



Accounting for uncertainty when mapping species distributions: The need for maps of ignorance

Progress in Physical Geography
35(2) 211–226

© The Author(s) 2011

Reprints and permission:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0309133311399491

ppg.sagepub.com



Duccio Rocchini*

IASMA Research and Innovation Centre, Italy

Joaquín Hortal*

Museo Nacional de Ciencias Naturales, Spain, and Imperial College London, UK

Szabolcs Lengyel

University of Debrecen, Hungary

Jorge M. Lobo

Museo Nacional de Ciencias Naturales, Spain

Alberto Jiménez-Valverde

The University of Kansas, USA

Carlo Ricotta

Università La Sapienza di Roma, Italy

Giovanni Bacaro

University of Siena, Italy

Alessandro Chiarucci

University of Siena, Italy

Abstract

Accurate mapping of species distributions is a fundamental goal of modern biogeography, both for basic and applied purposes. This is commonly done by plotting known species occurrences, expert-drawn range maps or geographical estimations derived from species distribution models. However, all three kinds of maps are implicitly subject to uncertainty, due to the quality and bias of raw distributional data, the process of map building, and the dynamic nature of species distributions themselves. Here we review the main sources of uncertainty suggesting a code of good practices in order to minimize their effects.

Corresponding authors (*both authors contributed equally to this work):

Duccio Rocchini, Fondazione Edmund Mach, Research and Innovation Centre, Department of Biodiversity and Molecular Ecology, GIS and Remote Sensing Unit, Via E. Mach 1, 38010 S. Michele all'Adige (TN), Italy

Email: duccio.rocchini@iasma.it, ducciorocchini@gmail.com

Joaquín Hortal, Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales (CSIC), C/José Gutiérrez Abascal 2, 28006, Madrid, Spain

Email: jhortal@mncn.csic.es

Specifically, we claim that uncertainty should be always explicitly taken into account and we propose the creation of maps of ignorance to provide information on where the mapped distributions are reliable and where they are uncertain.

Keywords

biodiversity, biogeography, conservation, macroecology, species distribution maps, species distribution modelling, uncertainty

A map must always reflect the quality of the information on which it is based. (Stott, 1981: 31)

I Introduction

Uncertainty is a direct consequence of the complexity of nature. All the facets of a complex phenomenon are impossible to measure with absolute precision, because the act of measuring itself affects its perception (Heisenberg, 1958). It follows that any measure has an associated measurement error of unknown extent (Klir and Wierman, 1999), and that measurement in general and the statistical quantification of uncertainty in particular are core parts of scientific inquiry. The quality of both data and results are central to the reliability of the answers to any question evaluated quantitatively (Elzinga et al., 2001). Hence, good practice in science necessarily includes the assessment, statement and acknowledgement of measurement error: any technical assessment, monitoring program or scientific research based on data should thus include an evaluation of the uncertainty of its results.

The description of the distribution of biodiversity at different spatial and temporal scales has long been the focus of ecology and biogeography. Reliable descriptions of species distributions are fundamental for conservation and research purposes (see Cox and Moore, 2004; Dormann, 2007b). Therefore, creating maps of where each species occurs is an essential goal of biogeography (Gillespie et al., 2008; Rushton et al., 2004). Given that the development of distribution maps is a way of measuring the geographical distribution of biodiversity, these maps are in fact

measures, and hence should have their associated uncertainty (Hortal, 2008).

Here we review the most important sources of such uncertainty, from the quality of the original data on species distributions, to the process of mapping distributions using the currently available methods, with a particular emphasis on *species distribution models* (SDMs). By SDMs we refer to both *potential* and *realized* distribution models sensu Jiménez-Valverde et al. (2008b). Potential distribution models (ecological niche models sensu Soberón, 2010) aim to identify the places where a species could live at a given moment of time; realized distribution models (distribution models sensu Soberón, 2010) try to recover the places where a species currently lives.

Based on our review, we argue that maps of species distributions should undergo more rigorous testing than is currently common, including the quantification of uncertainty, before they are widely used in macroecology, biogeography and conservation biology. In consequence, we propose and discuss different ways of quantifying and mapping error and uncertainty, which should be incorporated into good practice in biodiversity mapping.

II Quality and bias of distributional data

Raw data on the presence of species gathered from museums and herbaria and/or recorded from field surveys are increasingly used to build species distribution maps. For the sake of clarity, we herein will distinguish between (1) *collection data*, which refers to the original field data, and

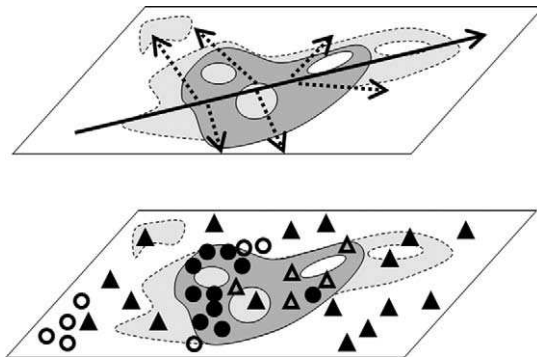


Figure 1. Common errors and biases in distributional data. The upper map shows the actual (dark grey, continuous border) and potential (light grey, discontinuous border) distributions of a species in a certain region. Arrows represent the main (continuous) and secondary (discontinuous) environmental gradients that shape the species response to the environment. By main gradient we refer to the spatial variation in the environmental factor(s) that constrain the most the distribution of the species in the studied area, typically due to its ecophysiological limitations and/or other ecological requirements. Secondary gradients would thus be spatial variations in other factors that shape the distribution of the species, although in a less important way. The lower map shows the data available to represent the distribution of the species. Circles represent recorded presences, and triangles represent absences, which often are assumed. Full symbols represent true presences/absences, and empty symbols are false ones. Note the important bias in the allocation of observed presences, which limits the representativeness of the data. In this case, both the distribution and environmental response of the species are only partly covered by the data. False presences placed near the actual distribution could come from positional errors and/or observations of vagrant individuals (or currently extinct populations). False presences placed far away from the actual distribution could represent taxonomic errors. True absences can be allocated in areas that are either unsuitable for the species, or that are suitable but currently do not host any population (i.e., environmental and contingent absences, respectively; *sensu* Lobo et al., 2010). False absences could be caused by lack of surveys and/or failure to detect the species during the survey.

(2) *Atlas data*, which refers to the direct use of field data for mapping (as in Hortal, 2008; see below). Most collection data available in natural history collections have been obtained without any planned sampling scheme, and their geographical locations are often uncertain and/or imprecise (Chefaoui et al., 2005; Dennis and Thomas, 2000; Dennis et al., 1999; Elith et al., 2006; Hortal et al., 2007, 2008; Lobo, 2008b; Loiselle et al., 2008; see Figure 1 and Table 1). As a result, such data are far from providing an accurate picture of the distribution of most species (the so-called ‘Wallacean shortfall’; Lomolino, 2004; Whittaker et al., 2005).

Uncertainty in distributional data can originate from three main sources: (1) the inherent complexity of species distributions; (2) the quality of the available data; and (3) the limitations of the measurement system used to gather information on the presence of the species (i.e. field surveys and taxonomic identifications). The distribution of a species is particularly difficult to describe because it is the outcome of many complex and dynamic phenomena that vary across spatial and temporal scales. Depending on the scale considered, the presence of a species is affected by its intrinsic population dynamics, environmental adaptations and/or dispersal capacity, as well as by landscape and environmental complexity (Chust et al., 2003; Holt et al., 2005; Nagendra, 2001; Ricotta and Anand, 2006; Rocchini and Ricotta, 2007; Soberón, 2007, 2010; Whittaker et al., 2001). While the uncertainty associated with such inherent complexity is difficult to assess, data quality and the limitations of the measurement system can be described and studied by examining data bias, completeness, precision and accuracy (see Kadmon et al., 2003; Walther and Moore, 2005).

I Data quality

The quality of biodiversity data has been extensively studied during the last decades (e.g. Dennis and Thomas, 2000; Dennis et al.,

Table 1. Summary of the main sources of uncertainty, the problem(s) related to each one of them and the possible solutions that would constitute good practices for mapping and modelling species distributions. We also indicate studies that have applied the solution proposed. See the text for further explanations; 'Input data' mainly refers to section 'II Quality and bias of distributional data', and 'Modelling and mapping procedures' to section 'III Uncertainty and biases in species distribution mapping and modelling'.

Source of uncertainty (problem)	Possible solution (good practice)
Input data	
Observer bias and species misidentifications.	Implementing concept taxonomies allows avoiding misinterpretation of presences and absences of species (Berendsohn, 1995).
Species records derived from museum or herbarium collections with no planned sampling scheme.	Possibly relying only on data deriving from robust sampling design developed based on statistical sampling theory (e.g. Funk et al., 2005; Hortal and Lobo, 2005; Hortal et al., 2007).
Uncertain coordinates for localities, e.g. toponyms.	Possibly relying only on data based on certain localization (coordinates and reference system) (e.g. Elith et al., 2006). Correcting the location of data based on additional information from field data or museum labels (e.g. Chefaoui et al., 2005).
Positional error associated to presence records. Different resolutions of input data.	Carrying out multiscale models and accounting for potential positional errors by downscaling data may allow detection of possible discrepancies (Johnson and Gillingham, 2008; Wisz et al., 2008).
Bias in the surveyed areas towards classical localities of particular natural beauty, localities that were species-rich in the past, or areas near the experts' residences and/or research centres.	These areas could be profitably used as well-sampled areas to assess model accuracy. Notice that it cannot be ruled out that a species is absent from many areas simply because these areas are less studied than the areas where the species has been recorded (e.g. Hortal et al., 2007).
Only presence of a species is recorded. No 'certain' absences are recorded. A fundamental source of uncertainty in distributional data comes from assuming the absence of a species from places where it is actually present but remains undetected.	Quantifying sampling intensity could in turn be used to determine the likelihood of absence from areas that were sampled with enough effort to provide reliable inventories (e.g. Hortal and Lobo, 2005; Lobo, 2008a).
Modelling and mapping procedures	
Broad-scale maps of species distributions are often more an art than a science.	Caution should be taken to avoid considering such maps as the 'truth' (e.g. Stohlgren, 2007).
Projections derived by different methods may differ so as to make them useless for diversity conservation. Model selection methods inform on the most adequate model for the data set used, not on their true accuracy.	The uncertainties in estimates and predictions should not be reported as if the model was correct.

(continued)

Table 1 (continued)

Source of uncertainty (problem)	Possible solution (good practice)
Range map data generally overestimate species richness at finer spatial grains (see Figure 2).	Testing different methods and reporting their associated errors may be more honest than assuming a single model to be the best choice a priori. It is crucial to report uncertainties in estimates and predictions as output statistics, maps of residuals, etc. (e.g. Beaumont et al., 2007; Kühn et al., 2006; Murphy et al., 2004; Palmer et al., 2005; Pearson et al., 2006). Keeping in mind that input data are not bias-free, it may be better to rely on input data or on grid superimposition to field records such as Atlas data. In fact, fine-scale distributions are much patchier than those achieved even by the most detailed expert-drawn range maps (e.g. Hurlbert and Jetz, 2007; see also Hortal, 2008).
Variable collinearity and selection of predictors.	In some cases, manual selection based on good ecological hypotheses may outperform statistically based procedures such as stepwise selection since it allows selection of predictors based on biological theory and ecological cause/effects expectations (e.g. Ginzburg et al., 2007).
The process of overlaying data itself contributes to error propagation if each data point has some error associated.	Explicitly mapping errors of input data and of the model's output (residuals) may provide information about the input error and the error propagated after overlaying procedures (e.g. Arbia et al., 1998).

1999; Kadmon et al., 2003; Prendergast et al., 1993). Data gathered in an unbiased fashion can potentially give a relatively accurate picture of species distributions. However, surveys have traditionally been influenced by the preconceived views and/or previous field experience of data collectors (Bacaro et al., 2009; Sastre and Lobo, 2009), although objective sampling schemes and cost-efficient survey protocols have long been available (Baffetta et al., 2007; Chiarucci et al., 2008; Funk et al., 2005; Hortal and Lobo, 2005; Rocchini et al., 2005). Field efforts by naturalists were largely biased to classical localities of particular natural beauty, localities of high species richness in the past, and to areas near the experts' residences and/or research centres (Dennis and Thomas,

2000; Lobo, 2008a, 2008b; see Table 1). Thus, while older collections made by amateur naturalists or taxonomists are concentrated in a few easily accessible areas surveyed in the past, more recent surveys conducted with a modern ecological focus provide a broader coverage of the environmental and geographical variability (Lobo et al., 2007). In consequence, the spatial coverage provided by the collection data stored in museums and herbaria is generally limited and spatially and environmentally biased (Hortal et al., 2007, 2008), particularly for hyperdiverse groups and regions (Ariño and Otegui, 2008).

The limited quality of the data has a profound impact on the accuracy of the species distribution maps. Consider, for example, a species for which most historical records are clustered in a

few National Parks and/or protected areas due to the bias in historical surveys. If these records are taken as an accurate representation of its distribution, the species will be thought to be restricted to a few scattered localities. More importantly, if these localities are located in an environmentally or spatially biased way (e.g. on mountain areas; see Hortal et al., 2008) the species will be thought to show a relatively restricted response to the environment. In the absence of a good assessment of sampling bias, it would not be possible to rule out that the species is absent from many areas simply because these areas have been less studied than those where the species has been recorded, instead of being a geographically rare or specialist species. With very few exceptions (see Griffiths et al., 1999; Prendergast et al., 1993), these problems are of paramount importance. Only a few atlases can be considered to provide a relatively adequate description of the distribution of species in a given moment of time. Good examples may include: the exhaustive *Flora of Berkshire* (Crawley, 2005), which provides a detailed distribution of all species at the small scale; the *Atlas of the Breeding Birds of the British Isles* (Gibbons et al., 1993), which includes comprehensive records in 10×10 km grid cells for such territory; and perhaps the IUCN's *Global Mammal Assessment* (Schipper et al., 2008; <http://www.iucnredlist.org/initiatives/mammals>), which describes the global distribution of all known mammal species at a coarser resolution, although in this case the lack of knowledge on many areas worldwide is acknowledged as a major source of error. In general, unbiased data for large numbers of species across wide geographical regions are the exception rather than the rule (Hortal et al., 2007), inhibiting the compilation of useful distribution maps for most species in need of conservation strategies (Cayuela et al., 2009).

2 Limitations of the measurement system

Field surveys often do not provide reliable information on all target species due to

differences in detectability, insufficient sampling effort, inadequacy of sampling techniques, geographic (i.e. positional) inaccuracy and collection bias (see Hortal and Lobo, 2005; Pellet and Schmidt, 2005; Soberón and Peterson, 2004; Yoccoz et al., 2001). Also, some data from museum collections or informal captures have an associated positional error because the only available information on the location is the name of the place, locality or area where the species was recorded (Fernández et al., 2009; Table 1). However, positional errors can affect even data recorded using a GPS (see Johnson and Gillingham, 2008). Hence, it is often necessary to use the information available in field records to correct for the position of the recorded presences of the species (Chefaoui et al., 2005; Garcillán et al., 2003).

Taxonomic identifications are also a potential source of imprecision, due either to misidentifications (Lozier et al., 2009; Soberón and Peterson, 2004; Stribling et al., 2003) or to lack of adequate taxonomic knowledge (including synonymies; see Baselga et al., 2010). For instance, Bacaro et al. (2009) demonstrated that subjectivity in acquiring species lists by field biologists increases error variance instead of enhancing the quality of the information on actual community diversity, resulting in huge differences in the species presences/absences obtained by different field biologists. Further problems may arise when combining data obtained from different sources, in particular when different taxonomic classifications are applied. Implementing concept taxonomies is mandatory to solve this issue (Berendsohn, 1995), as well as other related problems such as taxonomic inflation (Isaac et al., 2004; Knapp et al., 2005).

Misidentifications can place species in wrong places, leading to overestimations of their actual distributions especially in the case of hyper-diverse groups and/or poorly surveyed countries, for which appropriate

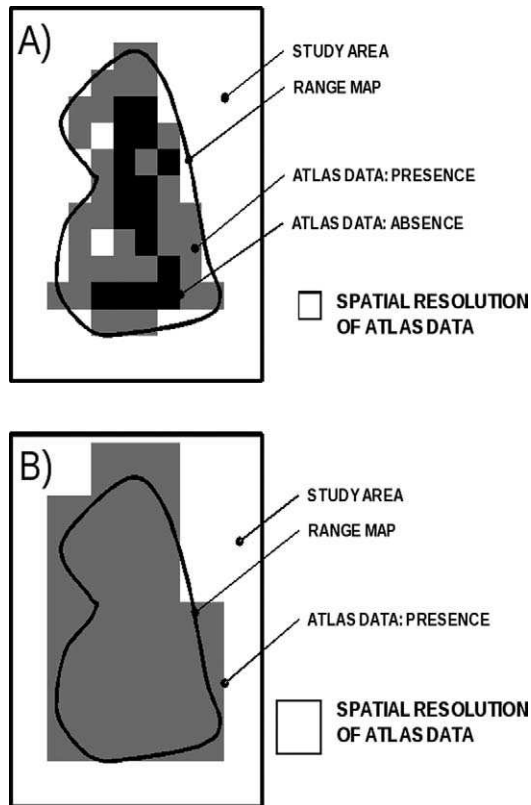


Figure 2. Example of discrepancies between Atlas data and range maps. In (A), the range map (expert drawn, bold line) overestimates the distribution of a species with respect to Atlas data where grey squares represent cells with the certain presence of the species and black squares represent well-sampled cells with a very high probability of absence. Depending on the resolution, an opposite pattern may also be observed based on the same data (B).

taxonomic revisions are lacking. Potential errors (i.e. isolated points that lie far from the main distributional areas) could be identified by screening the location of the records. However, taxonomic knowledge is also spatially structured (e.g. Baselga et al., 2007, 2010), so deciding the veracity of those outliers may be difficult when there is a substantial lack of information; discarding potential outliers using automatic procedures or rules or thumb could result in eliminating essential information.

III Uncertainty and biases in species distribution mapping and modelling

The straightforward approach to fill in the gaps of knowledge on the species distributions would be to formulate spatially explicit hypotheses, i.e. maps. The methods available to generate these hypotheses can be grouped into three broad categories (see Hortal, 2008): (1) *range maps* (i.e. expert drawn); (2) *Atlas data* (i.e. grid superimposition to the data collected in the field); and (3) *species distribution models* (i.e. interpolation techniques using records as input data). Although all three kinds of procedures are subject to uncertainty, their potential errors are often overlooked, particularly in the case of SDMs.

I Range maps

Range maps are created from information on the location of known presences and convex-hull techniques, from the opinion of experts, or by a mixture of both (Burgman and Fox, 2003). They are typically used to describe the realized distribution of species at coarse (i.e. regional, continental or even global) scales. However, they are known to have limited value at finer scales (Hawkins et al., 2008; Hortal, 2008; Hurlbert and Jetz, 2007; Lawes and Piper, 1998), overestimating the area of occupancy of individual species (see, for example, Figure 2 and Table 1 at Hurlbert and Jetz, 2007) and species richness values when overlapped (Hawkins et al., 2008). Fine-scale distributions are much patchier than any of the most detailed expert-drawn range maps can account for (see Graham and Hijmans, 2006; Hurlbert and Jetz, 2007), a human limit that has been already explored in other fields of cartography (Burnett and Blaschke, 2003). However, this kind of error can result in the misidentification of richness hotspots, and in serious overestimations of the actual distribution of the species if inappropriately fine scales are chosen.

2 Atlas data

Direct mapping from data collected in the field is usually done to create distributional atlases. Atlas data (sensu Hortal, 2008) are generally obtained by superimposing a grid to the known presence records. Some mapping schemes use grids defined a priori to standardize and plan survey efforts, which intend to collect all taxa within each cell (e.g. the *Atlas of the British Flora*; Preston et al., 2002). Presence in a given cell of the grid is counted when one or more records are found within it. This format allows most use of information gathered with very different survey schemes, although the lack of quality and completeness of the data collected in the field have an effect on atlas data. While most recorded occurrences are indisputable, absences are always uncertain. False absences (i.e. places where the species is present but has not been recorded; Gu and Swihart, 2004; Lobo et al., 2010) can arise from many different reasons. The species may be present but undetectable (e.g. species hidden in the seed bank), may suffer substantial population shifts from year to year (thus being temporarily absent when the surveys are carried out), may be missed simply because the habitat or site in question has not been accurately sampled (Gu and Swihart, 2004; Lobo et al., 2007; Stohlgren, 2007), or may have been misidentified (Berendsohn, 1995). Hence, atlas data will typically underestimate the area of occupancy of species, for it will contain a substantial (though unknown) number of false absences.

These problems for mapping Atlas data would be partially overcome if true absences could be identified. Unfortunately, most surveys record only the presence of species, not absence (the localities in which the species remain undetected after an adequate survey effort); this information is generally not available, or is only available from expert opinion, the latter being the best available to construct range maps (Hortal, 2008). Alternatively, reliable absences could

be obtained from comprehensive survey effort assessments (Hortal and Lobo, 2005; see below), by assuming that the species is absent from the well-surveyed areas where it has not been recorded. However, sampling intensity is often not planned in advance and is usually not quantified, so these areas cannot be identified easily either (Lobo, 2008b; see Table 1).

Errors in atlas data are also scale-dependent, as with range maps. Some collection data might not be suitable to use with fine-grained grids, which are also sensible to data with significant positional errors. Also, the number of false absences will diminish rapidly at coarser scales due to the diminution in the degree of patchiness (see La Sorte and Hawkins, 2007). Hence, maps constructed from atlas data cannot be assumed to be accurate either, at least at small scales (Hortal, 2008). By choosing a coarse enough grid size, the effect of these errors would be minimized, but with the cost of sacrificing the precision in the description of species distributions that could apparently be obtained from fine-grained presence records.

3 Species distribution models (SDMs)

SDMs, developed since the late 1980s, are a heterogeneous group of procedures which establish a relationship between distributional data and environmental variables. SDMs have become widely accepted tools to predict realized and/or potential species distributions, forecast global change impacts on biodiversity, or hindcast the past distributions of extant and extinct species (see, e.g. Araújo and Guisan, 2006; Guisan and Thuiller, 2005; Guisan and Zimmermann, 2000; Nogués-Bravo, 2009; Soberón and Peterson, 2004). Although SDMs are generally considered powerful tools, the data used to calibrate the models and the modelling process itself often introduce high levels of uncertainty in their results (Elith and Graham, 2009; Elith et al., 2002; Jiménez-Valverde et al., 2008b; Loiselle et al., 2003). A number of studies have

discussed and/or highlighted some of the problems of SDM (e.g. Araújo and Guisan, 2006; Heikkinen et al., 2006; Hortal and Lobo, 2006; Hortal et al., 2007, 2008; Jiménez-Valverde et al., 2008b; Kadmon et al., 2003, 2004; Lobo, 2008b; Lobo et al., 2007; Loiselle et al., 2003, 2008; Luoto et al., 2005; Palmer, 1995; Palmer et al., 2002; Pearson et al., 2006). Neither are the environmental variables used as predictors free from error, nor do all SDM techniques provide consistent results (Graham and Hijmans, 2006; Table 1). Nonetheless, distributional data are often not good enough to provide reliable descriptions of the environmental responses of most species either.

Data quality critically affects the quality of SDMs. Obviously, SDMs built from environmental and geographically biased distributional data will misrepresent the distribution of species as well as their environmental responses (see Hortal et al., 2007, 2008). If these biases are consistent across species (as they usually are), it is also quite likely that SDM errors will aggregate in space (Hortal and Lobo, 2006). Predictive accuracy depends also on the sample size available (e.g. Jiménez-Valverde et al., 2009; Wisz et al., 2008); for small sample sizes, no modelling approach can be robust, and this drawback seriously hinders their usefulness for the immense part of hyperdiverse regions and groups. Environmental variables used as predictors for SDMs are also a source of uncertainty. Those created from climatic models and/or geographical interpolations present a certain degree of error depending on the accuracy of these models (see, for example, Luoto et al., 2005; Thuiller et al., 2004; Figure 1). Due to this, SDMs developed using the same variables coming from different sources (e.g. different climate modelling procedures) provide very different results (Beaumont et al., 2007). In the same way, variables derived from remote sensing data may be prone to classification errors (Foody, 2005; Jiménez-Valverde et al., 2008a).

The quality of data and predictors are not the sole sources of error for SDMs. The best predictors can be discarded or not accounted for in either manual or automated model selection processes (see, for example, Diniz-Filho et al., 2008; Fattorini, 2005; Ginzburg et al., 2007; Heikkinen et al., 2006; Maggini et al., 2006). It is well known that using exactly the same data with different SDM techniques provides quite different results (e.g. Araújo and New, 2007; Araújo et al., 2005; Loiselle et al., 2003; Wisz et al., 2008). On the one hand, each time a different technique or data set is used, the frequent collinearity among variables or their spatial autocorrelation results in different output models and maps (Bini et al., 2009; Carl and Kühn, 2007; Dormann, 2007a; Graham, 2003; see Table 1). On the other, the same technique can be tuned up in numerous ways by selecting different options in the parameterization process. Additional uncertainty comes from the selection of the cut-off used to classify the continuous probability values that are the output of some SDMs into presences and absences. Selecting such cut-off is not straightforward; some general recommendations have been proposed for such selection (Jiménez-Valverde and Lobo, 2007), but the final choice still depends on the intended use. Given that the mapped results change considerably depending on the cut-off value, the uncertainty associated to this classification should be interpreted together with model predictions (Pearson et al., 2006).

Last, but not least, SDMs do not always provide predictions of the actual distribution of the species. Rather, they approximate either the *potential* or the *realized* distribution of the species depending on the modelling technique, predictors and data used (Jiménez-Valverde et al., 2008b). Realized species distributions are shaped not only by environmental conditions, but also by a number of other factors including biotic interactions (Heikkinen et al., 2007; Schweiger et al., 2008), biogeographical constraints and historical events (for detailed

discussion see Colwell and Rangel, 2009; Lobo et al., 2010; Jiménez-Valverde et al., 2008b; Soberón and Nakamura, 2009; Soberón, 2007, 2010), most of which are unknown and difficult to consider in the modelling procedure. Accounting for the contingent processes limiting species distributions needs not only predictors accounting for them, but also data on the absence from unoccupied but suitable places. In consequence, SDM results lie at an unknown point along the fuzzy gradient between the potential and realized distribution (Jiménez-Valverde et al., 2008b); that is, between determining where a species could live and where does it actually live, even in the absence of other sources of uncertainty.

IV Maps of ignorance and good practice in mapping species distributions

Mapped species distributions coming from range maps, atlas data or species distribution models are widely used, and the increasing availability of data through the Global Biodiversity Information Facility (GBIF, <http://www.gbif.org>; Guralnick et al., 2007), or other biodiversity databases such as PANGAEA (<http://www.pangaea.de>) or the USDA plant database (<http://plants.usda.gov>), will certainly increase their usage even more (Hortal et al., 2007; Soberón and Peterson, 2004). However, a code of good practices for the use of species distribution maps is, as yet, lacking.

From our review, it is clear that species distribution maps possess intrinsic uncertainties. Both input data and the process of mapping species distributions are, in themselves, important sources of uncertainty (Elith et al., 2002; Palmer et al., 2005). It follows that good practice in biogeography research should include at least a description of the uncertainty associated with species distribution maps. However, uncertainty is often ignored or not accounted for, and the potential errors in mapped species distributions are often overlooked. To date, it is commonly

thought that these problems are either unimportant or have not been studied (see, for example, the bold statements by Phillips et al., 2009: 181). As a consequence, the completeness and biases of the input data are rarely assessed and the uncertainty associated with the mapping process is rarely taken into account (see Clark, 2007). Apart for a few sporadic examples using robust (replicable) methods to infer map quality (e.g. Bierman et al., 2010, using Bayesian posterior probabilities), in most cases, researchers with good knowledge of the group studied decide how reliable maps are based on field experience (e.g. Highton, 1995: 580–581). While dealing with the geographical variation of pollination types (insect-, wind- and self-pollination), Kühn et al. (2006) showed that mapping spatial errors is crucial to identify the areas with higher uncertainty, and eventually determine the environmental factors leading to such uncertainty. Moreover, mapping bias may reveal spatial autocorrelation in the data. In fact, if such data were derived from different areas following different protocols they will tend to show clustering effects with higher similarity within regions and dissimilarity among them (Dormann et al., 2007; Mahecha and Schmidtlein, 2008).

Building on an old idea of Samuel Whittemore Boggs (1949), we thus propose the development of ‘maps of ignorance’, which will depict the areas where the reliability of mapped distributions is either known or unknown. In practice, we argue that the degree and spatial distribution of uncertainty should be assessed when creating species distribution maps, and not only their overall accuracy or model errors (in the case of SDMs). These maps of ignorance, which we herein call Distributional Uncertainty Maps, could be constructed by mapping the bias and incompleteness in the distributional data; that is, by studying the spatial distribution of the survey effort and hence how the completeness of the inventories varies in space (see, for example, Hortal et al., 2008; Loiselle et al., 2008). In the particular case of SDMs, Distributional

Uncertainty Maps may be constructed from several sources (i.e. layers), including: (1) the quality of the distributional data, as above; but also (2) maps of the bias and error in the predictions obtained with these models (Murphy et al., 2004; Palmer et al., 2005); (3) maps of the variation in the predictive power of each predictor across space obtained using, for example, Geographically Weighted Regressions (Fotheringham et al., 2002) or by estimating the geographical deviations of model parameters (Elith et al., 2002). Note, however, that, when SDM results are extrapolated (i.e. transferred in space and/or time), additional information is necessary to account for the uncertainty in the areas where models extrapolate beyond the environmental or geographical circumstances considered in the calibration process (see, for example, Fitzpatrick and Hargrove, 2009).

V Conclusions

Identifying the best way to construct Distributional Uncertainty Maps and which of the layers of uncertainty is to be used to build each particular map remain as matters for further discussion and research. However, we argue that these maps should be considered an essential component of good practice in species distribution mapping. The interpretation of species distribution maps should take into account a spatially explicit quantification of the uncertainty. Hence, maps of uncertainty in the distribution should be an integral part of the basic 'metadata' of any species distribution map, no matter what its origin or the mapping technique used. This would allow research to take into account the spatial uncertainty while interpreting species distribution maps, and also to identify areas that need further fieldwork to attain reliable knowledge on the distribution of a particular species or biological group. We therefore suggest an effort to develop transparent and easy-to-use methods for the construction of Distributional Uncertainty Maps; once these methods are available, these

maps should become common practice in biogeographical research.

It is important to note that we do not advocate discarding species distribution maps as a whole. Rather, we propose that uncertainty should be considered at each step of the production of such maps: from the collection of input data through data processing and management to data analysis, modelling and mapping. That all knowledge is subject to error, and therefore to doubt, is deeply embedded in the Cartesian thinking which provides the very basis for the current philosophical framework of natural sciences. In biogeography, as a geographical science, both knowledge and doubt should include a geographical (i.e. spatial) component. Only by knowing where we should and should not doubt our knowledge on the distribution of species will we be in the position of improving such knowledge. It is thus time that good practice in biogeography included maps of ignorance, which will help not only in the correct interpretation of biodiversity maps, but also in deciding where empirical validations based on independently collected data are needed. Further research is required to decide the most suitable way of describing uncertainty in distributional maps, but we believe that by mapping the different aspects discussed here, Distributional Uncertainty Maps would allow improved estimates of where our knowledge is reliable, and where it is not.

Acknowledgements

We wish to thank the Editor Nicholas Clifford, Ingolf Kühn and two anonymous referees for their suggestions. DR is partially funded by the Autonomous Province of Trento (Italy), ACE-SAP project (No. 23, June 12, 2008, of the University and Scientific Research Service). JH was funded by the UK Natural Environment Research Council and by a CSIC JAE-Doc grant. SL was supported by two Hungarian Scientific Research Fund/Norway Financing Mechanism grants (OTKA NNF 78887, 85562) and a Bolyai Research Fellowship from the Hungarian Academy of Sciences during manuscript preparation. AJ-V was supported by a MEC (Ministerio de Educación y

Ciencia, Spain) postdoctoral fellowship (Ref.: EX-2007-0381). AC and GB thank the Department of Environmental Science of the University of Siena for supporting the Research Group “BIOCONNET, Biodiversity and Conservation Network”.

References

- Araújo MB and Guisan A (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography* 33: 1677–1688.
- Araújo MB and New M (2007) Ensemble forecasting of species distributions. *Trends in Ecology and Evolution* 22: 42–47.
- Araújo MB, Whittaker RJ, Ladle RJ, and Erhard M (2005) Reducing uncertainty in projections of extinction risk from climate change. *Global Ecology and Biogeography* 14: 529–538.
- Arbia G, Griffith DA, and Haining RP (1998) Error propagation modelling in raster GIS: Overlay operations. *International Journal of Geographical Information Science* 12: 145–167.
- Ariño A and Otegui J (2008) Sampling biodiversity sampling. *Proceedings of TDWG*. Available at: <http://www.tdwg.org/proceedings/article/view/413>.
- Bacaro G, Baragatti E, and Chiarucci A (2009) Using taxonomic data to assess and monitor biodiversity: Are the tribes still fighting? *Journal of Environmental Monitoring* 11: 798–801.
- Baffetta F, Bacaro G, Fattorini L, Rocchini D, and Chiarucci A (2007) Using nested plot design to obtain unbiased estimates of species richness. *Community Ecology* 8: 119–127.
- Baselga A, Hortal J, Jiménez-Valverde A, Gómez JF, and Lobo JM (2007) Which leaf beetles have not been yet described? Determinants of the description of Western Palearctic *Aphthona* species (Coleoptera: Chrysomelidae). *Biodiversity and Conservation* 16: 1409–1421.
- Baselga A, Lobo JM, Hortal J, Jiménez-Valverde A, and Gómez JF (2010) Assessing alpha and beta taxonomy in Eupelmid wasps: Determinants of the probability of describing good species and synonyms. *Journal of Zoological Systematics and Evolutionary Research* 48: 40–49.
- Beaumont LJ, Pitman AJ, Poulsen M, and Hughes L (2007) Where will species go? Incorporating new advances in climate modelling into projections of species distributions. *Global Change Biology* 13: 1368–1385.
- Berendsohn WG (1995) The concept of ‘potential taxa’ in databases. *Taxon* 44: 207–212.
- Bierman SM, Butler A, Marion G, and Kühn I (2010) Bayesian image restoration models for combining expert knowledge on recording activity with species distribution data. *Ecography* 33: 451–460.
- Bini LM, Diniz-Filho JAF, Rangel TFLVB, Akre TSB, Albaladejo RG, Albuquerque FS, et al. (2009) Coefficient shifts in geographical ecology: An empirical evaluation of spatial and non-spatial regression. *Ecography* 32: 193–204.
- Boggs SW (1949) An atlas of ignorance: A needed stimulus to honest thinking and hard work. *Proceedings of the American Philosophical Society* 93: 253–258.
- Burgman MA and Fox JC (2003) Bias in species range estimates from minimum convex polygons: Implications for conservation and options for improved planning. *Animal Conservation* 6: 19–28.
- Burnett C and Blaschke T (2003) A multi-scale segmentation/object relationship modelling methodology for landscape analysis. *Ecological Modelling* 168: 233–249.
- Carl G and Kühn I (2007) Analyzing spatial autocorrelation in species distributions using Gaussian and logit models. *Ecological Modelling* 207: 159–170.
- Cayuela L, Golicher DJ, Newton A, Kolb M, de Albuquerque FS, Arets, EJMM, et al. (2009) Species distribution modelling in the tropics: Problems, potentialities, and the role of biological data for species conservation. *Tropical Conservation Science* 2: 319–352.
- Chefaoui RM, Hortal J, and Lobo JM (2005) Potential distribution modelling, niche characterization and conservation status assessment using GIS tools: A case study of Iberian *Copris* species. *Biological Conservation* 122: 327–338.
- Chiarucci A, Bacaro G, and Rocchini D (2008) Quantifying plant species diversity in a Natura 2000 network: Old ideas and new proposals. *Biological Conservation* 141: 2608–2615.
- Chust G, Pretus JL, Ducrot D, Bedòs A, and Deharveng L (2003) Identification of landscape units from an insect perspective. *Ecography* 26: 257–268.
- Clark JS (2007) *Models for Ecological Data: An Introduction*. Princeton, NJ: Princeton University Press.
- Crawley MJ (2005) *The Flora of Berkshire*. Harpenden: Brambley Books.
- Colwell RK and Rangel TF (2009) Hutchinson’s duality: The once and future niche. *Proceedings of the National Academy of Sciences USA* 106: 19651–19658.

- Cox CB and Moore PD (2004) *Biogeography: An Ecological and Evolutionary Approach*. Oxford: Blackwell.
- Dennis RLH and Thomas CD (2000) Bias in butterfly distribution maps: The influence of hot spots and recorder's home range. *Journal of Insect Conservation* 4: 73–77.
- Dennis RLH, Sparks TH, and Hardy PB (1999) Bias in butterfly distribution maps: The effects of sampling effort. *Journal of Insect Conservation* 3: 33–42.
- Diniz-Filho JAF, Rangel TFLVB, and Bini LM (2008) Model selection and information theory in geographical ecology. *Global Ecology and Biogeography* 17: 479–488.
- Dormann CF (2007a) Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography* 16: 129–138.
- Dormann CF (2007b) Promising the future? Global change projections of species distributions. *Basic and Applied Ecology* 8: 387–397.
- Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, Carl G, et al. (2007) Methods to account for spatial autocorrelation in the analysis of atlas data: A review. *Ecography* 30: 609–628.
- Elith J and Graham CH (2009) Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography* 32: 66–77.
- Elith J, Burgman MA, and Regan HM (2002) Mapping epistemic uncertainties and vague concepts in predictions of species distribution. *Ecological Modelling* 157: 313–329.
- Elith J, Graham H, Anderson P, Dudik M, Ferrier S, Guisan A, et al. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129–151.
- Elzinga LC, Salzer DW, Willoughby JW, and Gibbs JP (2001) *Monitoring plant and animal populations*. Malden, MA: Blackwell Science.
- Fattorini S (2005) A simple method to fit geometric series and broken stick models in community ecology and island biogeography. *Acta Oecologica* 28: 199–205.
- Fernández MA, Blum SD, Reichle S, Guo Q, Holzman B, and Hamilton H (2009) Locality uncertainty and the differential performance of four common niche-based modeling techniques. *Biodiversity Informatics* 6: 36–52.
- Fitzpatrick MC and Hargrove WW (2009) The projection of species distribution models and the problem of non-analog climate. *Biodiversity and Conservation* 18: 2255–2261.
- Foody GM (2005) Local characterization of thematic classification accuracy through spatially constrained confusion matrices. *International Journal of Remote Sensing* 26: 1217–1228.
- Fotheringham AS, Brunson C, and Charlton M (2002) *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester: Wiley.
- Funk VA, Richardson KS, and Ferrier S (2005) Survey-gap analysis in expeditionary research: Where do we go from here? *Biological Journal of the Linnean Society* 85: 549–567.
- Garcillán PP, Ezcurra E, and Riemann H (2003) Distribution and species richness of woody dryland legumes in Baja California, Mexico. *Journal of Vegetation Science* 14: 475–486.
- Gibbons DW, Reid JB, and Chapman RA (1993) *The New Atlas of Breeding Birds in Britain and Ireland: 1988–91*. London: T and AD Poyser.
- Gillespie TW, Foody GM, Rocchini D, Giorgi AP, and Saatchi S (2008) Measuring and modeling biodiversity from space. *Progress in Physical Geography* 32: 203–221.
- Ginzburg LR, Jensen CXJ, and Yule JV (2007) Aiming the 'unreasonable effectiveness of mathematics' at ecological theory. *Ecological Modelling* 207: 356–362.
- Graham CH and Hijmans RJ (2006) A comparison of methods for mapping species ranges and species richness. *Global Ecology and Biogeography* 15: 578–587.
- Graham MH (2003) Confronting multicollinearity in ecological multiple regression. *Ecology* 84: 2809–2815.
- Griffiths GH, Eversham BC, and Roy DB (1999) Integrating species and habitat data for nature conservation in Great Britain: Data sources and methods. *Global Ecology and Biogeography* 8: 329–345.
- Gu W and Swihart RK (2004) Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biological Conservation* 116: 195–203.
- Guisan A and Thuiller W (2005) Predicting species distribution: Offering more than simple habitat models. *Ecology Letters* 8: 993–1009.
- Guisan A and Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147–186.

- Guralnick RP, Hill AW, and Lane M (2007) Towards a collaborative, global infrastructure for biodiversity assessment. *Ecology Letters* 10: 663–672.
- Hawkins B, Rueda M, and Rodriguez MA (2008) What do range maps and surveys tell us about diversity patterns? *Folia Geobotanica* 43: 345–355.
- Heikkinen RK, Luoto M, Araújo MB, Virkkala R, Thuiller W, and Sykes MT (2006) Methods and uncertainties in bioclimatic envelope modelling under climate change. *Progress in Physical Geography* 30: 751–777.
- Heikkinen RK, Luoto M, Virkkala R, Pearson RG, and Körber J-H (2007) Biotic interactions improve prediction of boreal bird distributions at macro-scales. *Global Ecology and Biogeography* 16: 754–763.
- Heisenberg WK (1958) *Physics and Philosophy: The Revolution in Modern Science*. New York: Harper and Row.
- Highton R (1995) Speciation in eastern North-American salamanders of the genus *Plethodon*. *Annual Review of Ecology and Systematics* 26: 579–600.
- Holt RD, Keitt TH, Lewis MA, Maurer BA, and Taper ML (2005) Theoretical models of species' borders: Single species approaches. *Oikos* 108: 18–27.
- Hortal J (2008) Uncertainty and the measurement of terrestrial biodiversity gradients. *Journal of Biogeography* 35: 1335–1336.
- Hortal J and Lobo JM (2005) An ED-based protocol for the optimal sampling of biodiversity. *Biodiversity and Conservation* 14: 2913–2947.
- Hortal J and Lobo JM (2006) Towards a synecological framework for systematic conservation planning. *Biodiversity Informatics* 3: 16–45.
- Hortal J, Jiménez-Valverde A, Gómez JF, Lobo JM, and Baselga A (2008) Historical bias in biodiversity inventories affects the observed realized niche of the species. *Oikos* 117: 847–858.
- Hortal J, Lobo JM and Jiménez-Valverde A (2007) Limitations of biodiversity databases: Case study on seed-plant diversity in Tenerife (Canary Islands). *Conservation Biology* 21: 853–863.
- Hurlbert AH and Jetz W (2007) Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. *Proceedings of the National Academy of Science USA* 104: 13384–13389.
- Isaac NJB, Mallet J, and Mace GM (2004) Taxonomic inflation: Its influence on macroecology and conservation. *Trends in Ecology and Evolution* 19: 464–469.
- Jiménez-Valverde A and Lobo JM (2007) Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologica* 31: 361–369.
- Jiménez-Valverde A, Gómez JF, Lobo JM, Baselga A, and Hortal J (2008a) Challenging species distribution models: The case of *Maculinea nausithous* in the Iberian Peninsula. *Annales Zoologici Fennici* 45: 200–210.
- Jiménez-Valverde A, Lobo JM, and Hortal J (2008b) Not as good as they seem: The importance of concepts in species distribution modelling. *Diversity and Distributions* 14: 885–890.
- Jiménez-Valverde A, Lobo JM, and Hortal J (2009) The effect of prevalence and its interaction with sample size on the reliability of species distribution models. *Community Ecology* 10: 196–205.
- Johnson CJ and Gillingham MP (2008) Sensitivity of species-distribution models to error, bias, and model design: An application to resource selection functions for woodland caribou. *Ecological Modelling* 213: 143–155.
- Kadmon R, Farber O, and Danin A (2003) A systematic analysis of factors affecting the performance of climatic envelope models. *Ecological Applications* 13: 853–867.
- Kadmon R, Farber O, and Danin A (2004) Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications* 14: 401–413.
- Klir GJ and Wierman MJ (1999) *Uncertainty-Based Information: Elements of Generalized Information Theory*. New York: Physica Verlag.
- Knapp S, Lughadha EN, and Paton A (2005) Taxonomic inflation, species concepts and global species lists. *Trends in Ecology and Evolution* 20: 7–8.
- Kühn I, Bierman SM, Durka W, and Klotz S (2006) Relating geographical variation in pollination types to environmental and spatial factors using novel statistical methods. *New Phytologist* 172: 127–139.
- La Sorte FA and Hawkins BA (2007) Range maps and species richness patterns: Errors of commission and estimates of uncertainty. *Ecography* 30: 649–662.
- Lawes MJ and Piper SE (1998) There is less to binary maps than meets the eye: The use of species distribution data in the southern African sub-region. *South African Journal of Science* 94: 207–210.
- Lobo JM (2008a) Database records as a surrogate for sampling effort provide higher species richness estimations. *Biodiversity and Conservation* 17: 873–881.

- Lobo JM (2008b) More complex distribution models or more representative data? *Biodiversity Informatics* 5: 14–19.
- Lobo JM, Baselga A, Hortal J, Jiménez-Valverde A, and Gómez JF (2007) How does the knowledge about the spatial distribution of Iberian dung beetle species accumulate over time? *Diversity and Distributions* 13: 772–780.
- Lobo JM, Jiménez-Valverde A, and Hortal J (2010) The uncertain nature of absences and their importance in species distribution modelling. *Ecography* 33: 103–114.
- Loiselle BA, Howell CA, Graham CH, Goerck CH, Brooks T, Smith KG, et al. (2003) Avoiding pitfalls of using species distribution models in conservation planning. *Conservation Biology* 17: 1591–1600.
- Loiselle BA, Jorgensen PM, Consiglio T, Jiménez I, Blake JG, Lohmann LG, et al. (2008) Predicting species distributions from herbarium collections: Does climate bias in collection sampling influence model outcomes? *Journal of Biogeography* 35: 105–116.
- Lomolino MV (2004) Conservation biogeography. In: Lomolino MV and Heaney LR (eds) *Frontiers of Biogeography: New Directions in the Geography of Nature*. Sunderland, MA: Sinauer Associates, Inc, 293–296.
- Lozier JD, Aniello P, and Hickerson MJ (2009) Predicting the distribution of Sasquatch in western North America: Anything goes with ecological niche modelling. *Journal of Biogeography* 36: 1623–1627.
- Luoto M, Poyry J, Heikkinen RK, and Saarinen K (2005) Uncertainty of bioclimate envelope models based on the geographical distribution of species. *Global Ecology and Biogeography* 14: 575–584.
- Maggini R, Lehmann A, Zimmermann NE, and Guisan A (2006) Improving generalized regression analysis for the spatial prediction of forest communities. *Journal of Biogeography* 33: 1729–1749.
- Mahecha MD and Schmidtlein S (2008) Revealing biogeographical patterns by nonlinear ordinations and derived anisotropic spatial filters. *Global Ecology and Biogeography* 17: 284–296.
- Murphy JM, Sexton DMH, Barnett DN, Jones GS, Webb MJ, Collins M, et al. (2004) Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature* 430: 768–772.
- Nagendra H (2001) Defining local conservation priorities: Assessing the impact of landscape transformation. *Biodiversity and Conservation* 10(3): 353–365.
- Nogués-Bravo D (2009) Predicting the past distribution of species climatic niches. *Global Ecology and Biogeography* 18: 521–531.
- Palmer MW (1995) How should one count species? *Natural Areas Journal* 15: 124–135.
- Palmer MW, Earls P, Hoagland BW, White PS, and Wohlgemuth T (2002) Quantitative tools for perfecting species lists. *Environmetrics* 13: 121–137.
- Palmer TN, Shutts GJ, Hagedorn R, Doblus-Reyes FJ, Jung T, and Leutbecher M (2005) Representing model uncertainty in weather and climate prediction. *Annual Review of Earth and Planetary Sciences* 33: 163–193.
- Pearson RG, Thuiller W, Araújo MB, Martinez-Meyer E, Brotons L, McClean C, et al. (2006) Model-based uncertainty in species range prediction. *Journal of Biogeography* 33: 1704–1711.
- Pellet J and Schmidt BR (2005) Monitoring distributions using call surveys: Estimating site occupancy, detection probabilities and inferring absence. *Biological Conservation* 123: 27–35.
- Phillips SJ, Dudik M, Elith J, Graham CH, Lehmann A, Leathwick J, et al. (2009) Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications* 19: 181–197.
- Prendergast JR, Quinn RM, Lawton JH, Eversham BC, and Gibbons DW (1993) Rare species, the coincidence of diversity hotspots and conservation strategies. *Nature* 365: 335–337.
- Preston CD, Pearman DA, and Dines TD (2002) *New Atlas of the British and Irish Flora: An Atlas of the Vascular Plants of Britain, Ireland, the Isle of Man and the Channel Islands*. Oxford: Oxford University Press.
- Ricotta C and Anand M (2006) Spatial complexity of ecological communities: Bridging the gap between probabilistic and non-probabilistic uncertainty measures. *Ecological Modelling* 197: 59–66.
- Rocchini D and Ricotta C (2007) Are landscapes as crisp as we may think? *Ecological Modelling* 204: 535–539.
- Rocchini D, Andreini Butini S, and Chiarucci A (2005) Maximizing plant species inventory efficiency by means of remotely sensed spectral distances. *Global Ecology and Biogeography* 14: 431–437.
- Rushton SP, Ormerod SJ, and Kerby G (2004) New paradigms for modelling species' distributions? *Journal of Applied Ecology* 41: 193–200.
- Sastre P and Lobo JM (2009) Taxonomist survey biases and the unveiling of biodiversity patterns. *Biological Conservation* 142: 462–467.

- Schipper J, Chanson JS, Chiozza F, Cox NA, Hoffmann M, Katariya V, et al. (2008) The status of the world's land and marine mammals: Diversity, threat, and knowledge. *Science* 322: 225–230.
- Schweiger O, Settele J, Kudrna O, Klotz S, and Kühn I (2008) Climate change can cause spatial mismatch of trophically interacting species. *Ecology* 89: 3472–3479.
- Soberón J (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters* 10: 1115–1123.
- Soberón J (2010) Niche and area of distribution modeling: A population ecology perspective. *Ecography* 33: 159–167.
- Soberón J and Nakamura M (2009) Niches and distributional areas: Concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences USA* 106: 19644–19650.
- Soberón J and Peterson AT (2004) Biodiversity informatics: Managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society of London B* 359: 689–698.
- Stohlgren TJ (2007) *Measuring Plant Diversity: Lessons From the Field*. New York: Oxford University Press.
- Stott P (1981) *Historical Plant Geography*. London: Allen and Unwin.
- Stribling JB, Moulton SR II, and Lester GT (2003) Determining the quality of taxonomic data. *Journal of the North American Benthological Society* 22: 621–631.
- Thuiller W, Araújo MB, and Lavorel S (2004) Do we need land-cover data to model species distributions in Europe? *Journal of Biogeography* 31: 353–361.
- Walther BA and Moore JL (2005) The definitions of bias, precision, and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography* 28: 815–829.
- Whittaker RJ, Araújo MB, Jepson P, Ladle RJ, Watson JEM, and Willis KJ (2005) Conservation biogeography: Assessment and prospect. *Diversity and Distributions* 11: 3–23.
- Whittaker RJ, Willis KJ, and Field R (2001) Scale and species richness: Towards a general, hierarchical theory of species richness. *Journal of Biogeography* 28: 453–470.
- Wisz MS, Hijmans RJ, Li J, Peterson AT, Graham CH and Guisan A (2008) Effects of sample size on the performance of species distribution models. *Diversity and Distributions* 14: 763–773.
- Yoccoz NG, Nichols JD, and Boulinier T (2001) Monitoring of biological diversity in space and time. *Trends in Ecology and Evolution* 16: 446–453.