Tree migration in the dynamic, global vegetation model LPJ-GM 1.0: Efficient uncertainty assessment and improved dispersal kernels

Deborah Zani[1,2], Veiko Lehsten[1,2], Heike Lischke[1]

[1]Dynamic Macroecology/Land Change Science, Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland
[2]Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden

Correspondence to: Deborah Zani (deborah.zani01@gmail.com)

Abstract. The prediction of species geographic redistribution under climate change (i.e. range shifts) has been addressed by both experimental and modelling approaches and can be used to inform efficient policy measures on the functioning and services of future ecosystems. Dynamic Global Vegetation Models (DGVMs) are considered state-of-the art tools to understand and quantify the spatio-temporal dynamics of ecosystems at large scales and their response to changing environments. They can explicitly include local vegetation dynamics relevant to migration (establishment, growth, seed production), species-specific dispersal abilities and the competitive interactions with other species in the new environment. However, the inclusion of more detailed mechanistic formulations of range shift processes may also widen the overall uncertainty of the model. Thus, a quantification of these uncertainties is needed to evaluate and improve our confidence in the model predictions. In this study, we present an efficient assessment of parameter and model uncertainties combining low-cost analyses in successive steps: local sensitivity analysis, exploration of the performance landscape at extreme parameter values, and inclusion of relevant ecological processes in the model structure. This approach was tested on the newly-implemented migration module of the state-of-the-art DGVM, LPJ-GM 1.0. Estimates of post-glacial migration rates obtained from pollen and macrofossil records of dominant European tree taxa were used to test the model performance. The results indicate higher sensitivity of migration rates to parameters associated with the dispersal kernel (dispersal distances and kernel shape) compared to plant traits (germination rate and maximum fecundity) and highlight the importance of representing rare long-distance dispersal events via fat-tailed kernels. Overall, the successful parametrization and model selection of LPJ-GM will allow simulating plant migration with a more mechanistic approach at larger spatial and temporal scales, thus improving our efforts to understand past vegetation dynamics and predict future range shifts in a context of global change.

1 Introduction

It is widely accepted that climate change is affecting the geographic distribution of species worldwide (Pecl et al., 2017; Lenoir et al., 2020). Especially for taxa with slow thermal adaptation such as trees, the ability of a species to track its optimal environment (i.e. range shift) will likely be the primary limitation of response to rapid climate change (Berg et al. 2010; Thompson and Fronhofer 2019). Genetic and spatial studies of pollen and macrofossils show that tree species have successfully

responded to past climatic changes by following northward thermal shifts after the retreat of ice sheets at the end of the Last Glacial Maximum (Huntley and Birks, 1983). The responses of contemporary vegetation seem to follow a similar trend towards the poles and the summits of mountains in previously cooler latitudes and elevations, respectively (Lenoir et al., 2020). Importantly, differences in migration abilities will result in new assemblages of communities via species-specific range

35 contractions and expansions, where range shifts will determine the threat of invasion by alien species and extinction of local species, which may in turn alter ecosystem services, such as carbon sequestration and wood production (Pecl et al., 2017). Thus, species-specific range dynamics need to be accounted for to aid policy measures targeting the protection of the functioning and services of future ecosystems.

So far, the prediction of species range shifts in response to climate change has been addressed by both experimental and

40 modelling approaches. Dynamic Global Vegetation Models (DGVMs) are considered state-of-the art tools to understand and predict the spatio-temporal dynamics of ecosystems and their response to changing environments (Snell et al., 2014; Briscoe et al., 2019). In particular, DGVMs have the potential to explicitly account for the processes involved in migration, including the demographic components (e.g. fecundity, population establishment and growth) along with species-specific dispersal abilities and the competitive interactions with other species in the new environment (Snell et al., 2014; Shifley et al., 2017).

45 Though the inclusion of more detailed mechanistic representations of the migration process can potentially improve the predictive power of a model, the larger number of simulated equations and parameters, each with its inherent uncertainty, may also increase the overall uncertainty of the model predictions (Snowling and Kramer, 2001). An assessment of these errors is therefore needed to increase our confidence in the model predictions, and/or to identify parameters and representations of ecological processes that require further improvement. Such an assessment is achieved by comparing model outputs with

50 independent observations, while quantifying the impact of ecological parameters and processes on the model predictions.

The considerable complexity of DGVMs often leads to model performance being estimated considering various sources of errors, including data inaccuracy, lack of detailed information at high temporal and spatial resolution, and a limited knowledge of the processes underlying the modelled system. Specifically, uncertainty in DGVMs is mainly attributed to 1) the appropriate (mathematical) representation of the ecological processes underlying the model (model uncertainty), 2) the estimation of the

55 high number of parameters, whose values are not readily measurable or are derived from data of limited sample size (parameter uncertainty), and 3) the influence of stochastic processes that may render average model predictions uninformative (inherent uncertainty) (Higgins et al., 2003a). Model uncertainty can be addressed by a comparison of the performance among modelling frameworks where different processes or process formulations are tested for their impact on the model output (e.g. Cheaib et al., 2012). Specifically to migration modelling, model uncertainty can be controlled by accurately translating recent ecological

60 evidences on the components of migration (demographic and dispersal processes) and their key drivers (e.g. climate, landscape, species interactions) into the model (Alexander et al., 2018; Tomiolo and Ward, 2018) in order to have a more appropriate representation of the ecological processes underlying migration. Generally, the dispersal sub-model is likely to be inadequately represented both as phenomenological (describing seed dispersal patterns via seed kernels) and mechanistic (explicitly describing dispersal processes) models (Higgins et al., 2003a). A seed dispersal kernel (i.e. the probability density function

65   describing the distribution of seeds after dispersal) can be inaccurate because of its failure to represent rare long-distance
dispersal events or multiple dispersal modes (e.g. seed transport via both wind and water) (Higgins et al., 2003b; Rogers et al.,
2019). On the other hand, mechanistic dispersal sub-models may introduce further uncertainties with the inclusion of a larger
set of parameters to estimate (e.g. wind velocity, gut retention; increase of parameter uncertainty) and of stochastic processes
(e.g. animal movement; increase of inherent uncertainty; Higgins et al., 2003a; Nathan et al., 2012). With respect to migration

70   forecasts, parameter uncertainty is mainly linked to the uncertainty of empirical estimates of demographic parameters driven
by complex climatic-vegetation dynamics (e.g. fecundity). Additionally, it can be attributed to the small sampling size of some
dispersal parameters, i.e. the limited available data on the shape of seed dispersal at large spatial scales, and the lack of
information on mechanistic dispersal parameters for many species (e.g. gut retention) (Higgins et al., 2003a). A sensitivity
analysis (SA) can be employed to quantify parameter uncertainty by systematically changing input parameter values and

75   measuring the corresponding response of the model output (Saltelli et al., 2000). Information on the influence of each parameter
on the model predictions can then be used to identify relevant parameters to retain in the modelling framework ("model
reduction"; e.g. Loehle, 2004). This allows exploring the relationship between each input parameter and model output, both as
directionality and magnitude, so to highlight range values or directions where the error between model output and observations
are progressively reduced (e.g. McKenzie et al., 2019). This can help to inform a further parametrization of relevant parameters,

80   where the main task of the parametrization is to find the best set of parameter values that minimize the difference between
model outputs and observational data as much as the inherent uncertainty of the model and of the observations allow (Stork et
al., 2020). Finally, inherent uncertainty can be controlled to some extent by limiting the inclusion of stochastic processes. For
migration forecasts, inherent uncertainty tends to increase for mechanistic representation of seed dispersal and, to a lesser
degree, for fat-tailed dispersal kernels (Clark et al., 2003; Higgins et al., 2003a).

85   A number of studies have conducted thorough assessments of parameter estimates and model uncertainties in dynamic
vegetation models, though mainly focusing on metrics of vegetation composition and structure (e.g. Zaehle and Sitch, 2005;
Wramneby et al., 2008; Pappas et al., 2013). However, there are only a few examples of how uncertainty in parameter and
model selection may impact migration forecasts. For example, Dullinger et al. (2004) assessed the relative influence of
temperature increase, dispersal ability and competition on the range expansion of a single species (shrubby pine) and found a

90   significant effect of dispersal on its expansion rate. More recently, Petter et al. (2020) conducted ensemble simulations of four
widely used forest landscape models that implement migration for multiple species [LandClim (Schumacher et al., 2004),
LANDIS (Mladenoff, 2004), TreeMig (Lischke et al., 2006) and iLand (Seidl et al., 2012)] to quantify the uncertainties
underlying their different model structures. Petter et al. (2020) found that different formulations of seed dispersal contributed
little to explain the variance across all model simulations compared to, for example, the use of different climate scenarios.

95   However, the formulation of seed dispersal is comparatively similar among the four models, i.e. seed dispersal is simulated
with a phenomenological (non-mechanistic) probability distribution function (kernel) using either a single (LandClim) or two-
part (LANDIS-II, iLand, TreeMig) negative exponential function, thus possibly explaining the little effect of dispersal
formulation on the overall variance. On the other hand, the use of different dispersal scenarios (roughly with and without seed

limitation) had a consistent impact on species composition for all models, in agreement with recent empirical evidence (Albrich

100 et al., 2020; Scherrer et al., 2020). This significant effect found for a relatively short simulated time span (100 years) and at the landscape level suggests that the uncertainties linked to the migration dynamics should be carefully evaluated in modelling studies and possibly at larger spatial and temporal extents.

In a previous study we coupled a dynamic migration module to a process-orientated ecosystem modelling framework (LPJ-GUESS), resulting in the model LPJ-GM 1.0, which allows simulating the migration of tree species while simultaneously

105 simulating their inter-specific interactions (Lehsten et al., 2019). As the aim of Lehsten et al. (2019) was the technical implementation of the model, an assessment of model and parameter uncertainties has still to be conducted on the new migration module. Indeed, a test simulation of the migration speed in this first study showed a significant underestimation with respect to historical estimates for the species *Fagus sylvatica* (European beech), thus highlighting the need to evaluate and increase the performance of the model.

110 In this study, we present an efficient uncertainty assessment of model selection and parameter estimates for the newly-implemented migration module of LPJ-GM 1.0. As one of the main challenges for an efficient uncertainty assessment of complex DGVMs is the high computational cost (in terms of CPU time) associated with both DGVM simulations and the parametrization effort at increasing number of parameters (e.g. Pappas et al., 2013), our approach aims to improve the current configuration and parameter estimates of LPJ-GM 1.0 with as little computational demand as possible. To this end, we initially

115 conducted a species-specific local sensitivity analysis (LSA) to assess the influence of each migration parameter with respect to the model output (i.e. migration rate), both for magnitude and linearity, while providing a measure of parameter uncertainty. The advantage of a LSA with respect to a global approach (GSA) is that it is relatively less costly in terms of CPU demand and simpler to implement and interpret, as the magnitude and direction of sensitivity indexes refer to individual variables. Next, we used information from the LSA on the direction and linearity of the effect of each parameter on the model output to

120 formulate an extreme value analysis (EVA). This approach allows to explore the response of the model output to collective variations of all parameters at their extreme range values, and thus can help to shrink, to some extent, the performance landscape in order to inform a more efficient model parametrization. Additionally, an EVA requires a low computational cost as it is independent of the parameter size. Next, we modified the model structure of the dispersal sub-model of LPJ-GM 1.0 in order to include different phenomenological formulations of seed dispersal based on prior knowledge (literature and/or expert).

125 In the last step, we compared different combinations of parameter estimates and dispersal formulations in terms of model performance and uncertainty, where we selected plausible combinations of parameter values based on insights from EVA on the model performance at the extremes of the parameter space. We then identified the set of parameter estimates that minimized the difference between simulated and observed migration rates for each species (parametrization) and the model structure with higher utility (model selection). The optimized model structure and parameter sets resulted in the model LPJ-GM 1.1.

2 Material and Methods

2.1 Observational data: estimates of past migration rates

Our parametrization routine requires independent estimates of migration rates to compare against simulated migration rate values from LPJ-GM 1.0. These estimates should be ideally derived for each species (which is implemented in the model) from available empirical data. For example, genetic assignment tests can be used to fit seed dispersal patterns (Manel et al., 2005; Goto et al., 2006; Klein et al., 2006; Moran and Clark; 2011) and, combined with spatio-temporal information from seed traps and parent trees, to derive rates of population spread over time (Beckman et al., 2020). Alternatively, rates of recent migration (over the last several generations) can be directly estimated using Bayesian inference on genotypic data, given a sufficient differentiation among tree populations (Wilson and Rannala, 2003). However, the empirical estimation of migration rates from current tree distributions can be problematic for a number of reasons: 1) empirical procedures are costly, leading to a lack of estimates for some major species; 2) the presence of multiple and large source populations or of continuously distributed species can complicate the correct assignment of seeds to mother trees; 3) empirical estimates are generally conducted at the local level and can have an inherent uncertainty given by stochastic processes (e.g. animal movements or behavior in the case of seed dispersal by animals; Higgins et al., 2003a; Nathan et al., 2012; Beckman et al., 2020). Furthermore, it is not clear which factors determine the observed variation in contemporary migration rates and how important each one is for the total variability. Potential drivers may vary at the site level or over time and include: climate forcing, local topography, habitat suitability and fragmentation, inherent species-specific dispersal ability, physical factors linked to seed dispersal (e.g. wind speed for wind-dispersed species), and a number of biotic factors (competition with other plants, herbivory and human disturbance) (Alexander et al., 2018; Tomiolo and Ward, 2018). Unknown driving processes and the above-described problems with contemporary estimates can inflate both model and inherent uncertainties, thus decreasing our confidence in parameter estimates.

An alternative to empirical estimates from contemporary distributions is to derive independent migration rates from paleo-records of European forest expansion after the Last Glacial Maximum (LGM) (Huntley and Birks, 1983). Estimates of post-LGM vegetation spread have a number of advantages: 1) data may be available for different time points, thus allowing a direct inference of the speed; 2) estimates are available for most of the major European tree species; 3) initial sources of population spread are confined to specific areas, i.e. glacial refugia; 4) historical estimates are derived from long-term continental movements and are thus less biased by stochastic local processes; 5) the order of tree species expansion is roughly known and can help to define the role of competition in tree migration; 6) the role of climate in controlling the rates of tree expansion after the LGM is generally assumed to be less influential than intrinsic dispersal ability as indicated by the wide-spread phenomenon of migration lag (i.e. the delayed arrival of a species into a newly-suitable habitat; Feurdean et al., 2013; Huntley et al. 2013; Giesecke and Brewer, 2018), thus allowing to reduce the number of influential migration drivers during simulation. Nevertheless, some uncertainties may still remain in the interpretation of paleo-records. For example, the use of fossil pollen to provide records of post-glacial tree expansion presents temporal, spatial and taxonomic limitations as listed by MacDonald

(1993). 1) Radiocarbon dating of fossil pollen generally provides a coarse temporal resolution, ranging from several decades to more than a century. 2) Similarly, the source areas of parent trees may lay within tens to hundreds of kilometers from the

165   pollen deposits due to long-distance pollen dispersal, and 3) pollen analysis conducted via light microscopy can identify tree taxa mostly at the level of genus. 4) Additionally, percentage data of pollen records cannot provide a direct representation of tree abundance (though pollen accumulation rates may be used as a proxy), thus complicating the estimation of population growth rates. This is especially relevant in defining when a taxon firstly arrives near a deposit as a sharp increase in pollen representation can be associated with either tree establishment, or an exponential population growth of already established

170   trees (MacDonald, 1993; Giesecke and Brewer, 2018). Finally, the spatial uncertainty of glacial refugia and their relative contribution as source populations for each tree taxa may also complicate the estimation of past migration rates (Tzedakis et al., 2013; Nobis and Normand, 2014). Three main areas have been traditionally defined as sources of post-glacial tree expansion in Europe, i.e. Italy, the Iberian Peninsula and the Balkans (Huntley and Birks, 1983; Bennett et al., 1991). However, previous studies based on plant macrofossils and potential glacial tree distribution (e.g. Stewart and Lister, 2001) have also

175   hypothesized the presence of northern refugia during the LGM (above 45° N), which would yield lower rates of northward tree migration (Feurdean et al., 2013).

Despite the limitations associated with paleo-data, estimates of post-glacial migration rates are overall less likely to add uncertainties during model parametrization than contemporary estimates. Particularly, since current species distribution tend to be limited to one point in time, post-glacial range limits estimated over a continental scale and considering time intervals of

180   hundreds of years are more suitable to describe the spatio-temporal process of migration. Furthermore, taxonomic and spatial uncertainties contained in paleo-pollen can be corrected by the use of plant macrofossils (e.g. fossilized leaves or cones) and phylogeographic studies, since plant macrofossil remains at a site provide unambiguous evidence of the presence of an established individual, which can be identified at the species level (Binney et al., 2009).

Accordingly, we used estimates of post-glacial migration rates for the parametrization of the 17 major European tree species

185   implemented in LPJ-GM (Table 1). Upper and lower boundaries for the value ranges of migration rates were derived from different empirical studies based on the method employed for their estimation. Pollen-based estimates of maximum rates of spread for common European tree taxa were first summarized by Huntley and Birks (1983). Giesecke et al. (2017) revisited these estimates by correcting for the uneven distributions of pollen diagrams using interpolated maps of pollen percentages and threshold values to reduce the risk of false presence. Giesecke and Brewer (2018) further complemented pollen-based

190   estimates from Giesecke et al. (2017) with phylogeographic studies, where the variability and persistence of genetic markers in tree populations through time can be used to trace patterns of post-glacial migration (e.g. Petit et al. 2002). Feurdean et al, (2013) derived post-glacial migration rates from plant macrofossils, which can yield estimates of finer taxonomic and spatial resolutions than pollen-based analysis. Accordingly, we derived upper boundaries of migration rate estimates from maximum values of pollen-based studies, whereas estimates derived from macrofossils, or complemented by threshold values of pollen

195   presence and/or phylogeographic studies were used to define the lower boundaries (see legend of Table 1). This criterion was based on the assumption that pollen analysis, as above-mentioned, may overestimate the true rates of migration (i.e. determined

by first arrival followed by population expansion), especially in the case of tree taxa with long-distance pollen dispersal (LDD). That is, the detection of pollen on the spreading front might indicate an event of first arrival with tree establishment (from which a true rate of migration is calculated) as well as an LDD event with no further establishment and/or population expansion

200 (MacDonald, 1993; Giesecke and Brewer, 2018). In the latter scenario, pollen analysis will result in an apparent rate of migration, which will be an overestimation of the true rate. In this case, genetic studies, estimates from macrofossils and threshold values of pollen percentage can help clarify pollen-based studies and partially compensate for overestimations (Giesecke and Brewer, 2018). Finally, we decided to exclude estimates calculated assuming the presence of northern glacial refugia (Feurdean et al., 2013) based on the unlikely survival of temperate tree taxa north of 45° N during the LGM (Tzedakis

205 et al. 2013).

Table 1. Estimates of maximum post-glacial migration rates in meters per year (m yr⁻¹), and dispersal syndromes during post-glacial expansion for 17 major European tree species. Competitors in model simulations are *Betula pendula* (or *B. pubescens* in the case of the simulated spread of *B. pendula*) and boreal/temperate grasses (see Sect. 2.3). a = Estimates of maximum

210 migration rates by Giesecke and Brewer (2018) with pollen analysis corrected by phylogeographic studies. b = Estimates of maximum migration rates by Huntley and Birks (1983) with pollen analysis. c = Estimates of maximum migration rates by Feurdean et al. (2013) with fossil records, assuming spread from southern refugia (40–45° N latitude). d = Estimates of overall migration rates by Giesecke et al. (2017) derived from the increase in area of presence from interpolated pollen maps and threshold values for pollen presence. e = Dispersal syndromes as reported by Vittoz and Engler (2007) and TRY Database

215 (Kattge et al. 2020): W = wind; Wa = water; B = bird; LA = large mammal (deer, badger, cattle); SA = small animal (e.g. hoarding by rodents). For a comprehensive summary of estimated migration rates from the literature, see Table 5 of Birks (2019).

Species (Figures' notation)	Migration rates	Dispersal syndrome[e]
Abies alba (Abi_alb)	250[a] – 300[b]	W
Betula pendula (Bet_pen)	540[c] – 800[a]	W, Wa, B
Betula pubescens (Bet_pub)	540[c] – 800[a]	W, Wa, B
Carpinus betulus (Car_bet)	500[a] – 1000[b]	W, Wa, B, SA
Corylus avellana (Cor_ave)	1000[a] – 1500[b]	B, SA
Fagus sylvatica (Fag_syl)	200[b] – 300[b]	SA, LA
Fraxinus excelsior (Fra_exc)	200[b] – 500[b]	W, Wa, B, LA
Picea abies (Pic_abi)	150[d] – 500[b]	W, Wa, B, LA
Picea sitchensis (Pic_sit)	150[d] – 500[b]	W
Pinus sylvestris (Pin_syl)	600[a] – 1500[b]	W, Wa
Pinus halepensis (Pin_syl)	600[a] – 1500[b]	W

Quercus coccifera (Que_coc)	300[d] – 500[b]	SA
Quercus ilex (Que_ile)	300[d] – 500[b]	SA
Querucs pubescens (Que_pub)	300[d] – 500[b]	SA
Quercus robur (Que_rob)	300[d] – 500[b]	SA
Tilia cordata (Til_cor)	150[b] – 500[b]	SA
Ulmus glabra (Ulm_gla)	550[d] – 1000[b]	W, Wa

2.2 LPJ-GM 1.0: simulating migration in a dynamic vegetation model

220 LPJ-GM 1.0 (Lehsten et al., 2019) couples a dynamic migration module to the widely used DGVM, LPJ-GUESS (where LPJ-GM is short for LPJ-GUESS-MIGRATION). LPJ-GUESS employs a gap model approach for the simulation of ecophysiological processes and the structural dynamics of forests, including species composition and vertical and horizontal heterogeneity (Smith et al., 2001). Attributes of life-history strategy, phenology, physiology and bioclimatic limits are assigned to plant functional types (PFTs), which correspond to broad physiologically and/or biogeographically distinct groups of taxa

225 (e.g. needle-leaved summergreen), or individual plant species. The spatial domain of simulation is divided into grid cells (usually 0.5º x 0.5º longitude-latitude), each defined by different climatic conditions and soil properties. In LPJ-GUESS, vegetation dynamics are represented by a certain number of replicate patches per grid cell, each sharing the same climate while potentially differing in successional phases due to stochastic processes (e.g. timing of disturbances and mortality). Differently to LPJ-GUESS, LPJ-GM reduces the number of replicate units to one while using multiple explicitly placed patches per grid

230 cell (1 km^2 each) in order to give a spatially explicit representation of the migration processes. This is achieved by simulating seed exchanges among grid cells at the beginning of each simulation year. Contrary to LPJ-GUESS where all species are able to establish without seed limitation and no seed dispersal is explicitly simulated, LPJ-GM allows species to disperse simultaneously while interacting with each other and defines establishment as a function of the amount of seeds available at the patch-level given suitable environmental conditions.

235 Similarly to the TreeMig model implementation (Lischke et al., 2006), LPJ-GM simulates migration at a yearly time-step through four main processes: 1) seed production, 2) seed dispersal via a dispersal kernel, 3) seed bank dynamics, and 4) seedling establishment (Appendix A and Fig. A1). The migration rate depends on distance seeds travel, influenced by the parameters of the dispersal kernel, and on the number of seeds transported, surviving in the seed bank, germinating and finally growing up to a threshold biomass. Thus, we expect that the migration process as described by the LPJ-GM functions

240 (Appendix A) depends positively on the migration parameters: maximum fecundity (i.e. number of seeds produced per tree per year (FEC$_{max}$), seed germination rate (GERM$_p$), and the average short- and long- dispersal distances of seeds (SDD$_d$ and LDD$_d$, respectively; see Table 2 for more details on the migration parameters).

For more details on the model description and technical implementation, see Appendix A and Lehsten et al., 2019.

2.3 Simulation protocol

In order to calculate the migration rates for the 17 major tree species implemented in LPJ-GM 1.0, we simulated the spread of each species in a terrain already occupied by the early successional *Betula pendula* (or *B. pubescens* in the case of the simulated spread of *B. pendula*) and C3 grasses. This choice was based on evidences from pollen records (Birks and Birks, 2008) and phylogeographic studies (Palmé et al., 2003) suggesting a scenario of early colonization of treeless ground by *Betula* species within a very short time period (only hundreds of years) after the retreat of the ice sheet from northern Europe, followed by successive waves of colonization by later-successional tree species (Giesecke and Brewer, 2018). Simulations were performed for a total of 500 years and a spin-up time of 50 years, covering an area of 201 x 201 cells (each covering 1 km^2) with corridors located on the perimeter and the two major diagonals of the domain for a total of 1,197 simulated cells (see Fig. S1). After the spin-up phase, migrating tree species were allowed to establish freely in the upper-left corner of the simulated landscape (the starting point of migration). We applied a static suitable climate for all species (i.e. allowing each species to grow) and an entirely permeable terrain to all grid cells and across all simulation years in order to reproduce optimal environmental conditions for tree migration. This was done to ensure a comparison between simulated migration speeds and post-glacial observations (Table 1) since the latter are calculated as maximum rates, thus likely achieved under favorable climate and without geographical barriers (see Sect. 2.1). Finally, we used the Fast Fourier transform method (FFTM; see Appendix A5) to enhance the computational efficiency of seed dispersal and test for different species-specific seed dispersal kernels.

Since the underlying framework, LPJ-GUESS, has been already extensively validated for different metrics of vegetation composition and structure (e.g. Morales et al. 2007; Pappas et al. 2013), we focused our parametrization effort on the newly-added migration parameters – FEC$_{max}$, GERM$_p$, SDD$_d$ and LDD$_d$ (Table 2) – and evaluate the effect of alternative seed dispersal kernels (Table 3) on migration rates and the model uncertainty.

2.4 Parameter values and model assumptions

Species-specific estimates of the range of the four migration parameters were compiled by reviewing research articles and public databases as indicated in Table 2. We identified the species-specific minimum and maximum values in the literature for each parameter x_i as the range extremes to obtain the parameter uncertainty ($x_{i,max} - x_{i,min}$), and calculated the mean along with the 25th and 75th percentiles, assuming a normal distribution (see Table B1 for species-specific values and data sources). Where data was lacking to identify a minimum and maximum value for a species-specific parameter, we assumed 12.5 % variation from the default value in order to evaluate the model sensitivity to the specific parameter as suggested by Downing et al. (1985) (see Sect. 2.5.1 and Table B1).

For the uncertainty assessment of LPJ-GM 1.0, we assumed static parameters (e.g. species-specific germination rates do not change over time), a uniform seed dispersal in all directions (isotropy), and a proportion of LDD events (LDD$_p$) of 0.01, following the proportional definition of LDD as the 1 % of seed dispersal exceeding a certain quantile of dispersal distance (Schurr et al., 2009). Since we assumed maximum values of paleo-records obtained by classic pollen estimates to be generally

over-estimations of true migration rates (see Sect. 2.1), we used the 75th percentiles of the observed range of migration values as the target observational value to improve and assess the performance of the model across species (see also Sect. 2.6.2). Species-specific default parameters correspond to the values reported by Lischke and Löffler (2006).

Table 2. Description of the migration parameters and data source: a = Lischke and Löffler (2006); b = TRY database (Kattge et al., 2020); c = Royal Botanic Gardens Kew Seed Information Database (SID); d = Vittoz and Engler (2007); e = Tamme et al. (2014). See Table B1 for species-specific range values (default, minimum, mean, maximum and percentiles) and data sources, and Table S1 for species-specific parameter uncertainties ($x_{i,max} - x_{i,min}$; UI).

Parameter	Notation	Unit	References
Maximum fecundity per tree and year	FEC_{max}	no. seeds (in 100)	a, b, c
Seed germination rate	$GERM_p$	%	a, b, c
Average short dispersal distance	SDD_d	Meters	a, b, d
Average long dispersal distance (1 %)	LDD_d	meters	a, b, d, e

2.5 Evaluation of parameter uncertainty

2.5.1 Local sensitivity analysis (LSA)

As a first step, we applied a species-specific local sensitivity analysis (LSA). The LSA method quantifies the relative contribution of each parameter to the model output (first-order effect) by determining the effect of its variation on the output variability while keeping the other parameters fixed to a nominal value (Hamby, 1995). We followed the approach by Downing et al. (1985) and applied a 5-points LSA, where one parameter is adjusted to its minimum, mean, maximum, 25th and 75th percentile values, while the others are kept at their default values (Table B1). We quantified the response of the model to each parameter in terms of directionality, linearity and magnitude by four summary statistics: the Sensitivity Index (SI), two Importance Indexes (II1 and II2), and the Linearity Index (LI) (Downing et al., 1985; Hamby, 1995).

Assuming that y is the migration rate obtained by an LPJ-GM simulation and $x = \{x_1, ..., x_n\}$ is the vector representing the n migration parameters of the model (with $n = 4$; see Table 2), the first-order sensitivity SI_i of parameter x_i is the ratio of the change of the simulated output Δy to the corresponding change in the input parameter Δx_i:

$$SI_i = \frac{\Delta y}{\Delta x_i}, \tag{1}$$

where Δy is the average of the differences of the output values for the 5-points (Downing et al., 1985), and Δx_i is the corresponding 25 % value change for each input parameter x_i. A model output that shows a large change with respect to a small parameter change is considered sensitive to that parameter, where the parameter is said to be influential. While the

absolute value of SI is used to rank parameters based on the sensitivity of the model, the direction of SI (positive or negative) can help to determine whether the output is an increasing or decreasing function with regard to the individual input parameter. The first Importance Index $II1_i$ is used to assess whether a large increase in the parameter uncertainty may propagate into a large uncertainty in the model output, which may be sensitive to the parameter. Thus, $II1_i$ of parameter x_i is the product

305   between its uncertainty Ux_i and its effect on the model output SI_i:

$$II1_i = Ux_i \times SI_i \, ,\qquad\qquad(2)$$

where Ux_i is calculated as the normalized value range $\frac{x_{i,max}-x_{i,min}}{x_{i,max}}$, where $x_{i,max}$ and $x_{i,min}$ are the maximum and the minimum values found in the literature, respectively, for each parameter x_i.

The second Importance Index $II2_i$ is calculated as the percentage difference of the output y when varying the input parameter

310   x_i from its minimum to its maximum:

$$II2_i = \frac{y_{max}-y_{min}}{y_{max}} \, ,\qquad\qquad(3)$$

This index is useful in the case of a monotonic relationship between input parameters and output, and to account for the whole parameter range when calculating the sensitivity.

The Linearity Index LI_i indicates whether the relationship between each input parameter and the model output approach

315   linearity:

$$II2_i = \frac{y_{max}-y_{min}}{\sqrt{s^2}} \, ,\qquad\qquad(4)$$

where s^2 is the sample variance of the model output over the five points ($\sqrt{s^2}$ corresponds to the standard deviation for each parameter range, ca. $\frac{x_{max}-x_{min}}{2}$) and an exact linear relationship between model output and parameter corresponds to $LI_i = 1$. Additionally, we conducted species-specific linear regression analyses of the type $y \sim x_i$, such that every change of one

320   parameter x_i unit translates to a change of migration rate (y) given by the slope coefficient of the regression (i.e. a slope value of 1 should correspond to $LI_i \approx 1$).

2.5.2 Extreme value analysis (EVA)

Having evaluated the importance of the parameters, their direction and linearity with respect to the model output across species, we quantified the effect of species-specific parameter combinations at their extreme values on simulated migration rates.

325   Namely, we fixed all influential parameters at their minimum (all_MIN) and maximum (all_MAX) values, and calculated the corresponding errors as residuals to quantify the performance of the simulations for each species. Species-specific residuals (res) were calculated across the whole value range of observational values:

$$res = \frac{100}{mig_{obs}} \times (mig_{sim} - mig_{obs}) \, ,\qquad\qquad(5)$$

where mig_{sim} is the simulated migration rate for a species, mig_{obs} indicates all integer values from the lower to the upper

330   boundary of migration rate estimates, and $\overline{mig_{obs}}$ their average (Table 1). Thus, residuals determine whether simulated values are over- or under-estimations with respect to observed values (i.e. positive or negative residuals, respectively, above or below 10 m yr^{-1} from the average observed migration speed, where we assumed good estimates to fall within a threshold range of 10 m yr^{-1}), and whether an error is minimized or not for each species. In other words, if the simulated migration rate is an increasing function of all parameters, outputs generated with all parameters at their minimum (all_MIN) should never exceed

335   the upper boundary of observational values, and vice-versa in the case where all influential parameters are fixed to their maximum (all_MAX). Using this insight, we tried to identify a set of parameter values to minimize residuals when possible for each species. Additionally, we conducted a Mann-Whitney U test on the residuals obtained with all_MIN and all_MAX in order to assess whether there were significant differences in model performance between the two parameter settings at the species level (via p-values adjusted with Bonferroni correction for multiple comparisons).

340   This approach allows the exploration of the performance landscape at its extremes, thus allowing to shrink, to some extent, the parameter space, while reducing considerably the number of simulations (i.e. computational demand) as EVA is independent from the number of parameters to optimize.

2.6 Evaluation of model uncertainty

2.6.1 Implementation of fat-tailed seed dispersal kernels

345   Following insights from our previous results (see Sect. 3.2) and recommendation from the literature (e.g. Clark, 1998; for more details see Sect. 4), we decided to implement five additional dispersal kernels to better represent long-distance dispersal (LDD) events in the migration model (Table 3). We selected dispersal kernel functions without exponentially bounded tails (i.e. fat tails), using the exhaustive list of probability density functions provided by Nathan et al. (2012) and Bullock et al. (2017). Fat-tailed kernels are characterized by two parameters, i.e. the scale parameter a, which depends on the average dispersal distance,

350   and the shape parameter b, which defines the weight of the tail. Our objective was to find the species-specific shape parameter value and kernel function that provided a better representation of the migration process (i.e. migration rate) with respect to the default negative exponential function, while keeping the remaining migration parameters within realistic values. We therefore implemented the five additional fat-tailed kernels into the dispersal sub-model of LPJ-GM and ran simulations with default parameter values for the two linearly combined pdfs as calculated from mean dispersal distances (SDD$_d$) and maximum

355   distances (LDD$_d$) (see Appendix A2 for the kernel equation and Table B1 for species-specific dispersal parameter values). We varied the shape parameter b in a suitable range for each kernel, so that b values would define a mathematically significant and fat-tailed pdf (Table 3).

We evaluated the performance of the new kernels by calculating the error between simulated and observed migration rates for each species (residuals; see Eq. (5)) and across species (RMSE; see Eq. (7)). Additionally, we conducted a one-way ANOVA

360   with post-hoc Tukey's HSD test among errors generated by the newly-added kernels to verify whether one or more fat-tailed

kernels improved the predictions across all species (RMSEs) or at the species level (residuals). Finally, we selected the kernels with significantly minimized errors for each species.

Table 3. Probability density functions (pdf) for the default dispersal kernel in LPJ-GM (negative exponential) and five additional kernels. d = mean distance (in meters); a = scale parameter as a function of distance; b = shape parameter range to search for better representation of LDD events. Range boundaries are defined by the values for which pdf are mathematically significant and the corresponding tail is fat (Nathan et al., 2012). Table adapted from Nathan et al. (2012) and Bullock et al. (2017). Gamma function: $\mathbb{\Gamma}(n) = (n-1)!$, where n is any positive integer.

Kernel family	Probability density function	Scale parameter (a)	Shape parameter (b)	Weight of the tail
Negative exponential (NegExp)	$\dfrac{1}{2\pi a^2} \times exp\left(-\dfrac{d}{a}\right)$	$\dfrac{d}{2}$	–	Exponentially bounded
Exponential power (ExpPow)	$\dfrac{b}{2\pi a^2 \mathbb{\Gamma}\left(\frac{2}{b}\right)} exp\left(-\dfrac{d^b}{a^b}\right)$	$\dfrac{\mathbb{\Gamma}\left(\frac{2}{b}\right)}{\mathbb{\Gamma}\left(\frac{3}{b}\right)} \times d$	$0-1$	Fat-tailed (for b < 1) non-power law
Weibull	$\dfrac{b}{2\pi a^2} d^{b-2} exp\left(-\dfrac{d^b}{a^b}\right)$	$\dfrac{b}{\mathbb{\Gamma}\left(\frac{1}{b}\right)} \times d$	$0-2.5$	Fat-tailed non-power law
Bivariate Student's t (twoDt)	$\dfrac{b-1}{\pi a^2}\left(1+\dfrac{d^2}{a^2}\right)^{-b}$	$\dfrac{2}{\pi} \times \dfrac{\mathbb{\Gamma}(b-1)}{\mathbb{\Gamma}\left(b-\frac{3}{2}\right)} \times d$	$1-5$	Fat-tailed power law
Logistic	$\dfrac{b}{2\pi a^2 \mathbb{\Gamma}\left(\frac{2}{b}\right) \mathbb{\Gamma}\left(1-\frac{2}{b}\right)}\left(1+\dfrac{d^b}{a^b}\right)^{-1}$	$\dfrac{\mathbb{\Gamma}\left(\frac{2}{b}\right) \mathbb{\Gamma}\left(1-\frac{2}{b}\right)}{\mathbb{\Gamma}\left(\frac{3}{b}\right) \mathbb{\Gamma}\left(1-\frac{3}{b}\right)} \times d$	$2-5$	Fat-tailed power law
Log-hyperbolic secant (LogSec)	$\dfrac{1/(\pi^2 b d^2)}{\left(\frac{d}{a}\right)^{\frac{1}{b}} + \left(\frac{d}{a}\right)^{-\frac{1}{b}}}$	$\sim d$	$0-1$	Fat-tailed power law

2.6.2 Uncertainty analysis

Model uncertainty is assessed with respect to model complexity, error and sensitivity. Generally, more complex models tend to generate more accurate predictions (i.e. to minimize the error relative to observational data) by incorporating more components (e.g. parameters) into the modelling framework. However, given that each additional component may introduce sensitivity, more complex models are likely to be more sensitive too. Ideally, we would like to select a model structure where

375   both error and sensitivity are as low as possible, i.e. where model uncertainty is minimized (Snowling and Kramer, 2001).

Thus, model uncertainty can be quantify by a summary index of error and sensitivity:

$$U_i = \sqrt{2} - \sqrt{\left(\frac{S}{S_{max}}\right)^2 + \left(\frac{E}{E_{max}}\right)^2} \,, \tag{6}$$

where U_i is the utility of model i (between 0 and 1, where 1 is maximum utility), S_{max} and E_{max} are the maximum sensitivity

and error across models, where the error E is the root mean square error (RMSE) between the simulated migration rates

380   (mig_{sim}) and the observed values (mig_{obs}; species-specific 75th percentile of the value ranges in Table 1):

$$RMSE = \frac{100}{\overline{mig_{obs}}} \times \sqrt{\frac{1}{n}\sum_{m=1}^{n}(mig_{i,sim} - mig_{i,obs})^2} \,, \tag{7}$$

where n is the number of species, and $\overline{mig_{obs}}$ is the average of observations across species. We used the species-specific 75th

percentile of the observed values since upper boundaries (i.e. maximum values) are assumed to be potential over-estimations

of post-glacial migration speeds (see Sect. 2.1).

385   The sensitivity S was calculated according to Eq. (3) as the mean across species and parameters. Thus, we conducted a further

sensitivity analysis of minimum and maximum values for the model structure with newly-implemented and best-fitted fat-

tailed kernels.

Additionally, we calculated an index of model complexity based on the number of parameters and processes implemented in

each dispersal model structure:

390   $$C_i = \sum_{j=1}^{N}\sum_{l=1}^{n_j} p_i r_i \,, \tag{8}$$

where C_i is the complexity of model i, N is the number of state variable, n_j is the number of processes implemented for each

state variable j, p_l is the number of parameters of each process l, and r_l is the number of equations used to formulate each

process (see Fig. A1 for a graphical representation of the components of the migration module).

We defined two different model structures relative to the phenomenological formulation of seed dispersal, which is simulated

395   either as 1) a two-part negative exponential for all species ("Model 1: default kernel"), 2) a two-part best-fitted fat-tailed kernel

per species ("Model 2: fat-tailed kernels"). Thus, we run simulations for all species with the best set of migration parameters

found by EVA for Model 1, or according to the performance analysis of the newly-implemented fat-tailed kernels for Model

2. The evaluation of both indexes (U_i and C_i) should inform the choice of a model structure, which would ideally maximize its

utility while not being overly complex.

3 Results

3.1 Sensitivity analysis

According to the four summary statistics of the local sensitivity analysis (LSA), we identified the two most influential parameters as the mean (SDD_d) and maximum (LDD_d) dispersal distances for local and long-distance seed dispersal (LDD), respectively (Fig. 1 and Fig. S2). On the other hand, maximum seed fecundity (FEC_{max}) and especially seed germination rate ($GERM_p$) showed a smaller effect on the predictions of migration rate both across species (Fig. 1) and at the species level (Fig. S2 and Table S1). Overall, parameter ranking agreed among species, with SDD_d and LDD_d ranked first for 12 species, over 17 total species (see Supplement).

All parameters related to the dispersal kernel (SDD_d and LDD_d) showed an overall consistency across species regarding the positive sign of their relationship with migration rate, though of different magnitude relative to the species (Fig. 1, Fig. S2 and Table S1). Magnitude-wise and on average, an increase of 1 m of mean dispersal distance for SDD_d or maximum dispersal distance for LDD events lead to a corresponding increase of 0.58 or 0.30 m yr^{-1} of migration rate, respectively. On the other hand, maximum fecundity and germination rate had occasionally null or negative effects on the model output (Fig. S2), though the relationship was positive for most species, with an overall increase of 0.20 and 0.09 m yr^{-1} for each unit increase of FEC_{max} (100 seeds) and $GERM_p$ (1 %), respectively. These results were confirmed by the species-specific linearity index (LI) and slope coefficients, where values above or close to the unit indicated a strong and positive linear relationship (and thus proportional increase) of SDD_d and to a lesser extent of LDD_d with respect to the migration rate, whereas the sub-unit and mostly non-significant values of FEC_{max} and especially $GERM_p$ suggested little linear relationship with the model output (see Fig. 1 and Table S1 for LI values, and Fig. S2 and S3 for species-specific slope coefficients and species-specific shapes of sensitivity functions, respectively).

Uncertainty in parameter estimates from literature sources seemed to be highly species-specific (Table S1) though, overall, FEC_{max} and $GERM_p$ showed the highest and the lowest uncertainties, respectively. The effect of parameter uncertainty on simulated migration rates (II1) was overall greater for SDD_d due to the large sensitivity of the model output to this parameter (Fig. 1) though the uncertainty of SDD_d is moderate with respect to others (52.70 ± 41.06; see Supplement). On the other hand, the relatively high effect of FEC_{max} on migration rate (II2) was mainly due to its high uncertainty (3,470.82 ± 9,339.38; see Supplement), especially in the case of *Betula* spp. (28,275), *Carpinus betulus* (682) and *Ulmus glabra* (894) (Table S1). Similarly, the higher importance of LDD_d when accounting for the whole parameter range (II2) seems likely due to the large uncertainty associated with estimates of maximum dispersal distances (Table S1).

In summary, parameters linked to the seed dispersal kernel appear to be the most influential, while all parameters are overall positively related to migration rate over the range of each individual parameter for most species. Insights gained from the LSA and from the equations of the migration module (see Sect. 2.2 and 4, and Appendix A) indicate that the migration function is monotonic also over the entire parameter space, at least in the neighborhood of the default values. Thus, we expect that by

increasing the values of all parameters within acceptable ranges for each species, simulated migration rates will correspondingly increase to their potential maxima.

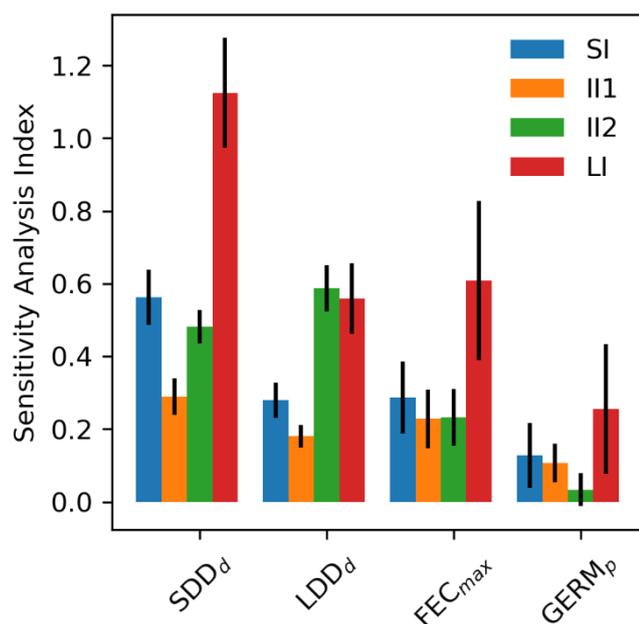
We applied this insight to the Extreme Value Analysis (EVA).

435

Figure 1: Mean sensitivity analysis (SA) indices (±2 standard errors) across 17 tree species for the four parameters of the newly-implemented migration module of LPJ-GM: average short dispersal distance (SDD$_d$, meters), 1 % average long dispersal distance (LDD$_d$, meters), maximum fecundity per tree (FEC$_{max}$, no. seeds in 100 per year), and seed germination rate (GERM$_p$, %). See Sect. 2.5.1 for the calculation of SA indices: Sensitivity Index (SI), Importance Indexes (II1 and II2), and Linearity Index (LI). See Table S1 for species-specific values.

440

3.2 Extreme value analysis

The probability density of residuals between simulated migration rates and observed values across their range, where migration parameters were varied simultaneously to their minimum (all_MIN) or maximum values (all_MAX) per species, is shown in Fig. 2. The range of the residual distribution quantifies the uncertainty of the observational data, where *Abies alba* and *Pinus*

445 spp. have the smallest (50 m yr^{-1}) and the largest (900 m yr^{-1}) data uncertainties, respectively. Negative residuals indicate that model outputs underestimate historical observations across all species, with the exception of the over-estimations obtained with the all_MAX setting for *Abies alba* and *Picea abies* (above 10 m yr^{-1} from the average observed migration speed; Fig. 2 and Table S2). Consequently, we looked for smaller parameter values (all_MAX_opt) to minimize the error for both species (RMSE$_{all\_MAX\_opt}$ = 2.66 % vs. RMSEall_MAX = 32.99 % for *Abies alba*, and RMSE$_{all\_MAX\_opt}$ = 9.21 % vs. RMSE$_{all\_MAX}$ =

450 47.45 % for *Picea abies*; see Supplement).

Overall, simulated migration rates showed high errors with respect to the 75th percentile of observational data for both extreme value settings, though migration rates generated by all_MAX were significantly less biased than estimates generated by

all_MIN (Mann-Whitney U test and Bonferroni adjusted p-value < 1e-3), both across species (RMSE$_{all\_MIN}$ = 125.2 %;

RMSE$_{all\_MAX}$ = 111.01 %) and at the species level, with the exception of *Pinus halepensis* (p-value = 1.0) and *Quercus*

455 *pubescens* (p-value = 0.8281; see Supplement for all species-specific p-values).

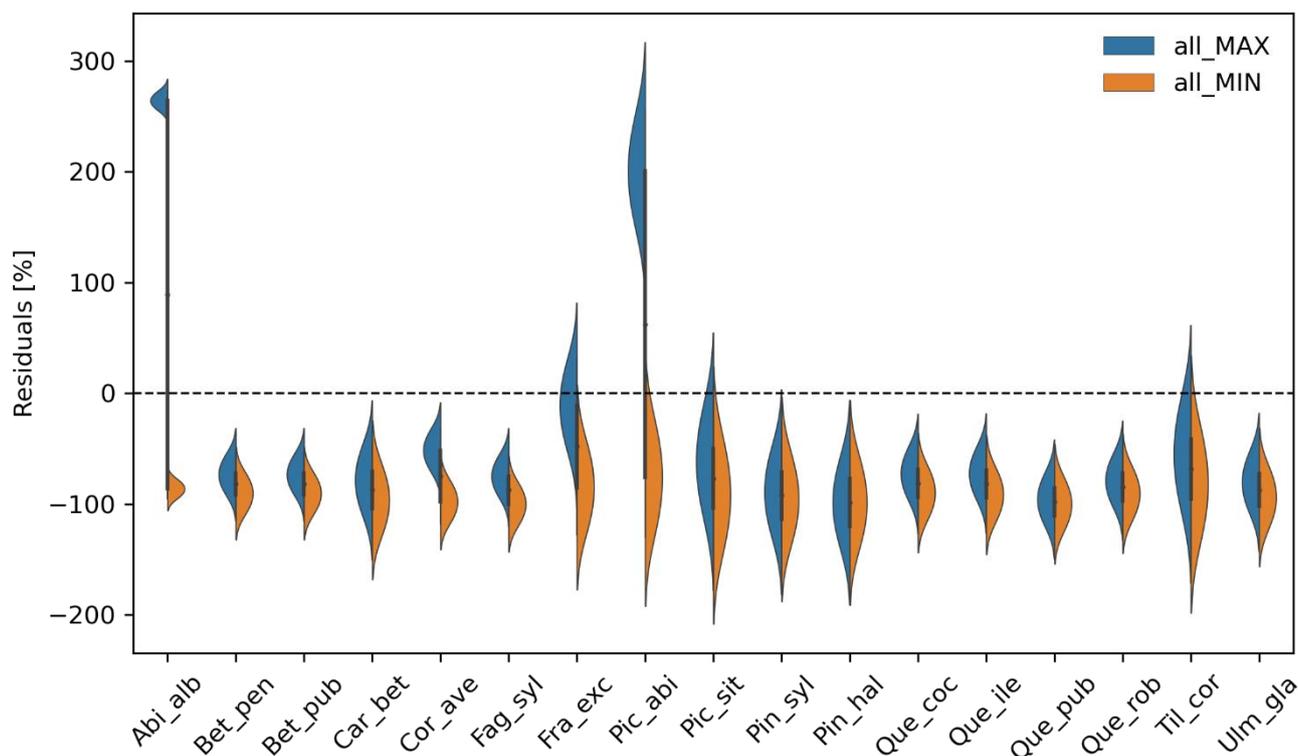



Figure 2: Species-level residuals of observed versus predicted migration rates for extreme parameter settings. The observed migration rates refer to the range value of historical estimates (Table 1), i.e. each violin represents the residual distribution, where the upper point of the distribution indicates the difference between the simulated migration rate and the minimum observed migration rate, whereas the lower
460 point corresponds to the difference between the simulated migration rate and the maximum observed migration rate. all_MIN: all migration parameters are fixed to their minimum values; all_MAX: all migration parameters are fixed to their maximum values. See Table B1 for species-specific extreme parameter values, and Table S2 for species-specific residuals.

3.3 Performance of fat-tailed kernels

465 Newly-implemented fat-tailed kernel function with examples of shape parameter b values and scale parameters a based on the

same LDD and mean dispersal distance data are shown in Fig. 3, compared to the default negative exponential. Exploring the

species- and kernel-specific parameter space of the shape parameter while keeping the remaining migration parameters within

realistic values (see Table 2), we found that overall simulated migration rates over-estimated observed values (with respect to

the 75th percentile) when generated by the logistic function in the shape parameter range [3–4], and by the log-hyperbolic

470 secant and twoDt functions around the 0.5 and 2 values, respectively (Fig. 4). On the other hand, the exponential power function

at lower b values [0.15–0.3] and the Weibull function across the whole parameter range selected [1.75–2.5] tended to under-estimate observed values (Fig. 4).

After selecting the best shape parameters per kernel (i.e., yielding the minimum residual; Table S3), we confirmed that the Weibull function under-estimated observed migration speed across all species, whereas the log-hyperbolic secant and logistic

475 functions were overall the best at minimizing residuals at the species level (Fig. 5). Nevertheless, we found significant differences in kernel performance based on the species (Fig. 5, Fig.6, and Table S4). For example, the exponential power and twoDt functions generated good estimates for *Fraxinus excelsior* (RMSE = 25.75 % and 25.83 %, respectively), while the log-hyperbolic secant and logistic functions tended to over-estimate migration rates by ≈30–45 % for both species. Conversely, good estimates of migration rates for *Quercus* spp. were generated by the log-hyperbolic secant and logistic functions (RMSE

480 ≈20 %), while the exponential power and twoDt functions produced under-estimations across the whole range of observed values by ≈25 % up to ≈95 %.

Overall, fat-tailed kernels were able to significantly improve model outputs relative to the default negative exponential both across species and at the species level, with the exception of *Abies alba* and *Picea abies* (see Sect. 3.2). Concerning the worst performances at the species level, the simulated migration rate obtained with the default kernel had higher error relative to the

485 observed 75th percentile (RMSE = 91.5 % for *Pinus halepensis*) compared to the significantly less biased estimate generated by the logistic function for *Pinus sylvestris* (34.32 %). Kernel functions can be ranked based on RMSE % calculated across species as follows: 1) log-hyperbolic secant (27.26 %), 2) logistic (48.07 %), 3) twoDt (63.19 %), 4) exponential power (73.67 %), 5) negative exponential (89.94 %), and 6) Weibull (97.72 %) (see Supplement). Finally, we selected the kernel function that significantly minimized residuals for each species as shown in Fig. 6 (see Table B2 for species-specific values of all

490 optimal parameters).

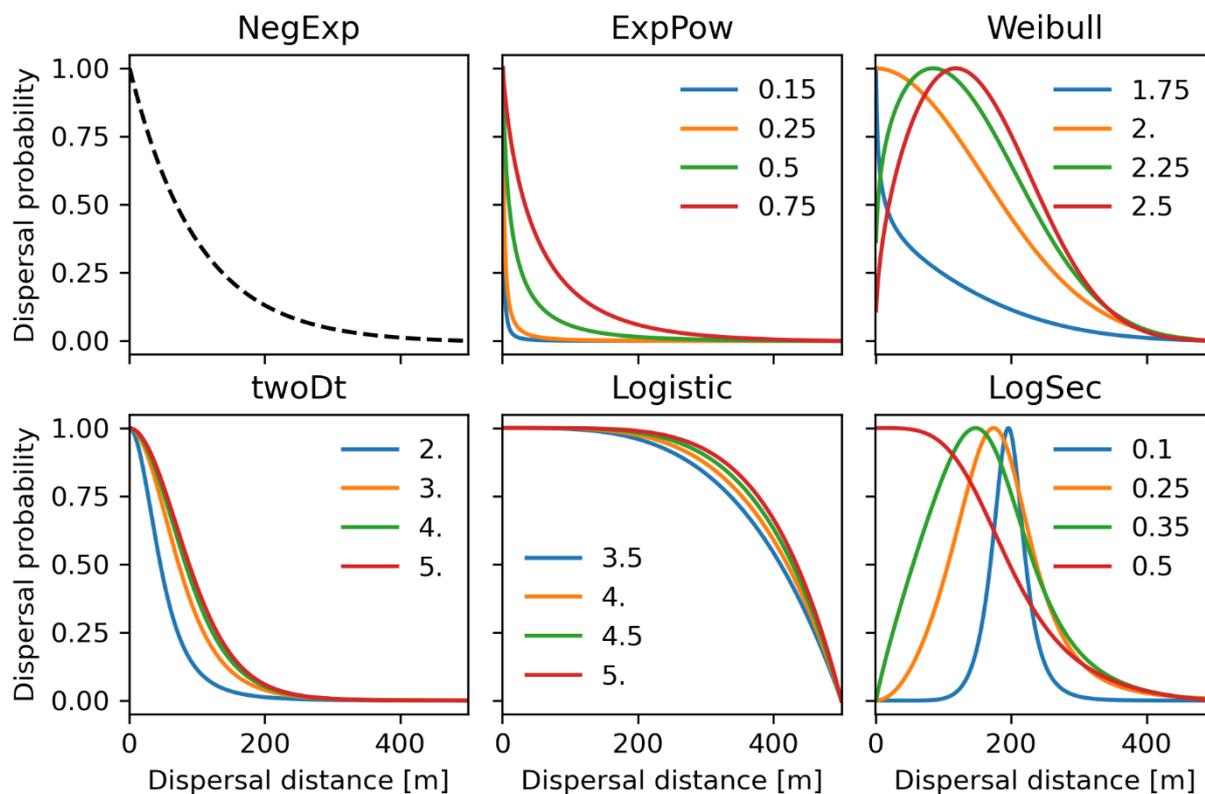



Figure 3: Probability density functions for the newly-implemented fat-tailed seed kernels. Fat-tailed kernels with examples of shape parameter values are compared to the default negative exponential (NegExp). Newly-implemented kernels are: ExpPow = exponential power; Weibull; twoDt = Bivariate Student's *t*; Logistic; LogSec = log-hyperbolic secant. See Table 3 for the kernel formulae and range values of the shape parameter for each kernel. The scale parameters were all calculated from the same average dispersal distance values (1–500 meters).

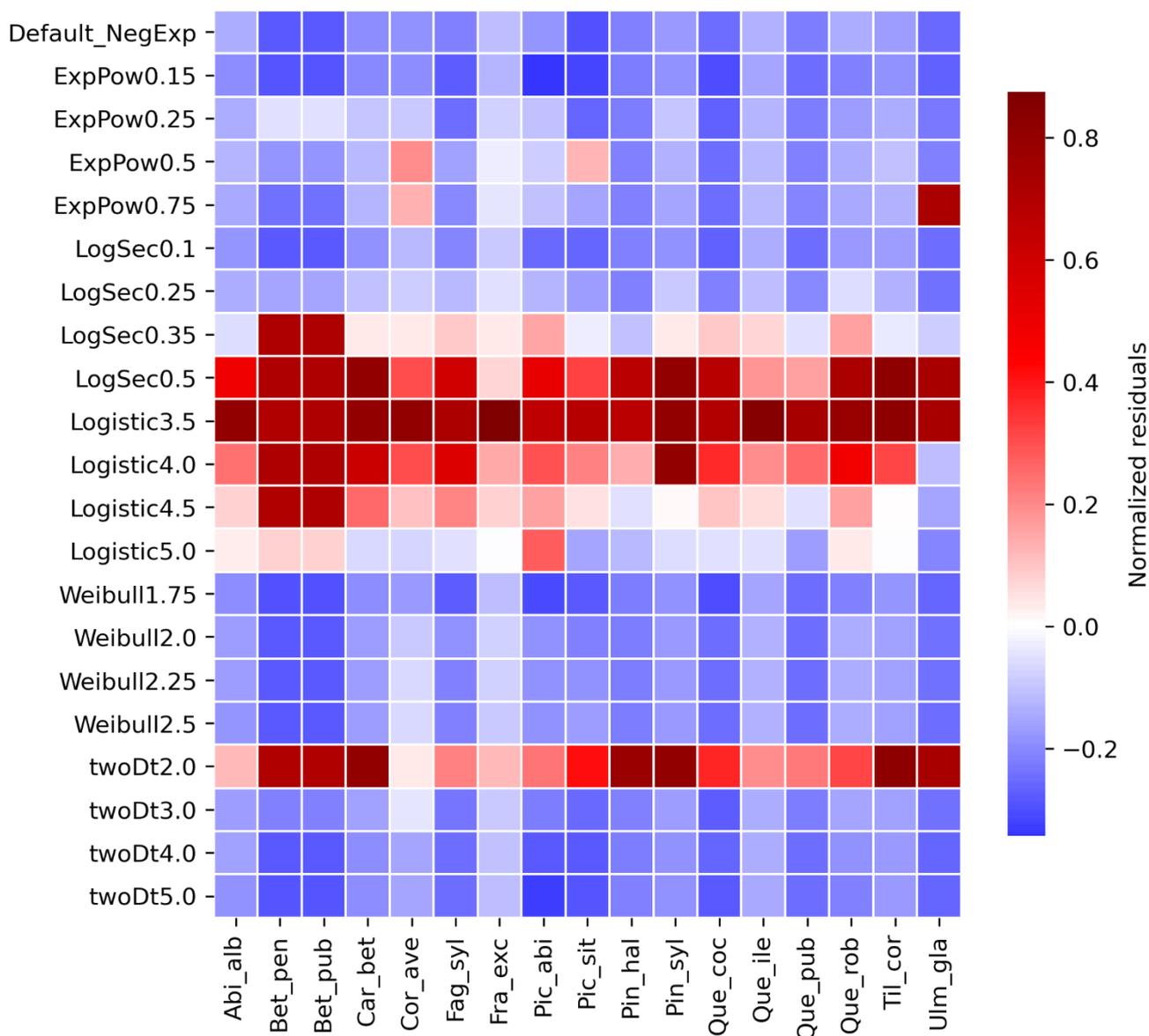



Figure 4: Species-specific performance of seed kernels for different shape parameters. The heatmap shows the species-specific normalized residuals between simulated migration rates and observed 75th percentile historical estimates for the default negative exponential function ("Default_NegExp") and the newly-implemented fat-tailed kernels across examples of shape parameter values ("{kernel name}+{b value}"; cf. Fig. 3). Red and blue indicate over- and under-estimations, respectively. See Table S3 for species- and kernel-specific residuals and simulation values for all tested shape parameter values.

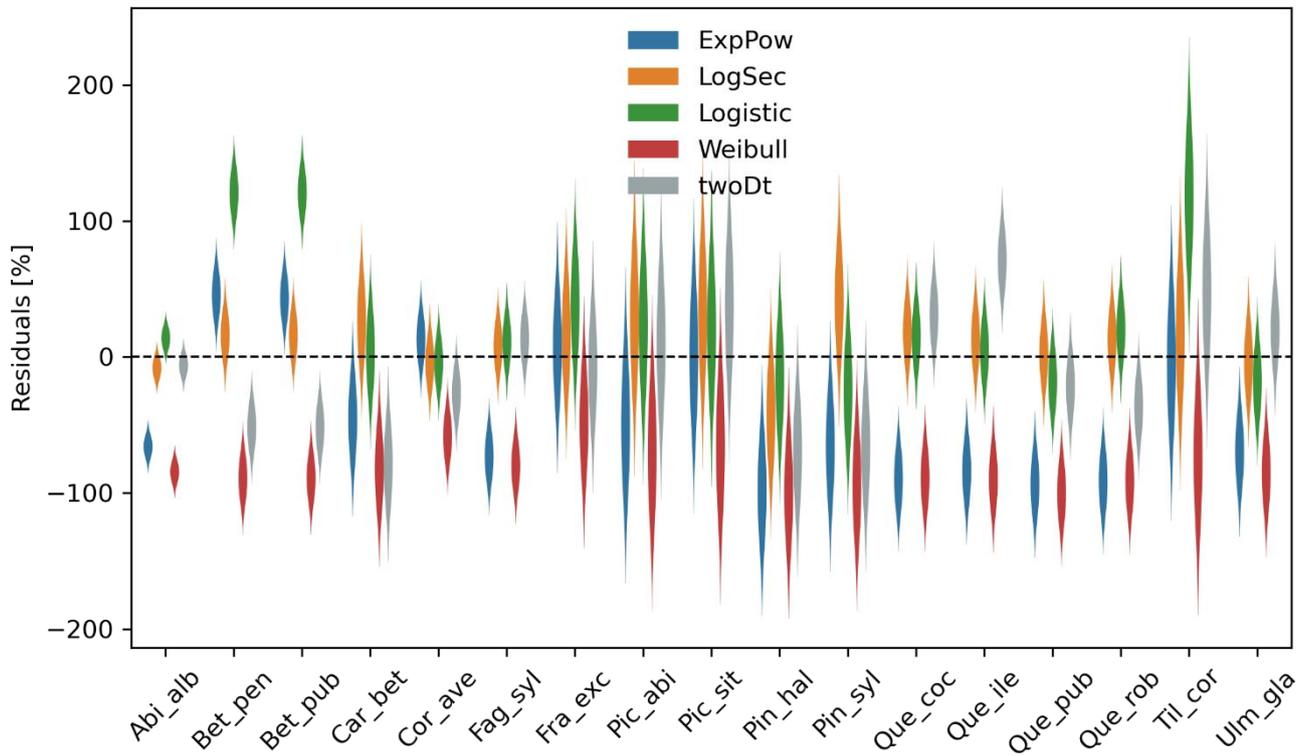



Figure 5: Performance evaluation of fat-tailed kernels. Comparison of species-specific residuals of observed versus simulated migration
505 rates across the range values of historical estimates (Table 1) among newly-implemented kernel functions. See Table S4 for means and
standard deviations of species-specific residuals.

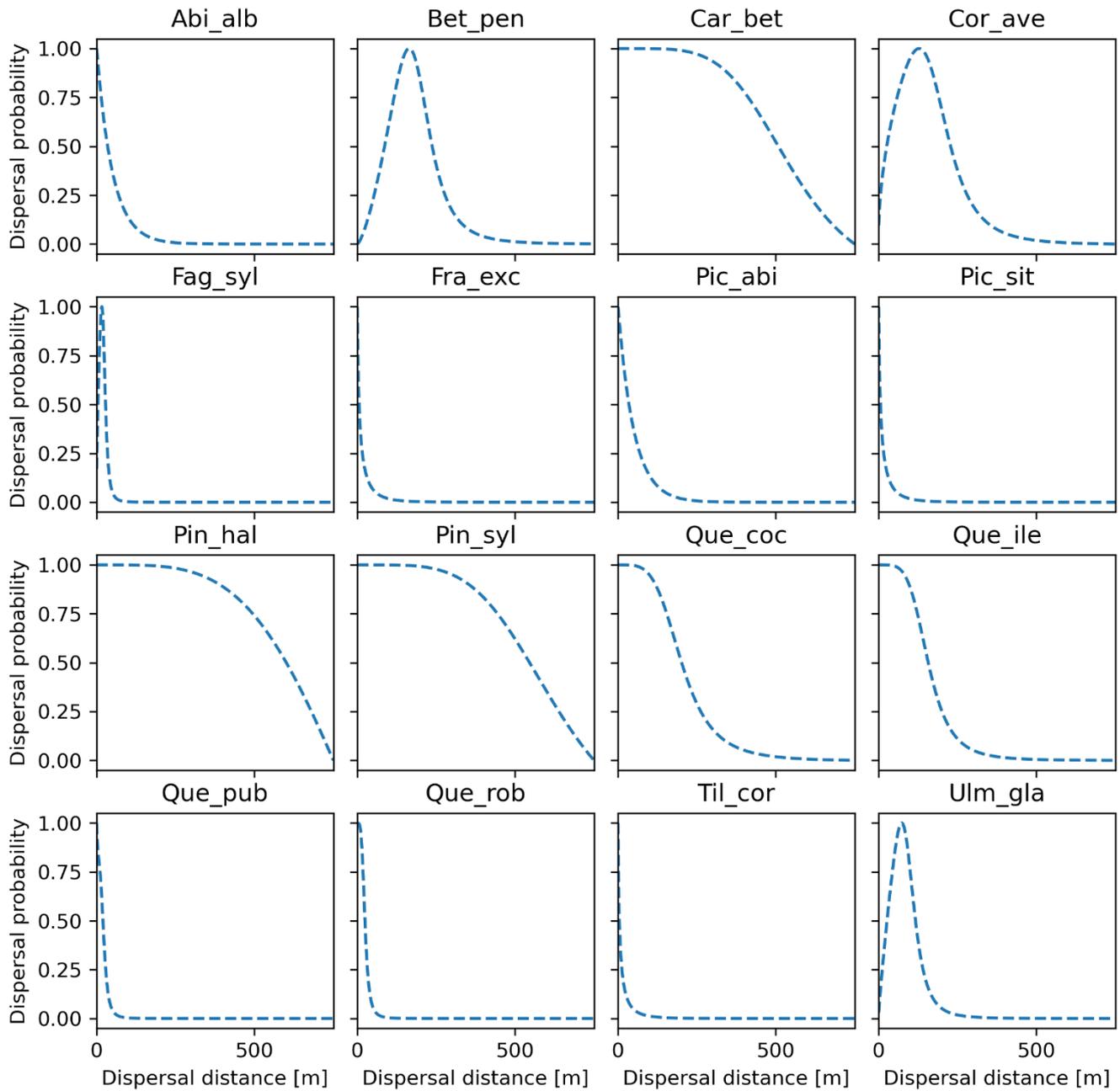



Figure 6: Best fat-tailed kernel shapes per species. See Table B2 for the corresponding parameter values: SDD_d, LDD_d and shape parameter b.

510

3.4 Uncertainty analysis

The results of the analysis of model uncertainty for the two modelling frameworks – seed dispersal simulated either with a negative exponential kernel across species (Model 1: "default kernel") or with species-specific fat-tailed kernels (Model 2: "fat-tailed kernels") – are shown in Fig. 7.

515 Model complexity is calculated from four processes – 1) seed production, 2) seed dispersal, 3) seed bank dynamics, and 4) seedling establishment – represented by one operation each, including 1) FEC_{max}, 2) SDD_d and LDD_d, 3) $GERM_p$ and 4) $GERM_p$ for Model 1 ($C_1 = 1 + 2 + 1 + 1 = 5$), and 1) FECmax, 2) SDD_d, LDD_d and b, 3) $GERM_p$ and 4) $GERM_p$ for Model 2 ($C_2 = 1 + 3 + 1 + 1 = 6$), respectively (see Appendix A and Fig. A1). Thus, the added unit of complexity for Model 2 results from the inclusion of the shape parameter b as weight for the fat-tailed kernels.

520 Confirming previous error analyses at the species level (Sect. 3.3), the use of fat-tailed kernels managed to reduce model error by ≈65 % across species compared to the default setting ($RMSE_1 = 89.94$ % vs. $RMSE_2 = 25.14$ %). As expected from theory (Snowling and Kramer, 2001; see also Sect. 2.6.2), model sensitivity increased along with complexity, with Model 2 being nearly two times more sensitive across all its parameters than Model 1 ($S_1 = 0.27$ vs. $S_2 = 0.42$). This was mainly due to the inclusion of the shape parameter for fat-tail kernels in Model 2 (b), to which the model output is highly sensitive (Fig. 7). On 525 the other hand, the ranking of parameters based on their effect on simulated migration rates is the same for both model structures, with dispersal distances for local (SDD_d) and long-distance dispersal (LDD_d) being the most influential parameters. Comparing parameter effects among the two model structures, simulated migration rates were more sensitive to germination rates ($GERM_p$) in Model 2 relative to Model 1, whereas SDD_d was more influential in Model 1 relative to Model 2 (Fig. 7). Overall, the inclusion of fat-tailed kernels improved model utility to a moderate level, $U_2 = 0.54$ (where U = 1 is the ideal 530 model utility), with respect to the default modelling framework ($U_1 = 0.37$).

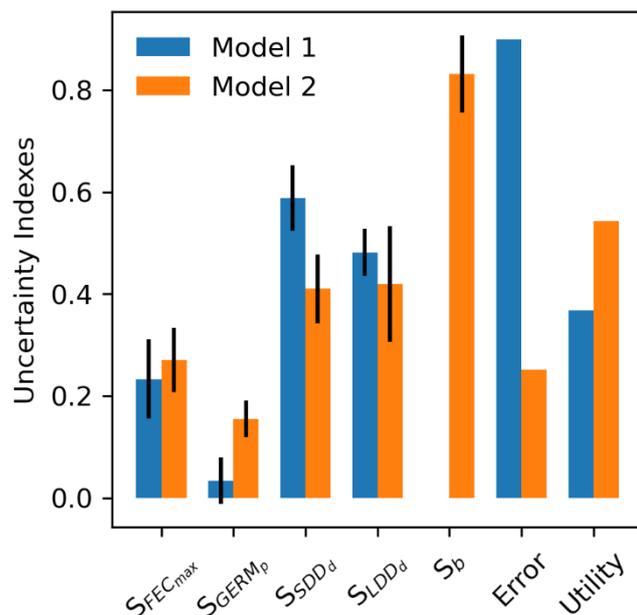



Figure 7: Comparison of parameter-specific sensitivities, model error and utility between two structures of the migration module of LPJ-GM. Model structures refer to seed dispersal simulated either with a linear combination of two negative exponential kernels across species (Model 1: "default kernel") or of species-specific fat-tailed kernels (Model 2: "fat-tailed kernels"). S_{br}: model sensitivity to the shape parameter b of fat-tailed kernels (see Table 2 for the description of other parameters). Note that the kernel formulation of Model 1 has no shape parameter (hence, S_b is NA). See Sect. 2.6.2 for the calculation of model utility.

4 Discussion

The sensitivity analysis (SA) classification showed that migration rates had lower sensitivity to local demographic traits (germination rate and maximum annual fecundity) than to parameters related to seed dispersal (average distance for local dispersal and rare LDD events) for most of the species (Fig.1 and Table S1). Similar results were found by two studies applying SA on mechanistic models for the simulation of seed dispersal (Soons et al., 2004; Nathan and Katul, 2005). In both cases, average dispersal distances or LDD were more influenced by parameters linked to the dispersal kernel (e.g. horizontal wind velocity) than to local demographic or environmental factors. The importance of dispersal parameters was also confirmed by the study of Lustenhouwer et al. (2017) on the relationship between dispersal ability, local demographic traits and migration speed across 80 plant species, including trees. Lustenhouwer et al. (2017) found that the intrinsic migration capacity of tree species (simulated by the Clark et al., 2001 model) was significantly and positively correlated to dispersal ability (i.e. decreasing rate of dispersal as a function of distance from the mother tree). On the other hand, local demographic traits such as fecundity had a weaker correlation to spread velocity (Lustenhouwer et al., 2017). Our metrics of the shape between migration rates and migration parameters (LI in Fig. 1 and slope coefficients in Fig. S2) further highlighted a linear effect of

dispersal distances (especially for local dispersal) on migration rate, whereas maximum fecundity and germination rate seemed to affect migration in a bounded way (see also Fig. S3). Similar shapes were found by Lustenhouwer et al. (2017) when analyzing the relationship of spread velocity with dispersal ability (linear) and fecundity (asymptotic). The shape of our relationships can be explained by the formulation of the migration process in the model LPJ-GM and the calculation of

555   migration rates. Migration speed is defined by the distance of an individual from the migration source (i.e. mother tree) when surpassing a certain biomass threshold (LAI = 0.5) and accounting for the time elapsed since the start of migration (see Sect. 2.2 and Appendix A). In turn, individual biomass at a location depends on the number of seeds produced by the individual each year (i.e. annual fecundity), by the distances at which seeds are dispersed (i.e. seed dispersal kernel described by SDD_d and LDD_d) and by the local biomass growth. On one hand, annual fecundity and biomass growth are determined by

560   environmental conditions, and inter- and intra-specific competition for light and space. Thus, we expect that the positive effect of maximum fecundity on migration rate will form a plateau when species reach their carrying capacity at a site. On the other hand, an increase of average dispersal distance would likely translate into an almost direct (i.e. linear) increase of migration speed, which corresponds to seed movement per time unit.

The overall increase in migration rate values given by simultaneously setting all parameters to their maximum values

565   (all_MAX) supported our assumption that the simulated migration rate is likely an increasing function of the four migration parameters. However, high and significant error values corresponding to all_MAX simulations suggested that the default model structure was unable to generate unbiased estimates within the acceptable range of parameter values for most species, with the only two exceptions being *Abies alba* and *Picea abies* (Fig. 2). This systematic underestimation might indicate that an ecological process significant to tree migration is missing or is incorrectly represented by the default model structure. In

570   this respect, a number of studies on plant migration modelling suggests that the asymptotic behavior of the dispersal function (i.e. the extent of the tail) is crucial for the simulation of population spreading rates, where kernel tails represent long-distance dispersal (LDD) events (Clark, 1998; Caswell et al., 2003). Our default model implements LDD events in the weighted linear combination of two negative exponential functions that represent SDD and LDD events with a 0.99 and 0.01 relative probability, respectively. However, dispersal kernels with high leptokurtosis, long or even fat-tails (i.e. not exponentially

575   bounded) are generally considered more accurate representations of LDD events than negative exponential functions (Clark et al., 1999; Bullock and Clarke, 2000; Nathan et al., 2012), especially in the case of species with a high probability of LDD due to the presence of active dispersal vectors, such as hoarding or migratory animals (Clark, 1998; Clark et al. 1999; Powell et al., 2004). This is in agreement with our species-specific best performance across seed kernels relative to the natural dispersal syndrome of the tree species (Table B2). Specifically, the only two species that could generate unbiased estimates of migration

580   rates with the default negative exponential function (*Abies alba* and *Picea abies*) are primarily dispersed by wind, a passive dispersal vector. On the other hand, fat-tailed kernels provided a better fit for animal- or wind-dispersed species with additional LDD mechanisms, such as water currents and migratory animals (e.g. birds; see Schurr et al., 2009 for a classification of LDD mechanisms). Overall, wind-dispersed species tended to have higher maximum dispersal distances than species primarily dispersed by animals (Fig. S4), in agreement with observational studies on vegetation spread in temperate and tropical forests

585 (Clark et al., 1999). From an ecological point of view, this can be explained by the average traits of wind-dispersed seeds, which are smaller and winged and thus more likely to be transported farther compared to the heavier animal-dispersed seeds. From a modelling point of view, larger LDD values would allow to simulate high migration rates even with an exponentially-bounded kernel function. On the other hand, animal-dispersed species might have shorter dispersal distances on average (Fig. S4) but can potentially reach higher migration rates via rare LDD events producing outlying individuals ahead of the migration

590 front. In this case, the use of fat-tail kernels will produce a noisy and accelerating vegetation spread relative to a step-wise and slower spreading front given by exponentially-bounded kernels (Clark, 1998). This seems to suggest that occasional LDD events are more important for tree migration than local dispersal driven by more common vectors, at least in order to achieve the spreading rates of the paleo-records used in this study. Supporting this, previous studies showed that migration rates generated by fat-tailed kernels are more compatible with historical estimates of post-glacial forest expansion than rates

595 obtained with Gaussian-like kernels or other functions that poorly represent LDD events (Cain et al., 1998; Caswell et al., 2003). The importance of LDD in post-glacial tree expansion has been long recognized as the most likely explanation to the Reid's paradox, i.e. the apparent discrepancy between contemporary plant dispersal potential and observed post-glacial migration rates (Reid, 1899; Cain et al., 1998). In the context of the Reid's paradox, dispersal potential mostly refers to the common (99 % of events) SDD relying on conventional dispersal vectors, whereas high post-glacial migration rates are

600 explained by rare (1 %) LDD events where trees colonized newly-emptied areas via less common vectors (e.g. water or large/migratory animal) (Higgins et al., 2003b; Vittoz and Engler, 2007; see also Sect. 4.4 of Birks, 2019 for a discussion of possible scenarios of post-glacial forest expansion).

The shape of the kernel tail can affect not only the migration speed but also its sensitivity to other migration parameters. In agreement with Clark (1998), the implementation of fat-tailed kernels enhanced the importance of ecological traits linked to

605 migration, especially germination rate, compared to negative exponential kernels (Fig. 7). Additionally, relatively small differences in the tail shape of fat-tailed kernels (i.e. shape parameter) had strong effects on the migration rates (Fig. 4 and Fig. 7). Such a high sensitivity might add to the uncertainty of parameter estimates and thus lower the confidence of model predictions with fat-tailed kernels. However, it might be observed that the tail shape is inherently uncertain, regardless of the kernel function used. That is, we cannot reliably fit the tail to empirical data of LDD and shape parameters as these are nearly

610 impossible to estimate with high accuracy, especially in the case of animal-dispersed species (Clark et al., 2003). For example, compared to the relatively easier recovery of seeds from traps for wind-dispersed species, it would be required to know in advance which animals would act as dispersal vectors and then visually follow or track them until seeds are deposited on the ground (Beckman et al., 2020). Furthermore, it has been observed that a single dispersal probability density function cannot be standardized for all species since kernel shapes depend on species-specific dispersal mechanisms (Bullock et al., 2017). At

615 the same time, a kernel shape cannot be built based on a specific dispersal mechanism as a given species can disperse using a variety of vectors (Nathan et al., 2008; Counsens et al., 2017). Under these considerations, our approach was to implement species-specific dispersal kernels that summarized dispersal modes (see "total dispersal kernel"; Nathan et al., 2008) and

provided a good representation of LDD events that could match the historical spreading rates, rather than find the kernel function that best-fitted experimental data of seed shadows.

620   Overall, our results seem to justify the use of simple functions to summarize multiple dispersal modes and simplify complex dispersal mechanisms, at least in a large-scale context. For example, in the case of continental-scale simulations over thousands of years (e.g. post-glacial forest expansion), it is suggested to avoid a more detailed representation of migration that would require the inclusion of e.g. wind properties, animal movement and behavior, seed retention and deposition. Such local stochastic processes are generally challenging to parametrize in models given the difficulty of observation and their case-

625   specific nature (Nathan et al., 2012). As such, the choice of species-specific phenomenological dispersal functions would allow reducing the inherent uncertainty of migration models compared to more explicitly mechanistic representations of seed dispersal, while still providing a representation of dispersal abilities and population dynamics (via germination rates and maximum fecundity) at the species level.

5 Limitations and future challenges

630   There are a number of reasons for the disagreement between model output and observations beside parameter and model uncertainties.

- Observational range values of migration rates and/or of migration parameters obtained from the literature might be incorrect or too uncertain. For instance, *Pinus* spp. have both the highest error relative to simulated migration rates ($RMSE_{allMAX}$ = 191.5 % for *P. halepensis*) and the largest uncertainty of observed migration rate across all species

635   (from 600 to 1500 m yr^{-1}).

- We might have assumed the wrong competitor (see Sect. 2.3), e.g. some late-successional species (e.g. *Fagus sylvatica*) might have competed with additional species beside the early-successional *Betula* spp. (Giesecke and Brewer, 2018).

- We decided to simulate a homogeneous terrain (for permeability) and climate to ensure ideal climatic/topographic

640   conditions for vegetation spread to match maximum migration speeds. This decision was based on the assumption of migration lag for the paleo-vegetation, i.e. trees were mainly limited by dispersal and not by climate during the post-glacial forest expansion of Europe. There is still controversy regarding this point (Normand et al., 2011; Svenning and Sandel, 2013; Sect. 4.4 of Birks, 2019).

- We assumed static parameters, whereas, for example, germination rate is usually dependent on temperature and light

645   regimes (Baskin and Baskin, 1998). Furthermore, there are some instances of evolutionary responses concerning dispersal ability (e.g. adaptation to the new environment may reduce the need to disperse). However, trees seem to have a slow plastic/evolutionary response with respect to other taxa and are more likely to respond to climate change with range shifts (Berg et al., 2010; Lenoir et al., 2020; see also Sect 6.3 and Table 11 of Birks, 2019 for the little evidence of plasticity/adaptation of terrestrial plants during the Quaternary).

650   • We did not conduct a formal and thorough model calibration by using an optimization algorithm, though we have
            still identified valuable parameter estimates (i.e. we minimized the error by staying within probable parameter
            combinations) and helped to reduce the parameter space (following the concept of Exploratory Landscape Analysis;
            Mersmann et al., 2011; Stork et al., 2020). A thorough optimization could be the next step and might suggest whether
            there is more than one local optimum (global optimization; Stork et al., 2020). However, these approaches are very
655   expensive for complex models such as LPJ-GM. Furthermore, obtaining a better fit to observations does not
            necessarily guarantee that a model is more realistically simulating vegetation dynamics (i.e. right predictions for the
            wrong reasons). Uncertainty assessment is still useful to identify model components to improve (e.g. parameter
            uncertainty to reduce by acquiring more empirical data). More generally, our efficient uncertainty assessment could
            be used to reduce the cost of solving optimization problems for computationally demanding models by allocating
660   computational resources on relevant model components (parameters or process formulations) and gain insights into
            model limitations for further improvement.

        • Predicting future trees' range shifts based on a model parametrized with mid-Holocene estimates might present some
            limitations since past and current global warming have different intensities and species are submitted to different
            conditions (e.g. more fragmented habitats and unlimited human-driven dispersal in present/future conditions with
665   respect to the mid-Holocene; Corlett and Wescott, 2013; see also Sect. 7.2.2 and 7.2.3 of Birks, 2019). Nevertheless,
            the outcome of our study is important not in the sense that post-glacial migration estimates are to be expected in the
            future, but rather that we identified important mechanisms controlling migration rate (LDD events) and implemented
            them into the model structure (fat-tail kernels) (Nogués-Bravo et al., 2018). The high velocity of isothermal shift
            predicted for the 21st century (up to 4–6 km yr^{-1}; Svenning and Sandel, 2013; Lenoir et al., 2020) suggests that trees
670   will be limited by their dispersal ability (i.e. migration lag), and especially by LDD events in a fragmented landscape
            with few available establishment sites. Thus, the model representation of LDD will likely be more important for the
            realistic simulation of future range shifts.

6 Conclusions

Model predictions of species range shifts have many sources of uncertainties, which is important to acknowledge and evaluate
675 as a first step for model improvement. This study aimed to provide an evaluation of the parameter and model uncertainties of
the migration module of a dynamic vegetation model, LPJ-GM 1.0. We used low-cost methodologies for estimating the
sensitivity of one model output, tree migration rate, to key migration parameters, and provided quantitative information about
the importance of mechanisms underlying the migration process across different tree species and a first guess of parameter
values used in the simulations. Overall, the model structure implementing fat-tailed dispersal kernels provided significantly
680 better predictions than the default modelling framework, while not being overly more complex, and can therefore be a good
candidate for an improved model structure (LPJ-GM 1.1). Though a reduced model error does not necessarily mean that

predictions of migration rates would be correct, identifying influential migration mechanisms (LDD events) and their efficient inclusion in the model structure (via fat-tail kernels) can improve our confidence in range shift predictions, especially in a context of global change where LDD will likely be more relevant.

685 **Appendix A. Migration module of the model LPJ-GM 1.0**

For more details on the model description and technical implementation, see Lehsten et al., 2019. See also Fig. A1 for a graphical summary of the migration module.

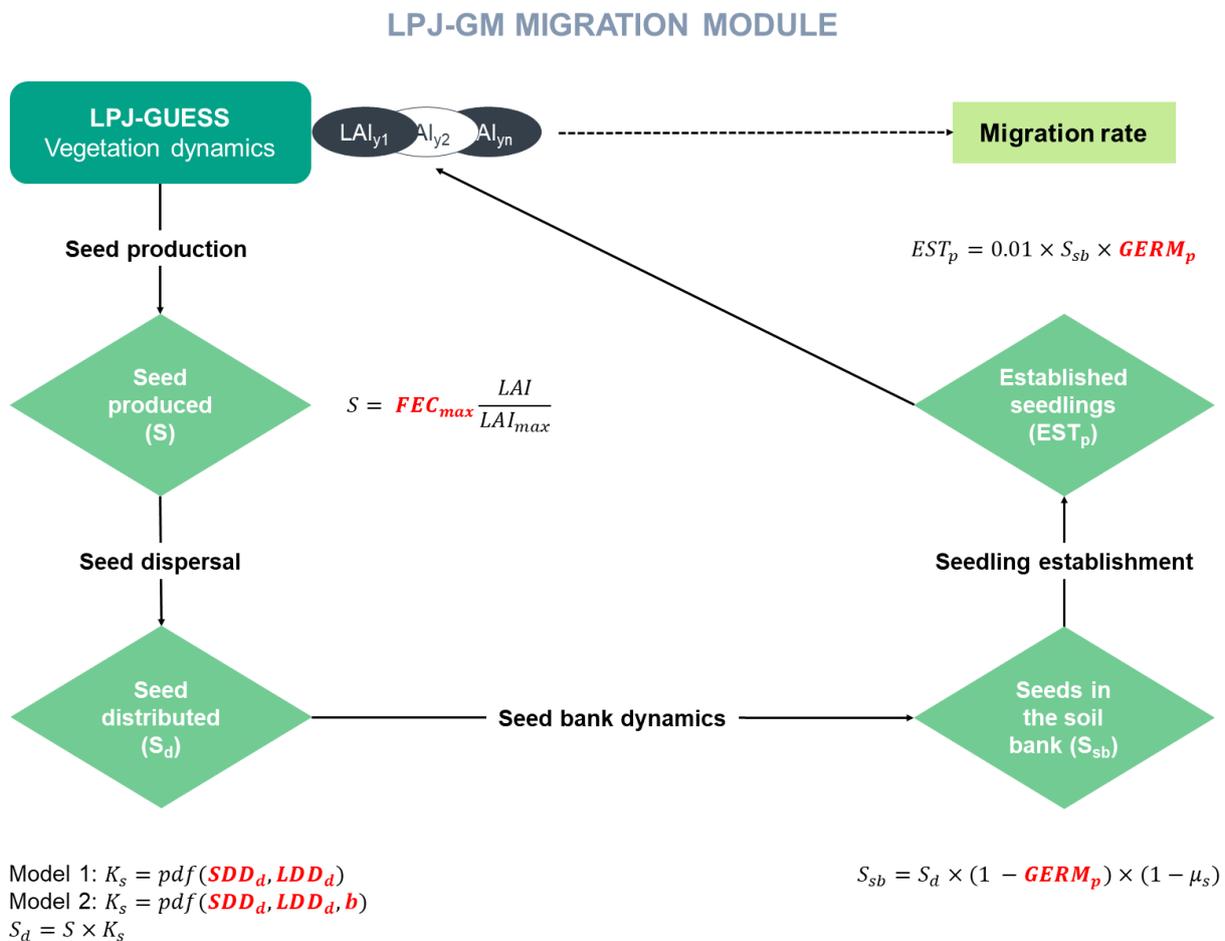



Figure A1: Migration module of the model LPJ-GM 1.0. Migration consists of four processes: 1) seed production, 2) seed dispersal via
690 a probability density function (pdf, dispersal kernel), 3) seed bank dynamics, and 4) seedling establishment. Migration parameters are highlighted in red. Model 1 and Model 2 refer to the default and modified model structures employing negative exponentials or species-specific fat-tailed kernels, respectively, for seed dispersal (see Sect. 2.6.2).

A1. Seed production

695 The number of seeds produced S is the product of maximum fecundity (FEC_{max}) and the proportion of current leaf area (LAI_{ind}) to the maximum (LAI_{max}, calculated following Eq. (3.20) and Eq. (3.21); Bugmann, 1994):

$$S = \text{FEC}_{max} \times \frac{\text{LAI}_{ind}}{\text{LAI}_{max}} , \tag{A1}$$

where species-specific LAI_{ind} is generated yearly for each simulated tree individual by the main vegetation dynamic model (LPJ-GUESS).

700 **A2. Seed distribution**

The seeds $S(x', y')$ produced at a location (x', y') are distributed according to a probability density function (pdf), i.e. the seed dispersal kernel k_s, so that the seed input $S_d(x, y)$ in location x, y is obtained by integrating over all other locations x', y':

$$S_d(x, y) = \int S(x', y') k_s(x - x', y - y') dx' dy' , \tag{A2}$$

The dispersal kernel k_s is a linear combination of two pdfs for short- (SDD) and long-distance dispersal (LDD):

705 $$k_s = (1 - LDD_p) \times pdf(z, SDD_d) + LDD_p \times pdf(z, LDD_d) , \tag{A3}$$

where LDD_p is the proportion of long-distance dispersal (the actual value is species dependent; for example, LDD_p of *Fagus sylvatica* corresponds to 0.01, i.e. 1 % of seed dispersal is attributed to long-distance transport), SDD_d and LDD_d are the average distances for SDD and LDD, respectively, and $z = \sqrt{(x - x')^2 + (y - y')^2}$ is the distance between a sink cell (x, y) and a source location (x', y').

710 In the default setting of LPJ-GM ("Model 1: default kernel" according to the model uncertainty analysis; Sect. 2.6.2), the pdfs for both SDD and LDD components are negative exponentials. In the modified model structure ("Model 2: fat-tailed kernels"; Sect. 2.6.2), the pdfs for both SDD and LDD components are species-specific fat-tailed kernels (see Table 3 for pdfs' formulae).

A3. Seed bank dynamics

715 The seed bank dynamic is defined by the yearly change of dormant seeds in the soil S_{sb} that can germinate in the following years. S_{sb} increases by the seed input S_d according to Eq. (A2) and decreases by the number of germinated seeds or by loss of seeds (μ_s):

$$S_{sb,t+1} = \widetilde{S_{sb,t+1}} \times (1 - GERM_p) \times (1 - \mu_s) , \tag{A4}$$

with

720 $$\widetilde{S_{sb,t+1}} = S_{sb,t} + S_{d,t+1} , \tag{A5}$$

where $GERM_p$ is the rate of germination, and the yearly loss of seeds from the seed bank μ_s is 0.8.

A4. Seedling establishment and calculation of migration rate

Finally, the probability of seedling establishment in a certain year EST_p depends on the number of available seeds for germination (S_{sb}) and on the germination rate:

$$EST_p = 0.01 \times S_{sb} \times GERM_p \, , \tag{A6}$$

New individuals are then established as saplings, where sapling numbers correspond to the rounded value for EST_p. The established seedlings grow, compete and die according to the LPJ-GUESS algorithm, and finally start producing seeds following Eq. (A1).

At the end of the simulation, the species-specific migration rate (in m y^{-1}) is calculated as the migration distance divided by migration time, i.e. the simulation time elapsed since the end of the spin-up phase when the vegetation, soil and litter pools develop from "bare ground" into a dynamic equilibrium. During this phase, all simulated species (i.e. the focal migrating species and its competitors; see Table 1) are allowed to establish without seed limitation, whereas migrating species are killed at the end of the spin-up phase throughout the simulation domain, with the exception of the starting point of migration. Migration distance is then obtained by the direct output of LPJ-GM, yearly and species-specific leaf area index (LAI), as the distance between the starting point of migration and the 95th percentile farthest point in the terrain where LAI exceeds 0.5.

A5. Methods to enhance dispersal simulations

LPJ-GM implements 1) the Fast Fourier transform method (FFTM) to enhance the computational efficiency of dispersal simulations, and 2) the use of corridors to reduce the spatial domain of simulations (i.e. number of grid cells where local dynamics are simulated). The FFTM employs the convolution theorem and the Fast Fourier transform (FFT) to evaluate the convolution of seed production and dispersal kernel (Eq. (A2)) at reduced computational cost, while allowing for the inclusion of different species-specific dispersal kernels and, thus potentially dispersal syndromes (e.g. via wind or animal transport).

Appendix B. Migration parameters

The migration parameters were compiled from various literature sources (Table B1) as described in Sect. 2.4. Optimal values (i.e. corresponding to minimized errors between predicted and observed migration rate estimates) are listed in Table B2.

Table B1. Species-specific migration parameters and data sources for minimum and maximum values found in the literature. SDD$_d$: average short dispersal distance (meters); LDD$_d$: 1 % average long dispersal distance (meters); FEC$_{max}$: maximum fecundity per tree (no. seeds in 100 per year); GERM$_p$: seed germination rate (%). Data sources: a = Lischke and Löffler (2006); b = TRY database (Kattge et al., 2020); c = Royal Botanic Gardens Kew Seed Information Database (SID); d = Vittoz

750 and Engler (2007); e = Tamme et al. (2014); * = species-specific variation of the 5-points parameter values is set at 12.5 % when data is lacking (e.g. for SDD, with a parameter range of 100 m, 25–125 m, the difference between each of the adjacent LSA-5-points is 12.5 m; see Sect. 2.4). Note that differently from the source literature for the default setting (Lischke and Löffler, 2006), we assumed the proportion of LDD events to be fixed at 1 % (LDD$_p$ = 0.01). In the case of absence of LDD events (k = 0 according to Lischke and Löffler, 2006), we assumed equal values for average short- and long- dispersal distances

755 (SDD$_d$ = LDD$_d$).

Parameter	Species	Default	Min	25th	Mean	75th	Max
FEC$_{max}$	Abi_alb	50[a]	19[a]	34.5	50	65.5	81[a]
	Bet_pen	11775[a]	1725[a]	8793.75	15862.5	22931.25	30000[a]
	Bet_pub	11775[a]	1725[a]	8793.75	15862.5	22931.25	30000[a]
	Car_bet	154[a]	23	193.5	364	534.5	705[b]
	Cor_ave	6[a]	6[a]	7.5	9	10.5	12[b]
	Fag_syl	29[a]	2[b]	17	32	47	62[a]
	Fra_exc	42[a]	42[a]	44	46	48	50[b]
	Pic_abi	97[a]	47[a]	76	105	134	163[a]
	Pic_sit	50[b]	25[b]	37.5	50	62.5	75[b]
	Pin_syl	22[a]	6[a]	15.25	24.5	33.75	43[a]
	Pin_hal	22[a]	6[a]	15.25	24.5	33.75	43[a]
	Que_coc	5[b]	3[b]	4.75	6.5	8.25	10[b]
	Que_ile	20[a]	10[a]	20	30	40	50[b]
	Que_pub	18[a]	10[a]	20	30	40	50[b]
	Que_rob	28[a]	15[a]	23.75	32.5	41.25	50[b]
	Til_cor	720[a]	540*	630	720[a]	810	900*
	Ulm_gla	372[a]	55[a]	278.5	502	725.5	949[a]
GERM$_p$	Abi_alb	46[a]	30[a]	37.5	45	52.5	60[a]
	Bet_pen	19[a]	10[a]	15	20	25	30[a]
	Bet_pub	19[a]	10[a]	15	20	25	30[a]
	Car_bet	67[a]	60[a]	65	70	75	80[c]
	Cor_ave	30[a]	30[a]	37.5	45	52.5	60[c]
	Fag_syl	71[a]	50[a]	57.5	65	72.5	80[a]
	Fra_exc	60[a]	50[a]	53.75	57.5	61.25	65[a]
	Pic_abi	76[a]	1[a]	24.5	48	71.5	95[a]
	Pic_sit	75[c]	70[c]	72.5	75	77.5	80[c]

	Pin_syl	91[a]	85[a]	87.5	90	92.5	95[a]
	Pin_hal	60[a]	60[c]	60	60	60	60[c]
	Que_coc	70[c]	65[c]	67.5	70	72.5	75[c]
	Que_ile	90[c]	85[c]	87.5	90	92.5	95[c]
	Que_pub	70[a]	60[a]	67.5	75	82.5	90[a]
	Que_rob	75[a]	60[a]	68.75	77.5	86.25	95[a]
	Til_cor	45[a]	20[a]	28.75	37.5	46.25	55[a]
	Ulm_gla	35[a]	30[a]	38.75	47.5	56.25	65[c]
SDD_d	Abi_alb	100[a]	75*	87.5	100[a]	112.5	125*
	Bet_pen	200[a]	150*	175	200[a]	225	250*
	Bet_pub	200[a]	150*	175	200[a]	225	250*
	Car_bet	100[a]	75*	87.5	100[a]	112.5	125*
	Cor_ave	25[a]	25[a]	68.75	112.5	156.25	200[e]
	Fag_syl	25[a]	4.13[d]	9.3475	14.565	19.7825	25[a]
	Fra_exc	100[a]	75*	87.5	100[a]	112.5	125*
	Pic_abi	100[a]	75*	87.5	100[a]	112.5	125*
	Pic_sit	100[a]	75*	87.5	100[a]	112.5	125*
	Pin_syl	100[a]	75*	87.5	100[a]	112.5	125*
	Pin_hal	100[a]	75*	87.5	100[a]	112.5	125*
	Que_coc	25[a]	25[a]	32.5	50	62.5	75*
	Que_ile	25[a]	25[a]	32.5	50	62.5	75*
	Que_pub	25[a]	25[a]	32.5	50	62.5	75*
	Que_rob	25[a]	25[a]	32.5	50	62.5	75*
	Til_cor	100[a]	75*	87.5	100[a]	112.5	125*
	Ulm_gla	100[a]	75*	87.5	100[a]	112.5	125*
LDD_d	Abi_alb	101[a]	101[a]	1825.75	3550.5	5275.25	7000[d]
	Bet_pen	201[a]	201[a]	269.5	338	406.5	475[e]
	Bet_pub	201[a]	201[a]	269.5	338	406.5	475[e]
	Car_bet	101[a]	101[a]	182	263	344	425[e]
	Cor_ave	200[a]	200[a]	525	850	1175	1500[d]
	Fag_syl	200[a]	32[d]	74	116	158	200[a]
	Fra_exc	101[a]	101[a]	257	413	569	725[d]
	Pic_abi	101[a]	101[a]	450.75	800.5	1150.25	1500[a]

Species						
Pic_sit	101[a]	101[a]	275.75	450.5	625.25	800[e]
Pin_syl	101[a]	101[a]	138.25	175.5	212.75	250[e]
Pin_hal	101[a]	101[a]	138.25	175.5	212.75	250[e]
Que_coc	200[a]	200[a]	225	250	275	300[d]
Que_ile	200[a]	200[a]	225	250	275	300[d]
Que_pub	200[a]	200[a]	225	250	275	300[d]
Que_rob	200[a]	200[a]	225	250	275	300[d]
Til_cor	101[a]	101[a]	169.25	237.5	305.75	374[e]
Ulm_gla	101[a]	101[a]	163.25	225.5	287.75	350[e]

Table B2. Species-specific optimal migration parameters with related RMSE % (relative to the 75th percentile of historical estimates; see Table 1) and dispersal syndromes as reported by Vittoz and Engler (2007) and TRY Database (Kattge et al. 2020): W = wind; Wa = water; B = bird; LA = large mammal (deer, badger, cattle); SA = small animal (e.g. hoarding by rodents).

Species	Kernel	B	FEC_{max}	$GERM_p$	SDD_d	LDD_d	RMSE %	Dispersal syndrome
Abi_alb	NegExp	NA	50	46	100	710	2.66	W
Bet_pen	LogSec	0.29	11775	19	200	475	19.71	W+Wa+B
Bet_pub	LogSec	0.29	11775	19	200	475	19.71	W+Wa+B
Car_bet	Logistic	4.75	705	80	100	425	19.63	W+Wa+B+SA
Cor_ave	LogSec	0.4	6	60	200	1500	12.21	B+SA
Fag_syl	LogSec	0.375	29	71	25	200	14.19	SA+LA
Fra_exc	ExpPow	0.5	42	60	100	725	25.75	W+Wa+B+LA
Pic_abi	NegExp	NA	163	80	100	780	9.21	W+Wa+B+LA
Pic_sit	ExpPow	0.5	50	75	100	800	31.15	W
Pin_hal	Logistic	4	22	60	100	250	29.20	W+Wa
Pin_syl	Logistic	4.5	43	91	100	250	34.32	W
Que_coc	Logistic	4.15	5	70	25	200	21.11	SA
Que_ile	Logistic	4.5	50	90	25	200	15.25	SA
Que_pub	LogSec	0.525	50	90	25	200	14.69	SA
Que_rob	LogSec	0.5	50	95	25	200	19.53	SA
Til_cor	ExpPow	0.45	720	55	100	374	31.41	SA
Ulm_gla	LogSec	0.35	725	65	100	350	17.12	W+SA

Code and data availability

The model LPJ-GM 2.0 (LPJ-GUESS 4.0 coupled to a dynamic migration module) was used for the simulations presented in the study. The model code is archived in a private repository (https://github.com/zanid90/LPJ-GMINT), along with detailed

765   instructions on how to compile and run the model version presented in this paper, where the model version is identified by the permanent version number v2.0-gm in the code repository. Access to the code is available upon request under license due to university policy (Lund University). Documentation and code base of the main-version of LPJ-GUESS is available upon request under license via the LPJ-GUESS home page: https://web.nateko.lu.se/lpj-guess/ (last access: 28 October 2021). An open-access educational version of the base code LPJ-GUESS is also available there. Input climate and landscape data filtered

770   for the simulation domain, grid list and instruction file with parameter settings for the LPJ-GM simulations are available at: https://doi.org/10.18161/20211127. Full climate and landscape data were provided by Armstrong et al., 2019.

Supplement

The Supplement contains: 1) the post-processing MATLAB script for the calculation of migration speed from the LPJ-GM output (post_processing/Postprocessing_plotMigrationRate.m with a sample output lai.out); 2) the .csv tables with migration

775   speed separated by analysis (Input_SA.csv and Input_SA_model2.csv for the Sensitivity Analysis of the default and improved model structure, Sect. 2.5.1 and 2.6.2, respectively; Input_EVA.csv for the Extreme Value Analysis, Sect. 2.5.2; Input_KA.csv for the Kernel Analysis, Sect. 2.6.1); 3) the Jupyter Notebook used to generate all figures and tables from the .csv data, and statistical analyses performed in the study, including detailed information on the setting and parameters used in the LPJ-GM simulations (Supplement_Information.ipynb); and 4) supplementary tables and figures (Table S1–4, and Fig. S1–4). The

780   supplement related to this article is available at: https://doi.org/10.18161/20211127.

Author contribution

DZ and VL designed the study, DZ performed the simulations and the statistical analysis, VL provided support with the coding, VL and HL provided statistical support. All authors contributed to the writing of the paper.

Competing interests

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

Acknowledgments

This study was funded by the Swiss National Science foundation project "EXTRAS" 205320M_182609.

References

Albrich, K., Rammer, W., and Seidl, R.: Climate change causes critical transitions and irreversible alterations of mountain forests. Glob. Change Biol., 26, 4013–4027, https://doi.org/10.1111/gcb.15118, 2020.

Alexander, J. M., Chalmandrier, L., Lenoir, J., Burgess, T. I., Essl, F., Haider, S., Kueffer, C., McDougall, K., Nuñez, M. A., Pauchard, A., Rabitsch, W., Rew, L. J., Sanders, N. J., Pellissier, L.: Lags in the response of mountain plant communities to climate change. Glob. Change Biol., 24, 563–579, https://doi.org/10.1111/gcb.13976, 2018.

Armstrong, E., Hopcroft, P. O., and Valdes, P. J.: A simulated Northern Hemisphere terrestrial climate dataset for the past 60,000 years. Sci Data, 6, 1–16, https://doi.org/10.1038/s41597-019-0277-1, 2019.

Baskin, C. C., and Baskin, J. M. Seeds: ecology, biogeography, and, evolution of dormancy and germination. Elsevier, 1998.

Beckman, N. G., Aslan, C. E., Rogers, H. S., Kogan, O., Bronstein, J. L., Bullock, J. M., Hartig, F., HilleRisLambers, J., Zhou, Y., Zurell, D., Brodie, J. F., Bruna, E. M., Cantrell, R. S., Decker, R. R., Efiom, E., Fricke, E. C., Gurski, K., Hastings, A., Johnson, J. S., Loiselle, B. A., Miriti, M. N., Neubert, M. G., Pejchar, L., Poulsen, J. R., Pufal, G., Razafindratsima, O. H., Sandor, M. E., Shea, K., Schreiber, S., Schupp, E. W., Snell, R. S., Strickland, C., and Zambrano, J.: Advancing an interdisciplinary framework to study seed dispersal ecology. AoB Plants, 12, https://doi.org/10.1093/aobpla/plz048, 2020.

Bennett, J. M., Steets, J. A., Burns, J. H., Durka, W., Vamosi, J. C., Arceo-Gómez, G., Burd, M., Burkle, L. A., Ellis, A. G., Freitas, L., Li, J., Rodger, J. G., Wolowski, M., Xia, J., Ashman, T-L., and Knight, T. M.: GloPL, a global data base on pollen limitation of plant reproduction. Scientific Data, 5, 1–9, https://doi.org/10.1038/sdata.2018.249, 2018.

Berg, M. P., Kiers, E. T., Driessen, G., Van Der Heijden, M., Kooi, B. W., Kuenen, F., Liefting, M., Verhoef, H. A., and Ellers, J.: Adapt or disperse: understanding species persistence in a changing world. Glob. Change Biol., 16, 587–598, https://doi.org/10.1111/j.1365-2486.2009.02014.x, 2010.

Binney, H. A., Willis, K. J., Edwards, M. E., Bhagwat, S. A., Anderson, P. M., Andreev, A. A., Blaauw, M., Damblon, F., Haesaerts, P., Kienast, F., Kremenetski, K. V., Krivonogov, S. K., Lozhkin, A. V., MacDonald, G. M., Novenko, E. Y., Oksanen, P., Sapelko, T. V., Väliranta, M., and Vazhenina, L.: The distribution of late-Quaternary woody taxa in northern Eurasia: evidence from a new macrofossil database. Quaternary Sci. Rev. 28, 2445–2464, https://doi.org/10.1016/j.quascirev.2009.04.016, 2009.

Birks, H. J. B.: Contributions of Quaternary botany to modern ecology and biogeography. Plant Ecol. Divers., 12, 189-385, https://doi.org/10.1080/17550874.2019.1646831, 2019.

Birks, H. J. B., and Birks, H. H.: Biological responses to rapid climate change at the Younger Dryas—Holocene transition at Kråkenes, western Norway. Holocene, 18, 19–30, https://doi.org/10.1177/0959683607085572, 2008.

Briscoe, N. J., Elith, J., Salguero-Gómez, R., Lahoz-Monfort, J. J., Camac, J. S., Giljohann, K. M., Holde, M. H., Hradsky, B. A., Kearney, M. R., McMahon, S. M., Phillips, B. L., Regan, T. J., Rhodes, J. R., Vesk, P. A., Wintle, B. A., Yen, J. D. L., and Guillera-Arroita, G.: Forecasting species range dynamics with process-explicit models: matching methods to applications. Ecol. Lett., 22, 1940–1956, https://doi.org/10.1111/ele.13348, 2019.

Bullock, J. M., and Clarke, R. T.: Long distance seed dispersal by wind: measuring and modelling the tail of the curve. Oecologia, 124, 506–521, https://doi.org/10.1007/PL00008876, 2000.

Bullock, J. M., Mallada González, L., Tamme, R., Götzenberger, L., White, S. M., Pärtel, M., and Hooftman, D. A.: A synthesis of empirical plant dispersal kernels. J. Ecol., 105, 6–19, https://doi.org/10.1111/1365-2745.12666, 2017.

825 Cain, M. L., Damman, H., and Muir, A.: Seed dispersal and the Holocene migration of woodland herbs. Ecol. Monogr, 68, 325–347, https://doi.org/10.1890/0012-9615(1998)068[0325:SDATHM]2.0.CO;2, 1998.

Caswell, H., Lensink, R., and Neubert, M. G.: Demography and dispersal: life table response experiments for invasion speed. Ecology, 84, 1968–1978, https://doi.org/10.1890/02-0100, 2003.

Cheaib, A., Badeau, V., Boe, J., Chuine, I., Delire, C., Dufrêne, E., François, C., Gritti, E. S., Legay, M., Pagé, C., Thuiller, 830 W., Viovy, N., and Leadley, P.: Climate change impacts on tree ranges: model intercomparison facilitates understanding and quantification of uncertainty. Ecology Letters, 15, 533–544, https://doi.org/10.1111/j.1461-0248.2012.01764.x, 2012.

Clark, J. S.: Why trees migrate so fast: confronting theory with dispersal biology and the paleorecord. The American Naturalist, 152, 204–224, https://doi.org/10.1086/286162, 1998.

Corlett, R. T., and Westcott, D. A.: Will plant movements keep up with climate change?. Trends Ecol. Evol., 28, 482–488, 835 https://doi.org/10.1016/j.tree.2013.04.003, 2013.

Clark, J. S., Silman, M., Kern, R., Macklin, E., and HilleRisLambers, J.: Seed dispersal near and far: patterns across temperate and tropical forests. Ecology, 80, 1475–1494, https://doi.org/10.1890/0012-9658(1999)080[1475:SDNAFP]2.0.CO;2, 1999.

Clark, J. S., Lewis, M., and Horvath, L.: Invasion by extremes: population spread with variation in dispersal and reproduction. Am. Nat., 157, 537–554, https://doi.org/10.1086/319934, 2001.

840 Downing, D. J., Gardner, R. H., and Hoffman, F. O.: An examination of response-surface methodologies for uncertainty analysis in assessment models. Technometrics, 27, 151–163, https://doi.org/10.1080/00401706.1985.10488032, 1985.

Dullinger, S., Dirnböck, T., and Grabherr, G.: Modelling climate change-driven treeline shifts: relative effects of temperature increase, dispersal and invasibility. J. Ecol., 92, 241–252, https://doi.org/10.1111/j.0022-0477.2004.00872.x, 2004.

Feurdean, A., Bhagwat, S. A., Willis, K. J., Birks, H. J. B., Lischke, H., and Hickler, T.: Tree migration-rates: narrowing the 845 gap between inferred post-glacial rates and projected rates. PLoS One, 8, e71797, https://doi.org/10.1371/journal.pone.0071797, 2013.

Giesecke, T., and Brewer, S.: Notes on the postglacial spread of abundant European tree taxa. Veg. Hist. Archaeobot., 27, 337–349, https://doi.org/10.1007/s00334-017-0640-0, 2018.

Giesecke, T., Brewer, S., Finsinger, W., Leydet, M., and Bradshaw, R. H.: Patterns and dynamics of European vegetation 850 change over the last 15,000 years. J. Biogeogr., 44, 1441–1456, https://doi.org/10.1111/jbi.12974, 2017.

Goto, S., Shimatani, K., Yoshimaru, H., and Takahashi, Y.: Fat-tailed gene flow in the dioecious canopy tree species Fraxinus mandshurica var. japonica revealed by microsatellites. Mol. Ecol., 15, 2985–2996, https://doi.org/10.1111/j.1365-294X.2006.02976.x, 2006.

Hamby, D. M.: A comparison of sensitivity analysis techniques. Health Phys., 68, 195–204, 1995.

855 Higgins, S. I., Clark, J. S., Nathan, R., Hovestadt, T., Schurr, F., Fragoso, J. M. V., Aguiar, M. R., Ribbens, E., and Lavorel, S.: Forecasting plant migration rates: managing uncertainty for risk assessment. J. Ecol., 91, 341–347, https://doi.org/10.1046/j.1365-2745.2003.00781.x, 2003a.

Higgins, S. I., Nathan, R., and Cain, M. L.: Are long-distance dispersal events in plants usually caused by nonstandard means of dispersal?. Ecology, 84, 1945–1956, https://doi.org/10.1890/01-0616, 2003b.

860 Huntley, B., and Birks, H. J. B. (Eds.): Atlas of past and present pollen maps for Europe: 0-13,000 years ago. Cambridge University Press, 1983.

Huntley, B., Allen, J. R., Collingham, Y. C., Hickler, T., Lister, A. M., Singarayer, J., Stuart, A. J., Sykes, M. T., and Valdes, P. J.: Millennial climatic fluctuations are key to the structure of last glacial ecosystems. PLoS One, 8, e61963, https://doi.org/10.1371/journal.pone.0061963, 2013.

865 Kattge, J., Bönisch, G., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., ... and Cuntz, M. TRY plant trait database–enhanced coverage and open access. Glob Change Biol., 26, 119–188, https://doi.org/10.1111/gcb.14904 , 2020.

Klein, E. K., Lavigne, C., Picault, H., Renard, M., and Gouyon, P. H.: Pollen dispersal of oilseed rape: estimation of the dispersal function and effects of field dimension. J. Appl. Ecol., 43, 141–151, https://doi.org/10.1111/j.1365-2664.2005.01108.x, 2006.

870 Lehsten, V., Mischurow, M., Lindström, E., Lehsten, D., and Lischke, H.: LPJ-GM 1.0: simulating migration efficiently in a dynamic vegetation model. Geosci. Model Dev., 12, 893–908, https://doi.org/10.5194/gmd-12-893-2019, 2019.

Lenoir, J., Bertrand, R., Comte, L., Bourgeaud, L., Hattab, T., Murienne, J., and Grenouillet, G.: Species better track climate warming in the oceans than on land. Nature Ecol. Evol., 4, 1044–1059, https://doi.org/10.1038/s41559-020-1198-2, 2020.

Lischke, H., and Löffler, T. J.: Intra-specific density dependence is required to maintain species diversity in spatio-temporal
875 forest simulations with reproduction. Ecol. Model., 198, 341–361, https://doi.org/10.1016/j.ecolmodel.2006.05.005, 2006.

Lischke, H., Zimmermann, N. E., Bolliger, J., Rickebusch, S., and Löffler, T. J.: TreeMig: a forest-landscape model for simulating spatio-temporal patterns from stand to landscape scale. Ecol. Model., 199, 409–420, https://doi.org/10.1016/j.ecolmodel.2005.11.046, 2006.

Lustenhouwer, N., Moran, E. V., and Levine, J. M.: Trait correlations equalize spread velocity across plant life histories.
880 Global Ecol. Biogeogr., 26, 1398–1407, https://doi.org/10.1111/geb.12662, 2017.

Loehle, C.: Challenges of ecological complexity. Ecol. Complex., 1, 3–6, https://doi.org/10.1016/j.ecocom.2003.09.001, 2004.

Manel, S., Gaggiotti, O. E., and Waples, R. S.: Assignment methods: matching biological questions with appropriate techniques. Trends Ecol. Evol., 20, 136–142, https://doi.org/10.1016/j.tree.2004.12.004, 2005.

MacDonald, G. M.: Fossil pollen analysis and the reconstruction of plant invasions. Adv. Ecol. Res., 24, 67–110,
885 https://doi.org/10.1016/S0065-2504(08)60041-0, 1993.

McKenzie, P. F., Duveneck, M. J., Morreale, L. L., and Thompson, J. R.: Local and global parameter sensitivity within an ecophysiologically based forest landscape model. Environ. Modell. Soft., 117, 1–13, https://doi.org/10.1016/j.envsoft.2019.03.002, 2019.

Mersmann, O., Bischl, B., Trautmann, H., Preuss, M., Weihs, C., and Rudolph, G.: Exploratory landscape analysis. In
890 Proceedings of the 13th annual conference on Genetic and evolutionary computation (pp. 829–836),
https://doi.org/10.1145/2001576.2001690, July 2011.

Mladenoff, D. J.: LANDIS and forest landscape models. Ecol. Modell., 180, 7–19,
https://doi.org/10.1016/j.ecolmodel.2004.03.016, 2004.

Morales, P., Hickler, T., Rowell, D. P., Smith, B., and Sykes, M. T.: Changes in European ecosystem productivity and carbon
895 balance driven by regional climate model output. Global Change Biology, 13, 108–122, https://doi.org/10.1111/j.1365-
2486.2006.01289.x, 2007.

Moran, E. V., and Clark, J. S.: Estimating seed and pollen movement in a monoecious plant: a hierarchical Bayesian approach
integrating genetic and ecological data. Mol. Ecol., 20, 1248–1262, https://doi.org/10.1111/j.1365-294X.2011.05019.x, 2011.

Nathan, R., and Katul, G. G.: Foliage shedding in deciduous forests lifts up long-distance seed dispersal by wind. PNAS, 102,
900 8251–8256, https://doi.org/10.1073/pnas.0503048102, 2005.

Nathan, R., Klein, E. K., Robledo-Arnuncio, J. J., and Revilla, E.: Dispersal kernels: review, in Dispersal Ecology and
Evolution, 187–210, 2012.

Nobis, M. P., and Normand, S.: KISSMig – a simple model for R to account for limited migration in analyses of species
distributions. Ecography, 37, 1282–1287, https://doi.org/10.1111/ecog.00930, 2014.

905 Normand, S., Ricklefs, R. E., Skov, F., Bladt, J., Tackenberg, O., and Svenning, J. C.: Postglacial migration supplements
climate in determining plant species ranges in Europe. P. Roy. Soc. B-Biol. Sci., 278, 3644–3653,
https://doi.org/10.1098/rspb.2010.2769, 2011.

Nogués-Bravo, D., Rodríguez-Sánchez, F., Orsini, L., de Boer, E., Jansson, R., Morlon, H., Fordham, D. A., and Jackson, S.
T.: Cracking the code of biodiversity responses to past climate change. Trends Ecol. Evol., 33, 765–77,
910 https://doi.org/10.1016/j.tree.2018.07.005, 2018.

Palmé, A. E., Su, Q., Rautenberg, A., Manni, F., and Lascoux, M.: Postglacial recolonization and cpDNA variation of silver
birch, Betula pendula. Mol. Ecol., 12, 201–212, https://doi.org/10.1046/j.1365-294X.2003.01724.x, 2003.

Pappas, C., Fatichi, S., Leuzinger, S., Wolf, A., and Burlando, P.: Sensitivity analysis of a process-based ecosystem model:
Pinpointing parameterization and structural issues. J. Geophys. Res-Biogeo., 118, 505–528,
915 https://doi.org/10.1002/jgrg.20035, 2013.

Pecl, G. T., Araújo, M. B., Bell, J. D., Blanchard, J., Bonebrake, T. C., Chen, I. C., Clark, T. D., Colwel, R. K., Danielsen, F.,
Evengård, B., Falconi, L., Ferrier, S., Frusher, S., Garcia, R. A., Griffis, R. B., Hobday, A. J., Janion-Scheepers, C., Jarzyna,
M. A., Jennings, S., Lenoir, J., Linnetved, H. I., Martin, V. Y., McCormack, P. C., McDonald, J., Mitchell, N. J., Mustonen,
T., Pandolfi, J. M., Pettorelli, N., Popova, E., Robinson, S. A., Scheffers, B. R., Shaw, J. D., Sorte, C. J. B., Strugnell, J. M.,
920 Sunday, J. M., Tuanmu, M.-N., Vergés, A., Villanueva, C., Wernberg, T., Wapstra, E., and Williams, S. E.: Biodiversity
redistribution under climate change: Impacts on ecosystems and human well-being. Science, 355, 10.1126/science.aai921,
2017.

Petit, R. J., Brewer, S., Bordács, S., Burg, K., Cheddadi, R., Coart, E., Cottrell, J., Csaikl, U. M., van Dam, B., Deans, J. D., Espinel, S., Fineschi, S., Finkeldey, R., Glaz, I., Goicoechea, P. G., Jensen, J. S., König, A. O., Lowe, A. J., Madsen, S. F., Mátyás, G., Munro, R. C., Popescu, F., Slade, D., Tabbener J., de Vries, S. G. M., Ziegenhagen, B., de Beaulieu J.-L., and Kremer, A.: Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. Forest Ecol. Manag., 156, 49–74, https://doi.org/10.1016/S0378-1127(01)00634-X, 2002.

Petter, G., Mairota, P., Albrich, K., Bebi, P., Brůna, J., Bugmann, H., Haffenden, A., Scheller, R. M., Schmatz, D. R., Seidl, R., Speich, M., Vacchiano, G., and Lischke, H.: How robust are future projections of forest landscape dynamics? Insights from a systematic comparison of four forest landscape models. Environ. Modell. Soft., 134, 104844, https://doi.org/10.1016/j.envsoft.2020.104844, 2020.

Reid, C.: The origin of the British flora. Dulau, London, 1899.

Roberts, N. (Eds.): The Holocene: an environmental history, John Wiley & Sons, 2013.

Rogers, H. S., Beckman, N. G., Hartig, F., Johnson, J. S., Pufal, G., Shea, K., Zurell, D., Bullock, J. M., Cantrell, R. S., Loiselle, B., Pejchar, L., Razafindratsima, O. H., Sandor, M. E., Schupp, E. W., Strickland, W. C., and Zambrano, J.: The total dispersal kernel: a review and future directions. AoB Plants, 11, https://doi.org/10.1093/aobpla/plz042, 2019.

Royal Botanic Gardens Kew, Seed Information Database (SID). Version 7.0. Available from: http://data.kew.org/sid/, last access: 19 December 2019.

Saltelli, A., Tarantola, S., and Campolongo, F.: Sensitivity analysis as an ingredient of modeling. Statist. Sci., 15, 377–395, https://doi.org/10.1214/ss/1009213004, 2000.

Scherrer, D., Vitasse, Y., Guisan, A., Wohlgemuth, T., and Lischke, H.: Competition and demography rather than dispersal limitation slow down upward shifts of trees' upper elevation limits in the Alps. J. Ecol., 108, 2416-2430, https://doi.org/10.1111/1365-2745.13451, 2020.

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

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

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

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

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

Soons, M. B., Heil, G. W., Nathan, R., and Katul, G. G.: Determinants of long-distance seed dispersal by wind in grasslands. Ecology, 85, 3056–3068, https://doi.org/10.1890/03-0522, 2004.

Snowling, S. D., and Kramer, J. R.: Evaluating modelling uncertainty for model selection. Ecol. Modell., 138, 17–30, https://doi.org/10.1016/S0304-3800(00)00390-2, 2001.

960   Stewart, J. R., and Lister, A. M.: Cryptic northern refugia and the origins of the modern biota. Trends Ecol. Evol., 16, 608–613, https://doi.org/10.1016/S0169-5347(01)02338-2, 2001.

Stork, J., Eiben, A. E., and Bartz-Beielstein, T.: A new taxonomy of global optimization algorithms. Nat. Comput., 1–24, https://doi.org/10.1007/s11047-020-09820-4, 2020.

Svenning, J. C., and Sandel, B.: Disequilibrium vegetation dynamics under future climate change. Am. J. Bot., 100, 1266–
965   1286, https://doi.org/10.3732/ajb.1200469, 2013.

Tamme, R., Götzenberger, L., Zobel, M., Bullock, J. M., Hooftman, D. A., Kaasik, A., and Pärtel, M.: Predicting species' maximum dispersal distances from simple plant traits. Ecology, 95, 505–513, https://doi.org/10.1890/13-1000.1, 2014.

Thompson, P. L., and Fronhofer, E. A.: The conflict between adaptation and dispersal for maintaining biodiversity in changing environments. PNAS, 116, 21061–21067, https://doi.org/10.1073/pnas.1911796116, 2019.

970   Tomiolo, S., and Ward, D.: Species migrations and range shifts: A synthesis of causes and consequences. Perspect. Plant Ecol., 33, 62–77, https://doi.org/10.1016/j.ppees.2018.06.001, 2018.

Tzedakis, P. C., Emerson, B. C., and Hewitt, G. M.: Cryptic or mystic? Glacial tree refugia in northern Europe. Trends Ecol. Evol., 28, 696–704, https://doi.org/10.1016/j.tree.2013.09.001, 2013.

Vittoz, P., and Engler, R.: Seed dispersal distances: a typology based on dispersal modes and plant traits. Bot. Helv., 117, 109–
975   124, https://doi.org/10.1007/s00035-007-0797-8, 2007.

Wallingford, P. D., Morelli, T. L., Allen, J. M., Beaury, E. M., Blumenthal, D. M., Bradley, B. A., Dukes, J. S., Early, R., Fusco, E. J., Goldberg, D. E., Ibáñez, I., Laginhas, B. B., Vilà, M., and Sorte, C. J.: Adjusting the lens of invasion biology to focus on the impacts of climate-driven range shifts. Nat. Clim. Change, 10, 398–405, https://doi.org/10.1038/s41558-020-0768-2, 2020.

980   Wilson, G. A., and Rannala, B.: Bayesian inference of recent migration rates using multilocus genotypes. Genetics, 163, 1177–1191, https://doi.org/10.1093/genetics/163.3.1177, 2003.

Wramneby, A., Smith, B., Zaehle, S., and Sykes, M. T.: Parameter uncertainties in the modelling of vegetation dynamics—effects on tree community structure and ecosystem functioning in European forest biomes. Ecol. Model., 216, 277–290, https://doi.org/10.1016/j.ecolmodel.2008.04.013, 2008.

985   Zaehle, S., Sitch, S., Smith, B., and Hatterman, F.: Effects of parameter uncertainties on the modeling of terrestrial biosphere dynamics. Global Biogeochem. Cy., 19, https://doi.org/10.1029/2004GB002395, 2005.