REVIEW AND SYNTHESIS

Tradeoffs of different types of species occurrence data for use in systematic conservation planning

Carlo Rondinini,[1]* Kerrie A. Wilson,[2] Luigi Boitani,[1] Hedley Grantham[2] and Hugh P. Possingham[2]

[1]*Department of Animal and Human Biology, Università di Roma La Sapienza, Viale dell'Università 32, 00185 Roma, Italy*
[2]*The Ecology Centre, The University of Queensland, Brisbane, 4072, Queensland, Australia*
Correspondence: E-mail: carlo.rondinini@uniroma1.it

Abstract

Data on the occurrence of species are widely used to inform the design of reserve networks. These data contain commission errors (when a species is mistakenly thought to be present) and omission errors (when a species is mistakenly thought to be absent), and the rates of the two types of error are inversely related. Point locality data can minimize commission errors, but those obtained from museum collections are generally sparse, suffer from substantial spatial bias and contain large omission errors. Geographic ranges generate large commission errors because they assume homogenous species distributions. Predicted distribution data make explicit inferences on species occurrence and their commission and omission errors depend on model structure, on the omission of variables that determine species distribution and on data resolution. Omission errors lead to identifying networks of areas for conservation action that are smaller than required and centred on known species occurrences, thus affecting the comprehensiveness, representativeness and efficiency of selected areas. Commission errors lead to selecting areas not relevant to conservation, thus affecting the representativeness and adequacy of reserve networks. Conservation plans should include an estimation of commission and omission errors in underlying species data and explicitly use this information to influence conservation planning outcomes.

Keywords

Commission error, geographic range, omission error, point data, predicted distribution data, reserve selection.

Ecology Letters (2006) **9**: 1136–1145

INTRODUCTION

Systematic conservation planning refers to the selection of areas for conservation action (including reservation, restoration and management) in order to achieve clearly stated conservation goals (Margules & Pressey 2000). Developments over the last 20 years have seen a shift in focus from the selection of a minimum set of areas to reserve that conserve a single representation of all species of interest (Kirkpatrick 1983) to ensuring that biodiversity persists in the selected areas in the long-term (Rodrigues *et al.* 2000a,b; Bonn *et al.* 2002; Araujo *et al.* 2004; Cabeza *et al.* 2004). Conservation planning decisions are generally based largely on biological and physical data. Often the data employed is that which is readily available and is typically incomplete. This can affect the accuracy of conservation planning outcomes and introduce uncertainty and subjectivity into the planning process (Larsen & Rahbek 2003).

A subjective decision made in every conservation planning process is the choice of surrogate for biodiversity. Recently, this issue has been the subject of intense debate, with candidates including species occurrences (Brooks *et al.* 2004a,b), spatial patterns in species richness and turnover (Ferrier *et al.* 2004), environmental diversity (Faith *et al.* 2004), and vegetation types or land classes (Lombard *et al.* 2003; Pressey 2004). The optimal surrogate, however, is likely to depend on many factors, including the broad goal and the target to be achieved by the areas selected for reservation (Margules & Pressey 2000), the scale of analysis (Ferrier 2002), and constraints associated with data availability.

Among the available types of biodiversity data, species occurrence data are widely used in conservation planning at

regional (e.g. Howard *et al.* 1998; Araujo 1999; Bonn *et al.* 2002), continental (Eken *et al.* 2004; Rondinini *et al.* 2005), and global scales (Rodrigues *et al.* 2004), often in conjunction with other biodiversity data (Pressey 2004). Three types of species occurrence data exist: point localities, the locations where a species has been recorded; geographic ranges, the (broad) geographic boundaries of the area where a species is known to occur (Gaston 1991); and predicted distribution data, representing areas where a species is likely to be present as inferred from the suitability of environmental conditions (Corsi *et al.* 1999; Guisan & Zimmermann 2000).

The aims of this paper are (1) to discuss the strengths and weaknesses of each type of species occurrence data (point, range and predicted) in their application to systematic conservation planning; and (2) to assess the effect of choice of species occurrence data on the conservation planning outcomes. In particular, we focus on four criteria, specifically comprehensiveness, the ability of the selected areas to capture inter-specific variation; adequacy, the ability of the selected areas to ensure species persistence; representativeness, the ability of the areas to capture intra-specific variation; and efficiency, the economic impact of the selected areas (Possingham *et al.* 2006).

EXTENT OF OCCURRENCE, AREA OF OCCUPANCY AND THEIR REAL-WORLD REPRESENTATION

Species distributions can be synthesized by two concepts, namely the extent of occurrence (EOO) and the area of occupancy (AOO) (Gaston 1991). The EOO identifies the region encompassing all localities where a species has been recorded (Gaston 2003); the AOO is a subset of the EOO, which excludes all areas within the EOO that are not occupied by the species, because they are unsuitable or presently not occupied (Gaston 1993). Although a good knowledge of the AOO of species is key information for conservation planners, much of the area actually occupied by species may represent sink habitat that is unable to sustain a population without the contribution of immigrating individuals (Tyre *et al.* 2001).

The representation of a species' distribution can be wrong in two ways: it can erroneously indicate that a species is present or absent, which are respectively referred to as errors of commission and omission (Fielding & Bell 1997; Morrison *et al.* 1998). The EOO is prone to commission errors because of an overestimation of the area occupied by a species. By definition the AOO should be free from both commission and omission errors. In practice AOOs are obtained by excluding the portions of the EOO that are perceived to be unsuitable for the species, therefore commission errors are reduced but this at the expense of

a potential increase in omission errors (Fielding & Bell 1997).
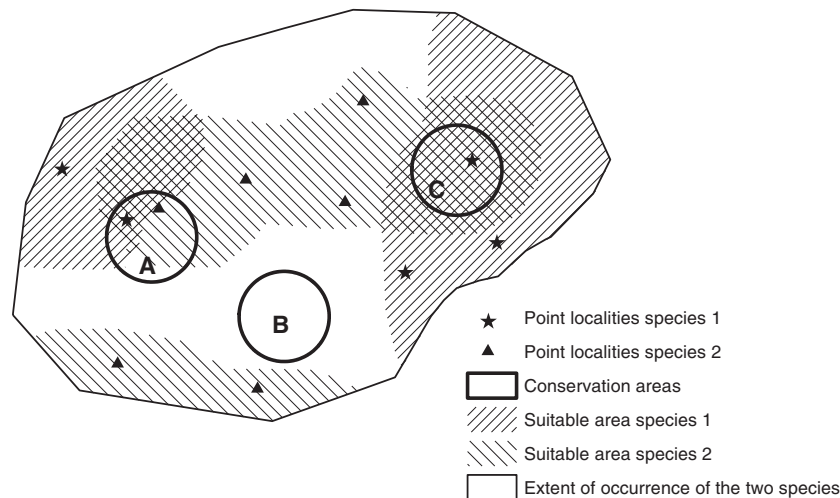
The two concepts of EOO and AOO are clearly distinct and their usefulness depends on the scale of analysis. However, any realization of a map of species distribution is positioned somewhere in a continuum between the EOO and the AOO; this is because at increasingly small scales more and more holes appear in the species distribution, which are overlooked at broad scales. The three types of species distribution data have different capabilities to represent the EOO and the AOO of a species.

Point localities are closer to the scale of the AOO because they originate from actual observations or specimen collections. Since point locality data are often sparse and discontinuous, they generally underestimate the actual AOO. In order to approximate the EOO of a species, point locality data are usually degraded to grid cells of various size (e.g. Gaston 1993; Larsen & Rahbek 2003; Araujo 2004).

Geographic ranges should coincide with the EOO if the exact limits of species distributions are known (Gaston 2003). However, ranges are typically delineated using the extreme localities where the species is known to occur and also (broad) ecological preferences: for example Rodrigues *et al.* (2004) limited the ranges of forest species to forest boundaries from land use maps. Therefore, geographic ranges tend to be positioned between a very broad AOO and the actual EOO.

The position of predicted distribution data in the continuum between the EOO and the AOO depends largely on the resolution of the data used for modelling. Broad-scale models based on expert knowledge of species–habitat relationships, and on relatively low-resolution maps of environmental variables (e.g. Boitani *et al.* 1999; Rondinini *et al.* 2005), can refine the EOO by removing from it patches of clearly unsuitable habitat. In comparison, local-scale and ecologically realistic models of well known species, where the species–habitat relationships are derived from point locality data and high-resolution environmental data, can reach a level of detail that allows the representation of the AOO (Scott *et al.* 2002).

The differences between the three types of species distribution data result in different rates of commission and omission errors, which in turn influence conservation planning outcomes (Fig. 1). The choice between using an estimator of the EOO or of the AOO of a species in conservation planning, ultimately depends on the scale and resolution of analysis (Reid 1998; Erasmus *et al.* 1999; Larsen & Rahbek 2003), the relative acceptability of omission and commission errors (Wilson *et al.* 2005), and constraints associated with data availability.

- ★ Point localities species 1
- ▲ Point localities species 2
- ☐ Conservation areas
- ▨ Suitable area species 1
- ▧ Suitable area species 2
- ☐ Extent of occurrence of the two species

Figure 1 Observed occurrences, area of occupancy (AOO) predicted by suitable area, and coincident geographic range of two hypothetical species (species 1 and species 2). For a conservation plan based on species' range each of the three areas, A, B and C, could be selected as these all occur with the extent of occurrence (EOO) of both species. Based on the species' predicted AOO, area C could be selected because it encompasses the area occupied by both species, area A is partially suitable for both species, and area B is outside of the predicted AOO of both species. Based on species' observed occurrences, area A could be selected because it contributes to the conservation of both species, area C contributes to the conservation of species 1, and area B gives no contribution.

POINT LOCALITY DATA

A point datum refers to a locational record where a species has been recorded (e.g. visually, audibly or through other evidence such as tracks and scats). Each locational record is assigned a geographic coordinate. Point locality data fall into two broad categories: those derived from museum collections (e.g. Graham *et al.* 2004) and those collected with a pre-determined sampling design, usually to compile atlases of species distribution (e.g. Harrison *et al.* 1997).

Strengths

The main strengths of point locality data relate to its availability, ease of use, and low rate of commission errors. Point locality data are increasingly available through on-line databases. These on-line databases typically contain museum data and it has been estimated that *c.* 5–10% of the 2.5 billion specimens held in natural history museums have been recorded in electronic format (Graham *et al.* 2004). Point locality data from on-line sources requires less user-manipulation than that required for generating geographic ranges and predicted distribution data. For example, ranges often need to be digitized from existing publications or from drawings made by species experts and predicted distribution data are generally obtained by combining point locality data with environmental data thought to determine the distribution of a species. The use of point locality data to obtain an understanding of the distribution of a species, therefore, offers not only time-saving advantages, but also a

potential reduction in errors associated with data manipulation (Burrough & McDonnell 1998). In addition, accurate and recent point locality data have relatively few commission errors. This low commission error rate minimises the uncertainty surrounding the adequacy of areas selected for reservation as it provides increased confidence that these areas will contain the species of interest.

Weaknesses

Point locality data can however contain commission errors if there is inaccurate recording of geographic co-ordinates or taxonomic errors, for example, due to species misidentification or changes in taxonomic nomenclature (Dunn *et al.* 1990; Williams *et al.* 2002). Rate of misclassification is high for some taxa: > 80% of the museum specimens of the genus *Euscelidia* (Insecta) analysed by Meier & Dikow (2004) had been previously assigned to the wrong species because of misidentification or synonymy. Additionally, some records of species presence may pertain only to the contemporaneous conditions; such as representing a unique event like the dispersal of a species into an area due to drought, or can be affected by long-term temporal species turnover (Lawes & Piper 1998; Rodrigues *et al.* 2000a).

No region has totally complete point locality data (Pressey 2004), because detailed surveys over the entire possible range of most species are almost never available due to lack of resources (Haila & Margules 1996). Even when extensive surveying is applied some species that are present may remain undetected resulting in errors of

omission (MacKenzie *et al.* 2002; Wintle *et al.* 2004). Additionally, there will likely be uncertainty surrounding the level of omission errors, since sites where a species is actually absent are rarely recorded. Even in such circumstances there will likely be uncertainty whether a species is actually absent. For example, Wintle *et al.* (2005) found that *c.* 18 independent visits were required to be 90% certain that a forest owl was absent from a site in south-eastern Australia. For these reasons and because point locality data do not include any informed interpolation or extrapolation of species occupancy, this type of species occurrence data contain more omission errors compared with the other types of species occurrence data. Several studies have investigated the sensitivity of conservation planning to incomplete data and found decreased efficiency (Freitag & Van Jaarsveld 1998), reduced flexibility in terms of the possible areas available to be selected for conservation action (Gladstone & Davis 2003), and selected areas missing rare species (Gaston & Rodrigues 2003).

Point locality data collected opportunistically, such as museum or herbarium records (Ponder *et al.* 2001; Funk & Richardson 2002; Williams *et al.* 2002), result in records being spatially biased towards places that are easily accessed, taxonomically biased towards species that are relatively conspicuous, and temporally biased, due to irregular recordings over time (Nelson *et al.* 1990; Lawes & Piper 1998; Keller & Scallan 1999; Polasky *et al.* 2000; Funk & Richardson 2002; Lombard *et al.* 2003). The spatial bias can reflect a bias towards easily accessible areas (Davis *et al.* 1990) or between different habitat types (Austin & Heyligers 1989) and can be different across taxonomic groups in the same area. For example, Freitag *et al.* (1998) found sampling in southern Africa to be biased towards protected areas for large mammals, but less biased for small mammals. Furthermore, species are often under-recorded in the core of their range, while outlying occurrences are recorded and as a consequence, further biases is introduced into point locality data. Biases inherent in point locality data can misdirect conservation planning effort to areas that are better surveyed and towards taxa that are relatively well known. For example, Reddy & Davalos (2003) revealed that priority areas for conservation action in Africa were geographically biased towards well-surveyed areas of vertebrates, and Nelson *et al.* (1990) found that botanical sampling intensity in the Amazon was directly correlated with areas designated as centres of plant endemism.

Point data derived from museum collections have generally been accumulated across large time spans and this should be accounted for in their analysis. Often, however, it is assumed that museum data are static, and that it represents a snapshot of species distribution at a given (usually the present) time. This ignores the dynamic nature of species distributions due to dispersal or shifts in distribution due to changed land use or environmental conditions. Such changes might result in the misallocation of conservation priorities, particularly in areas undergoing rapid environmental changes or that are highly threatened (Wessels *et al.* 2000).

It is impossible to assess commission and omission error rates associated with point locality data that has been collected opportunistically and that represents only locations where a species has been recorded as present. This requires an independent point data set; and even if such data were available, point data are by no means straightforward to validate, since the exact location of specimens and populations (especially of animals and annual plants) changes over time.

Point data are routinely extrapolated to a binary grid where species are recorded as present or absent (Hopkinson 2000). This is not an informed extrapolation, therefore, there is a trade-off with smaller grid cells lessening the likelihood of commission errors but increasing the likelihood of omission errors simply by chance. Species distributions, patterns of species richness, and consequently species representation in conservation areas, are sensitive to scaling (Stoms 1992; Lawes & Piper 1998; Lennon *et al.* 2001; Araujo *et al.* 2005). Using larger grid cells, or scaling up, reduces the impact of spatial biases, but results in a decreased resolution which in turn enhances the oasis effect, by which grid cells are assigned the same weighting regardless of the number of occurrences they contain (Lawes & Piper 1998). Binary grids derived from point data are thus more suitable to local- and regional-scale planning than to continental- and global-scale planning. In fact, even at a relatively small scale, the assessment of the contribution of conservation areas to conservation objectives is sensitive to the way binary grids are overlaid on selectable sites. Using distribution data of bryophytes in Portugal from 10×10 km grid cells, Araujo (2004) found that existing reserves represent more or fewer bryophytes than expected by chance depending on the rule used to assign species to sites.

GEOGRAPHIC RANGES

Geographic ranges are a very simple form of predicted distribution data, where the focus is on the identification of the external boundary of species distributions. Species are assumed to occur uniformly inside their range, and the position of the range edge is drawn on maps based on expert opinion of the known and/or presumed occurrence of the species (Gaston 2003).

Strengths

Like point data, geographic ranges of many species are becoming increasingly available through scientific literature,

atlases of the distribution of taxa, and specific worldwide efforts by international conservation organizations to collate range data on large groups of species including threatened birds (Stattersfield & Capper 2000); amphibians (IUCN, Conservation International and NatureServe 2004); various freshwater organisms, marine fish, reptiles, and mammals (through ongoing global assessments by the World Conservation Union Species Survival Commission).

Since ranges contain inferences on where species could be present but have not been recorded, they can reduce geographical biases and fill gaps that exist in point locality data. Although these inferences are generally made by experts without explicit rules, they are often perceived as more defensible than those used to generate predicted distribution data. This is because they involve only one step of data processing (the interpolation of geographic ranges from existing point data using expert knowledge) and therefore can avoid the potential propagation of errors, which can result from the many steps required to generate predicted distribution data.

Weaknesses

The qualitative nature of the subjective interpolation used to define geographic ranges makes them prone to error. Furthermore, geographic ranges provide only a presence–absence indication of the distribution of a species, as opposed to the more informative probabilities of occurrence or abundances that can be derived from statistical models used to generate predicted species distribution data.

The inclusion or exclusion of outliers inside range boundaries disproportionately influence the position of the range edge (IUCN 1994). The inclusion of outliers increases the risk of selecting sites that do not contain the species of interest. The position of the range edge has important implications for conservation planning outcomes since areas identified as priorities for conservation action tend to be inherently biased towards zones of ecological transition, where species from abutting assemblages overlap and result in higher species richness than in the surrounding area (Araujo & Williams 2001; Gaston *et al.* 2001). This phenomenon is not unique to geographic range data and is also observed when other types of species data are employed. Nonetheless, when geographic ranges are used in the planning process, there is an increased risk of selecting sites that do not contain the species of interest, because range peripheries that include the extremes of species distributions contain a high proportion of commission errors. Zones of ecological transition may, however, be important for conservation, because populations occurring along ecological gradients retain the maximum amount of adaptive variation (Smith *et al.* 2001; Araujo 2002). A partial solution might be the use of conservative geographic ranges

that exclude regions of uncertain species presence when identifying areas for conservation action, while simultaneously prioritising important zones of ecological transition (e.g. Pressey *et al.* 2003).

Geographic ranges are likely to have the highest commission error rates among the three types of species data. Furthermore, the rate of commission errors associated with geographic ranges is not uniform across species, with the geographic ranges of rare or uncommon species that occur spasmodically over large areas being the most overestimated. Commission errors have higher associated risk than omission errors for conservation planning, because their potential consequence is that areas selected for conservation will not contain the species of interest. Moreover, the high commission error rate associated with geographic ranges can mask differences between areas in terms of their relative importance for species conservation. Notwithstanding, to our knowledge neither the commission error associated with geographic ranges nor the implications of these errors for conservation planning have been assessed.

Finally, the quality of geographic range data is rarely quantified by ancillary information and relies only on the credibility of the expert responsible for producing the data. This is quite different from point data that are usually qualified by dates and geographic coordinates, and from predicted distribution data that are obtained through explicit algorithms.

Geographic ranges generally provide a poor basis for quantitative analyses in conservation planning. For example, gap analysis, which is an assessment of the degree of species protection, is sensitive to the fine-scale detail of species occurrence data and an optimistic assessment of an existing reserve network is produced if there is a high commission error rate. As a consequence, the additional area requiring protection in order to achieve targets for species can be underestimated (Rondinini *et al.* 2005). Overall, since geographic ranges are generally very coarse representations of species occurrence their application is largely limited to the identification of large-scale priority areas for conservation (e.g. Rodrigues *et al.* 2004).

PREDICTED DISTRIBUTION DATA

The relationship between species locality data and environmental variables can be used to predict the distribution of a species. These data can be generated using either deductive or inductive models. For deductive models, knowledge of the species–habitat relationships is derived from the existing literature and expert knowledge and the output is an ordinal scale of habitat suitability. For inductive models, knowledge of the species–habitat relationships is inferred from the habitat where the species has been recorded and the output

is a semi-quantitative or quantitative probability of species occurrence or abundance (Corsi *et al.* 1999; Rodrigues *et al.* 2004).

Strengths

Predictive distribution models have the ability to use incomplete information (spatially incomplete point locality data or spatially implicit knowledge on species habitat preferences), to generate spatially comprehensive predictions of species distributions, avoiding many of the problems of scale inherent in the manual construction of range maps and giving information about variation in likelihood of occurrence or abundance. Predicted distribution data can thus improve the knowledge of the AOO of a species through the extrapolation of incomplete point locality data, and improve the knowledge of its EOO by interpolating the likelihood of occurrence or abundance of a species inside the range.

Predicted distribution data can be useful in systematic conservation planning for a variety of reasons. They act as filters on habitats from a species–specific viewpoint. While broad geographic ranges tend to smooth the differences between adjacent habitats, predicted distributions emphasize differences in habitat suitability by providing spatially explicit probabilities or likelihoods of species occurrence or abundance, and can do so for multiple species concurrently. While predicted distribution data are often converted to presence–absence data before being used for conservation planning, methods exist to use the probabilities of presence without transformations (e.g. Moilanen 2005a,b; Wilson *et al.* 2005). Predicted distribution data can also be used to predict changes in species distributions associated with changes in land use and climate. This avoids the potential problem associated with using EOO data when areas identified as priorities for conservation action have, for example, been exploited and converted to human-dominated land use and as such are no longer suitable habitat.

Weaknesses

Irrespective of how detailed the model used to generate predicted distribution data is, it is unlikely to take into account all of the variables that affect the distribution of a species. The majority of species distribution models are limited to the relationships between species and the environment, and do not take into account historical and biogeographical factors affecting species distributions (Morrison *et al.* 1998). This problem is often circumvented by constraining the output of models to the known EOO of the species (Boitani *et al.* 1999) or can be accommodated through the inclusion of spatial variables (e.g. Leathwick

1998). Furthermore, probabilities of species occurrence generated through inductive models should not be interpreted as the likelihood of a species persisting in an area (Williams & Araujo 2000) as factors influencing persistence are unlikely to be included as explanatory variables in the model. Such factors might include population dynamics, inter-specific competition, and individual behaviour (Van Horne 1983; Austin 2002b). Indeed, point locality and geographic range data also do not indicate the probability of species persistence. Probabilities generated through inductive models represent the probability of finding a species conditional upon the approach used to obtain the species–habitat relationships (i.e. through field sampling or expert knowledge) and the modelling approach employed (i.e. the choice of model and the choice of explanatory variables).

Deductive approaches to species distribution modelling do not generate quantitative probabilities that a species is present, and their output is limited to a qualitative assessment of levels of habitat suitability in unsurveyed areas, usually reported on an ordinal scale. Therefore, uncertainty can exist around predicted suitability levels and a possible way to deal with this uncertainty is to use different thresholds of suitability to generate binary (presence/ absence) maps of species distribution, and then analyse the sensitivity of conservation planning outcomes to the threshold used (Rondinini *et al.* 2005).

The quality of predicted distribution data depends on the quality of the input data, which is determined by

(a) For inductive models, the quality of the point locality data used to infer the species–environment relationships and whether these points are representative of the range of environmental conditions that the species occupies; for deductive models, the quality of published information and expert knowledge on the species that are translated into species–habitat relationships.
(b) The availability of maps of environmental variables used to predict the distribution of the species, and the accuracy and resolution of these maps. It should be noted that the availability and quality of these maps also influences the quality of geographic range data, although to a lesser extent because this type of information is only used as a support for the identification of the geographic range boundaries.

Although predictive models have extraordinary value in accommodating varying intensities of sampling, the resulting predicted distributions may still reflect this bias (Kadmon *et al.* 2004). Predicted distribution data can also be affected by biases in the knowledge of species–environment relationships, either due to incomplete or inaccurate point locality data (Tyre *et al.* 2003; Vaughan & Ormerod 2003; Gu & Swihart 2004), or inaccurate or biased expert knowledge (Clevenger *et al.* 2002).

The availability of maps of environmental variables represents a limiting factor for the predictive power of large-scale species distribution models (Austin 2002a). This is best illustrated with the example of habitat suitability models for African amphibians (Rondinini *et al.* 2005). The distribution of many of these species is known to be dependent on the presence of small, temporary water bodies that are necessary for reproduction. However, maps of these water bodies are not available across large regions because their locations are very difficult to define and they fluctuate unpredictably over time. Therefore, while predicted distribution data for these species are more detailed than geographic ranges and thus have the potential to refine the EOO, they cannot approximate the AOO because a key variable is not mapped. The lack of this data layer also introduces commission errors into the predicted distribution data for these species.

Another limitation is related to the resolution at which environmental variables are mapped (Huston 2002). Unless the maps are compiled for the purpose of modelling the species of interest, their resolution often does not match the scale at which the species uses the environment. As a result, individuals of low-vagility species might be capable of inhabiting fragments of habitat that are much smaller than the resolution of existing maps, introducing errors of omission. These omission errors might be biased to particular types of land cover (e.g. fragmented habitats might be omitted as good habitat if existing maps are too coarse to identify small fragments), and can in turn result in geographic biases in the predicted distribution data (habitat that is fragmented due to anthropogenic reasons tend to occur in low altitude areas that are more readily accessible and therefore the distribution of the species might be incorrectly associated with high altitude areas).

Predicted distribution data overcome many of the problems associated with point locality data and geographic range data described above, including incomplete datasets and overestimation of the EOO, respectively. However, as any model is an abstraction or generalization of reality, the limitations and uncertainty associated with this data needs to be assessed, communicated and taken into consideration in conservation planning. Methods have been developed to evaluate separately various properties of the predictive capability of species distribution models (Pearce & Ferrier 2000; Vaughan & Ormerod 2005) and to incorporate uncertainty into model predictions (Wintle *et al.* 2003). Nonetheless, at present the consideration of the uncertainty associated with predicted species distribution data and the implications of this for conservation planning is still in its infancy (Polasky *et al.* 2000; Elith *et al.* 2002; Burgman *et al.* 2005).

IMPLICATIONS OF ERRORS IN SPECIES DATA FOR THE COMPREHENSIVENESS, REPRESENTATIVENESS, ADEQUACY AND EFFICIENCY OF SELECTED RESERVE SYSTEMS

Priority areas for conservation action identified using geographic ranges and predicted distribution data are likely to be more comprehensive and representative that those designed using point locality data. This is because point data are limited to surveyed areas, while ranges and models predicting the distribution of a species infer the presence of a species in unsurveyed areas. If point data are degraded to grid cells then the comprehensiveness and representativeness of selected areas might improve, but this will occur by chance alone. This scaling-up also occurs at the expense of an ability to identify areas important for the persistence of species, because the differences in species assemblages between adjacent grid cells decrease. This affect is also prevalent in range data. The representativeness of areas selected using predicted distribution data is affected by the accuracy and generality of the data. Areas selecting using predicted data generated from only a few sampled populations of a species are likely to be less representative than those generated from the knowledge of many populations. Reserve networks based on predicted distribution data for a subset of African vertebrate taxa conferred more protection to other vertebrate taxa than the corresponding geographic ranges (Rondinini & Boitani 2006). These results suggest more cross-taxon congruence (*sensu* Su *et al.* 2004) from predicted distribution data and that this type of data may perform better than geographic ranges in terms of comprehensiveness of selected reserve networks.

The adequacy of selected areas is related to the interaction between omission and commission error rates and the way the conservation goal is formulated. Conservation goals generally fall between two extremes: from identifying a fixed proportion of the planning area that maximises the amount of biodiversity conserved to preserving a proportion (fixed, or variable depending on species characteristics) of the area occupied by each species. With fixed-area goals, commission errors greatly affect the adequacy of the selected reserve network. This is because areas might be protected because a species was thought to be present, when in fact the species was absent. The level of protection afforded by existing protected areas is therefore overestimated (Loiselle *et al.* 2003) and the uncertainty associated with the adequacy of selected areas is increased (Wilson *et al.* 2005). With fixed-area goals, point locality data collected recently and using a well-designed sampling scheme should provide the safest solution for planners in terms of the adequacy of selected areas.

When the goal is to preserve a proportion of species AOO, commission errors also affect the adequacy of reserve

systems. For example, areas selected on the basis of geographic ranges have been shown to be inadequate because range data overestimate species occupancy (Rondinini *et al.* 2005). Moreover, when the conservation planning goal is specified as a proportion of the area occupied by species, omission errors lead to an underestimation of the species distribution and therefore reduce the adequacy of the selected reserve system. Predicted distribution data, with intermediate omission and commission error rates, could provide the least biased solutions to the planner when using this type of goal.

The efficiency of selected areas is directly related to the omission error rate of the underlying species data, because this reduces the number of options available to the planner to represent each species, and therefore reduces flexibility. Point data provide the least efficient solutions to planners because they do not extrapolate species presence to unsurveyed areas and therefore can limit the choice of sites for conservation action to those where the species has been recorded. On the other hand, the apparent higher efficiency of priority areas for conservation action identified using geographic ranges is in part an effect of the flexibility arising from an overestimation of species occurrence.

Systematic conservation planning is sensitive to errors in the underlying species data (Freitag & Van Jaarsveld 1998; Gaston & Rodrigues 2003; Rondinini *et al.* 2005; Wilson *et al.* 2005; Rondinini & Boitani 2006). Nonetheless, these errors are seldom quantified by planners. The commission and omission errors associated with species occurrence data used for conservation planning must be quantified using independent data sets (e.g. Elith *et al.* 2006 and references therein). When this is infeasible, the sensitivity of selected areas to the data employed should be assessed. We argue for greater consideration of commission and omission errors in systematic conservation planning.

ACKNOWLEDGEMENTS

REFERENCES

Araujo, M.B. (1999). Distribution patterns of biodiversity and the design of a representative reserve network in Portugal. *Divers. Distrib.*, 5, 151–163.

Araujo, M.B. (2002). Biodiversity hotspots and zones of ecological transition. *Conserv. Biol.*, 16, 1662–1663.

Araujo, M.B. (2004). Matching species with reserves – uncertainties from using data at different resolutions. *Biol. Conserv.*, 118, 533–538.

Araujo, M.B. & Williams, P.H. (2001). The bias of complementarity hotspots toward marginal populations. *Conserv. Biol.*, 15, 1710–1720.

Araujo, M.B., Williams, P.H., Cabeza, M., Thuiller, W. & Hannah, L. (2004). Would climate change drive species out of reserves? An assessment of existing reserve-selection methods. *Global Change Biol.*, 10, 1618–1626.

Araujo, M.B., Thuiller, W., Williams, P.H. & Reginster, I. (2005). Downscaling European species atlas distributions to a finer resolution: Implications for conservation planning. *Global Ecol. Biogeogr.*, 14, 17–30.

Austin, M.P. (2002a). Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecol. Model.*, 157, 101–118.

Austin, M.P. (2002b). Case studies of the use of environmental gradients in vegetation and fauna modeling: theory and practice in Australia and New Zealand. In: *Predicting Species Occurrences: Issues of Accuracy and Scale* (eds Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, V.A. & Samson, F.B.). Island Press, Washington, Covelo, London, pp. 73–82.

Austin, M.P. & Heyligers, P.C. (1989). Vegetation survey design for conservation: gradsect sampling of forests in north-east New South Wales. *Biol. Conserv.*, 50, 13–32.

Boitani, L., Corsi, F., De Biase, A., D'Inzillo Carranza, I., Ravagli, M., Reggiani, G. *et al.* (1999). *A Databank for the Conservation and Management of African Mammals*. Istituto di Ecologia Applicata, Roma.

Bonn, A., Rodrigues, A.S.L. & Gaston, K.J. (2002). Threatened and endemic species: are they good indicators of patterns of biodiversity on a national scale? *Ecol. Lett.*, 5, 733–741.

Brooks, T., Da Fonseca, G.A.B. & Rodrigues, A.S.L. (2004a). Species, data, and conservation planning. *Conserv. Biol.*, 18, 1682–1688.

Brooks, T.M., Da Fonseca, G.A.B. & Rodrigues, A.S.L. (2004b). Protected areas and species. *Conserv. Biol.*, 18, 616–618.

Burgman, M.A., Lindenmayer, D.B. & Elith, J. (2005). Managing landscapes for conservation under uncertainty. *Ecology*, 86, 2007–2017.

Burrough, P.A. & McDonnell, R.A. (1998). *Principles of Geographic Information Systems*. Oxford University Press, Oxford.

Cabeza, M., Moilanen, A., Araujo, M.B., Wilson, R.J., Thomas, C.D. & Cowley, M.J.R. (2004). Combining probabilities of occurrence with spatial reserve design. *J. Appl. Ecol.*, 41, 252–262.

Clevenger, A.P., Wierzchowski, J., Chruszcz, B. & Gunson, K. (2002). GIS-generated, expert-based models for identifying wildlife habitat linkages and planning mitigation passages. *Conserv. Biol.*, 16, 503–514.

Corsi, F., de Leeuw, J. & Skidmore, A.K. (1999). Modeling species distributions with GIS. In: *Research Techniques in Animal Ecology: Controversies and Consequences* (eds Boitani, L. & Fuller, T.K.). Columbia University Press, New York, pp. 389–434.

Davis, F.W., Stoms, D.M., Estes, J.E., Scepan, J. & Scott, J.M. (1990). An information systems approach to the preservation of biological diversity. *Int. J. G.I.S.*, 4, 55–78.

Dunn, R., Harrison, A.R. & White, R.C. (1990). Positional accuracy and measurement error in digital databases: an empirical study. *Int. J. Geogr. Inf. Sci.*, 4, 385–398.

Eken, G., Bennun, L., Fishpool, L.D.C., Brooks, T.M., Foster, M., Knox, D. *et al.* (2004). Key biodiversity areas as site conservation targets. *BioSci.*, 54, 1110–1118.

Elith, J., Burgman, M.A. & Regan, H.M. (2002). Mapping epistemic uncertainty and vague concepts in predictions of species distribution. *Ecol. Model.*, 157, 313–329.

Elith, J., Graham, C.H., Anderson, R.P., Dudìk, M., Ferrier, S., Guisan, A. *et al.* (2006). Novel methods improve predictions of species' distributions from occurrence data. *Ecography*, 9, 129–151.

Erasmus, B.F.N., Freitag, S., Gaston, K.J., Erasmus, B.H. & Van Jaarsveld, A.S. (1999). Scale and conservation planning in the real world. *Proc. R Soc. Lond. Ser. B – Biol. Sci.*, 266, 315–319.

Faith, D.P., Ferrier, S. & Walker, P.A. (2004). The ED strategy: how species-level surrogates indicate general biodiversity patterns through an 'environmental diversity' perspective. *J. Biogeogr.*, 31, 1207–1217.

Ferrier, S. (2002). Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? *Syst. Biol.*, 51, 331–363.

Ferrier, S., Manion, G., Mantle, K., Powell, G.V.N., Allnutt, T.F., Burgess, N.D. *et al.* (2004). Mapping more of terrestrial biodiversity for global conservation assessment. *BioSci.*, 54, 1101–1109.

Fielding, A.H. & Bell, J.F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.*, 24, 38–49.

Freitag, S. & Van Jaarsveld, A.S. (1998). Sensitivity of selection procedures for priority conservation areas to survey extent, survey intensity and taxonomic knowledge. *Proc. R. Soc. Lond. Ser. B – Biol. Sci.*, 265, 1475–1482.

Freitag, S., Hobson, C., Biggs, H.C. & Van Jaarsveld, A.S. (1998). Testing for potential survey bias: the effect of roads, urban areas and nature reserves on a southern African mammal data set. *Anim. Conserv.*, 1, 119–127.

Funk, V.A. & Richardson, K.S. (2002). Systematic data in biodiversity studies: use it or lose it. *Syst. Biol.*, 51, 303–316.

Gaston, K.J. (1991). How large is a species' geographic range? *Oikos*, 61, 434–438.

Gaston, K.J. (1993). *Rarity*. Chapman and Hall, London.

Gaston, K.J. (2003). *The Structure and Dynamics of Geographic Ranges*. Oxford University Press, Oxford, New York.

Gaston, K.J. & Rodrigues, A.S.L. (2003). Reserve selection in regions with poor biological data. *Conserv. Biol.*, 17, 188–195.

Gaston, K.J., Rodrigues, A.S.L., Van Rensburg, B.J., Koleff, P. & Chown, S.L. (2001). Complementary representation and zones of ecological transition. *Ecol. Lett.*, 4, 4–9.

Gladstone, W. & Davis, J. (2003). Reduced survey intensity and its consequences for marine reserve selection. *Biodiv. Conserv.*, 12, 1525–1536.

Graham, C.H., Graham, C.H., Moritz, C., Ferrier, S., Huettman, F. & Peterson, A.T. (2004). New developments in museum-based informatics and applications in biodiversity analysis. *Trends Ecol. Evol.*, 19, 497–503.

Gu, W. & Swihart, R.K. (2004). Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biol. Conserv.*, 116, 195–203.

Guisan, A. & Zimmermann, N.E. (2000). Predictive habitat distribution models in ecology. *Ecol. Model.*, 135, 147–186.

Haila, Y. & Margules, C.R. (1996). Survey research in conservation biology. *Ecography*, 19, 323–331.

Harrison, J.A., Allan, D.G., Underhill, L.G., Herremans, M., Tree, A.J., Parker, V. *et al.* (1997). *The Atlas of Southern African Birds*. BirdLife South Africa, Johannesburg.

Hopkinson, P. (2000). National-scale conservation assessments at an appropriate resolution. *Divers. Distrib.*, 6, 195–204.

Howard, P.C., Viskanic, P., Davenport, T.R.B., Kigenyi, F.W., Baltzer, M., Dickinson, C.J. *et al.* (1998). Complementarity and the use of indicator groups for reserve selection in Uganda. *Nature*, 394, 472–475.

Huston, M.A. (2002). Introductory essay: critical issues for improving predictions. In: *Predicting species occurrences: issues of accuracy and scale* (eds Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, V.A. & Samson, F.B.). Island Press, Washington, Covelo, London, pp. 7–21.

IUCN (1994). *IUCN Red List Categories*. IUCN (The World Conservation Union), Gland.

IUCN, Conservation International and NatureServe (2004). *Global Amphibian Assessment*. Available at http://www.globalamphibians.org.

Kadmon, R., Farber, O. & Danin, A. (2004). Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecol. Applic.*, 14, 401–413.

Keller, C.M.E. & Scallan, J.T. (1999). Potential roadside biases due to habitat changes along breeding bird survey routes. *Condor*, 101, 50–57.

Kirkpatrick, J.B. (1983). An iterative method for establishing priorities for the selection of natural reserves: an example from Tasmania. *Biol. Conserv.*, 25, 127–134.

Larsen, F.W. & Rahbek, C. (2003). Influence of scale on conservation priority setting – a test on African mammals. *Biodiv. Conserv.*, 12, 599–614.

Lawes, M.J. & Piper, S.E. (1998). There is less to binary maps than meets the eye: the use of species distribution data in the southern African sub-region. *S. Afr. J. Sci.*, 94, 207–210.

Leathwick, J.R. (1998). Are New Zealand's Nothofagus species in equilibrium with their environment? *J. Veg. Sci.*, 9, 719–732.

Lennon, J.J., Koleff, P., Greenwood, J.J.D. & Gaston, K.J. (2001). The geographical structure of British bird distributions: diversity, spatial turnover and scale. *J. Anim. Ecol.*, 70, 966–979.

Loiselle, B.A., Howell, C.A., Graham, C.H., Goerck, J.M., Brooks, T., Smith, K.G. *et al.* (2003). Avoiding pitfalls of using species distribution models in conservation planning. *Conserv. Biol.*, 17, 1591–1600.

Lombard, A.T., Cowling, R.M., Pressey, R.L. & Rebelo, A.G. (2003). Effectiveness of land classes as surrogates for species in conservation planning for the Cape Floristic Region. *Biol. Conserv.*, 112, 45–62.

MacKenzie, D.I., Nicholls, J.D., Droege, S., Royle, J.A. & Langtimm, C.A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83, 2248–2255.

Margules, C.R. & Pressey, R.L. (2000). Systematic conservation planning. *Nature*, 405, 243–253.

Meier, R. & Dikow, T. (2004). Significance of specimen databases from taxonomic revisions for estimating and mapping the global species diversity of invertebrates and repatriating reliable specimen data. *Conserv. Biol.*, 18, 478–488.

Moilanen, A. (2005a). Methods for reserve selection: interior point search. *Biol. Conserv.*, 124, 485–492.

Moilanen, A. (2005b). Reserve selection using nonlinear species distribution models. *Am. Nat.*, 165, 695–706.

Morrison, M.L., Marcot, B.G. & Mannan, R.W. (1998). *Wildlife–Habitat Relationships: Concepts and Applications*, 2nd edn. University of Wisconsin Press, Madison, WI.

Nelson, B.W., Ferreira, C.A.C., da Silva, M.F. & Kawasaki, M.L. (1990). Endemism centres, refugia and botanical collection intensity in Brazilian Amazonia. *Nature*, 345, 714–716.

Pearce, J. & Ferrier, S. (2000). Evaluating the predictive performance of habitat models developed using logistic regression. *Ecol. Model.*, 133, 225–245.

Polasky, S., Camm, J.D., Solow, A.R., Csuti, B., White, D. & Ding, R. (2000). Choosing reserve networks with incomplete species information. *Biol. Conserv.*, 94, 1–10.

Ponder, W.F., Carter, G.A., Flemons, P. & Chapman, R.R. (2001). Evaluation of museum collection data for use in biodiversity assessment. *Conserv. Biol.*, 15, 648–657.

Possingham, H.P., Wilson, K.A., Andelman, S.J. & Vynne, C.H. (2006). Protected areas: goals, limitations, and design. In: *Principles of Conservation Biology* (eds Groom, M.J., Meffe, G.K. & Carroll, C.R.). Sinauer Associates Inc., Sundaland, MA, pp. 509–533.

Pressey, R.L. (2004). Conservation planning and biodiversity: assembling the best data for the job. *Conserv. Biol.*, 18, 1677–1681.

Pressey, R.L., Cowling, R.M. & Rouget, M. (2003). Formulating conservation targets for biodiversity pattern and process in the Cape Floristic Region, South Africa. *Biol. Conserv.*, 112, 99–127.

Reddy, S. & Davalos, L.M. (2003). Geographical sampling bias and its implications for conservation priorities in Africa. *J. Biogeogr.*, 30, 1719–1727.

Reid, W.V. (1998). Biodiversity hotspots. *Trends Ecol. Evol.*, 13, 275–280.

Rodrigues, A.S.L., Gaston, K.J. & Gregory, R.D. (2000a). Robustness of reserve selection procedures under temporal species turnover. *Proc. R. Soc. Lond. Ser. B – Biol. Sci.*, 267, 49–55.

Rodrigues, A.S.L., Gaston, K.J. & Gregory, R.D. (2000b). Using presence-absence data to establish reserve selection procedures that are robust to temporal species turnover. *Proc. R. Soc. Lond. Ser. B – Biol. Sci.*, 267, 897–902.

Rodrigues, A.S.L., Andelman, S.J., Bakarr, M.I., Boitani, L., Brooks, T.M., Cowling, R.M. *et al.* (2004). Effectiveness of the global protected area network in representing species diversity. *Nature*, 428, 640–643.

Rondinini, C. & Boitani, L. (2006). Differences in the umbrella effects of African amphibians and mammals based on two estimators of the area of occupancy. *Conserv. Biol.*, 20, 170–179.

Rondinini, C., Stuart, S. & Boitani, L. (2005). Habitat suitability models reveal shortfall in conservation planning for African vertebrates. *Conserv. Biol.*, 19, 1488–1497.

Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, V.A. *et al.* (2002). *Predicting Species Occurences: Issues of Accuracy and Scale*. Island Press, Washington, Covelo, London.

Smith, T.B., Kark, S., Schneider, C.J., Wayne, R.K. & Moritz, C. (2001). Biodiversity hotspots and beyond: the need for preserving environmental transitions. *Trends Ecol. Evol.*, 16, 431.

Stattersfield, A.J. & Capper, D.R. (2000). *Threatened Birds of the World: The Official Source for Birds on the IUCN Red List*. BirdLife International, Cambridge.

Stoms, D.M. (1992). Effects of habitat map generalization in biodiversity assessment. *Photogramm. Eng. Remote Sens.*, 58, 1587–1591.

Su, C.J., Debinski, D.M., Jakubauskas, M.E. & Kindscher, K. (2004). Beyond species richness: community similarity as a measure of cross-taxon congruence for coarse-filter conservation. *Conserv. Biol.*, 18, 167–173.

Tyre, A.J., Possingham, H.P. & Lindenmayer, D.B. (2001). Inferring process from pattern: can territory occupancy provide information about life history parameters. *Ecol. Applic.*, 11, 1722–1737.

Tyre, A.J., Tenhumberg, B., Field, S.A., Possingham, H.P., Niejalke, D. & Parris, K. (2003). Improving precision and reducing bias in biological surveys: Estimating false-negative error rates. *Ecol. Applic.*, 13, 1790–1801.

Van Horne, B. (1983). Density as a misleading indicator of habitat quality. *J. Wildl. Manage.*, 47, 893–901.

Vaughan, I.P. & Ormerod, S.J. (2003). Improving the quality of distribution models for conservation by addressing shortcomings in the field collection of training data. *Conserv. Biol.*, 17, 1601–1611.

Vaughan, I.P. & Ormerod, S.J. (2005). The continuing challenges of testing species distribution models. *J. Appl. Ecol.*, 42, 720–730.

Wessels, K.J., Reyers, B. & Van Jaarsveld, A.S. (2000). Incorporating land cover information into regional biodiversity assessments in South Africa. *Anim. Conserv.*, 3, 67–79.

Williams, P.H. & Araujo, M.B. (2000). Using probability of persistence to identify important areas for biodiversity conservation. *Proc. R. Soc. Lond. Ser. B – Biol. Sci.*, 267, 1959–1966.

Williams, P.H., Margules, C.R. & Hilbert, D.W. (2002). Data requirements and data sources for biodiversity priority area selection. *J. Biosci.*, 27, 327–338.

Wilson, K.A., Westphal, M.I., Possingham, H.P. & Elith, J. (2005). Sensitivity of conservation planning to different approaches to using predicted species distribution data. *Biol. Conserv.*, 122, 99–112.

Wintle, B.A., McCarthy, M.A., Volinski, C.T. & Kavanagh, R.P. (2003). The use of Bayesian model averaging to better represent uncertainty in ecological models. *Cons. Biol.*, 17, 1579–1590.

Wintle, B.A., McCarthy, M.A., Parris, K. & Burgman, M.A. (2004). Precision and bias of methods for estimating point survey detection problems. *Ecol. Applic.*, 14, 703–712.

Wintle, B.A., Burgman, M.A., McCarthy, M.A. & Kavanagh, R.P. (2005). Estimating and dealing with detectability in occupancy surveys: forest owls and arboreal marsupials. *J. Wildl. Manage.*, 69, 905–917.