

Predicting the dynamics of establishing tree populations: A framework for statistical inference and lessons for data collection

Dominique Lamonica^{1,2}  | Jörn Pagel¹  | Frank M. Schurr¹ 

¹Institute of Landscape and Plant Ecology, University of Hohenheim, Stuttgart, Germany

²RiverLy, HYNES (Irstea-EDF R&D), INRAE, Villeurbanne, France

Correspondence

Dominique Lamonica
Email: dominique.lamonica@gmail.com

Funding information

D.L., J.P. and F.M.S. were funded by the German Research Foundation (DFG, SCHU 2259/7-1) in the framework of ERA Net BiodivERSA project Sponforest.

Handling Editor: Res Altwegg

Abstract

1. Under global change, there is an urgent need to forecast the dynamics of establishing tree populations. However, tree population dynamics are slow and historical data on these dynamics are rare. This raises the question whether tree population dynamics can be reconstructed from data collected at a single time point. Doing so poses challenges for modelling, data collection and model-data integration.
2. We present a Bayesian framework that uses multiple data types to parametrize an individual-based model (IBM) for the growth of establishing tree populations. The framework combines likelihood-based Bayesian inference and approximate Bayesian computation (ABC). Using this framework, we assess the information content of three data types (recruitment data, dendrochronological data describing individual growth and molecular markers characterizing within-population pedigrees) by comparing the bias and uncertainty of parameter estimates and model forecasts obtained under different simulated scenarios of data availability.
3. The combination of all data types leads to accurate forecasts of the future state of tree populations, despite large uncertainties in some parameter estimates. Dendrochronological data were the most informative of the examined data types. Combining data types improved forecasts of population state. Nevertheless, for a given parameter related to a given process, combining data types did not improve estimates compared to using only the data type most closely related to the process.
4. The presented Bayesian framework allows to infer the dynamics of establishing tree populations from data collected at a single time point. It helps to optimally allocate limited resources for data collection in order to rapidly improve the understanding and forecasting of tree population dynamics.

KEYWORDS

Approximate Bayesian computation, Bayesian model-data integration, dendrochronology, ecological forecasting, forest dynamics, Individualbased model, molecular markers, parameter estimation

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society

1 | INTRODUCTION

Forecasting the dynamics of non-equilibrium populations is of prime interest in the context of global change. In particular, the dynamics of establishing tree populations play a key role for phenomena as diverse as the spread of invasive tree species (Chornesky et al., 2006), the migration of tree species under climate change (Nathan et al., 2011), biome shifts from tundra to boreal forests or from grasslands to woodlands (Frost & Epstein, 2014) and carbon sequestration (Bastin et al., 2019). Tree population dynamics are also critical in the context of land use abandonment which can lead to the emergence of new forests (Lira et al., 2012) and forest recolonization in fragmented landscapes (Palmero-Iniesta et al., 2020). Assessing and ultimately managing these multiple facets of tree population dynamics under global change require a thorough understanding of forest establishment and its underlying ecological mechanisms, as well as reliable forecasts of tree population dynamics. This leads to challenges for modelling, data collection and model-data integration.

Modelling the dynamics of establishing tree populations is challenging because the demographic processes driving these dynamics (individual growth, fecundity, dispersal, recruitment, mortality) are size dependent and thus strongly vary among individuals in a population. Moreover, individuals interact with each other *via* distance-dependent processes such as gene flow and competition for resources. Individual-based models (IBMs) can describe the resulting feedbacks between the demographic performance of individuals and the spatial and size structure of populations (Grimm et al., 2005). Hence, IBMs are well suited for modelling the dynamics of tree populations.

Forecasts of tree population dynamics require not only models but also informative data (Clark et al., 2001). In particular, time series of tree population size structure need to be long enough to cover the long generation times of trees. Hence, when historical data on tree populations are lacking, it is an open question whether—within a short time period—one can collect sufficient data for inferring their past dynamics and forecasting their future dynamics. This may be feasible since several data types contain information about the past dynamics of tree populations

(Grimm et al., 2005). Of particular importance are data on size distribution and spatial structure of populations, dendrochronological data on annual growth increments of individual trees, data on tree fecundity and recruitment as well as ‘genetic data’ on molecular polymorphic markers that can be used for genotyping individuals and reconstructing pedigrees. For instance, rates of individual growth and competition can be inferred by relating dendrochronological data on annual growth increments of individual trees to the spatial structure of the tree stand (Canham et al., 2004; Lamonica et al., 2020). Seed production, seed dispersal and seedling recruitment can be estimated by relating recruitment data on spatial variation in seedling density to the size distribution and spatial structure of adult trees (Uriarte et al., 2005; Schurr et al., 2008). Additionally, high-resolution molecular markers have been used to infer within-population pedigrees that quantify effective pollen and seed production as well as rates of pollen (Klein et al., 2011) and seed dispersal (Klein et al., 2013). These data types differ in the time, material costs and skills required for data collection and analysis (Table 1). To design efficient data collection, it is thus important to quantify the amount of information that each data type provides for forecasting population dynamics.

The final challenge to predicting and forecasting tree population dynamics is the integration of suitable models (notably IBMs) with multiple data types. IBM parametrization still widely uses a ‘piecemeal’ approach, in which different submodels are parametrized independently (Clark & Gelfand, 2006; Moran & Clark, 2011). Bayesian inference allows the use of multiple data types, collected at different scales, as well as the simultaneous estimation of all parameters (Gopalswamy et al., 2012; Lamonica et al., 2016). Bayesian approaches to inference typically depend on having explicit functional forms for the likelihood. However, commonly used likelihood-based methods of parameter estimation are difficult to apply to IBMs, because the likelihood functions of stochastic IBMs usually cannot be calculated explicitly (Hartig et al., 2011, 2014). Alternatively, approximate Bayesian computation (ABC) methods (Beaumont, 2018) use model simulations instead of likelihood computations, and are becoming more and more popular in the field of ecology (Csilléry et al., 2010; Hartig et al., 2011). ABC has also

TABLE 1 Data types that can be collected at a single time point to parametrize an IBM of tree population establishment

Data	Notation	Field time	Lab and analysis time	Material costs	Skills
Size distribution and spatial structure (base scenario)	Base	Low	Low	Low	Low
Genetic	G	Low	High	High	High
Recruitment	R	High	Low	Low	Low
Dendrochronological	D	Low	High	Low	High

Notes. For each data type, the table gives a qualitative estimate of time investment, material costs and skills required for data collection and analysis. Time indications: personal communications of D. Bert, G. Gerzabek and E. Valdes-Correcher, based on the sample of 15 populations of *Quercus robur* L. with mean patch area of 2,500 m² and mean population size of 50 individuals, data being collected in the field and analysed in the laboratory by one person.

been applied to parametrize IBMs of tree population dynamics from historical forest inventory data (Hartig et al., 2014; Lagarrigues et al., 2015). However, the efficiency of ABC methods to parametrize IBMs of tree population dynamics from different data types remains to explore.

This paper has two aims: first, we develop a statistical framework that combines likelihood-based and ABC methods to parametrize an IBM of establishing tree populations which have not reached equilibrium yet and in which the initial founder tree may still be alive. This framework can integrate multiple data types collected at a single time point (size distribution and spatial structure, genetic, recruitment and dendrochronological data). Second, we quantify how informative these different data types are for understanding and forecasting tree population dynamics. These results provide guidelines for the optimal allocation of limited resources for data collection in order to improve the understanding and forecasting of tree population dynamics.

2 | MATERIALS AND METHODS

We formulated an individual-based model (IBM) for the dynamics of an establishing tree population and then developed a statistical framework for the estimation of model parameters from combinations of different data types (Figure 1). To test the framework and to quantify the amount of information provided by different data types, we used a 'virtual ecologist' approach (Zurell et al., 2010), collecting virtual data from simulated forest population dynamics. We consider the dynamics of an isolated establishing tree population (Lamonica et al., 2020; Ruiz-Carbayo et al., 2020) focusing on establishment of a new forest patch rather than gradual forest expansion (Clark et al., 2001; Palmero-Iniesta et al., 2020). We applied the framework to a range of scenarios of data availability that represent different combinations of collected data types and evaluated the bias and uncertainty of parameter estimates and of forecast forest dynamics under each scenario. To each data type, we chose a best-case scenario in terms of data availability and quality. For data simulation, Bayesian model parametrization and forecasting we used R 3.4.4 (R Core Team, 2018). R Code is supplied as Supporting Information.

2.1 | Simulation model

2.1.1 | General overview

The individual-based, spatially explicit and grid-based model describes the dynamics of forest establishment within a patch. The grid is divided in square cells of 1 m². Each tree is characterized by its position on the grid (the cell number n_i), its mother and father tree and by its size, namely diameter at breast height (DBH) in a given time step, $x(i, t)$. The following processes are modelled: growth, adult mortality, fecundity and pollen and seed dispersal, and seedling

survival. The time step of the model is 1 year. The model structure is shown in Figure 1 and model parameters are listed in Table 2.

2.1.2 | Processes

Individual growth and competition

Annual tree growth depends both on the current tree size and competition from neighbouring trees. The mean of the logarithm of the absolute growth rate $\mu_g(i, t)$ of individual i at time t is modelled following Uriarte et al. (2004):

$$\mu_g(i, t) = \log \left(a_1 \times \exp \left(\frac{-1}{2} \frac{\ln \frac{x(i, t)}{a_2}}{a_3} \right)^2 \right) \times \exp(-b_1 \times y_c(i, t)), \quad (1)$$

with a_1 the maximum growth rate, a_2 the size at maximum growth rate, a_3 the shape of the growth rate curve and b_1 the sensitivity to competition. $y_c(i, t)$ is the sum of competition kernels depending on the distance D_{ij} between the focal individual i and its neighbour j and the size of the neighbour $x(j, t)$ (Nottebrock et al., 2017):

$$y_c(i, t) = \sum_{j=1}^J x(j, t) \times \exp \left(\frac{-D_{ij}^2}{b_2^2} \right), \quad (2)$$

with b_2 the distance at which competition is 37% of maximum competition at zero distance.

The logarithm of growth increment $y_g(i, t)$ (cm/year) follows a normal distribution with mean $\mu_g(i, t)$ and standard deviation a_4 :

$$\log(y_g(i, t)) \sim N(\mu_g(i, t), a_4), \quad (3)$$

The size of individual i at the end of time step t is then:

$$x(i, t + 1) = x(i, t) + y_g(i, t), \quad (4)$$

Adult mortality

The probability $\tau(i, t)$ that a tree i dies in year t depends on the tree's current size $x(i, t)$ (Hülsmann et al., 2018):

$$\tau(i, t) = \frac{\exp(c_1 + c_2 \times \ln(x(i, t)))}{1 + \exp(c_1 + c_2 \times \ln(x(i, t)))}, \quad (5)$$

with c_1 the logit^{-1} regression intercept and c_2 the logit^{-1} regression linear coefficient. Whether individual i dies at the end of time step t follows a Bernoulli process with probability $\tau(i, t)$.

Pollen production, pollen dispersal and ovule fertilization

The model describes the relative contribution of pollen from each parent in the population and pollen immigration from outside the population. These relative pollen contributions are the basis for assigning fatherhood, but we do not consider the total number of fertilized seeds to be limited by pollen availability. The pollen production

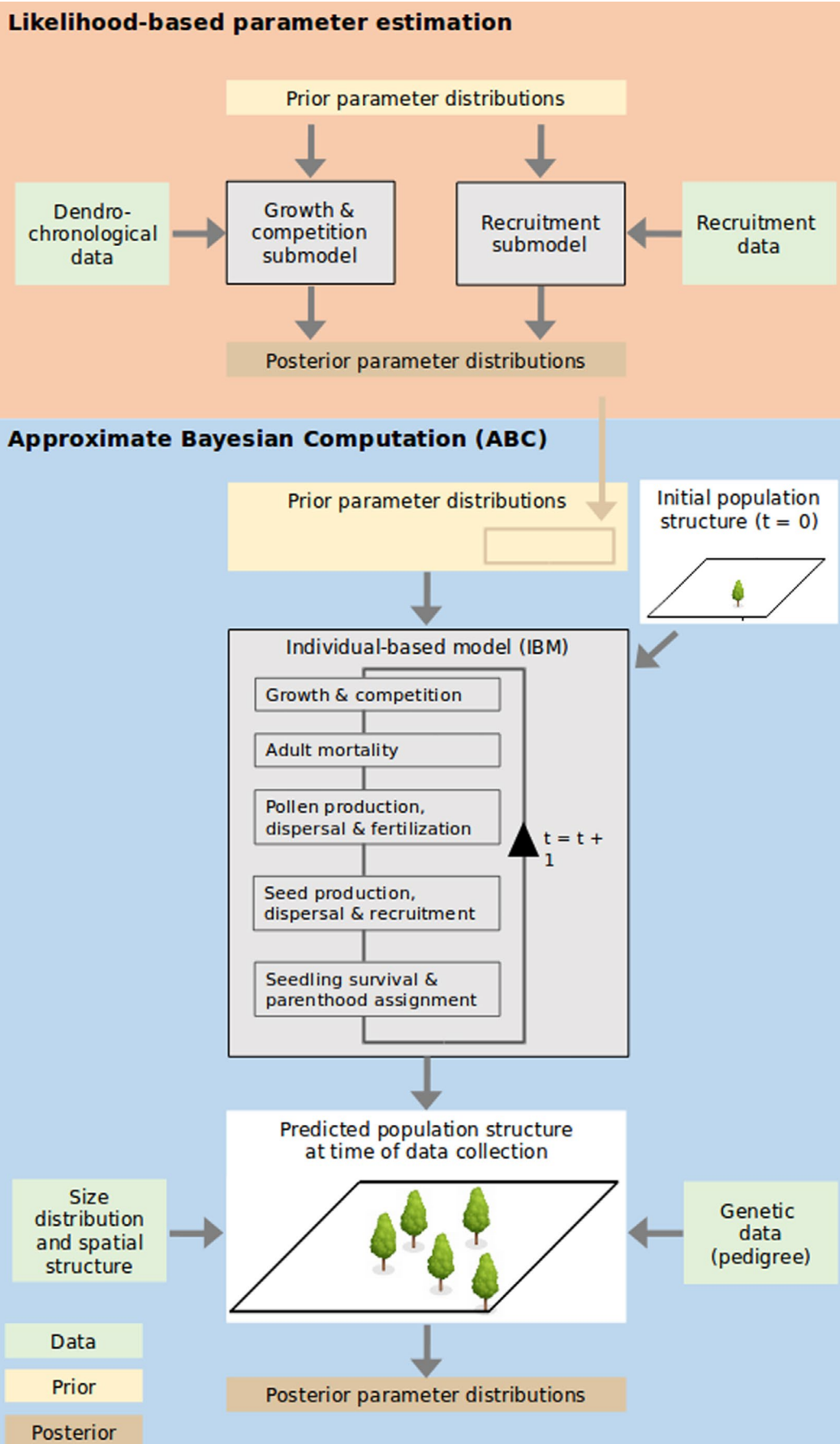


FIGURE 1 A Bayesian framework to parametrize an individual-based model (IBM) for the dynamics of establishing tree populations from data collected at a single time point. In a first step, submodels for recruitment and/or growth and competition are parametrized using likelihood-based Bayesian inference. Posterior parameter distributions of this first step then provide some of the priors for a second step in which the full IBM is parametrized with approximate Bayesian computation (ABC) methods using further data on size distribution and spatial structure as well as pedigrees. Note that the framework can either use only size distribution spatial structure data or it can combine these data with arbitrary combinations of the other data types (if dendrochronological and recruitment data are not used, the framework simplifies to classical ABC)

TABLE 2 Parameters of an individual-based model for the dynamics of establishing tree populations with reference values used for simulating data and prior distributions used for Bayesian parameter estimation

Symbol	Name	Description	Unit	Reference value	Prior distribution
a_1	Maximum growth rate	Potential maximum DBH growth rate without competitors	cm/year	3	$U(1, 5)$
a_2	Size at maximum growth	Size (DBH) at maximum growth rate without competitors	cm	12	$U(1.2, 30)$
a_3	Growth rate shape	Shape of the growth rate-DBH relationship	–	1.2	$U(0.8, 2)$
a_4	Growth rate sd	Standard deviation of log(absolute growth rate)	–	0.08	$\log_{10}(a_4) \sim U(-2, 0)$
b_1	Sensitivity to competition		–	0.02	$\log_{10}(b_1) \sim U(-2.7, -0.7)$
b_2	Spatial scale of competition	Inter-tree distance for which competition is 37% of maximum competition	m	8	$U(1, 20)$
c_1	Adult mortality intercept	Logit ⁻¹ regression intercept of adult mortality rate against size (DBH)	–	0.8	$\log_{10}(c_1) \sim U(-2.3, 0.7)$
c_2	Adult mortality slope	Logit ⁻¹ regression slope of adult mortality rate against size (DBH)	–	-3.5	$U(-8, -2)$
d_1	Size at maturity	Size (DBH) at which an individual starts producing seeds and pollen	cm	10	$\log_{10}(d_1) \sim U(0.7, 1.8)$
d_2	Allometric exponent	of fecundity-size relationship	–	1.8	$U(1, 2)$
e_1	Mean pollen dispersal distance	Mean of exponential pollen dispersal kernel	m	50	$U(20, 100)$
e_2	Mean seed dispersal distance	Mean of exponential seed dispersal kernel	m	15	$U(5, 25)$
f_1	Inverse seedling survival rate	Seed input per cell for which there is a 50% chance that a new tree establishes	Number of seed/m	700	$\log_{10}(f_1) \sim U(2, 3.7)$
g_1	Immigrant pollen contribution	Amount of immigrant pollen received by each tree (expressed as a multiple of the pollen production of a mature tree with DBH = 1 cm)	–	0.12	$\log_{10}(g_1) \sim U(-2, 0)$
g_2	Immigrant seed contribution	Mean number of immigrant seeds per grid cell	Number of seed	0.05	$\log_{10}(g_2) \sim U(-3, 0)$

Note. U stands for the uniform distribution. DBH stands for diameter at breast height. Reference values and prior distributions were chosen to yield realistic forest dynamics.

$y_p(i, t)$ of individual i at time t depends on current size $x(i, t)$ and is set to 0 if the individual is smaller than the size at maturity d_1 :

$$y_p(i, t) = \begin{cases} 0 & \text{if } x(i, t) < d_1 \\ x(i, t)^{d_2} & \text{if } x(i, t) > d_1, \end{cases} \quad (6)$$

with d_2 the allometric exponent of the fecundity-size relationship.

Ovules cannot be self-fertilized. Ovule fertilization on a tree is a function of the distance to the neighbouring trees and their respective pollen production. We use an exponential kernel (Klein

et al., 2006) $k_p(i, j)$ to model pollen dispersal from individual i to a neighbour j :

$$k_p(i, j) = \frac{1}{(2\pi e_1^2)} \times \exp\left(\frac{-D_{ij}}{e_1}\right), \quad (7)$$

with e_1 the mean pollen dispersal distance. The relative contribution $\theta(i, j, t)$ of pollen from individual i to the pollen cloud that can fertilize a seed of individual j at time t is:

$$\theta(i, j, t) = y_p(i, t) \times k_p(i, j). \quad (8)$$

Each individual receives a constant amount of immigrant pollen from outside the population g_1 .

Seed production, seed dispersal and recruitment

For each grid cell, the model describes the contribution of seeds from each parent and seed immigration from outside the population. As for pollen production, seed production $y_s(i, t)$ of individual i at time t depends on current tree size $x(i, t)$ and maturity threshold d_1 :

$$y_s(i, t) = \begin{cases} 0 & \text{if } x(i, t) < d_1 \\ x(i, t)^{d_2} & \text{if } x(i, t) > d_1, \end{cases} \quad (9)$$

The number of seeds $N(i, t)$ produced by individual i at time t follows a Poisson distribution with mean $y_s(i, t)$:

$$N(i, t) \sim P(y_s(i, t)). \quad (10)$$

We use an exponential kernel (Klein et al., 2006) $k_s(i, n)$ to model seed dispersal from individual i to grid cell n

$$k_s(i, n) = \frac{1}{(2\pi e_2^2)} \times \exp\left(\frac{-D_{i,n}}{e_2}\right), \quad (11)$$

with e_2 the mean seed dispersal distance and $D_{i,n}$ the distance between individual i and grid cell n .

The contribution of individual i to the seed pool of grid cell n is

$$\eta(i, n, t) = k_s(i, n) \times N(i, t) \times 1, \quad (12)$$

where $D_{i,n}$ is the distance between individual i and cell n . Finally, we assume that all seeds in a grid cell become seedlings (if a lower proportion of seeds became seedlings this would just alter the interpretation of seed production y_s).

Seedling survival and parenthood assignment

In each empty grid cell that currently does not hold a tree, a maximum of one sapling can establish. Lottery competition is used to determine which (if any) of the seedlings in the grid cell survives and becomes a new sapling (1 cm DBH).

The probability $\rho(i, n, t)$ that a new tree in grid cell n at time step t originates from mother tree i is

$$\rho(i, n, t) = \frac{\eta(i, n, t)}{f_1 + \sum_{j=1}^J \eta(j, n, t) + g_2}, \quad (13)$$

with f_1 the inverse seedling survival rate, i.e. the seedling number per cell for which there is a 50% chance that one seedling survives. The contribution of immigrant seeds to the seed pool is g_2 . Similarly, the probability that a new tree originates from immigrant seeds is

$\rho(n, t) = \frac{g_2}{f_1 + \sum_{j=1}^J \eta(j, n, t) + g_2}$. The probability that no seedling survives and grid cell n remains empty is

$$\rho(0, n, t) = 1 - \left(\rho(n, t) + \sum_{i=1}^I \rho(i, n, t) \right). \quad (14)$$

The mother of a new tree is drawn from a multinomial distribution with the following probabilities for each potential mother tree in the population i and for seed immigration from outside the population ($\rho(1, n, t), \dots, \rho(i, n, t), \dots, \rho(l, n, t), \rho(n, t)$). For new trees with mother tree i inside the population, the father is drawn from a multinomial distribution with the following probabilities for each potential father tree j in the population and for pollen immigration from outside the population ($\theta(1, i, t), \dots, \theta(j, i, t), \dots, \theta(J, i, t), g_1$) (after normalization of the contributions θ and g_1) (Oddou-Muratorio Davi, 2014). We assume that new trees originating from immigrant seeds have both their mother and father outside the population.

2.2 | Virtual data collection

For virtual data collection, we simulated a population with parameters set to their reference values (Table 2) (Lamonica et al., 2021). The size of the plot was 55 m by 55 m. The model was initialized with one individual of 2 cm DBH randomly located on the grid. After running the simulation for 70 years, we recorded for each individual with a DBH >3 cm the size, location on the grid, absolute growth rates in each year, as well as the father and mother trees. We also recorded the number of seedlings per grid cell.

Simulation outputs were used to generate four different data types (Table 1): (a) data on tree size distribution and spatial structure (denoted hereafter as 'Base' data), (b) genetic data determining the within-population pedigree (denoted by G), (c) recruitment data (R) and (d) dendrochronological data (D). The base data comprised the exhaustive mapping and DBH measurement of all individuals above 3 cm DBH. Genetic data comprised genotyping of each individual within the population (>3 cm DBH) and the resulting pedigree that identifies parental relationships between these trees. Dendrochronological data comprised annual absolute growth rates. Recruitment data comprised the number of seedlings in each grid cell in the current year. In Table 1, we indicate time investment, material costs and skills required to collect and analyse each data type, based on expert assessments.

To evaluate the type and amount of information provided by different combinations of data types we estimated parameters using eight data scenarios. A base scenario only included the size distribution and spatial structure and we additionally considered all possible combinations of the base data with genetic data, dendrochronological data and recruitment data. These data scenarios were denoted by the combination of letters representing the included data types.

2.3 | A Bayesian framework for IBM parametrization from different data types

The statistical framework for parameter estimation (Figure 1) combines likelihood-based Bayesian inference with ABC. For the base data scenario and scenario G, we estimated all parameters using only the ABC method based on summary statistics of the size distribution and spatial structure and/or summary statistics of the genetic data. When the data scenario included recruitment and/or dendrochronological data, we first estimated parameters related to fecundity and dispersal and/or growth with likelihood-based Bayesian methods. The posterior distributions obtained for those parameters were then used as the prior distributions for the subsequent ABC parameter estimation. As original priors we used uniform distributions that span the realistic ranges for each parameter (Table 2).

2.3.1 | Likelihood-based parameter estimation

Dendrochronological data were used to estimate posterior distributions for parameters of growth (a_1, a_2, a_3, a_4) and competition (b_1, b_2). The data comprised the growth increment $y_g(i, t)$ that were recorded for each living tree i in all years t and the sizes $x(i, t)$ (calculated as summed growth increments). The likelihood of the observed growth increment $y_g(i, t)$ was calculated from a log-normal distribution $\log(y_g(i, t)) \sim N(\mu_g(i, t), a_4)$ where parameter a_4 is the standard deviation and $\mu_g(i, t)$ is the expected growth (on a log-scale) predicted according to Equation (1).

Recruitment data were used to estimate posterior distributions for parameters of fecundity (d_1, d_2), seed dispersal (e_2) and seed immigration (g_2). The data comprised the recorded number of seedlings per grid cell n in the last simulated year Z_n . The likelihood of the observed number of seedlings Z_n was calculated from a Poisson distribution $Z_n \sim P(\sum_{i=1}^I \eta(i, n) + g_2)$ where parameter g_2 is the contribution of immigrant seeds to the seed pool and $\eta(i, n)$ follows Equation (12).

Likelihood-based parameter estimation used the `RJAGS` package (Plummer, 2009). Three independent chains with three different initial conditions were run in parallel using the `SNOW` and `DCLONE` packages (Sólymos, 2010; Tierney et al., 2016). The chains were run for 50,000 iterations to reach convergence, as verified with the Gelman Rubin (1992) convergence diagnostic (95% quantile of potential scale reduction factor below 1.02).

2.3.2 | Approximate Bayesian computation

For ABC, we sampled 100,000 parameter combinations from the prior distributions and simulated the model for each parameter combination (Lamonica et al., 2021). Subsequently, we computed summary statistics for each simulated population and compared them to the summary statistics calculated from the data. Five hundred parameter combinations were accepted and the posterior

distributions were drawn using the local linear regression of the `ABC` package (Csilléry et al., 2012). For size distribution and spatial structure (Base), we computed six summary statistics: the basal area, the standard deviation and 9th decile of size distribution, tree density, a clumping index (standard deviation of tree density distribution in 16 m² plots), the coefficient of the regression of individual size against the mean of the distances to neighbours divided by neighbour sizes. For genetic data (G), we computed eight summary statistics: the mean distance between offspring and mother trees, the mean distance between father and mother trees, the allometric exponent of the offspring number–size relationship, the percentage of local mother trees, the percentage of local father trees, the mean maternal and paternal sibship size (both computed with the `PEDANTICS` package Morrissey & Wilson, 2010) and a genetic relatedness index (non-zero F , computed with the `PEDANTICS` package).

2.4 | Evaluation

To compare data scenarios in terms of predictive accuracy, we used six variables characterizing the population state: the basal area, the population density, the standard deviation of size, the allometric exponent of the offspring number–size relationship and the proportions of local mothers and fathers. Using the posterior distributions for each data scenario, as well as the true parameter values and the prior distributions, we computed these variables after 70 years of model simulation. We also compared data scenarios in terms of the accuracy of forecasts into the future. To this end, we simulated population dynamics for 40 additional years (from 70 to 110 years, starting with the ‘true’ population structure). We then computed the above-mentioned six variables using the posterior distributions for each data scenario, the true parameter values and the prior distributions. As measures of the information content of each data scenario, we then calculated the standard deviation of forecast state variables and their bias compared to the ‘true’ future population state. To assess how much information each data scenario contains on individual parameters, we also calculated the standard deviation of posterior parameter distributions and their bias (compared to true parameter values).

3 | RESULTS

3.1 | Parameter estimation

The combination of all data types (DGR) leads to precise and accurate estimates of most model parameters (Figure 2, Supporting Information Table A1). The exception is parameters describing adult mortality (c_1 and c_2) and mean pollen dispersal distance (e_1), for which posterior distributions were rather wide. The posterior of the inverse seedling survival rate (f_1) was narrow but biased. The base data by themselves provided little information on model parameters,

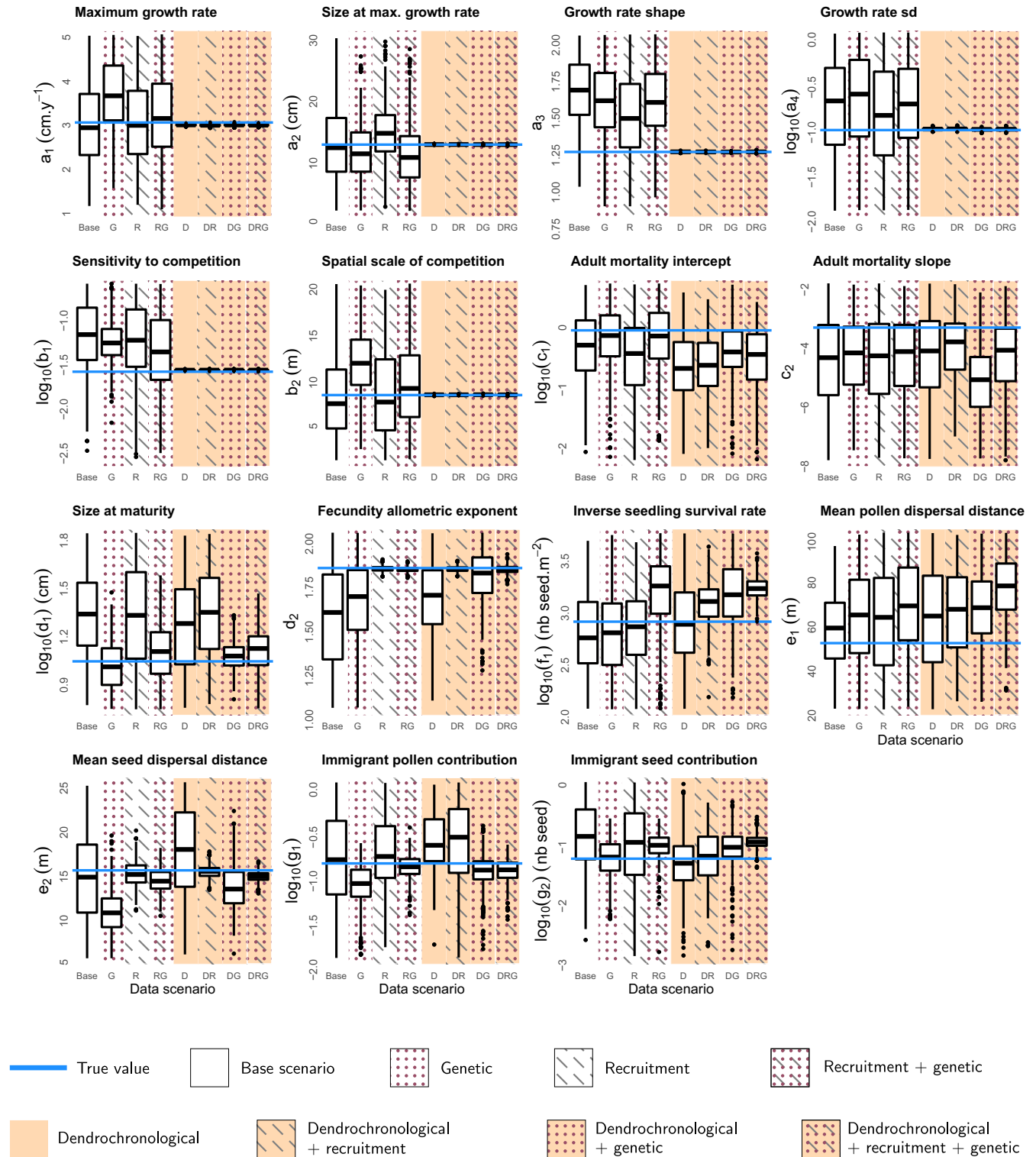


FIGURE 2 Posterior distributions of model parameters obtained under different data scenarios (Base: only size distribution and spatial structure, G: base plus genetic, R: base plus recruitment, RG: base plus recruitment plus genetic, D: base plus dendrochronological, DR: base plus dendrochronological plus recruitment, DG: base plus dendrochronological, plus genetic, DRG: all data). The blue lines represent the 'true' values of each parameter that were used for data simulation. See Table 1 for parameter definitions

leading to posteriors that were just slightly narrower than the prior distributions.

With genetic data (G), we obtained more precise estimates for size at maturity d_1 and immigration parameters (g_1 and g_2) than the

base scenario. Additionally, uncertainties were slightly reduced for growth parameter a_2 and for competition b_1 . Dendrochronological data (D) lead to very precise and accurate estimates for all growth and competition parameters (a_1 , a_2 , a_3 , a_4 and b_1 , b_2). Uncertainties

were also reduced for pollen and seed immigration (g_1 and g_2). Recruitment data (R) only informed on seed fecundity (d_2) and seed dispersal (e_2).

Combining genetic and dendrochronological data (DG) improved the estimation of parameters related to immigration, and to seed fecundity and dispersal. The combination of genetic and recruitment data (RG) reduced uncertainties on immigration parameters. Combining recruitment and dendrochronological data (scenario DR) did not improve estimates (except for the inverse seedling survival rate f_1).

3.2 | Forecasts of future dynamics

Forecasts of future population state are one way to summarize the information content of data scenarios across all model parameters. All data scenarios yielded largely unbiased forecasts of population density (Figures 3 and 4, Supporting Information Table A2) as well as the size and genetic structure of populations 40 years into the future (Figure 3). Forecasts of density (number of trees and basal area) and size structure (size distribution and offspring size relationship) were least uncertain when dendrochronological data were used for model parametrization (Figures 3 and 4). Forecasts of the proportion of local mothers and fathers became more certain when using genetic data in combination with other data. However, forecasts of the proportion of local mothers were also improved by using only dendrochronological data. In contrast, recruitment data helped little to reduce forecast uncertainty (Figures 3 and 4). We found similar results for reconstructions of past population dynamics (from 0 to 70 years, Supporting Information Figures A1 and A2). Overall, forecasts of future dynamics showed less uncertainty than predicted past dynamics, presumably because forecast simulations were initialized with the exact same starting point (the 'true' population structure in year 70) and were only run over 40 years.

4 | DISCUSSION

In this study, we developed a framework for parametrizing an IBM for establishing tree populations from different data types by combining ABC and likelihood-based estimation methods (Figure 1). When comparing the value of different data types for model parametrization, we found that dendrochronological data were particularly informative although genetic and recruitment data also improved estimates of certain parameters (Figure 2). Reliable forecasts of future population structure and density could be obtained from dendrochronological and genetic data, whereas recruitment data contributed little to improve forecasts (Figures 3 and 4). This holds not only for population states at the time horizon of forecasts (Figure 3) but also for the temporal dynamics up to this time point (Figure 4).

Data collected at a single time point can thus be remarkably informative about the past and future dynamics of tree populations. This is because (a) the considered data contain information about

past population dynamics (albeit being collected at a single time point) and (b) these data inform about the entire history of the considered establishing populations. Under these conditions, most model parameters could be estimated with little bias and variance (Figure 2). However, even when combining all data types we did not obtain reliable estimates of adult mortality and seedling survival (Figure 2). This is because these parameters partly compensate each other in their effect on the summary statistics used for model parametrization, causing moderately strong correlations between parameter estimates (Supporting Information Figure A3). Thus, different parameter combinations lead to the same population state, for example, a high adult mortality combined with a high seedling survival rate would lead to similar size distribution and spatial structure as lower adult mortality and lower seedling survival rate. Another limit to inference occurs for mean pollen dispersal distance e_1 . This parameter could not be estimated reliably because the data were not informative—only few father trees from the population produced offspring—and the true value (50 m) is close to the side length of the study site (55 m). Moreover, the sensitivity of summary statistics to this parameter was quite low.

We obtained reliable forecasts of population dynamics even though the marginal posterior distributions of adult mortality and seedling survival parameters were wide (Figures 3 and 4, Supporting Information Figures A1 and A2). This is due to the above-mentioned correlation of parameter estimates. It may not be necessary to resolve this trivariate parameter uncertainty if one is interested in forecasting population dynamics rather than learning about specific parameters (Sirén et al., 2018). However, if one needs precise estimates of individual parameters, certain data types are more valuable than others. In our study, for instance, dendrochronological data lead to good forecasts, but provided almost no information on fecundity parameters, whereas recruitment data did.

When designing data collection to infer the dynamics of establishing tree populations, dendrochronological data are the first choice (in addition to data on tree size distribution and spatial structure) given its high information content and the relatively low costs in terms of time, money and skills (Table 1). Combining either recruitment or genetic data with dendrochronological data informed fecundity, dispersal, recruitment and/or immigration processes, thus reducing further uncertainty. This combination of data types leads to better forecasts of population structure because different data types informed on different processes. Yet when focusing on a single parameter related to a given process, combining data types did not lead to a better estimate than using only the data type related to the process.

Our analysis permits to steer future data collection depending on (a) data costs and available skills, (b) processes and parameters of interest and (c) available prior information and knowledge on the study system. For instance, if the studied population is known to be very isolated or if the seed dispersal distance is unknown, it may be most informative to complement dendrochronological with recruitment data. On the contrary, if parenthood relationships are of major interest, it may not be necessary to collect recruitment data in addition to genetic data because the latter will provide enough information on parenthood.

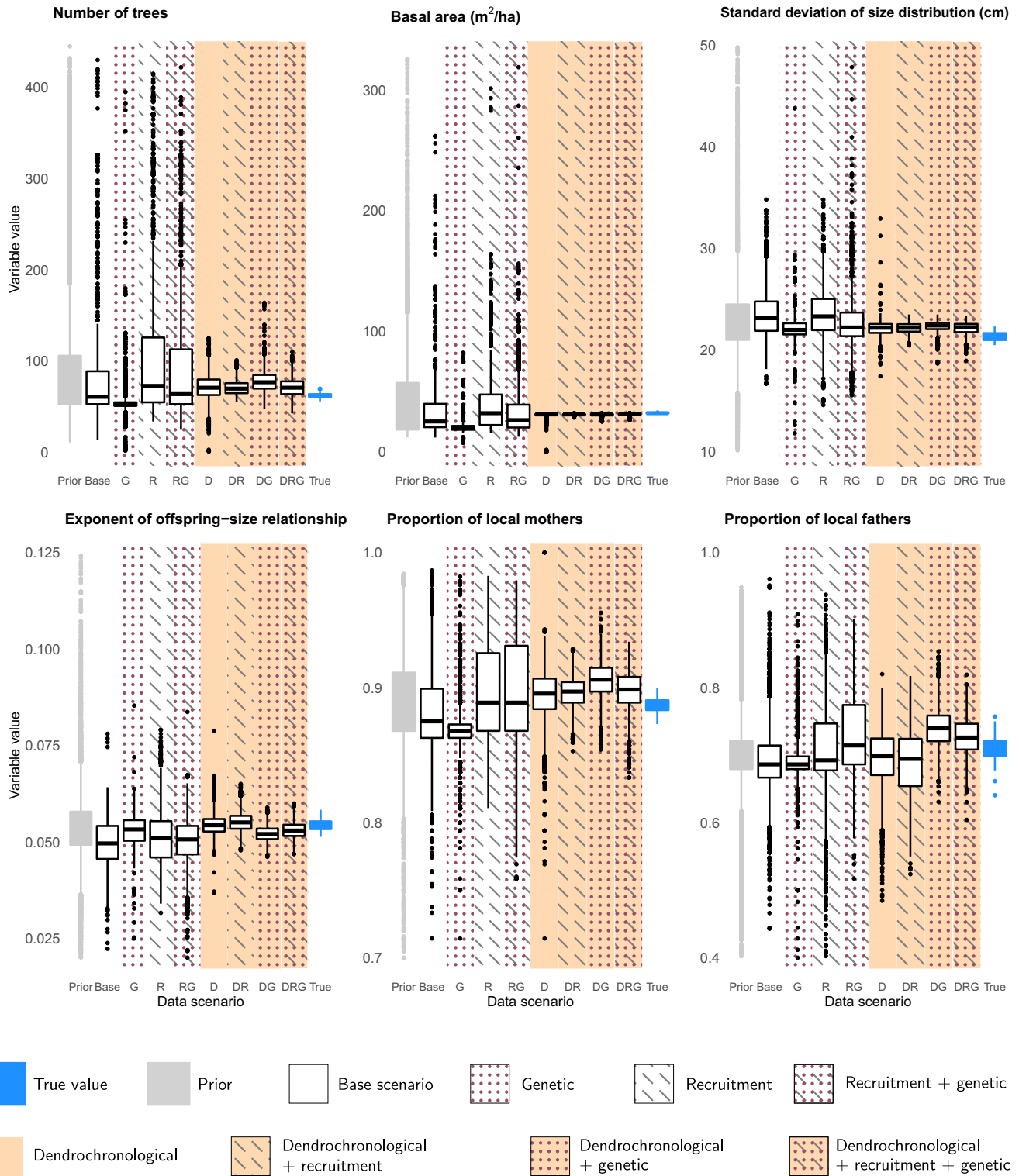


FIGURE 3 Forecasts of the state of tree populations 40 years into the future. Forecasts of eight population characteristics were generated by an IBM parametrized with different data scenarios (see Figure 2), from the prior distribution of parameters (grey) and with the ‘true’ parameter values used for data simulation (blue). Each boxplot represents variation in forecasts for 500 replicate simulations of the stochastic simulation model

We considered a ‘best-case scenario’ of data-driven modelling, in which (a) each data type is sampled exhaustively, (b) there are no measurement errors and (c) the ‘true’ model is fitted to the data. We chose this best-case approach since it enabled us to quantify the

potential value of each data type. Exhaustive sampling of genetic data is necessary since it may allow to retrieve the complete pedigree (Jones & Arden, 2003). The ability to reconstruct parentage and accommodate genotyping errors depends on genetic data resolution

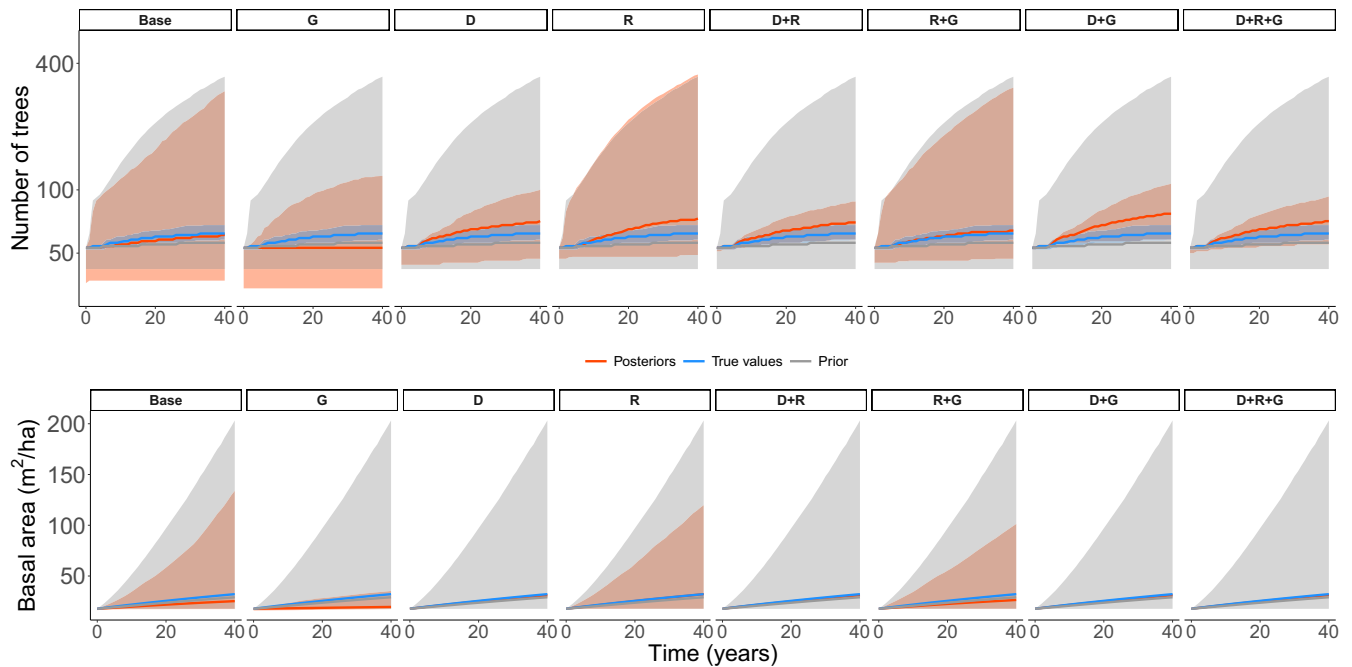


FIGURE 4 Forecasts of tree density (top) and basal area (bottom) over a time horizon of 40 years. Columns represent different data scenarios (see Figure 2). Lines represent medians and areas the 95% credibility intervals of forecasts obtained for prior distributions (light grey), posterior distributions under the given data scenario (orange) and 'true' parameter values used for data simulation (blue)

(number of single nucleotide polymorphisms or multi-allelic microsatellites) and on the method used to reconstruct parentage (exclusion, relatedness-based or likelihood-based methods) (Huisman, 2017). Genomics is a rapidly moving field, both in terms of resolution and data analysis, so that one can expect the quality of reconstructed pedigrees to get closer to the best-case scenario considered here. Exhaustive sampling of dendrochronological data is also necessary since it allows to infer past competition neighbourhoods (Lamonica et al., 2020). Importantly, the exhaustive sampling of genetic and dendrochronological data is feasible for populations comprising several hundreds of individuals (Gerzabek et al., 2020; Lamonica et al., 2020). Concerning recruitment data, censusing the entire patch area might be too time-consuming, so that one would rather sample a fraction of the patch. Yet it is possible to obtain reliable estimates of spatial recruitment parameters by sampling < 1% of the patch area (Schurr et al., 2008).

The estimation of IBM parameters with summary statistics-based ABC methods can lack efficiency (Hartig et al., 2014). Here, we found that the two-step combination of direct likelihood-based estimation and ABC provided very informative posterior distributions of certain parameters. Specifically, likelihood-based methods provided posterior distributions for submodels, which were then used as priors for ABC estimation of the full IBM. We obtained very good estimates of growth and competition parameters because dendrochronological data directly result from the growth-competition process. In contrast, other data types result from the combination of several processes, which increases uncertainty of individual parameter estimates but means that these data inform about a wider range of parameters (for instance, genetic data provide modest information about growth, competition, adult mortality and fecundity).

5 | CONCLUSIONS

Inferring past dynamics and parametrizing predictive models from data collected at a single time point is a prerequisite to forecast the population dynamics of long-lived organisms. This is because time is lacking to monitor long-term dynamics of these organisms, especially under global change. In the case of trees, dendrochronological and genetic data make it possible to reconstruct not only individual growth trajectories but also past population dynamics. While tree rings are specific to trees in climates with annual growth cycles, various other sessile organisms have morphological structures that can serve the same purpose, for instance annual growth rings in herb roots (Dietz & Ullmann, 1998), coral growth rings (Marschal et al., 2004) and annual stem growth increments in certain shrubs (Carlson et al., 2011). Annual growth increments can also be inferred from the otoliths of fish (Rountrey et al., 2014). With annual growth data, one can reconstruct individual growth trajectories that can help to infer population dynamics. The presented framework for data-driven modelling and forecasting of tree population dynamics could thus be modified to describe the population dynamics of other ecologically important long-lived sessile organisms (whereas application to mobile organisms such as fish would require major modifications in submodels for competition, reproduction and dispersal). Moreover, it will be exciting to extend the framework beyond population dynamics to the modelling of community and range dynamics (Evans et al., 2016; Pagel et al., 2020; Schurr et al., 2012), and the forecasting of invasion dynamics and species range expansions (Hastings & Wysham, 2010; Nathan et al., 2011).

AUTHORS' CONTRIBUTIONS

D.L., J.P. and F.M.S. conceived the ideas and designed methodology; D.L. analysed the data and led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/2041-210X.13656>.

[Correction added on 18 Aug 2021, after first online publication: Peer review history statement has been added.]

DATA AVAILABILITY STATEMENT

All data can be found in the Zenodo data repository (Lamonica et al., 2021) and R code for IBM simulation and parameter estimation is publicly available at the following address https://gitlab.com/lamonica_/data-predicting-the-dynamics-of-establishing-tree-populations-a-framework-for-statistical-inference-and-lessons-for-data-collection.

ORCID

Dominique Lamonica  <https://orcid.org/0000-0002-0054-0272>

Jörn Pagel  <https://orcid.org/0000-0002-4817-1920>

Frank M. Schurr  <https://orcid.org/0000-0002-4518-0602>

REFERENCES

- Bastin, J. F., Finegold, Y., Garcia, C., Mollicone, D., Rezende, M., Routh, D., Zohner, C. M., & Crowther, T. W. (2019). The global tree restoration potential. *Science*, 365(6448), 76–79. <https://doi.org/10.1126/science.aay8060>
- Beaumont, M. A. (2018). Approximate Bayesian computation. *Annual Review of Statistics and Its Application*, 6(1), 379–403. <https://doi.org/10.1146/annurev-statistics-030718-105212>
- Canham, C. D., Lepage, P. T., & Coates, K. D. (2004). A neighborhood analysis of canopy tree competition: Effects of shading versus crowding. *Canadian Journal of Forest Research*, 34, 778–787. <https://doi.org/10.1139/x03-232>
- Carlson, J. E., Holsinger, K. E., & Prunier, R. (2011). Plant responses to climate in the cape floristic region of South Africa: Evidence for adaptive differentiation in the proteaceae. *Evolution*, 65(1), 108–124. <https://doi.org/10.1111/j.1558-5646.2010.01131.x>
- Chornesky, E. A., Bartuska, A. M., Aplet, G. H., Britton, K. O., Cummings-Carlson, J., Davis, F. W., Eskow, J., Gordon, D. R., Gottschalk, K. W., Haack, R. A., Hansen, A. J., Mack, R. N., Rahel, F. J., Shannon, M. A., Wainger, L. A., & Wigley, T. B. (2006). Science priorities for reducing the threat of invasive species to sustainable forestry. *BioScience*, 55(4), 335. [https://doi.org/10.1641/0006-3568\(2005\)055\[0335:spfrrt\]2.0.co;2](https://doi.org/10.1641/0006-3568(2005)055[0335:spfrrt]2.0.co;2)
- Clark, J. S., Carpenter, S. R., Barber, M., Collins, S., Dobson, A., Foley, J. A., Lodge, D. M., Pascual, M., Pielke Jr, R., Pizer, W., Pringle, C., Reid, W. V., Rose, K. A., Sala, O., Schlesinger, W. H., Wall, D. H., & Wear, D. (2001). Ecological forecasts: An emerging imperative. *Science*, 293(5530), 657–660. <https://doi.org/10.1126/science.293.5530.657>
- Clark, J. S., & Gelfand, A. E. (2006). A future for models and data in environmental science. *Trends in Ecology & Evolution*, 21(7), 375–380. <https://doi.org/10.1016/j.tree.2006.03.016>
- Clark, J., Lewis, M., & Horvath, L. (2001). Invasion by extremes: Population spread with variation in dispersal and reproduction. *The American Naturalist*, 157(5), 537–554. <https://doi.org/10.1086/319934>
- Csilléry, K., Blum, M. G. B., Gaggiotti, O. E., & François, O. (2010). Approximate Bayesian computation (ABC) in practice. *Trends in Ecology & Evolution*, 25(7), 410–418. <https://doi.org/10.1016/j.tree.2010.04.001>
- Csilléry, K., François, O., & Blum, M. G. B. (2012). abc: An R package for approximate bayesian computation (ABC). *Methods in Ecology and Evolution*, 3(3), 475–479. <https://doi.org/10.1111/j.2041-210X.2011.00179.x>
- Dietz, H., & Ullmann, I. (1998). Ecological application of 'herbchronology': Comparative stand age structure analyses of the invasive plant *Bunias orientalis* L. *Annals of Botany*, 82(4), 471–480. <https://doi.org/10.1006/anbo.1998.0706>
- Evans, M. E. K., Merow, C., Record, S., McMahon, S. M., & Enquist, B. J. (2016). Towards process-based range modeling of many species. *Trends in Ecology & Evolution*, 31(11), 860–871. <https://doi.org/10.1016/j.tree.2016.08.005>
- Frost, G. V., & Epstein, H. E. (2014). Tall shrub and tree expansion in Siberian tundra ecotones since the 1960s. *Global Change Biology*, 20(4), 1264–1277. <https://doi.org/10.1111/gcb.12406>
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–511. <https://www.jstor.org/stable/2246093>
- Gerzabek, G., Oddou-Muratorio, S., & Hampe, A. (2020). Recruitment of a genotyped *Quercus robur* L. seedling cohort in an expanding oak forest stand: Diversity, dispersal, and performance across habitats. *Annals of Forest Science*, 77, 78. <https://doi.org/10.1007/s13595-020-00979-5>
- Gopalaswamy, A. M., Royle, J. A., Delampady, M., Nichols, J. D., Karanth, K. U., & Macdonald, D. W. (2012). Density estimation in tiger populations: Combining information for strong inference. *Ecology*, 93(7), 1741–1751. <https://doi.org/10.1890/11-2110.1>
- Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., Thulke, H.-H., Weiner, J., Wiegand, T., & DeAngelis, D. L. (2005). Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science*, 310(5750), 987–991. <https://doi.org/10.1126/science.1116681>
- Hartig, F., Calabrese, J. M., Reineking, B., Wiegand, T., & Huth, A. (2011). Statistical inference for stochastic simulation models – Theory and application. *Ecology Letters*, 14(8), 816–827. <https://doi.org/10.1111/j.1461-0248.2011.01640.x>
- Hartig, F., Dislich, C., Wiegand, T., & Huth, A. (2014). Technical note: Approximate bayesian parameterization of a process tropical forest model. *Biogeosciences*, 11, 1261–1272. <https://doi.org/10.5194/bg-11-1261-2014>
- Hastings, A., & Wysham, D. B. (2010). Regime shifts in ecological systems can occur with no warning. *Ecology Letters*, 13(4), 464–472. <https://doi.org/10.1111/j.1461-0248.2010.01439.x>
- Huisman, J. (2017). Pedigree reconstruction from SNP data: Parentage assignment, sibship clustering and beyond. *Molecular Ecology Resources*, 17(5), 1009–1024. <https://doi.org/10.1111/1755-0998.12665>
- Hülsmann, L., Bugmann, H., Cailleret, M., & Brang, P. (2018). How to kill a tree: Empirical mortality models for 18 species and their performance in a dynamic forest model. *Ecological Applications*, 28(2), 522–540. <https://doi.org/10.1002/eap.1668>
- Jones, A. G., & Ardren, W. R. (2003). Methods of parentage analysis in natural populations. *Molecular Ecology*, 12(10), 2511–2523. <https://doi.org/10.1046/j.1365-294X.2003.01928.x>
- Klein, E. K., Lavigne, C., & Gouyon, P. H. (2006). Mixing of propagules from discrete sources at long distance: Comparing a dispersal tail to an exponential. *BMC Ecology*, 6(3), 1–12. <https://doi.org/10.1186/1472-6785-6-3>
- Klein, E. K., Carpentier, F. H., & Oddou-Muratorio, S. (2011). Estimating the variance of male fecundity from genotypes of progeny arrays: Evaluation of the Bayesian forward approach. *Methods in Ecology and Evolution*, 2(4), 349–361. <https://doi.org/10.1111/j.2041-210X.2010.00085.x>

- Klein, E. K., Bontemps, A., & Oddou-Muratorio, S. (2013). Seed dispersal kernels estimated from genotypes of established seedlings: Does density-dependent mortality matter? *Methods in Ecology and Evolution*, 4(11), 1059–1069. <https://doi.org/10.1111/2041-210X.12110>
- Lagarrigues, G., Jabot, F., Lafond, V., & Courbaud, B. (2015). Approximate Bayesian computation to recalibrate individual-based models with population data: Illustration with a forest simulation model. *Ecological Modelling*, 306, 278–286. <https://doi.org/10.1016/j.ecolmodel.2014.09.023>
- Lamonica, D., Herbach, U., Orias, F., Clément, B., Charles, S., & Lopes, C. (2016). Mechanistic modelling of daphnid-algae dynamics within a laboratory microcosm. *Ecological Modelling*, 320, 213–230. <https://doi.org/10.1016/j.ecolmodel.2015.09.020>
- Lamonica, D., Pagel, J., Valdés-Correcher, E., Bert, D., Arndt, A., & Schurr, F. M. (2020). Tree potential growth varies more than competition among spontaneously established forest stands of pedunculate oak (*Quercus robur*). *Annals of Forest Science*, 77, 80. <https://doi.org/10.1007/s13595-020-00981-x>
- Lamonica, D., Pagel, J., & Schurr, F. M. (2021). Data from: Predicting the dynamics of establishing tree populations a framework for statistical inference and lessons for data collection. *Zenodo*, <https://doi.org/10.5281/zenodo.4773205>
- Lira, P. K., Tambosi, L. R., Ewers, R. M., & Metzger, J. P. (2012). Land-use and land-cover change in Atlantic Forest landscapes. *Forest Ecology and Management*, 278, 80–89. <https://doi.org/10.1016/j.foreco.2012.05.008>
- Marschal, C., Garrabou, J., Harmelin, J. G., & Pichon, M. (2004). A new method for measuring growth and age in the precious red coral *Corallium rubrum* (L.). *Coral Reefs*, 23(3), 423–432. <https://doi.org/10.1007/s00338-004-0398-6>
- Moran, E. V., & Clark, J. S. (2011). Estimating seed and pollen movement in a monoecious plant: A hierarchical Bayesian approach integrating genetic and ecological data. *Molecular Ecology*, 20(6), 1248–1262. <https://doi.org/10.1111/j.1365-294X.2011.05019.x>
- Morrissey, M. B., & Wilson, A. J. (2010). Pedantics: An R package for pedigree-based genetic simulation and pedigree manipulation, characterization and viewing. *Molecular Ecology Resources*, 10(4), 711–719. <https://doi.org/10.1111/j.1755-0998.2009.02817.x>
- Nathan, R., Horvitz, N., He, Y., Kuparinen, A., Schurr, F. M., & Katul, G. G. (2011). Spread of North American wind-dispersed trees in future environments. *Ecology Letters*, 14(3), 211–219. <https://doi.org/10.1111/j.1461-0248.2010.01573.x>
- Nottebrock, H., Schmid, B., Treurnicht, M., Pagel, J., Esler, K. J., Böhning-Gaese, K., Schleuning, M., & Schurr, F. M. (2017). Coexistence of plant species in a biodiversity hotspot is stabilized by competition but not by seed predation. *Oikos*, 126, 276–284. <https://doi.org/10.1111/oik.03438>
- Oddou-Muratorio, S., & Davi, H. (2014). Simulating local adaptation to climate of forest trees with a Physio-Demo-Genetics model. *Evolutionary Applications*, 7(4), 453–467. <https://doi.org/10.1111/eva.12143>
- Pagel, J., Treurnicht, M., Bond, W. J., Kraaij, T., Nottebrock, H., Schutte-Vlok, A. L., Tonnabel, J., Esler, K. J., & Schurr, F. M. (2020). Mismatches between demographic niches and geographic distributions are strongest in poorly dispersed and highly persistent plant species. *Proceedings of the National Academy of Sciences of the United States of America*, 117(7), 3663–3669. <https://doi.org/10.1073/pnas.1908684117>
- Palmero-Iniesta, M., Espelta, J. M., Gordillo, J., & Pino, J. (2020). Changes in forest landscape patterns resulting from recent afforestation in Europe (1990–2012): Defragmentation of pre-existing forest versus new patch proliferation. *Annals of Forest Science*, 77, 43. <https://doi.org/10.1007/s13595-020-00946-0>
- Plummer, M. (2009). *rjags: Bayesian graphical models using mcmc*. R package version 1.0.3-12.
- R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Rountrey, A. N., Coulson, P. G., Meeuwig, J. J., & Meekan, M. (2014). Water temperature and fish growth: Otoliths predict growth patterns of a marine fish in a changing climate. *Global Change Biology*, 20(8), 2450–2458. <https://doi.org/10.1111/gcb.12617>
- Ruiz-Carbayo, H., Pino, J., Bonal, R., James, P. M. A., Hampe, A., Molowny-Horas, R., & Espelta, J. M. (2020). Insect herbivory in novel *Quercus ilex* L. forests: The role of landscape attributes, forest composition and host traits. *Annals of Forest Science*, 77, 32. <https://doi.org/10.1007/s13595-020-00934-4>
- Schurr, F. M., Steinitz, O., & Nathan, R. (2008). Plant fecundity and seed dispersal in spatially heterogeneous environments: Models, mechanisms and estimation. *Journal of Ecology*, 96(4), 628–641. <https://doi.org/10.1111/j.1365-2745.2007.0>
- Schurr, F. M., Pagel, J., Cabral, J. S., Groeneveld, J., Bykova, O., O'Hara, R. B., Hartig, F., Kissling, W. D., Linder, H. P., Midgley, G. F., Schröder, B., Singer, A., & Zimmermann, N. E. (2012). How to understand species' niches and range dynamics: A demographic research agenda for biogeography. *Journal of Biogeography*, 39(12), 2146–2162. <https://doi.org/10.1111/j.1365-2699.2012.02737.x>
- Sirén, J., Lens, L., Cousseau, L., & Ovaskainen, O. (2018). Assessing the dynamics of natural populations by fitting individual-based models with approximate Bayesian computation. *Methods in Ecology and Evolution*, 9(5), 1286–1295. <https://doi.org/10.1111/2041-210X.12964>
- Sólymos, P. (2010). dclone: Data cloning in R. *The R Journal*, 2(2), 29–37.
- Tierney, L., Rossini, A. J., Li, N., & Sevcikova, H. (2016). *snow: Simple network of workstations*. R package version 0.4-2.
- Uriarte, M., Canham, C. D., Thompson, J., & Zimmerman, J. K. (2004). A neighborhood analysis of tree growth and survival in a hurricane-driven tropical forest. *Ecological Monographs*, 74(4), 591–614. <https://doi.org/10.1890/03-4031>
- Uriarte, M., Canham, C. D., Thompson, J., Zimmerman, J. K., & Brokaw, N. (2005). Seedling recruitment in a hurricane-driven tropical forest: Light limitation, density-dependence and the spatial distribution of parent trees. *Journal of Ecology*, 93(2), 291–304. <https://doi.org/10.1111/j.0022-0477.2005.00984.x>
- Zurell, D., Berger, U., Cabral, J. S., Jeltsch, F., Meynard, C. N., Münkemüller, T., Nehrbass, N., Pagel, J., Reineking, B., Schröder, B., & Grimm, V. (2010). The virtual ecologist approach: Simulating data and observers. *Oikos*, 119(4), 622–635. <https://doi.org/10.1111/j.1600-0706.2009.18284.x>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Lamonica, D., Pagel, J., & Schurr, F. M. (2021). Predicting the dynamics of establishing tree populations: A framework for statistical inference and lessons for data collection. *Methods in Ecology and Evolution*, 12, 1721–1733. <https://doi.org/10.1111/2041-210X.13656>