Use of ecological niche models to predict the distribution of invasive species: a scientometric analysis

Barbosa, FG.a,b,*, Schneck, F.a and Melo, AS.c

aPrograma de Pós-graduação em Ecologia, Instituto de Biociências, Universidade Federal do Rio Grande do Sul – UFRGS, CP 15007, CEP 91501-970, Porto Alegre, RS, Brazil
bLaboratório de Estatística Ambiental, Instituto de Matemática, Estatística e Física, Universidade Federal do Rio Grande – FURG, CP 474, CEP 96201900, Rio Grande, RS, Brazil
cDepartamento de Ecologia, Instituto de Ciências Biológicas, Universidade Federal de Goiás – UFG, CP 131, CEP 74001-970, Goiânia, GO, Brazil
*e-mail: fabibarbos@gmail.com

Received August 25, 2011 – Accepted January 25, 2012 – Distributed November 30, 2012
(With 5 figures)

Abstract

We conducted a scientometric analysis to determine the main trends and gaps of studies on the use of ecological niche models (ENMs) to predict the distribution of invasive species. We used the database of the Thomson Institute for Scientific Information (ISI). We found 190 papers published between 1991 and 2010 in 82 journals. The number of papers was low in the 1990s, but began to increase after 2003. One-third of the papers were published by researchers from the United States of America, and consequently, the USA was also the most studied region. The majority of studies were carried out in terrestrial environments, while only a few investigated aquatic systems, probably because important aquatic predictor variables are scarce or unavailable for most regions in the world. Species-occurrence records were mainly composed of presence-only records, and almost 70% of the studies were carried out with plants and insects. Twenty-three different distribution modelling methods were used. The Genetic Algorithm for Rule-set Production (GARP) was used most often. Our scientometric analysis showed a growing interest in the use of ENMs to predict the distribution of invasive species, especially in the last decade, which is probably related to the increase in species introductions worldwide. Among some important gaps that need to be filled, the relatively small number of studies conducted in developing countries and in aquatic environments deserves careful attention.

Keywords: biodiversity, biological invasions, scientific production, trends.

Uso de modelos de nicho ecológico para prever a distribuição de espécies invasoras: uma análise cienciométrica

Resumo

Conduziu-se uma análise cienciométrica para determinar as principais tendências e lacunas de estudos sobre o uso de modelos de nicho ecológico (MNEs) para prever a distribuição de espécies invasoras. Usou-se a base de dados Thomson Institute for Scientific Information (ISI). Foram encontrados 190 artigos publicados entre 1991 e 2010, em 82 periódicos. O número de trabalhos foi baixo na década de 1990, mas começou a aumentar após 2000. Um terço dos trabalhos foi publicado por pesquisadores dos Estados Unidos da América e, consequentemente, os Estados Unidos também foram a região mais estudada. A maioria dos estudos foi realizada em ambientes terrestres. Apenas alguns sistemas aquáticos foram investigados, provavelmente porque importantes variáveis preditoras aquáticas são escassas ou inexistentes na maioria das regiões do mundo. Registros de ocorrência de espécies foram compostos principalmente por registros de presença (ou seja, sem registros de ausência) e quase 70% dos estudos foram realizados com plantas e insetos. Vinte e três diferentes métodos de modelagem foram utilizados. O Algoritmo Genético para Produção de Conjunto de Regras (GARP) foi o mais utilizado. Esta análise cienciométrica mostrou um interesse crescente no uso de MNEs para prever a distribuição de espécies invasoras, especialmente na última década, estando esse aumento provavelmente relacionado ao crescimento na introdução de espécies em todo o mundo. Entre algumas lacunas importantes que precisam ser preenchidas, destaca-se o número relativamente pequeno de estudos realizados em países em desenvolvimento e em ambientes aquáticos.

Palavras-chave: biodiversidade, invasões biológicas, produção científica, tendências.
1. Introduction

The intensification of global trade is continuously increasing the number of exotic species (also known as non-native species or non-indigenous species) introduced intentionally or accidentally to a new area (Westphal et al., 2008). The majority of the species do not succeed in establishing in the areas where they were introduced (Mack et al., 2000), but once established, they may spread and cause ecological and/or economic problems (Pimentel et al., 2005), becoming invasive species (Mack et al., 2000).

Biological invasions are causing dramatic changes in global biodiversity, often leading to a decline and/or extinction of native species (Mack et al., 2000; Pimentel et al., 2005). The development and use of preventive measures to deal with invasive species are thus a priority in biodiversity conservation (Hulme, 2006). Preventive measures are more cost-effective than control and/or eradication measures (Leung et al., 2002). In this context, Ecological Niche Models (ENMs), also known as BIoclimatic Models, Climate Envelopes, Habitat Models, Species Distribution Models, Range Maps, and Resource Selection Functions (Elith and Leathwick, 2009), have been applied to predict the potential distribution of exotic species (Jiménez-Valverde et al., 2011). ENMs are fitted to data from a species’ native area and are then used to identify suitable areas for the establishment of the invasive species in a new region (Peterson and Vieglais, 2001). Models can also be built using data from the native and invaded areas to predict the potential distribution of invasive species (Brommimann and Guisan, 2008). The models are constructed using a variety of modelling methods and combine species-occurrence records (geographical coordinates of the occurrence records) with a set of predictor variables (e.g., climate, land use type, and salinity). Models are used to predict suitable habitats in which species are able to maintain a population in order to persist through time (see Guisan and Thuiller, 2005 and Mateo et al., 2011 for reviews). Modelling methods are classified into two groups based on the type of occurrence-records input used to create the models: i) methods that use presence-only records (e.g., BIOCLIM and DOMAIN), and ii) methods that use presence and absence records (e.g., logistic regression and generalised additive model (GAM)) (Tsoar et al., 2007).

Some methods use pseudo-absence data (see Engler et al., 2004 to a definition of pseudo-absence data and ways to generate these data) for model construction (e.g., Genetic Algorithm for Rule-set Production (GARP) and Maximum Entropy (MAXENT)), but these are still classified as methods that use presence-only records because there is no real use of absence records in the construction of the model (Tsoar et al., 2007).

Scientometric studies use quantitative analyses to identify irregularities, patterns, or trends that may exist in publications of a given field of scientific research (e.g., Melo et al., 2006). For instance, in the case of biological invasions, two scientometric studies found a growing academic interest in invasion ecology in recent decades (Pysek et al., 2006; Qiu and Chen, 2009). In the area of ENMs, Cayuela et al. (2009) used publications from the period 1995-2007 to perform a scientometric study on the applications of ENMs to support conservation planning in tropical areas.

We conducted a scientometric study focused on the use of ecological niche models to predict the distribution of invasive species. We analysed papers published in peer-reviewed scientific journals from 1991 to 2010. Our main questions were: i) Is the number of papers on the use of ENMs to predict the distribution of invasive species increasing? ii) Is there a temporal trend in the quality or visibility (Scarano et al., 2009) of the journals, measured by their impact factor, in which these papers were published? iii) Which countries are the major publishers of papers using ENMs to predict the distribution of invasive species? iv) What are the main characteristics of the studies on this subject (predictor variables, methods, organisms, and regions studied)? v) What are the main gaps in the studies on this subject?

2. Material and Methods

We used the database of the Thomson Institute for Scientific Information (ISI; www.isiknowledge.com) to search for papers. The analysis was based on papers published between 1991 and 2010 that contained in the title, abstract, or keywords the following combination of words: “invasion* and ecological niche model* or bioclimatic model* or climate envelope* or habitat model* or species distribution model* or resource selection function* or range map*”. We collected the data from the Thomson ISI in April 2011.

We analysed each paper according to (i) year of publication, (ii) journal of publication and impact factor of each journal, (iii) number of citations, (iv) first author’s country, (v) region covered by the study, (vi) type of species-occurrence records (presence-only, or presence and absence), (vii) biological groups (algae, amphibians, birds, fish, fungi, insects, mammals, other invertebrates, plants, and reptiles), (viii) type of predictor variables (aquatic, climatic, human, land cover, land use, soil properties, topographic, and vegetation), (ix) spatial scale of the study (global, continental, national, regional, and local), (x) environment covered by the study (aquatic or terrestrial), and (xi) methods used to generate the models in each study. We also obtained the journal impact factors from the Journal Citation Reports (JCR) published in the year of publication of each paper (JCR 1990-2009). We used the scheme presented by Pearson and Dawson (2003) to assign the environmental predictors to different spatial scales of study.

We used a regression tree to identify possible trends over time in the number of papers on the use of ENMs to predict the distribution of invasive species. This method partitions the predictor variable in segments that are composed by similar values of the response variable. Each segment is then partitioned again and the partition process...
continues until the number of observations is considered small (De’Ath and Fabricius, 2000; see Melo et al., 2006 for a similar use of this method). We used the relative contribution (×1000) of papers in relation to the total number of papers published in a given year in all journals in the ISI database. We conducted the analysis using the package rpart (Therneau and Atkinson, 2010) in the R environment (R Development Core Team, 2010).

To test if the impact factor of the journals in which the papers were published increased through the years, we standardized the journal impact factor in a given year to the maximum impact factor for a journal in the field of ecology in the same year. We initially conducted a linear regression. However, because of the triangular arrangement of the data in the scatter diagram, we conducted a permutation test to evaluate whether this pattern could be generated by chance (Bardsley et al., 1999). The test evaluated simultaneously whether the mean and the variation of the impact factors increased in recent years. We used the software Ecosim (Gotelli and Entsminger, 2001), module “Macroecology” to conduct the analysis.

3. Results

A total of 190 papers related to the use of ENMs to predict the distribution of invasive species were published between 1991 and 2010. From 1991 to 1999, few papers were published, and in several years no paper appeared on this subject. The regression tree analysis partitioned the predictor variable (i.e., year of publication) in two segments, before and after 2003.5. The segment from 1991 to 2003 corresponds to the period with a low and relatively constant proportion of papers on EMS to predict the distribution of invasive species. The second segment (2004 to 2010) reflects the period with a trend of increase in the percentage of papers published (Figure 1).

The studies were published in 82 journals, although 56 of them contained only one paper and 10 contained only two papers. The 16 journals that published more than two papers on the distribution of invasive species using ENMs accounted for 60% (114 papers) of the total number of papers (Figure 2a). The journal Diversity and Distributions published 21 papers, followed by Biological Invasions (17 papers) and Weed Research (9 papers). The mean and the variation of the impact factors of the journals that published papers which used ENMs to predict the distribution of invasive species increased over the years (test for triangular arrangement of data, \( P = 0.040 \); Figure 2b).

Many papers received only 1-5 citations (62 of 190 papers), while 13 papers were never cited (Figure 3a). The most cited article was by Peterson and Vieglais (2001), which received 242 citations. Other heavily cited papers were by Peterson (2003), by Thuiller et al. (2005) and by Broennimann et al. (2007), which received 222, 158 and 112 citations, respectively. The papers by Kearney and...
Porter (2009), Broennimann et al. (2007), and Pearman et al (2008) figured among the most cited after we standardised the number of citations by the year of publication (i.e., divided the number of citations by the number of years since their publication) (Figure 3b).

Researchers from 23 countries published papers on ENMs to predict the distribution of invasive species (Figure 4a). Sixty-four papers were published by researchers from the United States of America, followed by Australia (21 papers), New Zealand (14 papers), Canada, Spain, and South Africa (11 papers, each). Following the same trend, the region most studied was the United States of America (36 papers). Global studies (32 papers) and North America (20 papers) were the second and third most studied regions, respectively (Figure 4b).

Species-occurrence records (n = 178 papers) were composed mainly of presence-only records (85.40%), while records on presence and absence were used in only 14.60% of the articles. Almost half of the studies were carried out with plants (85 of 181 papers). Insects were the second most investigated biological group (29 papers), followed by other invertebrates (15 papers), amphibians (11 papers), fish and reptiles (10 papers, each), birds (8 papers), fungi (6 papers), mammals (5 papers), and algae (2 papers).

We identified eight types of predictor variables used to construct the ENMs to predict the distribution of invasive species. Climatic variables – such as temperature and precipitation – were used in 55.18% of the articles, followed by topographic variables (22.22%). Land cover (4.44%), land use and vegetation (4.07%, each), aquatic – such as salinity and dissolved oxygen – (3.70%), soil properties (2.96%), and human – such as human populations and footprints - (2.22%) were the other types of variables used. The climatic variables were most often used in the global and regional scales (above 50%), while in the national to local scales other environmental predictors were used (Table 1). Additionally, 81.66% of the studies (147 of 180) were carried out in terrestrial environments, while only 18.34% of the studies investigated aquatic systems. Most studies in freshwater environments used only terrestrial predictor variables (16 of 32 papers) rather than using aquatic variables (12 papers) or both types of variables (4 papers). In contrast, studies in marine or estuarine environments used only aquatic variables or both types of variables (4 papers) to generate ecological niche models of invasive species.
Twenty-three different methods were used in 180 papers of the 190 papers we analysed (Figure 5). The Genetic Algorithm for Rule-set Production (GARP) method was the most used, appearing in 59 papers, followed by the CLIMEX (33 papers), the Maximum Entropy (MAXENT) (30 papers), and the logistic regression (LR; 23 papers) methods.

4. Discussion

Our results showed an increase in the number of publications on ENMs over time, which is probably related to the increasing interest in invasive species in recent decades (Pysek et al., 2006). In parallel to the increasing interest in biological invasions, the application of ENMs in different areas of ecology has been widely used (Guisan and Thuiller, 2005), contributing to the growth of studies on ENMs to predict the distribution of invasive species in the last decade.

Although most of the papers were published in only a few journals, the majority of these journals have high impact factors and are among the main journals in the subject categories of Ecology and Biodiversity Conservation. Further, the growing interest of ENMs to predict the distribution of invasive species is also apparent in the increases of journals’ impact factors, including journals with high impact factors in more recent years.

Citation frequency is also a criterion to quantify the impact and quality of a paper, although controversial (see Leimu and Koricheva, 2005 and Padial et al., 2010 for discussions). According to Garfield (2006), most published papers are never cited or cited only a few times. However, our results do not support clearly this suggested pattern since 60% of the papers were cited more than 5 times. Among the most cited articles, the one by Peterson and Vieglais (2001) explores the applicability of new bioinformatic tools (GARP) to predict species invasions, and was published in the beginning of the last decade when interest in ENMs to predict the distribution of invasive species began to increase. The other papers that were highly cited are a review (Peterson, 2003) and articles that tested new approaches and new tools to predict invasions (Thuiller et al., 2005; Broennimann et al., 2007). For instance, Peterson and Vieglais (2001) and Peterson (2003) created ENMs using data of the native region of the species, and thus assumed niche conservation across space and time. On the other hand, Broennimann et al. (2007) demonstrated through ENMs and additional analyses that a species may alter its niche during the invasion process. This means that some models created with data of the native region of the species may not predict the total region of invasion. Therefore greater attention is needed in interpretation of model predictions. The paper by Thuiller et al. (2005) is a broad study that builds multispecies projections to examine global risks of species invasions, in contrast to

Table 1. Percentage of each predictor used in ENMs to predict the distribution of invasive species within different spatial scales.

<table>
<thead>
<tr>
<th>Type of predictor</th>
<th>Spatial scale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Global</td>
</tr>
<tr>
<td>Aquatic</td>
<td>2.78</td>
</tr>
<tr>
<td>Climatic</td>
<td>80.56</td>
</tr>
<tr>
<td>Human</td>
<td>2.78</td>
</tr>
<tr>
<td>Land cover</td>
<td>-</td>
</tr>
<tr>
<td>Land use</td>
<td>-</td>
</tr>
<tr>
<td>Soil properties</td>
<td>-</td>
</tr>
<tr>
<td>Topographic</td>
<td>13.88</td>
</tr>
<tr>
<td>Vegetation</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 5. Number of studies carried out with different methods to generate ecological niche models to predict the distribution of invasive species. GARP = genetic algorithm for rule-set production; MAXENT = maximum entropy; LR = logistic regression; GAM = generalised additive models; GLM = generalised linear models; CART = classification and regression tree models; ANN = artificial neural networks; ENFA = ecological niche factor analysis; BRT = boosted regression trees; GBM = generalised boosted models; RF = random forest; MDA = mixture discriminant analysis; ClimEnv = climatic envelope; SRE = surface range envelope; MARS = multivariate adaptive regression splines; FuzzyEnv = fuzzy envelope; SVM = support vector machines; ED = environmental distance.
previous studies that focused on creating models for one specific species.

The United States of America was the country that showed the largest number of first authors and concentrated most of the studies. The position of the United States of America reflects its high investment in infrastructure and research (Fazey et al., 2005), providing basic data for the development of studies on ENMs to predict high-risk areas for invasions. This may allow researchers and governments to focus on prevention rather than eradication or control strategies. Similarly, in a recent bibliometric study, Qiu and Chen (2009) showed that research on biological invasions is mostly conducted in developed countries, following the general pattern noted by Pysek et al. (2008). Also interesting is the strong contribution of Australia and New Zealand, which figured out among the countries with the largest number of authorships and studies. This is likely due to the problems caused by invasive species (specially vertebrates and plants) to the remarkable and endemic biota of the pacific islands, including Australia and New Zealand. For instance, invasive vertebrates contributed to the extinction of many mammals and birds in both countries (Kingsford et al., 2009). The low representation of developing countries in studies on this subject may have several explanations, such as fewer resources for scientific studies and scarcity of data on exotic species (see Nuñez and Pauchard, 2010 for more explanations).

Species-occurrence records serve as the primary data for ENMs (Mateo et al., 2011), and the large number of studies composed by presence-only records may be attributed to the fact that presence records are easier to obtain and more reliable, since they typically derive from herbarium specimens, museum collections, and field observations by experts (Mateo et al., 2011). In contrast, absence records are rarely available, since species absence is more difficult to confirm, and often a recorded absence is actually nothing more than an undetected presence (Elith and Leathwick, 2009). Additionally, recent advances in biodiversity informatics and the development of extensive databases on biodiversity available via the Internet (Mateo et al., 2011) have facilitated the acquisition of presence-only records to generate ENMs. The implications in using presence-only records or presence and absence to create ENMs was discussed by Mateo et al. (2011).

Plants and insects were the biological groups most often used for predict the distribution of invasive species. Invasive insects and plants can cause severe economic problems in cropland production systems, urban environments, or natural environments (Pimentel et al., 2005) and, according to Pysek et al. (2008), it is the impact of the invasive species that determines whether or not it is studied. For instance, Pimentel et al. (2005) estimated that about 30% of the US$ 120 billion annually spent for invasive species in USA is directed to invasive plants. Despite the larger amount of studies on invasive plants than on invasive animals (Pysek et al., 2008; Qiu and Chen, 2009; our results), invasive insects have a prominent place in the list of invasive exotic fauna worldwide (Kenis et al., 2009), and this will probably lead to increasing interest in ENMs to predict the distribution of invasive species during the coming years.

We found that climatic variables are the type of predictor most used. The applicability of predictor variables is influenced by the spatial scale in the modelling process: at global, continental and national scales the climate appears to be the dominant factor determining species distributions, while at regional to local scales topography and land use become more important (Pearson and Dawson, 2003). This pattern is probably related to the fact that small spatial scales are associated with fine data resolutions, while large scales are associated with coarse data resolutions (see Pearson and Dawson, 2003 and Mateo et al., 2011 for more explanations). Our results showed that the spatial scale was also the determining factor in the choice of environmental predictors used to create the ENMs.

Most of the studies on ENMs to predict the distribution of invasive species were developed in terrestrial environments, whereas studies in aquatic environments are few. For many years, the attention of governments and scientists was focused on terrestrial invasive species (Pysek et al., 2008). Therefore, the large number of studies on ENMs conducted in the terrestrial environments is due to greater availability of information on these organisms (e.g., species occurrence records). Additionally, Puth and Post (2005) conducted a scientometric study on the invasion process using publications of the periods 1995-2005 and found more studies in terrestrial environments than aquatic environments at all stages of the invasion process. Ecological niche models in aquatic environments are limited because the most important predictor variables to determine the presence of a species (e.g., water temperature, salinity, and dissolved oxygen) are scarce or unavailable for most regions in the world (Ready et al., 2010). In freshwater environments, the aquatic variables are usually restricted to few sampling points (e.g., water monitoring stations), hindering the creation of ENMs (McNyset, 2005; Oliveira et al., 2010). Therefore, many studies in freshwater environments have used terrestrial predictor variables to generate ENMs (McNyset, 2005). However, it is worth noting that the use of terrestrial predictor variables can produce robust models (e.g., Hopkins, 2009; Kumar et al., 2009). In marine environments, the ENMs have been generated using marine predictor variables, because of the availability of several global databases (e.g., Integrating Multiple Demands on Coastal Zones with Emphasis on Aquatic Ecosystems and Fisheries – Incofish; NOAA World Ocean Database). Moreover, in aquatic environments, species-occurrence records are scarce, and so far, only a fraction of aquatic invaders are known (Ready et al., 2010).

GARP, CLIMEX, MAXENT, and logistic regression were the methods most used to predict the distribution of invasive species (see Elith et al., 2006 to explanations on advantages and problems of each method), compared with others. In this case, wide availability may be an important factor, because the three methods most used are software.

packages of easy accessibility and use, contrasting methods that require specialised knowledge.

Moreover, GARP, CLIMEX, and MAXENT are methods that use presence-only records as the primary data to create ENMs, and therefore their use is favoured by the greater availability of presence records than of absence records. Comparative analyses of the statistical performance of GARP and MAXENT are available (Kumar et al., 2009; Colombo and Joly, 2010; Oliveira et al., 2010; Terrible et al., 2010). The CLIMEX method has been mainly applied to evaluate the invasion potential of exotic organisms (Kriticos et al., 2003). Logistic regression is also frequently used in ecological niche models, although it belongs to the group of presence-absence methods (Guisan and Thuiller, 2005).

Our scientometric analysis showed a growing interest and popularity in the use of ENMs to predict the distribution of invasive species, especially in the last decade. However, some important gaps need to be filled, such as the relatively small numbers of studies conducted in developing countries and in aquatic environments. The lack of studies on these two issues cannot be scientifically justified, since many developing countries harbor the highest biodiversity in the world (Nuñez and Pauchard, 2010), and invading species are a major concern in biodiversity conservation. Detailed data on invasive species distribution are usually not available in developing countries or its availability is limited and little disclosure (Rodríguez, 2001), forming a gap in the construction of predictive models for invasive species in these regions. However, it is worth noting that the few papers on ENMs to predict the distribution of invasive species carried out in developing countries found in the present study does not necessarily represent a total lack of studies, but may be related to the fact that such studies are only available in other small or regional databases. Moreover, aquatic environments are more vulnerable to invasive species than terrestrial environments (Ready et al., 2010). Ecological niche models can be used for strengthening the development and use of preventive measures to deal with invasive species. Therefore, basic information, records of occurrence of invasive species and predictor variables, are urgently needed to that researchers can devote more effort to studies of ENMs to predict the distribution of invasive species, both in developing countries and in all types of ecosystems.

Acknowledgements – We thank A.A. Pedal and three anonymous referees for suggestions that improved the manuscript. Janet Reid reviewed the English. FGB and FS received student fellowships from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). ASM received research grants (476304/2007-5; 474560/2009-0; 558187/2009-9) and a research fellowship (302482/2008-3) from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

References


READY, J., KASCHNER, K., SOUTH, AB., EASTWOOD, PD., REES, T., RIOUS, J., AGBAYANI, E., KULLANDER, S. and FROESE, R., 2010. Predicting the distributions of marine...


