUESS
164 is a flexible framework for modelling the dynamics of terrestrial ecosystems from landscape to global
165 scales (Sitch et al., 2003; Smith et al., 2001). This DGVM consists of a number of sub-modules
166 containing formulations of subsets of ecosystem processes at defined spatial and temporal scales. A
167 large body of publications describes the features of LPJ-GUESS in detail; here we concentrate on the
168 changes that were applied to LPJ-GUESS version 4.0. To differentiate between the original version of
169 LPJ-GUESS and our extended version (where we implemented the migration module) we refer to the
170 extended version as LPJ-GM (short for LPJ-GUESS-MIGRATION).

171 **2.2 Technical implementation**

172 Standard LPJ-GUESS simulations are typically performed at a computing cluster with cells running on
173 different nodes of the cluster without any interaction of the nodes. We implemented a distributed
174 simulation using MPI (Clarke et al., 1994) with the grid cells communicating with a master process.

175 Seeds are produced potentially in each grid cell at the end of each migration year. The number of seeds
176 produced is sent to the node computing the dispersal while all nodes wait for this node to finish the
177 calculation. This node sends the number of seeds that arrive at each grid cell back to all nodes to
178 continue the calculation.

179 Similar to the standard version of LPJ-GUESS (Sitch et al., 2003; Smith et al., 2001), in the first 100
180 years no seed dispersal is performed and all species are allowed to establish and grow without seed
181 limitation and without N-limitation to equilibrate the soil pools with carbon and nitrogen. This time
182 period is used to sample NPP given a certain N deposition and climate to subsequently equilibrate the
183 N pools of the soil and a fast spin-up of 40000 years approximated using the sampled rates of C
184 assimilation (Smith et al., 2014). After this initialisation period all vegetation is killed and succession
185 starts from a bare soil and now seed limitation is active.

186 In LPJ-GM seed dispersal is done on an annual basis which corresponds to the temporal resolution of
187 seed production. The amount of seeds produced is communicated to the master node at the end of each
188 year. The master node re-distributes seeds over the whole spatial domain according to the dispersal

189 ~~algorithm and communicates the amounts of arriving seeds back to each grid cell. Similar to the~~
190 ~~standard version of LPJ-GUESS (Sitch et al., 2003; Smith et al., 2001), in the first 100 years no seed~~
191 ~~dispersal is performed and all plant types are allowed to establish and grow without N limitation to~~
192 ~~equilibrate the soil pools with carbon and nitrogen. This time period is used to sample NPP given a~~
193 ~~certain N deposition and climate to subsequently equilibrate the N pools of the soil and a fast spin-up~~
194 ~~of 40000 years approximated using the sampled rates of C assimilation (Smith et al., 2014). After this~~
195 ~~initialisation period all vegetation is killed and succession starts from a bare soil and now seed~~
196 ~~limitation is active.~~

197 ~~In LPJ-GM seed dispersal is done on an annual basis which corresponds to the temporal resolution of~~
198 ~~seed production. The amount of seeds produced is communicated to the master node at the end of each~~
199 ~~year. The master node re-distributes seeds over the whole spatial domain according to the dispersal~~
200 ~~algorithm and communicates the amounts of arriving seeds back to each grid cell. Seeds transferred to~~
201 ~~the grid cells are added to the seed bank which determines establishment probability in~~
202 ~~environmentally-suitable cells (environmental suitability is determined by means of environmental~~
203 ~~envelopes, containing amongst others minimum survival and establishment temperatures; see; Smith et~~
204 ~~al. 2001).- All communications between the processes are done via MPI protocol (Clarke et al.,~~
205 ~~1994)(Clarke et al., 1994).~~

206 LPJ-GUESS is a gap model with the typical successional vegetation changes. To even out
207 successional based fluctuations in ecosystem properties and to be able to simulate disturbances most
208 previous applications simulate a certain number of replicate patches are simulated per grid cell. All
209 patches share the same climate but potentially differ in their successional stage due to different timing
210 of disturbances and stochastic mortality. Conceptually, each patch has a size of 1000 m² but represents
211 an area depending on the resolution of the grid cell. Patches have no spatial position with respect to
212 each other and do not interact (Smith et al., 2001). In LPJ-GM we reduced the number of patches to
213 one but achieved the representative averaging by using explicitly placed small grid cells instead of
214 statistical units (replicate patches). For each large grid cell in the climate grid we simulate a large
215 number of cells of 1km² area resulting in a more than sufficient averaging of successional stages. LPJ-
216 GUESS simulations are typically performed with patch numbers around 10 (e.g. Smith *et al.*, 2001)
217 but depending on the aim of the simulation patch numbers have been increased even to 500 (e.g.
218 Lehsten *et al.*, 2016). In our setup even with 50 km corridors (see below and a certain number of
219 replicate patches are simulated per grid cell. All patches share the same climate but potentially differ
220 in their successional stage due to different timing of disturbances and stochastic mortality.
221 Conceptually, each patch has a size of 1000 m² but represents an area depending on the resolution of
222 the grid cell. Patches have no spatial position with respect to each other and do not interact (Smith et
223 al., 2001). In LPJ-GM we reduced the number of patches to one but achieved the representative
224 averaging by using explicitly placed small grid cells instead of statistical units (replicate patches). For

225 ~~each large grid cell in the climate grid we simulate a large number of cells of 1km² area resulting in a~~
226 ~~more than sufficient averaging of successional stages. LPJ-GUESS simulations are typically~~
227 ~~performed with patch numbers around 10 (e.g. Smith *et al.*, 2001) but depending on the aim of the~~
228 ~~simulation patch numbers have been increased even to 500 (e.g. Lehsten *et al.*, 2016). In our setup~~
229 ~~even with 50 km corridors (see below) LPJ-GM represents a 0.5x0.5 degree~~ Fig. 3) LPJ-GM represents
230 ~~a 0.5x0.5 degree~~ cell with 200 simulation cells ranging at the higher end of the patch number per area
231 compared to previous simulations.

232 2.3 Migration processes

233 2.3.1 Seed production

234 ~~The seed production starts once the tree reaches maturity height and is scaled linearly with leaf area up~~
235 ~~to maximum LAI.~~

236 ~~Similar to TreeMig (Lischke *et al.*, 2006) we implemented seed production (depending on leaf area~~
237 ~~index; LAI) and seed bank dynamics.~~

238 The seed number produced ~~per tree within each grid cell~~ is calculated as the product of the maximum
239 fecundity multiplied by the proportion of the current LAI to the maximum LAI and multiplied by the
240 area per grid cell (Lischke *et al.*, 2006). For example, the maximum fecundity of beech is 29000, the
241 ~~maximum LAI is 5 m² *m⁻² and the maturity height is 14.4 m. Hence a tree of 15m height is above~~
242 ~~the maturity height, and with an LAI of 2.5 m² *m⁻² it will produce 29000*0.5/5=14500~~
243 ~~seeds. (Lischke *et al.*, 2006).~~ No specific age of maturity is taken into account.

244 All seeds of a species produced $S(x', y')$ at a location (x', y') within a year are available for seed
245 dispersal. Once seeds have entered the seed bank, no further dispersal is possible (they remain in the
246 seed bank). Though LPJ-GUESS keeps track of carbon allocated to the main plant compartments and
247 even allocates a certain amount of carbon to seeds (which is transferred to the litter pool, the soil pool
248 and finally the atmosphere), for simplicity we decided not to relate the seed production to the carbon
249 accounting at this point. Allocation rules including seed production and even mast fruiting effects
250 ~~(synchronised strong increases in seed production e.g. similar to Lischke *et al.* 2006)~~ could be
251 included in the future.

252 2.3.2 Seed dispersal

253 The produced seeds are distributed according to

$$254 S_d(x, y) = \int S(x', y') k_s(x - x', y - y') dx' dy' \quad (\text{eq. 1}).$$

255 $S(x', y')$ is the seed production, and $k_s(x - x', y - y')$ the seed dispersal kernel in euclidean
256 coordinates. The seed distribution $S_d(x, y)$, i.e. the input of seeds in location x, y is then obtained by
257 integrating over all possible locations x', y' for arriving at x, y .

258 | Thus, the seed distribution is given by the convolution (~~**~~) of the seed production and the seed
 259 | dispersal kernel:

260 | $S_d = S ** k_s.$ (eq. 2)

261 |
 262 | For this study we used the seed dispersal kernel and parameterization for *Fagus sylvatica* from
 263 | TreeMig (Lischke *et al.*, 2006). The seed dispersal kernel defines the probability of seeds arriving at a
 264 | sink cell (x,y) from the source cell (x',y') with a certain distance $z = \sqrt{(x - x')^2 + (y - y')^2}$.

265 | The kernel is specified in a polar coordinate system,

266 | $k_s(z, \theta) = k_s(z|\theta)k_s(\theta)$, with the radial distance z. The seeds follow a mixture of two exponential
 267 | distributions, the short and the long term dispersal, while the angular dispersion, θ , is uniform in all
 268 | directions (in our case the angular dispersion θ is uniform, but if one is interested e.g. in implementing
 269 | wind directions this can be changed). Thus, the radial component of the kernel is given by

270 | $k_s(z|\theta) = (1 - \kappa) \frac{1}{\alpha_{s,1}} e^{-\frac{z}{\alpha_{s,1}}} + \kappa \frac{1}{\alpha_{s,2}} e^{-\frac{z}{\alpha_{s,2}}}, \kappa \in (0,1)$ (eq. 3)

271 | while the angular term is given by

272 | $k_s(\theta) = \frac{1}{2\pi}$ for $\theta \in [0,2\pi]$ ~~$(\theta) = \frac{1}{2\pi} 1_{\{\theta \in [0,2\pi]\}}$~~ (eq.
 273 | 4.1)

275 | $k_s(\theta) = 0$ otherwise . (eq. 4.2)

277 | The dispersal kernel is defined by the species specific values for the proportion of long distanceterm
 278 | dispersal κ and the species expected dispersal distances $\alpha_{s,1}$ and $\alpha_{s,2}$ for the two kernels.

279 | The species specific values for these parameters (0.99 for κ_s and 25m and 200m for the two mean
 280 | dispersal distances k_s for *Fagus sylvatica*) were taken from by Lischke *et al.* (2006).

281 | ~~The species specific values for these parameters were roughly estimated by Lischke *et al.* (2006) to be~~
 282 | ~~0.99 for κ_s and 25m and 200m for the two mean dispersal distances k_s for *Fagus sylvatica*.~~

283 | **2.3.3 Seed bank dynamics**

284 | The number of the seeds in the seed bank (i.e. the dormant seeds in the soil that can germinate in
 285 | subsequent years in each cell) is increased by the influx S_d of seeds according to (eq. 1), and reduced

286 by the yearly loss of germinability (caused by decay of seeds) and the amount of germinated seeds at
287 the end of each simulated year, similar to TreeMig (Lischke et al., 2006).

288 For each grid cell and each year we prescribe whether the species requires seeds to establish. By not
289 requiring seeds in some cells for establishment or not requiring seeds for establishment for some
290 species for all cells we define refugia, or in the latter case we define that the species' seeds are known
291 to be very far dispersed and hence no explicit simulation of establishment by seeds is required for this
292 species. Technically this is implemented by reading in a list for each cell containing a year from which
293 onwards a species' establishment is not limited by the availability of seeds.

294 **2.3.4 Germination** The number of the seeds in the seed bank is increased by the influx S_d of seeds
295 according to (eq. 1), and reduced by the yearly loss of germinability and the amount of germinated
296 seeds at the end of each simulated year, similar to TreeMig (Lischke et al., 2006).

297 A year is defined for each species and grid cell before which seed bank constraints are ignored (hence
298 free establishment happens). This way we can specify in which areas each species is already
299 potentially established at the beginning of the simulation and this defines thereby the location of the
300 refugia. This parameter can also indicate that a species is not hindered in its establishment by
301 migration.

302 **2.3.4 Germination**

303 LPJ-GUESS is a gap model and in the original version the number of newly established saplings only
304 depends on the amount of light reaching the forest floor (given that the cell has a suitable climate). In
305 LPG-GM we additionally limit the establishment of seedlings depending stochastically ~~depending~~ on
306 the number of available seeds. Hence the seed limitation is applied before the light limitation. The
307 probability that a species establishes is given in equation 5.

$$308 \pi_{est} = S p_x \pi_{germ} \text{_____} \text{ (eq. 5)}$$

309 Where the π_{est} is the probability of the species establishing, S is ~~proportional to~~ the seed number and
310 π_{germ} is ~~in the seed bank multiplied by~~ the seed germination proportion. ~~The and an~~ extra parameter
311 p_x ~~takes which~~ (implicitly) ~~takes~~ the area of each grid cell into account. In our case we fixed this
312 parameter to 0.01 after initial testing. Hence if in a certain year 100 seeds are in the seed bank ~~per~~ and
313 the germination rate is 0.71 (value for *Fagus sylvatica*) the probability of ~~establishment an age cohort~~
314 establishing is $0.01 * 100 * 0.71 = 0.71$.

315 **2.4 Enhanced dispersal simulation**

316 One way to simulate seed dispersal is to calculate the convolution of ~~all seeds produced of~~ the matrix
317 containing the seed production and the seed dispersal kernel (specified in eq. 1 and eq. 3).- However,

318 evaluating the convolution explicitly can be computationally expensive for seed dispersal kernels with
319 long range.

320 2.4.1 Fast Fourier transformation method (FTM)

321 ~~An alternative is based on the convolution theorem and the Fast Fourier Transformation (FFT), a~~
322 ~~technique commonly used in physics, image processing and engineering (Strang, 1994), but rarely in~~
323 ~~ecology (see e.g. Shaw et al., (2006), Pueyo et al., (2008) or Powell, (2001)).~~

324 ~~An alternative is based on the convolution theorem and the Fast Fourier Transformation (FFT), a~~
325 ~~technique commonly used in physics, image processing and engineering (Strang, 1994).~~

326 This approach carries out the computations in the frequency domain, see Gonzales & Woods
327 ~~(2002).~~(2002). Here we use the notation $F\{S\} = \int e^{-iux-ivy}S(x,y) dx dy$ to denote the two
328 dimensional Fourier transform of S and correspondingly $F\{k_s\}$ the two dimensional Fourier transform
329 of k_s . It then follows that the Fourier transform of the convolution equals the product of the Fourier
330 transforms

$$331 \quad F\{S ** k_s\} = F\{S\}F\{k_s\} \quad (\text{eq. } \underline{65})$$

332 Thus, it is possible to compute the convolution by applying the inverse Fourier transform to the
333 products of the Fourier transforms

$$334 \quad S ** k_s = F^{-1}\{F\{S\}F\{k_s\}\} \quad (\text{eq. } \underline{76})$$

335 ~~This equation must be discretized before evaluating it on a computer. The discrete Fourier transform is~~
336 ~~computed using the Fast Fourier Transform (Cooley and Tukey, 1965), which has a computational~~
337 ~~cost of $O(N^2 \log^2(N))$ in two dimensions. The discrete approximation of S_d is then given by~~

$$338 \quad S_d = F^{-1}\{F\{S\} \odot F\{k_s\}\} \quad (\text{eq. } \underline{8})$$

339 ~~This equation must be discretized~~
340 ~~before evaluating it on a computer. The discrete Fourier transform is computed using the Fast Fourier~~
341 ~~Transform (Cooley and Tukey, 1965), which has a computational cost of $O(N^2 \log^2(N))$ in two~~
342 ~~dimensions. The discrete approximation of S_d is then given by~~

$$342 \quad S_d = F^{-1}\{F\{S\} \odot F\{k_s\}\} \quad (\text{eq. } \underline{7})$$

343 where \odot is the element-wise (Hadamard product) multiplication of matrices.

344 Nowadays, software packages for FFT typically only compute positive frequencies. That means that
345 we have to shift the frequencies prior to the element-wise multiplication of $F\{S\}$ and $F\{k_s\}$. This is
346 illustrated in Fig.1, see also supplementary material S.2.

347

348 <Figure 1 to be placed here>

349 While this method allows including different wind distributions by changing the seed dispersal kernel
350 (as long as they are valid for the whole simulated area), it does not allow to use different seed
351 dispersal kernels at different locations, e.g. due to prevailing wind directions in valleys, due to barriers
352 to animal transport like a motorway, or due to lower transport permeability in already forested areas.

353 **2.4.2 Seed matrix shifting method (SMSM)**

354 Another way to simulate seed dispersal is to simulate the seed movement between the cells explicitly
355 by shifting the matrix containing the produced seeds by one position (repeatedly in all directions of the
356 Moore neighbourhood; i.e. the surrounding eight cells) and simulating seed transport of a certain
357 proportion of the seeds into the next cell. Each move can be viewed as an independent random
358 variable. Repeating these moves thus corresponds to a random walk process. The Lindeberg's
359 condition for sequences for sums of independent random variables ensures that the kernel will be
360 Gaussian under general conditions ~~(Shiryaev, 2016)~~(Shiryaev, 2016), with the expected value given by
361 the sum of expected values for each random variable and similarly for the variance (see supplementary
362 material S.1 for a formal proof and a derivation of the parameters of the resulting normal distribution).

363 If this is done repeatedly it allows an easy implementation of spatial explicit differences in seed
364 dispersal kernel distributions, by adjusting the proportions of seeds being transported into the next cell
365 according to a similarly sized matrix containing the area roughness or permeability. By this approach,
366 barriers and even wind speeds in latitudinal and longitudinal directions can be implemented by
367 adjusting the dispersal probabilities accordingly. ~~After the distribution of the dispersed seeds is~~
368 ~~calculated, the seeds are added to the seed bank.~~ An example calculation of the first three steps of the
369 SMSM (in the final simulation 10 steps are performed) is given in the Supplement S.3.

370

371 <Figure 2 placed here>

372 After the distribution of the dispersed seeds is calculated, the seeds are added to the seed bank.

373 **2.5 Corridors**

374 Seed dispersal acts at a rather fine scale compared to the usual scale at which DGVMs are run (LPJ-
375 GUESS is typically run at a 0.5 to 0.1 degree longitude / latitude scale), though some regional
376 applications use finer grids (e.g. Scherstjanoi et al., 2014). Given that the average long distance seed
377 dispersal for example for *Fagus sylvatica* is 200 m, simulations at such a coarse scale will not be able
378 to capture this process.

379 As a compromise between currently available computing resources and required simulation detail we
380 choose a 1km scale at which we performed our simulations. However, even at this scale, simulating

381 | large areas for example within the European continent would result in a highan-extreme computational
382 | effort.

383 | Given that in some areas the landscape is rather homogenous while other areas have a variable terrain
384 | (or land use conditions), we test whether for ~~the~~ homogenous landscapes it is sufficient to simulate the
385 | local dynamics only in latitudinal, longitudinal and diagonal transects (i.e. north-south, east west, as
386 | well as, northeast-southwest and northwest-southeast corridors) and how this will influence the
387 | migration speed. The corridors are 1 grid cell wide and regularly placed in the simulation domain.
388 | Their density can be chosen by defining the distance between the latitudinal and longitudinal
389 | corridors.

390 | Although LPJ-GM only simulates local dynamics in the cells along the corridors, the seed matrix
391 | needed to be filled for the dispersal calculation using the FFTM or the SMSM algorithm. We applied a
392 | nearest neighbour interpolation of the seed production before performing the seed dispersal calculation
393 | (theoretical considerations show that a distance weighted average would strongly speed up the
394 | migration).

395 | **2.6 Simulation experiments**

396 | To test our newly developed migration module we simulated the spread of a single late successional
397 | species (*Fagus sylvatica*) through an area covered by an early successional species (*Betula pendula*).
398 | The species specific parameters for both species are given in the Supplement S.4. All grid cells and all
399 | years in the simulated area had a static climate suitable for both species. Though the simulated domain
400 | is quadratic in our case it could have any shape. Each cell in the simulated domain has been simulated
401 | independently (except for the influx and outflux of seeds) from each other. For one~~For a~~ specific
402 | simulation using the SMSM method we assumed differences in the dispersal ability (e.g. more or less
403 | permeable areas or physical barriers) while the climate on all grid cells is still static and favourable.
404 | The dispersal ability of the landscape is displayed in Fig. 2. Areas colored white have zero
405 | permeability, hence no seeds can reach these areas.

406 |
407 | <Fig. 2 placed here>

408 |
409 | Figure 3 demonstrates the sequence of local dynamics on the corridors, interpolation of seed
410 | production, seed dispersal on the entire grid and back via the seed input on the transects.

411 | <Fig. 3 placed here>

412 |

413 | Given the uniformity of the climate, there should be no variability in the migration speed caused by
414 | differences in climatic conditions. We simulated the spread of *F. sylvatica* from a single grid cell in
415 | the corner of the study area which represents the refugium. We tested several corridor distances
416 | (between the parallel and between the diagonal corridors) for their effect on the migration speed. ~~To~~
417 | ~~calculate~~~~We calculated~~ the migration speed we first determined the migration distance. This wasas the
418 | distance between the start point of the migration and the 95th percentile farthest point in the virtual
419 | landscape where the leaf area index (LAI) of *F. sylvatica* ~~was~~is larger than 0.5. This migration
420 | distance was subsequently divided by the simulated time elapsed since the start of the migration. To
421 | avoid founder effects we neglected the points within the first 5 km of the refugium. The simulations
422 | were performed over 3000 years and over an area of 100 by 100 cells of 1 km². Finally we ran one
423 | simulation where we did not calculate the seed dispersal (but performed all communication between
424 | cells ~~and one run even without the communication~~), hence allowing us to estimate the computation
425 | time demand for the seed dispersal calculation.

426 | **2.7 Performance evaluations**

427 | To estimate the performance of our methods against an implementation in which each grid cell
428 | exchanges seeds with each other we developed a Matlab® script, since initial testing had shown that
429 | such a procedure would be too slow to be implemented in LPJ-GUESS. Hence when evaluating the
430 | performance differences from the script one has to bear in mind that these are calculated in a different
431 | environment. However in a general sense we can see no reason why they should not reflect the
432 | performance differences between the algorithms. The whole Matlab® script testing the performance
433 | including the graphs is part of the Supplementary material.

434 | **3. Results**

435 | **3.1 Explicit seed dispersal**

436 | The study comparing the performance of different migration mechanisms without the vegetation
437 | dynamics, implemented in Matlab®, ~~has~~Pre-studies have shown that both the FFTM as well as the
438 | SMSM are performing ~~much~~ faster than ~~the~~an explicit dispersal from each grid cell to each other
439 | within the range of the dispersal (last figure Supplement 2).- This is especially pronounced if the area
440 | to be simulated is increased. Though faster thanInstead of comparing the explicit seed exchange with
441 | the FFTM using LPJ-GM we demonstrated this with a small Matlab™ script in the supplementary
442 | material S.2 which also allows demonstrating both the transformation of the seed dispersal method, the
443 | SMSM is still up to an order of magnitude slower thankernel as well as the FFTM. in detail.

444 **3.2 FFTM simulations**

445 Using the parameterization from TreeMig in a complete (no corridors) simulation area of 100 by 100
446 grid cells with the size of 1km² each results in a migration speed of 34 m per year for *Fagus sylvatica*
447 (Fig. 43).

448 <Figure 43 placed here>

449 Though the establishment is stochastic, the spread is relatively smooth. The corridor Using at a
450 distance of 10 km, 20 km and 50 km results in a ~~somewhat~~ reduced migration rate of 26, 28 and 28
451 m/year (compared to a simulation without corridors), respectively (Fig. 43, lower three rows of
452 panels). While in the simulation without corridors the variability of the migration speed is relatively
453 low (dots under the red line in upper left panel of ~~Fig. 43~~), this variability is strongly increased when
454 corridors are simulated. This is caused by *F. sylvatica* migrating along the diagonal, reaching the end
455 point of the diagonal and then migrating along the longitudinal and latitudinal corridors into cells
456 which have actually a shorter distance to the refugia than the endpoint of the diagonal.

457 The calculation time per grid cell in the whole area (range for which the seed dispersal is computed) is
458 increased by 12% ~~by simulating the FFTM~~, but by using the corridors it is reduced to 36%, 22% and
459 12%, compared to simulating the full area (Tab. 1, col. 7). The proportion of computation time used
460 to perform the FFTM increases from 11% without corridors to 18%, 29% and 29% for simulations
461 with corridors every 10, 20 and 50 km. This estimate only includes the required time for computing
462 the FFT-based seed dispersal since the control run without seed dispersal still contained all
463 communication between cells. For the control run seeds were produced and send to the master but the
464 master did not compute the seed dispersal, though still communicated with all other nodes to allow a
465 fair assessment of the computation time demand of the two methods (see Tab. 1). An additional run
466 without any communication resulted in a computation time similar to the run with communication.

467 **3.3 Shifting seed simulations**

468 Initial testing of the probability parameter for the SMSM suggested a value of $p=5 * 10^{-7}$ to generate a
469 migration speed comparable to the migration speed for the FFTM based on the TreeMig
470 parameterization. Using the derivation presented in supplement 2 it is possible to calculate this
471 parameter for a Gaussian dispersal kernel. One can approximate any dispersal kernel by adding several
472 Gaussian kernel, however this would increase calculation time since the SMSM would have to be
473 performed several times. Therefore we decided to choose a parameter for the SMSM approximating
474 the migration speed rather than the seed dispersal kernel used in Lischke et al. (2006).

475 This resulted in a migration speed of 39 m/year for the filled area and 27m/year respective 29 m/year
476 and 30m/year for the 10 km, 20 km and 50km corridors (Fig. 54).

477 <Figure 54 placed here>

478 Similarly to the FFTM simulations, the migration speed is reduced (see table 1 for a summary). Also
479 comparable to the FFTM based seed dispersal computation, calculation time per grid cell in the whole
480 area (range for which the seed dispersal is computed) is increased by 16% by the simulation of
481 dispersal, but reduced to 35%, 19% and 11% by using the corridors. The proportion of calculation
482 time spend for simulating the seed dispersal is comparable to the proportion using the FFT, it is 16%,
483 19%, close to 23% and 32% (see Tab. 1).

484 Since the SMSM allows adjusting the probability depending on the seed transport permeability of the
485 terrain we also simulated the migration within a non-homogenous dispersal area. The results of this
486 simulation are displayed in Fig [65](#).

487

488 <Figure [65](#) placed here>

489 Though all cells of the virtual landscape have a similar climate, some cells will never be occupied (see
490 Fig. [65](#)) because the seeds are not able to reach them (which might not be reasonable for real world
491 simulations but demonstrates the method). Migration speed is different in different parts of the
492 simulated area.

493 <Table 1 placed here>

494

495 **4. Discussion**

496 To our knowledge, in our study for we are the first ~~time that manage to implement~~ (tree) → species
497 migration is implemented in a DGVM in a way that allows ~~continental~~ simulations of simultaneously
498 migrating and interacting species for large areas.

499 **4.1 Performance of new migration methods**

500 The presented new methods for simulating migration in DGVMs show a promising performance in
501 different aspects.

502 The first is the gain of efficiency by the FFTM and the SMSM methods as compared to the traditional,
503 straightforward approach to evaluate the seed transport from each cell to each other (last Fig in S.2). A
504 two dimensional FFT can be obtained by successive passes of the one dimensional FFT, hence the
505 complexity will be the one-dimensional complexity squared (Gonzalez and Woods, 2002) (~~Gonzalez
506 and Woods, 2002~~). The computational complexity for the FFTM is $O(N^2 \log^2(N))$ for a $N \times N$ grid
507 discretizing the seed distribution, while the complexity of the direct implementation of the convolution
508 approach in the SMSM is $O(2KRN^2)$ for a $N \times N$ grid discretizing the seed distribution and $R \times R$
509 kernel with K being the number of iterations of the SMSM (for the derivation see supplementary
510 material S.1). This can be computationally comparable to the FFTM for kernels with short range of R .
511 Secondly, simulating the local dynamics only along the corridors instead of in the full area resulted in
512 a ~~very~~ similar migration pattern, and the simulated migration speed is similar to that of the simulation
513 with full grid cell cover (though it is ~~slightly~~ slower, caused by the stochasticity of the establishment,
514 see table 1), but needs much less computing time (reduction of 88% for the corridors every 50km).

515 **4.2 Comparison of the two dispersal methods**

516 In this study we present two alternative methods for simulating dispersal, which differ in their
517 properties. While the FFTM allows any type of seed dispersal kernel, the SMSM corresponds to a
518 normal distribution kernel. Although other shapes of dispersal kernels can be approximated by
519 weighted sums of normal distributions, of which each of them has to be simulated by an own SMSM,
520 which will cause strong increases in computational demand.

521 On the other hand, the advantage of the SMSM lies in its ability (contrary to the FFTM) to modify the
522 parameters of the seed dispersal kernel spatially, depending on the terrain. If instead of applying a
523 single permeability for all directions, a different permeability is applied for each of the 8 directions
524 (e.g. north, northeast, east, etc.) this method also allows a spatially explicit consideration of wind
525 directions (which is not possible for the FFTM, as it relies on a universal kernel applied to the ~~whole~~
526 entire area). Hence, depending on the aim of the analysis either one or the other or a combination of
527 the algorithms is most suitable.

528 While not implemented here, it should be theoretically possible to use the FFTM (preferably with
529 corridors) for some homogenous parts of the simulated area and the SMSM for the remaining part in a
530 single simulation. As long as the seed donor areas for both methods are exclusive, and the areas in
531 which the seeds are allowed to disperse overlap at least with the width of the kernel, we can see no
532 reasons why this should not be feasible.

533 **4.3 Comparison to other approaches**

534 Our new species migration submodule FFTM uses for the first time an algorithm based on Fast Fourier
535 Transformation to simulate dispersal in a DGVM. FFTM is due to its efficiency one of the
536 “workhorses” in mathematics, physics and signal processing (Strang, 1994). In ecology, there have
537 been a few applications using FFTs to simulate dispersal of pollen (e.g. for risk analysis, Shaw et al.
538 (2006), seeds (Pueyo et al., 2008) or even in a course compendium (Powell, 2001)), but not as a
539 standard technique in DGVMs.

540 The SMSM, in turn, mimics the seed transport process itself in a simple and straightforward way,
541 which to our knowledge has also not been implemented in DGVMs either.

542 Both approaches are combined with features of modelling species migration that are already used in
543 other dynamic vegetation models (cf. Snell, 2014).

544 The cellular automaton KISSMig (Nobis & Normand, 2014), e.g. simulates the spread of single
545 species driven by a spatio-temporal grid of suitability, and by transitions to the nearest neighbour cells,
546 which is similar to one iteration in the SMSM. The suitability based models CATS (Dullinger *et al.*,
547 2012) or MigClim (Engler and Guisan, 2009) simulate a simple demography of single species and
548 explicitly the spread based on a seed dispersal kernel.

549 To also account for ecophysiology, the CATS model was combined with LPJ-GUESS in a post-
550 processing approach (Lehsten *et al.*, 2014) which used a spatio-temporally explicit suitability
551 estimated from LPJ-GUESS simulated productivity of a single species, assuming the presence of the
552 other species. This suitability was subsequently used within CATS to simulate migration spread rates.
553 Such a post-processing approach however does not include interactions between several migrating
554 species.

555 Forest landscape models have been developed to integrate such feedbacks between species as well as
556 dispersal (He et al., 2017; Shifley et al., 2017). These models simulate local vegetation dynamics with
557 species interactions, and dispersal by explicit calculation of seed or seedling transport probabilities
558 with dispersal kernels of different shapes (e.g. LandClim (Schumacher *et al.*, 2004), Landis
559 (Mladenoff, 2004), Iland (Seidl et al., 2012)). To capture spatial heterogeneity, they run at a
560 comparably fine spatial resolution (about 20-100m grid cells), allowing only the simulation of
561 relatively small areas due to computational demands.

562 To overcome such computational limits, several approaches for a spatial upscaling of the models have
563 been put forward. For example, the forest landscape model TreeMig can operate at a coarser resolution
564 (grid cell size 1000m) because it aggregates the within-stand- heterogeneity by dynamic distributions
565 and height classes (Lischke *et al.*, 1998), which allows applications at a larger scale, e.g. over entire
566 Switzerland (Bugmann *et al.*, 2014) or on a transect through Siberia (Epstein *et al.*, 2007). Another
567 upscaling of TreeMig was achieved by the D2C method (Nabel, 2015; Nabel and Lischke, 2013)
568 which simulates local vegetation dynamics only in a subset of cells that are dynamically determined as
569 representative for classes of similar cells. This method led to a computing time reduction of 30-85% as
570 compared to the full simulation similar to our transect methods which resulted a computing time
571 reduction in a similar range depending on the configuration of the corridors.

572 ~~In DGVMs, Our new species migration submodule FFTM uses for the first time (to our best~~
573 ~~knowledge) in ecology an algorithm based on Fast Fourier Transformation to simulate dispersal in a~~
574 ~~DGVM, which due to its efficiency is one of the “workhorses” in mathematics, physics and signal~~
575 ~~processing (Strang, 1994). The SMSM, in turn, mimics the seed transport process itself in a simple~~
576 ~~and straightforward way, which to our knowledge has also not been implemented in DGVMs either.~~

577 ~~Both approaches are combined with features of modelling species migration that are already used in~~
578 ~~other DGVMs (cf. Snell, 2014).~~

579 ~~The cellular automaton KISSMig (Nobis & Normand, 2014), e.g. simulates the spread of single~~
580 ~~species driven by a spatio-temporal grid of suitability, and by transitions to the nearest neighbour cells,~~
581 ~~which is similar to one iteration in the SMSM. The suitability based models CATS (Dullinger *et al.*,~~
582 ~~2012) or MigClim (Engler and Guisan, 2009) simulate a simple demography of single species and~~
583 ~~explicitly the spread based on a seed dispersal kernel.~~

584 ~~To also account for ecophysiology, the CATS model was combined with LPJ-GUESS in a post-~~
585 ~~processing approach (Lehsten *et al.*, 2014) which used a spatio-temporally explicit suitability~~
586 ~~estimated from LPJ-GUESS simulated productivity of a single species, assuming the presence of the~~
587 ~~other species. This suitability was subsequently used within CATS to simulate migration spread rates.~~
588 ~~Such a post-processing approach however does not include interactions between several migrating~~
589 ~~species.~~

590 ~~Forest landscape models have been developed to integrate such feedbacks between species as well as~~
591 ~~dispersal (He *et al.*, 2017; Shifley *et al.*, 2017). These models simulate local vegetation dynamics with~~
592 ~~species interactions, and dispersal by explicit calculation of seed or seedling transport probabilities~~
593 ~~with dispersal kernels of different shapes (e.g. LandClim (Schumacher *et al.*, 2004), Landis~~
594 ~~(Mladenoff, 2004), Iland (Seidl *et al.*, 2012)). To capture spatial heterogeneity, they run at a~~

595 comparably fine spatial resolution (about 20–100m grid cells), allowing only the simulation of
596 relatively small areas due to computational demands.

597 To overcome such computational limits, several approaches for a spatial upscaling of the models have
598 been put forward. For example, the forest landscape model TreeMig can operate at a coarser resolution
599 (grid cell size 1000m) because it aggregates the within stand heterogeneity by dynamic distributions
600 and height classes (Lischke et al., 1998), which allows applications at a larger scale, e.g. over entire
601 Switzerland (Bugmann et al., 2014) or on a transect through Siberia (Epstein et al., 2007). Another
602 upscaling of TreeMig was achieved by the D2C method (Nabel, 2015; Nabel and Lischke, 2013)
603 which simulates local vegetation dynamics only in a subset of cells that are dynamically determined as
604 representative for classes of similar cells. This method led to a computing time reduction of 30–85% as
605 compared to the full simulation similar to our transect methods which resulted in 85% gain of
606 computing time.

607 In dynamic global vegetation models (DGVMs), the discretization problem resulting from the need to
608 upscale from the fine scale at which migration processes act to the scale at which DGVMs work is
609 veryis even more pronounced, because they are designed to operate on very large extents (continents
610 or the entire globe). Given the computational demands of the simulations, they are therefore typically
611 running at a coarse resolution for example 0.5 or 0.1 degree longitude / latitude, and simulate the
612 vegetation dynamics at the centre of each of these grid cells, assuming this point to be representative
613 for the entire cell.

614 Snell (2014)Snell (2014) approached the discretization problem for the DGVM LPJ-GUESS by also
615 using a reduced number of representative units (here patches) within each grid cell. She assumed that
616 the numerous replicates of the vegetation dynamics on a patch are randomly distributed over the area
617 of the grid cell. Migration within the grid cell is treated similar to an infection process, where the
618 probability of a patch becoming infected (e.g. of the migrating species being able to establish) depends
619 only on the number of already invaded patches within the grid cell. Only once a migrating species
620 managed to establish in a certain proportion of the patches of the simulated grid cell, further dispersal
621 (explicit via a dispersal kernel) into surrounding grid cells is possible. Yet, there is no spatial
622 orientation of the patches within the grid cell and all simulations in this approach are strongly
623 resolution dependent. Simulations of large areas such as continents remain computational challenging
624 with this approach.

625 Our transect approach, similarly to the approach of Snell (2014), uses smaller representative spatial
626 units, 1km-cells, for a spatial upscaling.Our transect approach, similarly to the approach of Snell
627 (2014), uses smaller representative spatial units, 1km-cells, for a spatial upscaling. Since these small
628 grid cells are arranged in contiguous corridors, the migration along these corridors can be simulated

629 without or with only a small discretization error. The results indicate that also the error potentially
630 introduced by the interpolation to the rest of the area is small.

631 Thus, with our approaches, we have combined several advantages of the before mentioned approaches:
632 the seed dispersal from forest landscape models, improved by the novel FFTM or SMSM and the
633 ecophysiology, structure and community dynamics of LPJ-GUESS. We furthermore found a
634 compromise between discretization and efficiency by the corridor method.

635 **4.4 Potential further improvements**

636 Despite the satisfying performance of the new methods in these first tests some aspects suggest further
637 development.

638 **4.4.1 Computation time**

639 Even with the computing time reduction by the corridor approach using a corridor of 50km distance,
640 the computing time required for the simulations including dispersal is still considerable. This is caused
641 by the number of cells on the corridors where the local dynamics is simulated being larger than the
642 number of replicates usually used in all the 1 or 0.5 degree grid cells simulated in traditional DGVMs.
643 For large-scale applications, the approach should be further optimized, e.g. by choosing corridors
644 even further apart from each other in homogenous areas and adapting the corridor density to the large
645 scale (between grid-cell) heterogeneity of the terrain. The within grid-cell heterogeneity in turn can be
646 accounted for by deriving seed dispersal permeability, that can be used in the SMSM approach.
647 Another area of improvement lies in the technical implementation of the seed dispersal algorithm. In
648 the current implementation, the seed dispersal is performed at a single cpu, while all other cpus wait
649 until they receive the seeds. There are certainly ways to perform the seed dispersal computation on
650 several nodes to decrease the waiting time. Furthermore, in multi-species simulations the dispersal has
651 to be calculated for each migrating species. In this case, the dispersal of different species should be
652 calculated on separate nodes. When evaluating the run times needed for the simulated areas in the
653 supplementary material it becomes obvious that sometimes larger areas resulted in shorter runtimes for
654 the FFTM (last Fig. in S.2). The differences are quite pronounced given that the time axis is
655 logarithmic. These decreases are caused by the effect that the calculation of a fft can be optimised in
656 case the domain has a size of 2^n .

657 **4.4.2 Migration speed reduction by corridor approach**

658 As expected, any sub-cell assumption results in discretisation errors. In our case the assumption of a
659 corridor ~~slightly~~ reduced the migration speed. This needs to be taken into account when evaluating the
660 result of such studies. The design of the corridors might also not ~~have been~~ optimal, maybe a
661 corridor wider than a single cell might result in less decrease of migration speed. However, these types
662 of analysis are outside the scope of this study. One other aspect of using the corridors is that while a
663 late successional species (in our case *F. sylvatica*) has certainly no problems to establish below the
664 early successional species, in the case of an early successional species (e.g. *B. pendula*) migrating into

665 an area occupied by a late successional species, the corridors might decrease the migration speed even
666 more. An early successional species can only establish after sufficient light reaches the ground, either
667 due to the senescence of a tree of the established species or a disturbance event. The narrow corridors
668 might strongly limit the availability of such grid cells. However since early successional species have
669 typically a good dispersal ability, this should not influence simulations of tree migration following
670 climate change (e.g. after the last glaciation).

671 **4.4.3 Parameterisation of dispersal kernels and other plant parameters**

672 In this study the focus is on developing and testing the novel methods, i.e. we do not attempt to
673 correctly simulate the spread of *F. sylvatica* over a defined time period. The calculated spread rates are
674 well below most of the spread rates in the literature. *F. sylvatica* has been estimated to migrate with ca
675 100 m per year based on pollen analyses by Bradshaw and Lindbladh (2005). Although such estimated
676 high migration speeds could also be the result of glacial refugia located further north than assumed
677 (Feurdean et al., 2013), our estimates of the migration speeds of 20-30 m/year still seem rather low.
678 However, in this paper we aimed to implement tree migration by using the parameterisation of
679 TreeMig in a DGVM and thereby allow large scale simulations. Our estimated migration rates of 20-
680 30 m per year are very close to the migration rates estimated for this parameterisation for TreeMig by
681 Meier et al., (2012) which estimated a value of 22 m per year. ~~*sylvatica* has been estimated to migrate~~
682 ~~with ca 100 m per year based on pollen analyses by Bradshaw and Lindbladh (2005). Although such~~
683 ~~estimated high migration speeds could also be the result of glacial refugia located further north than~~
684 ~~assumed (Feurdean et al., 2013), our estimates of the migration speeds of 20-30 m/year still seem~~
685 ~~rather low. However, in this paper we aimed to implement tree migration by using the~~
686 ~~parameterisation of TreeMig in a DGVM and thereby allow continental scale simulations. Our~~
687 ~~estimated migration rates of 20-30 m per year are very close to the migration rates estimated for this~~
688 ~~parameterisation for TreeMig by Meier et al., (2012) which estimated a value of 22 m per year. Hence,~~
689 though we implemented the migration module into a conceptually very different model, the resulting
690 migration rate remains comparably similar.

691 To perform modelling runs estimating the migration speed of any species would require a fine tuning
692 of the, age of maturity, seed production, dispersal parameters, germination rates, and seed survival
693 (which are very rough estimates in TreeMig ;Lischke et al., 2006) to generate the observed migration
694 e.g. by comparing to migration rates based on pollen records. Unfortunately, though all of these
695 parameters are most likely strongly influencing the migration rates, they are not only hard to find in a
696 study performed with similar methods for all tree species, they are likely to be highly variable
697 depending on growth conditions and even provenance of the individual tree. However for a large scale
698 application at least the sensitivity of these parameters should be evaluated.

699 While we limited us to use the same approach as Lischke et al. (2006) starting seed production at a
700 fixed height of maturity which accounts for a developmental threshold, but also growth and thus for
701 environmental conditions, other studies used age of maturity as a trigger to start seed production,
702 which has been shown to be important to determine tree migration rates (e.g Nathan et al., 2011). As
703 the aim of this study was not a full sensitivity analysis but a study showing that a similar approach as
704 Lischke et al. (2006) results in comparable migration rates, we will implement the option to use age of
705 maturity in the next version of LPJ-GM.

706 Applications of our approach to simulate migration in the future are only suitable if the migration
707 speed is substantially faster than the migration speed that we reach for *F. sylvatica* (due to typically
708 shorter simulation period) and if the species is not typically planted, which is common practice in
709 many commercial forests where alien species a planted.

710 ~~To perform modelling runs estimating the migration speed of any species would require a fine tuning~~
711 ~~of the seed production and dispersal parameters (which are both very rough estimates in TreeMig~~
712 ~~(Lischke et al., 2006) to generate the observed migration e.g. by comparing to migration rates based on~~
713 ~~pollen records.~~

714 **4.5 Potential for applications**

715 The test simulations were performed at a virtual landscape of 100km by 100km, but eventually the
716 method is aimed to allow ~~large scale~~~~continental~~ simulations over several millennia. Regarding
717 memory requirements, this is possible of currently available hardware: Test runs with landscapes of
718 4000 by 4000 grid cells (i.e. the size of Europe) performed without technical problems at least
719 regarding the memory requirement (given 62 GB of RAM). The considerable computational costs
720 however require a relatively high amounts of computing time, which might be reduced by efforts for
721 speeding up (due to efficient parallelisation) of the FFT (currently the FFTM is performed on a
722 single node while the remaining nodes are idle, one could use all nodes to perform the FFTM) or FFT
723 ~~as the local simulations, e.g.~~ by even further apart corridors.

724 **5. Conclusions**

725 The presented novel approaches offer ~~a~~ high potential to simulate the spatiotemporal dynamics of
726 species which are migrating and interacting with each other simultaneously. The approaches are not
727 restricted to LPJ-GUESS, but can in in principle be applied to other DGVMs or FLMs which simulate
728 seed (or seedling) production and explicit regeneration. ~~The~~~~While the~~ presented methods need to be
729 improved in terms of computing performance to~~will~~ allow simulations of~~in future to simulate~~ tree
730 migration at continental scale and over paleo time scales. ~~O~~~~ur~~ study also shows that the estimates for
731 seed dispersal kernels for the major tree species need to be revised to allow ~~truly mechanistic~~
732 simulations of forest development for example over the Holocene.

733 **6. Author contributions**

734 VL, DL and HL designed the study, VL performed the simulations and the statistical analysis. MM
735 and EL contributed to the study design, MM also performed large parts of the coding. EL developed
736 the formal proof in Supplementary material S.1 and the computation performance related estimates [in](#)
737 the Conclusion section. All authors contributed to the writing of the article.

738 **7. Competing interests**

739 The authors declare that they have no conflict of interest.

740 **8. Acknowledgements**

741 This study was funded by the Swiss National Science Foundation project CompMig, Nr .
742 205321_163223 .

743 **9. Code and Data availability**

744 The code generating the figures in Supplementary material 2 are part of the material. The used DGVM
745 LPJ-GUESS containing the migration module can be requested from the author.

746 The data behind all figures will be published on the DataGURU server (dataguru.lu.se) with an own
747 DOI upon acceptance of the paper.

748 **10. References**

- 749
- 750 Bradshaw, R. H. W. and Lindbladh, M.: Regional spread and stand-scale establishment of *Fagus*
751 *sylvatica* and *Picea abies* in Scandinavia, *Ecology*, 86(7), 1679–1686, doi:10.1890/03-0785, 2005.
- 752 Bugmann, H. K. M., Brang, P., Elkin, C., Henne, P., Jakoby, O., Lévesque, M., Lischke, H., Psomas,
753 A., Rigling, A., Wermelinger, B. and Zimmermann, N. E.: Climate change impacts on tree species,
754 forest properties, and ecosystem services, in CH2014-Impacts (2014): Toward Quantitative Scenarios
755 of Climate Change Impacts in Switzerland, edited by O. (Meteoswiss) Foen., 2014.
- 756 Clarke, L., Glendinning, I. and Hempel, R.: The MPI Message Passing Interface Standard, in
757 Programming Environments for Massively Parallel Distributed Systems, edited by R. R. . Decker
758 K.M., Birkhäuser, Basel., 1994.
- 759 Cooley, J. W. and Tukey, J. W.: An Algorithm for the Machine Calculation of Complex Fourier Series
760 Mathematics of Computation An Algorithm for the Machine Calculation of Complex Fourier Series,
761 Source Math. Comput., 19(90), 297–301, doi:10.2307/2003354, 1965.

762 Dullinger, S., Willner, W., Plutzer, C., Englisch, T., Schrott-Ehrendorfer, L., Moser, D., Ertl, S., Essl,
763 F. and Niklfeld, H.: Post-glacial migration lag restricts range filling of plants in the European Alps,
764 *Glob. Ecol. Biogeogr.*, 21(8), 829–840, doi:10.1111/j.1466-8238.2011.00732.x, 2012.

765 Engler, R. and Guisan, A.: MigClim: Predicting plant distribution and dispersal in a changing climate,
766 *Divers. Distrib.*, 15(4), 590–601, doi:10.1111/j.1472-4642.2009.00566.x, 2009.

767 Epstein, H. E., Yu, Q. Y. Q., Kaplan, J. O. and Lischke, H.: Simulating Future Changes in Arctic and
768 Subarctic Vegetation, *Comput. Sci. Eng.*, 9(4), 12–23, doi:10.1109/MCSE.2007.84, 2007.

769 Feurdean, A., Bhagwat, S. A., Willis, K. J., Birks, H. J. B., Lischke, H. and Hickler, T.: Tree
770 Migration-Rates: Narrowing the Gap between Inferred Post-Glacial Rates and Projected Rates, *PLoS*
771 *One*, 8(8), doi:10.1371/journal.pone.0071797, 2013.

772 Foley, J. A., Levis, S., Prentice, I. C., Pollard, D. and Thompson, S. L.: Coupling dynamic models of
773 climate and vegetation, *Glob. Chang. Biol.*, 4(5), 561–579, doi:10.1046/j.1365-2486.1998.00168.x,
774 1998.

775 Gonzalez, R. C. and Woods, R. E.: *Digital Image Processing*, 2nd ed., Prentice Hall., 2002.

776 He, H. S., Gustafson, E. J. and Lischke, H.: Modeling forest landscapes in a changing climate: theory
777 and application, *Landsc. Ecol.*, 32(7), 1299–1305, doi:10.1007/s10980-017-0529-4, 2017.

778 Lehsten, D., Dullinger, S., Huber, K., Schurgers, G., Cheddadi, R., Laborde, H., Lehsten, V., Francois,
779 L., Dury, M. and Sykes, M. T.: Modelling the Holocene migrational dynamics of *Fagus sylvatica* L.
780 and *Picea abies* (L.) H. Karst, *Glob. Ecol. Biogeogr.*, 23(6), 658–668, doi:10.1111/geb.12145, 2014.

781 Lehsten, V., Arneth, A., Spessa, A., Thonicke, K. and Moustakas, A.: The effect of fire on tree-grass
782 coexistence in savannas: A simulation study, *Int. J. Wildl. Fire*, 25(2), doi:10.1071/WF14205, 2016.

783 Lischke, H., Löffler, T. J. and Fischlin, A.: Aggregation of individual trees and patches in forest
784 succession models: Capturing variability with height structured, random, spatial distributions, *Theor.*
785 *Popul. Biol.*, 54(3), 213–226, doi:10.1006/tpbi.1998.1378, 1998.

786 Lischke, H., Zimmermann, N. E., Bolliger, J., Rickebusch, S. and Löffler, T. J.: TreeMig: A forest-
787 landscape model for simulating spatio-temporal patterns from stand to landscape scale, *Ecol. Modell.*,
788 199(4), 409–420, doi:10.1016/j.ecolmodel.2005.11.046, 2006.

789 Meier, E. S., Lischke, H., Schmatz, D. R. and Zimmermann, N. E.: Climate, competition and
790 connectivity affect future migration and ranges of European trees, *Glob. Ecol. Biogeogr.*, 21(2), 164–
791 178, doi:10.1111/j.1466-8238.2011.00669.x, 2012.

792 Mladenoff, D. J.: LANDIS and forest landscape models, *Ecol. Modell.*, 180(1), 7–19,
793 doi:10.1016/j.ecolmodel.2004.03.016, 2004.

794 Nabel, J. E. M. S.: Upscaling with the dynamic two-layer classification concept (D2C): TreeMig-2L,
795 an efficient implementation of the forest-landscape model TreeMig, *Geosci. Model Dev.*, 8(11), 3563–
796 3577, doi:10.5194/gmd-8-3563-2015, 2015.

797 Nabel, J. E. M. S. and Lischke, H.: Upscaling of spatially explicit and linked time- and space-discrete
798 models simulating vegetation dynamics under climate change., in *27th International Conference on*
799 *Environmental Informatics for Environmental Protection, Sustainable Development and Risk*
800 *Management, EnviroInfo 2013*, edited by B. Page, F. A. G. J. Göbel, and V. Wohlgemuth, pp. 842–
801 850, Hamburg., 2013.

802 Nobis, M. P. and Normand, S.: KISSMig - a simple model for R to account for limited migration in
803 analyses of species distributions, *Ecography (Cop.)*, 37(12), 1282–1287, doi:10.1111/ecog.00930,
804 2014.

805 Quillet, A., Peng, C. and Garneau, M.: Toward dynamic global vegetation models for simulating
806 vegetation–climate interactions and feedbacks: recent developments, limitations, and future
807 challenges, *Environ. Rev.*, 18(NA), 333–353, doi:10.1139/A10-016, 2010.

808 Sato, H. and Ise, T.: Effect of plant dynamic processes on African vegetation responses to climate
809 change: Analysis using the spatially explicit individual-based dynamic global vegetation model
810 (SEIB-DGVM), *J. Geophys. Res.*, 117(G3), G03017, doi:10.1029/2012JG002056, 2012.

811 Sato, H., Itoh, A. and Kohyama, T.: SEIB-DGVM: A new Dynamic Global Vegetation Model using a
812 spatially explicit individual-based approach, *Ecol. Modell.*, 200(3–4), 279–307,
813 doi:10.1016/j.ecolmodel.2006.09.006, 2007.

814 Schumacher, S., Bugmann, H. and Mladenoff, D. J.: Improving the formulation of tree growth and
815 succession in a spatially explicit landscape model, *Ecol. Modell.*, 180(1), 175–194,
816 doi:10.1016/j.ecolmodel.2003.12.055, 2004.

817 Seidl, R., Rammer, W., Scheller, R. M. and Spies, T. A.: An individual-based process model to
818 simulate landscape-scale forest ecosystem dynamics, *Ecol. Modell.*, 231, 87–100,
819 doi:10.1016/j.ecolmodel.2012.02.015, 2012.

820 Shifley, S. R., He, H. S., Lischke, H., Wang, W. J., Jin, W., Gustafson, E. J., Thompson, J. R.,
821 Thompson, F. R., Dijak, W. D. and Yang, J.: The past and future of modeling forest dynamics: from
822 growth and yield curves to forest landscape models, *Landsc. Ecol.*, 32(7), 1307–1325,
823 doi:10.1007/s10980-017-0540-9, 2017.

824 Shiryaev, A. N.: Probability, 3rd ed., Springer Verlag, New York., 2016.

825 Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J. O., Levis, S.,
826 Lucht, W., Sykes, M. T., Thonicke, K. and Venevsky, S.: Evaluation of ecosystem dynamics, plant
827 geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model, *Glob. Chang.*
828 *Biol.*, 9(2), 161–185, 2003.

829 Smith, B., Prentice, I. C. and Sykes, M. T.: Representation of vegetation dynamics in the modelling of
830 terrestrial ecosystems: comparing two contrasting approaches within European climate space, *Glob.*
831 *Ecol. Biogeogr.*, 10(6), 621–637, 2001.

832 Smith, B., Wärlind, D., Arneth, A., Hickler, T., Leadley, P., Siltberg, J. and Zaehle, S.: Implications of
833 incorporating N cycling and N limitations on primary production in an individual-based dynamic
834 vegetation model, *Biogeosciences*, 11(7), 2027–2054, doi:10.5194/bg-11-2027-2014, 2014.

835 Snell, R. S.: Simulating long-distance seed dispersal in a dynamic vegetation model, *Glob. Ecol.*
836 *Biogeogr.*, 23(1), 89–98, doi:10.1111/geb.12106, 2014.

837 Snell, R. S., Huth, A., Nabel, J. E. M. S., Bocedi, G., Travis, J. M. J., Gravel, D., Bugmann, H.,
838 Gutiérrez, A. G., Hickler, T., Higgins, S. I., Reineking, B., Scherstjanoi, M., Zurbriegen, N. and
839 Lischke, H.: Using dynamic vegetation models to simulate plant range shifts, *Ecography (Cop.)*,
840 37(12), 1184–1197, doi:10.1111/ecog.00580, 2014.

841 Strang, G.: Wavelets, *Am. Sci.*, 82(May-June), 250–255, 1994.

842 Yue, C., Ciais, P., Luysaert, S., Li, W., McGrath, M. J., Chang, J. and Peng, S.: Representing
843 anthropogenic gross land use change, wood harvest, and forest age dynamics in a global vegetation
844 model ORCHIDEE-MICT v8.4.2, *Geosci. Model Dev.*, 11(1), 409–428, doi:10.5194/gmd-11-409-
845 2018, 2018.

846

847 **1110. Contents of the supplementary material**

848

849 Derivation of the variance of the seed dispersal kernel for the SMSM S.1

850 Example evaluation of computation time difference between FFTM and the
851 traditional method S.2

852 In this appendix an example code for the FFTM is given together
853 with code demonstrating the required transformation of the seed
854 kernel for the FFTM

855

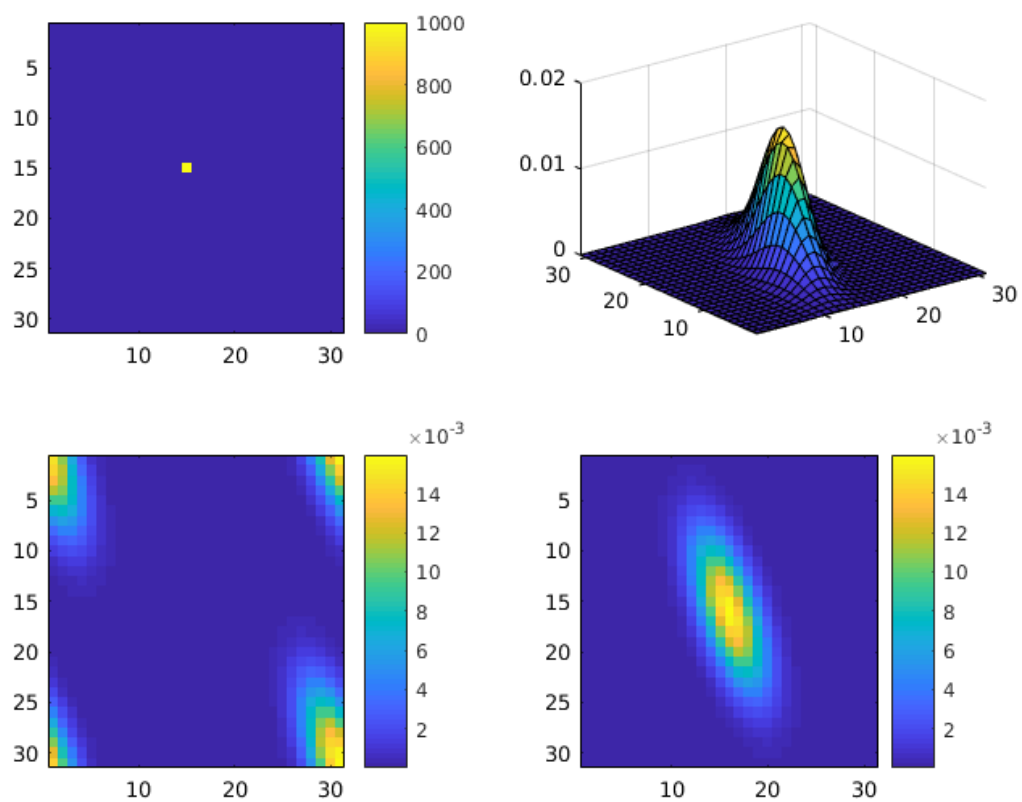
856 | [Example calculation of the SMSM](#) [S.3](#)

857 | [Species specific parameters within the simulation](#) [S.4](#)

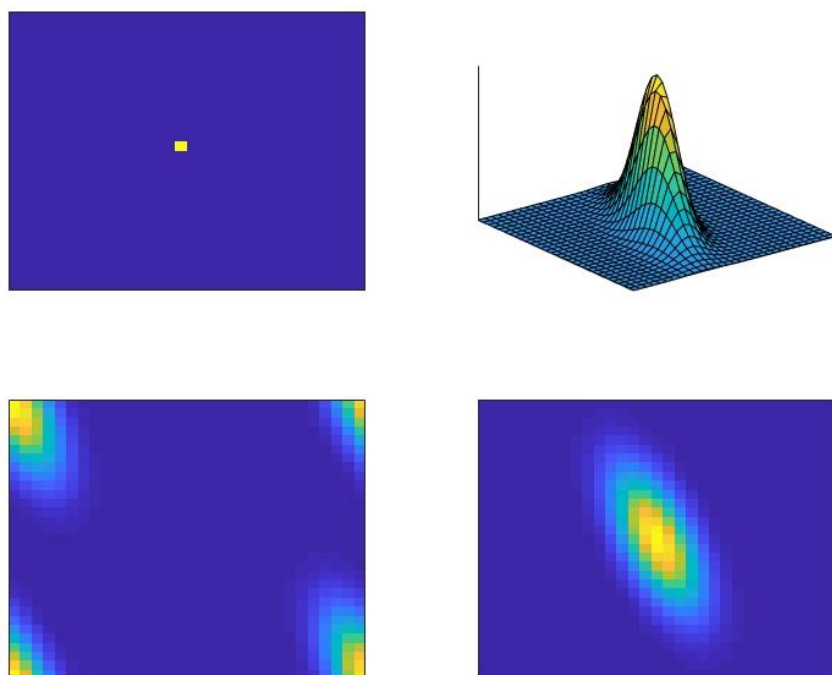
858

860 **11. Figures and tables**

861
862



863

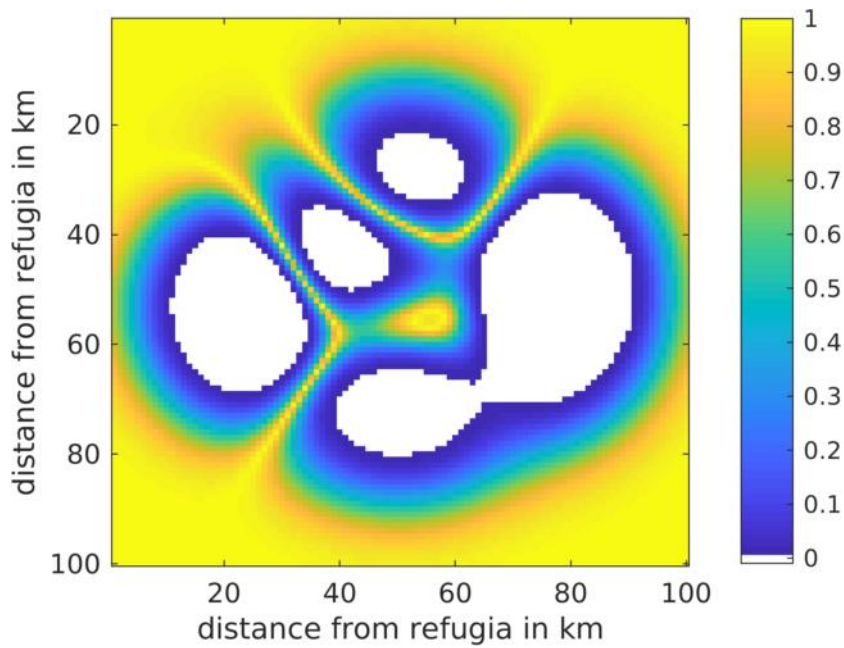


864

865 Fig. 1. Upper left panel: seed source. Upper right panel: example of a seed dispersal kernel (here a
866 non-symmetric kernel is assumed), lower left panel: transformed seed dispersal kernel, lower right
867 panel: seed distribution after convolution.

868

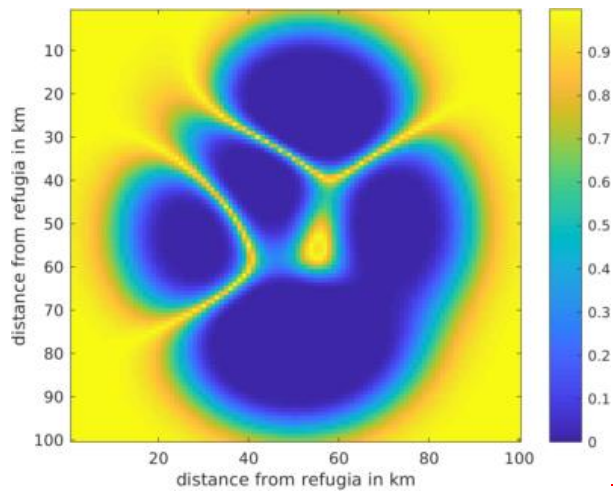
869



870

871

872

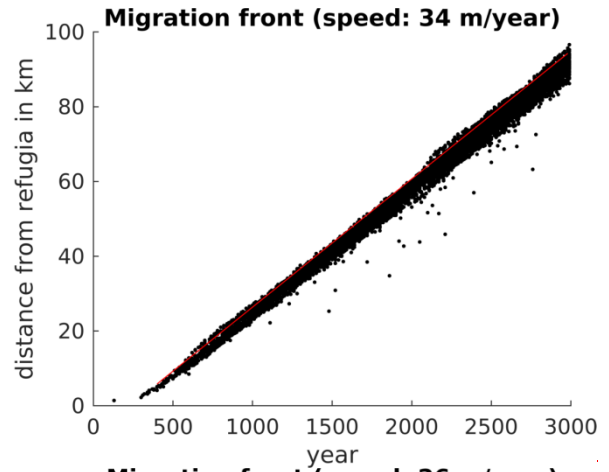
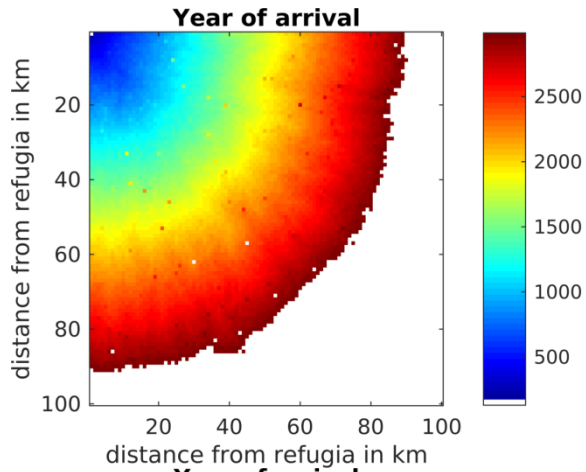


873

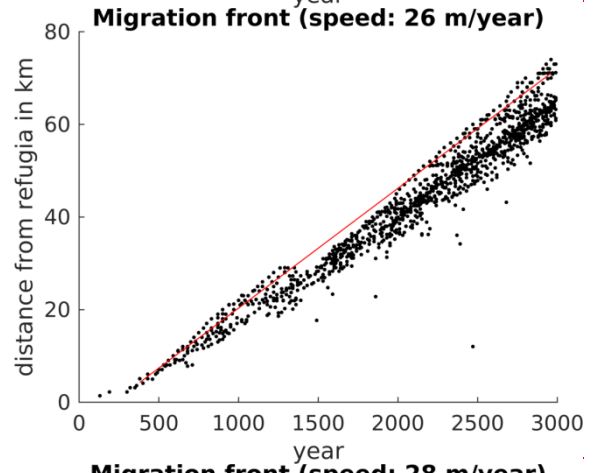
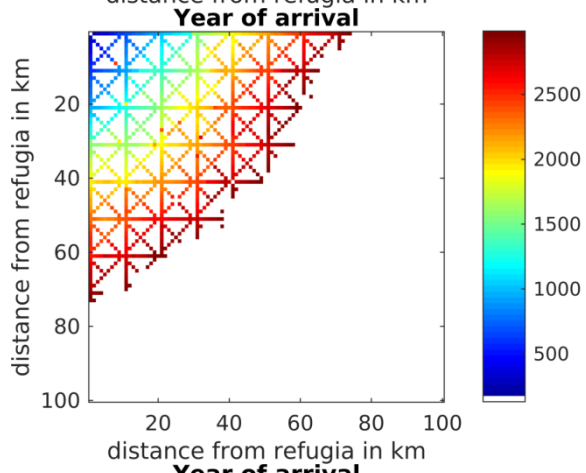
874 Fig. 2: Seed dispersal permeability for SMSM simulation tests. Each time the seed matrix is shifted,
875 the probability of entering the new cell (which in our test is set to $5 \cdot 10^{-7}$) is multiplied with the seed
876 dispersal permeability of the new potentially entered cell.

877

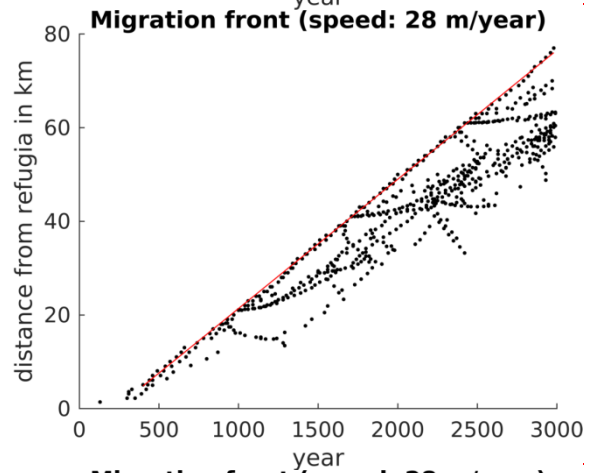
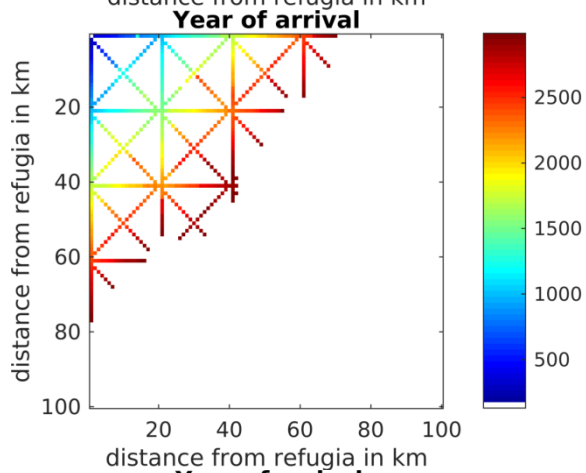
878



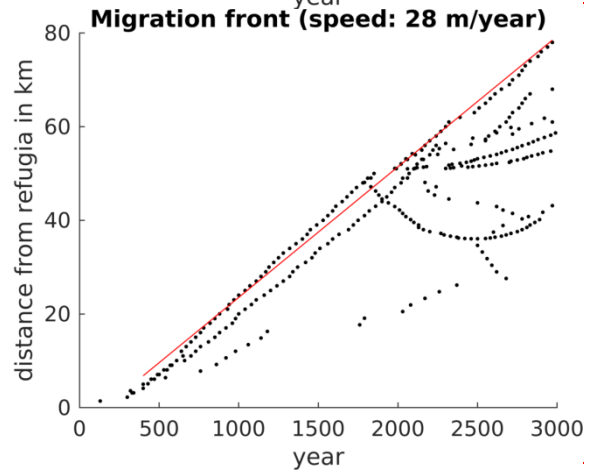
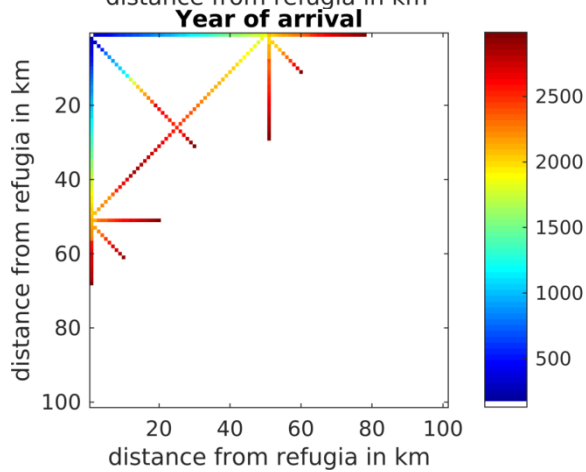
879



880



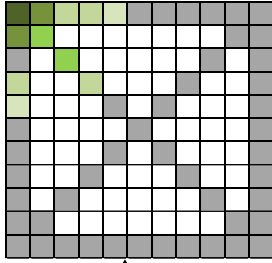
881



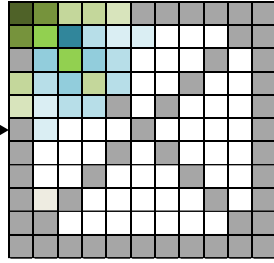
882

~~Fig.~~

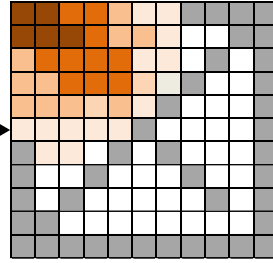
Local dynamics,
including seed
production,
on corridors



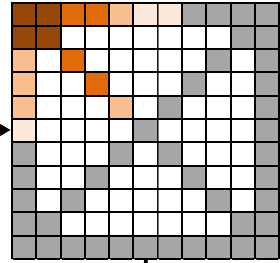
Seed production
interpolated (blue) ,
on grid



Seed dispersal,
on grid



Seed input,
on corridors



883

884

885

886

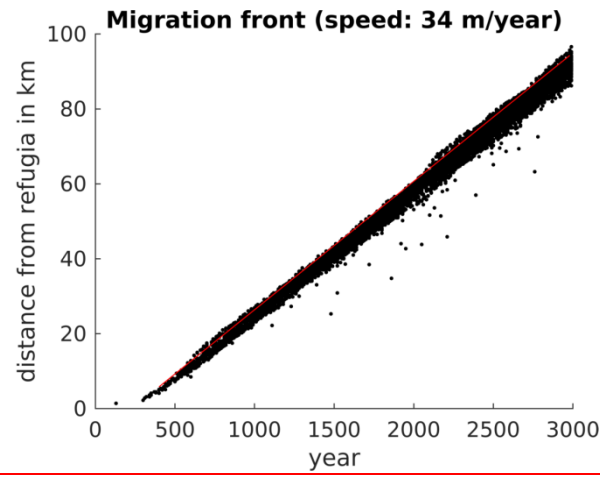
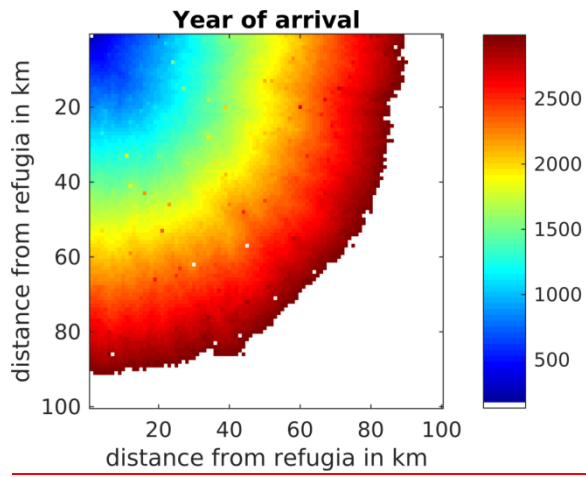
887

888

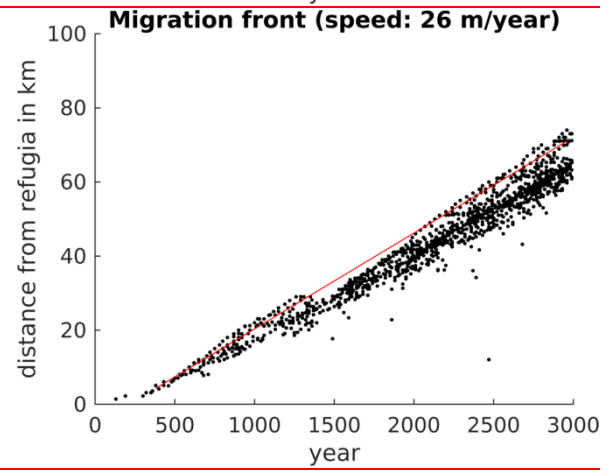
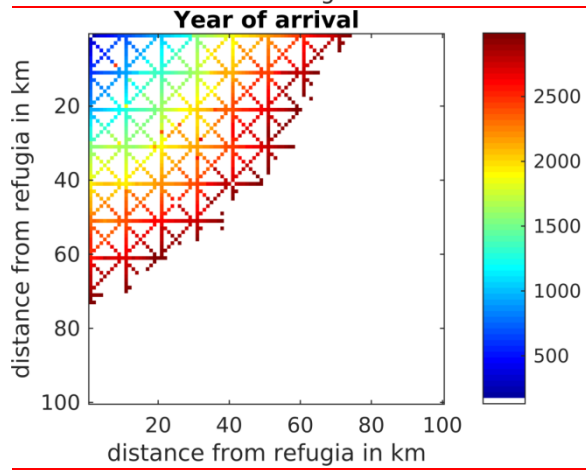
889

Fig. 3. Example of a simulated grid with transects (grey). In each time step the local vegetation dynamics including the seed production (green) is calculated on the transects. Then the seed production of each species is interpolated from the transects to all non-transect grid cells (blue) and then dispersed on the entire grid (brown). The seed input on the transect cell then enters the local dynamics in the next time step.

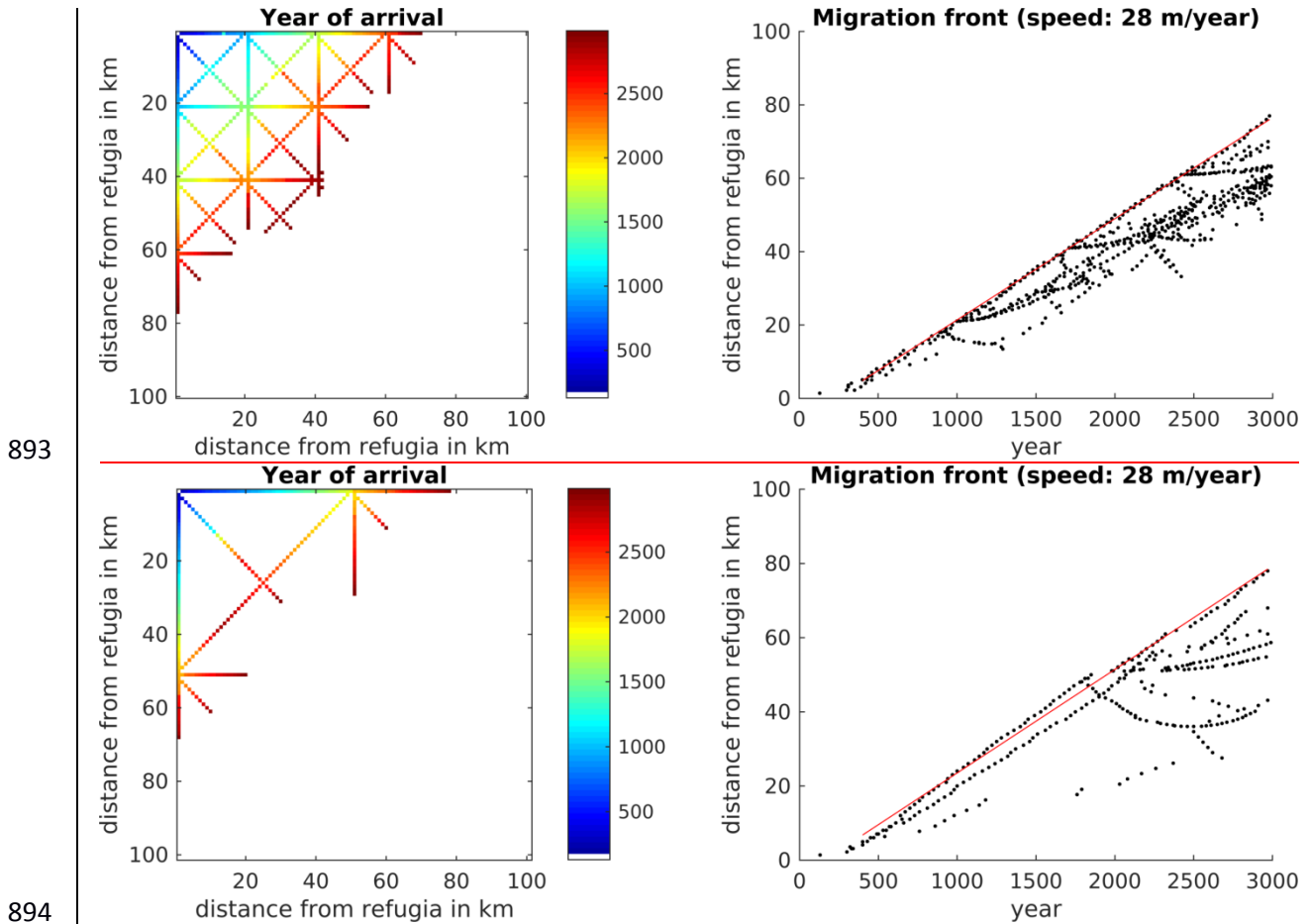
890



891



892

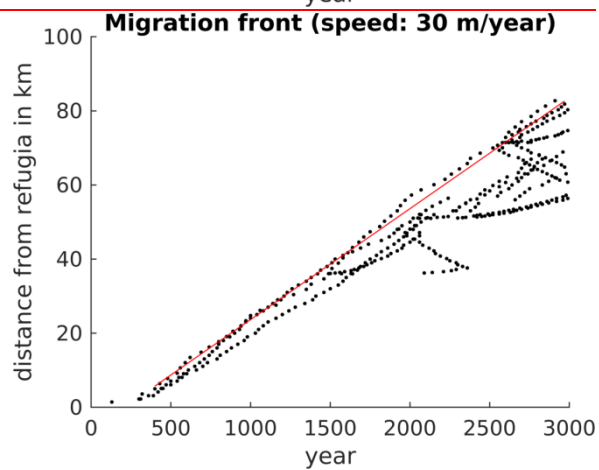
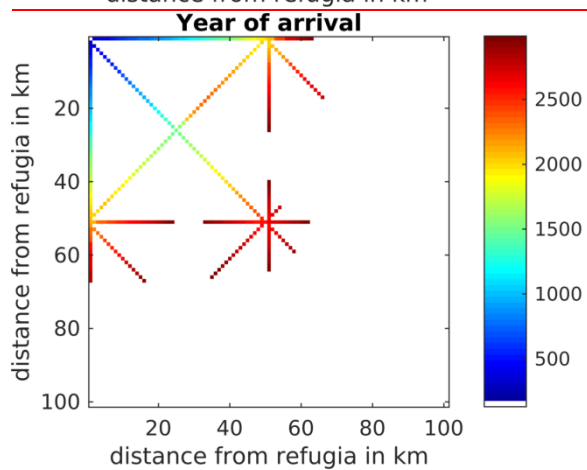
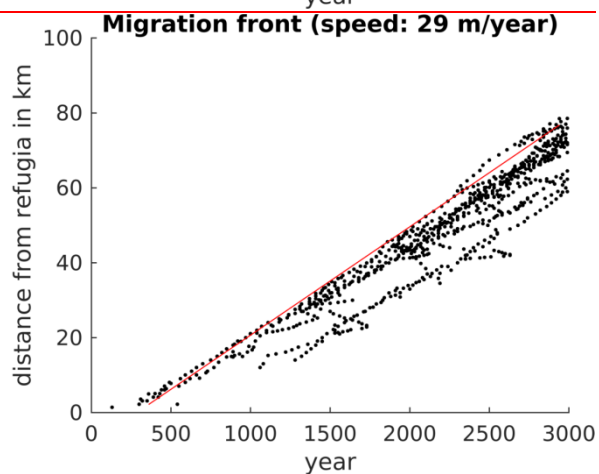
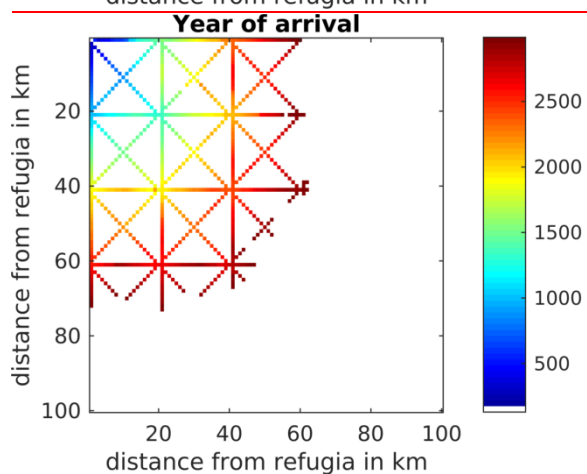
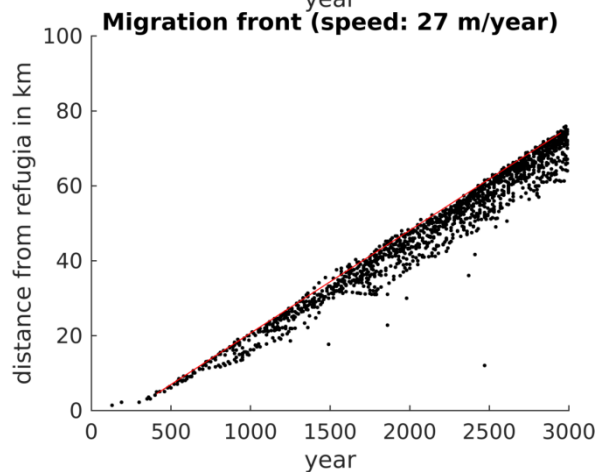
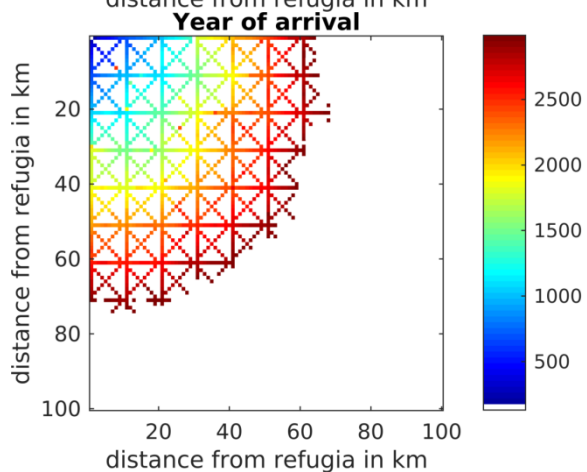
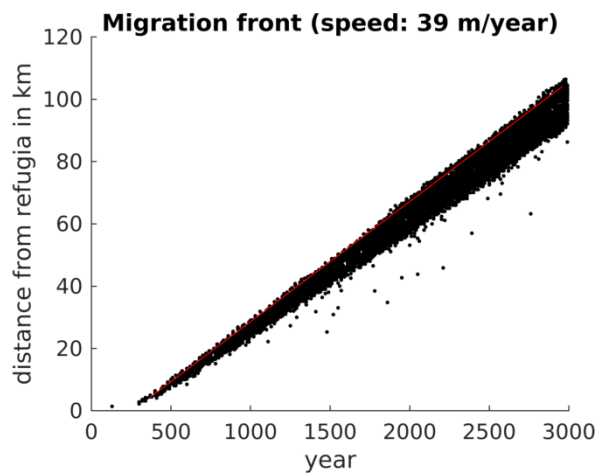
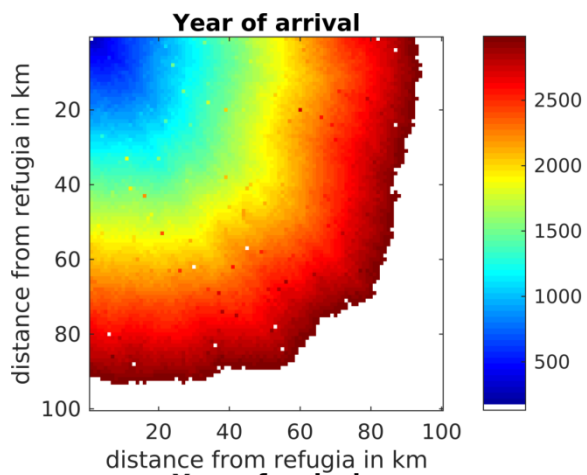


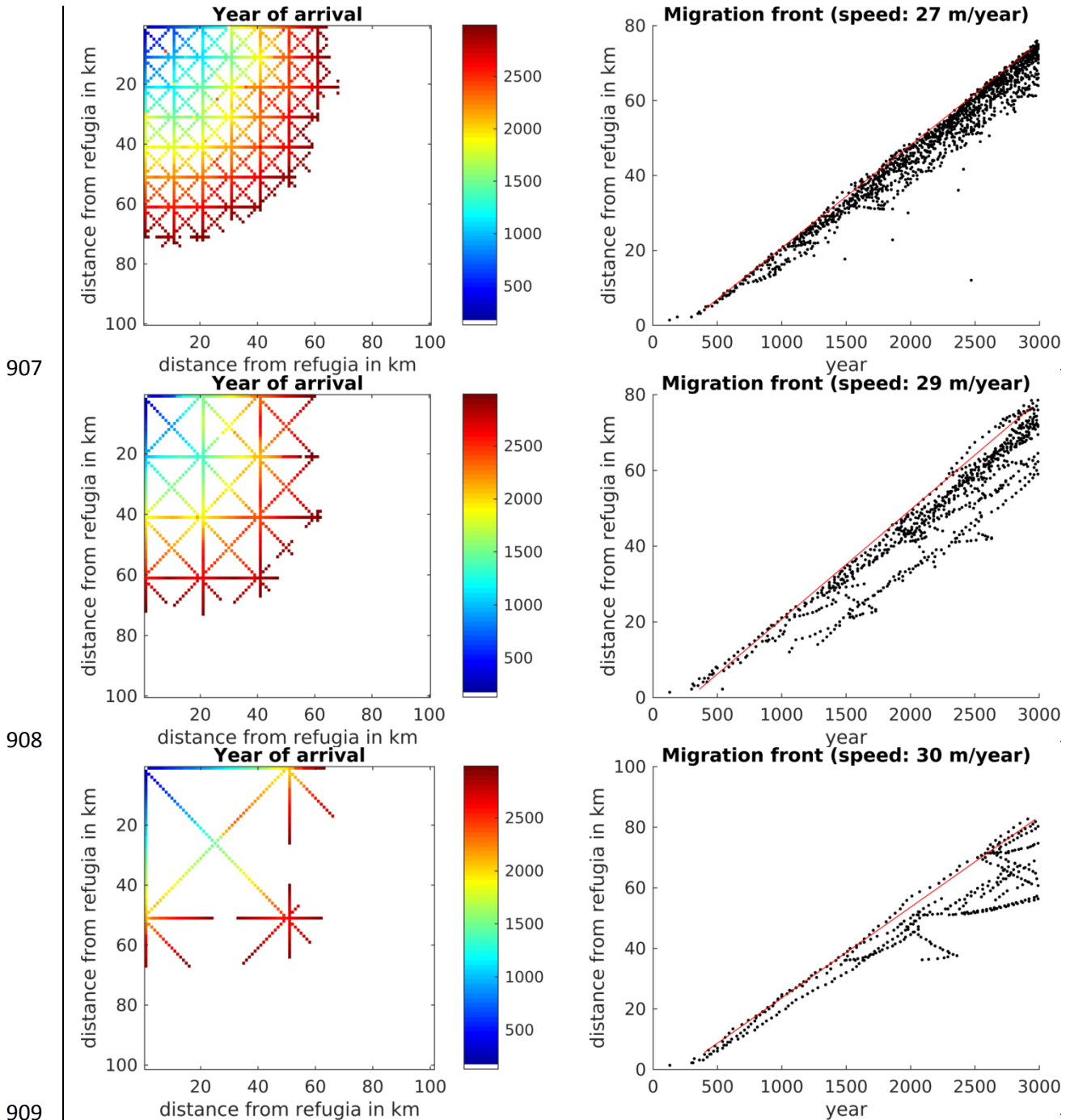
893

894

895 **Fig. 43** Spread of *Fagus sylvatica* through an area of 100 * 100 grid cells with static climate using the
 896 FFTM algorithm with no corridors or corridors every 10km, 20km or 50km. The left panels display
 897 the time when *F. sylvatica* first reached an LAI of 0.5. *F. sylvatica* is allowed to establish freely only
 898 in the upper left corner. The right panels show the distance of the grid cells with LAI 0.5 for *F.*
 899 *sylvatica* from the starting point. The red line indicates the 95 percentile of the grid cells farthest away
 900 from the starting point. The migration speed is calculated as slope of this line, taking only ~~for~~ grid
 901 cells at least 5 km away from the starting point into account to avoid some initial establishing effects.

902





907

908

909

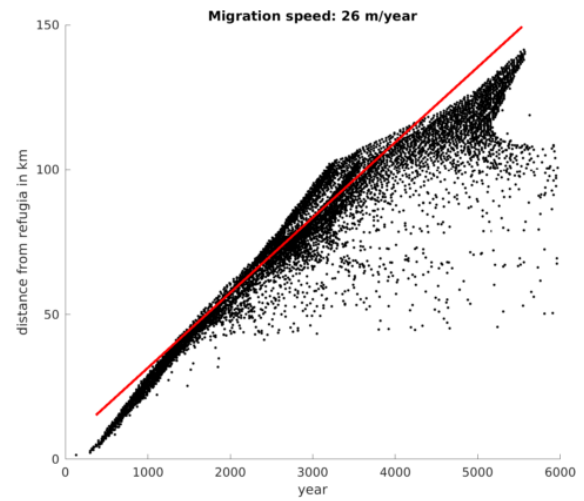
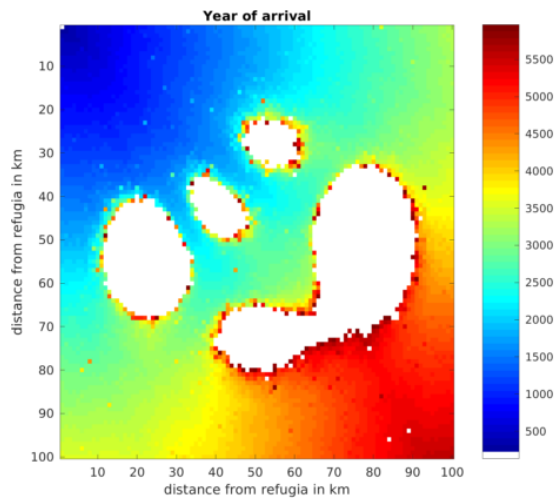
910

911

912

913

Fig. 54 Spread of *Fagus sylvatica* using the SMSM through an area of 100 * 100 grid cells with identical climate, using the full area (upper row of panels) or corridors every 10th, 20th or 50th cell. For more explanation see Fig. 3.



914

915 | Fig. 65 Spread of *Fagus sylvatica* using the SMSM method through an area of 100 * 100 grid cells
 916 with identical climate but probability of seed fall is set to 0.00005 multiplied with the spatially explicit
 917 seed dispersal permeability value as shown in Fig. 2. Note that we increased the simulation time to
 918 6000 years in order to have *F. sylvatica* establishing in all areas.

919

920 Table 1. Summary of migration speeds and calculation time. A corridor distance of 0 indicates no
 921 corridors but an area completely filled with grid cells. The simulated grid cells column lists the
 922 number of cells for which LPJ-GM calculates the population dynamics, in all simulations the
 923 simulation domain (for which the seed dispersal was calculated) had a size of 10000 grid cells and all
 924 simulations were performed over 3000 years. The last line lists a simulation identical to the others
 925 except that no seed dispersal was calculated to allow estimating the computation time demand for this
 926 operation.

Seed dispersal mode	Corridor distance (cells)	Simulated grid cells (corridor cells)	Migration speed, m/year	Computation time (CPU ^{epu} h)	Comp. time change per corridor grid cell compared to sim. without dispersal (CPU h)	Total comp. time change for whole domain compared to sim. without dispersal (CPU h)	Percentage of CPU time for dispersal	Decrease due to corridor simulation
FFTM	0	10000	34	1800	+12%	+12%	11%	
FFTM	10	3330	26	650	+22%	-59%	18%	67%
FFTM	20	1765	28	400	+41%	-75%	29%	78%
FFTM	50	977	27	220	+41%	-86%	29%	88%
SMSM	0	10000	39	2000	+25%	+19%	16%	
SMSM	10	3330	27	700	+31%	-59%	19%	65%
SMSM	20	1765	29	400	+41%	-77%	23%	81%
SMSM	50	977	30	220	+41%	-86%	32%	89%
Non	0	10000	0	1600	0%	0%	0%	

927