Dear Kirsten,

Thank you for your kind response.

We included all of the changes made to the manuscript and the updated figures. Please find below the point-by-point responses to your comments and those of the 4 reviewers along with the marked-up manuscript version as requested.

In the point-by-point responses, our response are placed in bold font below each of the reviewer's comments in italics. Followed by a citation of the changed text with a line statement that refers to the version of the manuscript with tracked changes.

With kind regards on behalf of the authors,

Stefan

Associate Editor Decision: Publish subject to minor revisions (review by editor) (21 Jan 2019) by Kirsten Thonicke

Comments to the Author:

1. in response to the reviewer comment (AC1): "Page 9, lines 195 f: Is there an explanation how ramet pairs can occur 30 m apart in Larix? How about the chance for full sibs to have identical genotypes?" you gave a detailed response, incl. additionally plotted data (Fig. R1). Please consider if your response can be partly included in the main text with the figure eventually included in the supplement.

> **We included large parts of our corresponding response to the main text in the discussion section 4.1. and added the additional figure the supplement (Fig. S6).**
>
> **"In this analysis, we needed to exclude the observed clonal groups that are consequences of exceptional reproduction. We are confident that these are true observations of clones as we minimized the chance of full sibs share the same genotype by using highly polymorphic nuclear microsatellites that are not in linkage disequilibrium (Kruse et al., 2018). Nevertheless, we cannot rule out that selfing or back-crossing have occurred that could yield to offspring being genetically identical to one of the parents. If those modes of inheritance regularly occur and would have caused a misidentification of full siblings as clones, we would expect to observe an continuously increasing number of transitional states from identical genotypes (0 different alleles) to sharing 50% of their alleles (8 different alleles). However, it sharply drops from the identified clonal groups to a very low value and increases again beginning at 3 to 4 differences (Fig. S6 in Supplement 1). This gives us confidence to classify such identical individuals as clones. An explanation for these could be, that wind thrown trees can survive or in non-favorable conditions producing horizontal branches rather than upright stems forming krummholtz (own observations, Wieczorek et al., 2017). By producing adventitious roots from branches touching the ground (Kajimoto, 2010; Cooper, 1911) and subsequent separation of the main stem or horizontal branch two genetically identical individuals can be found if both parts survive."**

2. Referring to the model description, AC3 asked for an explanation of what makes the model spatially explicit. In your response you explain the difference to dynamic vegetation models by explaining that you simulate plant growth from seed to sapling. But the explanation should also contain a sentence of what makes the model spatially explicit, i.e. a simulation grid of a certain cell size and how the adjacent grid cells are connected. Please double-check your model description in the methods section.

> **We checked the model description in section 2.4.1. and extended it by a sentence clarifying why this model is spatially explicit.**
>
> **"On a simulation area of user specified size, individuals grow where seeds settle and germinate, and competition among individuals is handled by a fine sub-grid of cells with an area of 20 x 20 cm."**

3. Please include the response to the reviewer comment (AC4) in 1 or 2 sentences in the methods section where you shortly describe the model: "Line 127: What is the ecological basis for changing the density competition to improve the on-site recruitment ratio? "

> **We included a short description about the simulated seed dispersal and how the effective seed dispersal distances can be adapted by changing model parameters.**
>
> **“With this and the detailed representation of competition the model realistically simulates, similar to Janzen’s (1970) and Connell’s findings (1971), that recruits have the highest chance to survive at intermediate distances to the producing tree, not directly at it. Fine-tuning the model parameters of involved processes, which includes the impact strength that competition has on smaller trees, allows adapting the effective seed dispersal distance.”**

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 #1

Received and published: 19 September 2018

We thank the reviewer for the review and 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:

The manuscript by Kruse et al. describes an empirical study of effective seed dispersal using molecular markers which is used to adapt and parametrize a simulation model on larch migration rates at the arctic treeline. The topic of this study is of general interest because estimates for the capability of species to shift their distribution ranges are important for assessing the impact of climate change on many ecosystems. Especially tree species are of interest since they are the foundation species of many ecosystems and exhibit life history traits, which make direct observations difficult. The study offers a nice example for a combined approach with empirical data and simulations although its direct implications are somewhat limited due to the lack of replications and the rather small study site. However, the authors acknowledge these limitations in their manuscript and it will be nice to see some replicates in the future to judge the range of migration rates possible at different locations.
The manuscript is overall well written and clearly structured. The applied methods are well chosen and experiments and data analyses are described in sufficient detail. The results are discussed in a concise way using the available body of literature and conclusions are well founded.
In general the manuscript is of high quality and I have not found any major flaws. Please find below a few specific comments.

Specific comments:

Page 2 lines 53 ff: These sentences are a bit hard to follow. Also it does not immediately become clear that the authors refer only to the Taymyr peninsula. Since there is so much literature on this available, it should maybe added somewhere that this region is well studied, which is a further argument for chosing this region for the study.

We added a reference to the region in focus here and tried to point out in the preceding sentence that this region was in focus for several treeline studies.

Now the text in line 52ff is:
“It represents an ideal study area because the treeline is formed of monospecific tree stands of *Larix* Mill. Taxa and was thus the focus of several treeline studies (IPCC, 2013; Naurzbaev et al., 2002; Sidorova et al., 2010). The response to warming seems to differ with time-scale:

while millennial-scale warming during the mid-Holocene is reflected by a treeline location 200 km further north on the Taymyr Peninsula (Andreev et al., 2002; Klemm et al., 2016; MacDonald et al., 2008), the decadal-scale ongoing warming generates no response (Niemeyer et al., 2015; Wieczorek et al., 2017), possibly because of low seed availability."

Page 3, lines 70 ff: I suggest to elaborate a bit more on the specific aims of the study here at the end of the introduction. Some aspects have been mentioned in earlier paragraphs but rather indirectly and not specifically related to this study.

As suggested, we edited the final paragraph of our introduction to enhance the visibility of the specific aims of our study.

Lines 75ff are now:
"With this study, we aim at improving seed dispersal and establishment processes in the simulation model LAVESI to make it applicable for simulating treeline migration rates. Therefore, we undertook a genetic parentage analysis of a treeline stand on the southern Taymyr Peninsula by applying an assay of eight nuclear microsatellites to get a reliable estimate of the effective seed dispersal distance (1). This information was used to improve the individual-based model LAVESI (2), which we then ran to simulate treeline advances into the tundra and estimate migration rates (3)."

Page 9, lines 195 f: Is there an explanation how ramet pairs can occur 30 m apart in Larix? How about the chance for full sibs to have identical genotypes?

We carefully checked the possibility of having sampled the same tree again and are confident to not have such a sampling error in our data.

Indeed, it seems a rather far distance, however, further own observations support the existence of ramets even on a longer distance than under the crown.

1. Wind thrown trees can survive (own observations) and most probably produce adventitious roots from their branches when touching the ground (Kajimoto, 2010; Cooper, 1911). If the main stem rots, two separate individuals could be found if both survive.

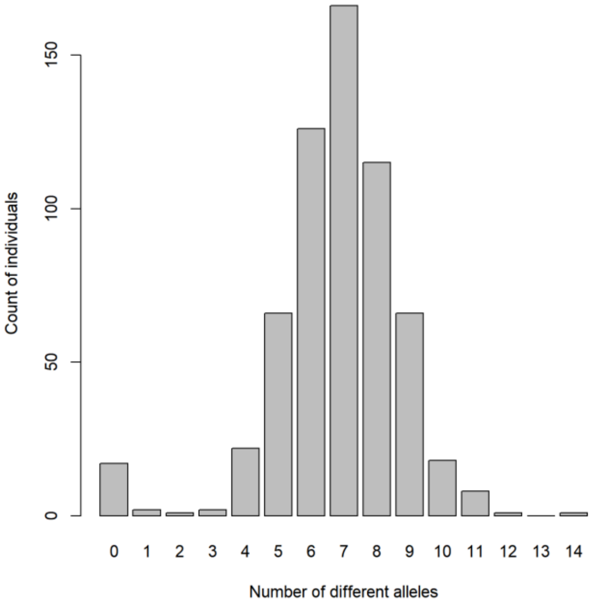
2. Under non-favorable conditions, larches can survive forming krummholtz individuals (Wieczorek et al., 2017, own observations). They grow by producing horizontal branches rather than an upright stem. At these, they could form adventitious roots and again after separating two individuals could be found sharing the same genotype.

Nevertheless, there is a chance that full siblings share the same genotype, but which is quite low especially because we used highly polymorphic nuclear microsatellites that are not in linkage disequilibrium (Kruse et al., 2018). Thus, the probability that two full siblings would have the same given genotype under pure Mendel inheritance (without mutations and recombination) is approximately ¼ per locus. In our case this leads to a low chance of 1 to ~65536 (¼ ^8) that full sibs share the same genotype.

Additionally, we cannot rule out that selfing or back-crossing have occurred that could yield to offspring being genetically identical to one

of the parents. If those modes of inheritance occur regularly and would have caused a misidentification of full siblings as clones, we would expect to observe an continuously increasing number of transitional states from identical genotypes (0 different alleles) to sharing 50% of their alleles (8 different alleles). However, it drops from the identified clonal groups to a very low value and increases again beginning at 3 to 4 differences (Fig. R1). This gives us confidence to classify such identical individuals as clones.

Figure R1. For each individual the smallest number of different alleles, binned into 0 to a maximum of 16 alleles.



References:
Cooper, W. S., 1911. Reproduction by layering among conifers. Botanical Gazette 52(5):pp. 369–379.
Kajimoto, T., 2010. Root system development of larch trees growing on siberian per- mafrost. In A. Osawa, O. A. Zyryanova, Y. Matsuura, T. Kajimoto & R. W.Wein, editors, Permafrost Ecosystems, volume 209, pages 303–330. Springer Nether- lands, Dordrecht.

Page 15, line 320: When I read this, I asked myself if the model includes the case of established individuals ahead of the treeline, which are not able to reproduce, yet, because conditions do not allow this at the moment. When the conditions change, the treeline might progress quite rapidly at first and then slow down. Since I am not familiar with the model in detail, I cannot judge if this is a point worth discussing or a scenario worth simulating.

For simplification and to clearly infer migration rates into tundra we did not allow in our transect simulation experiments survival of individuals ahead of the treeline until year 100.

At some places the presence of krummholtz may enhance migration if such tree island/refugials begin to reproduce sexually. Although not explicitly incorporated, this is partly covered by the homogenous forcing climate that allows long dispersed seedlings to survive with a higher chance ahead of the treeline than in the climate gradient scenario forming faster forest islands within the tundra (see simulation example

in Fig. 2). At the beginning of the simulations the migration rate is only slightly faster but this benefit accumulates over time until the positive effect can clearly be seen at the end of the simulation period (Fig. 6 & 7).

A detailed simulation study for a variety of latitudinal treelines might be worth considering in an extra study in which we could assess responses of different treeline types (e.g. sharp boundary vs. wide transition zone, and the presence/absence of krummholtz).

Page 15, lines 325 ff: The migration rates mentioned here, are they 20-60 m/20-50 m for the entire time period respectively or per year in these time periods? Is it possible to translate the elevational shift into a migration rate comparable to the model?

Migration rates at altitudinal treelines are hardly comparable to those of latitudinal treeline, the climate gradient is much steeper so that still seed sources are closer to the species limit so that climate improvements can lead to faster a migration response up the slope. In comparison, the same climate gradient is very likely thousand times longer on latitudes: 1 °C per ~150 m elevation compared to 1 °C per ~160.000 m latitude for the Taymyr Peninsula.

Because of not being strongly affected by dispersal limitations, they help to understand how treelines could ideally migrate when not limited by seed availability. Nevertheless, other restrictions might become more important for the migration process at these locations such as facilitation (e.g. Martínez et al., 2011).

References: Martínez, I., Wiegand, T., Camarero, J. J., Batllori, E. and Gutiérrez, E.: Disentangling the formation of contrasting tree-line physiognomies combining model selection and Bayesian parameterization for simulation models., Am. Nat., 177(5), E136–E152, doi:10.1086/659623, 2011.

Page 15, line 330: Establishment will for sure be affected not only by density-dependent mortality but also by abiotic conditions and their stochasticity in this extreme region of the planet.

That is right, therefore establishment of seeds dependent on weather forcing and their survival (mortality) implemented as a stochastic process (see details in Kruse et al., 2016).

Technical comments:

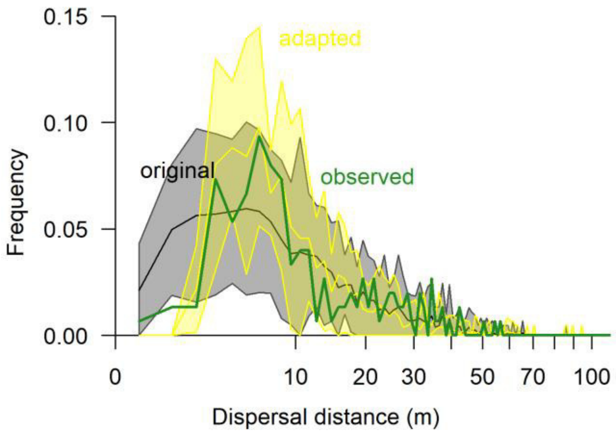
Page 5, line 102: “inferred” Why are the microsatellite data described as inferred? To me they seem quite directly measured.

The reviewer is right, the fragment lengths were measured and the parentage was inferred from these data. Accordingly, we deleted the word “inferred”.

Figure 5: The x axis is quite cramped in this figure. Maybe it could be stretched out a bit?

For a better visibility of the plotted results, we show now lines instead of points and stretched the scale of the x-axis.

The figure in line 236 is now:



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 #2
Received and published: 26 October 2018

We thank the reviewer for reviewing the 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 paper uses the LAVESI model to explore the rate of northward Larix migration on the southern portion of the Taimyr Peninsula in Siberia. The study uses recent field data collected in the area to drive the simulation work. The focus of the model parameterization/calibration was to improve the Larix dispersal functionality using genetic parentage, and then apply the model to understand how the Larix dispersal characteristics will play out in given some climate scenarios. The site of this modelling work is of particular interest, as it is near the northernmost forest stands where paleoecological records have indicated past presence of trees north of where they are currently found.
Given the unique nature of Larix forests (deciduous conifers linked to continuous permafrost distribution) and the fact that across their broad spatial domain (central and eastern Siberia) this class of trees coincides with marked changes in climate, and the potential for changes in tree distribution to alter the dynamics of high latitude systems (through changes to albedo and permafrost), this study is of great interest. I think the paper is for the most part clear and well-structured. Perhaps there can be some modification of the Discussion based on my main critique.

Main critiques:

1. At the scale of this study, are south-north assumptions of tree migration robust? At fine scales, the migrations may occur according to the patterns of favorable microsites, and primarily be confined to corridors with favorable active layer dynamics, and direct insolation. Such landforms seem important for explaining the current pattern of trees. A study with such an individual model that doesn't account for the conditions that are associated with the germination, survival, and growth of those individuals that are being dispersed should probably discuss in some detail this issue of the micro-site constraints that may contribute to broad error bars associated with migration rates. In other words, would the velocity of migration change between different microsites? If so, what is the relative prevalence of such favorable sites across the landscape, and how are they connected to the seed sources? These may be questions for follow-up work, and may be beyond the scope of this study, but I think a section in the Discussion could serve as a link between this study and some potentially viable next steps (one that incorporates landforms and micro-sites details).

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 R3 General comments.

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.”

2. This area is the northernmost forest ecotone. Some discussion for why this may be the case (paleoecological history) could be interesting and help contextualize predicttions of future treeline velocities.

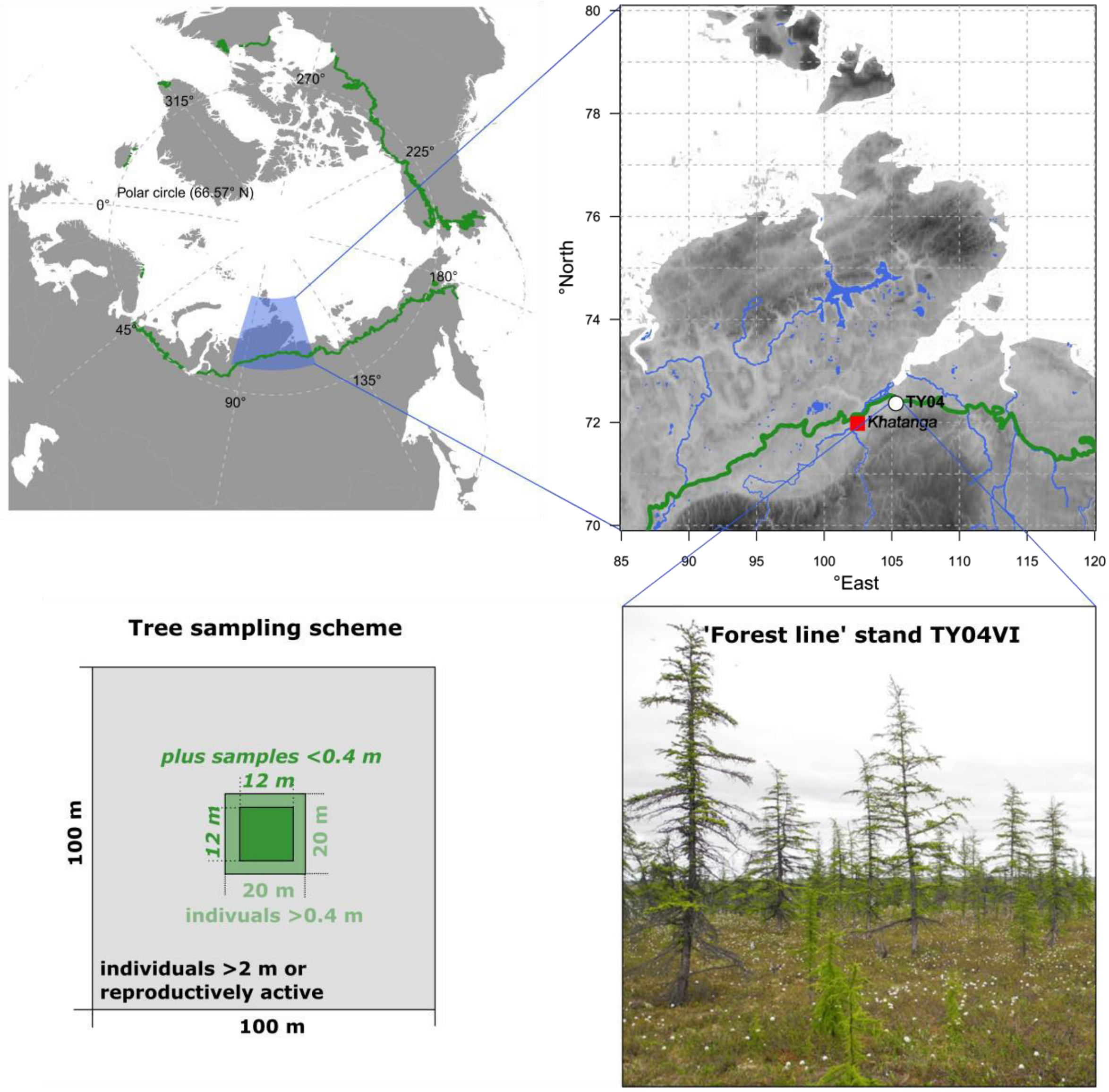
We added a short history of the treeline at the Taymyr Peninsula and giving likely explanations for the northward expansion and rate, ending with the modern situation.

In lines 350ff:
“During the Holocene Thermal Maximum boreal forests expanded on the Taymyr Peninsula to their northernmost position during the Holocene, which was likely assisted by glacial refugial populations ahead of the treeline (MacDonald et al., 2000, 2008). The treeline responded with a centennial lag to environmental improvement, for example solar insolation, and reached its maximum position at ~8000 to 4000 yr BP, and subsequently declined to reach its modern limits around 3000 yr BP (MacDonald et al., 2000). Recently, global warming is ameliorating conditions for *Larix* forests in Siberia and evidence can be found that treeline stands are starting to respond, but at a slower rate than one might expect given the strong increase in temperatures (Wieczorek et al., 2017; Harsch et al., 2009).”

3. In Figure 1, a simple schematic of the sampling design could be useful.

We added a simple sampling scheme of trees for the study site in Fig. 1.

The figure now is in line 93:



Other comments:

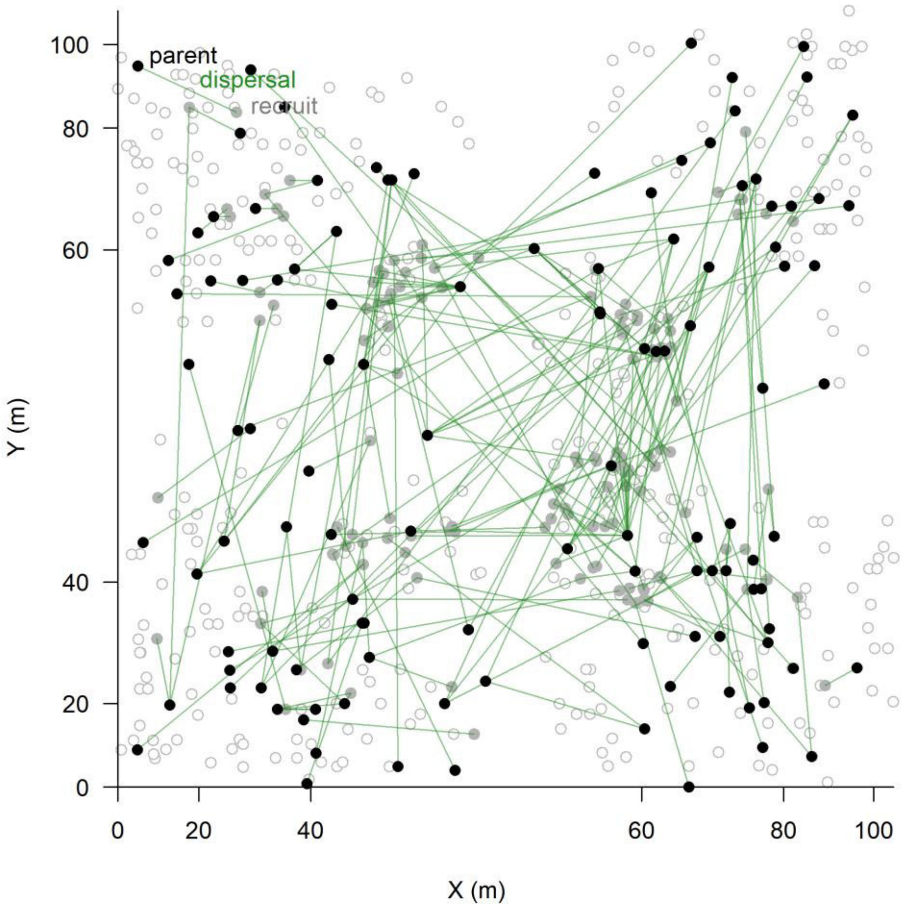
In the abstract, might it be possible to replace some of the technical wording associate with the genetic analysis with other more recognizable terminology that would be more likely to be understood by most of the readers of this journal?

This comment refers to the sentence in line 17. Following the suggestion of the reviewer, we edited this by deleting "highly polymorphic", which is already a requirement for the parentage analysis that we meet with our eight microsatellite loci. Additionally, we exchanged the word "loci" and used the synonym "marker".

The text in line 16ff is now:
"We inferred the effective seed dispersal distances of a typical open forest stand on the southern Taymyr Peninsula (north-central Siberia) from genetic parentage analysis using eight nuclear microsatellite markers."

Figure 4. Nice figure that shows how a few individuals dominate the reproduction. I think it needs to be graphically enhanced. Suggestion: Put an alpha (i.e., transparency) on the green lines, and put the dots on top.

As suggested we changed the transparency of the dispersal connections. Furthermore, we use now lines instead of arrows for a clearer view on the connections. Bringing the dots to the front caused many connections to be hidden. So that we decided to modify the coordinate system by stretched it to zoom in to the plot centre where most recruits and their connection to the parents can now be seen.



Section 4.3 Line 336 - “However, further processes...and should therefore [NOT?] be neglected in simulation studies:...”

Yes, this is a critical point for us. We corrected this mistake in writing.

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 […]”

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 #4
Received and published: 22 November 2018

> **We thank the reviewer for reviewing our manuscript and especially for a closer look on the supplementary material. The comments helped to improve the first version of our manuscript. This was revised 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 utilizes genotyping and parentage analysis of individual trees to improve larch seed dispersal simulation within an individual-based, spatially-explicit forest model. The study is carried out at a single 100 m x 100 m site in the Taymyr Peninsula in northern Siberia. LAVESI, the forest model used, is specifically designed for individual larch growth, mortality, and regeneration, and the updated model is used to simulate northward migration of the larch treeline and forestline under two different climate scenarios. The updated model performed well when compared to observation data, though it slightly overestimated the number of recruits close to the parent tree as well as an overestimation of very long dispersal. The south-north migration simulation under static climate resulted in a migration rate of 0.6 m/year and 1.6 m/year for the forest- and treelines, respectively. Under a climate scenario of decreasing temperature and slightly increasing temperature from south to north, the south-north migration rate was slower. They also found an accelerating rate of dispersal over the simulation time under the static climate scenario.
The study is important for field ecologists as well as the ecological modeling community. Currently, northward tree migration across the circumpolar boreal region is of crucial importance due to its potential impact and feedback to climate. However, most forest models do not adequately represent dispersal mechanisms. This study showcases an innovative way to determine in situ effective seed dispersal and incorporate such data into a forest model for calibration and application.

While the study is effective and well-structured, and shows how well the LAVESI model can perform at a local-scale, the model was tuned quite heavily to the small study area (only 100 m2), and the model output was compared only to data that was used in the tuning process. Before this model can be utilized at a larger scale I believe it will require more generalized parameter values. In particular, because the model produced fairly slow migration rates compared to other studies, I feel it may be overfitted to this study site and data, though only additional comparisons and simulations with the model will be able to determine if this is the case. It would be nice to see a sentence or two acknowledging this in the Conclusions. It would be nice in future studies to see this model compared to independent data at a separate site as *well. I would also be interested to see how the migration would play out under a climate change scenario, though this is likely planned for future work.*

Overall, I think this paper is well-written and the manuscript should be accepted with only a few minor revisions. This study is a great starting point for future work with this model and the equations developed within it. It should be of interest to other ecologists working on similar problems across the boreal region.

Response to the the centre part of the general comment in starting with "While […]". A similar comment came from R3. We added a short discussion about the "small" study area that is already challenging for such an analysis to the discussion in section 4.1. Nevertheless, it would be worth to undergo this work at more sites to compare the findings of this study to other treeline locations.

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."

Additionally, we extended our conclusion covering the comment on further studies that would help unravelling if our slow migration rate estimate is flawed by overfitting to only one study site or not, as requested by the reviewer.

Line 400ff:
"To find out if the estimated slow migration is an outlier coming from overfitting to only one study site or the general response rate under current warming, further similar studies at other treeline positions would be necessary."

Below are some minor comments and edit suggestions for consideration by the authors:

Line 92: Change "Subsequent" to "Subsequently,"

Response: Done

Line 120: Change "larch species" to "larch individuals"

Response: Done

Line 129: You say here and in the Supplementary Material that active layer depth influences tree mortality (which I am guessing is based on growth rate). However, it seems based on the information in the Supplementary Material that active layer depth directly influences tree growth, which in turn would also influence mortality (and potentially seed dispersal?).

We use the actual tree growth in comparison to the maximum potential growth of the same tree as currency for productivity and mortality.

The given information was not sufficient to explain how active layer depth influences trees (growth/mortality). In consequence, we edited the text in the Methods section for clarification.

Line 137ff:
“The original model of Kruse *et al.* (2016) was updated with the following processes (details in Suppelement 2): (i) seed dispersal distances now depend on species-specific traits (tree height, seed properties) and wind speed and direction (Kruse et al., 2018b), (ii) the tree diameter growth function is newly calibrated to the climate forcing (Epp et al., 2018), and (iii) the active-layer thaw depth directly influences the tree’s growth that is used to estimate it’s seed production and mortality.”

Lines 135-140: I’m not sure why some of these parameter descriptions are in quotes and some aren’t. In general this sentence is difficult to get through. You may want to consider just publishing a table instead of listing them in the text.

For clarification we decided to remove the listing of only some of the varied parameters and refer readers to the complete information in the supplementary material. A complete list and further detailed information on each parameter combination and the process can be found there.

Line 139: I’m not sure what “different modes to compute the competition” are

We tested the impact of several implementations of influence areas and strengths of competition on the trees diameter growth. The actual growth of an individual is the currency in the model by which other functionalities are based on (seed production/mortality).

We refer now the reader to the supplementary material, as there is the information on modified parameters/modes and tested model variables.

Line 151-152: Could you expand on the 20mx20m vs. the surrounding 100mx100m section? I’m not sure I follow where the spatial differences are coming from.

We needed to make the simulated data comparable to the inferred effective seed dispersal distances. Therefore, we followed directly our sampling scheme as described in Section 2.1 sample collection “[…] We sampled all individuals >0.4 m in height in a 20 x 20 m area as well as all trees >2 m high or bearing cones from the surrounding 100 x 100 m area (Fig. 3). Additionally, in the central 12 x 12 m area individuals <0.4 m were collected.”

Here we added a reference to the sampling scheme description in section 2.1 in the regarding sentence.

Line 166ff:
“We resampled these simulated distances to consider the same frequency of observed parenthoods in the central 20 x 20 m as in the surrounding 100 x 100 m area (sampling scheme details in section 2.1 sample collection).”

Line 224: Add “for this model” after “Mean dispersal” Line 229: change “have the smallest” to “has the smallest”

Response: Done

Supplement S2:

Line 74: Change "correspondingly" to "corresponding" and delete "roughly"

Response: Done

Line 76: Change "of Matlack" to "from Matlack"

Response: Done

Lines 76-79: I'm confused by what 0.86 m/s is referring to. I this Vd? Or w? Additionally this sentence is some- what awkward and I would recommend breaking it up into two sentences and clarifying.

The value 0.86 m/s is referring to the descent rate for seeds, which is abbreviated by Vd. We separated the sentences as suggested and edited it for clarification.

The corrected part of the text is now in line 74ff:
"The release height H_t is estimated at 75% of the individual's height. V_d is the descent rate for seeds and is estimated for *Larix gmelinii* by a linear regression using species data from Matlack (1987). For species having wing-scales attached to the seeds, this rate can be calculated by $V_d = 0.0032 * \sqrt{w} + 0.4807$ and is 0.86 m s^{-1}, with the wing loading w (Matlack 1987) for *L. gmelinii*. The variable w is calculated by dividing the average seed weight (in microdyne) of 3.5 mg (Heit and Eliason, 1940; Lukkarinen et al., 2009) by the propagule area of 0.2 cm^2 (Fu et al., 1999)."

Line 84: how did you obtain the sdist and the scaling parameter? I see that you tuned them variously but did you have initial starting values based on literature or data?

When implementing the seed dispersal kernel into the model (Kruse et al., 2016), we made a first guess for the resulting dispersal kernel based on literature values and tuned those values to observed patterns.

Line 88: Where did you obtain the data for the study showing no significant influence of temperature? Was it at the same study site? I am concerned about this growth function modification as it further "tunes" the model to a specific area, and may need to be re-tuned if the model is moved elsewhere

We used a tree ring series from Yamal of the National Climatic Data Center data bank for *Larix sibirica* and own data for *Larix gmelinii* from Khatanga near the study site and for both data from the nearest weather station. For further information, please see the supplement of Epp et al. (2018) published in Scientific Reports.

Regarding the second part of the comment. The modelled tree diameter growth in the current version of the model is adapted to weather in Taymyr and Yamal. Therefore, it has to be tuned for each species and region when using it for further applications.

Lines 93-98: See my above comment on permafrost-tree growth influence. It seems ALT impacts tree growth directly and mortality indirectly, though I may be wrong.

Yes, answered in the other comment above.

Line 97: What is the parameter fe?

> **It is a soil property parameter, see definition in Hinkel and Nicholas (1995).**

Table S4: I would suggest also adding variable symbols next to the parameter descriptions, especially if they are mentioned in this text or other published works.

> **We added for the model parameters the corresponding symbols. Corresponding changes were made in Table S5.**

Line 105: Why do you need to shift the dispersal peak by 2-3 m? Is this based on comparisons with the observation data? I would mention this here.

> **We tried to explore potential setting to align the modelled effective dispersal distances to the observations. For clarification, we edited the sentence and refer to the results presented in the main article.**
>
> **The text now line 105 reads:**
> **"To fit the simulated seed effective dispersal distance to observations (Fig. 5) we explored potential settings …"**

Line 120: What is the reference simulation? Additionally please expand on what you mean by "general performance."

> **We extended the statement of the reference simulation, which is the baseline simulation with the original model. Furthermore, we added for clarification of the "general performance" a reference to the correlation coefficients in Table S5.**
>
> **This sentences in line 122ff changed to:**
> **"This was improved by other simulations (qt-wJ) but their general performance (lower correlation coefficients, Table S5) was weaker than the reference simulation without parameter changes or adaptations of the model (a)."**

Line 123: I'm not sure what you mean by "In parts"

> **We deleted the confusing beginning of the first sentence of the regarding paragraph. In the following sentence we briefly state the achievements, but also at which results the best fitting model version deviated from the observed pattern.**
>
> **Text now in line 126:**
> **"We achieved a good fit when increasing the peak of the dispersal function in the model to longer distances."**

Line 127: What is the ecological basis for changing the density competition to improve the on-site recruitment ratio?

> **Similar to Janzen and Connell's findings, recruits have the highest chance to survive at intermediate distance to the producing tree, not directly at it. They are "pushed back" by the mother tree for a variety of reasons (shadow of the tree's crown, high pest pressure/seed predators,**

exhausted nutrients in the active layer, insulating accumulation of needles and other litter, etc.).

This is implicitly implemented in the model and can be manipulated by varying the competition density, e.g. by increasing the influence on smaller trees. With this, seedlings from farther distances could have a likely higher chance to establish.

Line 128: Delete "were" in between results and strongly

Done

Dispersal distances and migration rates at the arctic treeline in Siberia – a genetic and simulation based study

Authors: Kruse, Stefan[*1,2], Gerdes, Alexander[1], Kath, Nadja J.[1,2], Epp, Laura S.[1], Stoof-Leichsenring, Kathleen R.[1], Pestryakova, Luidmila A.[3], Herzschuh, Ulrike[1,2,4]

[1]Polar Terrestrial Environmental Systems Research Group, Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, 14473 Potsdam, Germany
[2]Institute of Biology and Biochemistry, University of Potsdam, 14476 Potsdam, Germany
[3]Institute of Natural Sciences, North-Eastern Federal University of Yakutsk, 677000 Yakutsk, Russia
[4]Institute of Earth and Environmental Science, University of Potsdam, 14476 Potsdam, Germany

Correspondence to: Stefan Kruse (stefan.kruse@awi.de)

Abstract. A strong temperature increase in the Arctic is expected to lead to latitudinal treeline shift. This tundra-taiga turnover would cause a positive vegetation-climate feedback due to albedo decrease. However, reliable estimates of tree migration rates are currently lacking due to the complex processes involved in forest establishment, which depend strongly on seed dispersal. We aim to fill this gap using LAVESI, an individual-based and spatially explicit *Larix* vegetation simulator. LAVESI was designed to simulate plots within homogeneous forests. Here, we improve the implementation of the seed dispersal function via field-based investigations. We inferred the effective seed dispersal distances of a typical open forest stand on the southern Taymyr Peninsula (north-central Siberia) from genetic parentage analysis using eight ~~highly polymorphic~~ nuclear microsatellite ~~loci~~markers.

The parentage analysis gives effective seed dispersal distances (median ~10 m) close to the seed parents. A comparison between simulated and observed effective seed dispersal distances reveals an overestimation of recruits close to the releasing tree and a shorter dispersal distance generally. We thus adapted our model and used the newly parameteris~~z~~ed version ~~it~~ to simulate south-to-north transects: a slow-moving treeline front was revealed. The colonisation of the tundra areas was assisted by occasional long-distance seed dispersal events beyond the treeline area. The treeline (~1 tree ha^{-1}) advanced by ~1.6 m yr^{-1}, whereas the forest line (~100 trees ha^{-1}) advanced by only ~0.6 m yr^{-1}.

We conclude that the treeline in north-central Siberia currently lags behind the current strong warming and will continue to lag in the near future.

1 Introduction

Changing climate forces species worldwide to migrate (Arctic Climate Impact Assessment, 2004; IPCC, 2013). This is exceptionally challenging for sessile organisms such as plants as they may strongly lag behind their moving climate envelope (Harsch et al., 2009; Loarie et al., 2009; Moran and Clark, 2012). Warming is particularly pronounced in the Arctic where the tundra-taiga ecotone demarks the transition from forest stands to treeless areas and which is expected to move northwards (Harsch et al., 2009; Holtmeier and Broll, 2005). Such tree range expansion is of major interest because the establishment of forest in the dwarf-shrub tundra would reduce the surface albedo and promote a positive feedback to global temperature (Bonan, 2008).

Trees migrate via seed dispersal and face several ecological barriers (Svenning et al., 2014): first, viable seeds need to be produced, second, these need to be dispersed and, third, seeds need to germinate and survive to grow to new individuals. This process, called ‘effective seed dispersal’ (Connell, 1971; Janzen, 1970), determines the speed and spatial pattern of a species’ response to climate change. For example, closely dispersed seeds and a long generation time result in a slow moving front, while a patchy pattern will form from many long-distance seed dispersal events (Clark, 1998; Nathan and Muller-Landau, 2000; Ritchie and MacDonald, 1986). Migration can speed up if there are relict trees from an earlier wider extent of forest which have survived in refugia ahead of the recent treeline (Stewart and Lister, 2001; Väliranta et al., 2011).

To project future species ranges, the potential migration rate under global warming is estimated via simulation studies (Kaplan and New, 2006; Roberts and Hamann, 2016; Snell and Cowling, 2015). However, these simulations depend strongly on the dispersal configuration of the model (Bhagwat and Willis, 2008; McLachlan et al., 2005; Stewart et al., 2010; Willis and Van Andel, 2004). Most empirical attempts to estimate historical migration rates are based on records of fossil pollen and macrofossils in sediment cores as indicators of species presence (MacDonald et al., 2008; Pisaric et al., 2001), but the interpretation is compromised because of a lack of knowledge about glacial refugia, particularly small ‘cryptic’ refugia that can be easily overlooked in the fossil record (e.g. Petit et al., 2008). Therefore, more reliable estimates of dispersal distances of tree taxa are needed in order to predict the treeline response under high-latitude warming (Snell, 2014; Snell and Cowling, 2015).

Understanding treeline changes on the southern Taymyr Peninsula is of particular relevance as the area is characterised by a strong warming trend (IPCC, 2013). It represents an ideal study area because the treeline is formed of monospecific tree stands of *Larix* Mill. Taxa and was thus the focus of several treeline studies (IPCC, 2013; Naurzbaev et al., 2002; Sidorova et al., 2010). The response to warming seems to differ with time-scale: while millennial-scale warming during the mid-Holocene is reflected by a treeline location 200 km further north on the Taymyr Peninsula (Andreev et al., 2002; Klemm et al., 2016; MacDonald et al., 2008), the decadal-scale ongoing warming generates no response (Niemeyer et al., 2015; Wieczorek et al., 2017), possibly because of low seed availability.

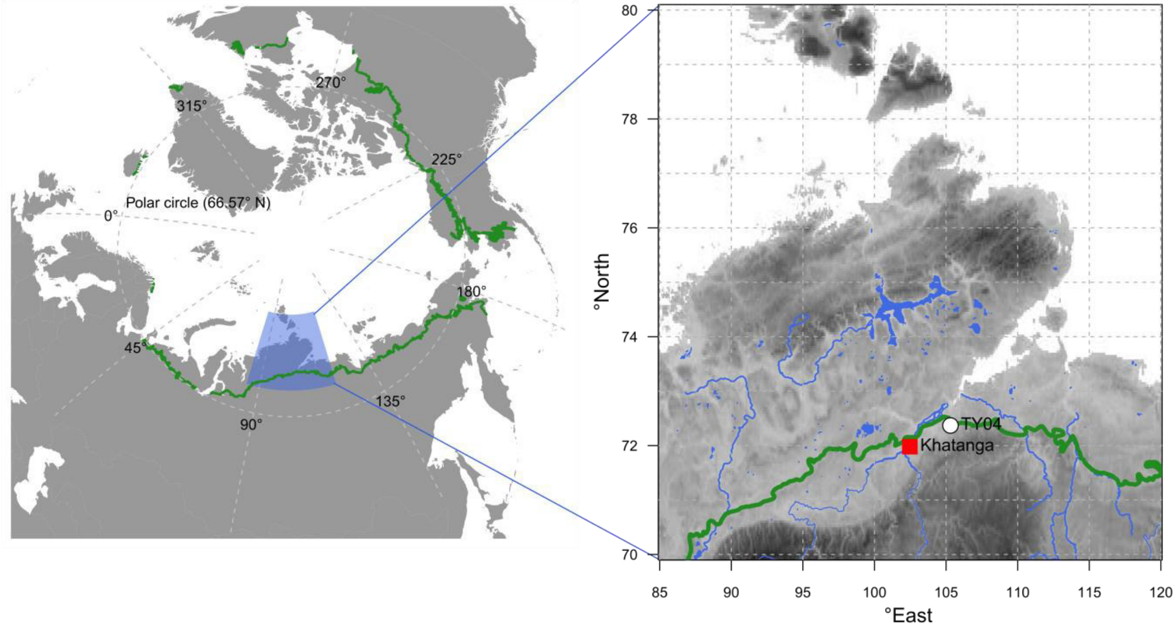
To study the responses and migration dynamics of treeline tree stands under climate change, LAVESI, an individual-based and spatially explicit simulation model for *Larix* (Kruse et al., 2016; Wieczorek et al., 2017), was developed. In comparison to other dynamic vegetation models, it handles each individual 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 ~~and~~ 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~~ling~~ influences of previously overlooked feedbacks (further details in Kruse et al., 2016; Wieczorek et al., 2017). However, the seed dispersal component had not been validated by observations. Traditional methods to track seed dispersal distances include seed traps and seed-bank analyses (Brown et al., 1988; Greene et al., 2004; Stoehr, 2000), which are time consuming and prone to underestimate distances (Ashley, 2010; Pairon et al., 2006). Fortunately, genetic analyses provide an alternative modern approach. Repetitive sequences in the nuclear genome (short sequence repeats, SSR, or microsatellites) are sufficiently variable genetic markers to resolve parentage (Ashley, 2010; Hartl and Clark, 2007; Schlötterer, 2000). Using such an approach, the dispersal of pollen and seeds in a landscape can be tracked and effective seed dispersal distances can be inferred (e.g. Pairon et al., 2006; Piotti et al., 2009; Pluess, 2011; Steinitz et al., 2011). For example, microsatellite studies have helped to elucidate the recruitment source of spruce juveniles and the dispersal patterns at an elevational treeline that recently shifted upwards (Piotti et al., 2009). Furthermore, a range expansion of larch following a glacier retreat could be tracked without a decrease in genetic diversity (Pluess, 2011). Genetic analyses can thus be used to provide a more realistic implementation of seed dispersal in simulation models.

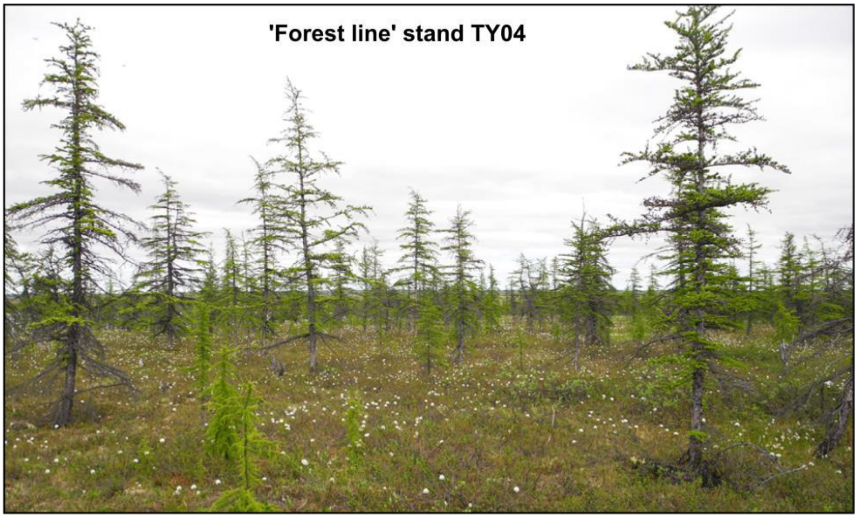
With this study, we aim at improving seed dispersal and establishment processes in the simulation model LAVESI to make it applicable for simulating treeline migration rates. Therefore, ~~W~~we undertook a genetic parentage analysis of a treeline stand on the southern Taymyr Peninsula by applying an assay of eight nuclear microsatellites to get a reliable estimate of the effective seed dispersal distance (1). This information was used to improve the individual-based model LAVESI (2), which we then ran to simulate treeline advances into the tundra and estimate migration rates (3).

2 Methods

2.1 Sample collection

Needle samples from larch individuals (*Larix gmelinii* (Rupr.) Rupr) were collected from a tree stand during fieldwork in the summer of 2013 on the southern Taymyr Peninsula, Krasnoyarsk Region, in northern-central Siberia (plot name: TY04VI; 72.409 °N and 105.448 °E; Fig. 1). The open canopy forest stand with ~300 trees ha^{-1} belongs to the forest tundra and has shown enhanced recent recruitment (site code FTe_1, Wieczorek et al., 2017). We sampled all individuals >0.4 m in height in a 20 x 20 m area as well as all trees >2 m high or bearing cones from the surrounding 100 x 100 m area (Fig. 3). Additionally, in the central 12 x 12 m area individuals <0.4 m were collected. Larch individuals from the 20 x 20 m plot were accurately mapped with a tape measure, while a standard GPS device (Garmin) was used to map the individuals in the 100 x 100 m area. We recorded the height of each individual and collected short twigs with needles and dried them in the field on silica gel.


270°
315°
225°
Polar circle (66.57° N)
0°
180°
45°
135°
90°
80
78
76
74
72
70
°North
TY04
Khatanga
85
90
95
100
105
110
115
120
°East


'Forest line' stand TY04

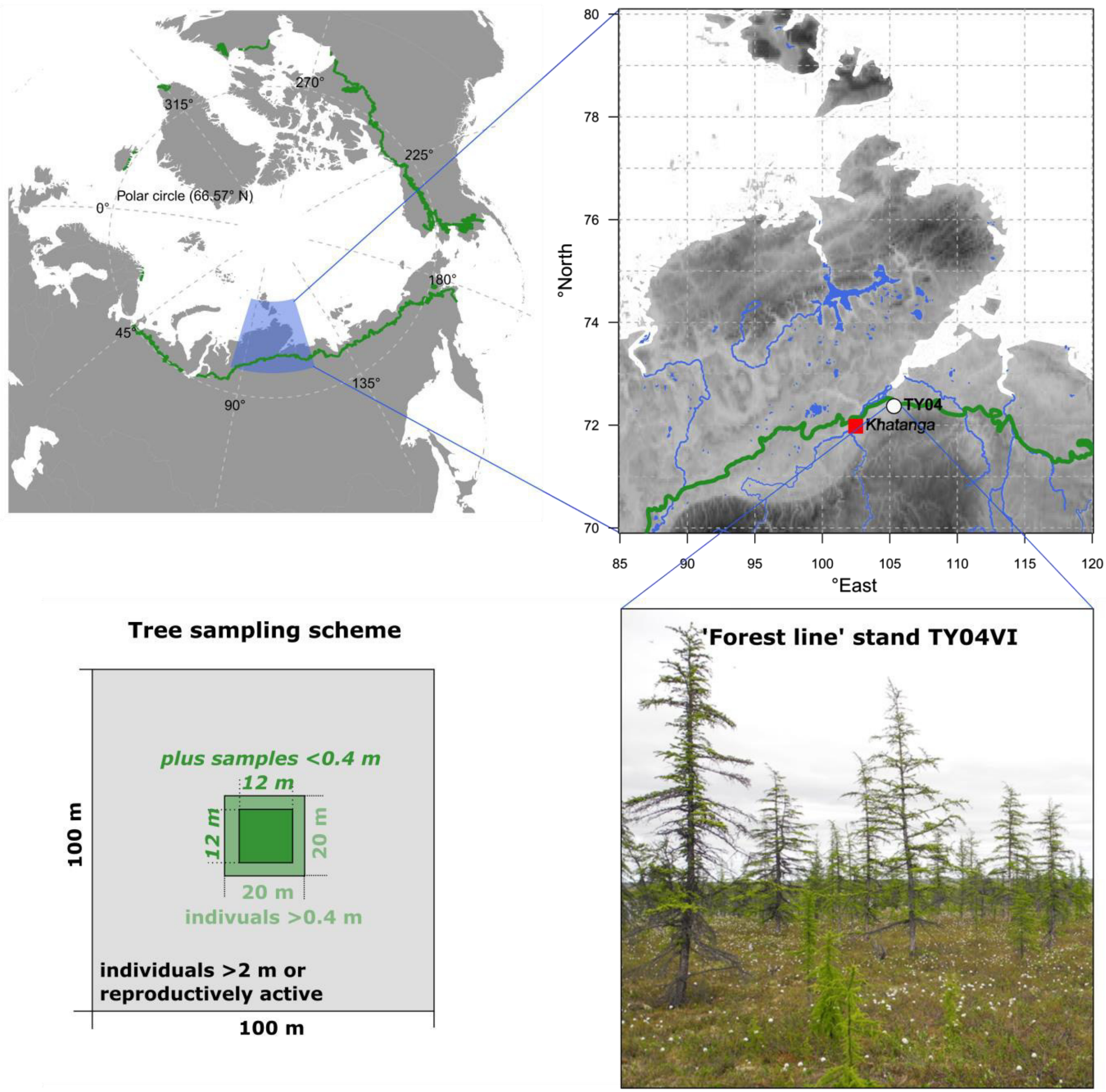


Figure 1: Overview of the larch forests (*Larix gmelinii*) growing at study site TY04VI on the southern Taymyr Peninsula and the sampling scheme. The green circumarctic line on the maps marks the modern treeline (Walker et al., 2005). Topography in the enlarged area ranges between 1 and 2521 m, (WorldClim1.4 Hijmans et al., 2005). Rivers and lakes are given in blue colours (GSSHS updated Version 2.2.2 01.01.2013 first published by Wessel and Smith, 1996). Photo: Stefan Kruse, 16.07.2013

2.2 Genotyping of individuals

Genomic DNA was extracted from 50–100 mg of dried needles after shock freezing in liquid nitrogen and grinding in a FastPrep® (MP BIOMEDICALS) with steel beads using the silica-column based extraction kit Invisorb® Spin Plant Mini Kit (STRATEC MOLECULAR). Subsequently, three multiplex PCR reactions (10 µl) were set up, including all eight primer pairs used in this study [bcLK211, bcLK253, Ld101, bcLK056, bcLK228, bcLK263 and bcLK189 (Isoda & Watanabe 2006) and Ld101, Ld42 and Ld56 (Wagner et al., 2012), further details in Supplement 1 Table S1] using the Multiplex PCR Master Kit (QIAGEN). Fragment length determination was done by SOURCEBIOSCIENCES (Oxford, UK). Allele sizes were scored in Geneious (version 7.1.5, BIOMATTERS LTD.) using the Microsatellite plugin (version 1.4.0). Raw allelic data were imported into R version 3.2.2 (R Core Team, 2015) and peaks binned to step sizes of two basepairs. The dataset was converted to the 'genind'-format using the package 'adegenet' (Version 1.4-2, Jombart, 2008; Jombart and Ahmed, 2011) for subsequent analyses. Details on the microsatellite primer selection, PCR-protocol and subsequent data analysis including binning of allele frequent lengths were described in Kruse *et al.* (Kruse et al., 2018a) and available online at https://doi.pangaea.de/10.1594/PANGAEA.885765.

2.3 Parentage analyses based on ~~inferred~~ microsatellite data

The effective seed dispersal function was estimated from the results of a parentage analysis. We used eight highly diverse nuclear loci; five of which were sufficiently informative to distinguish all individuals, as is mandatory for parental assignment studies (Figure S2 in Supplement 1). 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, w~~W~~e 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 ~~W~~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. Only those with positive assignments and a high confidence level exceeding 95% were retained for further analyses (Fig. 3b). However, we kept assignments where the parent-pair fell below the high confidence LOD threshold when one of the two parents could be assigned with high confidence. Assignments to identical genotypes were excluded from subsequent analyses (Fig. 3a).~~-~~ The goodness of the assignment was calculated as the mean of the exclusion probability over all tested offspring (which is one minus the non-exclusion probability calculated in CERVUS). Finally, we distinguished the more distant parent as the pollen donator (father) and the other as the seed source (mother) and all single parent cases as seed dispersal events following Dow and Ashley (1996).

2.4 Simulation study

2.4.1 The individual-based and spatially explicit ~~Larix~~ *Larix* model LAVESI

We ran simulations with the individual-based and spatially explicit model LAVESI (Kruse et al., 2016; Wieczorek et al., 2017). This model simulates the life-cycle of larch individuals~~species~~ from seeds to mature trees and was parameterised for *Larix gmelinii*. On a simulation area of user specified size, individuals grow where seeds settle and germinate, and competition among individuals is handled by a fine sub-grid of cells with an area of 20 x 20 cm. ~~It~~ The model was established to improve our understanding of past and future treeline displacements under changing climate. The relevant processes (growth, seed production and dispersal, establishment and mortality) are incorporated as submodules which are parameterised on the basis of field evidence (Wieczorek et al., 2017), complemented with data from literature. Seed dispersal into the environment is estimated by a Gaussian and fat-tailed dispersal function. With this and the detailed representation of competition the model realistically simulates, similar to Janzen's (1970) and Connell's findings (1971), that recruits have the highest chance to survive at intermediate distances to the producing tree, not directly at it. Fine-tuning the model parameters of involved processes, which includes the impact strength that competition has on smaller trees, allows adapting the effective seed dispersal distance. Simulation runs proceed in yearly time steps and are forced by monthly temperature and precipitation time series.

The original model of Kruse *et al.* (2016) was updated with the following processes (details in Suppelement 2): (i) seed dispersal distances now depend on species-specific traits (tree height, seed properties) and wind speed and direction (Kruse et al., 2018b), (ii) the tree diameter growth function is newly calibrated to the climate forcing (~~Kath~~Epp et al., 201~~6~~8), and (iii) the active-~~-~~layer thaw depth directly influences ~~directly~~ the tree's growth that is used to estimate it's seed production and mortality. Following these updates, the parameter settings of the original model were revised to simulate stands comparable to the site TY04VI (Kruse et al., 2018b).

2.4.2 Tuning the dispersal process in the model

We performed model runs to simulate larch stands in 100 x 100 m areas with closed boundaries. This means that seeds which leave the area on one side enter the field from the other side. To tune the model's processes in order to capture the observed effective seed dispersal distribution, ~~W~~we tested several combinations of model parameters and ~~newly~~ introduced new variables into formulae~~s~~ used in the program code of the model (listed in Table 1, details in Supplement 2 and in Table S5)~~, to tune the model's processes in order to capture the observed effective seed dispersal distribution~~. ~~The following model parameters were included in the tests (Supplement 2 and in Table S5): 'factor of seed production'~~ ~~f_s~~, ~~'factor of dispersal distance' Sdist, ratio of Gaussian and neg. exponential dispersal terms, variables determining the Gaussian dispersal function~~ ~~$r_{GaussExpDisp}$~~, ~~distance of the centre of the Gaussian dispersal term from the releasing tree~~ ~~$d_{GaussCentre}$~~, ~~exponent influencing the distance parameter of the Gaussian dispersal term~~ ~~$d_{GaussDistance}$~~, ~~'influencing factor for zone of trees'~~ ~~f_{HAI}~~, ~~'density influence factor on tree mortality'~~ ~~$f_{DensityMortality}$~~, ~~different modes to compute the competition among individuals.~~ Each simulation begins with a 2000 year-long spin-up phase, followed by an 80-year experimental phase (AD 1934–2013).

Climate forcing data was derived from the CRU TS 3.22 gridded data (0.5° resolution, Harris et al., 2014). We calculated the distance-to-centre weighted mean of all monthly temperature and precipitation values of the weather data of the TY04VI grid cell and its eight surrounding grid cells. May to August wind data were extracted from the ERA-Interim reanalysis data (AD 1979–2012; Bromwich et al., 2016). During the spin-up phase, weather data for each year was randomly sampled from the years AD 1939–2008 to allow the tree stand to reach a quasi-stable state. The period excludes a 5 year-long margin at the beginning and end of the climate observations available from the Khatanga station (see Kruse et al., 2016 for further information). For years not covered by wind data, the model randomly selected a year from the available wind data series. Simulations were run for 10 repeats and the outcomes of all individuals present in the final simulation year of AD 2013 were recorded.

The distance from the seed source tree for each established individual, i.e. the effective seed dispersal, was inferred from the simulation results. We resampled these simulated distances to consider the same frequency of observed parenthoods in the central 20 x 20 m as in the surrounding 100 x 100 m area (sampling scheme details in section 2.1 sample collection). We included only simulations which had at least 10 individuals present in the central 20 x 20 m area in further analyses. Distances were binned to classes of one metre steps. We evaluated the simulation results by calculating the Pearson's product moment correlation coefficient between simulated and observed dispersal distances. Furthermore, we reconstructed the proportion of on–site reproductive success in the final year as the ratio between on-site recruitment and all recruits. The differences between simulated and observed on-site recruitment ratios were tested by one-sided Student's *t*-tests.

2.4.3 Simulation experiments to depict migration rate

The best model that resulted from the tuning process was used to simulate larch migration in a hypothetical area of 1000 m (east-west) by 5000 m (north-south). Tree growth was initialised during the first 100 years by introducing ~~10~~20,000 seeds yr^{-1} into the southernmost 100-m area. This setting was run under two contrasting climate scenarios: first, with homogeneous temperature and precipitation forcing named 'EvenClim', and second, with linearly decreasing temperatures from south to north – mimicking the real south-to-north climate gradient – named 'GradClim'. The gradients of mean annual temperature and annual precipitation are described by $-6.24*10^{-6}$ °C m^{-1} and $+3.26*10^{-6}$ mm $(year\ m)^{-1}$, respectively, in a northward direction starting at TY04VI as inferred from an analysis of globally interpolated monthly climate data for 1960–1990 (Hijmans *et al.* 2005). Simulations under both scenarios were repeated 10 times and run for 2000 years using climate series from random years out of the available period of AD 1939–2008. Data for each established tree individual were collected every 10 years of the simulation.

We analysed stand densities for the entire simulated~~hypothetical~~ area. The number of trees (>2 m) was calculated from 100 x 100 m plots (=1 ha) by iteration over the entire area in 50 m steps (x and y). To reduce the errors introduced by the strict boundary conditions at the edges of the hypothetical area, the outer 100 m borders were excluded. The advance of larch stands into tundra was estimated from forest density by mean number of trees on east–west transects. We defined two relevant thresholds of tree densities (see also Fig. 2): (i) the ‘treeline’ when the mean tree density fell below 1 tree ha^{-1} and (ii) the ‘forest line’ when mean tree density fell below 100 trees ha^{-1}. These thresholds are in accordance with observations in treeline areas in Siberia (Kharuk et al., 2006; Montesano et al., 2016; Wieczorek et al., 2017). The migration rates of these thresholds in metres advanced per year were calculated as the slope of a linear model describing the position of the ‘treeline’ or ‘forest line’ as a function of simulation time. The rates were tested with a two sample *t*-test for significant differences. Analyses were performed in R version 3.2.2 (R Core Team, 2015).

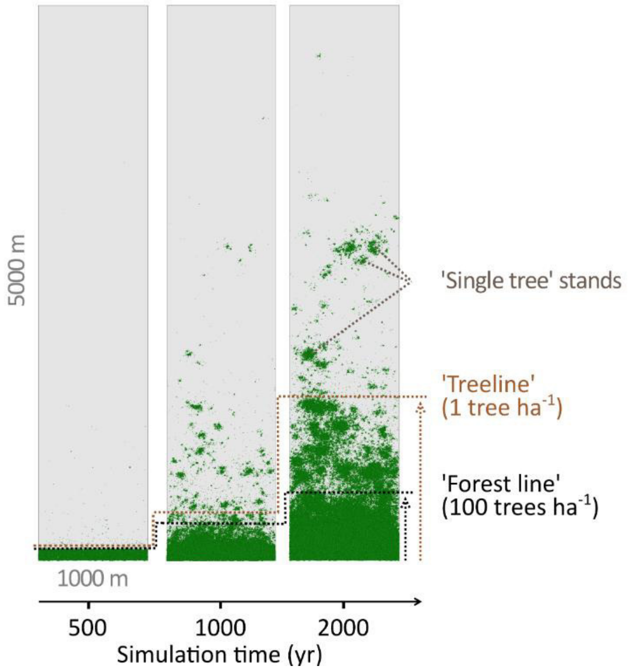


Figure 2: Simulated treeline advances along a hypothetical south-north transect across the modern ‘forest line’. Simulations were initialised with seeds in the southernmost 100 m area and run for 2000 years. In the beginning, trees established beyond the ‘treeline’ (mean density falling below 1 tree ha^{-1}) and formed ‘single tree’ stands in the tundra, which then acted as nuclei for further range expansion, so that in following years the ‘treeline’ and the ‘forest line’ (mean density falling below 100 trees ha^{-1}) could advance northwards forming open forest

3 Results

3.1 Effective seed dispersal distances as inferred from genetic parentage analysis

In total, 346 trees, 118 saplings and 148 seedlings were genotyped (Fig. 3). The mean distance between pairs of larch individuals was ~38 m (maximum: ~137 m). The mean tree height was 4.4 m (maximum: 9 m).

The eight chosen microsatellite loci were highly polymorphic and varied from 11 to 41 different alleles with only 0.49% missing alleles in total (Figure S1, Figure S4 in Supplement 1). The information content reached a plateau at four loci which could separate 597±2 individuals (>99%) and the power increased slightly towards 600±1 separated individuals with seven loci (Figure S2 in Supplement 1). Accordingly, we included all eight loci in the subsequent analyses to separate all individuals. In total, 601 sampled trees could be distinguished and ~~11~~22 individuals were identified as 10 clonal groups,~~es~~ of which 11 were subsequently excluded from further analyses (Fig. 3a, Supplement S1). The maximum distance between two individuals within ~~clonal~~ these groups was 30 m but mostly <5 m (Fig. 3a).

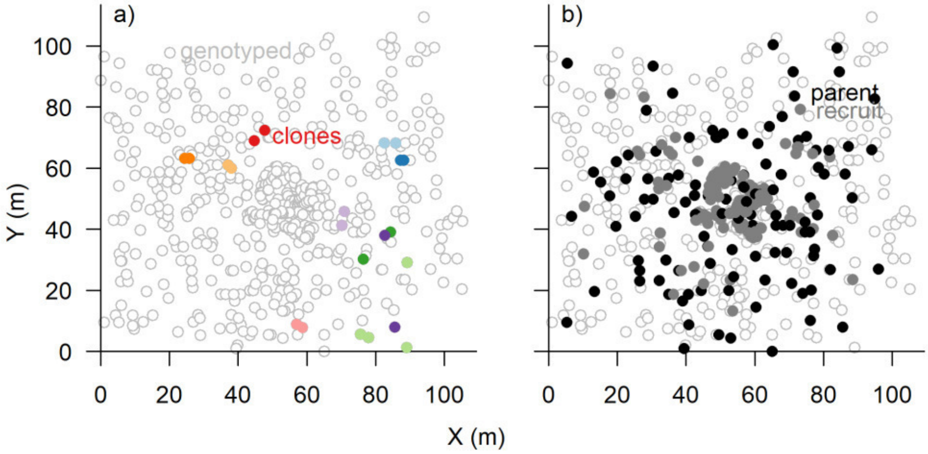
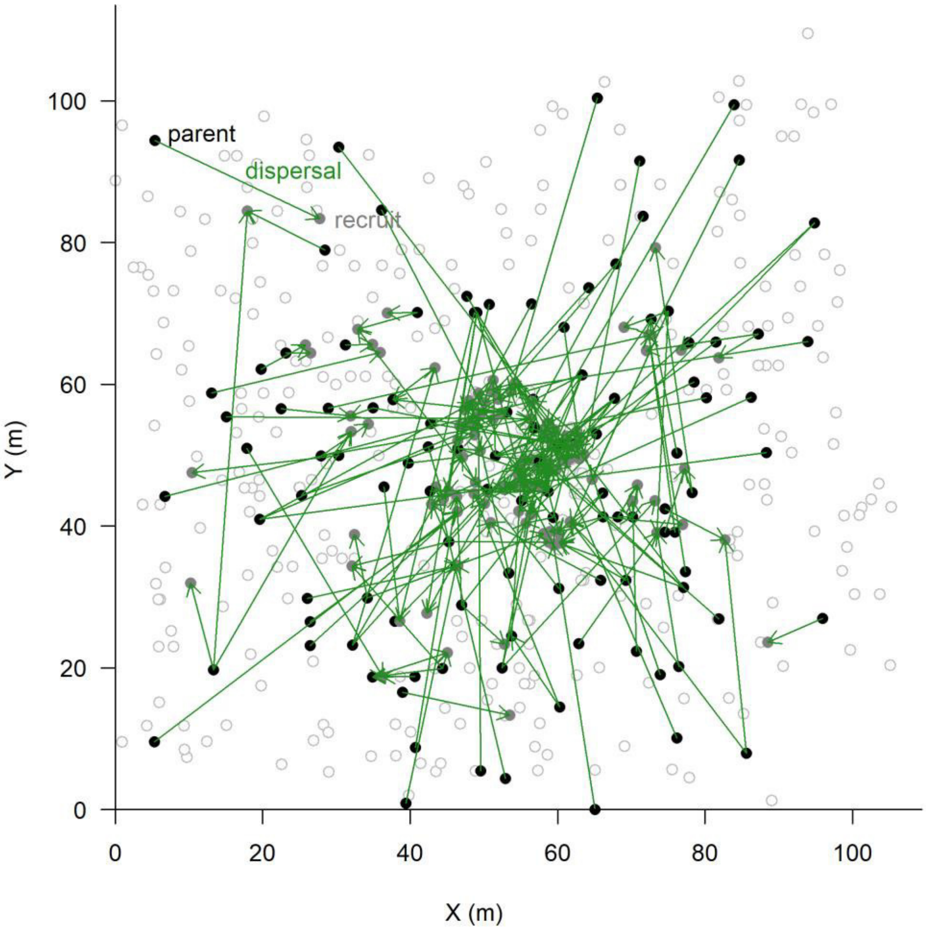


Figure 3: Map of genotyped *Larix gmelinii* individuals at site TY04VI, individuals sharing the same genotype (clones) are marked by filled points of the same colour in a) and local recruits are marked by filled bright circles whereas parents by filled black circles in b)

In the parenthood analyses, we aimed to find the parents for 266 individuals (<2 m) from among the remaining 464 individuals (>0.4 m). The exclusion probability for one parent was ~99.9910±0.0172% and for a pair of parents 100% in the used assay. A single parent or both parents were assigned for 151 individuals with a high confidence (>95%; LOD threshold for the parent pair of 15.13 and for a single parent 5.76, examples in Fig. 4). This is ~53% on-site recruitment in respect to all tested offspring. Among these, in 49 cases we found both parents (18.4%) and for 92 only a single parent (34.6%) was assigned. One of the largest trees (H=7 m) was assigned to 8% of the recruits (Fig. 4). Trees with many assignments are generally larger than those with few (Figure S5 in Supplement 1). Mostly parental trees exceeded their offspring's height, but in 3.7% cases recruits were higher than their assigned parents.


parent
dispersal
recruit
Y (m)
X (m)
0
20
40
60
80
100
0
20
40
60
80
100

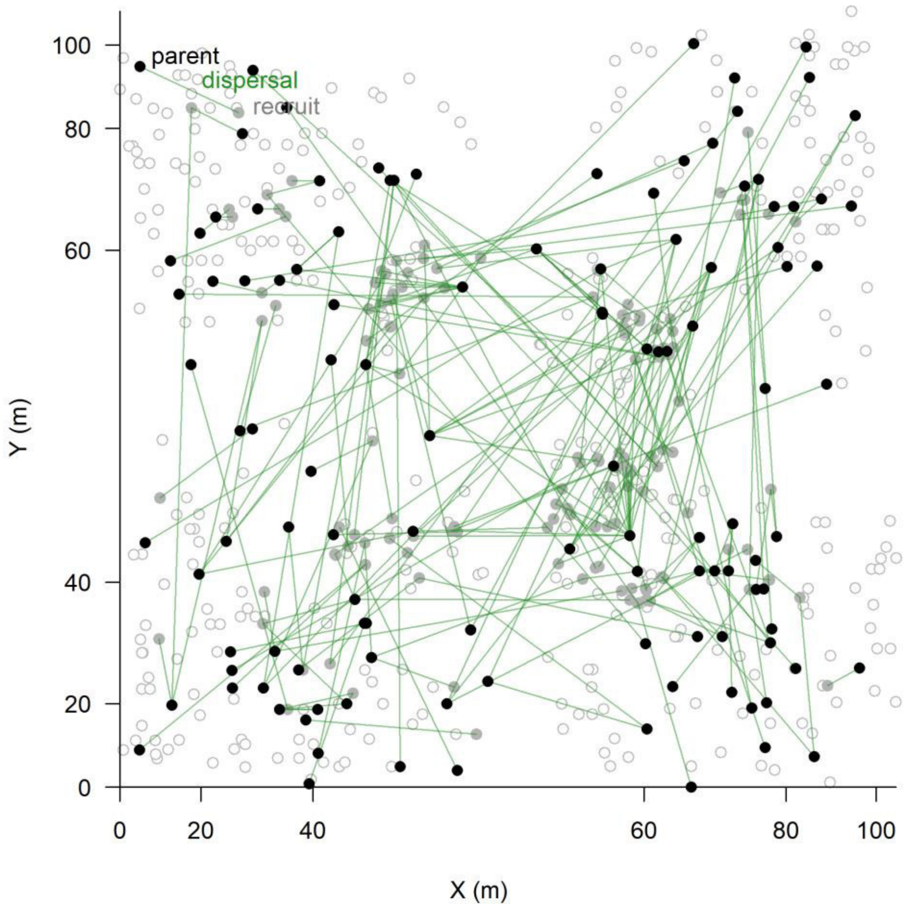


Figure 4: Map of assigned relationships between parents (filled black circles) and their offspring (filled bright circles) of *Larix gmelinii* individuals. Additional genotyped individuals are given as open circles. Note the non-linear scale of the coordinates

We identified 150 effective seed dispersal distances when assuming that the closest parent is the seed source when two parents were assigned or only a single parent was identified. The observed mean distance of effective seed dispersal is ~15.0 m (median of ~9.8 m), with a minimum of 0.8 m and a maximum of 56.1 m (Fig. 5).

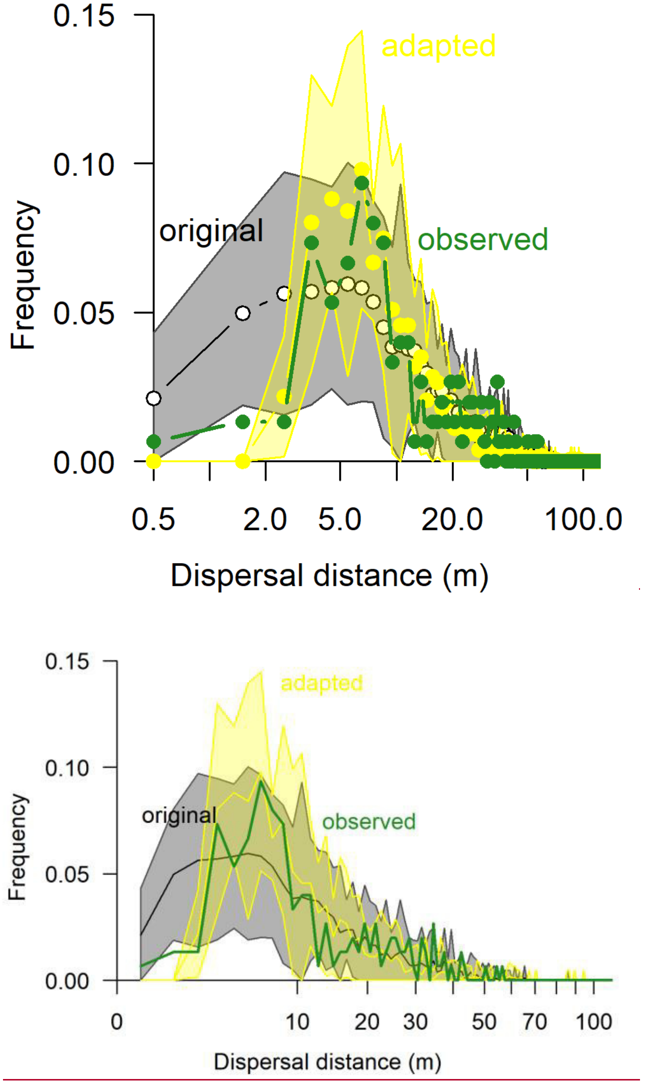


Figure 5: Seed dispersal of *Larix gmelniii* as simulated with the original simulation model LAVESI (open circles, grey band shows the 95% confidence interval) and the adapted model version (filled bright circles; yellow band shows the 95% confidence interval) compared to observed effective seed dispersal (filled dark circles). Dispersal distances were binned to one metre classes

3.2 The adapted effective seed dispersal in LAVESI

In the original model (Kruse et al., 2016; Wieczorek et al., 2017), effective seed dispersal distances follow a right-skewed distribution (Fig. 5). Mean dispersal for this model is 13.5 m (median 9.98 m), with a minimum of 0.6 m and a maximum of 60.3 m. In general, it captures the observed data, but it over-emphasis~~z~~es the recruits close to the mother tree at distances of 1–3 m. Furthermore, the simulated maximum probability peak between distances of ~4–7 m is roughly 3 m closer to the seed source than observed. Also, the observed tail approached zero probability faster at the far distances compared to the observed effective distances. These three deviations were used to guide the model adaptation.

The best-fitting model has effective dispersal distances that match well to the observed distances (r=0.93) and ha~~ve~~s the smallest sum of residuals (0.0063) compared to the other parameter sets (0.0066–0.1623). It is driven with a combination of parameters which increase the dispersal (Sdist=1, $r_{GaussExpDisp}$=1.0 and $d_{GaussCentre}$ =4.0) and seed production rate (f_s=11) compared to the original model (parameter set "I": Figure S7, Figure S8, Table S5 in Supplement 2). The resulting mean dispersal distance is 12.3 m (median 8.85 m) with a range of 2.7–71.1 m (Fig. 5).

The on-site recruitment ratio (~53%) was generally underestimated in the simulations (38.1–54.9%, Table S5). The parameter settings which produced similar ratios greatly overemphasised dispersal at short distances and, thus, the simulated seed dispersal pattern deviated strongly (correlation values of r=0.26–0.58, "tCJ"). The simulations with the best-fitting model overemphasis~~zs~~e effective dispersal distances so that more recruits immigrate into the plot (~7%, Supplement 2).

3.3 Simulating migration dynamics in the taiga-tundra transition zone

Simulations were run for a hypothetical 5000 m long south-north transect, initialised by introducing seeds to the southern 100-m wide area. In the homogeneous climate scenario, EvenClim, single trees spread up to ~3600 m during the 2000-year simulation and a 'treeline' (mean density falling below 1 tree ha^{-1}) formed at ~2000 m. A 'forest line' (mean density falling below 100 trees ha^{-1}) formed up to ~500 m further north than one forced by the climate gradient scenario ClimGrad (Fig. 7). Migration was first accelerated by isolated colonisation events above the 'forest line', so that the 'treeline' moved northwards by ~1.5 m yr^{-1} into treeless areas, but decelerated after a peak between 500–1000 years (Fig. 6). The advance of the 'forest line' on the other hand accelerated throughout the EvenClim simulation until becoming six times faster at the end of the simulation (1500–2000 yrs) than at the beginning (0–1000 yrs). The migration rates of the 'forest line' were roughly half in the GradClim scenario.

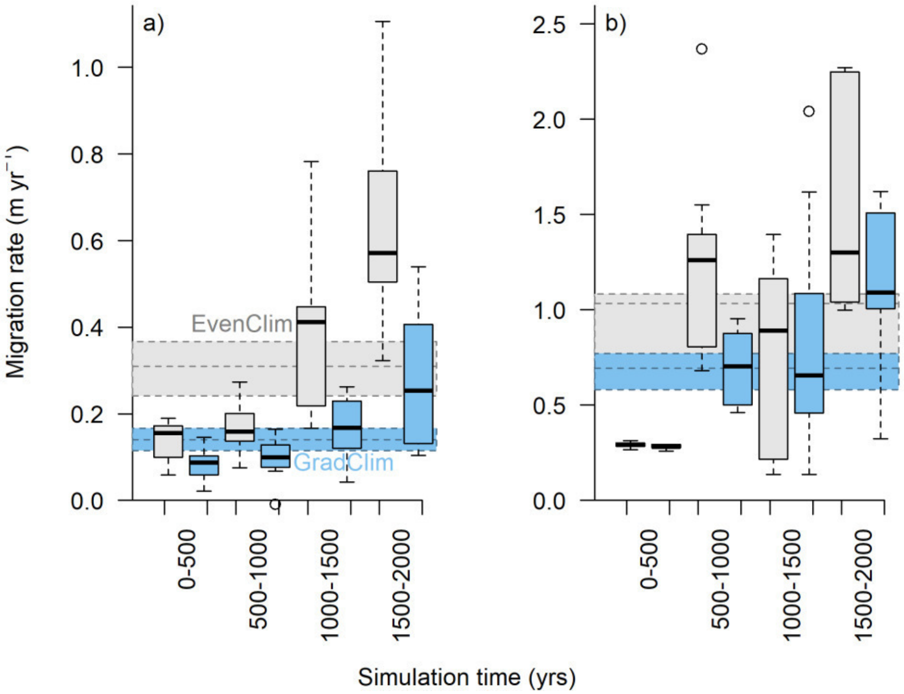


Figure 6: Simulated migration rates of the 'forest line' (mean density falling below 100 trees ha^{-1}) and the 'treeline' (mean density falling below 1 tree ha^{-1}) estimated from the best-fitting model. The simulations were forced by two contrasting climate scenarios, either homogeneous temperature across the area (EvenClim: grey shading) or linearly decreasing temperature from south to north (GradClim: darker blue shading)

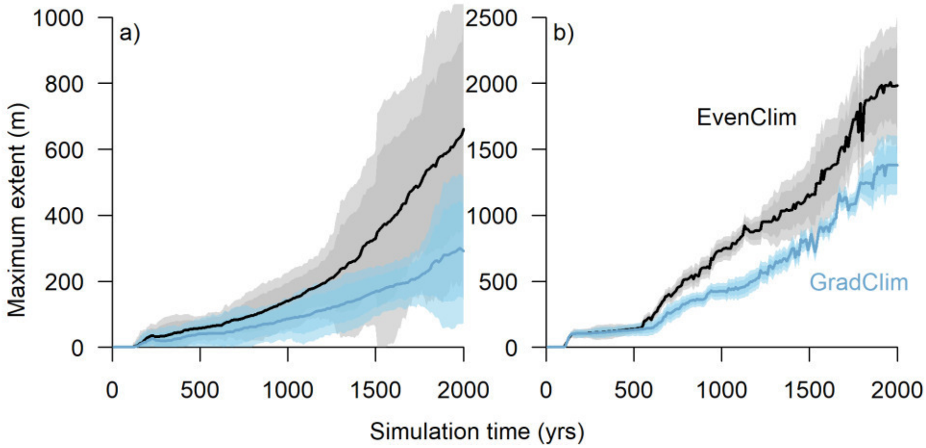


Figure 7: Simulations along a hypothetical transect at the taiga-tundra transition reveal the northward advance of a) the 'forest line' (mean density falling below 100 trees ha^{-1}) and b) 'treeline' (mean density falling below 1 tree ha^{-1}). Simulations were forced by two contrasting temperature scenarios (homogeneous temperature EvenClim (grey shading) and northwards linearly decreasing temperature and precipitation GradClim (darker blue shading)). Shaded areas give the 99 and 90 % confidence intervals around the mean value of 10 simulation repeats

4 Discussion

4.1 Seed dispersal distances inferred from genetic heritage analysis

Our assay of eight highly polymorphic microsatellites distinguished all 601 genotyped individuals allowing us to infer the local recruitment pattern of a tree stand and, from that, the effective seed dispersal distances. In this analysis, we needed to exclude the observed clonal groups that are consequences of exceptional reproduction. We are confident that these are true observations of clones as we minimized the chance of full sibs share the same genotype by using highly polymorphic nuclear microsatellites that are not in linkage disequilibrium (Kruse et al., 2018). Nevertheless, we cannot rule out that selfing or back-crossing have occurred that could yield to offspring being genetically identical to one of the parents. If those modes of inheritance regularly occur and would have caused a misidentification of full siblings as clones, we would expect to observe an continuously increasing number of transitional states from identical genotypes (0 different alleles) to sharing 50% of their alleles (8 different alleles). However, it sharply drops from the identified clonal groups to a very low value and increases again beginning at 3 to 4 differences (Fig. S6 in Supplement 1). This gives us confidence to classify such identical individuals as clones. An explanation for these could be, that wind thrown trees can survive or in non-favorable conditions producing horizontal branches rather than upright stems forming krummholtz (own observations, Wieczorek et al., 2017). By producing adventitious roots from branches touching the ground (Kajimoto, 2010; Cooper, 1911) and subsequent separation of the main stem or horizontal branch two genetically identical individuals can be found if both parts survive.

We were able to identify at least one parent for a majority of the offspring in the parentage analysis (53%: 18.4% both parents and 34.6% one parent), even though only those cases with a high degree of confidence (>95%) were regarded and an area of only 100 x 100 m was analysed. 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. The observed effective seed dispersal ranged between ~1 and 56 m (median: 10 m). This aligns with the short dispersal distances generally reported for larches. For example, it was found that most seeds (94%) fell within 18 m of the releasing trees in a study of *Larix laricina* in northern USA (Duncan, 1954). This result, however, is not directly comparable to effective seed dispersal, because all dispersed seeds are included in the estimation and not only those which germinated and established as a new individual. Effective seed dispersal distances of 2–48 m were found in dense forests of *Larix decidua* in the Swiss Alps using an approach similar to ours (Pluess, 2011). Higher effective seed dispersal distances have been observed, however, for other wind-dispersed tree taxa (e.g. 27–58 m for *Pinus pinaster*: González-Martínez et al., 2002; 39–833 m for *Picea* Piotti et al., 2009). One explanation for the observed differences might be the density of the tree stands because established trees reduce the wind speed. In our tree stand, which is comparatively denser than those studied above, shorter dispersal distances are more likely than in open areas (Antonovics and Levin, 2016). Furthermore, dispersal is dependent on the release height (Matlack, 1987), which in our stand, was rather low due to the shortness of the trees (mean: ~4.5 m, max: 9 m). Most cones occurred on branches at roughly half the tree's height (see Fig. 1), as is typical for open stands.

The observed amount of on-site recruitment is high compared to other studies (Piotti et al., 2009; Pluess, 2011), but lower than that found in orchards (Funda et al., 2008). We found both parents for one fifth of the offspring within the analysed area (compared to only 11.1% in Piotti et al., 2009). As a result, recruits effectively immigrated at a rate of 47% from the exterior of the analysed 100 x 100 m area, which is similar to a study of *Picea abies* in the Italian alps (43.3%: Piotti et al., 2009).

The parenthood in our investigated site was dominated by a few, relatively tall trees. For example, a 6 m-tall tree generated 8% (17 cases) of all identified recruits. This observation is reasonable, as the chance of producing viable offspring increases with age and size, and, once a tree is well established, its seeds are released into the local environment. Other studies make similar observations. For example, Dow & Ashley (1996) found that *Quercus* saplings often established close to the releasing tree and the majority of the offspring were assigned to four out of 62 mature trees. Schnabel *et al.* (1998) observed that three *Gleditsia* trees produced 58% of the offspring and Piotti *et al*. (2009) found that six local adults produced ~62% of *Picea* juveniles.

Overall, our results indicate that incorporating individual seed dispersal (such as implemented in LAVESI) rather than introducing a certain seed sum needs to be implemented in models to realistically model tree migration processes.

We were unfortunately unable to distinguish between fatherhood and motherhood using nuclear inherited markers (DeVerno et al., 1993; Szmidt et al., 1987). It is a valid criticism that we simply assume that single-parent assignments represent seed dispersal events (following Dow and Ashley, 1996; Moran and Clark, 2012; Piotti et al., 2009). In the extremely unlikely case that the more distant parent was instead the seed source (results not shown), the effective dispersal distance would increase to a median of ~23 m. This would lead to a further decrease in on-site recruitment which is already slightly underestimated. Furthermore, our approach risks assigning missing parents to extra-site recruitment if the local parents have died, leading to an overemphasis of the fat tail. We consider this risk to be low in our analysed tree stand as we found only a few dead trees or saplings within the 100 x 100 m area, and they were already largely decayed (Wieczorek et al., 2017).

4.2 Genetic-model comparison and model adaptation

Using observations of parentage from a treeline stand of *Larix gmelinii* we improved the seed dispersal function in LAVESI so that it will better represent larch migration.

The original dispersal function modelled seed distribution using a simple Gaussian density function with a fat tail (Kruse et al., 2016), as is implemented in a number of models (e.g. Levin et al., 2003; Snell, 2014), but, in contrast to most other models, dispersal in LAVESI is related to wind speed and direction (Kruse et al., 2018b). The most realistic simulation results are achieved via a combination of parameter adjustments, i.e. shifting the implemented Gaussian distribution term by 4 m away from the centre, increasing the factor scaling the distance by roughly six times the original value and increasing the influence of the Gaussian term twice (model "I" in Table S5 in Supplement 1). With these adjustments, the simulated effective seed dispersal distance aligns fairly well to the observed values. The new function slightly overestimates dispersal and therefore allows ~7% more recruits to immigrate into the plot. This discrepancy, however, might also be an artefact related to the shortcomings of the genetic parentage analyses. Regardless, with the dispersal function parameterised to the observed effective seed dispersal, simulations are now more realistic than with the original version of the model (Kruse et al., 2016).

4.3 Treeline migration rates

We performed simulation experiments with the best-fitting model to estimate the potential migration rate of the treeline on the southern Taymyr Peninsula. Under the scenario of even temperature and precipitation (EvenClim), the northwards migration rate of the 'forest line' is ~0.6 m year^{-1} and the 'treeline' ~1.6 m year^{-1}. Under the more realistic climate gradient scenario of northwards decreasing temperature and slightly increasing precipitation, an even slower advancing 'treeline' and 'forest line' is implied. Overall, we find an astonishingly low migration rate, even though the best-fitting model slightly overestimates immigration at the stand level. Our simulations may yet be conservative as we cannot rule out that the dispersal function underestimates far distance dispersal at the same time as overestimating intermediate distance dispersal. Nevertheless, the slow recruitment ahead of the 'treeline' is in accordance with field observations at northernmost single-tree stands in the tundra, which show creeping growth-forms (krumholtz) and no apparent recruitment (Wieczorek et al., 2017). Our estimated migration rate is quite slow compared to the observed spread of larch individuals into the tundra by 3–11 m year^{-1}, as mapped by Kharuk ~~*et al*~~et al. (2006). However, the stand Kharuk ~~*et al*~~et al. (2006) investigated is an exceptional open-forest island close to a river ahead of the modern 'forest line' where winds might be stronger leading to higher dispersal rates (Antonovics and Levin, 2016; Duncan, 1954). Another field-based study reports a treeline expansion of 50 m year^{-1} in arctic Alaska (Lloyd, 2005), whereas an elevational range shift for larch in the Polar Urals of 20–60 m during the last century is reported by Devi *et al.* (2008) and a general upward shift of 20–50 m between 1910 and 2000 of open forest in this mountainous area (Shiyatov et al., 2005; Shiyatov and Mazepa, 2012). During the Holocene Thermal Maximum boreal forests expanded on the Taymyr Peninsula to their northernmost position during the Holocene, which was likely assisted by glacial refugial populations ahead of the treeline (MacDonald et al., 2000, 2008). The~~s~~ treeline responded with a centennial lag to environmental improvement, for example solar insolation, and reached its maximum position at ~8000 to 4000 yr BP, and subsequently declined to reach its modern limits around 3000 yr BP (MacDonald et al., 2000). Recently, global warming is ameliorating conditions for *Larix* forests in Siberia and evidence can be found that treeline stands are starting to respond, but at a slower rate than one might expect given the strong increase in temperatures (Wieczorek et al., 2017; Harsch et al., 2009). -A possible explanation for the slow advance ~~given here,~~ may be because we report the advance of a forest line rather than single trees. Furthermore, we analysed only one tree stand and effective dispersal rates will likely differ among sites depending on a variety of abiotic or biotic factors (Moran and Clark, 2012). The actual dispersal distance depends on~~, for example,~~ stand density, amongst others, as more trees reduce the wind speed (Antonovics and Levin, 2016) and establishment will be affected by local density-dependent mortality due to seed predation close to the releasing tree (Janzen–Conell effect, Janzen (1970) and Conell (1971)). Furthermore, the probability of seeds surviving and forming a seedbank and the survival rates of seedlings strongly determine the colonis~~z~~ation 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 ~~is~~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.

Our results show that small uncertainties in the implementation of dispersal in a model impact~~s~~ the timing and shape of the simulated tundra colonisation. This is in accordance with a simulated lag in vegetation response to climate change when seed dispersal in a global dynamic vegetation model is constrained rather than using the usual unlimited seed bank approach (Snell, 2014; Snell and Cowling, 2015). However, further processes on smaller scales can constrain the response of tree stands and should therefore not be neglected in simulation studies: an advancing front is shaped by short-distance dispersal and spatially-explicit processes, such as competition between individuals. A simulation model with spatially-explicit seed dispersal combined with a representation of small-scale population processes helps to give realistic estimates on the migration rates. We have demonstrated that the LAVESI model allows a realistic implementation and parameterisation of dispersal processes.

In summary, our results suggest that the current climate change will lead to a lagged response by decades to centuries. In particular, the first step of migration will be slow, although the subsequent infilling could be rapid. It seems likely, therefore, that recent strong warming will cause a highly nonlinear response in forest and treeline advance.

5 Conclusions

We parameterised and applied the individual-based model and spatially explicit LAVESI to estimate migration rates of the treeline and forest line advance under current climate conditions. First, we inferred the effective seed dispersal distance from a genetic parentage analysis based on nuclear microsatellites, and second, we improved the dispersal process of the model according to the observed dispersal pattern.

In our genetic analyses, we found a genetically diverse tree population at a location within the treeline close to the tundra in Siberia. The parentage analysis revealed that the majority of recruits (~60%) have a local origin. Knowing the positions of the parent trees, we could estimate the effective seed dispersal distances between parent and offspring, which are mostly short (~10 m), although longer distances (up to ~60 m) are possible. Simulations with the adapted LAVESI model improved our knowledge about the likely treeline migration response. The rate is surprisingly slow: just a few metres northwards per year. To find out if the estimated slow migration is an outlier coming from overfitting to only one study site or the general response rate under current warming, further similar studies at other treeline positions would be necessary. The simulated migration pattern also showed that occasional long-distance seed dispersal events far beyond the treeline area assisted the colonisation of the tundra. Our migration rate estimates are in the lower range of those observed and significantly slower than those inferred from palaeoecological studies or from simulated vegetation responses to climate change in dynamic global vegetation models. These findings indicate that the treeline in north-central Siberia will lag behind the recent strong warming (which is moving by ~1,000 m yr^{-1}, Loarie ~~*et al.*~~et al, 2009) but if isolated trees occasionally establish in the tundra, they could become nuclei for a rapid colonisation of the tundra. Should this rapid colonisation occur, the albedo of these populated tundra areas will reduce and thus a positive feedback to climate warming will follow the lagged response of tundra-taiga transition. Such a scenario could be run in a large-scale simulation experiment using the improved version of the LAVESI model in an attempt to learn more about the impacts of such a vegetation-climate feedback in the upcoming decades.

Code availability

The source code of the original model LAVESI is available at GitHub https://github.com/StefanKruse/LAVESI/releases/tag/v1.01, and stored in the zenodo database http://doi.org/10.5281/zenodo.1155486. The updated version with a wind-dependent seed dispersal kernel named LAVESI-WIND is available in the first version 1.0 and accessible at GitHub at https://github.com/StefanKruse/LAVESI/tree/v1.0 and stored at http://doi.org/10.5281/zenodo.1165383.

Sample availability

Sampling locations, morphological data and microsatellite genotype data will be made publicly available at PANGAEA after acceptance of this manuscript.

Author contribution

SK and UH designed the study and SK performed the experiments. SK, LSE, KRS-L and LAP conducted fieldwork. SK generated molecular analysis. SK, AG and NK implemented the model and performed model simulations. SK, LE and UH analysed the data. SK wrote a first version of the manuscript that all co-authors commented on~~; other authors provided editorial advice~~.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

We thank our Russian colleagues from the joint Russian-German expedition 2013 for support in the field. Special thanks to Alexey Kolmogorov, Bastian Niemeyer and Xenia Schreiber for their help in sampling and tree measurements. We would like to thank Cathy Jenks for proofreading and improving the manuscript. Furthermore, we thank the Associate Editor Kirsten Thonicke and four anonymous reviewers for their comments improving earlier versions of the paper. ~~-~~The PhD position of Stefan Kruse was funded by the Initiating and Networking Fund of the Helmholtz Association and the ERC consolidator grant Glacial*Legacy* of Ulrike Herzschuh (Grant No. 772852).~~and by the German Research Foundation~~.

References

Andreev, A. A., Siegert, C., Klimanov, V. A., Derevyagin, A. Y., Shilova, G. N. and Melles, M.: Late Pleistocene and Holocene Vegetation and Climate on the Taymyr Lowland, Northern Siberia, Quat. Res., 57(1), 138–150, doi:10.1006/qres.2001.2302, 2002.

Antonovics, J. and Levin, D. A.: The Ecological and Genetic Consequences of Density-Dependent Regulation in Plants, Annu. Rev. Ecol. Syst., 11(1980), 411–452 [online] Available from: http://www.jstor.org/stable/2096915, 2016.

Arctic Climate Impact Assessment: Impacts of a Warming Arctic-Arctic Climate Impact Assessment, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., 2004.

Ashley, M. V.: Plant parentage, pollination, and pispersal: How DNA microsatellites have altered the landscape, CRC. Crit. Rev. Plant Sci., 29(3), 148–161, doi:10.1080/07352689.2010.481167, 2010.

Bhagwat, S. A. and Willis, K. J.: Species persistence in northerly glacial refugia of Europe: A matter of chance or biogeographical traits?, J. Biogeogr., 35(3), 464–482, doi:10.1111/j.1365-2699.2007.01861.x, 2008.

Bonan, G. B.: Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests, Science (80-.)., 320(5882), 1444–1449, doi:10.1126/science.1155121, 2008.

Bromwich, D. H., Wilson, A. B., Bai, L.-S., Moore, G. W. K. and Bauer, P.: A comparison of the regional Arctic System Reanalysis and the global ERA-Interim Reanalysis for the Arctic, Q. J. R. Meteorol. Soc., 142(695), 644–658, doi:10.1002/qj.2527, 2016.

Brown, K. R., Zobel, D. B. and Zasada, J. C.: Seed dispersal, seedling emergence, and early survival of *Larix laricina* (DuRoi) K. Koch in the Tanana Valley, Alaska, Can. J. For. Res., 18(3), 306–314, doi:10.1139/x88-047, 1988.

Clark, J. S.: Why trees migrate so fast: confronting theory with dispersal biology and the paleorecord., Am. Nat., 152(2), 204–224, doi:10.1086/286162, 1998.

Connell, J. H.: On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees, in Dynamics of populations, vol. 298, edited by P. J. Den Boer and G. Gradwell, p. 312, Centre for Agricultural Publishing and Documentation, Wageningen, Netherlands., 1971.

Cooper, W. S.: Reproduction by layering among conifers. Botanical Gazette 52(5), 369–379, 1911.

DeVerno, L., Charest, P. and Bonen, L.: Inheritance of mitochondrial DNA in the conifer *Larix*, Theor. Appl. Genet., 86(2), 383–388, doi:10.1007/BF00222106, 1993.

Devi, N., Hagedorn, F., Moiseev, P., Bugmann, H., Shiyatov, S., Mazepa, V. and Rigling, A.: Expanding forests and changing growth forms of Siberian larch at the Polar Urals treeline during the 20th century, Glob. Chang. Biol., 14(7), 1581–1591, doi:10.1111/j.1365-2486.2008.01583.x, 2008.

Dow, B. D. and Ashley, M. V.: Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*, Mol. Ecol., 5(5), 615–627, doi:10.1111/j.1365-294X.1996.tb00357.x, 1996.

Duncan, D. P.: A study of some of the factors affecting the natural regeneration of tamarack (*Larix laricina*) in Minnesota, Ecology, 35(4), 498–521, doi:10.2307/1931040, 1954.

Epp, L. S., Kruse, S., Kath, N. J., Stoof-Leichsenring, K. R., Tiedemann, R., Pestryakova, L. A. and Herzschuh, U.: Temporal and spatial patterns of mitochondrial haplotype and species distributions in Siberian larches inferred from ancient environmental DNA and modeling, Sci. Rep., 8(1), 17436, doi:10.1038/s41598-018-35550-w, 2018.

Funda, T., Chen, C. C., Liewlaksaneeyanawin, C., Kenawy, A. M. A. and El-Kassaby, Y. A.: Pedigree and mating system analyses in a western larch (*Larix occidentalis* Nutt.) experimental population, Ann. For. Sci., 65(7), 705, doi:10.1051/forest:2008055, 2008.

Germino, M. J., Smith, W. K., Resor, A. C., Ecology, S. P., Terms, J. and Smith, K.: Conifer Seedling Distribution and Survival in an Alpine-Treeline Ecotone Conifer seedling distribution, Plant Ecol., 162(2), 157–168, doi:10.1023/A:1020385320738, 2011.

González-Martínez, S. C., Gerber, S., Cervera, M. T., Martínez-Zapater, J. M., Gil, L. and Alía, R.: Seed gene flow and fine-scale structure in a Mediterranean pine (*Pinus pinaster* Ait.) using nuclear microsatellite markers, Theor. Appl. Genet., 104(8), 1290–1297, doi:10.1007/s00122-002-0894-4, 2002.

Greene, D. F., Canham, C. D., Coates, K. D. and Lepage, P. T.: An evaluation of alternative dispersal functions for trees, J. Ecol., 92(5), 758–766, doi:10.1111/j.0022-0477.2004.00921.x, 2004.

Harris, I., Jones, P. D., Osborn, T. J. and Lister, D. H.: Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset, Int. J. Climatol., 34(3), 623–642, doi:10.1002/joc.3711, 2014.

Harsch, M. A., Hulme, P. E., McGlone, M. S. and Duncan, R. P.: Are treelines advancing? A global meta-analysis of treeline response to climate warming, Ecol. Lett., 12(10), 1040–1049, doi:10.1111/j.1461-0248.2009.01355.x, 2009.

Hartl, D. L. and Clark, A. G.: Principles Of Population Genetics, 4th ed., Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts, USA., 2007.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. and Jarvis, A.: Very high resolution interpolated climate surfaces for global land areas, Int. J. Climatol., 25(15), 1965–1978, doi:10.1002/joc.1276, 2005.

Holtmeier, F.-K. and Broll, G.: Sensitivity and response of Northern Hemisphere altitudinal and polar treelines to environmental change at landscape and local scales, Glob. Ecol. Biogeogr., 14(5), 395–410, doi:10.1111/j.1466-822x.2005.00168.x, 2005.

IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., 2013.

Janzen, D. H.: The American Society of Naturalists Herbivores and the Number of Tree Species in Tropical Forests, Am. Nat., 104(940), 501–528, doi:10.1086/282687, 1970.

Jombart, T.: Adegenet: A R package for the multivariate analysis of genetic markers, Bioinformatics, 24(11), 1403–1405, doi:10.1093/bioinformatics/btn129, 2008.

Jombart, T. and Ahmed, I.: adegenet 1.3-1: New tools for the analysis of genome-wide SNP data, Bioinformatics, 27(21), 3070–3071, doi:10.1093/bioinformatics/btr521, 2011.

Kajimoto, T.: Root System Development of Larch Trees Growing on Siberian Permafrost in Permafrost Ecosystems, edited by A. Osawa, O. A. Zyryanova, Y. Matsuura, T. Kajimoto, and R. W. Wein, Springer Netherlands, Dordrecht., doi:10.1007/978-1-4020-9693-8, 2010.

Kalinowski, S. T., Taper, M. L. and Marshall, T. C.: Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment, Mol. Ecol., 16(5), 1099–1106, doi:10.1111/j.1365-294X.2007.03089.x, 2007.

Kaplan, J. O. and New, M.: Arctic climate change with a 2 °C global warming: Timing, climate patterns and vegetation change, Clim. Change, 79(3–4), 213–241, doi:10.1007/s10584-006-9113-7, 2006.

~~Kath, N. J.: Spatio-temporal competition dynamics of larch species in North Central Siberia, Master's thesis, University of Potsdam., 2016.~~

Kharuk, V. I., Ranson, K. J., Im, S. T. and Naurzbaev, M. M.: Forest-tundra larch forests and climatic trends, Russ. J. Ecol., 37(5), 291–298, doi:10.1134/S1067413606050018, 2006.

Klemm, J., Herzschuh, U. and Pestryakova, L. A.: Vegetation, climate and lake changes over the last 7000 years at the boreal treeline in north-central Siberia, Quat. Sci. Rev., 147(1–13), 422–434, doi:10.1016/j.quascirev.2015.08.015, 2016.

Kruse, S., Wieczorek, M., Jeltsch, F. and Herzschuh, U.: Treeline dynamics in Siberia under changing climates as inferred from an individual-based model for *Larix*, Ecol. Model., 338, 101–121, doi:10.1016/j.ecolmodel.2016.08.003, 2016.

Kruse, S., Epp, L. S., Wieczorek, M., Pestryakova, L. A., Stoof-Leichsenring, K. R. and Herzschuh, U.: High gene flow and complex treeline dynamics of *Larix* Mill. stands on the Taymyr Peninsula (north-central Siberia) revealed by nuclear microsatellites, Tree Genet. Genomes, 14(2), 19, doi:10.1007/s11295-018-1235-3, 2018a.

~~Kruse, S., Gerdes, A., Kath, N. J. and Herzschuh, U.: Implementing spatially explicit seed and pollen dispersal in the individual-based larch simulation model: LAVESI-WIND 1.0, Geosci. Model Dev. Discuss., 2018, 1–23, doi:10.5194/gmd-2018-31, 2018b.~~Kruse, S., Gerdes, A., Kath, N. J. and Herzschuh, U.: Implementing spatially explicit wind-driven seed and pollen dispersal in the individual-based larch simulation model: LAVESI-WIND 1.0, Geosci. Model Dev., 11(11), 4451–4467, doi:10.5194/gmd-11-4451-2018, 2018b.

Levin, S. A., Muller-Landau, H. C., Nathan, R. and Chave, J.: The ecology and evolution of seed dispersal: A theoretical perspective, Annu. Rev. Ecol. Evol. Syst., 34(1), 575–604, doi:10.1146/annurev.ecolsys.34.011802.132428, 2003.

Lloyd, A. H.: Ecological histories from Alaskan tree lines provide insight into future change, Ecology, 86(7), 1687–1695, 2005.

Loarie, S. R., Duffy, P. B., Hamilton, H., Asner, G. P., Field, C. B. and Ackerly, D. D.: The velocity of climate change, Nature, 462(7276), 1052–1055, doi:10.1038/nature08649, 2009.

MacDonald, G. M., Kremenetski, K. V. and Beilman, D. W.: Climate change and the northern Russian treeline zone, Philos. Trans. R. Soc. B Biol. Sci., 363(1501), 2283–2299, doi:10.1098/rstb.2007.2200, 2008.

MacDonald, G. M., Velichko, A. a., Kremenetski, C. V., Borisova, O. K., Goleva, A. a., Andreev, A. a., Cwynar, L. C., Riding, R. T., Forman, S. L., Edwards, T. W. D., Aravena, R., Hammarlund, D., Szeicz, J. M. and Gattaulin, V. N.: Holocene Treeline History and Climate Change Across Northern Eurasia, Quat. Res., 53(3), 302–311, doi:10.1006/qres.1999.2123, 2000.

Maher, E. L. and GerminO, M. J.: Microsite differentiation among conifer species during seedling establishment at alpine treeline, Écoscience, 13(3), 334–341, doi:10.2980/i1195-6860-13-3-334.1, 2006.

Marshall, T. C., Slate, J., Kruuk, L. E. B. and M., P. J.: Statistical confidence for likelihood-based paternity, Mol. Ecol., 7, 639–655, 1998.

Matlack, G. R.: Diaspore size, shape, and fall behavior in wind-dispersed plant species, Am. J. Bot., 74(8), 1150–1160, doi:10.2307/2444151, 1987.

McLachlan, J. S., Clark, J. S. and Manos, P. S.: Molecular indicators of tree migration capacity under rapid climate change, Ecology, 86(8), 2088–2098, doi:10.1890/04-1036, 2005.

Montesano, P. M., Sun, G., Dubayah, R. O. and Ranson, K. J.: Spaceborne potential for examining taiga-tundra ecotone form and vulnerability, Biogeosciences, 13(13), 3847–3861, doi:10.5194/bg-2015-575, 2016.

Moran, E. V. and Clark, J. S.: Between-site differences in the scale of dispersal and gene flow in red oak, PLoS One, 7(5), e36492, doi:10.1371/journal.pone.0036492, 2012.

Nathan, R. and Muller-Landau, H. C.: Spatial patterns of seed dispersal, their determinants and consequences for recruitment, Trends Ecol. Evol., 15(7), 278–285, doi:10.1016/S0169-5347(00)01874-7, 2000.

Naurzbaev, M. M., Vaganov, E. a., Sidorova, O. V. and Schweingruber, F. H.: Summer temperatures in eastern Taimyr inferred from a 2427-year late-Holocene tree-ring chronology and earlier floating series, The Holocene, 12(6), 727–736, doi:10.1191/0959683602hl586rp, 2002.

Neilson, R. P., Pitelka, L. F., Solomon, A. M., Nathan, R., Midgley, G. F., Fragoso, J. M. V., Lischke, H. and Thompson, K.: Forecasting Regional to Global Plant Migration in Response to Climate Change, Bioscience, 55(9), 749, doi:10.1641/0006-3568(2005)055[0749:FRTGPM]2.0.CO;2, 2005.

Niemeyer, B., Herzschuh, U. and Pestryakova, L. A.: Vegetation and lake changes on the southern Taymyr peninsula, northern Siberia, during the last 300 years inferred from pollen and *Pediastrum* green algae records, The Holocene, 25(4), 596–606, doi:10.1177/0959683614565954, 2015.

Pairon, M., Jonard, M. and Jacquemart, A.-L.: Modeling seed dispersal of black cherry, an invasive forest tree: how microsatellites may help?, Can. J. For. Res., 36(6), 1385–1394, doi:10.1139/x06-018, 2006.

Petit, R. J., Hu, F. S. and Dick, C. W.: Forests of the past: a window to future changes., Science, 320(5882), 1450–1452, doi:10.1126/science.1155457, 2008.

Piotti, A., Leonardi, S., Piovani, P., Scalfi, M. and Menozzi, P.: Spruce colonization at treeline: where do those seeds come from?, Heredity (Edinb)., 103(2), 136–45, doi:10.1038/hdy.2009.42, 2009.

Pisaric, M. F. J., Macdonald, G. M., Cwynar, L. C. and Velichko, A. A.: Pollen Modern and Conifer Stomates from Siberian Lake Their Sediments : Use in Interpreting Late Fossil Pollen Quaternary Assemblages, Arctic, Antarct. Alp. Res., 33(1), 19–27, 2001.

Pluess, A. R.: Pursuing glacier retreat: genetic structure of a rapidly expanding *Larix decidua* population, Mol. Ecol., 20(3), 473–485, doi:10.1111/j.1365-294X.2010.04972.x, 2011.

R Core Team: R: A Language and Environment for Statistical Computing, [online] Available from: https://www.r-project.org/, 2015.

Resler, L. M., Butler, D. R. and Malanson, G. P.: Topographic Shelter and Conifer Establishment and Mortality in an Alpine Environment, Glacier National Park, Montana, Phys. Geogr., 26(2), 112–125, doi:10.2747/0272-3646.26.2.112, 2005.

Ritchie, J. C. and MacDonald, G. M.: The Patterns of Post-Glacial Spread of White Spruce, Source J. Biogeogr. J. Biogeogr., 13(13), 527–540 [online] Available from: http://www.jstor.org/stable/2844816, 1986.

Roberts, D. R. and Hamann, A.: Climate refugia and migration requirements in complex landscapes, Ecography (Cop.)., 39(12), 1238–1246, doi:10.1111/ecog.01998, 2016.

Schlötterer, C.: Evolutionary dynamics of microsatellite DNA, Chromosoma, 109(6), 365–371, doi:10.1007/s004120000089, 2000.

Schnabel, A., Nason, J. D. and Hamrick, J. L.: Understanding the population genetic structure of *Gleditsia triacanthos* L.: Seed dispersal and variation in female reproductive success, Mol. Ecol., 7(7), 819–832, doi:10.1046/j.1365-294x.1998.00397.x, 1998.

Shiyatov, S. G. and Mazepa, V. S.: Climate-driven dynamics of the forest-tundra vegetation in the Polar Ural Mountains, Contemp. Probl. Ecol., 4(7), 758–768, doi:10.1134/S1995425511070071, 2012.

Shiyatov, S. G., Terent'ev, M. M. and Fomin, V. V.: Spatiotemporal dynamics of forest-tundra communities in the Polar Urals, Russ. J. Ecol., 36(2), 69–75, doi:10.1007/s11184-005-0051-9, 2005.

Sidorova, O. V. O. V. O. V., Siegwolf, R. T. W., Saurer, M., Naurzbaev, M. M., Shashkin, A. V. A. V. and Vaganov, E. a.: Spatial patterns of climatic changes in the Eurasian north reflected in Siberian larch tree-ring parameters and stable isotopes, Glob. Chang. Biol., 16(3), 1003–1018, doi:10.1111/j.1365-2486.2009.02008.x, 2010.

Slate, J., Marshall, T. C. and Pemberton, J. M.: A retrospective assessment of the accuracy of the paternity inference programm CERVUS, Mol. Ecol., 9, 801–808, doi:10.1046/j.1365-294X.2000.00930.x, 2000.

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

Snell, R. S. and Cowling, S. A.: Consideration of dispersal processes and northern refugia can improve our understanding of past plant migration rates in North America, J. Biogeogr., 42(9), 1677–1688, doi:10.1111/jbi.12544, 2015.

Steinitz, O., Troupin, D., Vendramin, G. G. and Nathan, R.: Genetic evidence for a Janzen-Connell recruitment pattern in reproductive offspring of *Pinus halepensis* trees., Mol. Ecol., 20(19), 4152–4164, doi:10.1111/j.1365-294X.2011.05203.x, 2011.

Stewart, J. R. and Lister, A. M.: Cryptic northern refusia and the origins of the modern biota, Trends Ecol. Evol., 16(11), 608–613, 2001.

Stewart, J. R., Lister, A. M., Barnes, I. and Dalen, L.: Refugia revisited: individualistic responses of species in space and time, Proc. R. Soc. B Biol. Sci., 277(1682), 661–671, doi:10.1098/rspb.2009.1272, 2010.

Stoehr, M. U.: Seed production of western larch in seed-tree systems in the southern interior of British Columbia, For. Ecol. Manage., 130(1–3), 7–15, doi:10.1016/S0378-1127(99)00173-5, 2000.

Svenning, J.-C., Gravel, D., Holt, R. D., Schurr, F. M., Thuiller, W., Münkemüller, T., Schiffers, K. H., Dullinger, S., Edwards, T. C., Hickler, T., Higgins, S. I., Nabel, J. E. M. S., Pagel, J. and Normand, S.: The influence of interspecific interactions on species range expansion rates, Ecography (Cop.)., 37(12), 1198–1209, doi:10.1111/j.1600-0587.2013.00574.x, 2014.

Szmidt, A. E., Aldén, T. and Hällgren, J. E.: Paternal inheritance of chloroplast DNA in *Larix*, Plant Mol. Biol., 9(1), 59–64, doi:10.1007/BF00017987, 1987.

Väliranta, M., Kaakinen, A., Kuhry, P., Kultti, S., Salonen, J. S. and Seppä, H.: Scattered late-glacial and early Holocene tree populations as dispersal nuclei for forest development in north-eastern European Russia, J. Biogeogr., 38(5), 922–932, doi:10.1111/j.1365-2699.2010.02448.x, 2011.

Wagner, S., Gerber, S. and Petit, R. J.: Two highly informative dinucleotide SSR multiplexes for the conifer *Larix decidua* (European larch)., Mol. Ecol. Resour., 12(4), 717–25, doi:10.1111/j.1755-0998.2012.03139.x, 2012.

Walker, D. A., Raynolds, M. K., Daniëls, F. J. A. A., Einarsson, E., Elvebakk, A., Gould, W. A., Katenin, A. E., Kholod, S. S., Markon, C. J., Melnikov, E. S., Moskalenko, N. G., Talbot, S. S., Yurtsev, B. A. and Team, the other members of the C.: The circumpolar Arctic vegetation map, J. Veg. Sci., 16(3), 267–282, doi:10.1658/1100-9233(2005)016[0267:TCAVM]2.0.CO;2, 2005.

Wessel, P. and Smith, W. H. F.: A global, self-consistent, hierarchical, high-resolution shoreline database, J. Geophys. Res.

Solid Earth, 101(B4), 8741–8743, doi:10.1029/96JB00104, 1996.

Wieczorek, M., Kruse, S., Epp, L. S., Kolmogorov, A., Nikolaev, A. N., Heinrich, I., Jeltsch, F., Pestryakova, L. A., Zibulski, R. and Herzschuh, U.: Dissimilar responses of larch stands in northern Siberia to increasing temperatures-a field and simulation based study, Ecology, 98(9), 2343–2355, doi:10.1002/ecy.1887, 2017.

Willis, K. J. and Van Andel, T. H.: Trees or no trees? The environments of central and eastern Europe during the Last Glaciation, Quat. Sci. Rev., 23(23–24), 2369–2387, doi:10.1016/j.quascirev.2004.06.002, 2004.

S1 Genetic analyses

Table S1. Summary table of the eight loci applied for the 11 subpopulations from the Taymyr Peninsula, sorted by decreasing population genetic differentiation value F_{ST}. Observed and expected heterozygosity are given by HO and HE, respectively.

No.	***Locus*** [1]	**Multiplex** [2]	**TAG** [3]	**Observed fragment length (bp)**	***Number of alleles***
1	bcLK253	1	Q3	211-247	16.99±0.39
2	Ld101	1	Q4	196-236	15.74±0.79
3	bcLK228	2	Q4	133-269	18.70±0.66
4	bcLK189	3	Q2	152-242	33.39±1.50
5	bcLK211	1	Q2	194-250	22.97±1.09
6	Ld42	3	Q4	187-201	7.86±0.35
7	bcLK056	2	Q1	154-256	31.79±1.05
8	bcLK263	2	Q2	198-280	39.77±0.96

[1] *Locus – marker names beginning 'bcLK' are developed by Isoda and Watanabe* (2006) *and those with 'Ld' by Wagner et al.,* (2012); [2] *Multiplex – number indicates the three primer mixes applied in a simultaneous PCR;* [3] *TAG - TAG – tailing sequence at forward primer: Q1 = TGTAAAACGACGGCCAGT* (Schuelke, 2000)*; Q2 = TAGGAGTGCAGCAAGCAT; Q3 = CACTGCTTAGAGCGATGC; Q4 = CTAGTTATTGCTCAGCGGT (Q2–Q4, after* Culley *et al.* (2008)*)).*

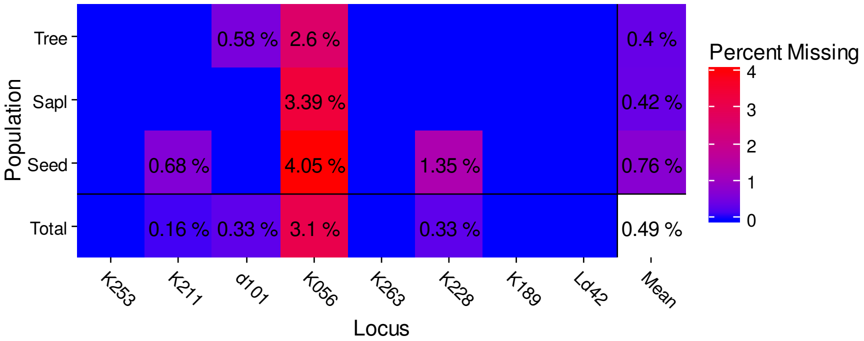


Figure S1. Fraction of missing alleles for each of three height classes – tree, sapling (Sapl), and seedling (Seed) (y-axis) and locus (x-axis) within each height class and the average value.

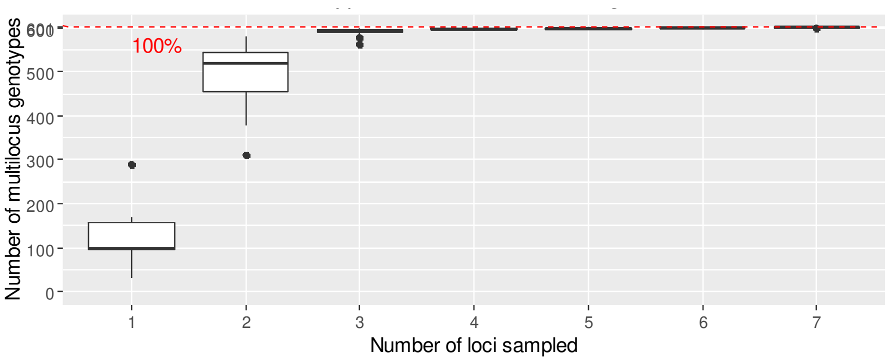


Figure S2. Genotype accumulation curve showing convergence at 5-7 loci from which nearly all 601 tested individuals can be differentiated.

S1.1 Allele diversity

S1.1.1 Introduction

The number of alleles per loci was analysed separately in three height classes: ‘seedling’ <0.4 m, ‘sapling’ – taller than seedlings but <2 m, and ‘tree’ >2 m. For the analyses, we resampled the dataset to avoid errors introduced by sample size. This was achieved by constructing 100 datasets from 30 randomly selected individuals of each height class. To check whether the loci were under the null expectation of the Hardy-Weinberg equilibrium, χ^2-tests were performed on the observed allele frequencies (‘hw.test’-function in ‘pegas’-library version 0.9 (Paradis, 2010)).

To exclude errors introduced by clonal reproduction we used clone-censored datasets for the analyses. By using all eight loci we could distinguish between all genotyped individuals. We identified 601 separate individuals and 11 clones (Fig. 3a). The members of one genetically identical group were up to 30 m distant from each other; similar distances were found for black spruce stunted forms (Gamache et al., 2003; Laberge et al., 2000).

S1.1.2 Results

The number of alleles per locus was nearly equal among all height classes with two exceptions at locus bcLK189 and bcLK263, at which the allele number was slightly smaller for seedlings. Individuals of all height classes showed significant heterozygote deficits with an observed mean of ~0.69 and an expected heterozygosity of ~0.86 (p<0.001, Table S2, Fig.~~ure~~ S4). At two loci (bcLK253 and bcLK263) observed values were close to the expected ratio and thus did not differ significantly from the Hardy-Weinberg equilibrium (Table S2).

S1.1.3 Discussion

The analysed tree stand is characterised by a high gene diversity (number of alleles and expected heterozygosity of ~86%) compared to other studies which used the same or parts of the same markers (Babushkina et al., 2016; Oreshkova et al., 2013; Pluess, 2011). Nevertheless, we observe a heterozygote deficit, which results in significant deviations from Hardy-Weinberg equilibrium, even though the analysed trees grew in a large area (one hectare). This was observed in the treeline area, spanning from dense forest to single-tree stands on the southern Taymyr Peninsula and which seems to be unaffected by the sampling area (Kruse et al., 2018). In general, this can be indicative of a higher degree of inbreeding among individuals and thus local recruitment outweighs immigration (Arenas et al., 2012; Hartl and Clark, 2007), although no straightforward pattern arises from the comparison of heterozygosity values (mean over all loci) among the three height classes (trees, saplings, seedlings). Nevertheless, in detail, the amount of alleles in seedlings is lower at two loci, for which also the observed heterozygosity is lower than for the other two height classes. This trend was expected for seedlings at all loci, because younger cohorts typically show depressed heterozygosity, caused by the higher probability of local reproduction (Addisalem et al., 2016; Moran and Clark, 2012). Subsequently, due to self-thinning, selection takes place, generally preferring fitter individuals – assuming heterozygotes are generally fitter (heterosis effect, for example Babushkina et al., 2016) one expects the older an individual is, the fitter it is compared to other competitors.

Table S2. Heterozygosity values for each locus by height class. The analyses are based on 100 resampled datasets, rarefied to 30 individuals.

Locus	Trees		Saplings		Seedlings	
	HO [%]	HE [%]	HO [%]	HE [%]	HO [%]	HE [%]
Ld101	55.9±8.5	77.4±4.8	53.8±8	79.7±3.8	60.8±8.2	78.8±3.6
bcLK056	62.5±8.9	91±1.4	55.2±8.2	90.1±1.6	64.2±7.4	91±1.1
bcLK189	79.1±6.8	88.8±2	70.2±7	89.3±1.6	72.8±6.5	88.3±1.5
bcLK211	68.6±7.2	88.5±2.5	63.7±6.7	89.4±1.9	66.4±6.7	89.2±1.8
bcLK228	70.8±7.8	87.9±1.6	68.5±7.5	88.4±1.3	65.4±7.5	89±1.4
bcLK253	80.1±8	83.8±2.5	80.3±6.1	83.8±2.6	81.2±6.4	83.3±2.7
bcLK263	90.1±4.8	93.8±0.8	89.6±4.5	93.8±0.7	86.3±5.3	92.6±0.9
Ld42	54.3±8.6	76±3.2	64.3±8.1	74.9±2.9	53±7.8	77.5±2.7
All	**70.2±7.6**	**85.9±2.3**	**68.2±7**	**86.2±2.1**	**68.7±7**	**86.2±2**

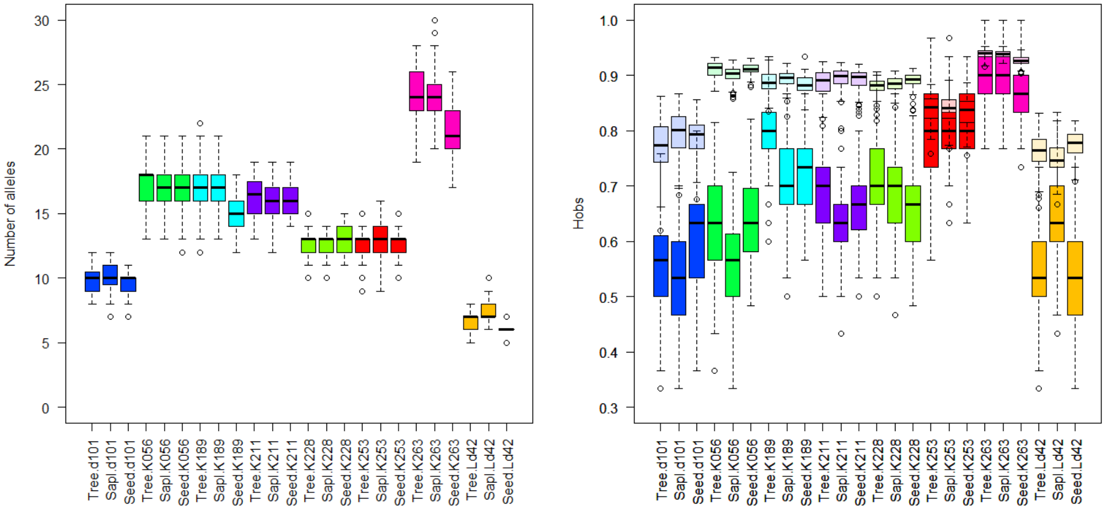


Figure S4. Left: Number of alleles, Right: Observed (H_O) and expected (H_E) heterozygosity. Based on a rarefied dataset of 30 individuals.

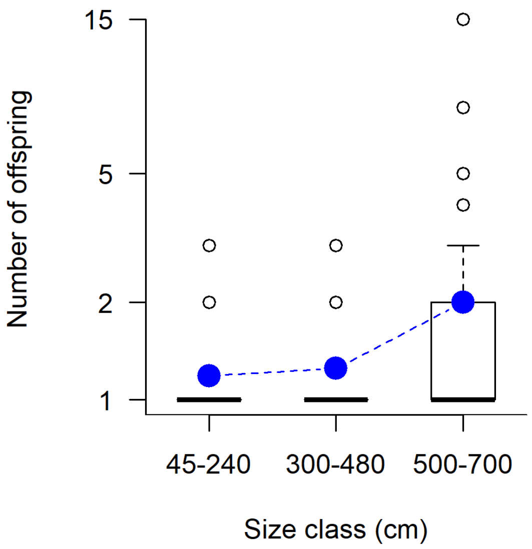


Figure S5. Number of offspring assigned to a single parent in three size classes. Filled circles: mean values.

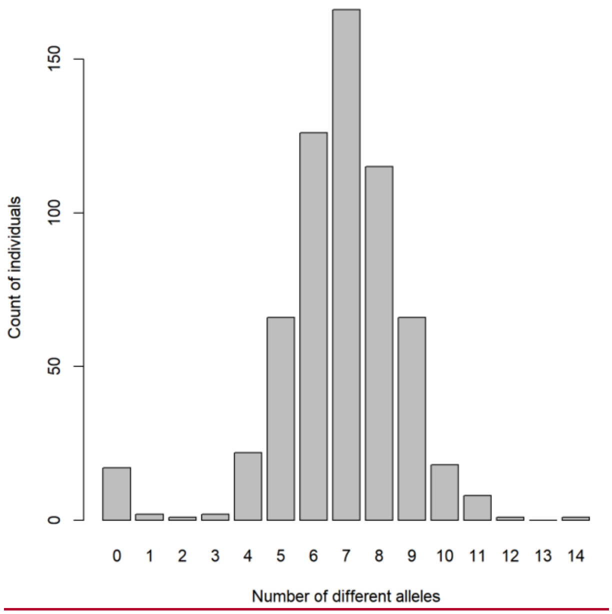


Figure S6. For each genotyped individual sample the smallest number of different alleles to the other samples was binned into 0 to a maximum of 16 alleles.

S2 Model adaptation

S2.1 Program code adaptation

S2.1.1 Seed dispersal function improvements

For each dispersed seed the wind direction is randomly drawn from vegetation period wind data of the year of its dispersal. The ballistic maximal flight distance E_0 (Equation 1) is estimated by species-specific size parameters following the approach of Matlack (1987), where V_h is defined as the horizontal wind speed and is chosen corresponding~~ly~~ to the wind direction in the model. The release height H_t is ~~roughly~~ estimated at 75% of the individual's height. V_d is the descent rate for seeds and is estimated for *Larix gmelinii* by a linear regression using species data from~~of~~ Matlack (1987). F~~ f~~or species having wing-scales attached to the seeds,~~: –~~ this rate can be calculated by $V_d = 0.0032 * \sqrt{w} + 0.4807$~~–~~ and is 0.86 m s^{-1}, with the wing loading w (Matlack 1987) for *L. gmelinii*. ~~,~~The variable w is calculated by dividing the average seed weight (in microdyne) of 3.5 mg (Heit and Eliason, 1940; Lukkarinen et al., 2009) by the propagule area of 0.2 cm^2 (Fu et al., 1999).

$$E_0 = V_h \frac{H_t}{V_d} \qquad (1)$$

This variable E_0 controls the standard deviation of the Gaussian term in the dispersal function of the model which is named originally '*width*' in Equation 5 in Kruse *et al.* (2016)), consisting of the two dispersal function terms $DL_{gaussian}(rn) = \sqrt{2 * {E_0}^2 * -1 * log(rn)}$ and $DL_{fat-tailed}(rn) = rn^{(-1*(1+\propto))}$, with rn – random number uniformely distributed between 0 and 1, $sdist$ – distance parameter for fitting and $\propto$ - scaling parameter for the fat tail of the function:

$$DL_{gaussian,fat-tail}(rn) = sdist * 0.5 * \left(\left(0.5 * DL_{gaussian}(rn)\right) + \left(2 * DL_{fat-tailed}(rn)\right)\right) \qquad (2)$$

S2.1.2 Growth function

The tree growth now depends only on July temperature, because climate-tree ring-width comparisons showed no significant influence of precipitation (~~data not shown~~Epp et al., 2018). With the species-specific linear regression coefficients we estimate the simulated tree growth in a year by $julindex = \left(\frac{0.078}{1+e^{14,825-Julytemperature}}\right) + 0.108$, which was further processed to the scale factor $weatherfactor = \frac{julindex-minimalgrowth}{maximalgrowth-minimalgrowth}$.

S2.1.3 Active layer thickness influences mortality

The influence of the active layer thaw depth on the diameter growth of larch trees is estimated based on the results of Nakai *et al.* (2008). It describes a linear relationship allowing 100% diameter growth at 100 cm thaw depth and only 10% when reaching 10 cm, which is the minimum value for *L. gmelinii*. The active layer thickness ALT (Equation 3) is estimated in metres for each year with the Stefan Formula, following simplifications by Hinkel and Nicholas (1995). It is determined by soil property parameter~~ies~~ fe=0.050 (Global Land Cover Characterization, Zhang et al., 2005) and the cumulative sum of daily temperatures exceeding the freezing points DDT:

$$ALT(year) = 1.0 - fe * \sqrt{DDT(year)} \quad (3)$$

Table S4. Overview of model parameters and processes for *L. gmelinii* individuals that differ from the original version (Kruse *et al.* 2016).

Parameter		Value and dimension	References
Growth			
Quadratic term of the equation for diameter growth rate	b	-0.003 ln(cm) cm^{-2}	data-based estimate similar to Fyllas *et al.* (2010)
Linear term of the growth function	a	0.030 ln(cm) cm^{-1}	
Constant term of the growth function	c	-1.98 ln(cm)	
Seed production, dispersal and establishment			
Factor of seed productivity	f_S	8	literature-based estimate (Kruklis & Milyutin, 1977, cited in Abaimov, 2010)
Background germination rate	$f_{Background\ Germination}$	0.01	estimated
Horizontal seed dispersal distance depended on actual wind, or for at wind speed of 10 km/h	E_0	variable, 60.1 m	estimated after Matlack (1987)
Seed descent rate	V_d	0.86 m s^{-1}	estimated descent rate based on Matlack (1987)
Mortality			
Background mortality rate	$m_{Background}$	0.0001 yr^{-1}	data-based estimate
Current tree growth influence factor on tree mortality	$f_{Growth\ Mortality}$	0.0	estimated
Weather influence factor on tree mortality	$f_{Weather\ Mortality}$	0.1	estimated
Density influence factor on tree mortality	$f_{Density\ Mortality}$	2.0	estimated
Seed fertility	$age_{max,seeds,L.gmelinii}$	2 yrs	Ban *et al.* (1998)
Mean temperature of the coldest month (January) at the border of the species‘ geographical range	$T_{min,L.gmelinii}$	-45 °C	Shugart *et al.* (1992)
Exponent scaling the height influence on tree mortality	$yexp$	0.2	estimated
Weather processing			
Exponent scaling the influence of surrounding density for a tree	$e_{density,basal\ influence}$	0.1	estimated
Exponent scaling the density value	$e_{density,tree-tile}$	0.5	estimated

S2.2 Simulation results

To fit the simulated seed effective dispersal distance to observations (Fig. 5) ~~To~~ we explored potential settings (I) to decrease the amount of recruitment close to the mother trees, (II) to shift the effective dispersal peak by 2-3 m and (III) to increase the effective recruitment at medium distances (~30-40 m). Therefore, we tuned two kinds of processes: parameters that determine the seed dispersal (model code: bcdopwxyzDEFGHIJK, Table S5) and tree density and parameters that set the impact of the tree's mortality (efghijklmnAC), or both (qrstuv) (details on individual adaptations in Table S5).

Of all 36 different simulations, some parameters decreased the amount of near mother effective seed dispersal (I) (cdo-zD-I) of which only (o-z) decreased the distance of up to 2 m based on the shifted dispersal function, while an increase in the distance parameter of the Gaussian-function peak improves the simulated function strongly (D-I) (Figure S8). Of these a shift towards farther distances and an increase in medium distances (II+III) was achieved with adaptations of the dispersal function only (oq-sx-zD-I), whereas the others only shifted the peak to ~5 m with a decrease for medium distances (pt-w). Nevertheless, the sum of squares of deviations from the observed pattern was improved in these candidates by a few sets (yzD-I) to within 66% to 82% of the reference run. The model performed best with parameter set "I" which is a combination of an adjusted dispersal function and increased seed production rate (Figure S7, Figure S8, Table S5). These sets increase the distance of dispersed seeds from the mother tree and the probability of a recruit growing at medium distances from its source was increased as well. Still the ratio of on-site recruitment was lower than observed (between 45.70 and 46.70 compared to observed 56.77%). This was improved by other simulations (qt-wJ) but their general performance (lower correlation coefficients, Table S5) was weaker than the reference simulation without parameter changes or adaptations of the model (a).

S2.3 Discussion of the simulation improvement

~~In parts w~~We achieved a good fit when increasing the peak of the dispersal function in the model to longer distances. The models where the distance from the centre of the distribution was shifted to 4 m improved the simulated effective seed dispersal distances best ("I"). However, the ratio between on-site recruitment and introductions from the exterior is around 10% lower than observed. This was not improved by the best-performing parameter set, but could be improved when changing the density competition, especially for small life stages ("t"). Combinations of both were tested but results ~~were~~ strongly deviated from observed situations ("C"). This smaller ratio points to an unrealistically high long-distance dispersed seed fraction. Here, we focus on the effective seed dispersal distance at short distances. Nevertheless, long-distance dispersal should be improved too, especially if one aims to conduct simulation studies over larger extents. It could be improved by decreasing the fat tail probability of the exponential part in the dispersal function, or by manipulating the implementation of the wind speed influence to a nonlinear process, decreasing the distances for strong winds. We analysed only one area at the treeline, which improved our understanding of the processes incorporated in the simulation model, but this may overemphasise the effective seed dispersal of one subpopulation. Therefore further validation by more plot-based analyses is needed for the general function parameters.

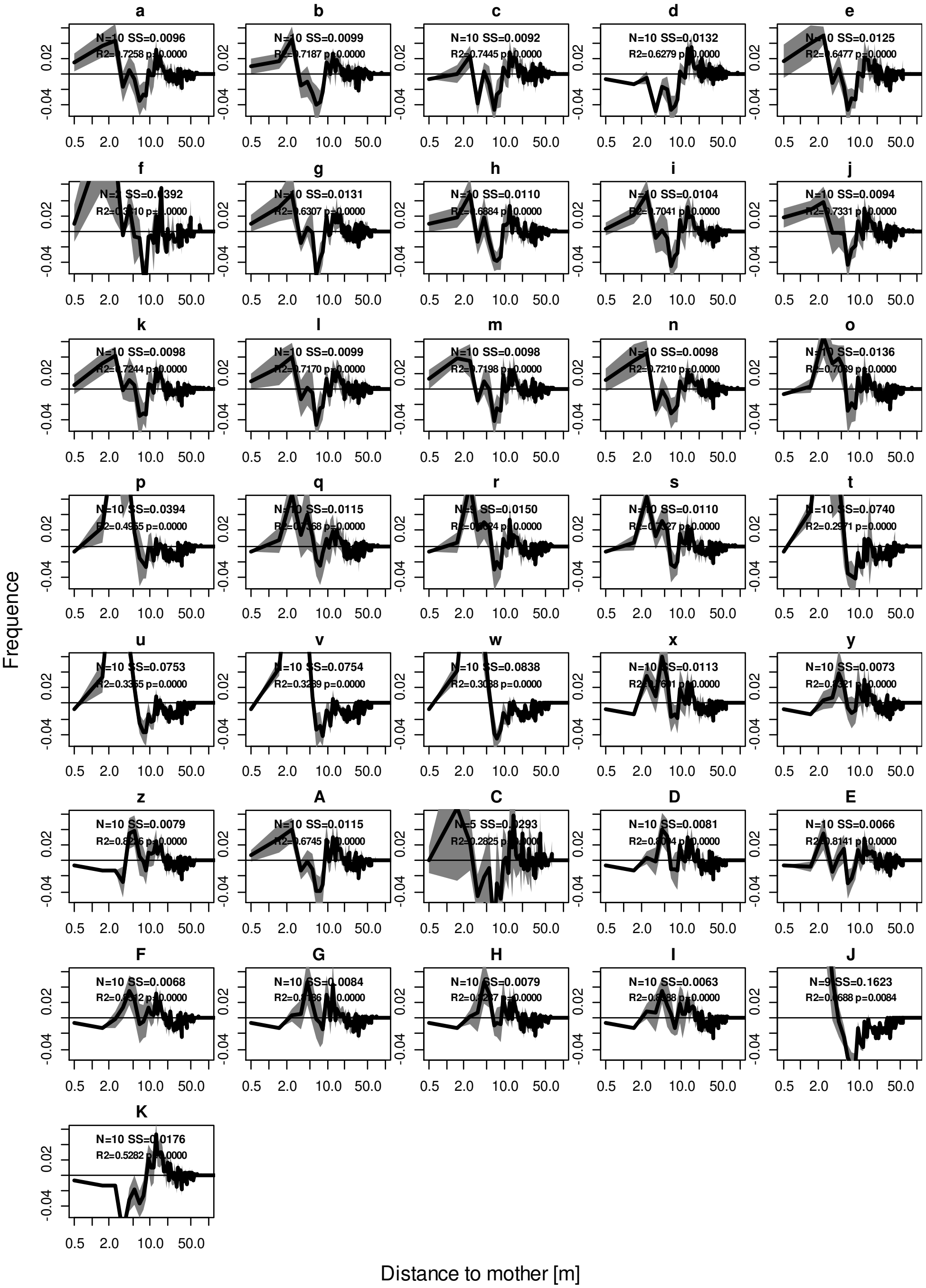
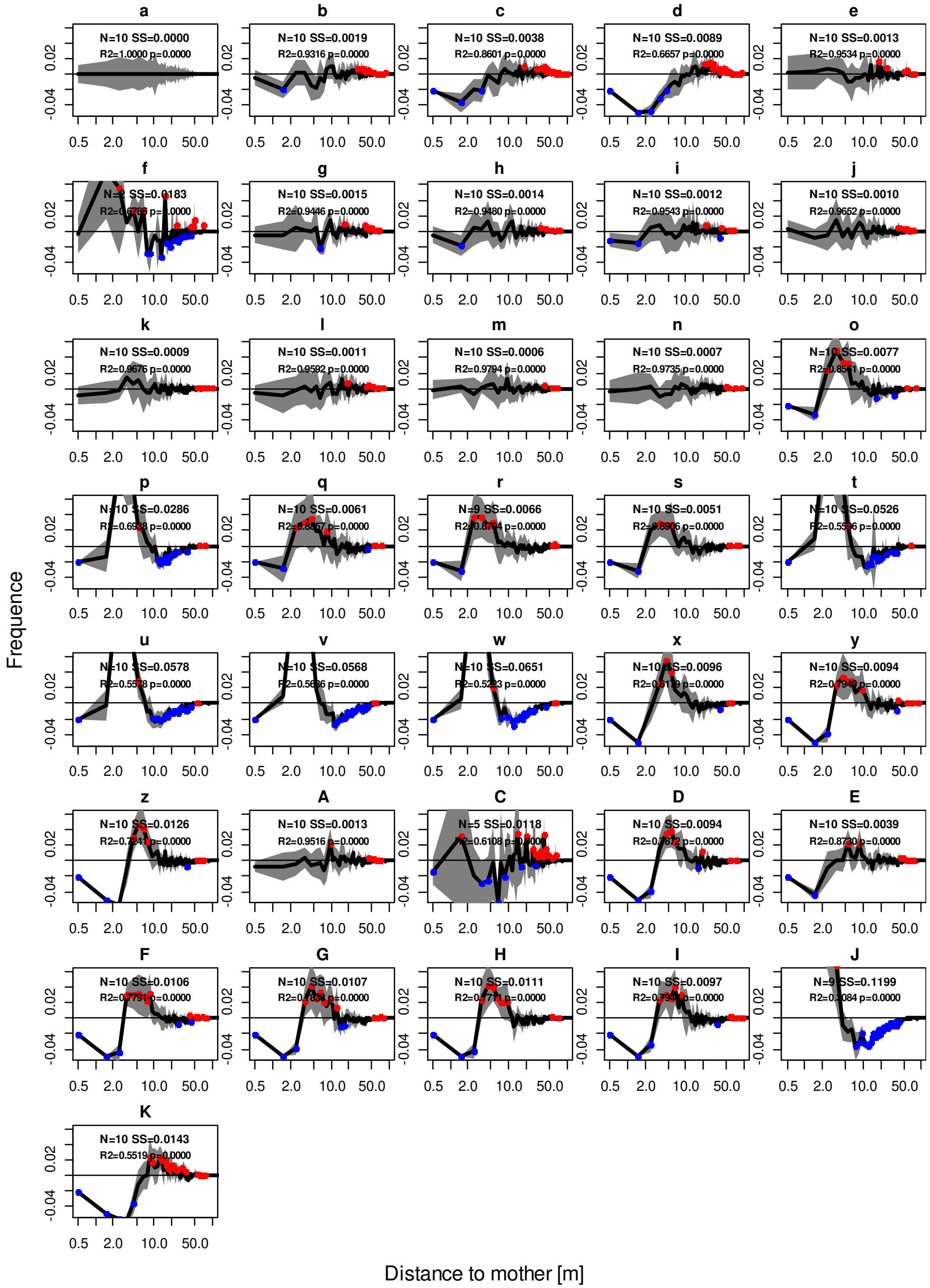


Figure S7. Deviations of simulated versus observed effective seed dispersal. SS – sum of squares, R^2 – square of correlation between the mean simulated series and the observed value, p – significance of correlation coefficient. Letters (a-zA-K) refer to a special simulation run s. Table S5 for details.



150 **Figure S8. Deviations of effective seed dispersal distances of all runs from the reference simulation "a". Grey areas are the standard deviation of all runs. Red and blue dots indicate values outside (above and below respectively) the standard deviation of the base run. Letters (a-zA-K) refer to a special simulation run s. Table S5 for details.**

Table S5. Results of effective seed dispersal in adapted simulations. Changes identifier: 1-dispersal function manipulation, 2-density calculation manipulation, 3-combinations of 1 and 2.

		Dispersal function			On-site recruitment ratio				Model parameters	
ID	**Changes**	**SS**	**r**	**r2**	**Ratio**	**SD**	**N>10 in center**	**p diff from obs**	**Adaptation and expected outcome**	**Parameter[1]**
a	-	0.0096	0.8519	0.7258	46.8%	1.3%	10	0.0000	*- reference run -*	Kruse *et al.* (2016)
b	1	0.0099	0.8478	0.7187	45.4%	1.4%	10	0.0000	longer dispersal distances	Sdist=1
c	1	0.0092	0.8629	0.7445	41.0%	1.2%	10	0.0000	longer dispersal distances	Sdist=5
d	1	0.0132	0.7924	0.6279	38.1%	1.0%	10	0.0000	longer dispersal distances	Sdist=10
e	2	0.0125	0.8048	0.6477	46.3%	1.7%	10	0.0000	larger distance to mother trees	$f_{DensityMortality}$~~mdichte~~=3
f	2	0.0392	0.5754	0.3310	45.6%	6.7%	3	**0.1023**	smaller distance to mother trees	f_{HAI}=5
g	2	0.0131	0.7942	0.6307	44.8%	2.3%	10	0.0000	larger distance to mother trees	f_{HAI}=15
h	2	0.0110	0.8297	0.6884	45.9%	2.6%	10	0.0000	less exclusion close to mother tree	$e_{density,basal\ influence}$~~dichtewertmanipulator exp~~=0.05
i	2	0.0104	0.8391	0.7041	48.4%	2.1%	10	0.0000	higher exclusion close to mother tree	$e_{density,basal\ influence}$~~dichtewertmanipulator exp~~=0.15
j	2	0.0094	0.8562	0.7331	46.2%	1.7%	10	0.0000	higher exclusion close to mother tree	densitysmallweighing=1
k	2	0.0098	0.8511	0.7244	47.0%	1.2%	10	0.0000	less exclusion close to mother tree	densitytreetile=0
l	2	0.0099	0.8467	0.7170	51.5%	2.2%	10	0.0000	higher exclusion close to mother tree	densitytreetile=1
m	2	0.0098	0.8484	0.7198	46.9%	0.7%	10	0.0000	higher exclusion close to mother tree	densitytiletree=1
n	2	0.0098	0.8491	0.7210	47.7%	1.3%	10	0.0000	higher exclusion close to mother tree	densitymaxreduction=1
o	1	0.0136	0.8390	0.7039	47.9%	2.2%	10	0.0000	more distant from centre and more intense peak	Sdist=1 + $r_{GaussExpDisp}$=1.0 + $d_{GaussCentre}$=2.0
p	1	0.0394	0.7039	0.4955	49.9%	3.2%	10	0.0001	shorter dispersal distances	o + $d_{GaussDistance}=D*0.5$
q	3	0.0115	0.8584	0.7368	52.0%	2.0%	10	0.0000	higher exclusion close to mother tree	o + $e_{density,tree-tile}$~~densitytreetile~~=1
r	3	0.0150	0.8139	0.6624	49.3%	2.8%	9	0.0000	less exclusion close to mother tree	o + $e_{density,basal\ influence}$~~dichtewertmanipulator exp~~=0.15
s	3	0.0110	0.8560	0.7327	44.7%	1.2%	10	0.0000	higher exclusion close to mother tree	o + $e_{density,tile-tree}$~~densitytiletree~~=3
t	3	0.0699	0.5795	0.3358	52.4%	4.4%	10	**0.0113**	increased tree density	o + treedensity^0.9
u	3	0.0706	0.5957	0.3548	50.9%	2.1%	10	0.0000	increased tree density	o + treedensity^0.95
v	3	0.0788	0.5627	0.3166	51.0%	1.3%	5	0.0006	weakened tree density	o + treedensity^1.1
w	1	0.0838	0.5512	0.3038	52.0%	2.0%	10	0.0000	shortened dispersal distance	o + $d_{GaussDistance}=D^{0.5}$
x	1	0.0113	0.8718	0.7601	47.7%	1.4%	10	0.0000	more distant from centre and more intense peak	Sdist=1 + $r_{GaussExpDisp}$=1.0 + $d_{GaussCentre}$ =3.0
y	1	0.0073	0.9122	**0.8321**	47.4%	1.3%	10	0.0000	more distant from centre and more intense peak	Sdist=1 + $r_{GaussExpDisp}$=1.0 + $d_{GaussCentre}$ =4.0
z	1	0.0079	0.9070	0.8226	46.7%	2.2%	10	0.0000	more distant from centre and more intense peak	Sdist=1 + $r_{GaussExpDisp}$=1.0 + $d_{GaussCentre}$ =5.0
A	2	0.0115	0.8213	0.6745	48.6%	2.9%	10	0.0000	higher exclusion close to mother tree	linear density 0-200 cm 1-0 extra mortality
C	2	0.0293	0.5315	0.2825	50.2%	7.5%	7	**0.0613**	higher exclusion close to mother tree	negative quadratic density 0-200 cm 5-0 extra mortality
D	1	0.0081	0.8997	0.8094	45.9%	0.9%	10	0.0000	increased seed production higher on-site reproduction	y + f_s=16 (twice standard)
E	**1**	**0.0066**	0.9023	0.8141	45.7%	1.8%	10	0.0000	more distant shifted dispersal peak	o + $d_{GaussDistance}=D*1.5$
F	**1**	**0.0068**	0.9226	**0.8512**	46.2%	1.2%	10	0.0000	increased seed production higher on-site reproduction	y + f_s=12
G	1	0.0084	0.9048	0.8186	46.6%	1.6%	10	0.0000	increased seed production higher on-site reproduction	y + f_s=9
H	1	0.0079	0.9103	0.8287	45.7%	1.2%	10	0.0000	increased seed production higher on-site reproduction	y + f_s=10
I	**1**	**0.0063**	**0.9267**	**0.8588**	46.4%	1.7%	10	0.0000	increased seed production higher on-site reproduction	y + f_s=11

J	1	0.1623	0.2624	0.0688	54.9%	4.0%	9	0.2023	higher on-site reproduction	a + no exponential dispersal
K	1	0.0176	0.7268	0.5282	44.8%	1.3%	10	0.0000	higher on-site reproduction	I + no exponential dispersal

1 – abbreviations following Kruse *et al.* (2016), Epp et al. (2018), and, Kruse et al. (2018)

References

Abaimov, A. P.: Permafrost Ecosystems, edited by A. Osawa, O. A. Zyryanova, Y. Matsuura, T. Kajimoto, and R. W. Wein, Springer Netherlands, Dordrecht., 2010.

Addisalem, A. B., Duminil, J., Wouters, D., Bongers, F. and Smulders, M. J. M.: Fine-scale spatial genetic structure in the frankincense tree *Boswellia papyrifera* (Del.) Hochst. and implications for conservation, Tree Genet. Genomes, 12(5), 86, doi:10.1007/s11295-016-1039-2, 2016.

Arenas, M., Ray, N., Currat, M. and Excoffier, L.: Consequences of range contractions and range shifts on molecular diversity, Mol. Biol. Evol., 29(1), 207–218, doi:10.1093/molbev/msr187, 2012.

Babushkina, E. A., Vaganov, E. A., Grachev, A. M., Oreshkova, N. V., Belokopytova, L. V., Kostyakova, T. V. and Krutovsky, K. V.: The effect of individual genetic heterozygosity on general homeostasis, heterosis and resilience in Siberian larch (*Larix sibirica* Ledeb.) using dendrochronology and microsatellite loci genotyping, Dendrochronologia, 38, 26–37, 2016.

Ban, Y., Xu, H., Bergeron, Y. and Kneeshaw, D. D.: Gap regeneration of shade-intolerant Larix gmelini in old-growth boreal forests of northeastern China, J. Veg. Sci., 9(4), 529–536, doi:10.2307/3237268, 1998.

Culley, T. M., Weller, S. G., Sakai, A. K. and Putnam, K. A.: Characterization of microsatellite loci in the Hawaiian endemic shrub *Schiedea adamantis* (Caryophyllaceae) and amplification in related species and genera, Mol. Ecol. Resour., 8(5), 1081–1084, doi:10.1111/j.1755-0998.2008.02161.x, 2008.

Epp, L. S., Kruse, S., Kath, N. J., Stoof-Leichsenring, K. R., Tiedemann, R., Pestryakova, L. A. and Herzschuh, U.: Temporal and spatial patterns of mitochondrial haplotype and species distributions in Siberian larches inferred from ancient environmental DNA and modeling, Sci. Rep., 8(1), 17436, doi:10.1038/s41598-018-35550-w, 2018.

Fu, L., Li, N. and Mill, R. R.: Sections on Cephalotaxaceae, Ginkgoaceae and Pinaceae, in Flora of China, edited by W. Zheng-yi and P. H. Raven, pp. 11–52, Science Press; St. Louis: Missouri Botanical Garden, Beijing., 1999.

Fyllas, N. M., Politi, P. I., Galanidis, A., Dimitrakopoulos, P. G. and Arianoutsou, M.: Simulating regeneration and vegetation dynamics in Mediterranean coniferous forests, Ecol. Model., 221(11), 1494–1504, doi:10.1016/j.ecolmodel.2010.03.003, 2010.

Gamache, I., Jaramillo-Correa, J. P., Payette, S. and Bousquet, J.: Diverging patterns of mitochondrial and nuclear DNA diversity in subarctic black spruce: imprint of a founder effect associated with postglacial colonization., Mol. Ecol., 12(4), 891–901 [online] Available from: http://www.ncbi.nlm.nih.gov/pubmed/12753210, 2003.

Hartl, D. L. and Clark, A. G.: Principles Of Population Genetics, 4th ed., Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts, USA., 2007.

Heit, E. C. and Eliason, E. J.: Coniferous tree seed testing and factors affecting germination and seed quality, Tech. Bull. New York State Agric. Exp. Stn., 255, 1–45, 1940.

Hinkel, K. M. and Nicholas, J. R. J.: Active layer thaw rate at a boreal forest site in central Alaska, USA, Arct. Alp.

Res., 1, 72–80, 1995.

Isoda, K. and Watanabe, A.: Isolation and characterization of microsatellite loci from *Larix kaempferi*, Mol. Ecol. Notes, 6(3), 664–666, doi:10.1111/j.1471-8286.2006.01291.x, 2006.

Kruse, S., Wieczorek, M., Jeltsch, F. and Herzschuh, U.: Treeline dynamics in Siberia under changing climates as inferred from an individual-based model for *Larix*, Ecol. Model., 338, 101–121, doi:10.1016/j.ecolmodel.2016.08.003, 2016.

Kruse, S., Epp, L. S., Wieczorek, M., Pestryakova, L. A., Stoof-Leichsenring, K. R. and Herzschuh, U.: High gene flow and complex treeline dynamics of *Larix* Mill. stands on the Taymyr Peninsula (north-central Siberia) revealed by nuclear microsatellites, Tree Genet. Genomes, 14(2), 19, doi:10.1007/s11295-018-1235-3, 2018.

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

Laberge, M.-J., Payette, S. and Bousquet, J.: Life span and biomass allocation of stunted black spruce clones in the subarctic environement, J. Ecol., 88, 584–593, 2000.

Lukkarinen, A. J., Ruotsalainen, S., Nikkanen, T. and Peltola, H.: The Growth Rhythm and Height Growth of Seedlings of Siberian (*Larix sibirica* Ledeb.) and Dahurian (*Larix gmelinii* Rupr.) Larch Provenances in Greenhouse Conditions, Silva Fenn., 43(1), 5–20, 2009.

Matlack, G. R.: Diaspore size, shape, and fall behavior in wind-dispersed plant species, Am. J. Bot., 74(8), 1150–1160, doi:10.2307/2444151, 1987.

Moran, E. V. and Clark, J. S.: Between-site differences in the scale of dispersal and gene flow in red oak, PLoS One, 7(5), e36492, doi:10.1371/journal.pone.0036492, 2012.

Nakai, Y., Matsuura, Y., Kajimoto, T., Abaimov, a. P., Yamamoto, S. and Zyryanova, O. a.: Eddy covariance CO2 flux above a Gmelin larch forest on continuous permafrost in Central Siberia during a growing season, Theor. Appl. Climatol., 93(3–4), 133–147, doi:10.1007/s00704-007-0337-x, 2008.

Oreshkova, N. V., Belokon, M. M. and Jamiyansuren, S.: Genetic diversity, population structure, and differentiation of Siberian larch, Gmelin larch, and Cajander larch on SSR-marker data, Russ. J. Genet., 49(2), 178–186, doi:10.1134/S1022795412120095, 2013.

Paradis, E.: Pegas: An R package for population genetics with an integrated-modular approach, Bioinformatics, 26(3), 419–420, doi:10.1093/bioinformatics/btp696, 2010.

Pluess, A. R.: Pursuing glacier retreat: genetic structure of a rapidly expanding *Larix decidua* population, Mol. Ecol., 20(3), 473–485, doi:10.1111/j.1365-294X.2010.04972.x, 2011.

Schuelke, M.: An economic method for the fluorescent labeling of PCR fragments A poor man's approach to genotyping for research and high-throughput diagnostics., Nat. Biotechnol., 18(February), 233–234, 2000.

Shugart, H. H., Leemans, R. and Bonan, G. B.: A Systems Analysis of the Global Boreal Forest, edited by H. H. Shugart, R. Leemans, and G. B. Bonan, Cambridge University Press, Cambridge, UK., 1992.

Wagner, S., Gerber, S. and Petit, R. J.: Two highly informative dinucleotide SSR multiplexes for the conifer *Larix decidua* (European larch)., Mol. Ecol. Resour., 12(4), 717–25, doi:10.1111/j.1755-0998.2012.03139.x, 2012.

Zhang, T., Frauenfeld, O. W., Serreze, M. C., Etringer, A., Oelke, C., McCreight, J., Barry, R. G., Gilichinsky, D., Yang, D., Ye, H. and others: Spatial and temporal variability in active layer thickness over the Russian Arctic drainage basin, J. Geophys. Res., 110(D16), D16101, doi:10.1029/2004JD005642, 2005.