Interactive comment on “Dispersal distances and migration rates at the arctic treeline in Siberia – a genetic and simulation based study” by Stefan Kruse et al.

Anonymous Referee #3

Received and published: 20 November 2018

We thank the reviewer for reviewing our manuscript and for the helpful comments. We revised our manuscript at the corresponding positions for each specific comment below.
Our response are placed in bold font below each of the reviewer’s comments in italics. Followed by a citation of changed text with a line statement that refers to the version of the manuscript with tracked changes.

General comments:

This study presents seed dispersal improvements made to the LAVESI individual-based, spatially explicit vegetation simulator to investigate larch migration rates on the Taymyr Peninsula in Northern Siberia. The seed dispersal equation updates were based on a field-based study of genetic parentage of trees in the study area, which found that prior to improvement, the model tended to underestimate dispersal distances in general and overestimate the numbers of recruits close to the parent tree.
The updated model was used to simulate south-to-north transects and the rate of tree line advance was found to differ from the rate of forest line advance by ~1m per year. The study is well-structured and presented, and it addresses an important and poorly understood topic related to environmental change in the region.

However, two key points that I think need to be clarified or addressed are topographic gradients and mortality. Microsite effects brought on by topographic variation seem to not be considered here, though they are an important consideration in a study measuring the rate and manner of treeline advancement. In addition, there is no discussion of mortality rates, both in seedlings and seeds. Seedling dispersal and seed/ling mortality are tightly interconnected and should be at least discussed if not reported. Overall however, this manuscript should be accepted with some of these modifications addressed.

We added a discussion about the important microsite effects and seedling survival rates to section 4.3. “Treeline migration rates”. Basically, we implicitly take account for them with our parameterization approach (Kruse et al., 2016). An explicit implementation would of course improve the realisticity of the model’s outcome, but also increase the already high demand of parameters and finding good estimates for them. Nevertheless, testing for microsite effects and implementing them in the model would allow for a detailed study of their impact, but this is out of the scope of this manuscript.

Same response as to comment R2 1.

Line 361ff:
“Furthermore, the probability of seeds surviving and forming a seedbank and the survival rates of seedlings strongly determine the colonisation speed. This is linked to the availability of microsites where seedlings benefit from shelter, thus lowering their mortality rates (e.g. Resler et al., 2005; Maher et al., 2006; Germino et al., 2011). These effects are not explicitly simulated but implicitly taken account of by our model parameterisation (Kruse et al., 2016). Migration corridors along rivers are not taken into account but they likely assist colonisation in these landscapes because of deeper active-layer depths close to the rivers and also from downstream seed dispersal (Neilson et al., 2005; Wieczorek et al., 2017). Nevertheless, the positive impact of an increased survivorship on migration rates can be observed in our migration simulation experiments.

The mortality rate ahead of the treeline is lower under homogeneous climate than in the linearly decreasing climate gradient scenario with the consequence that the migration enters the exponential phase earlier (Fig. 6 & 7). In addition, we based our model adaptations on an area that is only one hectare in size and with this we cannot directly assess the long-distance seed dispersal to which to fit our implemented kernel. To account for these cases, we implemented a Gaussian dispersal kernel combined with an exponential shaped with a fat tail (Kruse et al., 2016). In this study, this allows numerous seeds to be dispersed to far distances and led to a higher immigration into the simulated forest plot than observed. In consequence, the simulated migration rate tends to be overestimated.

This comprehensive study from genetic analyses to a model application is a first attempt showing the importance of undertaking these timely model parameterisation studies and should be enhanced by, for example, inferring the parentages for other positions in the treeline ecotone on the southern Taymyr Peninsula.”

Here are some other comments/questions/edits:

What makes LAVESI a spatially explicit model? It would be good if the authors could explain this in a few sentences. Even though the model has been previously published, it helps orient the reader to explain the model and what makes the model unique.

We added a short descriptions what our model makes it an individual-based model and explained the advantages of such a detailed approach.

The edited text can be found in line 58ff:
“To study the responses and migration dynamics of treeline tree stands under climate change, LAVESI, an individual-based and spatially explicit simulation model for *Larix* (Kruse et al., 2016; Wieczorek et al., 2017), was developed. In comparison to other dynamic vegetation models, it handles each individual larch tree beginning from a seed to an established seedling until becoming a mature tree and producing seeds itself and thus starting a new generation. This model includes wind-

> **dependent seed dispersal and density-dependent growth and mortality processes. The representation of the full life cycle allows in-detail simulation experiments to unravel the influences of previously overlooked feedbacks (further details in Kruse et al., 2016; Wieczorek et al., 2017).”**

This parameterization as well as the improvements made to LAVESI concerning seed dispersal rates and distances were made based on data collected over a 100m x 100m plot. The size of this plot is quite small to base landscape scale conclusions on. The disadvantages of this plot size are not well discussed in the discussion.

> **We extended the discussion about the plot size of 100x100 m. This area is at the upper edge to be manageable during expeditions to these remote areas. Several people needed days to record and sample these <1000 individuals. However, at more densely populated forests plots we sampled >3000 individuals at similar areas or even on smaller plots.**
>
> **We decided to use the northernmost plot close to the species line as this is the likely area responding most strongly and very likely “preparing” for a northwards migration triggered by recent climate warming.**
>
> **A larger area does mean more work and we think that the knowledge gain does not scale with effort. We added here sentences and also under 4.3 first paragraph at the end.**
>
> **In line 284ff:**
> **“Unfortunately, the labour-intensive sample collection and genetic analyses restricted the analysis to a rather small area in comparison to the large area of the treeline transition zone. Assessing the parentage across a broader scale and for different positions in the treeline ecotone would further help to understand dispersal dynamics at the treeline but the additional knowledge gain does not scale with effort.”**
> **In line 378ff:**
> **“This comprehensive study from genetic analyses to a model application is a first attempt showing the importance of undertaking these timely model parameterisation studies and should be enhanced by, for example, inferring the parentages for other positions in the treeline ecotone on the southern Taymyr Peninsula.”**

What about topography? Topography is not mentioned and is a very important feature with respect to treeline advancement, seed dispersal rates/distances and seed viability. Microsite climate effects caused by topography are also not addressed. These too are very important to consider here.

> **See our response to the general comment above.**

The methods surrounding how the needle genotyping (2.2) was used to infer microsatellite data (2.3), and was then used to update seed disperal rates and distances in LAVESI are very confusing. It is unclear what was simulated and how, versus what was measured in the field. For ex., line 108, “We simulated the heritage for 10,000 seeds...” How was this simulated? With LAVESI? With a statistical model? With CERVUS 3.0.7? These sections are very confusingly written, readers would not be able to use them to reproduce your study. Please explain more clearly the steps that were taken to go from needle collection to LAVESI updates.

We checked the sequence of the regarding methods and edited section 2.3. to made more clear that the observed parentages were estimated in the program CERVUS and not with our model LAVESI. Following the first method sections about the field data and subsequent analysis until estimating effective seed dispersal distances, we introduce the model tuning steps in section 2.4.2.

Line 115ff:
“We determined parents from allele frequency data with a likelihood-based approach implemented in CERVUS version 3.0.7 (Kalinowski et al., 2007). During the analyses, we allowed for 1% of errors in genotyping and a minimum of seven loci typed in the final analysis. All individuals (612 in total) were analysed and we searched for parents of recruits (height <2 m) from among all potential tree individuals (height >0.4 m). Following the program documentation we simulated in CERVUS the heritage for 10,000 seeds with a chance of 10% of a parent sampled and 1% error (Marshall et al., 1998; Slate et al., 2000) to determine thresholds for the ‘log of the overall likelihood ratio’ (LOD) scores in this analysis.”

Lines 19-24: The writing is not clear whether the comparison was done before the model code updates or if the model was run on transects to address the shortcoming.

We clarified which model version we used for the transect simulation.

Now text in line 21f:
“We thus adapted our model and used the newly parameterised version to simulate south-to-north transects: a slow-moving treeline front was revealed.”

Section 2.4.1: The model though published elsewhere should be explained in a few more sentences here. Why is it considered spatially explicit? What does that mean for this study in particular? How were the listed updates implemented?

We added the requested details in the introduction. In addition, we edited the methods section 2.4.2, but we refer the reader to the supplement 2 for the technical description of the model tuning by modifying parameters or newly introduced variables.

Line 58ff:
“In comparison to other dynamic vegetation models, it handles each individual larch tree beginning from a seed to an established seedling until becoming a mature tree and producing seeds itself and thus starting a new generation. This model includes wind-dependent seed dispersal and density-dependent growth and mortality processes. The representation of the full life cycle allows in-detail simulation experiments to unravel the influences of previously overlooked feedbacks (further details in Kruse et al., 2016; Wieczorek et al., 2017).”

Line 145ff:
“To tune the model’s processes in order to capture the observed effective seed dispersal distribution, we tested several combinations of model parameters and introduced new variables into formulae used in the program code of the model (listed in Table 1, details in Supplement 2 and in Table S5).”

Line 169: “Simulated” is more colloquial terminology than “hypothetical”

> **Done, changed to “simulated”**

Lines 190, 195-196: Are these two different results? What is the difference between “pairs of larch individuals” and “two individuals within a clonal group”?

> **The 11 individuals are those that are the excluded individuals from further analyses, which were part of the 10 clonal groups consisting of 22 individuals (9x2 and 1x4 individuals). We edited the text for clarification.**
>
> **Line 210ff:**
> **“In total, 601 sampled trees could be distinguished and 22 individuals were identified as 10 clonal groups, of which 11 were subsequently excluded from further analyses (Fig. 3a, Supplement S1). The maximum distance between two individuals within these groups was 30 m but mostly <5 m (Fig. 3a).”**

Line 237: Overemphasize is misspelled

> **Done**

Lines 322-332: Were these other study’s all simulated or field-based results?

> **They were all field-based studies and we added a reference to that in the sentence for clarification.**
>
> **Line 346ff start now with:**
> **“Another field-based study reports […]”**