ECOLOGY LETTERS

Ecology Letters, (2017) **20**: 426–435 doi: 10.1111/ele.12741

LETTER

Predicting the spread of all invasive forest pests in the United States

Emma J. Hudgins,[1]* Andrew M. Liebhold[2] and Brian Leung[1]

[1]*Biology Department, McGill University, Montreal, QC, Canada*
[2]*Northern Research Station, USDA Forest Service, Morgantown, WV, USA*

**Correspondence: E-mail: emma.hudgins@mail.mcgill.ca*

Abstract

We tested whether a general spread model could capture macroecological patterns across all damaging invasive forest pests in the United States. We showed that a common constant dispersal kernel model, simulated from the discovery date, explained 67.94% of the variation in range size across all pests, and had 68.00% locational accuracy between predicted and observed locational distributions. Further, by making dispersal a function of forest area and human population density, variation explained increased to 75.60%, with 74.30% accuracy. These results indicated that a single general dispersal kernel model was sufficient to predict the majority of variation in extent and locational distribution across pest species and that proxies of propagule pressure and habitat invasibility – well-studied predictors of establishment – should also be applied to the dispersal stage. This model provides a key element to forecast novel invaders and to extend pathway-level risk analyses to include spread.

Keywords

Dispersal kernel, habitat invasibility, macroecology, propagule pressure, spatially explicit.

Ecology Letters (2017) **20**: 426–435

INTRODUCTION

The number of invasions by non-indigenous forest pests is increasing worldwide due to growing travel and trade (Liebhold 2012). Pest invasions consist of three phases: arrival at a site, establishment at that location and subsequent spread (Elton 2000). Predictive models of each process allow managers to take targeted actions, decreasing the number of pests completing each phase (Simberloff & Gibbons 2004).

Arguably, many of the drivers of invasion stages should be common among invading species. However, whether a generalised predictive model is possible or whether species must be modelled and fit separately to account for idiosyncrasies remains unanswered (Leung *et al*. 2012). To this effect, we focus on the third phase, and build a general predictive tool for invasive forest pest spread within the United States. This is of fundamental interest in ecology – to determine whether general rules govern dispersal sufficiently to quantitatively make predictions across a suite of species using a common model. Spread at a macroecological scale is particularly relevant, given that there are conflicting opinions as to whether generalities exist across invasions and there have been few largescale studies of potential generalities (Cadotte *et al*. 2006). Such general rules could also have considerable applied value, allowing spread to be predicted *a priori*. Current spread models are typically derived after a pest has already arrived and spread (e.g. Gilbert *et al*. 2004; Morin *et al*. 2009; Tisseuil *et al*. 2015; Walter *et al*. 2015), potentially lessening the efficacy of management, for instance when implementing containment, rapid response and/or eradication efforts of novel species (Lovett *et al*. 2016).

The spread process has been extensively modelled using population ecology models that employ general growth and dispersal equations (Shigesada *et al*. 1995; Neubert & Parker 2004; Skarpaas & Shea 2007). Such models have been typically parameterised for individual species separately, and do not incorporate moderating variables for habitat invasibility, propagule pressure and pest life history, which have been highlighted within the invasion biology literature, would aid generalisation (Leung *et al*. 2012), but have not been examined in an explicit spread model. For instance, invasion biologists have theorised that certain habitat suitability characteristics may allow for greater establishment success, or greater habitat invasibility (Simberloff 2009). Logically, habitat invasibility could be incorporated into a dispersal kernel model, where certain habitat characteristics accelerate dispersal into or out of cells. In addition to environmental factors, proxies of propagule pressure, or the number of pest individuals introduced, have been associated with higher probabilities of establishment (Lockwood *et al*. 2005; Bradie *et al*. 2013). Likewise, higher numbers of individuals dispersed should accelerate secondary spread. Variables such as human population density may moderate propagule pressure, as humans are often vectors for pest spread, especially over long distances (Haack *et al*. 2010). These factors could also be included in the dispersal kernel to repartition dispersal preferentially into or out of more frequently visited locations. Finally, life history characteristics may be important in determining differential spread rates across species. For instance, the body size of forest pests may be a predictor of their spread rates, as larger species may disperse farther.

Thus, an ideal generalised model of spread would predict invasions across a suite of species using a single model. It would do so both in terms of the extent of invasion as well as the spatial locations of pest occurrences, and would integrate various factors influencing pest invasions. However, the extent to which such a general model is predictive, and which (if any) factors have common influence across an array of pest species remains an open question.

In this study, we developed a generalised dispersal kernel (GDK) model and tested two hypotheses: First, we hypothesised that the spread of biological invaders proceeds following similar processes across species, and so we predicted that a single general model of pest spread can fit well for all forest pests in the United States. Secondly, we hypothesised that pest life history, propagule pressure, and habitat invasibility can be meaningfully integrated into a dispersal kernel, and lead to improved predictions in a general model.

MATERIAL AND METHODS

Description of data

To build our model of pest spread, we used county-level species occurrence data, habitat suitability factors, and propagule pressure proxies from Liebhold *et al*. (2013), the Alien Forest Pest Explorer (Liebhold 2012, http://www.nrs.fs.fed.us/tools/afpe/maps) and 27 sources for pest characteristics (Table S1). Year of each pest species' first detection from Liebhold *et al*. (2013) was used as a proxy for year of establishment. Pest occurrence data consisted of 75 pest species distributions, comprising insects, mites and tree pathogens. We modelled each species as an independent unit and we did not consider interactions between species. We did consider additional predictors, consisting of propagule pressure proxies [human population density per decade (km^{-2}), per capita income in 1999 (USD), road length (km)], habitat invasibility proxies [host species richness, tree density ($m^3\ km^{-2}$), forested land (km^2), host tree density ($m^3\ km^{-2}$)] and pest life history traits (taxa: arthropods versus tree pathogens, number of host species, Eurasian versus non-Eurasian native range, maximum body length (mm) (with a separate intercept fit for tree pathogens, as body length was not applicable)] (Table 1, Table S2). All predictor variables included had a correlation $r < 0.7$ with other predictors (Table S3). Our discrete time dispersal model was fit in decadal increments to achieve computational feasibility and because our human population data were decadal (although we examined sensitivity by using 5-year increments, and found that model fit did not differ substantially, Fig. S4). Each species' time since detection was rounded down to the nearest decade, and so we limited our analysis to species that had been present in the United States for at

Table 1 Results of stepwise regression for the dispersal kernel model fit to United States data using habitat invasibility (HI), propagule pressure (PP) and pest life history (LH) factors

Term	Type	Description	Entry order	Estimate (β_p)	MET score (km)	R^2_{MSE}	Δ MET score (km)
Intercept	NA	NA	1	1.1248	113.07	0.6794	NA
Forested land (km^2)*	HI	Sum of land area covered by forest	2	−0.8438	68.11	0.7231	−44.96
Human population (km^{-2})*	PP	Current human population density at each time step	3	−0.1378	60.44	0.7560	−7.67
Forested land (km^2)†	HI	Sum of land area covered by forest	NA	0.4106	57.69	0.7519	−2.75
Host density ($km^3\ km^{-2}$)*	HI	Host tree volume for that particular pest per grid cell	NA	−0.0205	60.04	0.7558	−0.40
Host density ($km^3\ km^{-2}$)†	HI	Host tree volume for that particular pest per grid cell	NA	0.0031	60.41	0.7517	−0.03
Host species count*	HI	Number of tree species that are hosts of any pest present in grid cell	NA	−0.0384	60.04	0.7442	−0.40
Host species count†	HI	Number of tree species that are hosts of any pest present in grid cell	NA	0.1186	60.28	0.7524	−0.16
Tree density ($m^3\ km^{-2}$)*	HI	Total tree volume by grid cell	NA	0.3067	59.67	0.7485	−0.77
Tree density ($m^3\ km^{-2}$)†	HI	Total tree volume by grid cell	NA	0.1060	60.26	0.7488	−0.18
Body size (mm)	LH	Pest body length (separate intercept fit for fungi)	NA	0.0011; 0.1464	60.24	0.7572	−0.20
Continent of origin	LH	Eurasian vs. Non-Eurasian	NA	0	60.44	0.7560	0
Feeding guild	LH	Pathogens vs. Arthropods	NA	0.0163	60.25	0.7489	−0.19
Number of hosts	LH	Number of host species possessed by pest	NA	−0.0012	60.41	0.7557	−0.03
Human population (km^2)†	PP	Current human population density at each time step	NA	0.0217	60.32	0.7527	−0.12
Income (USD)*	PP	Per capita income in 1999	NA	0	60.44	0.7560	0
Income (USD)†	PP	Per capita income in 1999	NA	0	60.44	0.7560	0
Road length (km)*	PP	Total length of all major roads in grid cell	NA	−0.0148	60.17	0.7463	−0.27
Road length (km)†	PP	Total length of all major roads in grid cell	NA	0.4935	58.64	0.7587	−1.8

*Parameters influencing the probability of dispersal into a cell (Z_I).
†Parameters influencing the probability of dispersal out of a cell (Z_O).
Negative estimates indicate positive influences on dispersal and vice versa. Since all variables were standardised, the relative influence of each fitted parameter on dispersal can be determined by its magnitude (magnitude of 'Estimate' in the table). Conversely, the relative importance of each parameter on minimum energy test (MET) is determined by its entry order in our generalised dispersal kernel model (See Fig. S3). Our best model had $\delta = 2.4321$ and $\Phi = 0.0006227$ with a jackknifed MET score of 60.44 km per species and a jackknifed R^2_{MSE} of 0.7579. Terms with entry order 'NA' did not meet our variable importance threshold for inclusion, and their associated data is for their proposed inclusion as a fourth term in our model.

least 10 years at the time of data collection ($N = 64$). Years of first detection spanned between 1790 and 1997 (1–21 time steps).

To avoid issues with spread dynamics across unequal county sizes, we first converted county-level presence/absence data to a 50 × 50 km lattice using distances measured with the United States Equidistant Conic Projection calculated using ArcMap10.2 (ESRI 2011). Environmental variables within each grid cell were calculated using the area-weighted average of the US counties it encompassed. Further, all explanatory variables were centred and scaled to have a mean of zero and variance of one to ease the interpretation of each variable's contribution to dispersal.

Generalised dispersal kernel model

Dispersal kernels estimate the probability of pest dispersal across space based on the distance between source and destination cells. In our model, we tested additional predictors of spread through modifying the dispersal parameter (discussed below). While we recognise that the predictors may influence processes other than dispersal, our focus was explicitly on the dispersal process in this study. We fit a negative exponential dispersal kernel model using discrete time simulations (Fig. 1), where at each (decadal) time step, pests dispersed to surrounding patches according to:

$$T_{i,j} = \frac{e^{-d_{i,j}f(Z)}}{\sum_j e^{-d_{i,j}f(Z)}} \quad (1)$$

$$d_{i,j} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \quad (2)$$

where $T_{i,j}$ is the proportion of pests dispersing from cell i to cell j, normalised such that proportions sum to one across all cells j for a particular species, $d_{i,j}$ is distance, and $f(Z)$ is a combination of species (Z_S) and cell (dispersal into a cell = Z_I, dispersal out of a cell = Z_O) specific predictors and parameter values influencing the dispersal probabilities. As a special case, we tested a constant dispersal model, where $f(Z)$ was a constant ($f(Z) = \alpha$). In our full GDK model, $f(Z)$ was calculated using:

$$f(Z) = 2\alpha \frac{e^{Z_S+Z_I+Z_O}}{1 + e^{Z_S+Z_I+Z_O}} \quad (3)$$

and the equation for Z_S, Z_I and Z_O followed the general formulation:

$$Z_V = \sum_{p=1}^{k} \beta_p X_p \quad (4)$$

where β_p are parameters associated with variables X – either pest life history variables (S) or environmental variables at either destination cells (I) or source cells (O), thereby allowing these variables to influence dispersal both into and out of a cell.

Additionally, species were only able to invade grid cells where their host species was known to be present, as this was the most comprehensive information available, and logically should be important for pest distributions. For each species, we initiated spread in the central grid cell within each pest's known host distribution. This central point was chosen as a reasonable starting point, given (1) the uncertainty around the true origin for most pest species (Siegert *et al.* 2014), (2) where records of origin exist, they were sometimes found outside of the known host distribution ($n = 21$, possibly due to the presence of undocumented urban hosts), and hence there was no good way to incorporate these records and (3) the lack of urban tree information within our host data. Thus, inferences from the model should be limited to spread throughout natural host distributions.

Next, we assumed that dispersal would increase with time since establishment in each source cell i, because propagules

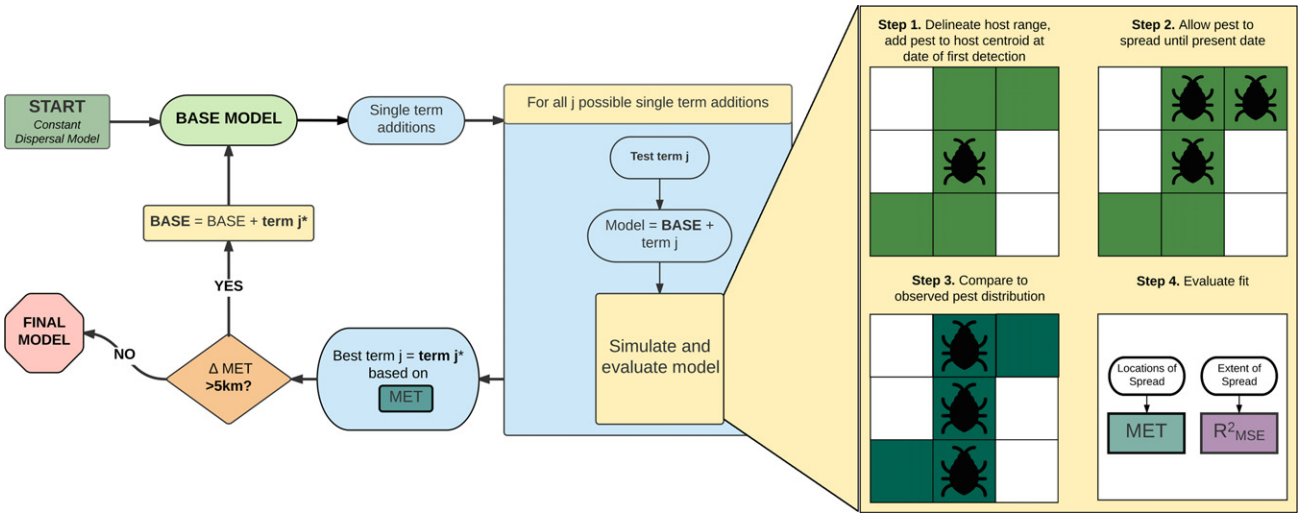

Figure 1 Conceptual framework for model building (main loop) and simulation procedure (expanded yellow box). The model is built by a forward selection procedure, where starting with the intercept-only constant dispersal model, all possible j single term additions are simulated and fit. The best term j* is chosen to be added to the model if it improves the minimum energy test score by at least 5 km and the process is repeated, otherwise the model building procedure is halted and the current model is kept as the final model.

should increase as populations grow, up to some maximum number of individuals. Because we built the model using presence/absence data, we only modelled 'relative propagule pressure'. Specifically, the net influx of propagules and growth at each time step was formulated as an integrodifference equation (Kot *et al*. 1996), where the overall accumulation of propagules at cell i at timestep $t+1$ was equal to the relative propagule pressure at time t, minus emigration, plus immigration, multiplied by the growth rate δ:

$$X_{i,t+1} = (X_{i,t} - \sum_j T_{i,j}X_{i,t} + \sum_k T_{k,i}X_{k,t})\delta \quad (5)$$

Our inclusion of dispersal and growth within a single time step is atypical of integrodifference models, but was needed given that our time steps are decadal.

Cells were considered 'presences' capable of being a source of propagules above a threshold population size Φ, to prevent immediate dispersal across the entire landscape. Both δ and Φ were fitted constants. However, the maximum propagule pressure in a cell was set to 1, because the maximum value was arbitrary and relative to the value of Φ (i.e. maximum 1 with threshold = Φ has identical dynamics as maximum 100 with threshold =100*Φ). These equations allowed us to capture some of the important temporal characteristics associated with population growth, but we do not interpret them as demographic rates, as only presence/absence data were available.

Metrics of fit

We used the minimum energy test (MET) as our metric of model fit between predicted versus observed distributions (Aslan & Zech 2005, Data S1). MET accounts for distances between predicted and true presences, which constitutes a higher information content than exact matches of presence/absence (i.e. mismatch of 50 km is better than 1000 km). Lower MET scores represent models with better goodness-of-fit.

We built our full dispersal model (GDK) using a forward selection procedure (Fig. 1). Starting from the intercept-only constant dispersal model, we determined the MET score for every possible two-term model built by adding the remaining 18 terms individually to the intercept-only model. The term producing the largest improvement was then added to the model, and forward selection repeated with higher-term models, until further additions of terms did not improve the MET score by 5 km. 5 km was chosen as an arbitrary threshold of improvement to prevent overfitting.

We report two metrics of model performance. First, we compared predicted and observed locational distributions, using MET. We also report locational accuracy, which is defined as the proportion of correctly assigned presences and absences across the number of possible presence sites. As a spatial null comparison model, we also used the observed number of infested cells, but simulated randomised occurrences within the host distribution for each species (random allocation model) and took the mean MET score of 1000 simulated pest distributions for each species. As our second metric, we compared predicted to observed range sizes, to evaluate the ability to predict the extent of invasions, using mean squared error, MSE, as a proportion of the variation in observed range sizes; R^2_{MSE} (discussed in Data S5).

As a statistical comparison model for the extent of spread, we regressed the area occupied by all pests against the time since they were first discovered in the United States (Liebhold *et al*. 2013). Though a regression of pest radius and time follows more logically from the invasion literature (Skellam 1951), we wanted to keep predictions comparable across models, and results were very similar between pest radius and pest area (within *c*. 2% variation explained), and did not change our conclusions.

RESULTS

We found that even our statistical comparison model – a simple regression model of pest range area as a function of time – had substantial predictive ability ($R^2_{MSE} = 0.2837$), suggesting promise for a common predictive model across pest species (Fig. 2). Next, we tested a simple dispersal kernel model with a constant dispersal parameter, without consideration of other predictive factors. We found that using a constant dispersal model improved the ability to predict pest range area by more than twofold compared to the simple regression model ($R^2_{MSE} = 0.6794$). The constant dispersal model also performed well comparing the locational predictions against observed infestations (mean MET score = 113.07 km, with 68.00% locational accuracy) more than halving the MET score expected by chance (298.52 km for random allocation null model). These results suggest that (1) there is considerable similarity in dynamics among pest species, (2) using a process-based dispersal kernel model that accounted for host distribution yielded substantial benefits compared to a purely statistical approach.

Next, we tested for common predictive factors, which could modify the dispersal parameter (GDK) (Table 1, Fig. 2). We found that forested land improved the fit more than any other variable, giving a model with $R^2_{MSE} = 0.7231$ and a MET score of 68.11 km (corresponding to an R^2_{MSE} improvement of 0.0437 and a MET score improvement of 44.96 km) and that human population density was the next most important, increasing the model fit to $R^2_{MSE} = 0.7560$ and a MET score of 60.44 km (corresponding to an R^2_{MSE} increase of 0.0328 and a MET score improvement of 7.67 km). The inclusion of forested land and human population density in recipient cells improved MET scores by 52.63 km (Fig. 3), and resulted in a model with 74.30% locational accuracy. Beyond these two, the addition of the other 16 variables tested improved fit by < 5 km in terms of MET (Table 1). Our model building results suggest that (1) there are general predictive factors of pest spread across species, which included both habitat invasibility variables (forested land) as well as proxies for propagule pressure (human population density), (2) the examined species traits other than host associations were not important for spread and (3) relatively simple models explained more than three-fourths of the variation in extent of pest ranges, as well as overlapped three-fourths of the geographical locations across pest species.

When we analysed individual species' contributions to the aggregate MET scores, fewer species were predicted incorrectly by the GDK than the constant dispersal model (as

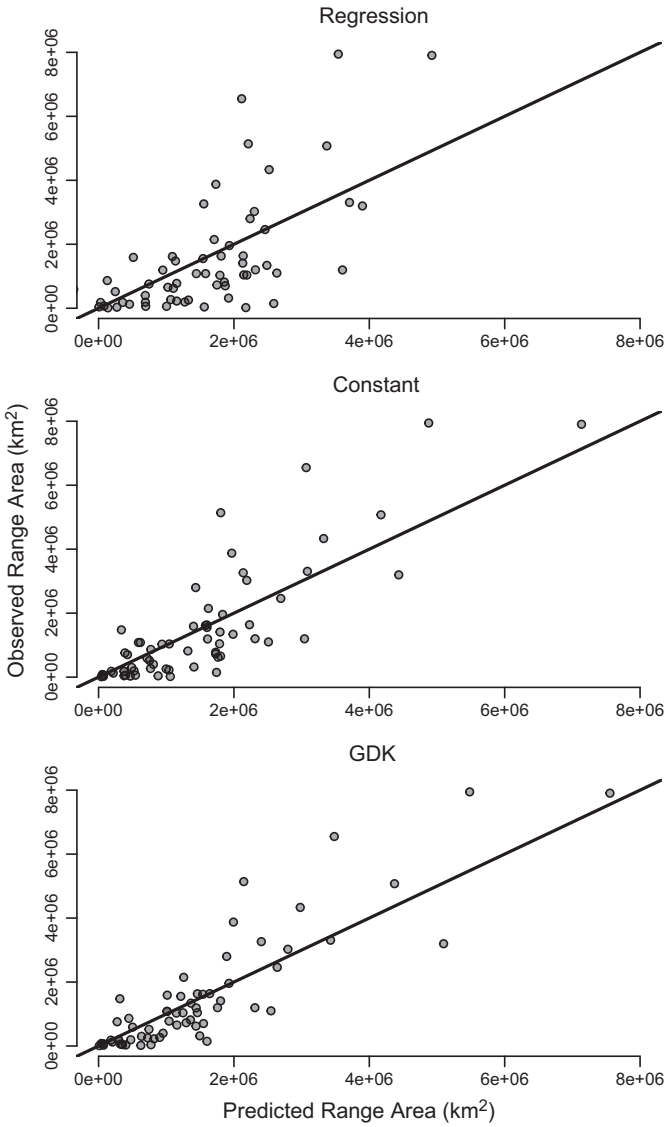

Figure 2 Predicted versus observed species range sizes for the analogous regression model (top, $R^2_{MSE} = 0.2837$; $R^2_{MSE} = 0.3072$ for a regression of pest radius instead of area), constant dispersal model (middle, $R^2_{MSE} = 0.6794$) and full generalised dispersal kernel model (bottom, $R^2_{MSE} = 0.7559$).

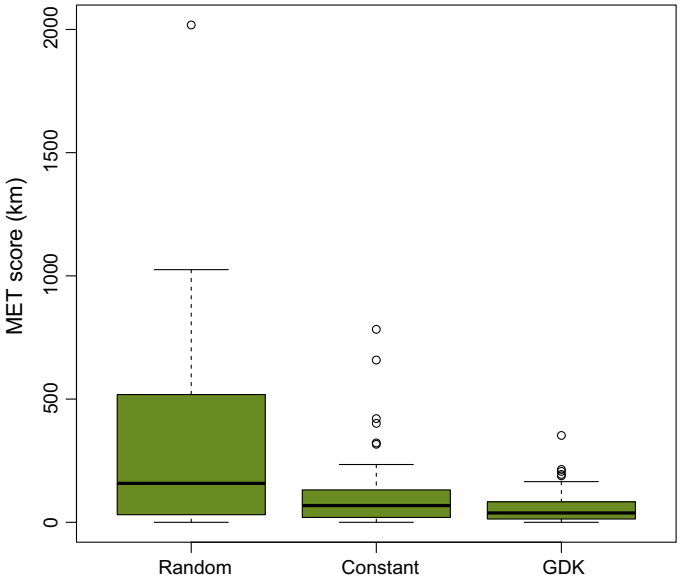

Figure 3 Distribution of minimum energy test scores across all species expected at random (left, mean = 207.70 km), and modelled for the constant dispersal (centre, mean = 113.07 km) and generalised dispersal kernel (right, mean = 68.11 km) models.

shown by the reduced outliers, Fig. 3). These mismatches occurred for both predicted presences as well as absences (see Fig. 4 for examples of individual fit, Fig. 5 for aggregated predictions and Fig. S2 for all individual species fits). We also found that MET was inversely related to the extent of spread ($r = -0.3406$, $P = 0.0059$).

DISCUSSION

Currently, most predictions of invasive species spread use species-specific models (Muirhead *et al*. 2006; Carrasco *et al*. 2010; Gagnon *et al*. 2015). Here, we showed that a common dispersal kernel can capture much of the variation in pest extent across all known damaging forest pests in the United States. As such, it appears that generalities are possible. Further, our GDK model has ramifications for invasion biology as a predictive science, including forecasting the spread of new invaders, a demonstration of predictive improvements using semi-mechanistic models and the incorporation of general predictors identified in the invasion literature (Leung *et al*. 2012). We discuss each in turn.

Generalities in ecology

Our results suggest that the rules governing the spread of pest species are sufficiently general to obtain strong quantitative macroecological predictions using a common model. As a line of future inquiry, a search for the underlying processes of these common spread patterns could yield fundamental insights into biological invasions and spread ecology. Parallels to such analyses have occurred across other fields of ecology, where predictive relationships have been discovered that transcend idiosyncratic species relationships and are discernable at large scales, such as those within the maximum entropy theory of ecology (Frank 2009) and the metabolic theory of ecology (West & Brown 2004). These relationships form so-called 'efficient' theories (Marquet *et al*. 2014).

It seems likely that over large scales, an efficient theory may exist for the dispersal of invasive species. The strong predictive power of the constant dispersal model was unexpected and interesting; it suggests that dispersal is largely occurring by one or more analogous spread mechanisms. We hypothesise that these analogous mechanisms are the various forms of human transport. It is well known that humans are important vectors of long-distance dispersal, but it appears that the various types of human transport (e.g. live plant trade, firewood movement) are occurring at roughly the same rate over large scales, allowing a constant dispersal parameter to capture the majority of variation in spread. In addition to this constant human transport, there also appears to be a small but

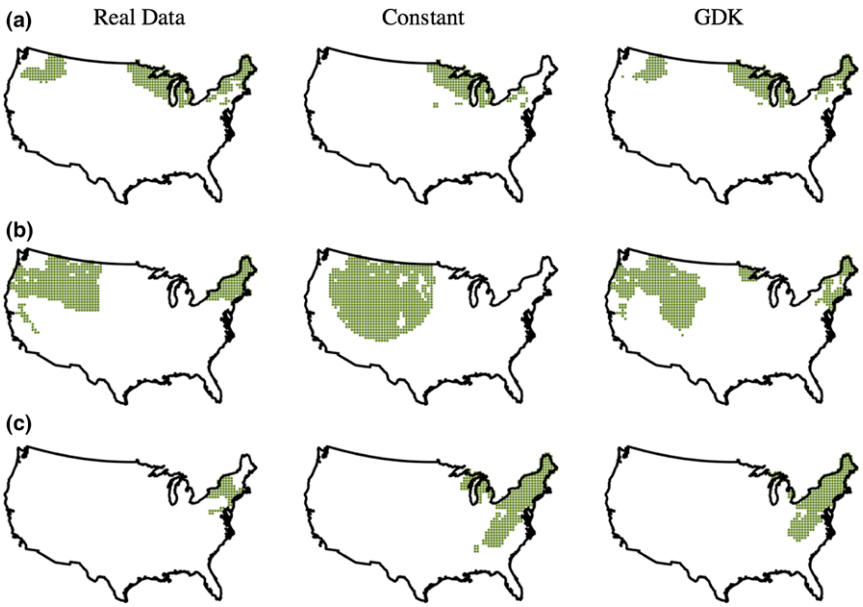

Figure 4 A selection of model predictions for individual species (a. *Coleophora laricella*, b. *Leucoma salicis*, c. *Nuculaspis tsugae*), showing the true presence data (left column), constant dispersal model predictions (centre column) and generalised dispersal kernel predictions (right column) as green areas. These distributions are only a small selection and do not show the full variation in model predictions across species, which are included in Fig. S2.

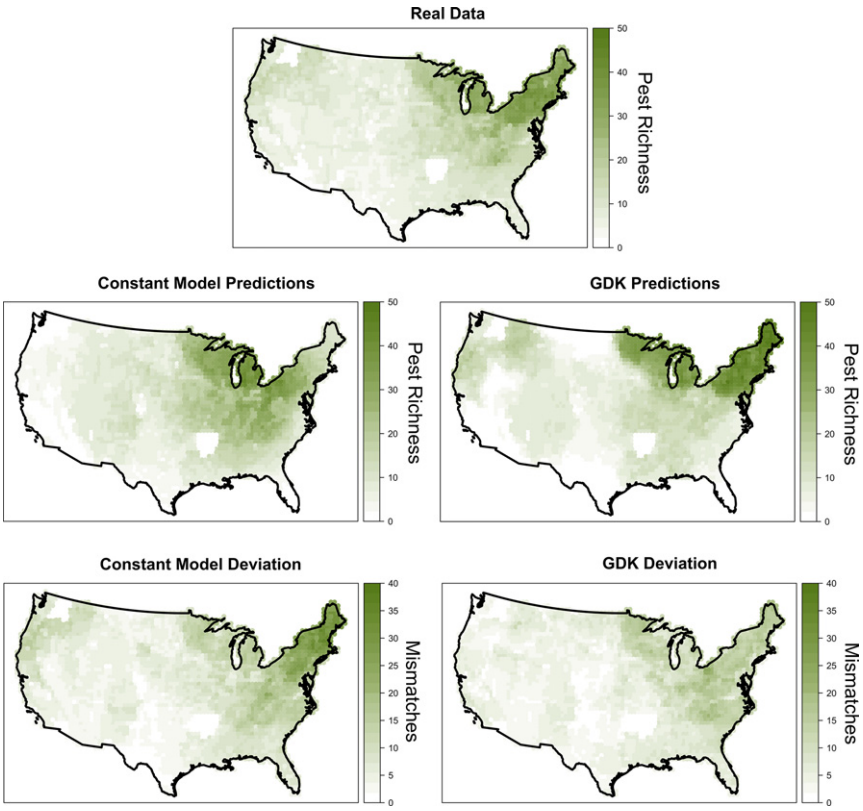

Figure 5 True observations of pest richness (top panel), predicted pest richness for the constant dispersal (middle left) and generalised dispersal kernel (GDK) (middle right) models, and the Euclidean distance or number of mismatched pest presences in the constant dispersal (bottom left), and GDK (bottom right) models. For the true observations and predicted richness maps, deeper green indicates higher richness. For the mismatch maps, deeper green indicates a higher degree of mismatch (false presences + false absences).

important component of preferential pest dispersal to areas of high population density, given that our GDK model predicts discrete patches of spread surrounding metropolitan areas.

Traditionally, dispersal has been conceptualised as the result of natural processes based on life history traits and habitat suitability as it relates to individual species' constraints. However, if analogous transport mechanisms and preferential pest dispersal to metropolitan areas occur, then human agency may be overshadowing traditional ecological mechanisms, such as natural flight capacity (Taylor *et al*. 2010), wind-driven dispersal (Aylor 1990) and community assembly mechanisms (e.g. Belyea & Lancaster 1999). More formally, we hypothesise that 'anthropogenic replacement' is occurring, wherein certain natural processes are essentially being overridden by anthropogenic ones, and that predictable generalities that operate across entire suites of species arise as a consequence of these processes' broad effects. The lack of explanatory power of life history parameters and preferential

dispersal into high human population density areas are arguably controversial findings, and further study is required to test the generality of such anthropogenic replacement. Additionally, we acknowledge that life history parameters are important for other stages of invasion (e.g. establishment, Forsyth *et al*. 2004; Bradie & Leung 2015), and could be important in different model formulations or components beyond dispersal distances. Nonetheless, these are important ideas, and replacement of natural processes may have ramifications beyond invasive species dispersal to other systems dominated by anthropogenic processes.

Forecasting spread of novel invaders

In concert with models that predict invader establishment in a country (e.g. Bradie & Leung 2015), our complementary analysis suggests that entry locations and host distributions will provide good predictors of pest spread, given that the GDK model explained 75.60% variation in pest range and had 74.30% locational accuracy, and that life history traits did not explain substantial variation in spread. The GDK results also indicate that future invasions to the United States will be characterised by preferential dispersal into areas of high human population density. Such forecasts allow for the refinement of pathway level analyses of invasion risk to include a spread component (e.g. the solid wood packaging materials pathway and introduction of wood borers, Haack 2006; Leung *et al*. 2014). Pests that have not been studied extensively can be included in these projections, given some knowledge of their host distributions, and information on establishment rates from the pathway in order to join this spread model to an establishment model (e.g. Aukema *et al*. 2010; Brockerhoff *et al*. 2014; Leung *et al*. 2014).

The importance of semi-mechanistic models

Several alternative models are available for predicting spread (Shigesada & Kawasaki 1997; Hastings *et al*. 2005), the simplest being a purely statistical approach, such as a general linear model of area as a function of time (Liebhold *et al*. 2013). We argue that the additional complexity of the semi-mechanistic dispersal kernel is well worth inclusion for making predictions. Comparing the simplest models in each, a constant dispersal kernel model essentially doubled the variation explained compared to a regression of pest area against time.

Further, our semi-mechanistic approach can easily incorporate spatial predictors of spread in predictive contexts, while it is less clear how general linear models can utilise spatial variables for prediction. One problem is that incorporating additional spatial variables requires the calculation of a single spatial value per pest in fitting (e.g. average human density across pest distribution). In a predictive context, where future distributions are unknown, these spatial predictors change as a pest spreads, which in turn affects their rate of spread, making these models impractical for prediction. Additionally, while regression models can relate time since discovery to the size of pest ranges, they do not predict spatial locations in their parameterisations of spread, thereby providing less information content than the GDK. Since these models have lesser utility, we did not test additional predictors of pest spread beyond time within a regression context.

Additionally, semi-mechanistic models have the potential to better account for issues of spatial autocorrelation and non-stationarity by replicating a spatially autocorrelated dispersal process and allowing for differential dispersal across different environments. Many statistical modelling approaches are unable to account for spatial autocorrelation and in particular, non-stationarity, where the effect of spatial autocorrelation varies across space (Dormann *et al*. 2007). Failing to account for these phenomena when present can result in model misspecification and invalid inferences and subsequent predictions relating to spatial data (Miller 2012).

General predictive factors

Invader life history, habitat invasibility and propagule pressure have been studied extensively in the invasion biology literature (Leung *et al*. 2012). We have shown here how these factors can be incorporated into an explicit dispersal model, and found that the predictors that were important and the magnitudes of effects were different than previous studies. In their review of the risk assessment literature, Leung *et al*. (2012) compiled over 200 models of the stages of species invasions used to predict risk. In contrast to our model, the majority of previously published models of the spatial distributions of invasive species have been largely formulated as models of pest establishment success (Inglis *et al*. 2006; Catford *et al*. 2011; Compton *et al*. 2012). Others have been formulated as gravity models (Gertzen *et al*. 2011; Potapov *et al*. 2011), while some studies have analysed the total richness of pests across space (Stohlgren *et al*. 2006; Liebhold *et al*. 2013; Iannone *et al*. 2015).

Although related, these various modelled processes and their patterns are subtly different. We modelled the factors that promote rapid spread of invasive species in general, rather than the factors influencing the establishment success of individual pest species. Further, in contrast to models of pest presence or richness that attempt to determine differential establishment across space, our model allowed predictor variables to modify the probability of dispersal across space and time, thereby influencing the level of propagule pressure reaching each cell based on the level of propagule pressure at surrounding cells and the associated predictor variables.

Given the difference in underlying phenomena modelled, it may not be surprising that the predictors and their magnitudes also differed between our study and previous results. We found that pest life history parameters were not important in our model, though previous models have found factors such as fruit size (Pyšek *et al*. 2009) and wind dispersal (Gassó *et al*. 2009) to be predictive of invasive plant distributions. It is unknown whether differences in our findings reflect differences in system (plant versus pest) or differences in invasion stage (establishment success versus spread). Regardless, our model's lack of pest life history traits enforces the importance of spatial factors for spread in this system, although there may be other important life history factors that were not considered.

Across previous studies of establishment, richness and gravity models, propagule pressure metrics tended to have higher

explanatory power than habitat invasibility metrics (Leung & Mandrak 2007; Catford *et al*. 2011), and these additional factors explained the majority of the variation in pest presences (Inglis *et al*. 2006; Compton *et al*. 2012). Conversely, we have found that a constant dispersal kernel model already explained the majority of the variation in pest spread (> 65%), and therefore that the unique explanatory power of habitat and propagule pressure was moderate (7.65% variation explained, 6.3% increase in locational accuracy), with habitat invasibility explaining more of the variation in pest spread than propagule pressure proxies. However, we suggest that our constant dispersal model is not independent of habitat invasibility and propagule pressure processes, but actually incorporates the parts of these processes that are consistent across space and time. As such, the aspects of propagule pressure and habitat invasibility that have constant spatiotemporal influences on dispersal are contained within the intercept in our constant dispersal model. This spatiotemporal invariance may differ depending on the system (cf. Leung & Mandrak 2007; Gertzen *et al*. 2011; Compton *et al*. 2012 which examined discrete lakes). Nonetheless, based on the variation explained by the constant dispersal kernel model, it appears as though the majority of their influence is constant across space and time, facilitating future predictions given an initial establishment.

Caveats and future directions

Scale-dependence occurs when a model's driving factors vary with the grain (spatial and/or temporal) at which it is fit (Pauchard & Shea 2006; Fridley *et al*. 2007). Our results are at the country scale where, importantly, long-distance anthropogenic dispersal may dominate over natural pest dispersal, possibly explaining why life history traits were not predictive. Hence, while we were able to strongly predict spread across counties, and across the United States, predicting local spread within counties requires additional models, where life history traits could potentially dominate the dispersal signal.

Scale-dependence can also occur when a model's driving factors and fit vary with temporal roughness. The observed correlation between MET score and spread extent is potentially consistent with temporal scale-dependence, as shorter spread time would be more affected by discrete decadal time units. However, sensitivity analyses using 5-year units yielded virtually identical results (Fig. S1). As an alternative *post hoc* explanation, MET was significantly lower (better) for species with introduction locations inside known host range ($t_{0.05\ (2),62} = 2.793$, $P = 0.0070$), suggesting that our lack of urban tree data is instead responsible for outliers.

The inability to predict pest presences outside of forested areas is a key shortcoming of our model. The positive influence of population density on pest spread further suggests that urban habitats have strong influences on pest dispersal. Clearly the collection of urban tree data at a large scale should be a priority. Additionally, our model structure assumed a single central introduction of each pest, though we know that several pests have had multiple independent introductions to the United States. Initiating spread from the host range centroid and limiting it to the natural host range likely worsens the estimation of early spread rates, as it adds a spatial mismatch in initial spread to the existing temporal mismatch from the date of establishment to that of first detection. Thus, the formulation of a detection model is an important future direction to understand this system mechanistically and alleviate the latter source of mismatch. Regardless, the strong predictive power of the model indicates robustness to these details.

We used a general negative exponential dispersal kernel in our model, though other dispersal models exist (e.g. 'fat-tailed' dispersal kernels Shigesada *et al*. 1995; Kot *et al*. 1996). However, several factors suggest that our model is sufficient to describe this system, including its strong predictive ability. Additionally, the breadth of durations of invasion within our dataset (1790–2008) suggests that we are able to capture both recent and longer-term invasion patterns. Finally, moderation of the dispersal kernel parameter provides an alternative formulation of long-distance dispersal. For instance, the inclusion of population density in our full dispersal model allows for urban centres to attract pests from distant locations. This is arguably a more process-driven formulation of long-distance dispersal than using fixed, alternative shapes for dispersal kernels, if we believe that humans are important vectors causing pest spread.

Our model was based on current and historical conditions. However, climate change could alter environmental suitability due to its influence on host tree species and on local abiotic conditions (Hellmann *et al*. 2008). Climatic variables such as temperature were not parameterised in our model. However, the most important factor for pest persistence is arguably the presence of viable hosts. In some cases, pests are instead limited by temperature [e.g. hemlock woolly adelgid (*Adelges tsugae*), which has reached its climatic limit, despite the presence of hosts to the north, Paradis *et al*. 2008], and these cases could contribute to the remaining error in the model. Additionally, climate change may influence human distributions, which, given their inclusion in our model, should also be forecasted. Finally, forest management could also alter the future distribution of tree species (e.g. which species are planted or cut), which will alter invasible host distributions. Likewise, land use change could alter forested land and urbanisation patterns, both of which would affect pest spread. In sum, forecasting into the future will require additional considerations and submodels, although the GDK can play a key role.

CONCLUSIONS

We have found that a single spread model for all invasive United States forest pests is predictive of both the extent and locational accuracy of pest distributions. This model provides a key element to forecast pest spread, thereby facilitating rapid responses to new pests. On a more fundamental level, the predictability across pest species suggests generality, and advances the possibility of a unified macroecological theory for invasive species spread by suggesting that common mechanisms underlie spread across species, beyond simple identification to the actual quantification of these mechanisms.

ACKNOWLEDGEMENTS

EH would like to thank L. Della Venezia, A. Sardain, V. Reed, A. Sellers, J. Bradie, K. Enciso, A. Pietrobon, D.

Nguyen, T. J. Davies, P. M. A. James, and the members of the McGill Stats-Bio Exchange group for their helpful discussion, L. Blackburn for help with acquiring data, and two anonymous reviewers for their valuable comments. This work was supported by an NSERC CGS-M fellowship awarded to EH and an NSERC Discovery grant to BL.

AUTHORSHIP

BL and EH designed the modelling framework and analyses. EH wrote the R scripts and performed all analyses. AL obtained the data and contributed to interpretation of results. All authors contributed to writing the manuscript.

DATA ACCESSIBILITY STATEMENT

The data supporting these results have been archived in Dryad: Hudgins E, Liebhold A, Leung B. Data from: Predicting the spread of all invasive forest pests in the United States. Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.75265.

REFERENCES

Aslan, B. & Zech, G. (2005). New test for the multivariate two-sample problem based on the concept of minimum energy. *J Stat. Comput. Sim.*, 75, 109–119.

Aukema, J.E., McCullough, D.G., Von Holle, B., Liebhold, A.M., Britton, K. & Frankel, S.J. (2010). Historical accumulation of nonindigenous forest pests in the continental United States. *Bioscience*, 60, 886–897.

Aylor, D.E. (1990). The role of intermittent wind in the dispersal of fungal pathogens. *Annu. Rev. Phytopathol.*, 28, 73–92.

Belyea, L.R. & Lancaster, J. (1999). Assembly rules within a contingent ecology. *Oikos*, 86, 402–416.

Bradie, J. & Leung, B. (2015). Pathway-level models to predict non-indigenous species establishment using propagule pressure, environmental tolerance and trait data. *J. Appl. Ecol.*, 52, 100–109.

Bradie, J., Chivers, C. & Leung, B. (2013). Importing risk: quantifying the propagule pressure–establishment relationship at the pathway level. *Divers. Distrib.*, 19, 1020–1030.

Brockerhoff, E.G., Kimberley, M., Liebhold, A.M., Haack, R.A. & Cavey, J.F. (2014). Predicting how altering propagule pressure changes establishment rates of biological invaders across species pools. *Ecology*, 95, 594–601.

Cadotte, M.W., Murray, B.R. & Lovett-Doust, J. (2006). Ecological patterns and biological invasions: using regional species inventories in macroecology. *Biol. Invasions*, 8, 809–821.

Carrasco, L.R., Mumford, J.D., MacLeod, A., Harwood, T., Grabenweger, G., Leach, A.W. *et al.* (2010). Unveiling human-assisted dispersal mechanisms in invasive alien insects: integration of spatial stochastic simulation and phenology models. *Ecol. Model.*, 221, 2068–2075.

Catford, J.A., Vesk, P.A., White, M.D. & Wintle, B.A. (2011). Hotspots of plant invasion predicted by propagule pressure and ecosystem characteristics. *Divers. Distrib.*, 17, 1099–1110.

Compton, T.J., deWinton, M., Leathwick, J.R. & Wadhwa, S. (2012). Predicting spread of invasive macrophytes in New Zealand lakes using indirect measures of human accessibility. *Freshw. Biol.*, 57, 938–948.

Dormann, C.F., M McPherson, J., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G. *et al.* (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30, 609–628.

Elton, C.S. (2000). *The ecology of Invasions by Animals and Plants, 2*. University of Chicago Press, Chicago.

ESRI (2011). *ArcGIS Desktop: Release 10*. Redlands, CA: Environmental Systems Research Institute.

Forsyth, D.M., Duncan, R.P., Bomford, M. & Moore, G. (2004). Climatic suitability, life-history traits, introduction effort, and the establishment and spread of introduced mammals in Australia. *Conserv. Biol.*, 18, 557–569.

Frank, S.A. (2009). The common patterns of nature. *J. Evol. Biol.*, 22, 1563–1585.

Fridley, J.D., Stachowicz, J.J., Naeem, S., Sax, D.F., Seabloom, E.W., Smith, M.D. *et al.* (2007). The invasion paradox: reconciling pattern and process in species invasions. *Ecology*, 88, 3–17.

Gagnon, K., Peacock, S.J., Jin, Y. & Lewis, M.A. (2015). Modelling the spread of the invasive alga *Codium fragile* driven by long-distance dispersal of buoyant propagules. *Ecol. Model.*, 316, 111–121.

Gassó, N., Sol, D., Pino, J., Dana, E.D., Lloret, F., Sanz-Elorza, M. *et al.* (2009). Exploring species attributes and site characteristics to assess plant invasions in Spain. *Divers. Distrib.*, 15, 50–58.

Gertzen, E., Leung, B. & Yan, N. (2011). Propagule pressure, Allee effects and the probability of establishment of an invasive species (*Bythotrephes longimanus*). *Ecosphere*, 2, 1–17.

Gilbert, M., Grégoire, J.C., Freise, J.F. & Heitland, W. (2004). Long-distance dispersal and human population density allow the prediction of invasive patterns in the horse chestnut leafminer Cameraria ohridella. *J. Anim. Ecol.*, 73, 459–468.

Haack, R.A. (2006). Exotic bark-and wood-boring Coleoptera in the United States: recent establishments and interceptions. *Can. J. For. Res.*, 36, 269–288.

Haack, R.A., Petrice, T.R. & Wiedenhoeft, A.C. (2010). Incidence of bark- and wood-boring insects in firewood: a survey at Michigan's Mackinac Bridge. *J. Econ. Entomol.*, 103, 1682–1692.

Hastings, A., Cuddington, K., Davies, K.F., Dugaw, C.J., Elmendorf, S., Freestone, A. *et al.* (2005). The spatial spread of invasions: new developments in theory and evidence. *Ecol. Lett.*, 8, 91–101.

Hellmann, J.J., Byers, J.E., Bierwagen, B.G. & Dukes, J.S. (2008). Five potential consequences of climate change for invasive species. *Conserv. Biol.*, 22, 534–543.

Iannone, B.V., Oswalt, C.M., Liebhold, A.M., Guo, Q., Potter, K.M., Nunez-Mir, G.C. *et al.* (2015). Region-specific patterns and drivers of macroscale forest plant invasions. *Divers. Distrib.*, 21, 1181–1192.

Inglis, G.J., Hurren, H., Oldman, J. & Haskew, R. (2006). Using habitat suitability index and particle dispersion models for early detection of marine invaders. *Ecol. Appl.*, 16, 1377–1390.

Kot, M., Lewis, M.A. & Van Den Driessche, P. (1996). Dispersal data and the spread of invading organisms. *Ecology*, 77, 2027–2042.

Leung, B. & Mandrak, N.E. (2007). The risk of establishment of aquatic invasive species: joining invasibility and propagule pressure. *Proc. Roy. Soc. B.*, 274, 2603–2609.

Leung, B., Roura-Pascual, N., Bacher, S., Heikkilä, J., Brotons, L., Burgman, M.A. *et al.* (2012). TEASIng apart alien species risk assessments: a framework for best practices. *Ecol. Lett.*, 15, 1475–1493.

Leung, B., Springborn, M.R., Turner, J.A. & Brockerhoff, E.G. (2014). Pathway-level risk analysis: the net present value of an invasive species policy in the US. *Front. Ecol. Environ.*, 12, 273–279.

Liebhold, A.M. (2012). Forest pest management in a changing world. *Int. J Pest Manage.*, 58, 289–295.

Liebhold, A.M., McCullough, D.G., Blackburn, L.M., Frankel, S.J., Von Holle, B. & Aukema, J.E. (2013). A highly aggregated geographical distribution of forest pest invasions in the USA. *Divers. Distrib.*, 19, 1208–1216.

Lockwood, J.L., Cassey, P. & Blackburn, T. (2005). The role of propagule pressure in explaining species invasions. *Trends Ecol. Evol.*, 20, 223–228.

Lovett, G.M., Weiss, M., Liebhold, A.M., Holmes, T.P., Leung, B., Lambert, K.F. *et al.* (2016). Nonnative forest insects and pathogens in

the United States: impacts and policy options. *Ecol. Appl.*, 26, 1437–1455.
Marquet, P.A., Allen, A.P., Brown, J.H., Dunne, J.A., Enquist, B.J., Gillooly, J.F. *et al.* (2014). On theory in ecology. *Bioscience*, 64, 701–710.
Miller, J.A. (2012). Species distribution models: spatial autocorrelation and non-stationarity. *Prog. Phys. Geog.*, 36, 681–692.
Morin, R.S., Liebhold, A.M. & Gottschalk, K.W. (2009). Anisotropic spread of hemlock woolly adelgid in the eastern United States. *Biol. Invasions*, 11, 2341–2350.
Muirhead, J.R., Leung, B., Overdijk, C., Kelly, D.W., Nandakumar, K., Marchant, K.R. *et al.* (2006). Modelling local and long-distance dispersal of invasive emerald ash borer Agrilus planipennis (Coleoptera) in North America. *Divers. Distrib.*, 12, 71–79.
Neubert, M.G. & Parker, I.M. (2004). Projecting rates of spread for invasive species. *Risk Anal.*, 24, 817–831.
Paradis, A., Elkinton, J., Hayhoe, K. & Buonaccorsi, J. (2008). Role of winter temperature and climate change on the survival and future range expansion of the hemlock woolly adelgid (*Adelges tsugae*) in eastern North America. *Mitigation Adaptat. Strateg. Glob. Chang.*, 13, 541–554.
Pauchard, A. & Shea, K. (2006). Integrating the study of non-native plant invasions across spatial scales. *Biol. Invasions*, 8, 399–413.
Potapov, A., Muirhead, J., Yan, N., Lele, S. & Lewis, M. (2011). Models of lake invasibility by *Bythotrephes longimanus*, a non-indigenous zooplankton. *Biol. Invasions*, 13, 2459–2476.
Pyšek, P., Krivanek, M. & Jarosik, V. (2009). Planting intensity, residence time, and species traits determine invasion success of alien woody species. *Ecology*, 90, 2734–2744.
Shigesada, N. & Kawasaki, K. (1997). *Biological Invasions: Theory and Practice*. Oxford University Press, UK.
Shigesada, N., Kawasaki, K. & Takeda, Y. (1995). Modeling stratified diffusion in biological invasions. *Am. Nat.*, 146, 229–251.
Siegert, N.W., McCullough, D.G., Liebhold, A.M. & Telewski, F.W. (2014). Dendrochronological reconstruction of the epicentre and early spread of emerald ash borer in North America. *Divers. Distrib.*, 20, 847–858.
Simberloff, D. (2009). The role of propagule pressure in biological invasions. *Annu. Rev. Ecol. Evol. Syst.*, 40, 81–102.
Simberloff, D. & Gibbons, L. (2004). Now you see them, now you don't! – population crashes of established introduced species. *Biol. Invasions*, 6, 161–172.
Skarpaas, O. & Shea, K. (2007). Dispersal patterns, dispersal mechanisms, and invasion wave speeds for invasive thistles. *Am. Nat.*, 170, 421–430.
Skellam, J.G. (1951). Random dispersal in theoretical populations. *Biometrika*, 38, 196–218.
Stohlgren, T.J., Barnett, D., Flather, C., Fuller, P., Peterjohn, B., Kartesz, J. *et al.* (2006). Species richness and patterns of invasion in plants, birds, and fishes in the United States. *Biol. Invasions*, 8, 427–447.
Taylor, R.A.J., Bauer, L.S., Poland, T.M. & Windell, K.N. (2010). Flight performance of *Agrilus planipennis* (Coleoptera: Buprestidae) on a flight mill and in free flight. *J. Insect Behav.*, 23, 128–148.
Tisseuil, C., Gryspeirt, A., Lancelot, R., Pioz, M., Liebhold, A.M. & Gilbert, M. (2015). Evaluating methods to quantify spatial variation in the velocity of biological invasions. *Ecography*, 39, 409–418.
Walter, J.A., Johnson, D.M., Tobin, P.C. & Haynes, K.J. (2015). Population cycles produce periodic range boundary pulses. *Ecography*, 38, 1200–1211.
West, G.B. & Brown, J.H. (2004). Life's universal scaling laws. *Phys. Today*, 57, 36–43.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

Editor, Regan Early
Manuscript received 29 August 2016
First decision made 12 October 2016
Manuscript accepted 3 January 2017