Research

Ecography 40: 1187–1197, 2017
doi: 10.1111/ecog.02194

Subject Editor: Matthew Fitzpatrick. Editor-in-Chief: Miguel Araújo. Accepted 25 August 2016

Accounting for imperfect observation and estimating true species distributions in modelling biological invasions

Thomas Mang, Franz Essl, Dietmar Moser, Gerhard Karrer, Ingrid Kleinbauer and Stefan Dullinger

T. Mang (http://orcid.org/0000-0001-5206-2981) (thomas_mang@univie.ac.at), D. Moser and I. Kleinbauer, Vienna Inst. for Nature Conservation and Analyses, Vienna, Austria. – F. Essl, S. Dullinger, TM and DM, Division of Conservation Biology, Vegetation Ecology and Landscape Ecology, Dept of Botany and Biodiversity Research, Univ. of Vienna, Vienna, Austria. FE also at: Environment Agency Austria, Vienna, Austria, and Centre for Invasion Biology, Dept of Botany and Zoology, Stellenbosch Univ., Matieland, South Africa. – G. Karrer, Inst. of Botany, Univ. of Natural Resources and Life Sciences, Vienna, Austria.

The documentation of biological invasions is often incomplete with records lagging behind the species' actual spread to a spatio-temporally heterogeneous extent. Such imperfect observation bears the risk of underestimating the already realised distribution of the invading species, misguiding management efforts and misjudging potential future impacts. In this paper, we develop a hierarchical modelling framework which disentangles the determinants of the invasion and observation processes, models spatio-temporal heterogeneity in detection patterns, and infers the actual, yet partly undocumented distribution of the species at any particular time. We illustrate the model with a case study application to the invasion of common ragweed *Ambrosia artemisiifolia* in Austria. The invasion part of the model reconstructs the historical spread of this species across a grid of ~ 6 × 6 km^2 cells as driven by spatio-temporal variation in physical site conditions, propagule production, dispersal, and 'background' introductions from unknown sources. The observation part models the detection of the species' occurrences based on heterogeneous sampling efforts, human population density, and estimated local invasion level. We fitted the hierarchical model using a Bayesian inference approach with parameters estimated by Markov chain Monte Carlo (MCMC). The actual spread of *A. artemisiifolia* concentrated on the climatically well-suited lowlands and was mainly driven by spatio-temporal propagule pressure from source cells with long-distance dispersal occurring rather frequently. Annual detection probabilities were estimated to vary between about 1 and up to 28%, depending mainly on sampling intensity. The model suggested that by 2005 about half of the actual distribution of the species was not yet documented. Our hierarchical model offers a flexible means to account for imperfect observation and spatio-temporal variability in detection efficiency. Inferences can be used to disentangle aspects of the invasion dynamics itself from patterns of data collection, develop improved future surveying schemes, and design more efficient invasion management strategies.

The human-mediated global movement of species has accelerated rapidly during recent decades (Hulme et al. 2009, Essl et al. 2011, van Kleunen et al. 2015). Naturalisation and subsequent spread of some alien species have already caused severe ecological and economic damage (Simberloff et al. 2013). Managing these threats and reducing potential costs may be greatly facilitated by distinguishing areas that are likely to be invaded by a particular species at a particular time (Epanchin-Niell and Hastings 2010, Richter et al. 2013). However, such forecasts require knowledge of the species' biology as well as appropriate quantitative tools for analysing spatio-temporal range dynamics. In this context, modelling the spread of invasive alien species has been an active field of interplay between theoretical and applied ecological research for a long time (Skellam 1951, Kot et al. 1996, Hastings et al. 2005, Travis et al. 2011, Mouquet et al. 2015).

A frequently applied approach to modelling the range expansion of alien species is based on the statistical analysis of past changes in the species' recorded geographical distribution patterns (Kadoya and Washitani 2010, Smolik et al. 2010, Marion et al. 2012, Václavík et al. 2012). However, invasive spread often results in rapid changes of species distributions. As a consequence, the documentation of species spread is often incomplete, in particular where species are inconspicuous (Aikio et al. 2010). Imperfect detection introduces two sorts of errors into spatio-temporal distribution data of range-expanding species: firstly, sites may appear currently uninvaded although the species is already present, resulting in false absences; and secondly, the documented time of first colonisation of a site may lag considerably behind the actual invasion of the species. Besides potentially biasing parameter estimates of invasive spread models, these errors bear the risk of underestimating the current extent of

the invasion, misguiding management efforts, and misjudging potential future impacts (Stanaway et al. 2011).

Accounting for observation errors and properly integrating them into quantitative analyses remains an understudied issue in ecology. Of relevance here are hierarchical modelling frameworks, where the biological process of interest and its human observation are represented by separate model layers (Cressie et al. 2009, Latimer et al. 2009, Royle and Dorazio 2009, Pagel and Schurr 2012). In this setup, the observation layer models sampling schemes and represents the translation from the actual yet incompletely observed biological process to the available data. Several models of this kind have been successfully applied to invasive species spread and focus on handling different types of incomplete data (Wikle 2003, Bled et al. 2011, Catterall et al. 2012, Ibáñez et al. 2014, Broms et al. 2016).

Distribution data of invading species may not only be spatially sparse and temporally lagged, but available species records may also have been collected non-systematically or under different, inconsistent observation schemes (Delisle et al. 2003, Ruiz and Carlton 2003, Meinesz 2007). In general, the larger the spatio-temporal extent of a study, the more likely it will be that species records stem from different sources that do not share a standardised sampling design (Feeley and Silman 2011, Berec et al. 2015). Put another way, over large geographical extents and long time frames detection rates of species may show considerable heterogeneity and thereby induce multiple data biases. Nevertheless, making use of information from all available sources is likely to improve both our mechanistic understanding of the invasion dynamics and the accuracy of predictions of the respective species' current or future distribution.

Integrating detection uncertainty into spread models has several further potential advantages. First, factors promoting invasion and detection, respectively, can easily become confounded, so models that explicitly account for the observation process can help to distinguish recording bias patterns from the actual biological process of interest. Secondly, research interests may focus on the observation process itself, for example to assess spatio-temporal variability in detection efficiency of particular surveying schemes. Last but not least, coupling invasion and observation processes allows for probabilistic inferences of the species' actual current distribution, and differences between (inferred) actual and documented distributions may guide future surveys or suggest a reassessment of management efforts.

In this paper we develop and demonstrate a generally applicable framework for the integration of imperfect detection into spatio-temporal modelling of biological invasions. Based on spatially incomplete and temporally lagged presence-only data, and basic information on the different species sampling schemes, the model estimates parameters of both the invasion and observation processes and infers the unknown actual distribution. We illustrate the framework by modelling the invasion of common ragweed *Ambrosia artemisiifolia* (Asteraceae) in Austria, central Europe, as a case study.

Methods

Ambrosia artemisiifolia invasion in central Europe

Ambrosia artemisiifolia is an annual herb native to North America. The first European records stem from the 19th century, but the species did not start to spread and naturalise until the first decades of the 20th century (Essl et al. 2015 and references therein). Range expansion accelerated considerably in the late 20th century. In temperate Europe, the species is so far mostly limited to lowland habitats. Climatic conditions, particularly temperature, have been demonstrated to be an important invasion filter (Chapman et al. 2014). The species mainly thrives in disturbed open habitats and agricultural fields. Long-distance transport of its propagules occurs via trade of contaminated goods and vehicle traffic (Essl et al. 2009). The *A. artemisiifolia* invasion is of considerable public health concern as the species' wind-blown pollen is highly allergenic (Smith et al. 2013). Seeds are persistent in the soil seed bank for several decades and a site once colonised hence remains occupied for a long time (Fumanal et al. 2008).

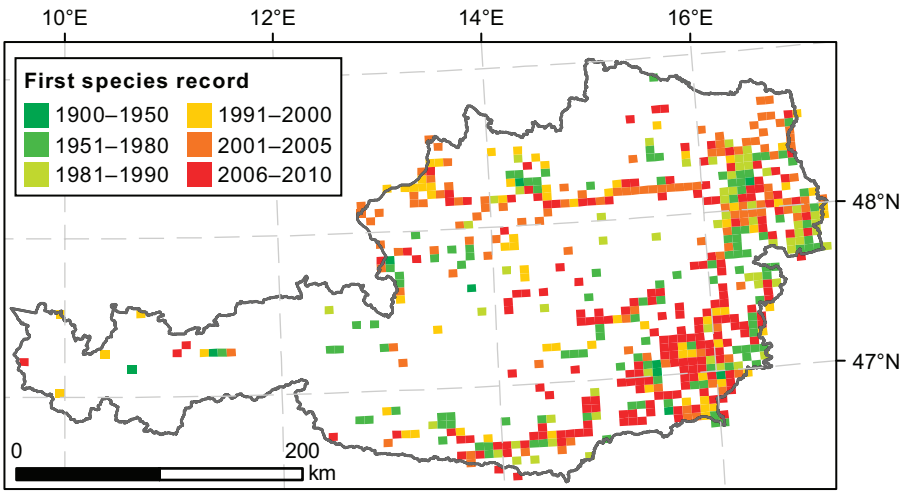

Figure 1. The chronology of first records of *Ambrosia artemisiifolia* in Austria in a 5′ × 3′ (~ 6 × 6 km²) grid.

Study area and period

We modelled the historical spread of *A. artemisiifolia* in Austria (central Europe) (Fig. 1) at annual time steps using the period 1900–2005 for model fitting and the period 2006–2010 for model validation. Austria is a land-locked country about 84 000 km² in size, and covers a range of climates with lowland continental areas in the east and south-east, sub-oceanic climate towards the north, and montane and alpine zones in the central and western parts. In the model, the country's territory is represented by a lattice system with 2626 cells of 5 × 3 geographical minutes (~ 6 × 6 km²), corresponding to the grid of the Central European Floristic Mapping Project (Niklfeld 1998).

Species data

Species records (= spatio-temporal occurrence data) were compiled from many different sources (e.g. floristic mapping projects, floristic publications, major herbaria, unpublished records of the authors and of colleagues) and mapped to one of the 2626 grid cells. Observation dates (years) were extracted from the original source (Fig. 1).

Sampling intensity varied both in time and across the study region. Firstly, observations of the species increased pronouncedly with the onset of the Central European Floristic Mapping Project in 1970 (Niklfeld 1998). Secondly, several scientific projects studying aspects of the *A. artemisiifolia* invasion have increased detection rates in selected regions and years, particularly recently (Karrer et al. 2011), resulting in a distinct recent rise in the total number of available records as well as in the number of cells documented as invaded (Fig. 2). However, record numbers increased much faster than the numbers of cells documented as invaded because most of these recent projects focused on geographically restricted regions. For the model developed here we used the earliest record of the species in a cell as the relevant date of detection. Spatio-temporal variation in sampling intensity is, among other factors, modelled in the part of the hierarchical model dedicated to the observation process.

Hierarchical model

Spatial heterogeneity in the conditions at different sites is represented by assigning each of the 2626 grid cells attributes describing the physical environment, in particular climate and land use, as well as putative determinants of species detection rates, such as human population density and sampling intensity. All these attributes may vary in each cell over the modelling period.

An initially uninvaded grid cell must undergo two subsequent changes of its status to yield a species record: in the first step it is invaded, and in the second step the local occurrence of the species is detected (and recorded). In our model, we assume that detection may lag behind the actual invasion into a particular cell by an unlimited period of time. Consequently, the first record of the species from a particular cell simultaneously represents the latest possible time by which the cell has become invaded; and cells where the species has not been documented so far may nevertheless be already invaded. The hierarchical model is therefore comprised of invasion and observation process layers which separately model the time of initial cell invasion and first detection, respectively. The two layers are linked by the cell's invasion time, which is the random variate of the invasion process and concurrently the start time of the observation process. In our model we assume that the species persists in a cell once invaded.

Actual cell invasion times are unknown. They hence become latent variables which are estimated during model fitting. Bayesian inference approaches (Gelman et al. 2004) with data augmentation are particularly well suited for fitting models with a large number of unknown quantities. Let $\boldsymbol{x}$ be the actual invasion times of all cells, $\boldsymbol{y}$ be the times when occurrences in these cells were first detected, $\boldsymbol{\theta}$ be all model parameters used in the invasion process, and $\boldsymbol{\delta}$ be all model parameters used in the observation process; then the

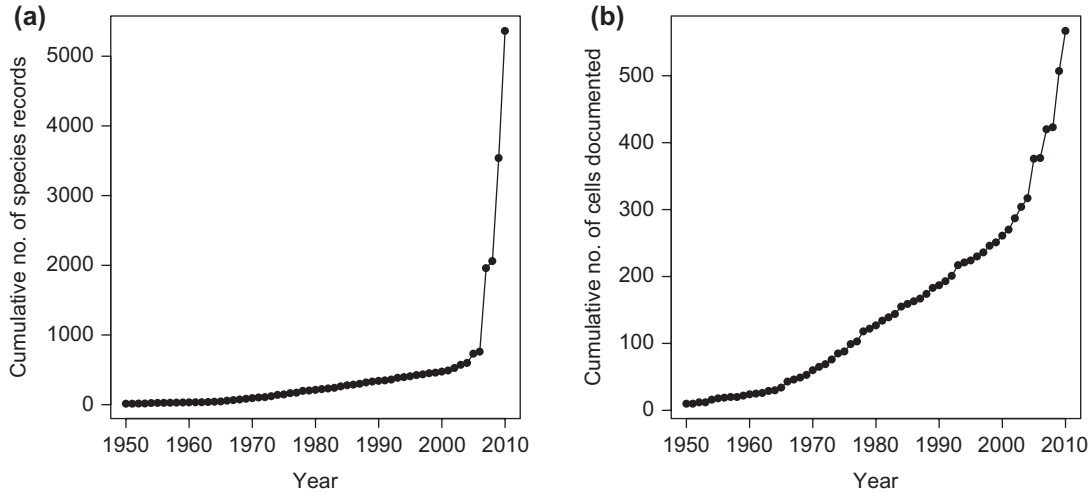

Figure 2. (a) Cumulative number of *Ambrosia artemisiifolia* species records in Austria until 2010, and (b) cumulative number of grid cells documented as invaded.

Bayesian posterior distribution of the hierarchical model is given by

$$p(\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{x} \mid \boldsymbol{y}) = \frac{p(\boldsymbol{\theta}, \boldsymbol{\delta})\, p(\boldsymbol{x} \mid \boldsymbol{\theta})\, p(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\delta})}{p(\boldsymbol{y})} \quad (1)$$

where $p(\boldsymbol{\theta}, \boldsymbol{\delta})$ is the (joint) prior distribution of $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$, $p(\boldsymbol{x} \mid \boldsymbol{\theta})$ is the invasion process likelihood, $p(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\delta})$ is the observation process likelihood, and $p(\boldsymbol{y})$ is the marginal distribution of $\boldsymbol{y}$ (a normalising constant). The unknown actual invasion times are regarded as additional parameters in the posterior distribution (Marion et al. 2012).

For this model, presence-only data (documented occurrences) are sufficient but precise dates associated with each species record are assumed (precise in terms of the chosen model unit). In general, however, distribution data such as those extracted from published and unpublished manuscripts or online databases may only provide information on eligible detection time intervals (e.g. between start of a sampling campaign and publication of records). We hence also developed an extension to Eq. 1 using interval censored detection times (Supplementary material Appendix 1).

Invasion process

To represent the actual (yet incompletely observed) invasion process, we model the cells' initial invasion times (Cook et al. 2007). Populations in invaded cells produce and disperse propagules. The resulting propagule pressure on yet uninvaded recipient cells is modelled as a cell-specific invasion risk function which defines the invasion time distribution. The magnitude of this risk depends on the fecundity of populations in source cells, i.e. those cells already invaded at a particular time, the spatial distance of a potential recipient cell to the source cells, and the environmental suitability of a recipient cell. In our model, we further use 'background' introductions which cause invasion risk independent of invaded source cells in the study area, for example due to human-mediated spread over very long distances (including introduction from the species' native range).

We represent relative environmental cell suitability as a log-linear model component using

$$S_i(t) = \mathrm{e}^{\mathbf{v}_{i,t}\boldsymbol{\beta}}\, \frac{a_i}{\bar{a}} \quad (2)$$

where $\mathbf{v}_{i,t}$ is the vector of environmental attributes of cell i at time t, $\boldsymbol{\beta}$ is the vector of associated weighting parameters (to be estimated), a_i is the cell's terrestrial area, and $\bar{a}$ is the mean terrestrial area across all cells. We used the following six spatio-temporal environmental variables: mean temperature and total precipitation of the growing season (April–October); the proportion of cropland area and of urban area; and the length (scaled relative to area) of motorway and railway networks (Dullinger et al. 2009). Non-climate variables were log-transformed to improve symmetry and reduce the impact of outlier values. All environmental variables were standardised and the magnitude of the associated parameter estimate is hence representative of the relative effect size. For details on data sources and data processing see Supplementary material Appendix 2.

For dispersal from source cells we started with a leptokurtic, one-dimensional kernel function from the power-law family of form $f_{1\mathrm{D}}(d_{j,i}) = {d_{j,i}}^{-\alpha}$, where $d_{j,i}$ is the Euclidean distance between the centroids of cells j and i (in km), and α is a shape parameter. This function was then projected into two-dimensional space and normalised (Supplementary material Appendix 3) to yield the two-dimensional kernel function $f_{2\mathrm{D}}(d_{j,i})$.

By modelling both invaded cells and background introductions as propagule sources, and taking account of environmental suitability, for recipient cell i the invasion risk as a function of time is:

$$g_i(t) = \left[\sum_{j \in \boldsymbol{\Omega}(t-1)} R_j(t-1)\, \eta\, f_{2\mathrm{D}}\left(d_{j,i}\right) + \lambda + \lambda_{\mathrm{b}}\, \mathbf{1}\left(t = t_{\mathrm{s}}\right) \right] S_i(t) \quad (3)$$

where $\boldsymbol{\Omega}(t-1)$ is the set of cells already invaded at the given time (i.e. potential source cells; for *A. artemisiifolia* an offset of one year applies as their seeds were produced in the year preceding germination in the recipient cell), $R_j(t-1)$ is a source cell's estimated invasion level at that time, the parameter η is the rate of produced and dispersed propagules, and the parameters λ and λ_{b} define the rate of background introductions. Abundance data of local *A. artemisiifolia* populations were not available, but as the annual species is capable of rapid initial population growth we used $R_j(t-1) = S_j(t-1)$ as an approximation. Even though this certainly overestimates cells' invasion levels in the first years after colonisation, it reflects long-term average differences in invasion levels among cells which differ in environmental suitability. Specifically, as the number of propagules produced in a cell is assumed to be directly proportional to its invasion level for the annual species *A. artemisiifolia*, the purpose of this term here is hence to quantify differences in source strength among invaded cells. The parameter λ defines the generic background introduction rate, and λ_{b} provides an additional boost applicable only to the model start time, t_{s} (via the indicator function $\mathbf{1}$). This boost corresponds de facto to the invasion risk accumulated up to when modelling begins. Background introductions are of particular importance during early invasion when sparse introductions into the study area may have dominated over autochthonous local spread. Unless the location and establishment time of the initial invasion focus (or foci) are specified it is actually also a compulsory model component since it allows for a non-zero invasion risk even in the absence of any source cell and hence enables initial invasion of the study area. In numerical terms, Eq. 3 defines the instantaneous rate of invasion at time t provided that a cell has so far remained uninvaded, and will be used to parameterise the invasion time distribution.

Observation process

In addition to the species' spread itself, we model the observation of this spread in terms of the delay between the species' initial invasion of a cell and the first documented record of the species from this cell. We use a spatio-temporal detectability function to define the distribution of this delay (see below). For invaded cell i detectability as a function of time is:

$$h_i(t) = \gamma\, R_i(t)^{\varphi}\, \mathrm{e}^{\mathbf{m}_{i,t}\cdot\boldsymbol{\rho}} \quad (4)$$

where the parameter γ defines the base detection rate, the parameter φ relates modelled invasion level to detectability, $\mathbf{m}_{i,t}$ is the vector of sampling intensity attributes of the cell at time t, and $\boldsymbol{\rho}$ is the vector of associated weighting parameters (to be estimated). We used the following three spatio-temporal variables as attributes: whether a year was prior to the start of the Central European Floristic Mapping Project in 1970 (a binary indicator variable); the degree of intensified sampling for *A. artemisiifolia*, measured continuously in [0,1] as a cell's area share of political districts with intensified sampling; and the human population density in and around a given cell and year (log-transformed and standardised). For details on data sources and data processing see Supplementary material Appendix 2. In numerical terms, Eq. 4 defines the instantaneous rate of detection at time t provided that the species' occurrence in a cell has so far remained undetected, and will be used to parameterise the detection time distribution.

Waiting-time distribution for invasion and detection

Invasion and detection times can be equivalently expressed as waiting-times since model start time until invasion and since first invasion into a cell until detection, respectively. The invasion risk and detectability functions described above quantify these waiting-times by parameterising the generalised waiting-time distribution (Arens et al. 2009). This distribution is central to survival analysis where objects become exposed to some hazard, and the random variable of interest is the residual time from initial hazard exposure until failure occurs. The magnitude of this hazard imposed at a given time is quantified by the hazard function. In our model, grid cells are initially (= model start time) uninvaded but exposed to an invasion risk; the invasion risk function $g_i(t)$ from Eq. 3 is therefore the hazard function for a cell's invasion time distribution. Where this invasion occurs, populations in the respective cells are subsequently detectable; the detectability function $h_i(t)$ from Eq. 4 is therefore the hazard function for a cell's detection time distribution.

Let W be a general waiting-time random variable, $b(t)$ be its associated hazard function, and t_0 be the start time of hazard exposure; then the probability density function of a continuous random variable W is given by

$$f_W\left(w;b(t),t_0\right) = b(w)\,\mathrm{e}^{-\int_{t_0}^{w} b(t)\,\mathrm{d}t} \tag{5}$$

(for $w \geq t_0$). Conveniently, the hazard function $b(t)$ only needs to be ≥ 0, otherwise no constraints are imposed onto its shape. Specifically, at any time it can be increasing, decreasing, or constant. For $b(t) = 0$ no hazard is imposed by definition and hence the event of interest must not occur at time t. Higher values of $b(t)$ correspond to earlier expected failure time (i.e. in our model earlier invasion or detection time).

Equation 5 uses continuous time. In practice, however, invasion or detection times may frequently be measured in discrete units (e.g. years, like in our case study application). Let t_{s_k} and t_{e_k} denote the start and end time of the k-th discrete modelling sub-period (e.g. a given year), respectively, then the probability mass function of a discrete random variable W is given by

$$f_W\left(w_k;b(t),t_0\right) = \mathrm{P}\left(W = w_k\right) = \mathrm{e}^{-\int_{t_0}^{t_{s_k}} b(t)\,\mathrm{d}t} - \mathrm{e}^{-\int_{t_0}^{t_{e_k}} b(t)\,\mathrm{d}t} \tag{6}$$

As Eq. 6 depends exclusively on the integral of the hazard function it makes no difference whether this hazard function is specified as a truly continuous function or as a step function stating an averaged value per sub-period. The step function approach may greatly facilitate alignment with environmental data and/or ease the computing implementation, and was hence also used in our case study application (i.e. the invasion risk and detectability functions changed values at annual intervals and thus effectively also had a discrete resolution).

In survival analysis, the complementary cumulative distribution function is referred to as the survival function as it states the probability that an object could retain its original state until time w, and is given by

$$S_W\left(w;b(t),t_0\right) = \mathrm{P}\left(W > w;b(t),t_0\right) = \mathrm{e}^{-\int_{t_0}^{w} b(t)\,\mathrm{d}t} \tag{7}$$

(for $w \geq t_0$). Evaluated for the model end time t_e, that is $w = t_e$, the survival function hence states the probability that a cell in the study area remained uninvaded (invasion process) or the probability that an invaded cell remained undetected (observation process).

Invasion process likelihood

The invasion process likelihood considers the invasion times of all cells: let $\boldsymbol{\Psi}_\mathbf{m}$ be the set of cells which became invaded during the modelling period; then for each cell in $\boldsymbol{\Psi}_\mathbf{m}$ the likelihood assesses its specific invasion time, x_i, by using either the probability density function given by Eq. 5 (if invasion times are modelled in continuous time) or the probability mass function given by Eq. 6 (for discrete time). Conversely, let $\boldsymbol{\Psi}_\mathbf{e}$ be the complementary set of cells still uninvaded at the model end time; then for each cell in $\boldsymbol{\Psi}_\mathbf{e}$ the likelihood assesses the probability that it has not been invaded yet by using the survival function given by Eq. 7. Let t_s be the model start time, t_e be the model end time, and let each cell's invasion risk function be dependent on all invasion process parameters $\boldsymbol{\theta}$ (as above), then the invasion process likelihood is therefore:

$$p\left(\boldsymbol{x}|\boldsymbol{\theta}\right) = \prod_{i\in\boldsymbol{\Psi}_\mathbf{m}} f_{X_i}\left(x_i;g_i(t),t_s\right) \times \prod_{i\in\boldsymbol{\Psi}_\mathbf{e}} S_{X_i}\left(t_e;g_i(t),t_s\right) \tag{8}$$

Observation process likelihood

The observation process likelihood considers the detection times of all cells which were invaded until the model end time: let $\boldsymbol{\Phi}_\mathbf{d}$ be the subset of these cells for which the species has also been recorded during the modelling period; then for each cell in $\boldsymbol{\Phi}_\mathbf{d}$ the likelihood assesses the cell's specific detection time, y_i, using either the probability density function given by Eq. 5 (if detection times are modelled in continuous time) or the probability mass function given by Eq. 6 (for discrete time). Conversely, let $\boldsymbol{\Phi}_\mathbf{u}$ be the complementary subset of cells which were invaded but for which no record

is available by the model end time; then for each cell in $\mathbf{\Phi_u}$ the likelihood assesses the probability that detection has not occurred yet by using the survival function given by Eq. 7. Let each invaded cell's detectability function be dependent on all observation process parameters $\boldsymbol{\delta}$ (as above), then the observation process likelihood is therefore:

$$p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\delta}) = \prod_{i\in\mathbf{\Phi_d}} f_{Y_i}\left(y_i; h_i(t), x_i\right) \times \prod_{i\in\mathbf{\Phi_u}} S_{Y_i}\left(t_e; h_i(t), x_i\right) \quad (9)$$

Model fitting using MCMC

We fitted the hierarchical Bayesian model using Markov chain Monte Carlo (MCMC) (Gelman et al. 2004, Brooks et al. 2011). Vague (marginal) prior distributions were used for all model parameters (Supplementary material Appendix 4, Table A1) and the posterior distribution was hence virtually exclusively determined by the data. To obtain the results we sampled 100 000 iterations of a single chain after a burn-in period of 10 000 iterations. This required about 77 h of execution time on a workstation with an Intel® Core i7-3930K processor and an AMD Tahiti device. Convergence was assessed by the Gelman–Rubin diagnostic using three separate chains. For further MCMC details see Supplementary material Appendix 4.

Estimated invasion and observation process parameters were summarised by the (marginal) posterior distribution median and the 95% (central) credible interval. For parameters with a null-hypothesis value located in the interior of the corresponding distribution support, the Bayesian kind of significance testing assesses if the credible interval overlaps the null-hypothesis value. In our model this applied to $\boldsymbol{\beta}$ and $\boldsymbol{\rho}$, with 0 as the null-hypothesis value for a neutral effect.

Estimated actual invasion times were summarised by the probability that a cell had already been invaded by a given reference year, calculated as the proportion of MCMC iterations in which x_i was less than or equal to this reference year.

Model validation

For model validation we used 10 000 draws from the posterior distribution (every tenth iteration), including cells' estimated actual invasion states by 2005 (= the last year of the MCMC fitting period), to simulate both the invasion and observation process further until 2010 (= the end of the validation period). We then calculated the 1) invasion probability and 2) detection probability by 2010 as the proportion of simulation runs in which a cell was predicted to be invaded and predicted to be detected as invaded, respectively, by 2010. For all cells which were not documented as already invaded by 2005 (2236 cells in total) we then compared these probabilities against the occurrence data collected during the validation period (2006–2010) using the area under the receiver operating characteristic (ROC) curve, AUC. The AUC can take on values from 0 to 1, where 1 corresponds to perfect classification, 0.5 to a match as expected by random choice, and 0 to perfect misclassification.

Simulation study

We further evaluated the model's inference characteristics and ability to deliver correct parameter estimates by means of a simulation study. We first performed simulations of the combined invasion and observation processes. Subsequently, we used the simulated species occurrence documentations (i.e. the species records data set generated by the simulation of the observation process) to estimate model parameters and the actual invasion state using MCMC. We tested three different scenarios of detection, and thus also available spread documentation: for the first and second scenario we used rather low annual baseline detection probabilities for invaded cells (as determined by γ) of 1 and 2.5%, respectively; both scenarios further assumed detectability to be lowered by a factor of 2.5 for years prior to 1970, and increased by a factor of 5 for intensified sampling. The third scenario assumed higher detectability throughout since 1970 by using a 10% annual baseline detection probability. For individual cells detection rates varied with environmental attributes, and were up to about three times higher than these baseline figures under exceptionally good conditions. Our parameterisations were chosen to represent (common) low-intensity sampling campaigns. A full list of all parameter values applied in the simulations is provided in Supplementary material Appendix 5, Table A2. For each of the three scenarios we performed 12 stochastic model realisations and sampled 50 000 MCMC iterations per realisation, discarding the first 10 000 for burn-in. To reduce computational costs, simulations started in 1950 (with λ_b of Eq. 3 increased accordingly). For each scenario and parameter we 1) averaged the median estimate and relative bias (difference between median estimate and true value, divided by true value) across the 12 stochastic model realisations, and 2) assessed the number of stochastic model realisations for which the 95% credible interval contained the true value.

Results

Simulation study

The simulation study demonstrated that the model accurately estimates parameters (Supplementary material Appendix 5, Table A2). The only parameter substantially overestimated was λ. In addition, under the lower detection rates (scenarios 1 and 2) parameters of environmental suitability were slightly underestimated and η and φ slightly overestimated. Nevertheless, the credible intervals contained the true value for both scenarios 2 and 3 with the expected frequency. Only for the scenario of lowest detectability (scenario 1) did the credible intervals contain the true value somewhat less often than expected, indicating that for this extreme scenario the prior distribution had a noticeable impact. Accuracy generally improved with increasing detection rates (and hence more cells documented as invaded). The bias in λ was therefore considerably reduced in scenario 2, and further reduced in scenario 3, showing that an improved documentation of the actual invasion sequence considerably helps to separate the role of sparse background introductions from subsequent autochthonous local spread.

For all three scenarios the true number of cells invaded was well re-captured by our model, with scenarios 2 and 3 showing deviations of only a few percent (scenario 1: on average 331.1 cells documented as invaded, 779.6 cells truly invaded and 866.2 estimated by the model; scenario 2: 517.0 cells documented, 791.3 cells truly invaded and 826.8 estimated by the model; scenario 3: 699.2 cells documented, 806.1 cells truly invaded and 792.8 estimated by the model).

Ambrosia artemisiifolia invasion process

Except for the proportion of cropland area, all variables used to characterise a cell's environmental suitability to *A. artemisiifolia* invasion were significant (Table 1; for graphical summaries of the posterior distribution see Supplementary material Appendix 6, Fig. A1). Climate variables were more important than measures of land use and human disturbance, with temperature being by far the most influential predictor. Consequently, the lowland regions in the east and central north of Austria were identified as most susceptible to invasion (Fig. 3a, b). Invasion susceptibility was intermediate in valleys of the Alps, and minimal in the high mountain areas. Spatial environmental variability strongly dominated over temporal variability (Fig. 3b; Supplementary material Appendix 6, Fig. A2).

During the earlier decades of the invasion, background introductions caused the establishment of multiple, spatially rather scattered invasion foci (Supplementary material Appendix 6, Fig. A3). The background introduction rate is approximately equal in magnitude to the propagule pressure from a single invaded neighbour cell with favourable environmental conditions elevating source strength. Consequently, the long-term invasion pattern is driven by an exponential acceleration as more and more cells become invaded and hence function as additional sources of further local spread. The dispersal kernel parameter estimate suggests considerable uncertainty with respect to dispersal distances but long-distance dispersal occurs rather frequently (Table 1, Fig. 4).

Ambrosia artemisiifolia observation process

Prior to the onset of the Central European Floristic Mapping Project in 1970 detectability was lower than post 1970 by a factor of about 2.9 (credible interval [CI]: 1.8–4.7) (Table 1). Human population density in and around a cell significantly increased detectability. However, by far the most important effect was intensified sampling for *A. artemisiifolia* in particular areas and years which increased detectability by a factor of about 7.9 (CI: 5.2–11.8). By contrast, modelled invasion level had only a modest effect on detectability (Table 1).

Translating the detectability function of cells realised to be invaded into annual detection probabilities yielded a very low annual detection probability of only 1.4 percent (CI: 0.8–2.5) before 1970, and an increase to 4.1 percent (CI: 2.5–7.3) afterwards. Under intensified sampling, however, detection probabilities were about 28.0 percent (CI: 17.2–48.6). Average expected delays between actual invasion time and the first species record therefore ranged from a few years under intensified recent sampling efforts up to decades (pre-1970).

Estimated actual spread and model validation

According to the model, about 1.8 (CI: 1.4–2.3) times more cells were already invaded by 2005 than were documented by species records (Fig. 5). Indeed, by 1990 it is likely that the same number of cells had already been invaded as were documented by 2005. Most cells with suggested undocumented species occurrences are concentrated in the climatically suitable lowlands in the east and south-east of Austria

Table 1. Parameter estimates of the invasion and observation process. Significance tests apply only to environmental suitability and sampling intensity parameters, with significant results marked by *.

		95% credible interval	
Parameter	Median	Lower	Upper
Invasion process			
Environmental suitability parameters			
$\beta_{temperature}$*	1.49	1.06	1.96
$\beta_{precipitation}$*	0.39	0.26	0.53
$\beta_{cropland\ area}$	–0.15	–0.32	0.01
$\beta_{urban\ area}$*	0.29	0.10	0.48
$\beta_{motorways}$*	0.10	0.01	0.19
$\beta_{railways}$*	0.19	0.06	0.31
Dispersal, α	0.27	0.03	0.68
Propagule production rate, η	0.0091	0.0039	0.019
Background introduction rate, λ	0.000080	0.000011	0.00021
Background introduction start boost, λ_b	0.0027	0.0006	0.0074
Observation process			
Detection rate, γ	0.029	0.016	0.053
Detection dependence on invasion level, φ	0.056	0.005	0.20
Sampling intensity parameters			
$\rho_{pre\text{-}1970}$*	–1.05	–1.55	–0.57
$\rho_{intensified\ sampling}$*	2.07	1.65	2.47
$\rho_{human\ population\ density}$*	0.29	0.07	0.60

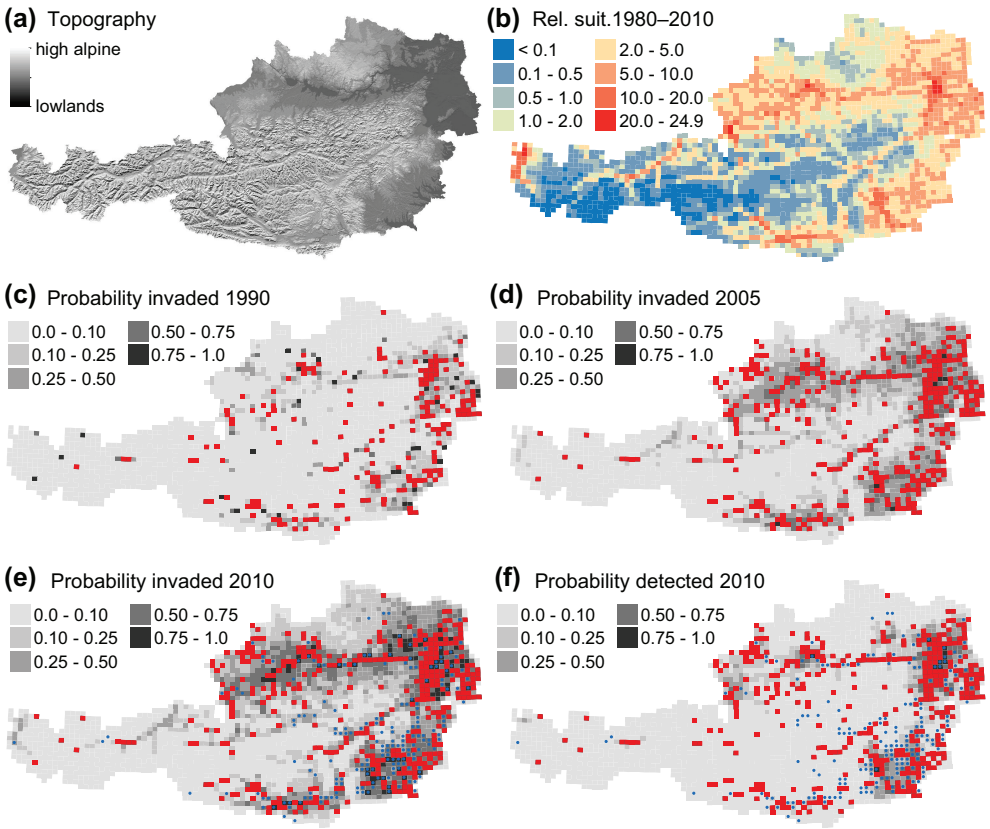

Figure 3. (a) Topography of Austria. (b) Relative environmental suitability of grid cells for *Ambrosia artemisiifolia* invasion during the recent decades 1980–2010. (c) Model-estimated probabilities of grid cells having been invaded by *A. artemisiifolia* by 1990, and (d) by 2005 (= last year for MCMC fitting). Cells in which the species occurrence has actually been documented by records up to the given year are shown in red. (e) Probabilities from model validation of cells being invaded by 2010 (= last year of the validation period), and (f) being detected as invaded by 2010. Cells in which the species occurrence has actually been documented before 2006 (= first year of the validation period) are shown in red, and cells with occurrence first documented during the validation period are marked by blue points.

and near cells where the species had already been recorded (Fig. 3c, d; for more snapshot maps see Supplementary material Appendix 6, Fig. A3).

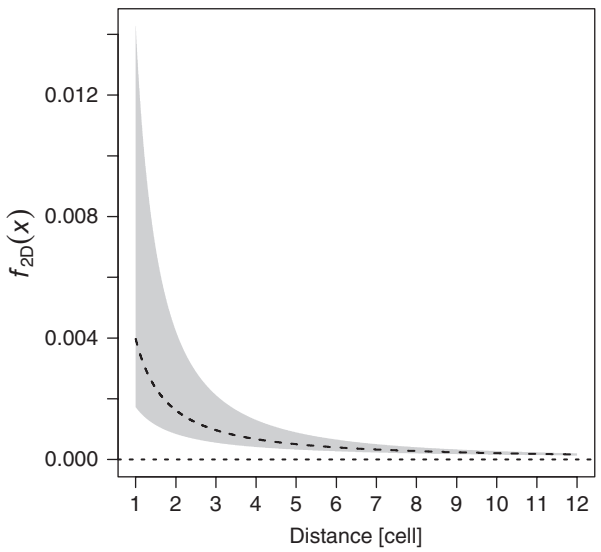

Figure 4. The kernel function for dispersal from invaded source to uninvaded recipient cells. The dashed line shows the median estimate of the shape parameter α (cf. Table 1), and the shaded area the 95% credible interval.

Comparing the (new) occurrence data from the validation period against the modelled probabilities of being 1) invaded and 2) detected as invaded by 2010 (Fig. 3e, f) yielded an AUC value of 0.87 and 0.85, respectively, thus showing good congruence between the model and the data. Many of these new records stem from cells which, according to the model, had likely already been invaded by 2005.

Discussion

For most larger-scale biological invasions, accurate and complete spatio-temporal distribution data are rarely available (Delisle et al. 2003, Chauvel et al. 2006) as species recording frequently suffers from detection errors and biases (Rocchini et al. 2011, Chen et al. 2013). In this paper we show, however, that numerical techniques like hierarchical modelling with estimation of latent invasion states, together with the advance of computer power that enables an application of these techniques to larger study systems, provide efficient means for analysing spatio-temporal invasion patterns despite imperfect detection. Our model follows a statistical approach in which parameters of both the invasion and its observation are estimated based on historical species records and limited information on the spatio-temporal variation of observation intensity. Using waiting-time random variables

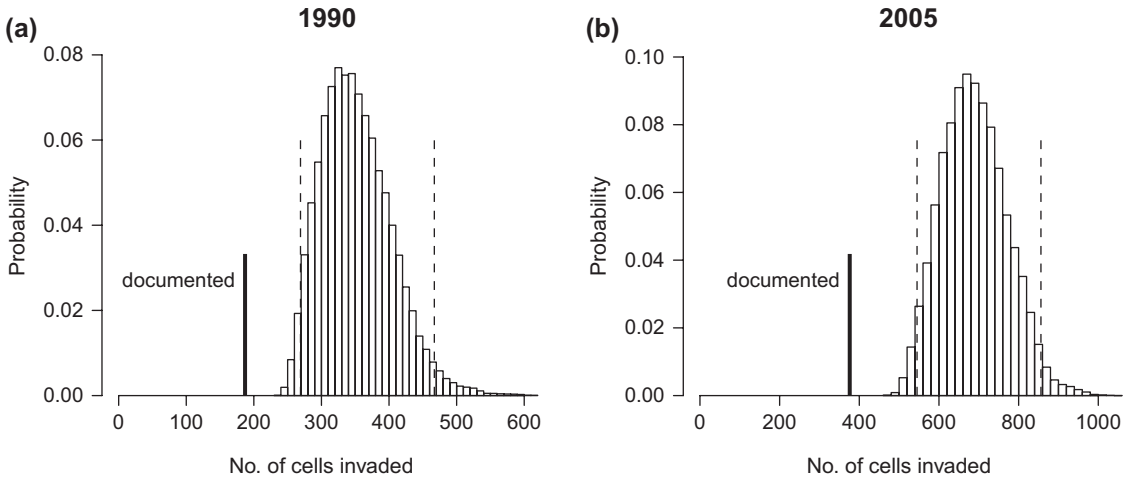

Figure 5. Posterior distributions of the number of grid cells (out of 2626 cells in total) estimated to be invaded by *Ambrosia artemisiifolia* by (a) 1990, and (b) 2005. Dashed lines mark the 95% credible interval, and the solid line marks the number of cells documented as invaded according to species records up to the given year.

for both processes eases the statistical definition of this combined modelling approach. Effectively, our hierarchical modelling approach relaxes data requirements since dated presence-only records are sufficient for parameterisation. Delayed recording is reflected by considering dates as upper (= latest) boundaries of possible actual invasion times, and by assigning an uncertain occurrence status to cells lacking a species record. As a trade-off, however, the model assumes persistent occurrence of the invading species in a cell after a cell's successful invasion, and inferences are contingent on the assumptions made about the detectability of these occurrences. Moreover, considering records as outcomes of the observation process implies that precise parameter estimation of the underlying invasion process requires a rather high number of such records. Based on our simulation study results we suggest that for a comparable study system the number of recorded cells should be at least 300. This data requirement will usually be met for regions with a good documentation tradition of species distributions, such as many European countries with floristic mapping projects. Data sets are also likely to be larger for invading species to which historically higher attention has been paid because they have a prominent socio-economic or environmental impact – which are also those species for which robust inferences on spread dynamics and actual distribution may be most critical.

Invasion and observation process

A reliable reconstruction of the actual spread pattern requires that both the invasion and the observation process are properly specified. Put another way, the invasion process layer should include the relevant drivers of species spread like the factors determining propagule pressure and site suitability, and the observation process layer should reflect the sampling schemes that were actually applied to collect the available species records. In our case study application, the model suggests that the invasion of *A. artemisiifolia* was strongly determined by environmental heterogeneity. In particular, climatic conditions represented an effective filter of colonisation, and accidental human movement of contaminated goods is suggested as the main driver of long-distance seed redistribution. These results corroborate earlier findings on the spread of *A. artemisiifolia* in different parts of Europe (see Essl et al. 2015 for a summary).

Detection odds of invaded cells were estimated statistically from the spatio-temporal sequence of species records and a number of putative determinants, or 'predictors', of detection efficiency. Nevertheless, the detectability function that we use is flexible and allows for the integration of many different possible surveying schemes for species: for example, if surveys were conducted with low intensity, yet in a spatio-temporally homogeneous way, a constant function would be sufficient to reflect observation lag times; as another example, periodically intensified sampling without spatial bias can be modelled by assigning the function higher values during these periods. Finally, a detectability function approaching infinity represents the special case of a complete survey for a given cell and time, where the lack of a species record corresponds to a confirmed absence and effectively places a lower boundary constraint on the cell's actual invasion time.

In our case study, the model suggests specific surveys undertaken in particular regions during particular years as most important for the detection of *A. artemisiifolia* occurrences. While these surveys delivered a large amount of species records, their strong spatio-temporal bias resulted in a highly skewed data set. Such biased observation patterns are probably characteristic for many documented invasions because species records frequently correlate with preferential sampling of selected locations due to easy accessibility, higher attractiveness, or location near research centres, field sites or living places of committed amateurs (Dennis and Thomas 2000, Reddy and Dávalos 2003, Romo et al. 2006, Merckx et al. 2011, Aikio et al. 2012, Yañez-Arenas et al. 2014). As a corollary, the actual biological process of interest and its observation may easily become confounded. In the case of invasive plants this is particularly likely because many of them prefer habitats characterised by high human disturbance frequency and/or intensity (Chytrý et al. 2008)

and for which human-driven propagule pressure is therefore also often high (Hulme et al. 2008, Pyšek et al. 2009). These sites are, simultaneously, those where detection odds may be particularly high because they are easily accessible and more frequently visited than more remote places. Consequently, easier detectability might contribute to the putative higher occurrence frequency in these habitats (Chytrý et al. 2008) to a certain yet unknown extent. The presented hierarchical model offers a tool for separating the possible effect of an observation bias from the actual invasion preferences of a species. For the weed species *A. artemisiifolia* we have shown that local human presence fosters invasion by providing disturbed habitats, but at the same time also increases recording chances. Differentiation of these processes might become even more important in the future with a potentially increasing use of data from citizen science projects for observing, analysing (and managing) alien species invasions (Dickinson et al. 2012, Crall et al. 2015, Lin et al. 2015): while citizen science might be a powerful tool to increase observation density, the resulting data are likely to be biased in one way or the other because the 'sampling scheme' will neither follow a systematic nor a random pattern.

Estimated actual spread

Fitting the hierarchical model to spatio-temporal observation data implies an estimation of observation lag times and hence also the unknown actual invasion times ($\boldsymbol{x}$). These estimates can subsequently be used to infer the actual historical and current distribution of the species (as opposed to the observed one) and identify unrecognised invasion hotspots. Based on these inferences, surveys to fill gaps in the knowledge about the species' current distribution (Hirzel and Guisan 2002, Guisan et al. 2006), or assessments of management measures (Richter et al. 2013), can potentially be designed in an efficient way. In our case study, estimated undetected species occurrences actually matched well with new species records from subsequent years. Ideally, inferences from such models, expert knowledge and applied field work can be combined to develop more efficient invasion management strategies (Yemshanov et al. 2010, Richter et al. 2013) in a similar way to such collaborative efforts in conservation science (Doswald et al. 2007, Fourcade et al. 2013, Guisan et al. 2013).

Limitations

Though generally applicable and powerful, the presented hierarchical model is also subject to practical limitations. Firstly, the need to estimate the cells' actual invasion times implies considerable computational costs. The time to fit a model certainly depends on the study system's dimensionality, model specification and programming implementation, but can realistically extend to days or even weeks. Secondly, the high number of model parameters (the joint set of $\boldsymbol{\theta}$, $\boldsymbol{\delta}$ and $\boldsymbol{x}$) implies a certain risk of undetected MCMC convergence failure within the finite number of iterations. Consequently, a critical assessment of results, ideally supplemented by expert knowledge, is a crucial step in model development. And finally, a full model validation for an empirical invasion is very difficult as accurate species data (from perfect detection) are typically unavailable.

Acknowledgements – This project received financial support from the Climate and Energy Fund and was carried out within the framework of the 'ACRP' Program (RAG-Clim, B068662). TM was the recipient of a DOC-fellowship (22636) of the Austrian Academy of Sciences at the Vienna Inst. for Nature Conservation and Analyses. We are indebted to Anna Mclvor for improving the English.

References

Aikio, S. et al. 2010. Lag-phases in alien plant invasions: separating the facts from the artefacts. – Oikos 119: 370–378.
Aikio, S. et al. 2012. The vulnerability of habitats to plant invasion: disentangling the roles of propagule pressure, time and sampling effort. – Global Ecol. Biogeogr. 21: 778–786.
Arens, T. et al. 2009. Ergänzungen und Vertiefungen zu Arens et al., Mathematik. – Spektrum Akademischer Verlag.
Berec, L. et al. 2015. Designing efficient surveys: spatial arrangement of sample points for detection of invasive species. – Biol. Invasions 17: 445–459.
Bled, F. et al. 2011. Hierarchical modeling of an invasive spread: the Eurasian collared-dove *Streptopelia decaocto* in the United States. – Ecol. Appl. 21: 290–302.
Broms, K. M. et al. 2016. Dynamic occupancy models for explicit colonization processes. – Ecology 97: 194–204.
Brooks, S. et al. 2011. Handbook of Markov chain Monte Carlo. – Chapman and Hall/CRC.
Catterall, S. et al. 2012. Accounting for uncertainty in colonisation times: a novel approach to modelling the spatio-temporal dynamics of alien invasions using distribution data. – Ecography 35: 901–911.
Chapman, D. S. et al. 2014. Phenology predicts the native and invasive range limits of common ragweed. – Global Change Biol. 20: 192–202.
Chauvel, B. et al. 2006. The historical spread of *Ambrosia artemisiifolia* L. in France from herbarium records. – J. Biogeogr. 33: 665–673.
Chen, G. et al. 2013. Imperfect detection is the rule rather than the exception in plant distribution studies. – J. Ecol. 101: 183–191.
Chytrý, M. et al. 2008. Habitat invasions by alien plants: a quantitative comparison among Mediterranean, subcontinental and oceanic regions of Europe. – J. Appl. Ecol. 45: 448–458.
Cook, A. et al. 2007. Bayesian inference for the spatio-temporal invasion of alien species. – Bull. Math. Biol. 69: 2005–2025.
Crall, A. W. et al. 2015. Citizen science contributes to our knowledge of invasive plant species distributions. – Biol. Invasions 17: 2415–2427.
Cressie, N. et al. 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. – Ecol. Appl. 19: 553–570.
Delisle, F. et al. 2003. Reconstructing the spread of invasive plants: taking into account biases associated with herbarium specimens. – J. Biogeogr. 30: 1033–1042.
Dennis, R. L. H. and Thomas, C. D. 2000. Bias in butterfly distribution maps: the influence of hot spots and recorder's home range. – J. Insect Conserv. 4: 73–77.
Dickinson, J. L. et al. 2012. The current state of citizen science as a tool for ecological research and public engagement. – Front. Ecol. Environ. 10: 291–297.
Doswald, N. et al. 2007. Testing expert groups for a habitat suitability model for the lynx *Lynx lynx* in the Swiss Alps. – Wildl. Biol. 13: 430–446.

Dullinger, S. et al. 2009. Niche based distribution modelling of an invasive alien plant: effects of population status, propagule pressure and invasion history. – Biol. Invasions 11: 2401–2414.
Epanchin-Niell, R. S. and Hastings, A. 2010. Controlling established invaders: integrating economics and spread dynamics to determine optimal management. – Ecol. Lett. 13: 528–541.
Essl, F. et al. 2009. Changes in the spatio-temporal patterns and habitat preferences of *Ambrosia artemisiifolia* during its invasion of Austria. – Preslia 81: 119–133.
Essl, F. et al. 2011. Socioeconomic legacy yields an invasion debt. – Proc. Natl Acad. Sci. USA 108: 203–207.
Essl, F. et al. 2015. Biological Flora of the British Isles: *Ambrosia artemisiifolia*. – J. Ecol. 103: 1069–1098.
Feeley, K. J. and Silman, M. R. 2011. Keep collecting: accurate species distribution modelling requires more collections than previously thought. – Divers. Distrib. 17: 1132–1140.
Fourcade, Y. et al. 2013. Confronting expert-based and modelled distributions for species with uncertain conservation status: a case study from the corncrake (*Crex crex*). – Biol. Conserv. 167: 161–171.
Fumanal, B. et al. 2008. Seed-bank dynamics in the invasive plant, *Ambrosia artemisiifolia* L. – Seed Sci. Res. 18: 101–114.
Gelman, A. et al. 2004. Bayesian data analysis. – Chapman and Hall/CRC.
Guisan, A. et al. 2006. Using niche-based models to improve the sampling of rare species. – Conserv. Biol. 20: 501–511.
Guisan, A. et al. 2013. Predicting species distributions for conservation decisions. – Ecol. Lett. 16: 1424–1435.
Hastings, A. et al. 2005. The spatial spread of invasions: new developments in theory and evidence. – Ecol. Lett. 8: 91–101.
Hirzel, A. and Guisan, A. 2002. Which is the optimal sampling strategy for habitat suitability modelling. – Ecol. Model. 157: 331–341.
Hulme, P. E. et al. 2008. Grasping at the routes of biological invasions: a framework for integrating pathways into policy. – J. Appl. Ecol. 45: 403–414.
Hulme, P. E. et al. 2009. Will threat of biological invasions unite the European Union? – Science 324: 40–41.
Ibáñez, I. et al. 2014. Integrated assessment of biological invasions. – Ecol. Appl. 24: 25–37.
Kadoya, T. and Washitani, I. 2010. Predicting the rate of range expansion of an invasive alien bumblebee (*Bombus terrestris*) using a stochastic spatio-temporal model. – Biol. Conserv. 143: 1228–1235.
Karrer, G. et al. 2011. Ausbreitungsbiologie und Management einer extrem allergenen, eingeschleppten Pflanze – Wege und Ursachen der Ausbreitung von Ragweed (*Ambrosia artemisiifolia*) sowie Möglichkeiten seiner Bekämpfung. – BMLFUW (Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft; Österreich).
Kot, M. et al. 1996. Dispersal data and the spread of invading organisms. – Ecology 77: 2027–2042.
Latimer, A. M. et al. 2009. Hierarchical models facilitate spatial analysis of large data sets: a case study on invasive plant species in the northeastern United States. – Ecol. Lett. 12: 144–154.
Lin, Y.-P. et al. 2015. Uncertainty analysis of crowd-sourced and professionally collected field data used in species distribution models of Taiwanese moths. – Biol. Conserv. 181: 102–110.
Marion, G. et al. 2012. Parameter and uncertainty estimation for process-oriented population and distribution models: data, statistics and the niche. – J. Biogeogr. 39: 2225–2239.
Meinesz, A. 2007. Methods for identifying and tracking seaweed invasions. – Bot. Mar. 50: 373–384.
Merckx, B. et al. 2011. Null models reveal preferential sampling, spatial autocorrelation and overfitting in habitat suitability modelling. – Ecol. Model. 222: 588–597.
Mouquet, N. et al. 2015. Review: predictive ecology in a changing world. – J. Appl. Ecol. 52: 1293–1310.
Niklfeld, H. 1998. Mapping the flora of Austria and the eastern Alps. – Revue Valdôtaine d'Histoire Naturelle 51: 53–62.
Pagel, J. and Schurr, F. M. 2012. Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. – Global Ecol. Biogeogr. 21: 293–304.
Pyšek, P. et al. 2009. Planting intensity, residence time, and species traits determine invasion success of alien woody species. – Ecology 90: 2734–2744.
Reddy, S. and Dávalos, L. M. 2003. Geographical sampling bias and its implications for conservation priorities in Africa. – J. Biogeogr. 30: 1719–1727.
Richter, R. et al. 2013. Spread of invasive ragweed: climate change, management and how to reduce allergy costs. – J. Appl. Ecol. 50: 1422–1430.
Rocchini, D. et al. 2011. Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. – Prog. Phys. Geogr. 35: 211–226.
Romo, H. et al. 2006. Identifying recorder-induced geographic bias in an Iberian butterfly database. – Ecography 29: 873–885.
Royle, J. A. and Dorazio, R. M. 2009. Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities. – Elsevier Academic Press.
Ruiz, G. M. and Carlton, J. T. 2003. Invasion vectors: a conceptual framework for management. – In: Ruiz, G. M. and Carlton, J. T. (eds), Invasive species: vectors and management strategies. Island Press, pp. 459–504.
Simberloff, D. et al. 2013. Impacts of biological invasions: what's what and the way forward. – Trends Ecol. Evol. 28: 58–66.
Skellam, J. G. 1951. Random dispersal in theoretical populations. – Biometrika 38: 196–218.
Smith, M. et al. 2013. Common ragweed: a threat to environmental health in Europe. – Environ. Int. 61: 115–126.
Smolik, M. G. et al. 2010. Integrating species distribution models and interacting particle systems to predict the spread of an invasive alien plant. – J. Biogeogr. 37: 411–422.
Stanaway, M. A. et al. 2011. Hierarchical Bayesian modelling of early detection surveillance for plant pest invasions. – Environ. Ecol. Stat. 18: 569–591.
Travis, J. M. J. et al. 2011. Improving prediction and management of range expansions by combining analytical and individual-based modelling approaches. – Methods Ecol. Evol. 2: 477–488.
Václavík, T. et al. 2012. Accounting for multi-scale spatial autocorrelation improves performance of invasive species distribution modelling (iSDM). – J. Biogeogr. 39: 42–55.
van Kleunen, M. et al. 2015. Global exchange and accumulation of non-native plants. – Nature 525: 100–103.
Wikle, C. K. 2003. Hierarchical Bayesian models for predicting the spread of ecological processes. – Ecology 84: 1382–1394.
Yañez-Arenas, C. et al. 2014. Predicting species' abundances from occurrence data: effects of sample size and bias. – Ecol. Model. 294: 36–41.
Yemshanov, D. et al. 2010. Detection capacity, information gaps and the design of surveillance programs for invasive forest pests. – J. Environ. Manage. 91: 2535–2546.

Supplementary material (Appendix ECOG-02194 at <www.ecography.org/appendix/ecog-02194>). Appendix 1–6.