Global Ecology and Biogeography, (Global Ecol. Biogeogr.) (2012) **21**, 1126–1136

Global Ecology and Biogeography
A Journal of Macroecology

MACROECOLOGICAL
METHODS

# Invasive species distribution models – how violating the equilibrium assumption can create new insights

Laure Gallien[1]*, Rolland Douzet[2], Steve Pratte[1], Niklaus E. Zimmermann[3] and Wilfried Thuiller[1]

[1]*Laboratoire d'Ecologie Alpine, CNRS UMR 5553, Université Joseph Fourier, BP 53, 38041 Grenoble Cedex 9, France,* [2]*Station Alpine Joseph Fourier, CNRS UMS 2925, Université Joseph Fourier, BP 53, 38041 Grenoble Cedex 9, France,* [3]*Landscape Dynamics, Swiss Federal Research Institute for Forest, Snow and Landscape Research, 8903 Birmensdorf, Switzerland*

*Correspondence: Laure Gallien, Laboratoire d'Ecologie Alpine, CNRS UMR 5553, Université Joseph Fourier, BP 53, 38041 Grenoble Cedex 9, France.
E-mail: laure.gallien@gmail.com

## ABSTRACT

**Aim** Two core assumptions of species distribution models (SDMs) do not hold when modelling invasive species. Invasives are not in equilibrium with their environment and niche quantification and transferability in space and time are limited. Here, we test whether combining global- and regional-scale data in a novel framework can overcome these limitations. Beyond simply improving regional niche modelling of non-native species, the framework also makes use of the violation of regional equilibrium assumptions, and aims at estimating the stage of invasion, range filling and risk of spread in the near future for 27 invasive species in the French Alps.

**Innovation** For each invader we built three sets of SDMs using a committee averaging method: one global model and two regional models (a conventional model and one using the global model output to weight pseudo-absences). Model performances were compared using the area under the receiver operating characteristic curve, the true skill statistic, sensitivity and specificity scores. Then, we extracted the predictions for observed presences and compared them to global and regional models. This comparison made it possible to identify whether invasive species were observed within or outside of their regional and global niches.

**Main conclusions** This study provides a novel methodological framework for improving the regional modelling of invasive species, where the use of a global model output to weight pseudo-absences in a regional model significantly improved the predictive performance of regional SDMs. Additionally, the comparison of the global and regional model outputs revealed distinct patterns of niche estimates and range filling among the species. These differences allowed us to draw conclusions about the stage of invasion and the risk of spread in the near future, which both correspond to experts' expectations. This framework can be easily applied to a large number of species and is therefore useful for control of biological invasions and eradication planning.

**Keywords**

**Adaptation, biological invasions, colonization, ecological niche, equilibrium, invasion stage, non-native plant species, spatial scale.**

## INTRODUCTION

Invasive plant species pose significant challenges with regard to managing and maintaining indigenous biodiversity in natural ecosystems (Olden *et al.*, 2004). Given that once introduced species become established they are often extremely difficult to eradicate (Rejmánek *et al.*, 2005), preventing their introduction is by far the most cost-effective form of management. To this end, a range of modelling tools have been developed in order to understand the drivers of species invasions and project the potential distribution of naturalized and invasive species (*sensu* Richardson *et al.*, 2000; to simplify, both naturalized and invasive species will be referred to as *invasives* hereafter) in space or time (Peterson, 2003; reviewed in Gallien *et al.*, 2010). Among

DOI: 10.1111/j.1466-8238.2012.00768.x

these tools, species distribution models (SDMs; Guisan & Thuiller, 2005), i.e. phenomenological models that statistically relate observed species occurrences to environmental variables, have been used prolifically (see the review inGallien *et al*., 2010). They rely on the ecological niche concept and use observed occurrences and thereby model the realized niches of the focal species in the region studied (Pulliam, 2000; Soberòn, 2007).

In the context of invasion ecology it is important to specify that there are (at least) three possible views of an invasive species' niche (Gallien *et al*., 2010). Firstly, the *global niche* corresponds to the broad abiotic and, to a lesser extent, the biotic conditions, under which the species persists. It is built from all data collected across a species' range (i.e. the sum of all its realized niches) and is the most complete estimate of the entire ecological niche without laboratory experimentation (Vetaas, 2002). Secondly, at the scale of the study region, the *regional niche at equilibrium* is limited by both the small-scale abiotic conditions and biotic interactions (e.g. competition, predation, pathogens) in this region (but we often don't have access to it). Thirdly, the *realized regional niche* differs from the regional equilibrium niche when the invader is not in quasi-equilibrium with the regional environment, and is thus limited by abiotic conditions, biotic interactions, invasion history and dispersal constraints (Wilson *et al*., 2007).

Following this differentiation between invasive species niches, SDMs have been used to predict the potential distributions of invasives in adventive regions (using the realized regional niche; e.g. Rouget *et al*., 2004), at the global scale (using the global niche; e.g. Beaumont *et al*., 2009) and even under environmental change scenarios (e.g. Roura-Pascual *et al*., 2004). However, the suitability of SDMs for modelling invasive species can be questioned on the grounds that two of their critical assumptions are usually seriously flawed. First, SDMs assume that the species' ecological niche is stable in space and time. In other words, the invasive species in its adventive region occupies similar environmental conditions as in the native range. Second, to ensure reliability, SDMs assume that the species of interest is at quasi-equilibrium with the environment in which it occurs. In other words, the invasive species has already reached all suitable places and is absent from all unsuitable sites (Guisan & Thuiller, 2005).

Concerning niche stability in space, realized regional niches may differ significantly between the native and invaded ranges (e.g. Broennimann *et al*., 2007; Gallagher *et al*., 2010). Compared to a species' global niche, the realized regional niche corresponding to a new adventive region can either: (1) occupy only a reduced part of the global niche (e.g. due to local biotic limitations such as competitors, predators or pathogens) similar to the realized niche in the native range; (2) occupy only a reduced part of the global niche, different from the realized niche in the native range; or (3) partly occupy areas outside of the global niche thanks to rapid genetic adaptations in the adventive range (Fig. 1 in Gallien *et al*., 2010). In order to address this issue and to project the potential distribution of an invasive species in an adventive range, it has been recommended that all data available throughout the world (for both native and invasive ranges) are used in order to estimate its ecological niche or at least the full-range biotic and abiotic niche requirements of the species (i.e. *the global niche*; Beaumont *et al*., 2009; Ibáñez *et al*., 2009). However, a global description of the niche does not account for the specificities of local adventive ranges (local environment, local biotic interactions and specific human uses). This explains why, when predicting the potential distribution of the species of interest in a specific invaded region, some researchers prefer to use occurrence data from the invaded range only (e.g. Dullinger *et al*., 2009). Therefore, using a combination of both global and realized regional niches has the potential to produce improved estimates of the potential distribution of a given invader in a study region (e.g. Roura-Pascual *et al*., 2009).

In theory, the equilibrium between the invader and the environment varies according to the stage of invasion: *introduction*, *colonization* or *establishment* (Theoharides & Dukes, 2007). These invasion stages are themselves strongly influenced by five elements: (1) the introduction history (e.g. propagule pressure, position of founder populations, time of residence; Wilson *et al*., 2007); (2) the spatial distribution of suitable habitats (Alofs & Fowler, 2010); (3) the invader's characteristics (e.g. dispersal capabilities; Aikio *et al*., 2010); (4) the invader's potential for rapid adaptation (Travis *et al*., 2009); and (5) interactions between the invader and the native communities (Davies *et al*., 2010). These factors make it difficult to infer the stage of regional invasion by simply using distribution data, without prior expert knowledge. Consequently, the observed distribution of the species does not always inform its potential distribution in the region.

Here, we take advantage of the difference between the *realized regional niche* and the *global niche* and propose a novel framework capable of producing more reliable predictions of the distribution of an invasive species in its adventive range of interest, on the one hand, and an improved estimation of its invasion stage and risk, on the other. This framework consists of two development phases (Fig. 1).

Firstly, we develop a hierarchical approach to improve regional SDM performance while simultaneously accounting for both global and regional information (Fig. 1, steps 1–2). When observed absences or pseudo-absence data are used to build a SDM, it is generally assumed that they represent 'true' absences (i.e. sites where the species cannot survive). In the case of invasive species, it is likely that some represent 'false' absences because the species is not at equilibrium (i.e. sites where the species could survive but is currently absent due to dispersal limitations; Le Maitre *et al*., 2008). This problem can be partly overcome by exploiting the estimation of the invader's global niche to attribute a weight to each (pseudo-)absence, i.e. if the pseudo-absence obtains a low probability of suitability in the global model then it will have a higher weighting in the regional model because it is more likely to be a 'true' absence, and vice versa. This is a first step towards closing the gap between the *realized regional niche* and the *regional niche at equilibrium*.

Secondly, we compare the invader's global (where the species could spread) and realized regional (where the species is already

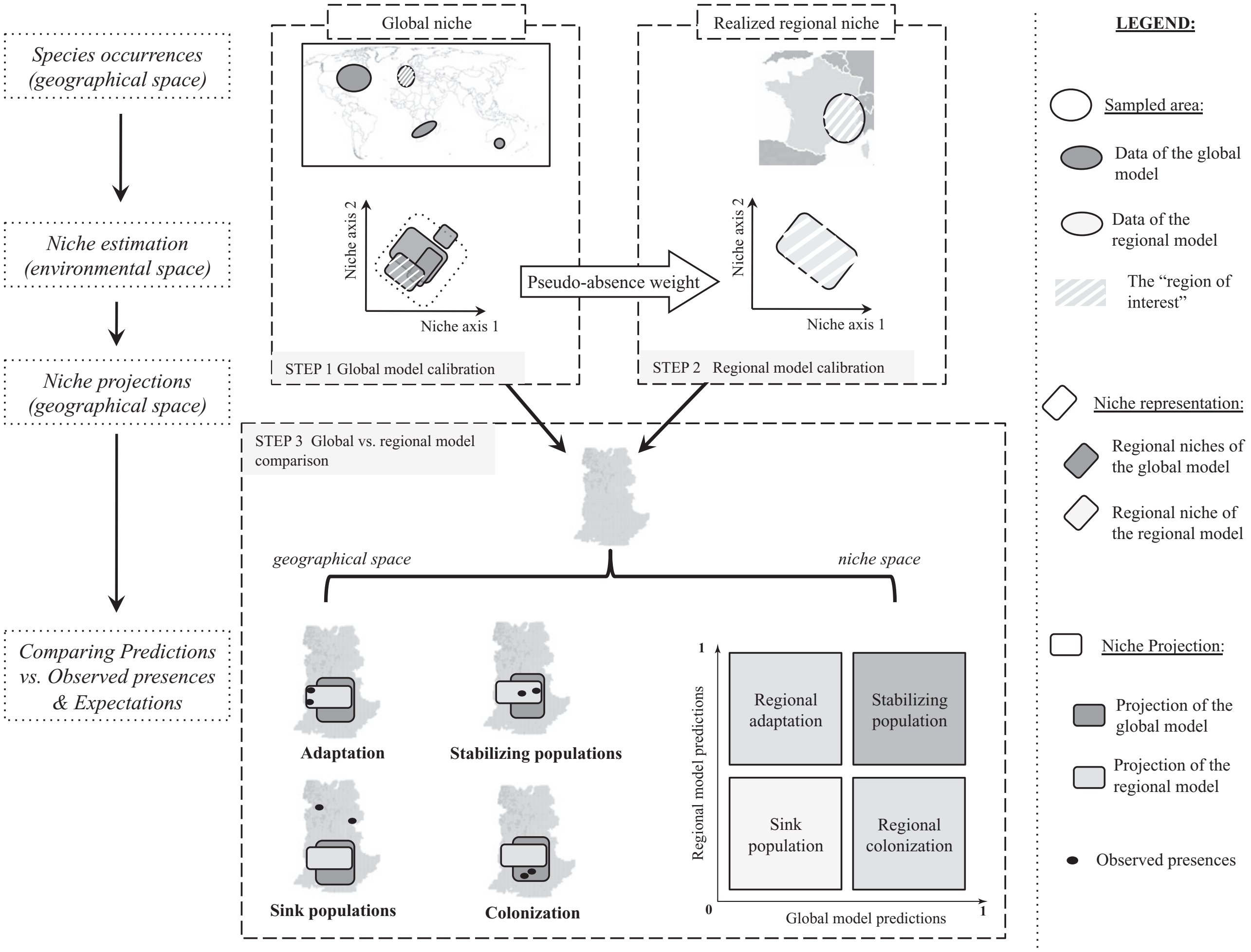

Figure 1 Flowchart summarizing the three steps proposed in our analysis. Step 1 represents the global model calibration and step 2 represents the regional model calibration (integrating the output of the global model to weight the regional pseudo-absences); both steps relate to the methodological innovation for modelling invasive species. Then, step 3 presents the theoretical framework developed to inform about the stages of invasion of an invasive species' population. If a species is observed (1) within its regional niche only, then it is potentially a population that develops adaptation to novel environmental conditions. If however a species is observed (2) within both its global and regional niches, then it is likely to be a stabilizing population. If a species is observed (3) only within in its global niche, then it may represent a population that participates to the colonization process of various environmental conditions in the region. If finally a species is observed (4) outside both niches, then it probably represents a 'sink' population.

observed) niches with the observed presences to take advantage of both the disequilibrium and the global niche estimation. Theoretically, this comparison allows us to infer both the stage of invasion for each population in the ecological niche space and the degree of regional range filling of the invading species in geographical space (Fig. 1, step 3). In the niche space of an adventive region (assuming the best set of explanatory variables and no data bias), a species is at quasi-equilibrium when at the same time its observed presences are located within both the global and realized regional niches and they fully fill the regional niche range. However, if the regional niche range is not filled then populations are approaching but still away from stabilization. When the species observations instead cover the global niche but not the realized regional niche (i.e. the regional model cannot predict some of the observed presences), then regional quasi-equilibrium is not yet reached (e.g. colonization from different sources in various environments). Alternatively, if some populations within the realized regional niche are outside the global niche, this indicates that these populations may have adapted to new (abiotic and/or biotic) environments (e.g. rapid local adaptations; Lavergne & Molofsky, 2007). Finally, if a species occurs regionally outside of both the global and the regional niches, then it is probable that the observed presences belong to sink populations (e.g. those introduced into unsuitable areas that are unlikely to provide opportunities for stable population development). Extending these comparisons from niche space to geographical space allows us to infer the degree of range filling for both a species' *global niche* and *realized regional niche*. Combining the information on species' population stages during invasion and range filling has the potential to provide interesting insights into the invader's future dynamics and potential threat.

Here, we use 27 invasive plant species in the French Alps to test and illustrate the overall development of the approach. The results are then consolidated with the expert knowledge of national botanists. Finally, we make suggestions regarding the further use of the framework to generate testable hypotheses of interest in invasion biology.

## MATERIALS AND METHODS

### The study system

Since 1980, 142 non-native plant species have been identified in the French Alps (source: National Botanical Conservatory of the Alps and Mediterranean, CBN). We used this list of species to create a global and a regional occurrence database. At the global scale, we extracted species occurrences from the Global Biodiversity Information Facility (GBIF; http://data.gbif.org) at a minimum resolution of 2.5′ (*c.* 4.5 km). We post-processed the data to remove all records from botanical gardens or those with unrealistic coordinates. At a regional scale, we used the occurrence records from the CBN containing 30 years of botanical surveys at a minimum resolution of 100 m (Boulangeat *et al.*, 2011). We only selected species with more than 200 records in the French Alps to avoid biased estimations of species environmental preferences and to remove casual species. This left us with a dataset of 27 species.

### Modelling framework

At both global and regional scales and for each species we built a set of SDMs – with presence records and randomly generated pseudo-absences – and applied a committee averaging method (Box 1) to extract a single output.

#### *The global model*

The committee averaging method (Box 1) was used to construct an estimation of the species' global niche for each of 27 invasives, using the both the GBIF and CBN presence data with two datasets of 20,000 random pseudo-absences each and the WorldClim climatic database (Hijmans *et al.*, 2005; http://www.worldclim.org/). The high number of pseudo-absences artificially reduces the prevalence in the models, which in return influences the probabilities of occurrence of the models (reducing the overall probability values). To avoid this problem here, we did not use the raw probabilities from the models but instead we transformed the probabilities into binary presence/absence data via a threshold (see below). In order to restrict the choice of pseudo-absences to realistically reachable locations, we created a buffer zone of 20 km around any of the presence records used and we randomly allocated absences inside these buffer zones. In this way we avoid areas where invasive species have not been inventoried. This strategy follows the one advocated by Phillips *et al.* (2009) who recommended following the same sampling design for selecting pseudo-absences as for selecting presences. From the 19 available bioclimatic variables we selected the five which had the lowest pair-wise correlations for our dataset (Spearman rank-correlation < 0.6), i.e. (1) maximal temperature in the warmest month, (2) annual temperature range, (3) mean temperature in the coldest quarter, (4) precipitation in the wettest month, and (5) precipitation in the driest month. To model species distributions, we used five algorithms available in the BIOMOD library (version 1.1–6.3; Thuiller *et al.*, 2009) in R (R Development Core Team, 2010): (1) a regression method [generalized additive model (GAM) with four degrees of smoothing and a stepwise variable selection based on the Akaike information criterion (AIC); Hastie & Tibshirani, 1990]; (2) a classification method [classification tree analysis (CTA), with a 50-fold cross-validation; Breiman *et al.*, 1984]; (3) a mix

**Box 1**

Committee averaging method. The committee averaging method is an ensemble forecasting method (Araújo & New, 2007) based on the use of different model algorithms (e.g. regressions, classification trees, machine learning). The rationale of ensemble forecasting is that different algorithms have different levels of accuracy under different circumstances and there is no single perfect algorithm (Elith et al., 2006). In the committee averaging method, predicted probability maps of species presences from the different algorithms are not averaged, but instead are transformed into binary maps (using for each model the threshold that maximizes both sensitivity and specificity) which are then averaged to obtain one single map of the final output. In other words, each model 'votes' for each site whether it forecasts a species' presence or not. It is therefore not a probability of occurrence that is measured but rather a percentage of agreement on species presence between the various algorithms. The main advantage of the committee averaging method is the use of 'comparable outputs' (binary presence–absences) instead of the raw algorithm outputs (continuous probabilities) that do not necessarily have the same meaning or the same range of variation. This method can easily incorporate the use of: (1) various model algorithms; (2) multiple selections of pseudo-absence data (minimizing the bias due to a specific set of selected pseudo-absences); and (3) several repetitions of cross-validation procedures (calibration and evaluation procedures are repeatedly carried out on different subdatasets). It is also possible to keep only the best performing models (i.e. reliable models only) for the final output by setting a selection threshold based on predictive accuracy metrics. In the end, the number of potential 'voting maps' is: number of algorithms × number of pseudo-absence datasets × number of cross-validation runs.

between regression and classification methods [multivariate adaptive regression splines (MARS); Friedman, 1991]; (4) a boosting algorithm [boosted regression trees (BRT) with the optimal number of trees selected by cross-validation; Ridgeway, 1999]; and (5) a machine learning method [artificial neural networks (ANN) with the best amount of weight decay and the number of units in the hidden layer selected using five-fold cross-validation; Ripley, 1996].

We implemented a split-sample cross-validation procedure to avoid circular reasoning for evaluating the models (i.e. different data portions are used to construct and to evaluate the model; Araújo *et al.*, 2005). Specifically, models were calibrated for each species on 70% of the initial data and then evaluated on the remaining 30% with the true skill statistic (TSS; Allouche *et al.*, 2006) and the area under the receiver operating characteristic curve (AUC; Swets, 1988). The AUC scores vary from 0 for a model whose predictions are systematically wrong, over 0.5 for a random fit, to 1 for a model achieving perfect agreement with the observed data. The TSS has a range of −1 to +1, with −1 and +1 representing systematically wrong predictions and systematically right predictions, respectively, and 0 representing a random fit.

For each algorithm and for each species, two pseudo-absence datasets were randomly selected, and four cross-validations performed. Forty different models were therefore calculated in total. Of these models, only those obtaining both a TSS and AUC score above 0.6 and 0.8, respectively, were used to build the committee averaging map. There are no specific guidelines for both TSS and AUC scores as they depend on the extent and (obviously) on the goal of the study. We chose 0.6 and 0.8, respectively, based on visual inspections of the output and on the proposed thresholds used in the literature (e.g. Araújo *et al.*, 2005; Engler *et al.*, 2011).

### *The regional models*

At the scale of the French Alps we used the committee averaging method to model the regional distribution of each invasive species using two different approaches: a conventional approach which used only the data from the French Alps to construct the models, and our proposed approach which additionally integrates global niche information (from the global model built in the section above). The single difference between the two approaches concerns the weights attributed to the pseudo-absences ('true' absence data are not available), after they had been randomly generated.

For both procedures, models were built for the French Alps region at a 100-m spatial resolution using the CBN occurrence records and a set of five environmental variables (four pedoclimatic and one land cover) known to be important for species establishment and spread. Four climatic variables originated from the meteorological model Aurelhy (Bénichou & Le Breton, 1987), based on interpolated measurements at a resolution of 100 m × 100 m, summarizing climatic information over the last 30 years (here 1971–2000). These variables were: (1) mean annual solar radiation; (2) maximum temperature of the summer as an index of extreme temperatures (this variable was highly correlated to minimum temperature in winter); (3) standard deviation of annual precipitation as an index of seasonality; and (4) soil water-holding capacity. Land-cover information was extracted from the CORINE Land Cover Map for Europe (as suggested in Polce *et al.*, 2011). Six statistical algorithms (used to attribute different weights to the data) were selected: GLM (generalized linear model), GAM, CTA, GBM, RF (random forest) and ANN. As for the global-scale models, we also ran the regional models with two random selections of pseudo-absence data, followed by four cross-validation repetitions (70–30% as for the global model), giving a total of 48 models per species for each procedure. In order to remove inaccurate models, only those which obtained TSS and AUC scores of over 0.6 and 0.8, respectively, were used to build the committee averaging map.

For the conventional approach, models were constructed using observed species presences and two sets of random pseudo-absence data (10,000 absences per dataset). Traditionally, pseudo-absence data have the same weight as presence data. In other words, the algorithms attribute equal confidence to the pseudo-absence data as to the observed presence data (i.e. pseudo-absences are considered as 'true' absences).

In our proposed approach, models were built using the same observed presences and pseudo-absence data as for the conventional approach, but without assuming that all pseudo-absences represent 'true' absences (i.e. we assume that some pseudo-absences probably reflect environmental conditions where the species cannot survive, while others reflect locations where the species has not yet arrived due to dispersal limitations). We used the global model projections applied to the region to weight each pseudo-absence. Where the global model showed a high level of agreement with the pseudo-absence (i.e. a low habitat suitability) we attributed a high weight to the pseudo-absence (i.e. it probably represents a 'true' absence), and vice versa. The weight was given by means of an inverse logistic transformation (equation 1) to obtain stronger discrimination between the predictions of absences and presences:

$$\text{Weight}\,(x) = \frac{1}{1+\left(\dfrac{\text{projG}(x)}{\text{projG}(x)-1}\right)^{2}} \tag{1}$$

where Weight($x$) is the weight attributed to the pseudo-absence $x$, which depends on projG($x$) the global model prediction at the location of $x$ [if projG($x$) = 1 then Weight($x$) = 0].

## Interpreting outputs

The global and regional prediction accuracy was estimated using the two aforementioned indices (AUC and TSS). At the regional scale, we also compared the regional model predictions with the global model predictions both for the observed presences and for the pseudo-absences separately. The comparison looked at sensitivity (proportion of presences correctly predicted) and specificity (proportion of absences correctly predicted), respectively. Note that a threshold value had to be selected in order to convert continuous model predictions into a discrete prediction of presences or absences. A threshold of 0.5 is not usually optimal when model outputs are true probabilities (Santika, 2011), but given that the outputs were the agreements between models, we selected this value in order to make sure that the majority of model agreements were decided (i.e. at least half of the models agree), because we wanted to create neither conservative nor liberal predictions.

The estimation of a species' stage of invasion was inferred according to the theoretical framework displayed in Fig. 1 (step 3). It is based on a comparison of the predictions made by the models at the global and at the regional scale for each observed presence in the French Alps. Whether the observed presence is predicted as a presence or an absence in one or both models can theoretically relate to the stage of invasion for each invader's population. In addition, for each species we estimated its range filling in the French Alps (regional scale). We compared three pairs of data in the geographical space: (1) the observed presences against the regional model predictions (Svenning & Skov, 2004); (2) the observed presences against the global model predictions; and (3) the regional against the global model predictions. We used these comparisons as proxies to assess: (1) how well species currently fill their projected regional niche; (2) how well species currently fill their projected global niche; and (3) how (dis-)similar are the regional and the global niche projections. In order to have comparable inputs (occurrence, regional and global projection resolutions), we used a grid at 2.5′ resolution (i.e. the global model resolution) where the pixel occupancy of the regional model at 100-m resolution and the observations were scaled up: if at least one of the pixels at 100-m resolution was occupied then the aggregated pixel was considered to be occupied.

## RESULTS

### Model performances

Global models showed good performances at the global scale (AUC > 0.8 and TSS > 0.6 for all species), but low to moderate performances at the regional scale (AUC between 0.15 and 0.8; TSS between 0 and 0.5; Fig. 2a). In other words, an invader's observed presences at regional scale did not fill the ranges predicted by the global models (absences within and/or presences outside of the projected niches). Global-scale information alone was thus not sufficient to predict regional-scale distribution of invaders.

The comparison of the performance of the two regional modelling approaches (weighted versus unweighted) revealed that in 78% of the cases, weighting the pseudo-absences significantly improved discrimination between areas where the species was observed as being present and where it was not recorded (pseudo-absences) (Fig. 2b). This difference was essentially due to the fact that: (1) presences were generally equally predicted by the unweighted model when they occurred outside of the global niche, but (2) absences were better predicted by the weighted model (Fig. 2c).

### Stage of invasion

The comparison of the performance of the global and regional models with the aim of inferring the stage of invasion was only carried out using the regional model with weighted pseudo-absences given that its performance was better (or equivalent) to the unweighted model in all comparisons (Fig. 2b).

Using the proposed theoretical framework (Fig. 1) we compared the invader's (global and realized regional) niches with observed presences and inferred the current stage of invasion for our studied species from this comparison. In general, four situations were distinguished (Fig. 1): (1) species with stabilizing populations in the region (e.g. *Panicum capillare*), for which the observed presences were included in both observed regional and global niches; (2) species that are probably undergoing rapid local adaptation (e.g. *Ailanthus altissima*, *Artemisia annua*), for which many observed presences were within the regional niche but outside of the global niche; (3) species that are engaged in the colonization process (e.g. *Solidago gigantea, Buddleja davidii*), for which many observed presences were outside of the regional niche but within the global niche; and (4) species with populations at different stages of invasion (e.g. *Sorghum halepense*), for which observed presences were both within and outside of both the regional and global niches. None of the species was found to only have sink populations in the region, probably because we selected species with at least 200 occurrences (see Appendix S1 in Supporting Information for the results on the 27 species).

### Range filling and risk of invasion

The level of range filling in the French Alps complemented the information on the stage of invasion for all species (Fig. 3, Appendix S1). Four situations were distinguished: (1) species at quasi-equilibrium with stabilizing populations widely distributed over the entire regional niche projection (e.g. *P. capillare*), representing a low risk of invasion in the near future; (2) species with stabilizing populations but not filling the entire regional niche projection, representing species that could be classified at risk of invasion in the future because not all suitable environments have yet been occupied; (3) species in disequilibrium due to local adaptations outside of their global niche; (4) species in disequilibrium due to a colonization process. In the specific case of species with populations undergoing local adaptations, if the regional niche projection was filled by presences (e.g. *Ailanthus altissima*) these species could be considered as having a higher risk of spread than if it was not filled (e.g. *Artemisia annua*). The assumption behind it is that a filled niche projection provides more opportunities for gene flow between populations that could increase the spread of the adaptations. In the case of species undergoing a colonization process, if the regional niche projection was filled by presences (e.g. *B. davidii*), then these species could be considered as having a lower risk of spread than if the regional niche projection was not filled (e.g. *Solidago gigantea*) because the suitable but still unoccupied sites were rare.

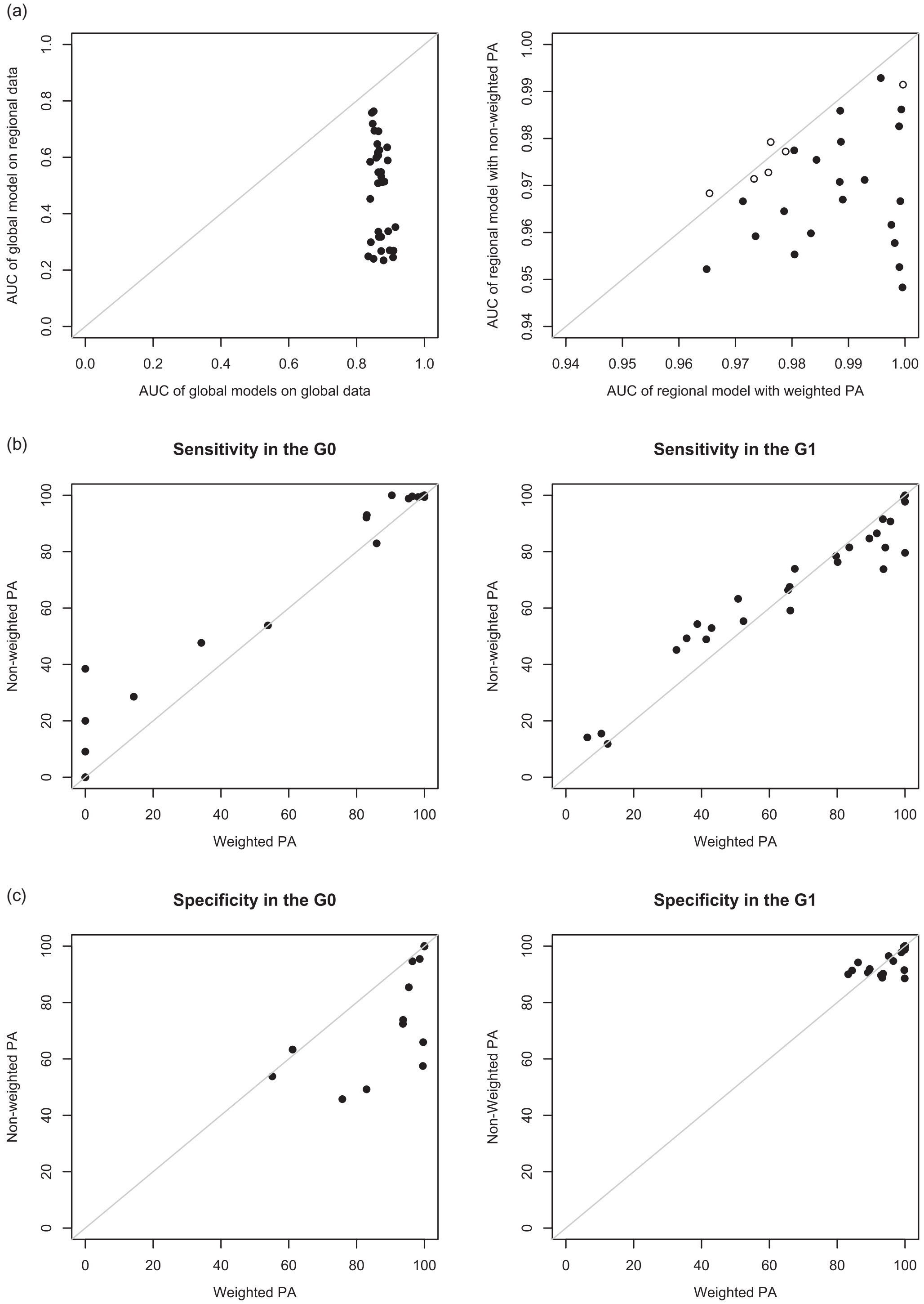

Figure 2 Comparison of model performance. (a) Global model performance (area under the receiver-operating characteristic curve, AUC) for global versus regional data. (b) Regional model performance (AUC) for weighted versus unweighted pseudo-absences (PA), where filled circles represent significant differences and open circles non-significant differences. (c) Sensitivity and specificity of the two regional models within (G1) and outside (G0) of each species' global niche. The diagonal lines show the values of identical performance between the models compared.

## DISCUSSION

The conceptual and statistical improvement proposed herein relies on the recognition and use of the multiple regional niches the species exhibited in their global distribution (Gallien *et al.*, 2010). It makes it possible to obtain the most complete estimate of a species' ecological niche from observational data. The use of the global niche allows us to simultaneously obtain a better

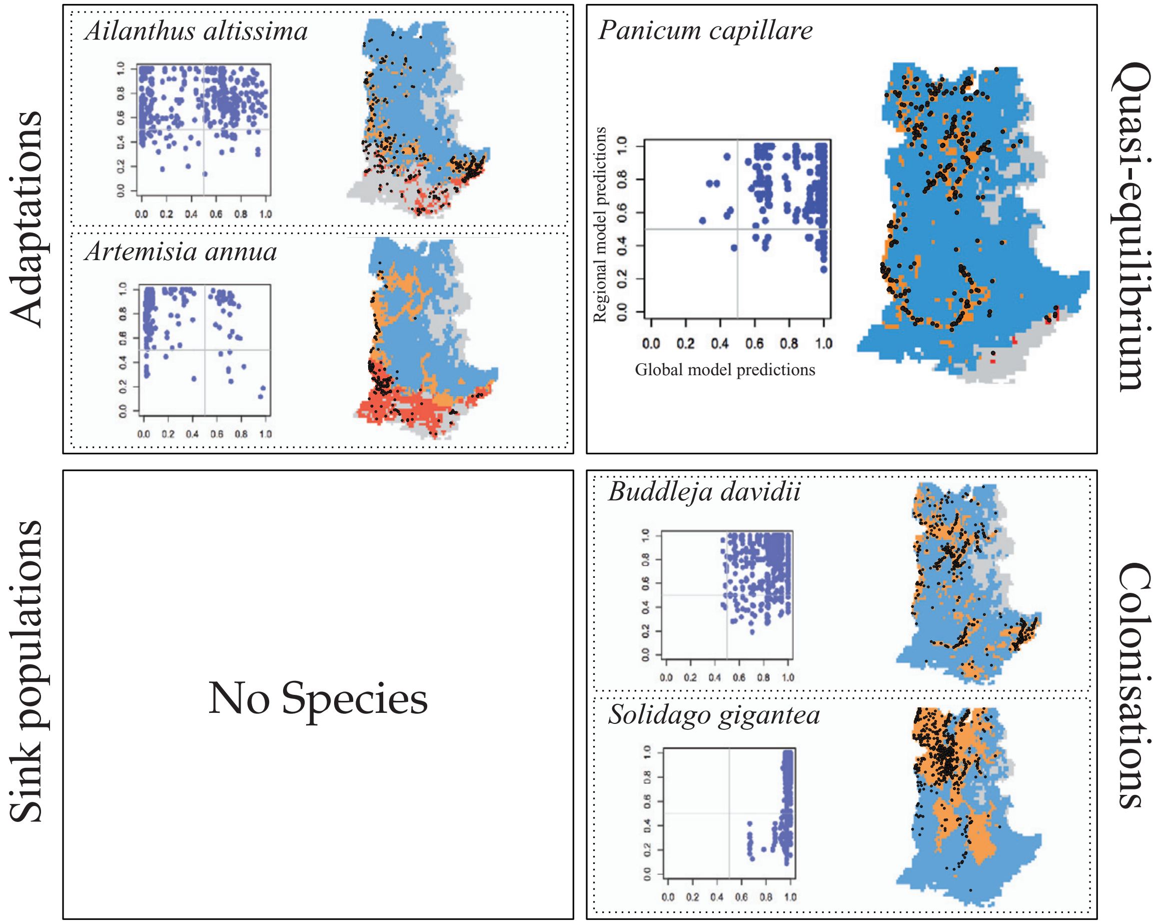

Figure 3 Invasion stages for five invasive species according to our novel methodological framework (Fig. 1). For each species' observed occurrence we plotted the values of the global and regional model predictions, and mapped their geographical representation. Global niches are shown in blue, overlapping regional and global niches in orange, regional niches in red, areas outside of both niches in grey, and observations of occurrences are represented with black dots.

estimation of a species' climatic limitations, and to realistically remove potentially false absences in the regional pseudo-absence datasets. Indeed, because both biotic and abiotic conditions differ between regions and because species climatic tolerances may change (e.g. local adaptation, genetic drift or phenotypic plasticity) the observed niches estimated in different regions (both native and invasive) may vary substantially. Using only one estimate of realized regional niches (e.g. the native range) may misrepresent the species' environmental preferences and result in incomplete predictions (Broennimann *et al.*, 2007; Fitzpatrick *et al.*, 2007).

Weighting the pseudo-absences at a regional scale using information from the global model increased the predictive accuracy of the regional models by: (1) decreasing the influence of (regional) false absences, (2) invariably accepting true (regional) presences, and (3) letting the regional climate, soil and land use refine the regional niche estimation. Other approaches have been proposed to constrain pseudo-absence selection, for example choosing only those found outside the species' climatic tolerance (e.g. Le Maitre *et al.*, 2008), or selecting pseudo-absence data with the same sampling bias as the one of the observed presence datasets (e.g. Phillips *et al.*, 2009). Although these approaches have the advantage of considering the species' global niche, or diminishing the sampling bias, the former does not ensure the model to matches observed regional distribution more closely and the latter does not resolve the problem of false absences.

Our proposed framework is more than just a model improvement as it has the potential to advance our understanding of species invasion and associated risk. Based on theoretical expectations, under the assumption that novel climatic conditions do not largely exceed those present in the global distribution, and when using adequate data, the framework provides insights into characteristics of invader populations in a region, as to whether they are: (1) at quasi-equilibrium, (2) potentially adapting to new local conditions, (3) still in the colonization process, or (4) represent sink populations (Fig. 3). These four different cases can then be formally tested using observations or common-garden experiments.

For instance, *Panicum capillare* L. (an annual tropical gram, Poaceae family) is judged currently to be at low risk of expansion in the French Alps. Most of the species' populations are 'stabilizing' and fill the regional range (Fig. 3). This corroborates the history of the species, introduced 250 years ago together with corn seeds. It does not spread a long way from the cornfields but has large seed banks and long dormancy making the populations relatively stable.

Amongst the set of investigated invasive species, some are well predicted by the regional model but also seem to occur outside of their global niches, such as *Ailanthus altissima* (Mill.) Swingle and *Artemisia annua* L. (Simaroubaceae and Asteraceae family, respectively; Fig. 3). One basic explanation for this pattern could be methodological. The estimated global niche might not be sufficient for describing a more complex niche (e.g. it may be missing important variables). Another non-exclusive explanation is that individuals within these populations have managed to modify their environmental preferences. This could be due to increased genetic variation from multiple introductions generating genetic novelties through recombination (e.g. Lavergne & Molofsky, 2007), genome characteristics for rapid adaptation (e.g. neopolyploidy; Ramsey & Schemske, 2002) or hybridization with adapted native congeneric species (cf. Dietz & Edwards, 2006). These populations are of particular concern as their distributions are likely to continue to expand in the near future. Interestingly, *Ailanthus altissima* fills a larger part of its regional range than *Artemisia annua*, suggesting that populations of *Ailanthus altissima* that rapidly adapted outside of the global niche have more chance of subsisting and spreading along the colonization front (e.g. via mutation surfing; Travis *et al*., 2010). This fits with the characteristics of *A. altissima*, a tree introduced 300 years ago for ornamental reasons, which has a high potential of tilling for efficient short-distance dispersal. In contrast, *Artemisia annua* is known to have sporadic populations that are less likely to adapt locally.

Similarly, the risk of future spread can be relatively different for two colonising species such as *S. gigantea* Aiton (Asteraceae) and *B. davidii* Franchet (Buddlejaceae). *Solidago gigantea* has a wider projected regional niche but it has filled it regional niche less than does *B. davidii*. This broad geographical scatter of the former probably results from multiple independent introduction events.

Finally, individual populations of the same species may be at various stages of invasion, like for example *Sorghum halepense*, a subtropical graminoid species. This species harbours a combination of functional trait advantages (perennial, C4 metabolism and vegetative reproduction) and is able to spread and colonize away from cultivated fields, and eventually develop adaptations to new environmental conditions (with the possibility of hybridisation with *Sorghum bicolor* that is cultivated in the region, Morrell *et al*., 2005).

There are obviously possible pitfalls associated with the proposed framework that need to be carefully discussed. For instance, the capacity of the framework to determine the degree of range filling will ultimately depend on the heterogeneity of the regional environment. In highly heterogeneous environments, a model calibrated at the coarse (global) resolution could fail to fully capture the environmental variability that is observable at the fine (regional) resolution. In our case study, we up-scaled our projections from the regional model to the same resolution as the global model. This ascertains that the number of occupied pixels becomes comparable, and it gives a first approximation of a species' range filling. However, in a perfect case, the global and regional models should be calibrated at the same resolution with exactly the same variables (climate, land cover, soil information). Having the species distribution and the climate data available at a 100-m spatial resolution at the global scale is currently too demanding. In addition, sampling bias at both global and regional scales could also influence the species niche estimations and thus affects the overall relevance of the analysis. In the case of a strong sampling bias, we would advise the collection of additional data in order to lessen the bias before applying the proposed methodological framework.

In conclusion, we discuss how the proposed framework could also be used to generate testable hypotheses that link the concept of ecological niche to invasion ecology.

**1.** What makes a good invader? Identifying which functional traits characterize invasive species has a long history in ecology, and some key traits have been consistently reported as favouring invasion (e.g. clonality, high seed production, hybridization potential; Pyšek & Richardson, 2007). Our framework could make it possible to identify, for a large number of species, whether invaders suspected of having evolved their niches (e.g. Beaumont *et al*., 2009) have certain specific characteristics (e.g. mean functional traits, large intra-specific functional trait variability, phylogenetic position in a rapidly evolving clade, generalist versus specialist species).

**2.** Could we retrieve the invasion history and dynamics? Being able to re-create the invasion history of a particular species and understand its dynamics is crucial. This is especially true for pest species undergoing evolutionary adaptation in their adventive range. Our framework is capable of identifying the populations that are likely to undergo rapid adaptation, which could further be sampled for trait measurements and genetic analyses (e.g. Albert *et al*., 2010). In the case of a known invasion history, our framework can help identify which factors contribute to colonization and adaptation or understand whether stabilizing populations are acting as source of gene flow toward non-equilibrium populations.

**3.** The interactions between the invader and the recipient native community. The reasons why some native communities are more resistant to invasion than others have been under investigation since Darwin's time (Darwin 1859), but this area still requires research (Thuiller *et al*., 2010). Our proposed framework can be used here to test new hypotheses about the interactions between the native and invasive species, such as whether the types of interaction differ according to the population's stage of invasion.

## ACKNOWLEDGEMENTS

This work was funded by ANR SCION (ANR-08-PEXT-03) project. Thanks to T. Münkemüller, C. H. Albert, K. Schiffers, F. Boucher, Núria Roura-Pascual and two anonymous referees for comments that greatly improved earlier versions of the manuscript. We would also like to thank the Alpine and Mediterranean National Botanical Conservatories for providing the observed data at regional scale and the list of invasive plant species in the region.

## REFERENCES

Aikio, S., Duncan, R.P. & Hulme, P.E. (2010) Herbarium records identify the role of long-distance spread in the spatial distribution of alien plants in New Zealand. *Journal of Biogeography*, **37**, 1740–1751.

Albert, C.H., Edwards, T., Jr, Graham, C.H., Yoccoz, N.G., Zimmermann, N.E. & Thuiller, W. (2010) Applied sampling in ecology and evolution – integrating questions and designs. *Ecography*, **33**, 1028–1037.

Allouche, O., Tsoar, A. & Kadmon, R. (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, **43**, 1223–1232.

Alofs, K.M. & Fowler, N.L. (2010) Habitat fragmentation caused by woody plant encroachment inhibits the spread of an invasive grass. *Journal of Applied Ecology*, **47**, 338–347.

Araújo, M.B. & New, M. (2007) Ensemble forecasting of species distributions. *Trends in Ecology and Evolution*, **22**, 42–47.

Araújo, M.B., Pearson, R.G., Thuiller, W. & Erhard, M. (2005) Validation of species–climate impact models under climate change. *Global Change Biology*, **11**, 1504–1513.

Beaumont, L.J., Gallagher, R.V., Thuiller, W., Downey, P.O., Leishman, M.R. & Hughes, L. (2009) Different climatic envelopes among invasive populations may lead to underestimations of current and future biological invasions. *Diversity and Distributions*, **15**, 409–420.

Bénichou, P. & Le Breton, O. (1987) Prise en compte de la topographie pour la cartographie des champs pluviométriques statistiques. *La Météorologie*, **7**, 23–34.

Boulangeat, I., Lavergne, S., Van Es, J., Garraud, L. & Thuiller, W. (2011) Niche breadth, rarity and ecological characteristics within a regional flora spreading over large environmental gradients. *Journal of Biogeography*, **39**, 204–214.

Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. (1984) *Classification and regression trees*. Chapman & Hall, New York.

Broennimann, O., Treier, U.A., Muller-Scharer, H., Thuiller, W., Peterson, A.T. & Guisan, A. (2007) Evidence of climatic niche shift during biological invasion. *Ecology Letters*, **10**, 701–709.

Darwin, C.R. (1859) *The origin of species*. John Murray, London.

Davies, K.F., Cavender-Bares, J. & Deacon, N. (2010) Native communities determine the identity of exotic invaders even at scales at which communities are unsaturated. *Diversity and Distributions*, **17**, 35–42.

Dietz, H. & Edwards, P.J. (2006) Recognition that causal processes change during plant invasion helps explain conflicts in evidence. *Ecology Letters*, **87**, 1369–1367.

Dullinger, S., Kleinbauer, I., Peterseil, J., Smolik, M. & Essl, F. (2009) Niche based distribution modelling of an invasive alien plant: effects of population status, propagule pressure and invasion history. *Biological Invasions*, **11**, 2401–2414.

Elith, J., Graham, C.H., Anderson, R.P. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Engler, R., Randin, C.F., Thuiller, W. *et al.* (2011) 21st century climate change threatens mountain flora unequally across Europe. *Global Change Biology*, **17**, 2330–2341.

Fitzpatrick, M.C., Weltzin, J.F., Sandersn, N.J. & Dunn, R.R. (2007) The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? *Global Ecology and Biogeography*, **16**, 24–33.

Friedman, J. (1991) Multivariate adaptive regression splines. *Annals of Statistics*, **19**, 1–141.

Gallagher, R.V., Beaumont, L.J., Hughes, L. & Leishman, M.R. (2010) Evidence for climatic niche and biome shifts between native and novel ranges in plant species introduced to Australia. *Journal of Ecology*, **98**, 790–799.

Gallien, L., Münkemüller, T., Albert, C.H., Boulangeat, I. & Thuiller, W. (2010) Predicting potential distributions of invasive species: where to go from here? *Diversity and Distributions*, **16**, 331–342.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.

Hastie, T.J. & Tibshirani, R. (1990) *Generalized additive models*. Chapman and Hall, London.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Ibáñez, I., Silander, J.A., Wilson, A.M., Lafleur, N., Tanaka, N. & Tsuyama, I. (2009) Multivariate forecasts of potential distributions of invasive plant species. *Ecological applications*, **19**, 359–375.

Lavergne, S. & Molofsky, J. (2007) Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proceedings of the National Academy of Sciences USA*, **104**, 3883–3888.

Le Maitre, D.C., Thuiller, W. & Schonegevel, L. (2008) Developing an approach to defining the potential distributions of invasive plant species: a case study of *Hakea* species in South Africa. *Global Ecology and Biogeography*, **17**, 569–584.

Morrell, P.L., Williams-Coplin, T.D., Lattu, A.L., Bowers, J.E., Chandler, J.M. & Patterson, A.H. (2005) Crop-to-weed introgression has impacted allelic composition of Johnsongrass populations with and without recent exposure to cultivated sorghum. *Molecular Ecology*, **14**, 2143–2154.

Olden, J.D., Poff, N.L., Douglas, M.R., Douglas, M.E. & Fausch, K.D. (2004) Ecological and evolutionary consequences of biotic homogenization. *Trends in Ecology and Evolution*, **19**, 18–24.

Peterson, A.T. (2003) Predicting the geography of species' invasions via ecological niche modeling. *Quarterly Review of Biology*, **78**, 419–433.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J.R. & Ferrier, S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, **19**, 181–197.

Polce, C., Kunin, W.E., Biesmeijer, J.C., Dauber, J., Phillips, O.L. & The ALARM Field Site Network (2011) Alien and native plants show contrasting responses to climate and land use in Europe. *Global Ecology and Biogeography*, **20**, 367–379.

Pulliam, H.R. (2000) On the relationship between niche and distribution. *Ecology Letters*, **3**, 349–361.

Pyšek, P. & Richardson, D.M. (2007) Traits associated with invasivness in alien plants: where do we stand? *Biological invasions* (ed. by W. Nentwig), pp. 97–122. Springer-Verlag, Berlin.

R Development Core Team (2010) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.

Ramsey, J. & Schemske, D.W. (2002) Neopolyploidy in flowering plants. *Annual Review of Ecology and Systematics*, **33**, 589–639.

Rejmánek, M., Richardson, D.M., Higgins, S.I., Pitcairn, M.J. & Grotkopp, E. (2005) Ecology of invasive plants: state of the art. *Invasive alien species: a new synthesis* (ed. by H.A. Mooney, R.N. Mack, J.A. Mcneely, L. Neville, P.J. Schei and J. Waage), pp. 104–161. Island Press, Washington, DC.

Richardson, D.M., Pyšek, P., Rejmánek, M., Barbour, M.G., Panetta, F.D. & West, C.J. (2000) Naturalization and invasion of alien plants: concepts and definitions. *Diversity and Distributions*, **6**, 93–107.

Ridgeway, G. (1999) The state of boosting. *Computing Science and Statistics*, **31**, 172–181.

Ripley, B.D. (1996) *Pattern recognition and neural networks*. Cambridge University Press, Cambridge.

Rouget, M., Richardson, D.M., Nel, J.L., Le Maitre, D.C., Egoh, B. & Mgidi, T. (2004) Mapping the potential ranges of major plant invaders in South Africa, Lesotho and Swaziland using climatic suitability. *Diversity and Distributions*, **10**, 475–484.

Roura-Pascual, N., Suarez, A.V., Gómez, C., Pons, P., Touyama, Y., Wild, A.L. & Peterson, A.T. (2004) Geographical potential of Argentine ants (*Linepithema humile* Mayr) in the face of global climate change. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 2527–2534.

Roura-Pascual, N., Bas, J.P., Thuiller, W., Hui, C., Krug, R.M. & Brotons, L. (2009) From introduction to equilibrium: reconstructing the invasive pathways of the Argentine ant in a Mediterranean region. *Global Change Biology*, **15**, 2101–2115.

Santika, T. (2011) Assessing the effect of prevalence on the predictive performance of species distribution models using simulated data. *Global Ecology and Biogeography*, **20**, 181–192.

Soberòn, J. (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, **10**, 1115–1123.

Svenning, J.-C. & Skov, F. (2004) Limited filling of the potential range in European tree species. *Ecology Letters*, **7**, 565–573.

Swets, K.A. (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285–1293.

Theoharides, K.A. & Dukes, J.S. (2007) Plant invasion across space and time: factors affecting nonindigenous species success during four stages of invasion. *New Phytologist*, **176**, 256–273.

Thuiller, W., Lafourcade, B., Engler, R. & Araújo, M.B. (2009) BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography*, **32**, 369–373.

Thuiller, W., Gallien, L., Boulangeat, I., de Bello, F., Münkemüller, T., Roquet-Ruiz, C. & Lavergne, S. (2010) Resolving Darwin's naturalization conundrum: a quest for evidence. *Diversity and Distributions*, **16**, 461–475.

Travis, J.M.J., Mustin, K., Benton, T.G. & Dytham, C. (2009) Accelerating invasion rates result from the evolution of density-dependent dispersal. *Journal of Theoretical Biology*, **259**, 151–158.

Travis, J.M.J., Münkemüller, T. & Burton, O.J. (2010) Mutation surfing and the evolution of dispersal during range expansions. *Journal of Evolutionary Biology*, **23**, 2656–2667.

Vetaas, O.R. (2002) Realized and potential climate niches: a comparison of four *Rhododendron* tree species. *Journal of Biogeography*, **29**, 545–554.

Wilson, J.R.U., Richardson, D.M., Rouget, M., Procheş, Ş., Amis, M.A., Henderson, L. & Thuiller, W. (2007) Residence time and potential range: crucial considerations in plant invasion ecology. *Diversity and Distributions*, **13**, 11–22.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

**Appendix S1** Stage and risk of invasion for 27 invasive species in the French Alps.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

## BIOSKETCHES

**Laure Gallien** is currently a PhD student at the Laboratory of Alpine Ecology (LECA) in France. She is interested in alpine plant invasion, where she especially focuses on evolutionary biogeography, community ecology and population genetics.

Editor: José Alexandre F. Diniz-Filho