**Understanding invasions with linked distribution and demographic models**

**Matthew E. Aiello-Lammens**

Pace University

Environmental Studies and Science Department

Pleasantville

New York

USA

maiellolammens@pace.edu

ORCID ID: 0000-0002-6180-5959

**Abstract**

The temporal and spatial patterns of a species invasion are determined by life history characteristics, biological and environmental conditions of the novel range, and human activities. Untangling the contributions of these factors based on current ecology and distributional knowledge for individual invasive species has proven difficult. A new analytical approach to this problem is to use linked distribution and demographic models and the past spatio-temporal patterns of an invasion to test alternative hypotheses.  Recent advances in modeling species ranges and dynamics allow directly linking observations of species occurrences, life history characteristics and environmental gradients, increasing our ability to analyze and predict the processes that shape patterns of biogeography. Application of these methods is increasing in conservation biology, but underutilized in invasion biology. To demonstrate the utility of these methods, I developed a linked demographic and species distribution model for the invasive species *Frangula alnus*. Using a global sensitivity analysis approach, I created simulations representing an extensive parameter uncertainty space, and identified regions in this space that best explain historic occurrence patterns for this species. Investigations of parameter uncertainty space facilitate invasion hypothesis testing, as parameter values can be associated with specific hypotheses *a priori*. In the case of *F. alnus,* simulations parameterized with moderate fecundity, high survival, and extensive long-distance dispersal values best-predicted historic occurrences. Strong interaction effects between fecundity and long-distance dispersal indicate that a successful invasion required both of these processes, which occur at different spatial scales. Lastly, long-distance dispersal was most likely facilitated by human transport. Results from linked models improved our understanding of the roles various population processes played in shaping patterns of *F. alnus* biogeography in its naturalized range. Wider application of these methods will further our understanding of species invasions more generally.

**Keywords**

*Frangula alnus*; linked models; demographic model; species distribution model; invasion dynamics

**Introduction**

Understanding the processes associated with successful establishment and spread of novel species is a major goal of invasion ecology. Over the past 50 years, numerous theories explaining why some species are invasive and other are not have been proposed, tested, and refined (Lowry et al. 2012). Fundamentally the processes governing dynamics of invasive species are the same as those governing native species (Gurevitch et al. 2011). As such, using the well-established principles and frameworks of ecological modeling, properly parameterized models should be able to reproduce patterns of occurrence associated with the spread of an invasive species. However, for such models to be successful, they must incorporate both local and regional ecological dynamics (Pyšek and Hulme 2005).

Demographic models link a mathematical representation of the life-cycle of an individual organism to a population of that organism (Caswell 2006), providing a way to examine population dynamics and forecast population growth and/or decline. A relatively large amount of data is required to properly parameterize these models, limiting most studies to a particular geographic focus or research question (e.g. examining density dependence effects in a particular forest area), or to informing management decisions involving one or few populations (e.g., Davis et al. 2006; Harris et al. 2009). But these models can also be used to “predict” historical species occurrence patterns, allowing us to investigate the demographic processes that likely lead to observed patterns. Species distribution models (SDMs) provide a way of estimating habitat suitability for a species by establishing a statistical relationship between observed species occurrences and the environmental conditions associated with those occurrence locations (Peterson et al. 2011). Integrating demographic models with models of landscape characteristics (e.g., SDMs) results in spatially informed simulations that incorporate species demographic processes such as growth, survival, and reproduction with spatial process that include dispersal between populations, as well as population colonization and extinction (Akçakaya 2000, 2001; Akçakaya et al. 2004; Franklin 2010). Further, dynamic habitat models can be used to incorporate changes to a landscape through time (Urban et al. 2007; Keith et al. 2008; Brook et al. 2009; Aiello-Lammens et al. 2011; Fordham et al. 2012, 2013; Franklin et al. 2013; Pearson et al. 2014). Thus, these models integrate local and regional processes governing species abundance and occurrence patterns over time. While these models are seeing increased usage in the field of conservation (Briscoe et al. 2019), they are underutilized in the study of invasive species.

Comparisons between the results of ecological models and empirical observations can provide insights into processes of population dynamics (Hastings et al. 2005). The processes governing the spatial spread and population increases of non-native species occur over multiple spatial and temporal scales (Sakai et al. 2001; Theoharides and Dukes 2007; Blackburn et al. 2011), and thus the utility of analysis tools that allow for investigation across scales cannot be overstated. While most studies using integrated demographic and distribution modeling primarily focus on forecasting future patterns, retrospective analyses offers an important tool for testing hypotheses of species spread. Such analyses have already been applied to understand causes of species declines and extinctions (Prowse et al. 2013; Stanton 2014). Theoretical and modeling work has also addressed cross-scale investigations. For example, there has been extensive investigations of the influence of propagule pressure on invasion success (Simberloff 2009), leading some to suggest that mathematical parameterizations of propagule pressure could serve as a null model of invasion (Colautti et al. 2006). Several cross-scale mechanisms have been proposed, and in some cases simulated, to explain temporal lags between stages of invasion, incorporating a mix of population ecology, biogeography, and evolution (Kowarik 1995; Crooks and Soulé 1999; Sakai et al. 2001; Pyšek and Hulme 2005; Theoharides and Dukes 2007). Similarly, in simulation work, (With 2002, 2004) demonstrated that landscape level characteristics, such as distance between suitable patches and habitat fragmentation, strongly influence species invasions. Integrated modeling methods provide a way to test these theoretical predictions with empirical data.

In this study, I constructed a linked demographic and species distribution model for *Frangula alnus* to examine the demographic processes resulting in areal growth matching its pattern of spread through the 20th century. Analysis of *F. alnus* occurrence records suggest an extended lag from its first observation in 1879 to approximately 1910 to 1920, followed by relatively rapid expansion throughout northeastern North America (Aiello-Lammens 2014). I used linked demographic and distribution models to test multiple hypotheses regarding this invasion. First, that life-history characteristics, including high fecundity and survival, were integral to the expansion of *F. alnus.* Second, that 20th century land-use change, particularly conversion from intensive agriculture to old-field and forest habitat in northeastern United States, was positively associated with *F. alnus* range expansion. Third, that long-distance dispersal events were an important contributor to its rapid spread. I estimated model parameters using demographic data collected in field surveys from two geographic locations in northeast North America over three years, as well as data from peer-reviewed and grey literature. I compared patterns of the simulated spatial spread for *F. alnus* to patterns of observed occurrences through time. Exploring the parameter uncertainty space for this model and examining deviations of the simulated spread from observed occurrences allowed for both testing of my initial hypotheses and development of additional explanations. While this model focuses specifically on the dynamics of *F. alnus,* it provides a framework to apply to other invasive species.

**Methods**

**Study species and field observations**

*Frangula alnus* is a woody plant native to Eurasia. It grows to approximately 8 m in height and produces fleshy fruit, each containing two to three seeds, with individual plants able to produce several hundred fruit per year (Godwin 1943; Medan 1994). Seeds are dispersed by birds (Howell and Blackwell 1977; Hampe et al. 2003; Hampe 2008), small mammals (Godwin 1936), and hydrochory (Hampe 2004). *F. alnus* spread rapidly throughout North America during the 20th century (Howell and Blackwell 1977; Catling and Porebski 1994; Frappier et al. 2003a), where its ability to dominate understory and canopy layers negatively impacts native flora (Possessky et al. 2000; Frappier et al. 2003a; Fagan and Peart 2004).

I collected demographic data for *F. alnus* in two geographic locations in the invaded range – Long Island, New York and coastal New Hampshire, from the summer of 2010 until the Fall of 2012. Data were collected at three different sites per location. Sites were separated by 1 to 10 km. Upland, old field, and wetland habitat types were represented in the sites. Within each site I establish 15 or 20 randomly located 2 x 2 meter plots in which *F. alnus* plants were measured and tagged for multi-year monitoring. Measurements included diameter at basal (ankle) height (DAH) in centimeters (cm), height (cm), number of basal stems, and number of fruit on the plant. For plants with more than one basal stem, the DAH of the largest stem was measured. All measurements were made between September and November each year. In total, I measured and monitored 815 plants. I marked an additional 385 seedlings in 2011 and re-surveyed them in 2012 to estimate seedling survival values. All measurements were used to parameterize a demographic model of *F. alnus*. Additional field-sampling details can be found in Online Resource – Appendix 1.

**Demographic model**

I constructed an integral projection model (IPM) to simulate the demographic processes of *F. alnus* (Easterling et al. 2000; Ellner and Rees 2006; Merow et al. 2014). At the core of an IPM is a function describing the transition of the population distribution from time *t* to *t +* 1, given a continuous structuring variable:

(1)

where *k(y,x)* is the IPM kernel representing all possible transitions of individuals from size *x* to size *y*. The kernel is analogous to the projection matrix in matrix projection models (MPM) (Easterling et al. 2000). Assuming the structuring variable is size, integrating the kernel multiplied by the size distribution at year *t* (i.e., *n(x, t)*) over all possible sizes, *Ω*, results in the size distribution at year *t* + 1, (i.e., *n(y, t* + 1*)*).

The IPM kernel is composed of sub-kernels representing different demographic processes (e.g, survival, growth, and fecundity). Using the field observations, I estimated the relationship between size and annual survival via a logistic regression and the relationship between size and growth rate via a linear regression. Combined, these yield the survival-growth sub-kernel. I calculated the fecundity sub-kernel using ANCOVA to determine the relationship between the number of fruit produced versus plant size and a measure of population density. Population density can greatly affect demographic processes (e.g., Comita and Hubbell 2009; Pardini et al. 2009; Dwyer et al. 2010; Ramula and Buckley 2010), but these effects are often ignored in plant demographic models (Menges 2000; Crone et al. 2010). I estimated population density for each plant based on the total number of plants within its 2 x 2 meter plot that were in a similar size class or larger. I define this density estimate as the *effective density.* The fecundity sub-kernel also includes parameters for the number of fruit per seed, seed germination, and establishment rates, which I estimated using information from published literature and field observations. Additional details on the demographic model construction, including further details on the effective density approach, can be found in Online Resource – Appendix 2.

Taking advantage of the similarities between IPMs and MPMs (Easterling et al. 2000; Merow et al. 2014), I used RAMAS Metapop (Akçakaya 2002), an MPM simulation program, to approximate the integral of the IPM kernel. This is identical to using a mid-point approximation rule generally applied in IPMs (Merow et al. 2014).

**Habitat suitability through time**

I used the Maxent species distribution modeling (SDM) software to estimate habitat suitability of *F. alnus* in space and time throughout its North American range (Phillips et al. 2006; Phillips and Dudík 2008; Elith et al. 2010). I compiled 2929 occurrence locations for *F. alnus* using published literature, publicly accessible biodiversity and invasive species databases, and regional herbarium records. For environmental predictor variables, I chose a set of temporally dynamic climate and land-use variables and static environmental variables based on *a priori* assumptions of how environmental conditions affect vital rates (Table 1) and minimizing highly correlated variables.

To optimize Maxent model tuning parameters (i.e., regularization and feature selection), I constructed multiple SDMs for a 50-year timeframe from 1960 to 2010. I calculated 50-year average values for all dynamic climate and land-use variables and included the static variables CTI, pH Top, Crop, Grass, and Potential Veg, resulting in a total of 11 predictor variables (9 continuous and 2 categorical). I examined the influence of increasing model complexity (i.e., adding additional features) and regularization (i.e., increasing the penalty on complex models) on model testing values area under the receiver operating curve (AUC) and omission rate at maximum specificity and sensitivity values using 5-fold cross-validation. Balancing these two metrics, I chose final model parameters of a regularization multiplier equal to 1 (i.e., default regularization) and Linear + Quadratic + Product features. Further details are noted in the Online Resource – Appendix 3.

To incorporate changes to the spatial structure of the *F. alnus* metapopulation through time, I constructed an SDM with 10-year aggregated layers for all dynamic predictor variables. For each occurrence record I extracted the values from the corresponding 10-year aggregate layers, which included the ten years leading up to, and including, the year the occurrence record was recorded. Additionally, I extracted random background points, proportional to the number of occurrences represented by each year. Occurrence locations were thinned within years such that only one occurrence per 5 x 5 arc minute grid cell was used. Values for static predictor variables (i.e., potential vegetation and topsoil pH) were also extracted for both occurrence and background points. Model tuning parameters were identical to the static model described above and the resulting habitat suitability layers were projected to a Lambert Equal Area projection with grain size of 20 x 20 km.

**Linking population demography and species distribution models**

I linked the demographic model with the SDM results to construct a metapopulation model using RAMAS GIS. In this model each 20 x 20 km grid was considered a potential habitat patch if it had a logistic output value greater than the 10% omission rate value across all modeled years. I assumed patch carrying capacity was proportional to the logistic output projected for a given year. To examine the sensitivity of my model outcomes to this assumption I tested three different carrying capacity scenarios, as described below. Lastly, to examine the role of land-use change in the spread of *F. alnus* I constructed a static metapopulation structure assuming no change in patch carrying capacity after 1910 and compared results from these simulations to those using a dynamic metapopulation structure.

I incorporated dispersal by first calculating the distance between patches as the distance between centers (i.e., 20 km for adjacent cells). I then parameterized a dispersal distance function using reported dispersal rates for a similar invasive species, *Celastrus orbiculatus* (Merow et al. 2011), and data for *F. alnus* reported by Berg (2011) and Hampe (2004, 2008). Though there is little information on *F. alnus* dispersal rates, existing data suggest that dispersal by avian frugivores in limited to a local region. However, as with many invasive species, there can be random long-distance dispersal (LDD) events, potentially facilitated by humans. I incorporated such events in my simulations in two ways. First, I assumed random LDD allowing for equal probability of translocation of plants to any patch on the landscape. Second, I assumed weighted random LDD, in which the probability of a patch being the target of translocation was weighted by the human population density within that patch. The latter scenario assumes that humans are the primary vectors for LDD.

Estimates of patch carrying capacity are necessary to incorporate density dependence processes. The effects of density dependence were calculated at a spatial scale of 2 x 2 m (see Online Resources – Appendix 2 for details) and therefore carrying capacity was measured as the number of 2 x 2 m cells within the 20 x 20 km patch that have suitable *F. alnus* habitat. I assumed that the maximum carrying capacity of a patch was the total number of 2 x 2 m cells contained by a 20 x 20 km cell (i.e., 100,000,000 cells). This value was multiplied by the SDM logistic output value of the patch to yield the total number of cells with suitable *F. alnus* habitat. The logistic output of SDMs can adequately estimate population carrying capacity (VanDerWal et al. 2009). To test how sensitive my model was to this assumption, I treated these values as a high carrying capacity scenario, and created medium and low scenarios by multiplying this number by 0.5 and 0.25, respectively. Related estimates for carrying capacity were calculated for an alternative ceiling type density dependence model by multiplying the number of cells deemed suitable in a patch by 40, which is approximately the number of individuals observed in the most dense 2 x 2 m plots in my field observations.

**Exploring parameter space via global sensitivity analysis**

I applied a global sensitivity analysis (GSA) approach to examine parameter uncertainty space and find regions of this space resulting in simulations closely matching the patterns of historical occurrences (Aiello-Lammens and Resit Akçakaya 2017). Additionally, I examined the relative influence of input parameter variation on measures of simulation fit. I generated 500 random parameter sets comprised of eight continuous and one categorical input parameters (Table 2). Parameter bounds were estimated via analysis of collected field data, estimates culled from published literature, and descriptions of *F. alnus* ecology. Additional details on these estimates are provided in the Online Resources – Appendix 4. To examine the impact of two structural model changes, the functional description of density dependence and changes in land-use through time, I created four simulations for each of the 500 parameter sets representing all model structure combinations: ceiling density dependence + land-use change, ceiling density dependence + no land-use change, effective density dependence + land-use change, and effective density dependence + no land-use change. In total, 2000 matched simulation models were constructed. Each simulation was run for 100 years (1910 to 2010) with ten replications, which were used to estimate within simulation variability. In addition to these 2000 simulations, I constructed and ran 1000 additional simulations, which focused on comparing the two LDD scenarios. For these simulations, 500 new parameter sets were sampled and two simulations each were run, both using effective density dependence + land-use change, but differing on the LDD scenario applied. Thus, these two sets of 500 simulations were matched to each other, but not to the other 2000 simulations.

Model evaluation metrics, described below, were calculated for each simulation. Additionally, changes in model endpoints due to structural differences in the models were calculated based on pair-wise comparisons - i.e., comparisons among the four simulations with matched input parameter sets *sensu* (Aiello-Lammens and Resit Akçakaya 2017). I used boosted regression tree analysis to calculate the relative influence of varying input parameters on model evaluation metrics and to establish univariate and bivariate response curves of the relationships between input parameter values and model evaluation metrics. Boosted regression tree analysis is a flexible machine learning analysis method that performs well at fitting complex response curves with mixed types of predictor variables (Elith et al. 2008). Previous sensitivity analyses (Prowse et al. 2013; Coutts and Yokomizo 2013; Aiello-Lammens and Resit Akçakaya 2017) have demonstrated this methods utility in this context. I examined the response curves and relative influence values to determine the parameter values that best matched the historic occurrence patterns for *F. alnus.*

I used a confusion matrix approach (Fielding and Bell 1997) to evaluate how well simulations matched the historic pattern of spread of *F. alnus* (as described in Aiello-Lammens 2014). Using historic occurrence records as true presences, I calculated positive predictive power and sensitivity for each time step in the model simulations, resulting in 100 values of each per simulation. For model evaluation purposes, I calculated the mean of these measures, yielding a mean sensitivity and mean positive predictive power value for each simulation. I calculated an additional model fit metric that combined sensitivity and positive predictive power. In the best model fit scenarios, both sensitivity and positive predictive power would be close to one. However, it is entirely possible to achieve sensitivity values close to 1 with simulations that predict *F. alnus* occurrence in nearly all patches. In this case, the positive predictive power would be very small (i.e., ). Thus, I defined a new binary model fit metric based on combined sensitivity and positive predictive power values, and called this the combined metric. The combined metric took a value of 1 if mean sensitivity was greater than or equal to 0.5 *and* the difference between mean sensitivity and mean positive predictive power was less than or equal to 0.1, and 0 otherwise. This metric balances sensitivity with over prediction.

A major factor that can effect model evaluation measures is the likelihood that a patch that is occupied by *F. alnus* is identified as such. With simulation data, it is possible to be certain that a patch is occupied or not (i.e., patch population size is greater than zero); however, such perfect knowledge is not realistic in the field. Species occurrence information gathered from natural history collections (e.g., herbariums) are subject to collection biases that may mis- or under-represent a species range (Graham et al. 2004; Anderson 2012; Lavoie 2012). The detection and collection of invasive species in particular may lag behind its establishment in a novel region (Crooks 2005). This can have a major effect on whether or not historical data adequately represents *F. alnus* occurrence through time. For example, *F. alnus* occurrence may not have been observed because plant density was not high enough for detection. To account for this I used three occupancy thresholds, in which a patch was considered occupied if it had a population size of at least 1, 1000, and 2000 individuals. The values were chosen to represent perfect observation, to match the initial population sizes used in simulations, and to represent a density of approximately one reproductive plant per square kilometer in a 20 x 20 km patch, respectively.

**Results**

*Demographic model and IPM kernel*

The IPM kernel was comprised of three regression models. The logistic regression of plant survival onto size resulted in fitted curves of and for high and low survival respectively (high – P = 0.152, R2 = 0.08, df = 910; low – P < 0.001, R2 = 0.12, df = 910). The low survival calculations treated plants that were not found in subsequent surveys as mortality events (Fig. 1a), whereas high survival calculations censored these observations (Online Resources Fig. A4). Survival values were calculated as: , with the high and low estimates as bounds in the GSA. The linear regression of growth onto size resulted in regression coefficients of *b0* = 0.041 and *b1* = 1.035 (P < 0.001, R2 = 0.984, df = 856) (Fig. 1b). I calculated the probability density of an individual of size *x* transitioning to size *y* as: , where and = the standard deviation of the residuals of the linear regression fit. For the fecundity kernel, the number of fruit produced declined as an exponential function of effective density: , where is the number of fruit produced with no density effect, is a parameter governing the rate of exponential decline, is analogous to the population carrying capacity, and is the effective density for the size class being modeled. I estimated and using results from the ANCOVA model (Fig. 1c, Table 3), in which slopes and intercepts were allowed to vary freely. For each size class I set as the exponent of the corresponding intercept value and as the slope for that size class in the ANCOVA model. was set *a priori* based on the population size and environmental conditions, as described above. was calculated at each time step during a run of a demographic model.

Combining these sub-kernels yielded the full model as:

(3)

, where *i* represents the size classes (A to D; Online Resources Table A3).

To solve this integral, I divided the range of DAH sizes, 0.032 to 7.337 cm, into 50 equal sized bins. This range included all observed values (0.04 to 6.67 cm). The mid-point for each bin was determined and a two-dimensional mesh grid of these values was constructed. Column values were treated as plant size at time *t* and row values as plant size at time *t* + 1. The first row of this grid was set as the seedling size at time *t* + 1. At each mesh point, equation 3 was evaluated. This process resulted in a projection matrix with 50 stages (Fig. 1d), which I used as the transition matrix in the RAMAS Metapop program. The population growth rate (λ) for this matrix was 1.04.

*Habitat suitability through time*

Examining SDM predictor variable importance, I found that extreme cold temperatures along the northern margin and high summer temperatures along the southern margin limit the range of *F. alnus*. Soil pH (topsoil) was consistently ranked as one of the most important predictors of *F. alnus* occurrence, suggesting that it plays a substantial role in shaping the species distribution. Presence of pasture was also consistently an important predictor, with relatively small amounts of pasture within a grid cell associated with low values of suitability. However, from a whole range perspective, very few grid cells had high values of percent cover of pasture. Both low and high values of growing degree days (GDD) were associated with low habitat suitability. Low GDD values may be associated with limited growth during short growing seasons. On the other hand, the low suitability associated with high GDD may be indicative of the lack of competitive ability of *F. alnus* with species better adapted to areas with long growing seasons. Projecting these relationships onto the combined dynamic and static layers resulted in a spatial structure with the locations of high habitat suitability changing through time (Fig. 2). Regions that experienced increases in land conversion to cropland appear to account for the greatest declines in suitability. Overall declines in habitat suitability resulted in a general decrease in metapopulation carrying capacity. In total, 3423 grid cells were considered suitable at least once between 1910 and 2010 based on the 10% omission threshold. Each of these grid cells was considered a potential patch that could be occupied by *F. alnus* in model simulations.

*Exploration of parameter space*

Randomly generated parameter sets resulted in good sampling coverage of input parameter uncertainty space (Online Resources Fig. A10), with population growth rates varying between 0.92 and 1.17. Assuming an occupancy threshold value of 1000 individuals, differences in model structure (land-use change versus no land-use change and ceiling density dependence versus effective density dependence) had no significant effects on measures of simulation sensitivity or positive predictive power (paired t-tests; Online Resources Table A8). Additionally, expected minimum population sizes were not significantly different. However, simulations parameterized with effective plot density versus ceiling type density dependence resulted in greater final metapopulation abundance values (t = -6.483, P < 0.001, df = 499), as did simulations with no land-use change versus land-use change (t = -2.813, P < 0.01, df = 499). Similar results were found assuming an occupancy threshold of 1 and 2000 individuals. Given the lack of difference in the key measures used to compare simulations to patterns of historical occurrence records, the subsequent results presented are for simulations considering land-use change and effective plot density, as these model structures are ecologically more realistic than the alternatives.

Simulations parameterized with weighted random LDD had significantly higher sensitivity values when compared with paired simulations parameterized with complete random LDD (t = 27.032, df = 499, P << 0.001). There were no significant differences in positive predictive power assuming an occupancy threshold of 1000. However, for an occupancy threshold of 1, the weighted LDD simulations returned significantly higher positive predictive power values compared to the random LDD simulations (t = 21.940, df = 499, P << 0.001). These results suggest that simulations incorporating LDD weighted by human population density provide better fits to historic occurrence patterns than those using complete random LDD. Subsequent results represent simulations using weighted LDD, in addition to land-use change and effective density.

*Simulation sensitivity values* ***–*** Sensitivity values varied widely through time both within and among simulations (Fig. 3). They were also greatly affected by occupancy threshold values; sensitivities were higher for the occupancy threshold of 1 individual compared to thresholds of 1000 and 2000 individuals (Fig. 3). The latter two thresholds showed no significant differences between them. It is notable that model sensitivity declined immediately after simulations begin, and for occupancy threshold values of 1000 and 2000, most show poor predictive ability (low sensitivity) during the middle years of the simulation (Fig. 3). These results indicate an inability of both the local and long-distance dispersal parameterizations to adequately track the historic spread of *F. alnus*. While model fit metrics calculated assuming an occupancy threshold of 1 yielded relatively high sensitivity throughout the duration of simulations (Fig. 3), they also had high values for the number of patches predicted as occupied (Fig. 4). This results in lower positive predictive power compared to the occupancy threshold of 1000 and 2000 measures. Considering only the occupancy threshold of 1000, there was a clear tradeoff between sensitivity and positive predictive power (Fig. 5). Results were similar using thresholds of 1 and 2000.

*Input parameters best explaining historical patterns of F. alnus occurrence -* The binary combined metric was calculated on results using each of the three threshold occupancy values. Considering the occupancy threshold of 1000 individuals, 94 of the 500 simulation models yielded a value of 1 (Fig. 5).Results of the boosted regression tree analysis showed that fecundity had the highest relative influence (32.2%), followed by number of LLD events (21.5%), and metapopulation initial abundance (14.2%) on distinguishing between a combined metric value of 0 or 1. The remaining parameters, with the exception of carrying capacity, all had influence values between approximately 5% and 7%. Examining the relationships between input parameter values and the prediction of the combined metric (logit (p)) suggests that simulations with moderate to high values for both fecundity and LDD values best predict historic occurrence patterns (mean fecundity of 9.25 (2.36 SD) and mean LDD of 291 (116 SD); (Figs. 6 & 7). However, extremely high values of both parameters do not result in simulations with good measures of fit. Simulations yielding good model fit include those with high survival-growth values. Examination of interaction plots (Fig. 8) confirms these findings and shows the strong interactions between fecundity and LDD values, as well as moderate interactions between fecundity and metapopulation initial abundance and fecundity and variability (standard deviation) of fecundity. Additionally, there were moderate interactions between LDD and metapopulation initial abundance. Moderate to high values of fecundity yield simulations with good model fit, provided these simulation also include moderate to high LDD values.

**Discussion**

I hypothesized that life-history characteristics, including high fecundity and survival, were integral to the expansion of *F. alnus*, that 20th century land-use change was positively associated with *F. alnus* range expansion, and that long-distance dispersal events were an important contributor to its rapid spread. The invasion of *F. alnus* in North American was replicated with simulations parameterized with moderate fecundity values, high survival values, and extensive long-distance dispersal, providing mixed support for each of these hypotheses. Using the combined sensitivity and positive predictive power fit metric, I conclude that *F. alnus* had relatively high fecundity during its expansion with a mean population growth rate of λ = 1.13 (0.01 SD). This is higher than the mean for 179 native species (λ = 1.05) reported in a review of matrix projection models, but lower than the mean of 21 invasive species (λ = 1.47) (Ramula et al. 2008). A relatively large number of LDD events was also necessary to match patterns of historic *F. alnus* spread – a mean of 291 (116 SD) dispersal events during its range expansion. The LDD events were likely associated with human facilitated transport. Interactions between fecundity and LDD also played a role in this invasion. *F. alnus* had *both* moderate to high fecundity and a moderate to high LDD; neither process alone predicts historic occurrence patterns well.

Simulations that best predicted historic occurrences had moderate to high mean fecundity values (9.25 [2.36 SD] recruits per reproductive plant) and moderate to high LDD values (286 [118 SD] dispersal events). Higher mean survival and moderate to high local dispersal values were also associated with better model fits. These trait values are all consistent with our expectations of invasive species. Interactions between parameters also influenced how well simulations predicted historic occurrences. While moderate to high fecundity values were associated with better model fit values, this was the case only in combination with moderate to high LDD. There was an interesting interaction between fecundity and metapopulation initial abundance, with lower values of initial abundance resulting in better model fit. This may be the result of the penalty applied in the combined metric to simulations with low positive predictive power, which is the result of over prediction of *F. alnus* occurrence. Further, it is likely that population sizes were relatively small at the beginning of this invasion. A similar relationship between metapopulation initial abundance and LDD was also observed. Finally, there was evidence for an interaction between mean fecundity and the variability (i.e., standard deviation) of fecundity, with higher values of the latter yielding better model fit. At the local scale, individual plants show high year-to-year variability in reproductive output (personal observations and (Medan 1994), and this may emerge at the patch level as well. Given the high probability of survival for established *F. alnus* plants, high variability in fecundity may indicate that occasional pulses in reproductive output can facilitate spread, but declines in reproductive output do not hinder population persistence.

Incorporating land-use changes associated with farming and livestock production (Klein Goldewijk et al. 2010) resulted in declines in habitat suitability throughout the western and southern parts of the study region over the 20th century. These changes caused declines in total metapopulation carrying capacity, but had no effect on how well simulations predicted observed *F. alnus* occurrences, suggesting that land-use change did not contribute to the spread of this species. This is surprising given the general role landscape disturbances are considered to have on invasions (Davis et al. 2000; Cadotte and Lovett-Doust 2001). In fact, disturbance has been linked to the success of some invasive species in the study region (e.g., Foster and Gross 1999; Searcy et al. 2006; McDonald et al. 2008; Mosher et al. 2009). However, in these cases it was not the conversion of natural lands to agriculture that were associated with invasions, but rather subsequent transitions from agriculture to natural lands, such as old-fields.

One explanation for the lack of an effect of land-use change is that few occurrence records came from areas showing the greatest decline in habitat suitability. Therefore, these changes would not affect the predictive ability of the simulations. An alternative explanation is that disturbance may facilitate *F. alnus* spread at a local spatial scale, but not directly affect spread at the regional, or continental, scale. It readily establishes in undisturbed wetlands and fens (Mills et al. 2009, 2012; Berg 2011) and intact upland forests (Frappier et al. 2003b; Fagan and Peart 2004), but also shows rapid population growth following disturbance, such as logging (Burnham and Lee 2009; Lee and Thompson 2012). At the spatial grain size of my simulations (20 km2), most disturbances are likely to manifest at a sub-patch scale, with each patch likely to contain areas of both undisturbed and disturbed habitat. Thus, while disturbance may have an effect on the patch population size (i.e., patches with more disturbance will have more individuals), it will have a smaller effect on whether a patch is successfully invaded, and subsequently classified as occupied. For species for which dispersal to new patches is strongly dependent on propagule pressure, increases in population size may lead to increases in spatial spread. However, for species that experience long-distance dispersal not dependent on population size, the influence of disturbance is lessened (With 2004). Through its associations with humans as long-distance dispersal vectors, *F. alnus* is in the latter category.

In sum, two important points emerge. First, the demographic characteristics of *F. alnus*, including population growth,appear intermediate between those reported for native and invasive plants (Ramula et al. 2008; Knight et al. 2011). Mean fecundity values are indicative of lower seed per plant per year values than those of both woody invasive and native species reported in (Mason et al. 2008). However, recruitment can be high in some habitats, resulting in *F. alnus* outpacing recruitment of native species (Medan 1994; Lee and Thompson 2012; Mills et al. 2012). Another characteristic likely contributing to the success of *F. alnus* is its high survival rate, particularly for plants beyond the seedling stage. Survival did not strongly influence predictive ability of my simulations, but this is likely because there was little variability in this parameter. The combination of high survival and the potential for high recruitment both contribute to the ability of *F. alnus* to persist in patches it invades, which is consistent with processes observed in the field (Cunard and Lee 2008).

These characteristics however, do not necessarily result in a rapidly spreading species. Thus, the second point is that *F. alnus* required extensive long-distance dispersal to spread throughout its novel range. Both local and long-distance dispersal influence the spatial spread of invasive species (Pyšek and Hulme 2005) and can be critical to attaining good model fits in grid-based simulations (Merow et al. 2011). LDD in particular has been shown to be important for predicting species spread (Clark et al. 2001; Hastings et al. 2005; Nehrbass et al. 2006). In fact, in an analysis of simulation results with occupancy threshold equal to 1, LDD was the most influential parameter on measures of sensitivity (results not shown). However, measures of *F. alnus* dispersal distance are very short compared to the patch size used in this simulation (on the order of tens to hundreds of meters; (Frappier et al. 2003b; Hampe 2004, 2008; Berg 2011). Further, though birds consume fruit and disperse seeds via defecation (Godwin 1943; Hampe and Bairlein 2000; Hampe 2008), because of laxative properties of the fruit flesh, avian seed dispersal is also limited in distance. This suggests that human facilitated long-distance dispersal played a major role in the invasion of *F. alnus.* Additionally, regardless of the occupancy threshold applied, human population density weighted LDD showed significantly better model fit metrics compared to random LDD. This is further supported by pattern of spread of *F. alnus*, rapid spread from New York to metropolitan areas of Chicago, Boston, Washington D.C., and London, Ontario in the earliest stages of it invasion, followed by slower spread out from those urban centers. Ultimately, spatial spread for *F. alnus* was likely driven by human assisted transport, as is the case for many invasive species (Mack et al. 2000; Herron et al. 2007; Jongejans et al. 2008; Gavier-Pizarro et al. 2010; Larkin 2011; Sullivan et al. 2012).

I made several assumptions during model construction that could have affected the simulation results and subsequent interpretations. First, the IPM framework used for the demographic model does not readily incorporate the effects of demographic stochasticity (Ellner and Rees 2007; Rees and Ellner 2009). Rather, populations are assumed to be large enough such that demographic stochasticity does not greatly affect outcomes. This may be unrealistic for the purposes of modeling spreading species that are establishing new populations. Demographic stochasticity has the greatest impact on small populations, and during spread of an invasive species, it is likely that new introductions will be small in population size. Ignoring these effects, my simulations may overestimate the success of very small populations, potentially overestimating the rate of spread throughout the landscape. Some work has been done to better integrate demographic stochasticity into IPMs (Vindenes et al. 2011), but further development is needed.

Second, the parameters used in the SDM affected the spatial structure of the metapopulation. A MaxEnt model constructed with a smaller number of features (i.e., a simpler model) yielded a greater number of populations than a more complicated model, but with overall lower habitat suitability values. I chose the simpler model to avoid potential effects of over-fitting (Syfert et al. 2013; Merow et al. 2013). The decrease in habitat suitability resulted in lower patch carrying capacities. However, variations in carrying capacity were consistently the least influential parameter in simulation models, so this likely had minimal effect on model fit measures, though it may have influenced final metapopulation abundances.

Third, the method I used to calculate carrying capacity (the number of 2 x 2 m cells within the patch) is a novel approach that scales up the effects of demographic processes playing out between individual plants and their neighbors to the population (patch; 20 x 20 km). In doing this I assumed that all 2 x 2 m grid cells are equal in the population. That is, the quality of 2 x 2 m grid cells is uniform. At a microsite level, this is certainly an over-simplification. However, I estimated density dependent effects on fecundity using field observations from 2 x 2 m sample plots that varied in habitat quality, and should therefore represent processes in the average 2 x 2 m grid cell. Second, I assumed that the number of grid cells available in a 20 x 20 km patch was proportional to habitat suitability. The logistic output from MaxEnt is an adequate measure of habitat suitability if the prevalence of the species on the landscape is approximately 50%. This was not the case when I compared the total number of grid cells with habitat suitability values greater than the 10% omission threshold versus the total number of grid cells with known occurrences (3423 versus 459, respectively). However, considering only patches were either *F. alnus* or native species with similar habitat requirements were observed (974 patches), the prevalence was approximately 47%.

Lastly, the application of global sensitivity analysis provided a means to examine the influence of most model assumptions. For example, the GSA demonstrated that simulation results were not highly influenced by carrying capacity values. Fecundity, on the other hand, was very influential. Fecundity was also one of the most difficult parameter to estimate. *F. alnus* remains a relatively understudied species, with many unanswered questions regarding life-history strategies and characteristics. One that could substantially impact my model results is the degree to which seeds germinate during the summer and fall in which they are produced, rather than after the following winter. My field observations were not well setup to answer this question, and future research should address this. Most studies of seed germination found that germination in green house conditions was very high, while germination for seeds left out overwinter was very low (Adams 1927; Godwin 1936; Berg 2011). None of these studies examined germination in field conditions for fruit that drop early in the growing season. A *F. alnus* plant can have flowers, unripe fruit, and ripe fruit simultaneously, thus it seems plausible that fruit could drop early enough in the growing season to emerge during that season, provided cold stratification is not necessary for emergence. If within season seedling emergence is common, my model parameterization will have underestimated fecundity values. In this case, simulations that fit the historic occurrence pattern of *F. alnus* spread and have high mean fecundity rates may best represent the demographic properties of this species.

Another process not include in my calculations of fecundity was seed bank dynamics. *F. alnus* seeds remain viable for upwards of three years (Godwin 1943; Granstrom 1988) and there is some evidence for a role for a seed bank after experimental removal (Frappier et al. 2004). However others have found no evidence for a role of seed banks in the population dynamics of this species (Kostel-Hughes et al. 1998; Hampe 2004). If seed bank dynamics do play a role in *F. alnus* demography, I may have overestimated fecundity by overestimating germination rates. This would suggest that simulations with moderate levels of fecundity best represent the demography of *F. alnus*, which would further support the role of moderate to high LDD events, assuming that the combined sensitivity and positive predictive power metric is more informative of simulation fit. Future field studies should address the role of seed bank dynamics.

**Conclusions**

I used an integrated demographic and species distribution model to simulate the spread of invasive *F. alnus* throughout its novel North American range. Comparing patterns of spatial spread resulting from 3000 simulations to historic occurrence records I concluded that the success of *F. alnus* can largely be attributed to its relatively high fecundity and the fact that it experienced extensive long-distance dispersal, mostly likely the result of humans. However, under most ecological conditions *F. alnus* does not demonstrate extraordinary population growth, as other invasive plants do (e.g., kudzu, multiflora rose, Japanese barberry). Ultimately, the factors that lead to the successful spread of this species are its ability to survive in varied habitats and have occasional periods of high reproductive output, combined with human proclivity to transport it to new parts of the novel range. Given the prevalence of non-native species that have been introduced via the horticulture industry, it is likely that these dynamics, and the patterns they produce, are indicative of what we will see with continued monitoring and investigation of invasive plants.

The integrated modeling methods I used here are increasingly being applied in conservation related studies, but are less common among invasion biology studies. They provide a means of connecting demographic characteristics and population level processes to species range patterns. In this study I found that interactions between population level processes influence how species spread and establish in novel regions. Such knowledge is important if we want to understand what makes an invasion successful and how to avoid facilitating such invasions in the future. However there are also important conservations lessons to be learned in these studies. As environments continue to be influenced by human impacts, resulting in novel ecological conditions where species can, and in some cases must, spread, it is vital that we learn how interactions among processes affect species spread. As shown in this study, these methods are a promising way to achieve this goal.

**Acknowledgments**

I thank C. H. Graham, J. Gurevitch, T. M. Knight and H. R. Akçakaya for helpful comments on earlier versions of this manuscript. M.A.-L. C. Mattrick helped me to establish field sites in the White Mountains National Forest of New Hampshire. T. Lee and S. Eisenhaure at the University of New Hampshire helped me establish field sites in the Durham, NH region. W. Bohach helped me obtain permits to establish field sites at Caleb Smith State Park in New York.

**References**

Adams J (1927) The germination of the seeds of some plants with fleshy fruits. American Journal of Botany 415–428

Aiello-Lammens ME (2014) Patterns and Processes of the Invasion of *Frangula alnus*: An Integrated Model Framework. Ph.D., Stony Brook University

Aiello-Lammens ME, Chu-Agor ML, Convertino M, et al (2011) The impact of sea-level rise on Snowy Plovers in Florida: integrating geomorphological, habitat, and metapopulation models. Global Change Biology 17:3644–3654. https://doi.org/10.1111/j.1365-2486.2011.02497.x

Aiello-Lammens ME, Resit Akçakaya H (2017) Using global sensitivity analysis of demographic models for ecological impact assessment. Conservation Biology 31:116–125. https://doi.org/10.1111/cobi.12726

Akçakaya HR (2000) Population viability analyses with demographically and spatially structured models. Ecological Bulletins 23–38

Akçakaya HR (2001) Linking population-level risk assessment with landscape and habitat models. The Science of the total environment 274:283–291

Akçakaya HR (2002) RAMAS GIS: Linking Spatial Data with Population Viability Analysis. Applied Biomathematics, Setauket

Akçakaya HR, Radeloff VC, Mladenoff DJ, He HS (2004) Integrating landscape and metapopulation modeling approaches: viability of the sharp‐tailed grouse in a dynamic landscape. Conservation Biology 18:526–537

Anderson RP (2012) Harnessing the world’s biodiversity data: promise and peril in ecological niche modeling of species distributions. Annals of the New York Academy of Sciences 1260:66–80. https://doi.org/10.1111/j.1749-6632.2011.06440.x

Berg J (2011) Susceptibility of five wetland community types to invasion by Glossy Buckthorn (*Frangula alnus* Mill.). Masters of Science, University of Wisconsin-Milwaukee

Blackburn TM, Pyšek P, Bacher S, et al (2011) A proposed unified framework for biological invasions. Trends in Ecology & Evolution 26:333–339. https://doi.org/10.1016/j.tree.2011.03.023

Briscoe NJ, Elith J, Salguero‐Gómez R, et al (2019) Forecasting species range dynamics with process‐explicit models: matching methods to applications. Ecol Lett ele.13348. https://doi.org/10.1111/ele.13348

Brook BW, Akçakaya HR, Keith DA, et al (2009) Integrating bioclimate with population models to improve forecasts of species extinctions under climate change. Biology Letters 5:723–725. https://doi.org/10.1098/rsbl.2009.0480

Burnham KM, Lee TD (2009) Canopy gaps facilitate establishment, growth, and reproduction of invasive *Frangula alnus* in a *Tsuga canadensis* dominated forest. Biological Invasions 12:1509–1520. https://doi.org/10.1007/s10530-009-9563-8

Cadotte MW, Lovett-Doust J (2001) Ecological and taxonomic differences between native and introduced plants of southwestern Ontario. Ecoscience 8:230–238

Caswell H (2006) Matrix Population Models: Construction, Analysis, and Interpretation, Second. Sinauer Associates, Sunderland, MA

Catling PM, Porebski ZS (1994) The history of invasion and current status of glossy buckthorn, *Rhamnus frangula*, in southern Ontario. Canadian field-naturalist 108:305–310

Clark JS, Lewis M, Horvath L (2001) Invasion by extremes: population spread with variation in dispersal and reproduction. The American Naturalist 157:537–54. https://doi.org/10.1086/319934

Colautti RI, Grigorovich IA, MacIsaac HJ (2006) Propagule Pressure: A Null Model for Biological Invasions. Biological Invasions 8:1023–1037. https://doi.org/10.1007/s10530-005-3735-y

Comita L, Hubbell S (2009) Local neighborhood and species’ shade tolerance influence survival in a diverse seedling bank. Ecology 90:328–334

Coutts SR, Yokomizo H (2013) Meta-models as a straightforward approach to the sensitivity analysis of complex models. Population Ecology 56:7–19. https://doi.org/10.1007/s10144-013-0422-1

Crone EE, Menges ES, Ellis MM, et al (2010) How do plant ecologists use matrix population models? Ecology Letters 14:1–8. https://doi.org/10.1111/j.1461-0248.2010.01540.x

Crooks J (2005) Lag times and exotic species: the ecology and management of biological invasions in slow-motion. Ecoscience 12:316–329

Crooks JA, Soulé ME (1999) Lag times in population explosions of invasive species: causes and implications. In: Sandlund OT, Schei PJ, Viken A (eds) Invasive species and biodiversity management. Kluwer Academic Dordrecht, The Netherlands, pp 103–125

Cunard C, Lee TD (2008) Is patience a virtue? Succession, light, and the death of invasive glossy buckthorn (*Frangula alnus*). Biological Invasions 11:577–586. https://doi.org/10.1007/s10530-008-9272-8

Davis AS, Landis DA, Nuzzo V, et al (2006) Demographic models inform selection of biocontrol agents for garlic mustard (*Alliaria petiolata*). Ecological Applications 16:2399–2410

Davis M, Grime J, Thompson K (2000) Fluctuating resources in plant communites: a general theory of invasibility. Journal of Ecology 88:528–534

Dwyer JM, Fensham RJ, Fairfax RJ, Buckley YM (2010) Neighbourhood effects influence drought-induced mortality of savanna trees in Australia. Journal of Vegetation Science 21:573–585. https://doi.org/10.1111/j.1654-1103.2009.01167.x

Easterling MR, Ellner SP, Dixon PM (2000) Size-specific sensitivity: applying a new structured population model. Ecology 81:694–708

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. Journal of Animal Ecology 77:802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Elith J, Phillips SJ, Hastie T, et al (2010) A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17:43–57. https://doi.org/10.1111/j.1472-4642.2010.00725.x

Ellner SP, Rees M (2006) Integral projection models for species with complex demography. The American Naturalist 167:410–428

Ellner SP, Rees M (2007) Stochastic stable population growth in integral projection models: theory and application. Journal of mathematical biology 54:227–56. https://doi.org/10.1007/s00285-006-0044-8

Fagan ME, Peart DR (2004) Impact of the invasive shrub glossy buckthorn (*Rhamnus frangula* L.) on juvenile recruitment by canopy trees. Forest Ecology and Management 194:95–107. https://doi.org/10.1016/j.foreco.2004.02.015

Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24:38–49

Fordham DA, Akçakaya HR, Araújo MB, Brook BW (2012) Modeling Range Shifts for Invasive Vertebrates in Response to Climate Change. In: Brodie (ed). pp 1–31

Fordham DA, Mellin C, Russell BD, et al (2013) Population dynamics can be more important than physiological limits for determining range shifts under climate change. Global change biology 19:3224–37. https://doi.org/10.1111/gcb.12289

Foster BL, Gross KL (1999) Temporal and spatial patterns of woody plant establishment in Michigan old fields. The American Midland Naturalist 142:229–243

Franklin J (2010) Moving beyond static species distribution models in support of conservation biogeography. Diversity and Distributions 16:321–330. https://doi.org/10.1111/j.1472-4642.2010.00641.x

Franklin J, Regan HM, Syphard AD (2013) Linking spatially explicit species distribution and population models to plan for the persistence of plant species under global change. Environmental Conservation 1–13. https://doi.org/10.1017/S0376892913000453

Frappier B, Eckert RT, Lee TD (2003a) Potential impacts of the invasive exotic shrub *Rhamnus frangula* L. (glossy buckthorn) on forests of southern New Hampshire. Northeastern Naturalist 10:277–296

Frappier B, Eckert RT, Lee TD (2004) Experimental removal of the non-indigenous shrub *Rhamnus frangula* (glossy buckthorn): Effects on native herbs and woody seedlings. Northeastern Naturalist 11:333–342

Frappier B, Lee TD, Olson KF, Eckert RT (2003b) Small-scale invasion pattern, spread rate, and lag-phase behavior of *Rhamnus frangula* L. Forest Ecology and Management 186:1–6. https://doi.org/10.1016/S0378-1127(03)00274-3

Gavier-Pizarro GI, Radeloff VC, Stewart SI, et al (2010) Housing is positively associated with invasive exotic plant species richness in New England, USA. Ecological Applications 20:1913–1925

Godwin H (1943) *Frangula alnus* Miller. Journal of Ecology 31:77–92

Godwin H (1936) Studies in the ecology of Wicken Fen: III. the establishment and development of Fen Scrub (Carr). The Journal of Ecology 82–116

Graham CH, Ferrier S, Huettman F, et al (2004) New developments in museum-based informatics and applications in biodiversity analysis. Trends in Ecology & Evolution 19:497–503. https://doi.org/10.1016/j.tree.2004.07.006

Granstrom A (1988) Seed banks at six open and afforested heathland sites in southern Sweden. Journal of Applied Ecology 297–306

Gurevitch J, Fox GA, Wardle GM, et al (2011) Emergent insights from the synthesis of conceptual frameworks for biological invasions. Ecology Letters 14:407–418. https://doi.org/10.1111/j.1461-0248.2011.01594.x

Hampe A (2008) Fruit tracking, frugivore satiation, and their consequences for seed dispersal. Oecologia 156:137–145. https://doi.org/10.1007/s00442-008-0979-0

Hampe A (2004) Extensive hydrochory uncouples spatiotemporal patterns of seedfall and seedling recruitment in a ‘bird‐dispersed’ riparian tree. Journal of Ecology 92:797–807

Hampe A, Arroyo J, Jordano P, Petit RJ (2003) Rangewide phylogeography of a bird-dispersed Eurasian shrub: contrasting Mediterranean and temperate glacial refugia. Molecular Ecology 12:3415–3426. https://doi.org/10.1046/j.1365-294X.2003.02006.x

Hampe A, Bairlein F (2000) Modified dispersal‐related traits in disjunct populations of bird‐dispersed *Frangula alnus* (Rhamnaceae): a result of its Quaternary distribution shifts? Ecography 23:603–613

Harris CM, Park KJ, Atkinson R, et al (2009) Invasive species control: Incorporating demographic data and seed dispersal into a management model for *Rhododendron ponticum*. Ecological Informatics 4:226–233. https://doi.org/10.1016/j.ecoinf.2009.07.005

Hastings A, Cuddington K, Davies KF, et al (2005) The spatial spread of invasions: new developments in theory and evidence. Ecology Letters 8:91–101. https://doi.org/10.1111/j.1461-0248.2004.00687.x

Herron PM, Martine CT, Latimer AM, Leicht-Young SA (2007) Invasive plants and their ecological strategies: prediction and explanation of woody plant invasion in New England. Diversity and Distributions 13:633–644. https://doi.org/10.1111/j.1472-4642.2007.00381.x

Howell JA, Blackwell WHJr (1977) The history of *Rhamnus frangula* (glossy buckthorn) in the Ohio flora. Castanea 42:111–115

Jongejans E, Shea K, Skarpaas O, et al (2008) Dispersal and demography contributions to population spread of *Carduus nutansin* its native and invaded ranges. Journal of Ecology 96:687–697. https://doi.org/10.1111/j.1365-2745.2008.01367.x

Keith DA, Akçakaya HR, Thuiller W, et al (2008) Predicting extinction risks under climate change: coupling stochastic population models with dynamic bioclimatic habitat models. Biology Letters 4:560–563. https://doi.org/10.1098/rsbl.2008.0049

Klein Goldewijk K, Beusen A, Van Drecht G, De Vos M (2010) The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. Global Ecology and Biogeography 20:73–86. https://doi.org/10.1111/j.1466-8238.2010.00587.x

Knight TM, Havens K, Vitt P (2011) Will the use of less fecund cultivars reduce the invasiveness of perennial plants? BioScience 61:816–822. https://doi.org/10.1525/bio.2011.61.10.11

Kostel-Hughes F, Young TP, McDonnell MJ (1998) The soil seed bank and its relationship to the aboveground vegetation in deciduous forests in New York City. Urban Ecosystems 2:43–59

Kowarik I (1995) Time lags in biological invasions with regard to success and failure of alien species. In: Pyšek P, Prach K, Rejmánek M, Wade M (eds) Plant invasions: General aspects and special problems. SPB Academic Publishing, Amsterdam, The Netherlands

Larkin DJ (2011) Lengths and correlates of lag phases in upper-Midwest plant invasions. Biological Invasions 14:827–838. https://doi.org/10.1007/s10530-011-0119-3

Lavoie C (2012) Biological collections in an ever changing world: herbaria as tools for biogeographical and environmental studies. Perspectives in Plant Ecology, Evolution and Systematics 15:68–76. https://doi.org/10.1016/j.ppees.2012.10.002

Lee TD, Thompson JH (2012) Effects of logging history on invasion of eastern white pine forests by exotic glossy buckthorn (*Frangula alnus* P. Mill.). Forest Ecology and Management 265:201–210. https://doi.org/10.1016/j.foreco.2011.10.035

Lowry E, Rollinson EJ, Laybourn AJ, et al (2012) Biological invasions: a field synopsis, systematic review, and database of the literature. Ecology and Evolution 3:182–96. https://doi.org/10.1002/ece3.431

Mack RN, Simberloff D, Mark Lonsdale W, et al (2000) Biotic invasions: causes, epidemiology, global consequences, and control. Ecological Applications 10:689–710

Mason R a. B, Cooke J, Moles AT, Leishman MR (2008) Reproductive output of invasive versus native plants. Global Ecology and Biogeography 17:633–640. https://doi.org/10.1111/j.1466-8238.2008.00402.x

McDonald RI, Motzkin G, Foster DR (2008) Assessing the influence of historical factors, contemporary processes, and environmental conditions on the distribution of invasive species. The Journal of the Torrey Botanical Society 135:260–271

Medan D (1994) Reproductive biology of *Frangula alnus* (Rhamnaceae) in southern Spain. Plant Systematics and Evolution 193:173–186. https://doi.org/10.1007/BF00983549

Menges ES (2000) Population viability analyses in plants: challenges and opportunities. Trends in Ecology & Evolution 15:51–56

Merow C, LaFleur N, Silander Jr JA, et al (2011) Developing dynamic mechanistic species distribution models: predicting bird-mediated spread of invasive plants across northeastern North America. The American Naturalist 178:30–43. https://doi.org/10.1086/660295

Merow C, Latimer AM, Wilson AM, et al (2014) On using integral projection models to generate demographically driven predictions of species’ distributions: development and validation using sparse data. Ecography 37:1167–1183. https://doi.org/10.1111/ecog.00839

Merow C, Smith MJ, Silander Jr JA (2013) A practical guide to MaxEnt for modeling species’ distributions: what it does, and why inputs and settings matter. Ecography 36:1058–1069. https://doi.org/10.1111/j.1600-0587.2013.07872.x

Mills JE, Meyer GA, Reinartz JA (2012) An exotic invasive shrub has greater recruitment than native shrub species within a large undisturbed wetland. Plant Ecology 213:1425–1436

Mills JE, Reinartz JA, Meyer GA, Young EB (2009) Exotic shrub invasion in an undisturbed wetland has little community-level effect over a 15-year period. Biological Invasions 11:1803–1820. https://doi.org/10.1007/s10530-008-9359-2

Mosher ES, Silander Jr JA, Latimer AM (2009) The role of land-use history in major invasions by woody plant species in the northeastern North American landscape. Biological Invasions 11:2317–2328. https://doi.org/10.1007/s10530-008-9418-8

Nehrbass N, Winkler E, Müllerová J, et al (2006) A simulation model of plant invasion: long-distance dispersal determines the pattern of spread. Biological Invasions 9:383–395. https://doi.org/10.1007/s10530-006-9040-6

Pardini EA, Drake JM, Chase JM, Knight TM (2009) Complex population dynamics and control of the invasive biennial *Alliaria petiolata* (garlic mustard). Ecological Applications 19:387–397

Pearson RG, Stanton JC, Shoemaker KT, et al (2014) Life history and spatial traits predict extinction risk due to climate change. Nature Climate Change 4:217–221. https://doi.org/10.1038/NCLIMATE2113

Peterson AT, Soberón J, Pearson RG, et al (2011) Ecological niches and geographic distributions (MPB-49). Princeton University Press, Princeton

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecological Modelling 190:231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Phillips SJ, Dudík M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31:161–175. https://doi.org/10.1111/j.2007.0906-7590.05203.x

Possessky SL, Williams CE, Moriarity WJ (2000) Glossy buckthorn, *Rhamnus frangula* L.: A threat to riparian plant communities of the northern Allegheny Plateau (USA). Natural Areas Journal 20:290–292

Prowse TAA, Johnson CN, Lacy RC, et al (2013) No need for disease: Testing extinction hypotheses for the thylacine using multi-species metamodels. Journal of Animal Ecology 82:355–364. https://doi.org/10.1111/1365-2656.12029

Pyšek P, Hulme PE (2005) Spatio-temporal dynamics of plant invasions: linking pattern to process. Ecoscience 12:302–315

Ramula S, Buckley YM (2010) Management recommendations for short-lived weeds depend on model structure and explicit characterization of density dependence. Methods in Ecology and Evolution 1:158–167. https://doi.org/10.1111/j.2041-210X.2010.00022.x

Ramula S, Knight TM, Burns JH, Buckley YM (2008) General guidelines for invasive plant management based on comparative demography of invasive and native plant populations. Journal of Applied Ecology 45:1124–1133. https://doi.org/10.1111/j.1365-2664.2008.01502.x

Rees M, Ellner SP (2009) Integral projection models for populations in temporally varying environments. Ecological Monographs 79:575–594

Sakai AK, Allendorf FW, Holt JS, et al (2001) The population biology of invasive species. Annual Review of Ecology and Systematics 32:305–332

Searcy KB, Pucko C, McClelland D (2006) The distribution and habitat preferences of introduced species in the Mount Holyoke Range, Hampshire Co., Massachusetts. Rhodora 108:43–61. https://doi.org/10.3119/04-16.1

Simberloff D (2009) The Role of Propagule Pressure in Biological Invasions. Annual Review of Ecology, Evolution, and Systematics 40:81–102. https://doi.org/10.1146/annurev.ecolsys.110308.120304

Stanton JC (2014) Present-day risk assessment would have predicted the extinction of the passenger pigeon (*Ectopistes migratorius*). Biological Conservation 180:11–20. https://doi.org/10.1016/j.biocon.2014.09.023

Sullivan MJP, Davies RG, Reino L, Franco AM a. (2012) Using dispersal information to model the species-environment relationship of spreading non-native species. Methods in Ecology and Evolution 3:870–879. https://doi.org/10.1111/j.2041-210X.2012.00219.x

Syfert MM, Smith MJ, Coomes DA (2013) The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. PLoS ONE 8:e55158. https://doi.org/10.1371/journal.pone.0055158

Theoharides KA, Dukes JS (2007) Plant invasion across space and time: factors affecting nonindigenous species success during four stages of invasion. New Phytologist 176:256–273. https://doi.org/10.1111/j.1469-8137.2007.02207.x

Urban MC, Phillips BL, Skelly DK, Shine R (2007) The cane toad’s *(Chaunus [Bufo] marinus)* increasing ability to invade Australia is revealed by a dynamically updated range model. Proceedings of the Royal Society B: Biological Sciences 274:1413–9. https://doi.org/10.1098/rspb.2007.0114

VanDerWal J, Shoo LP, Johnson CN, Williams SE (2009) Abundance and the environmental niche: environmental suitability estimated from niche models predicts the upper limit of local abundance. The American Naturalist 174:282–91. https://doi.org/10.1086/600087

Vindenes Y, Engen S, Saether B-E (2011) Integral projection models for finite populations in a stochastic environment. Ecology 92:1146–56. https://doi.org/10.1890/10-0500.1

With K a. (2002) The Landscape Ecology of Invasive Spread. Conservation Biology 16:1192–1203. https://doi.org/10.1046/j.1523-1739.2002.01064.x

With KA (2004) Assessing the risk of invasive spread in fragmented landscapes. Risk Analysis 24:803–815. https://doi.org/10.1111/j.0272-4332.2004.00480.x

**Table 1.** Bioclimatic and environmental predictor variables use in the SDM of *F. alnus*. Dynamic variables listed first, followed by static variables.

|  |  |  |  |
| --- | --- | --- | --- |
| Variable source or data Set | Variable Name | Description | Ecological Basis for Inclusion |
| Natural Resources Canada International Climate Modeling Project (McKenney et al. 2006, 2011) | BIO4 | Temperature seasonality | Indicative of evenness of temperature throughout the year |
|  | BIO5 | Max temperature of warmest month | Indicative of plant heat tolerance |
|  | BIO6 | Min temperature of coolest month | Indicative of plant cold tolerance (e.g., frost) |
|  | BIO12 | Annual precipitation | Proxy for total water input into a region and general plant moisture requirements |
|  | BIO15 | Precipitation seasonality | Indicative of evenness of rainfall throughout the year |
|  | BIO18 | Precipitation of the warmest quarter | Indicative of plant drought tolerance |
|  | BIO19 | Precipitation of the coldest quarter | Proxy for effects of snow pack, potentially important of *F. alnus* growth at northern edge of species range |
|  | SG 2 | Julian day number at start of growing season |  |
|  | SG 3 | Julian day number at end of growing season |  |
|  | SG 10 | Growing Degree Days (above 5 C°) | A measure of total growing time |
| Historic Land Use (Klein Goldewijk et al. 2010) | Crop | Percent crop cover | Proxies for land use and land disturbance |
|  | Grass | Percent pasture cover |  |
| Hydro1K North America | CTI | Compound Topographic Index – General wetness index | *F. alnus* has been reported to invade areas with moist soils (e.g., wetlands) |
| Potential Vegetation (REF Ramunkuty) | Potential Veg | Potential Vegetation | A measure of suitability in specific vegetation communities |
| FAO-UN Soil (REF) | pH Top | Soil pH of top soil | *F. alnus* has been reported to do well in acidic soils |
|  | pH Sub | Soil pH below top soil layer |  |

**Table 2.** The eight continuous, and one categorical, parameters that were varied in a global sensitivity analysis (GSA) of the demographic model for *F. alnus*. Each parameter was estimated based on field observations, information culled from literature, or estimates based on studies of similar species.

|  |  |  |
| --- | --- | --- |
|  | **Parameter** | **Estimate information** |
| Continuous parameters | |  |
|  | Survival-Growth | Field observations - Combination of logistic regression of survival on size and linear regression of growth on size; Direct observation of seedlings |
|  | Fecundity | Field observations and literature - ANCOVA of fruit per plant by size and effective plot density; Seeds per fruit and germination rates from literature |
|  | Variability in Survival | Field observations - Inter-annual variation in growth |
|  | Variability in Fecundity | Field observations - Coefficient of variation of number of observed fruit over three years at field sites |
|  | Dispersal | Literature - Modified model of dispersal of *Celastrus* *orbiculatus* (reported by Merow et al. 2011) to fit observed dispersal of *F. alnus* (as reported by Hampe 2004, 2008; Berg 2011) |
|  | Long-distance dispersal (LDD) | Randomly generated - each simulation could have between 1 and 500 LDD events, corresponding to a maximum of approximately 5 per year (shown to be appropriate for spread of *Celastrus orbicuatus* by Merow et al. 2011) |
|  | Initial abundance | Varied between 1,000 and 10,000 |
|  | Inter-population correlation | Literature - Inter-population correlation was assumed to be similar for values reported for all of North America reported by Pearson et al. (2014) |
|  |  |  |
| Categorical parameter | |  |
|  | Carrying capacity scenario | Three scenarios - 100%, 50%, and 25% of patch (20 x 20 km) could contain 2 x 2 m plots. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Parameter | Estimate | Std. Error | t value | P |
| Size Class B: Intercept | -1.702 | 0.09842 | -17.29 | 1.375e-57 |
| Size Class C: Intercept | 0.901 | 0.2328 | 11.18 | 3.705e-27 |
| Size Class D: Intercept | 1.632 | 0.3478 | 9.586 | 9.946e-21 |
| Size Class B: Slope | -0.005839 | 0.007915 | -0.7377 | 0.4609 |
| Size Class C: Slope | -0.07356 | 0.03281 | -2.242 | 0.02524 |
| Size Class D: Slope | -0.3441 | 0.2031 | -1.694 | 0.09064 |

**Table 3.** Intercept and slope parameter estimates resulting from ANCOVA; = 0.2826, = 66.51, *df* = 5 and 844, p < 0.001.

**Figure 1.** Component regression models (A-C) combined to parameterize the IPM kernel (D). The IPM kernel is a function of survival, growth, and fecundity. (A) Survival probability was fit as a logistic regression model, with plant size at time *t* as a predictor variable of survival to time *t+1*. (B) Growth was fit as a linear regression, treating size at time *t* as the predictor of size at time *t+1*. (C) Number of fruit per individual was fit using an ANCOVA model, allowing for unequal slope and intercept values. Plants were assigned into categorical sizes and plot effective density was treated as a continuous predictor variable, with log(Fruit count) as the response variable. (D) A full IPM kernel function was parameterized using results from these regressions and other parameters estimated from literature (see equation (3)). This kernel was discretized yielding a 50 x 50 matrix analogous to a 50-stage transition matrix.

**Figure 2.** Habitat suitability (HS) through time. (A) Study region (red square) with respect to northeast and middle North America, and all occurrence records. (B) HS for six representative time points. Black dots are *F. alnus* occurrence records up to, and including, the HS layer year. (C) Change in total metapopulation carrying capacity through time under high, medium, and low scenarios.

**Figure 3.** Sensitivity of simulation models through time. Simulations plotted here were those using plot effective density dependence, land-use change, and long-distance dispersal weighted by human population density. The sensitivity of each model with respect to historical occurrence records was calculated for each year of the simulation (1911 to 2010). A sensitivity value of 1 indicates that the simulation predicted occurrences in all populations where *F. alnus* was observed in that year or prior. Results from the three different occupancy thresholds are represented by three different colors.

**Figure 4.** Cumulative occupied area (measured as the number of 20 x 20 km patches classified as occupied) curves for all simulation models. Simulations presented here are those using plot effective density dependence, land-use change, and long-distance dispersal weighted by human population density. The black line represents cumulative occupied area for *F. alnus* based on historical occurrence records. Results from three different occupancy thresholds are represented by three different colors. The red line represents the square-root of the maximum number of patches occupied, based on the historical observations.

**Figure 5.** Sensitivity versus positive predictive power for simulations using plot effective density dependence, land-use change, and long-distance dispersal weighted by human population density. Model fit measures were calculated based on an occupancy threshold of 1000 individuals. Simulations classified as having a combined sensitivity and positive predictive power metric value of 1 (90 of 500 simulations) are black, all others are grey.

**Figure 6.** Response curves of predictor variables versus combined sensitivity and positive predictive power metric from BRT analysis. Each parameters relative influence value is in parentheses after the parameter name. Higher logit(p) values indicate better fit.

**Figure 7.** Input parameter density plots with dark grey densities representing all 500 randomly generated parameter sets and light gray densities representing parameter sets with combined sensitivity and positive predictive power metric equal to 1. Occupancy threshold was set as 1000 individuals in a patch.

**Figure 8.** Four most important interactions among input parameters, as determined by BRT analysis. Higher logit (p) values indicate better fit.