The application of predictive modelling of species distribution to biodiversity conservation

Jon Paul Rodríguez1*, Lluís Brotons2,3, Javier Bustamante1 and Javier Seoane6

1Centro de Ecología, Instituto Venezolano de Investigaciones Científicas, Apdo. 21827, Caracas 1020-A, Venezuela, 2Centre Tecnològic Forestal de Catalunya, Area de Biodiversitat, Pujada del Seminari s/n, 25280 Solsona, Spain, 3Institut Català d’Ornitologia, Museu de Zoologia, Paseig Picasso s/n, 08003 Barcelona, Catalonia, Spain, 4Department of Applied Biology, Estación Biológica de Doñana, Consejo Superior de Investigaciones Científicas, Avda. María Luisa S/N, 41013 Sevilla, Spain, 5Departamento Interuniversitario de Ecología, Facultad de Ciencias, Universidad Autónoma de Madrid, 28049 Madrid, Spain

*Correspondence: Jon Paul Rodríguez, Centro de Ecología, Instituto Venezolano de Investigaciones Científicas, Apdo. 21827, Caracas 1020-A, Venezuela. E-mail: jonpaul@ivic.ve

INTRODUCTION

Predictive modelling of species geographical distributions is a thriving ecological and biogeographical discipline. Major advances in its conceptual foundation and applications have taken place recently, as well as the delineation of the outstanding challenges still to be met (Araújo & Guisán, 2006; Guisan et al., 2006; Moisen et al., 2006; accompanying papers). In this article, we focus on the application of predictive distribution modelling to biodiversity conservation. We begin by outlining a series of ‘real-life’ conservation problems that have been addressed with predictive modelling. Next, we consider the need to integrate these techniques with the development of systematic baseline data sets and models, in order to effectively monitor biodiversity change. The following section considers the availability of remotely sensed ecological data for use in predictive modelling, and introduces the readers to their major on-line sources. These general remarks provide a backdrop for a series of articles collected in this issue of Diversity and Distributions that emerged from a workshop on Predictive Modelling of Species Distributions, New Tools for the XXI Century, carried out under the auspices of the Universidad Internacional de Andalucía, sede Antonio Machado in Baeza, Spain, on 2–4 November 2005. The final section of this article provides an introduction to the collection of papers.

APPLICATIONS TO ‘REAL-LIFE’ PROBLEMS

Keeping up-to-date with the conceptual and technical advances of predictive modelling of species distribution is a challenge in itself, because it comprises a rapidly evolving set of versatile tools with a wide range of applications (Guisan & Zimmermann, 2000; Guisan & Thuiller, 2005). Predictive models have been successfully used in a variety of conservation biology studies, with the main purpose of addressing pressing conservation problems (e.g. Brotons et al., 2004; Linkie et al., 2006). Confronted with extremely limited data on the distribution, abundance, and dynamics of most species on Earth (Rodríguez, 2003; Mace et al., 2005), models allow for the extrapolation of relatively few field samples and museum records to the entire potential range of a species, thus creating predicted distribution maps (e.g. Guisan & Zimmermann, 2000; Heglund, 2002; Anderson et al., 2003; Guisan & Thuiller, 2005). To carry out this task, detailed and reliable environmental data are needed to establish the species–environment relationships required to estimate species distribution outside sampling sites. The collection of field data using environmentally stratified sampling designs assures that the range of environmental variation is covered, and improves these predictions (Austin & Heyligers, 1991; Margules & Austin, 1991; Hortal et al., 2001; Hortal & Lobo, 2005; J.R. Ferrer-Paris & J.P. Rodríguez, unpubl. data). Once data on species distribution have been collected in the field or derived from museum collections, and adequate environmental models have been fitted to the data, it is possible to generate predictive distribution maps. A number of potential applications of these generated maps rapidly arise in the field of biodiversity conservations and planning (provided that beyond statistical evaluations, also managers and planners are satisfied with the predictions provided by modellers). It is possible, for example, to assess the representation of species in nature reserve networks, evaluate their response to global climate change (of both the species and the reserves), quantify the impact of expected land cover change, and evaluate the influence of expanding exotic species, among other applications.

Growth and potential of predictive modelling techniques and approaches is boosting the applications of biodiversity mapping in recent years. A good example of structure construction of a discipline that has grown in parallel to biodiversity mapping is reserve design. Inspired by pioneering work carried out in Australia in the 1980s (Kirkpatrick, 1983), the Gap Analysis Program...
(GAP) in the USA provided a major thrust to the application of predicted modelling to conservation (Scott et al., 1993; Jennings, 2000; Peterson & Kluza, 2003). This programme has the main objective of creating and updating maps of the predicted distribution of native vertebrate species in the USA. By collating information from field research programmes and modelling their distribution using a variety of tools, GAP has become a powerful source of basic cartographic information for a number of planning and management decisions at the local, state, and national scale (Pearls tine et al., 2002; Noon et al., 2003). After generating predicted ranges, the next step in the process consists of overlaying these distributions with the geographical extent of the protected areas network, thus identifying existing ‘gaps’ — species that are not represented in the network — and providing potential locations for the future expansion or creation of new protected areas.

Gap analysis has been proven highly successful and has become a very popular conservation tool. Its use is not limited to predicting geographical distributions, but also takes advantage of other sources of data on species ranges as well (e.g. De Klerk et al., 2004; Rodrigues et al., 2004).

The ample literature on reserve selection has been primarily devoted to optimizing the number, size, location, and cost of reserve networks, in order to assure that they include the largest possible fraction of biodiversity within the smallest amount of protected land and, ultimately, identify priority areas for conservation (Margules et al., 2002; Wilson et al., 2006; Alagador & Cerdeira, in press). Predictive modelling has enriched the discussion by including habitat suitability as an important characteristic to consider in the design of a network of reserves, because habitat suitability has been shown to correlate with the persistence of populations (Araújo et al., 2002; Cabeza et al., 2004). In an analogous way, habitat suitability has been incorporated into metapopulation models, which are primarily concerned with patch shape, size, and isolation (Akcakaya & Atwood, 1997; Kramer-Schadt et al., 2005).

Regardless of how well the current reserve networks represent extant biodiversity, however, they are all threatened by global change, a series of complex forces that combine both climate change and land-use change (Araújo, 2004; Thomas et al., 2004; Whittaker et al., 2005; Araújo & Rahbek, 2006; Araújo et al., 2006; Root & Schneider, 2006; Schlesinger, 2006; Sutherland, 2006; Thuiller et al., 2006a,b). A number of studies, many based on habitat modelling, have raised the concern that species distributions may shift in response to climate change (mainly towards higher latitudes and altitudes), so that some may be forced to abandon reserves that were designed to protect them, and move to unprotected areas — if they can. The problem is that the current distribution of a particular species may be constrained by unsuitable surrounding habitats (e.g. perturbed areas, topographical barriers) or there may be a physical limit to their range, as is the case of species that live on mountains and cannot continue to shift upwards indefinitely (Dirnbock et al., 2003; Thuiller et al., 2006a). In fact, extensive modelling exercises predict an overall loss of biodiversity over the next 50 years as a result of climate change, in particular for mountain species (Bakkenes et al., 2002; Thomas et al., 2004). These predictions have been challenged, however, because the results tend to vary with the technique applied (Thuiller et al., 2004; Pearson et al., 2006), and broad-scale models do not capture local interactions between organisms and their environment (Hampe, 2004), such as dispersal, interspecific interactions, or the existence of refuges at a finer spatial resolution than the climate change scenarios on which the models are based (Post et al., 1999; Araújo & Pearson, 2005; Pearson, 2006). Although most studies are pessimistic about the extent of future range contractions, this is still an ongoing debate that relies heavily on habitat modelling.

Another promising but less explored line of research is the use of predictive habitat models to forecast conflicts between human activities and biodiversity conservation, and to try to minimize these conflicts. This is the case of the assessment of the impact of land-use changes linked to evolving agricultural practices. In Europe, for example, the simultaneous abandonment of traditional agriculture (areas that harbour several endangered species) and the intensification of agriculture on more profitable lands have led to a net decrease of biodiversity (Donald et al., 2002; Newton, 2004). Environmental and agricultural policies tend to aim at different targets, and predictive models have been helpful for decision-makers to identify areas where these opposing forces may clash (Brotons et al., 2004; Nogues-Bravo & Agirre, 2006; Seoane et al., 2006). Similarly, maps that predict the likelihood of conflicts between carnivores and humans (due to livestock predation) have also been developed, guiding planners towards geographical locations where predation events are less likely to occur (Treves et al., 2004). Another common conflict arises when human infrastructures, such as airports, roads, and power lines, are in the path of the regular movement of animals. For example, large birds pose a threat to airplanes because of bird strikes, particularly during the migratory season when birds gather in large flocks (http://ecogrid.sara.nl/bambas/distribution/index.php). Collision with power lines and road kills have been linked to the local decline of endangered species (Ferrer & Janss, 1999; Guzmán et al., 2004), while some studies have modelled higher risk areas where corrective actions should be prioritized (Garthe & Huppop, 2004; Malo et al., 2004). Again, predictive models stand out as a useful tool for land-use planners seeking to make better decisions about biodiversity management and conservation.

Invasive species are one of the major drivers of global change, threatening numerous other species with extinction (Vitousek et al., 1996; Chapin et al., 2000). Biologists and agronomists initially began to search for traits that made species more aggressive colonizers, and for characteristics of ecosystems that made them more susceptible and vulnerable to invaders, in order to predict, mitigate, and perhaps prevent the negative consequences of biological invaders on native taxa (Crawley, 1986, 1987; D’Antonio & Vitousek, 1992; Crawley et al., 1996; Mack, 1996; Veltman et al., 1996; Williamson & Fitter, 1996; Williamson, 1999; Mack et al., 2000; Huston, 2004; Richardson, 2004; Callaway & Maron, 2006). More recently, predictive habitat models have been used to map areas of high risk of being invaded (or that are currently more affected) both regionally (Mercado-Silva et al., 2006; Muñoz & Real, 2006) and globally (Thuiller et al., 2005). The
resulting picture can hardly be more worrying: six global biodiversity hotspots are highly susceptible of being invaded (Thuiller et al., 2005).

On a relatively less ambitious scale, there are a number of additional applied questions that are being addressed with predictive models. Ecologists and field biologists, for example, are keen to use predictive models to generate potential habitat maps to plan for the reintroduction of species, and to evaluate re-introduction schemes underway (Mladenoff et al., 1999; Schadt et al., 2002; Hirzel et al., 2004; Pellet et al., 2004). Although re-introducing species into a portion of their historical range tends to be more successful than placing them elsewhere (Griffith et al., 1989), the ‘original’ habitat is not always available (in fact this may be the cause of the disappearance of the species), and it may be necessary to look for new sites. But even when the original range is still a viable option, maps of potential suitable habitat help to recognize barriers to the dispersal of released individuals, to identify isolated areas unlikely to be colonized naturally, and to quantify the carrying capacity of the managed territory. These maps have also helped to detect species behavioural responses to fragmentation that could be used as early warnings of human perturbation (Laiolo & Tella, 2005).

**BASELINE MONITORING AND SPATIAL MODELLING OF SPECIES DISTRIBUTION**

Conservation, planning, and species management need basic monitoring data in order to gather critical information about where a species occurs and how its abundance changes through time (Underhill & Gibbons, 2002). Basic monitoring is aimed at sampling populations in order to describe their distribution in space (Donald & Fuller, 1998) and time (Bibby et al., 2000).

Monitoring of species distribution has a strong link to predictive, habitat-based modelling techniques because both have the explicit objective of producing a spatial representation of a species’ range. In fact, the development of predictive, habitat based models has been possible due the large amounts of data accumulated in large-scale distribution monitoring programmes such as atlases (Osborne & Tigar, 1992; Tobalske & Tobalske, 1999; Araújo & Pearson, 2005). But predictive distribution modelling has not only grown due to the availability of monitoring data. More recently, these techniques have started to play a key role in producing basic quantitative information on species distributions, allowing atlases move beyond their classic ‘black and white’ maps derived from grid mapping of presence–absence data. These new methodological approaches also offer testable predictive models that allow a more adequate representation of species distributions in a context of recognized but often unquantified uncertainty (Elith et al., 2002; Barry & Elith, 2006).

Predictive models have already been applied to a number of atlases, for example, mammals (Hauser, 1995) and birds (Estrada et al., 2004). Since atlases are a primary source of species distribution data for planning and species management conducted by governmental agencies, non-governmental organizations (NGOs), academia, and the private sector, the progressive inclusion of predictive habitat modelling will allow for the production of an increasing number of detailed and testable maps, therefore increasing their potential impact and reliability.

For a number of conspicuous species, monitoring often involves counts to estimate their abundance in a given area. Population size is a fundamental ecological parameter, often required by local and international organizations and legislations, to assess the conservation status of a species, on the grounds that it is inversely correlated with the probability of extinction (IUCN, 2001; O’Grady et al., 2004). However, the development of a standard sampling scheme to correctly make inferences from the numbers observed in the field to the entire study area is sometimes unfeasible. For example, studies on sea mammals commonly resort to the affordable technique of sampling from fishing or recreational boats (so-called platforms of opportunity), whose bearings are neither random nor systematic, and consequently any design-based estimate of abundance may be biased (Williams et al., 2006). In contrast, model-based estimates are able to tackle the problem and have the further advantage that they provide maps at a higher spatial resolution (map units in model-based schemes are normally smaller than strata in design-based ones), as well as potential insights into the variables affecting abundance (Clarke et al., 2003; Hedley & Buckland, 2004). These spatial modelling methods applied to field data have a large number of applications to conservation (see for example http://www.ruwp.st-and.ac.uk/Research/SpatialModelling/index.htm).

Another promising application of habitat-based predictive modelling techniques to basic monitoring is less obvious. Typically, long-term monitoring programmes aim at detecting changes in population sizes. However, since they are generally based on a spatial sample of locations, these data also have the potential for producing species maps through interpolation or habitat modelling techniques (Sauer et al., 1995; J.R. Ferrer-Paris & J.P. Rodriguez, unpubl. data). Recent attempts of using this kind of data have offered promising results as mapping tools (Carrascal et al., 2006; Brotons et al., 2007), including the incorporation of sampling biases via modern Bayesian estimation techniques (Thogmartin et al., 2004, 2006).

Habitat models have another immediate link to monitoring efforts, since model predictions may help fill gaps in unsurveyed areas (as in atlas works) or to identify promising areas for future sampling, such as, for example, potential breeding nuclei for species in places yet to be explored (e.g. Seoane et al., 2003; Jarnevich et al., 2006; Brotons et al., 2007). These spatial models have been shown to be an improvement on traditional atlas and reference works (Bustamante & Seoane, 2004; Bini et al., 2006). Along the same line, drawing maps for sibling species whose identification relies on molecular characteristics is particularly demanding, and building habitat models with unequivocal genetic information is a cost-effective way to depict distributions (Real et al., 2005).

The potential increase in the application of monitoring data to spatial mapping of species distribution suggests a trend towards more blurred boundaries between the objectives that these two activities cover. A promising approach is the integration of spatial mapping, including predictive modelling techniques, into basic, temporal monitoring of population trends in order to generate useful information on both changes in numbers and spatial...
distribution of the monitored species (Travaini et al., 2007). This may allow, for instance, the development of spatial alerts linked to temporal alerts, indicating worrying changes in species distribution and abundance (Greenwood, 2003). Traditional mapping projects, such as atlases, appear in this context as a fundamental complement, offering a detailed spatial picture of species distributions over a relatively small temporal scale.

In areas where sampling is expensive due to lack of observers or poor communications, representative sampling covering existing environmental variability offers an optimal allocation of limiting resources for monitoring (Hortal et al., 2001; J.R. Ferrer-Paris & J.P. Rodríguez, unpubl. data). Such a sampling strategy is known to allow a potential representative estimation of species distributions through predictive modelling (Faith & Walker, 1996). In addition, such data may stand as baseline information for future sampling aimed at both better estimation of species distribution patterns or changes in population numbers or occurrence rates.

We think that the evidence shows that the use of predictive modelling of species distribution has been enormously useful and that conservation would benefit from a more explicit and structured integration of these techniques into basic monitoring programmes. We encourage governmental agencies, NGOs, academia, and the private sector, to develop basic monitoring programmes that incorporate as one of their cornerstones habitat modelling approaches to create and update species distributions.

A fundamental limitation that still exists, however, is that current biodiversity monitoring efforts are extremely limited in terms of their spatial and taxonomic coverage, survey methods are not standardized, and data sets collected at different scales lack integration (Pereira & Cooper, 2006). Worldwide monitoring initiatives such as the Global Amphibian Assessment (Stuart et al., 2004) and the Biodiversity Intactness Index (Scholés & Biggs, 2005), rely heavily on expert opinion rather than on field data, museum collections, or predicted distribution modelling (Mace, 2005). For only the better-known taxonomic groups, such as birds and mammals, are there maps of their geographical distributions (Patterson et al., 2003; Ridgely et al., 2003), and only in a small number better-known locations is there information on their relative abundance and temporal dynamics (e.g. Gibbons et al., 1993; Sauer et al., 2005). The challenge is particularly large in the tropics, which harbour the majority of the world’s species, but lack the financial resources and personnel for conservation (Rodríguez, 2003; Rodríguez et al., 2005). The design, implementation, and funding of systematic biodiversity surveys at the scale of entire tropical countries, which minimize the costs and time required to perform them, apply standardized methods, quantify both distribution and abundance, and generate data that are adequate for species distribution modelling, are a major gap in global biodiversity monitoring that needs to be addressed.

**AVAILABILITY OF ENVIRONMENTAL CARTOGRAPHY**

Using species distribution models to generate maps relies to a great extent on the availability of environmental cartography. There can be a strong correlation between an environmental variable and the distribution of a species, but if this variable is not available as a map, it cannot be applied to predict the distribution of the species. Traditionally, the availability of environmental cartography has been very dependent on the economic development of a country. Also, even if environmental maps exist, there has been a wide variability in the access to those data, especially if they were considered to hold commercial or strategic value. During the last two decades several changes in the status quo have quickly improved access to high-quality environmental cartography. Earth observation remote sensing mission has increased steadily since 1972 (Fig. 1), and sensors on those satellites provide environmental data at a rate never imagined before. Although some data sets still need to be purchased at fairly high costs, others are readily available (see Appendix S1 in Supplementary Material). Some require the users to demonstrate a scientific pursuit (rather than a commercial one), but most can be freely downloaded. Other sources of data, such as global climate maps based on modelling of records from meteorological stations are also accessible on the Internet (e.g. Worldclim, http://www.worldclim.org/, Hijmans et al., 2005). Although the spatial and temporal resolution of most data sets are adequate for species distribution modelling for both terrestrial and marine organisms, they have not been widely used. Notable exceptions include the work of Munoz et al. (2004), in which data of the NASA Wind Scatterometer were used to model landmass connectivity for plant dispersal in the southern hemisphere, and Travaini et al. (2007), in which Spot-Vegetation and Shuttle Radar Topographic Mission data were used to model guanaco distribution. Species distribution models, coupled with wider availability of environmental cartography, and relatively inexpensive and user-friendly geographical information systems (GIS), will surely result in a significant and rapid change in our knowledge of the distribution and abundance of biodiversity in the decades to come.

**INTRODUCTION TO THE INVITED PAPERS**

This special volume includes both methodological and applied works. Among those of a methodological nature, Ferrier et al.
(2007) introduce the technique of Generalized Dissimilarity Modelling (GDM), a new statistical method focused on modelling the distribution of biotic communities based on compositional dissimilarity. Ferrier et al. (2007) describe for the first time the technique in detail and present several examples of its use. Although originally designed to predict distributions of communities, it is a flexible approach that can also be accommodated to model individual species. Variations of the method can be used to predict the spatial distribution of phylogenetic dissimilarity, to regionalize areas based on species distribution patterns, or to include information on barriers to dispersal in the modelling. Elith & Leathwick (2007) explore the use of Multivariate Adaptive Regression Splines (MARS) to model presence-only data. MARS constitutes an extension of the better-known generalized additive models (GAMs), directed at modelling communities instead of individual species. Both GDM and MARS are examples of new techniques directed at modelling species groups to solve the common problem of species lacking enough information to be modelled as individual entities (Pearce & Boyce, 2006).

Two other new techniques: Geographically Weighted Regression (GWR) and Varying Coefficient Modelling (VCM) are explored by Osborne et al. (2007) to address the problem of non-stationarity in ecological systems. These new methods are complementary to global techniques, and provide details and habitat associations at local scales that global methods miss (Foody, 2005; Jetz et al., 2005). Guisan et al. (2007) also consider aspects of scale, and explore the effect of grain size in species distribution models to try to answer the question of whether there is an optimal grain size. They present the first extensive test of the effect of coarsening the resolution of environmental variables in species distribution models, and show that, although there is a general trend towards degradation of model performance with resolution coarsening, it does not affect equally all data set, species, or modelling techniques. In their analyses, a new modelling technique, Boosted Regression Trees (BRT), ranks as the best among the 10 they used (including building environmental envelopes and regression methods). The papers of Elith & Leathwick (2007) and Guisan et al. (2007) consider the important aspect of building distribution models with presence-only data from herbaria and museum collections, often the only source of information on historical distribution changes and sometimes an important source of information for rare species.

A very important aspect considered in this special issue is the practical application of species distribution models. Travaini et al. (2007) provide an example on how new technologies can be integrated to reduce the costs of mapping species distributions. New technologies include statistical modelling techniques, as well as the use of freely available remote sensing data, accurate georeferencing of species observations with global positioning systems (GPS) and laser range-finders, optimal data gathering using personal digital assistants (PDAs) and customized software, and the use of GIS for extrapolation and map production. Their results show that new technologies constitute very promising tools for large and remote regions that lack accurate species distribution maps. Even in areas where apparently reliable distribution data exist, species distribution modelling techniques have been shown to provide more accurate maps (Bustamante & Seoane, 2004). The article by Brotons et al. (2007) highlights the potential role of species distribution modelling techniques to improve and update our knowledge of species ranges using data from long-term monitoring programmes, which tend to be common in developed countries. Anadón et al. (2007) provide a practical example on the use of a hierarchical species distribution model to identify the causes of decline of an endangered species. Richardson & Thuiller (2007) use climate matching models to successfully identify species that may become potential invaders, and provide a technique to potentially design preventive measures for some species. As they emphasize, the most cost-effective way of reducing future problems is to prevent the introduction of high-risk invasive species.

REFERENCES


**SUPPLEMENTARY MATERIAL**

The following supplementary material is available for this article:

Appendix S1 Environmental predictors of global extent accessible through the Internet.

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1472-4642.2007.00356.x

(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.