Received: 30 June 2020 | Accepted: 12 October 2020

DOI: 10.1111/1365-2664.13782

POLICY DIRECTION

Journal of Applied Ecology | BRITISH ECOLOGICAL SOCIETY

Species distribution modelling is needed to support ecological impact assessments

David J. Baker[1] | Ilya M. D. Maclean[1] | Martin Goodall[2] | Kevin J. Gaston[1]

[1]Environment and Sustainability Institute, University of Exeter, Cornwall, UK

[2]Cornwall Wildlife Trust, Cornwall, UK

Correspondence
David J. Baker
Email: d.baker2@exeter.ac.uk

Handling Editor: Joseph Bennett

Abstract

1. Legislation commonly mandates the mitigation of impacts to biodiversity in planning and development processes, with potential impacts identified through some form of ecological impact assessment. Yet, protections for biodiversity are frequently undermined because the distributions of priority species are poorly known in most locations at the spatial scales required to inform planning decisions (i.e. *c*. 1–100 ha).
2. Planning applications are often screened against opportunistic records to determine potential impacts to priority species. However, raw occurrence records provide information only on where a species has been detected and cannot be used to indicate if a species is likely to be absent from a site.
3. Inferences drawn from these data on the likelihood of a species being present at a site can only be correctly interpreted through an appropriate species distribution modelling (SDM) framework that ensures assumptions about the data and models are formalised and documented. We argue that SDM frameworks must be integrated into ecological impact assessments to improve support for biodiversity protections within planning and development processes.
4. Biases and uncertainties in opportunistic data create modelling challenges, but recent methodological advances can minimise their impacts on predictions. We advocate co-production with practitioners of SDM tools, mapping products and best-practice guidelines specific to planning processes.
5. *Policy implications*. The integration of species distribution modelling frameworks into ecological impact assessments will strengthen biodiversity protections in planning and development processes by ensuring methodological transparency and rigour in the interpretation of species occurrence data.

KEYWORDS
environmental impact assessment, habitat suitability, incidental observations, species occurrence, threatened species

1 | INTRODUCTION

Biodiversity underpins many of the key ecosystem services provided by multifunctional landscapes, including food and water security and human health and wellbeing (Isbell et al., 2017). The integration of species protection within planning and development processes is a legal requirement in the United Kingdom under the National Planning Framework and is also a goal of the government's 25-year environment plan (DEFRA, 2018). Similar protections are in place in many other countries (Glasson & Therivel, 2013). Legislation typically mandates that potential impacts are identified during the planning stages of a proposed development and that appropriate action is taken to mitigate these impacts. It is, however, only possible to provide mitigation to reduce or remove impacts if the species that will be negatively affected by a proposed development are identified during an ecological impact assessment. Yet, knowing where species occur in a landscape is a fundamental challenge that hampers efforts to protect biodiversity (Jetz et al., 2019).

The distributions of species from a few taxonomic groups (e.g. Aves, Mammalia, Lepidoptera, and Vascular plants) in some geographical regions (e.g. populous areas of Europe and North America) are increasingly well-documented at landscape scales (e.g. with a precision of 10–100 km). However, the precise patterns of occupancy at finer scales are not known for the majority of species in the majority of locations. Planning decisions are typically made at field scales (i.e. 1–100 ha) and, although they may draw on information from larger scales, ignorance of occupancy patterns at these resolutions is a major obstacle for species protection. To protect biodiversity effectively, information on species' distributions is needed at the appropriate scale, particularly for species requiring special legal protection. This provides practitioners with the information necessary to implement effective mitigation early in the planning stages, to target surveys to minimise the likelihood of overlooking priority species, and to design landscapes to maintain connectivity and environmental flows. We propose that opportunistic records can help deliver information at the appropriate scale, but only if they are integrated into planning processes using a species distribution modelling (SDM) framework. An SDM framework encompasses the entire model building and mapping process, allowing for the evaluation of the data and the modelling process at multiple points (e.g. amount and quality of species data, model fit, cross-validation performance) to determine the reliability of the information produced. The SDM framework should be designed to ensure a transparent and reproducible methodology, with assumptions about the data formalised, and a standardised reporting structure.

Here, we provide the rationale for integrating SDM frameworks into ecological impact assessments aimed at supporting planning and development processes and highlight the importance of co-production between modellers, ecologists and other stakeholders for delivering this information.

2 | SHORTCOMINGS IN THE USE OF OCCURRENCE RECORDS IN PLANNING PROCESSES

Field-scale information on species that might be affected by a proposed development can be obtained by conducting an on-site environmental impact assessment (EIA). However, these are not only expensive but also of variable quality (e.g. dependent on ecologist's skill level) and utility (e.g. dependent on when they are conducted) and often incur costs relatively late in planning processes (Glasson & Therivel, 2013). In determining whether an EIA is required or how it should be focused (e.g. taxa or timing), the presence of priority species is often determined through the screening of local species occurrence records against planning proposals to identify species that have been recorded in close proximity to a site (Figure 1). Vast numbers of species occurrence records are now being collated online (Kays et al., 2020), with the overwhelming majority collected opportunistically (i.e. without a standardised sampling strategy), trading-off quality for quantity, both in terms of geographical and taxonomical coverage (Jetz et al., 2019). These records often provide the necessary level of precision (i.e. field scale) and have the considerable advantage they can be screened early in the planning process while incurring minimal costs.

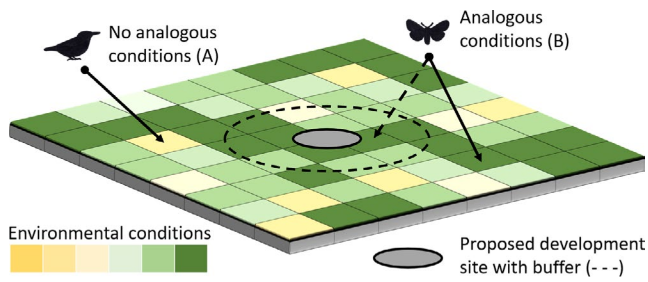

FIGURE 1 The screening of opportunistic species occurrence records against development plans is often carried out to establish whether any priority species are likely to be impacted by a development. Screening is often conducted using a buffer around the proposed site (- - -). However, the absence of opportunistic records within the buffer is not sufficient to determine that a species does not occur there. It would be an error to conclude the absence of a species from a site based only on occurrence records if suitable environmental conditions (i.e. conditions analogous to those in which the species is known to occur locally) are found at the site (species A vs. B)

Occurrence records alone, however, can only provide information on the locations where a species has previously been detected and provide absolutely no information in areas where records are absent without analysis of observer effort (Figure 2a). In screening records, clarity is required on how the absence of records at a location should be interpreted. Specifically, the absence of records does not equate to the absence of the species, particularly when there is little distinction between the environmental conditions at a proposed development site and the conditions in which a particular species is known to occur locally (Figure 1). The effects of these errors are to weaken protections for biodiversity in planning and development processes and to increase project costs when the need for mitigation is belatedly discovered.

3 | ADDING VALUE TO OPPORTUNISTIC BIOLOGICAL RECORDS USING AN SDM FRAMEWORK

Raw opportunistic species occurrence information is inadequate for identifying potential ecological impacts, but analysis of those same data through an SDM framework can provide considerable

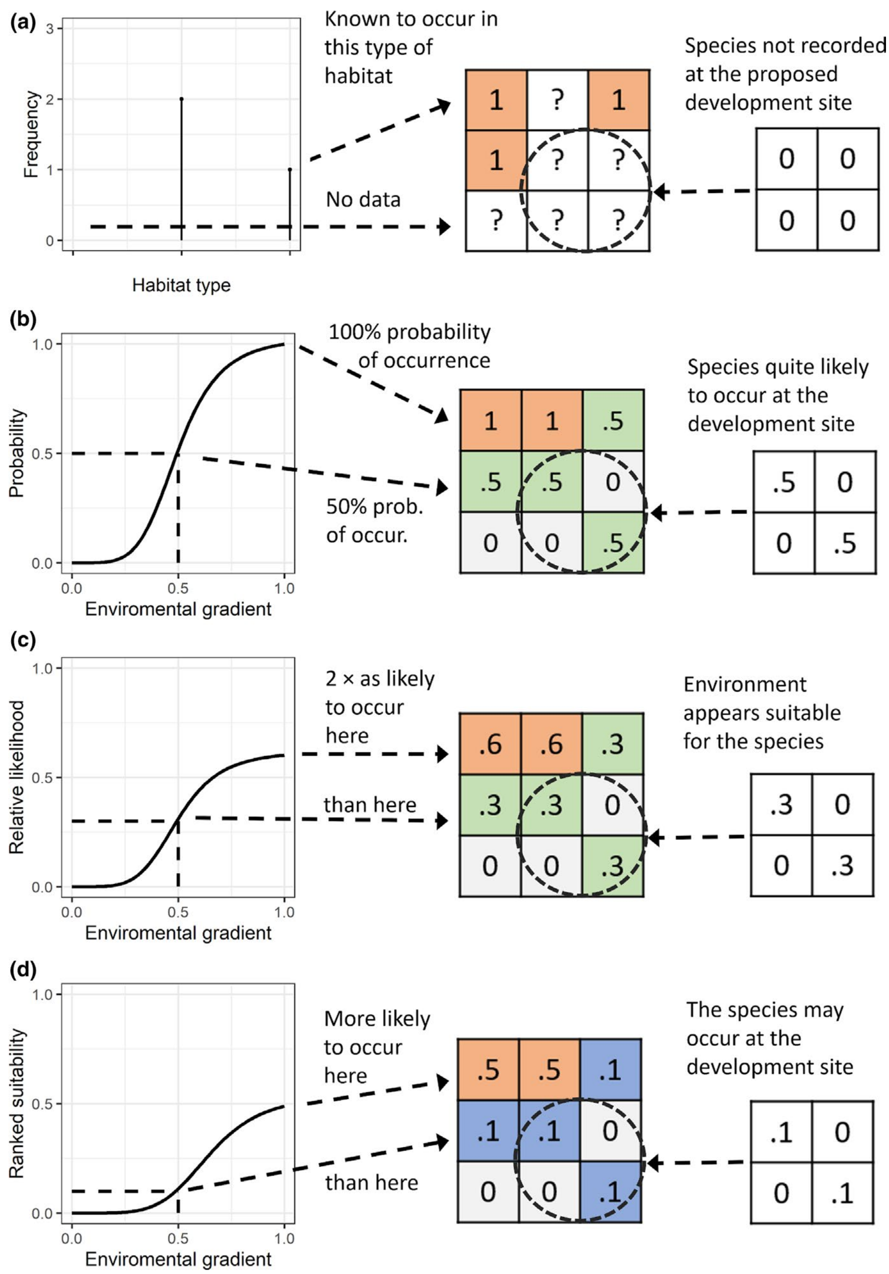

FIGURE 2 Species occurrence records alone provide limited information on species' distributions (a), but species distribution models can add value to these data by (depending on the data and model) providing a spatially explicit estimate of the: (b) probability of occurrence (POC); (c) relative likelihood of occurrence; or (d) ranked suitability. The POC indicates the true probability of the species occurring at a site with particular environmental conditions. Across sites with POC = 0.5, we would expect the species to occur at 50% of these sites. The relative likelihood of occurrence is proportional to the probability of occurrence but not calibrated to give the exact probability of finding the species at a site (for this we need to know the prevalence of the species in the landscape). The ranked suitability indicates the ranked order of suitable sites, but is not proportional to the probability of occurrence and so does not necessarily indicate how much more or less suitable a site is for a species

added value by attempting to infer the occurrence of the species at a site, or the potential suitability of a site for a species, including sites previously unsurveyed, based on local environmental conditions. Opportunistic records are prone to a broad range of biases and SDMs must be developed carefully to produce unbiased estimates of species distributions from opportunistic records (Phillips et al., 2009). The knowledge, computational tools, environmental data and species information now exist to improve field scale knowledge of species distributions and to integrate this information into planning processes.

The exact quantity estimated by an SDM and how this quantity can be used to inform planning and development decisions depends strongly on the underlying data. Under rare circumstances, opportunistic data are of sufficient quality to provide estimates of the probability of occurrence (POC; Figure 2b). Short of census data, POC estimates provide the highest quality information on species distributions, with well-calibrated models providing accurate information on the true POC at a site (Guillera-Arroita et al., 2015). This information is highly valuable for risk assessment, cost-benefit analyses and targeting monitoring, but is difficult to acquire at fine scales and for many species, requiring repeat site surveys to enable species detectability to be estimated. More often, model outputs from opportunistic records provide a relative suitability index (Figure 2c) or ranking of suitability (Figure 2d).

Estimates of relative suitability are obtained when the prevalence of the species is unknown, but environmental space is sampled in proportion to its availability and the detectability of the species is consistent across habitats and conditions (Guillera-Arroita et al., 2015). Relative suitability estimates are proportional to the true POC and, consequently, provide excellent information for discriminating the most from the least suitable sites (Figure 2c). These outputs provide a clear indication of sites that are entirely unsuitable for the species and, therefore, where risks to the species from a proposed development are minimal (i.e. zeros in Figure 2c). This information is invaluable in the early stages of planning to help avoid areas where a planning proposal is most likely to require mitigation (e.g. in routing a road) and is useful for targeting in situ monitoring to identify populations at risk (Fois et al., 2018). With careful application of models to opportunistic data (i.e. to address sampling biases), it is possible to estimate the relative suitability of the landscape for many conspicuous species.

Suitability rankings provide an estimate of the ranked suitability of sites, although the suitability estimates are not proportional to the true POC (Figure 2b vs. d). Although this type of information is less valuable than the POC or relative suitability, value is added to the raw occurrence data (Figure 2a vs. d). In particular, the model output provides information in areas that have not been surveyed, such as 'site X is ranked among the most suitable sites' or 'the environmental conditions at site X appear to be at least as suitable as locations with known occurrences'. Information on absences is less clear with suitability rankings, especially if the sampling is biased (i.e. available habitat is not surveyed in proportion to its availability) and the bias is not adequately corrected. In these circumstances, suitable but unsampled environmental space may appear to have the same ranking as entirely unsuitable environmental space (Figure 2b vs. d). This is a shortcoming, and these models must be interpreted carefully.

Regardless of the output, viewing occurrence records through an SDM framework usually adds value over the raw occurrence data because it ensures assumptions about the data are clearly stated, the methods are fully documented, models are evaluated using consistent criteria, and uncertainty in predictions is quantified. For many species and in many locations, it is likely that the available information will be insufficient to enable reliable inferences to be drawn regarding occupancy at sites where the species has not previously been detected. This, however, would be a considerable improvement on the current situation because support for the conclusion would be clearly evidenced. Such model insuffiencies could be determined by performance in cross-validation, evaluation against new field data, or by examination of the confidence intervals on estimates of POC or suitability. Ultimately, adding rigour to the interpretation of biodiversity data and its use within planning processes can only serve to enhance biodiversity protections.

4 | INTEGRATING SDMS IN PLANNING PROCESSES

To integrate SDMs into planning processes, it is important to consider processes end-to-end, including feedbacks between gaps and uncertainties in the model outputs and field data collection. At the initial stages of integration, we propose that the major focus should be on developing consistent modelling protocols, the co-production of mapping products and SDM best-practice guidelines specific to planning processes. We discuss some of these issues below.

4.1 | Data and modelling

Opportunistic records are the least information-rich data source for SDMs because they typically provide limited direct information on the distribution of survey effort (c.f. species inventories within a known search area and time period), and—in part because of this—provide no direct information on species absences. In place of absence information, models must instead be conditioned on locations with known presences and a sample of background locations, with the background data serving as a means of estimating the relative suitability of the available environmental space (Phillips & Elith, 2013). While simple in construct, naïve presence-background (PB) models can produce highly biased predictions if several challenges are not addressed in model building (Phillips et al., 2009), including spatial sampling biases, unknown prevalence and variations in detectability.

To some degree, the effects of these challenges can be overcome or reduced through careful model building. For example, spatial sampling biases (e.g. towards roads, nature reserves and urban areas) can be corrected to some extent using approaches such as target group background sampling (Phillips et al., 2009) and covariate-based methods (Warton et al., 2013). Unknown species prevalence in the landscape, which will occur unless, by chance, the ratio of presences to background points is equal to the ratio that would be expected in a random sample of sites, biases model prediction relative to the POC (Guillera-Arroita et al., 2015). However, additional information on species prevalence, even for just a few locations or from expert opinion, can be used to adjust the ratio of presence to background points appropriately (Golding, 2013). Methods have even been developed to correct for non-detections by integrating small replicated monitoring datasets designed to estimate detectability with opportunistic data (Koshkina et al., 2017). Where these issues persist, the challenge lies in creating outputs that encourage the correct interpretation of mapping products (see below).

A reasoned decision must also be made on the spatial scales used to build SDMs. There is no universal 'best' grain or extent, and the choice will vary depending on the quality of the occurrence information (e.g. number and geographical precision), ecology of the species (e.g. home range size) and the scales at which planning decisions are required (e.g. dwelling vs. highway). For example, motile species are likely to use different components of the landscape for different activities (e.g. sheltering, foraging and movement) and the extent and grain of the model, and the occurrence data itself, may capture only some of these components. The occurrence of species in a landscape is often conceptualised as a hierarchy, starting with a species' geographical range and ending with the selection of resources from within available habitat components (Johnson, 1980). Meyer and Thuiller (2006) suggest that the particular level of the hierarchy being analysed (e.g. local population or home range) should dictate the grain of the model. The wider landscape context of a site must also be considered if this affects occurrence. For example, it might be necessary to consider the habitat structure surrounding a site where, for example, fragmentation or distance from source populations may affect population persistence. Multigrain models have been shown to help constrain finer-grained models within the wider landscape context and provide considerable performance gains over single-grain models (Meyer & Thuiller, 2006). In many cases, the number of occurrence records will set a lower limit on the grain that should be used to build SDMs, and these constraints might preclude the building of models entirely or make the predictions too coarse for planning purposes. This information in itself is useful because it highlights data gaps and cautions against using existing data for a particular species to underpin decisions.

An additional challenge of occurrence records is that they have often been accrued over long time periods. There is a trade-off between constraining the period over which records are used—which will increase their relevance—and losing potentially valuable data. For opportunistic data, this can be a real problem, as collection has seldom been done at an even rate, with the periods of activity of individuals or groups of enthusiasts being critical. Temporal sampling biases will affect models correlating occurrence with environmental conditions if the species has been recorded at a site and then subsequently extirpated due to the site becoming environmentally unsuitable (rather than for stochastic reasons or deliberate removal). Where data permit, the spatio-temporal occupancy dynamics or data structure could be explicitly modelled (Martínez-Minaya et al., 2018). However, solutions are more likely to involve excluding records that fall outside of a time period considered stationary with respect to the environment. The specification of temporal windows should be underpinned by analysis or expert knowledge of regional environmental changes that may have affected species occurrence at a location (i.e. changes in land use practices).

4.2 | Easily interpretable outputs

A challenge facing the successful integration of SDMs into the planning process is ensuring that the outputs are easy to interpret. This challenge is not unique to SDMs, as species occurrence records are routinely misinterpreted in the planning process when used outside of a model framework. In fact, without viewing these data through an SDM framework, the interpretation of species occurrence records will never be clear or transparent. The use of SDMs provides an opportunity to develop outputs that effectively convey the best available knowledge on a species' distribution. However, even in optimal circumstances, conveying uncertain information to non-experts is not straightforward. Evidence suggests that end users frequently misinterpret probabilistic weather forecast information (e.g. probability of precipitation) by, for example, interpreting an 80% chance of rain to mean that rain will fall over 80% of the area or for 80% of the time (Joslyn et al., 2009). The exact quantity estimated by SDMs varies (Figure 2) and the interpretation of these quantities once they deviate from POC becomes even less intuitive than simple weather forecast probabilities. It is surprising that little effort has been made to evaluate the interpretation of SDM outputs by end users, and only sparse efforts have been made to communicate uncertainty visually in SDM maps (Golding, 2013). The co-production of mapping products with relevant end users and stakeholders is therefore critical to ensure that there is clarity and confidence in model outputs (Sofaer et al., 2019).

4.3 | Model co-production

Ecologists, modellers and stakeholders must work together to ensure that planning decisions are supported by the best available biodiversity information, including through the development and validation of SDMs. Close collaboration will be required to develop and validate models for each species initially; these could then be updated at regular intervals (e.g. annually) or after new records have been submitted and periodically reviewed by experts and evaluated against new data (Sofaer et al., 2019). The model outputs themselves could be used to inform the strategic collection of validation data. Co-production should, however, occur within a standardised framework (e.g. developed and implemented at a national level) to ensure that the modelling process is transparent and fully auditable, and to minimise disagreements over the use and interpretation of models and data during decision processes.

5 | PATHWAY FORWARD

The integration of an SDM framework into ecological impact assessments aimed at identifying potential impacts to biodiversity from proposed development projects will help improve protections for biodiversity in multifunctional landscapes. The availability of occurrence records and several decades of SDM development means that many historical barriers to the widespread use of SDMs can now be overcome. We suggest that the most important factor for the successful integration of SDMs into the planning process is co-production. Through all stages of development—including software, maps, metadata and best-practice standards—relevant stakeholders must be involved in decisions, and the final products must have considerable buy-in from all parties. Confidence will be gained through methodological transparency, particularly as this will force stakeholders to confront existing shortcomings in the use of occurrence data. The protection of biodiversity in planning and development processes must now reap the benefits of academic advances in modelling species' distributions.

ACKNOWLEDGEMENTS

We are grateful to the editor and two anonymous reviewers for comments on an earlier version of this work. This work is a contribution to the UKRI funded Landscape Decisions Programme.

AUTHORS' CONTRIBUTIONS

All authors contributed to the conception of the article, and D.J.B. led the writing with contributions from all authors.

DATA AVAILABILITY STATEMENT

We (I) will not be archiving data because this manuscript does not use data.

ORCID

David J. Baker https://orcid.org/0000-0002-0466-8222
Ilya M. D. Maclean https://orcid.org/0000-0001-8030-9136
Kevin J. Gaston https://orcid.org/0000-0002-7235-7928

REFERENCES

DEFRA. (2018). *A green future: Our 25 year plan to improve the environment*. HM Government London.

Fois, M., Cuena-Lombraña, A., Fenu, G., & Bacchetta, G. (2018). Using species distribution models at local scale to guide the search of poorly known species: Review, methodological issues and future directions. *Ecological Modelling*, *385*, 124–132. https://doi.org/10.1016/j.ecolmodel.2018.07.018

Glasson, J., & Therivel, R. (2013). *Introduction to environmental impact assessment*. Routledge.

Golding, N. (2013). *Mapping and understanding the distributions of potential vector mosquitoes in the UK: New methods and applications* (PhD thesis). University of Oxford.

Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., McCarthy, M. A., Tingley, R., & Wintle, B. A. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology & Biogeography*, *24*(3), 276. https://doi.org/10.1111/geb.12268

Isbell, F., Gonzalez, A., Loreau, M., Cowles, J., Díaz, S., Hector, A., & Larigauderie, A. (2017). Linking the influence and dependence of people on biodiversity across scales. *Nature*, *546*(7656), 65–72.

Jetz, W., McGeoch, M. A., Guralnick, R., Ferrier, S., Beck, J., Costello, M. J., Fernandez, M., Geller, G. N., Keil, P., Merow, C., Meyer, C., Muller-Karger, F. E., Pereira, H. M., Regan, E. C., Schmeller, D. S., & Turak, E. (2019). Essential biodiversity variables for mapping and monitoring species populations. *Nature Ecology and Evolution*, *3*(4), 539–551. https://doi.org/10.1038/s41559-019-0826-1

Johnson, D. H. (1980). The comparison of usage and availability measurements for evaluating resource preference. *Ecology*, *61*(1), 65–71. https://doi.org/10.2307/1937156

Joslyn, S., Nadav-Greenberg, L., Nichols, R. M., Berg, L. N., & Nichols, R. M. (2009). Probability of precipitation: Assessment and enhancement of end-user understanding. *Bulletin of the American Meteorological Society*, *90*(2), 185–194. https://doi.org/10.1175/2008BAMS2509.1

Kays, R., McShea, W. J., & Wikelski, M. (2020). Born-digital biodiversity data: Millions and billions. *Diversity and Distributions*, *26*(5), 644–648. https://doi.org/10.1111/ddi.12993

Koshkina, V., Wang, Y., Gordon, A., Dorazio, R. M., White, M., & Stone, L. (2017). Integrated species distribution models: Combining presence-background data and site-occupany data with imperfect detection. *Methods in Ecology and Evolution*, *8*(4), 420–430.

Martínez-Minaya, J., Cameletti, M., Conesa, D., & Pennino, M. G. (2018). Species distribution modeling: A statistical review with focus in spatio-temporal issues. *Stochastic Environmental Research and Risk Assessment*, *32*(11), 3227–3244. https://doi.org/10.1007/s00477-018-1548-7

Meyer, C. B., & Thuiller, W. (2006). Accuracy of resource selection functions across spatial scales. *Diversity and Distributions*, *12*(3), 288–297. https://doi.org/10.1111/j.1366-9516.2006.00241.x

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications*, *19*(1), 181–197. https://doi.org/10.1890/07-2153.1

Phillips, S. J., & Elith, J. (2013). On estimating probability of presence from use-availability or presence-background data. *Ecology*, *94*(6), 1409–1419. https://doi.org/10.1890/12-1520.1

Sofaer, H. R., Jarnevich, C. S., Pearse, I. S., Smyth, R. L., Auer, S., Cook, G. L., Edwards, T. C., Guala, G. F., Howard, T. G., Morisette, J. T., & Hamilton, H. (2019). Development and delivery of species distribution models to inform decision-making. *BioScience*, *69*(7), 544–557. https://doi.org/10.1093/biosci/biz045

Warton, D. I., Renner, I. W., & Ramp, D. (2013). Model-based control of observer bias for the analysis of presence-only data in ecology. *PLoS ONE*, *8*(11), e79168. https://doi.org/10.1371/journal.pone.0079168

How to cite this article: Baker DJ, Maclean IMD, Goodall M, Gaston KJ. Species distribution modelling is needed to support ecological impact assessments. *J Appl Ecol*. 2021;58:21–26. https://doi.org/10.1111/1365-2664.13782