Transferring Biodiversity Models for Conservation: Opportunities and Challenges

AMM Sequeira[1], PJ Bouchet[2], KL Yates[3], K Mengersen[56], MJ Caley[56]

Running title: Roadmap for improved model transfers in ecology

Word count: 7010

[1] IOMRC and Australian Institute of Marine Science, UWA Oceans Institute and School of Biological Sciences, M096, The University of Western Australia, Crawley, Western Australia, Australia

[2]Centre for Marine Futures, School of Biological Sciences, The University of Western Australia, Western Australia 6009, Australia

[3]School of Environment and Life Sciences, University of Salford, Manchester, UK

[4]School of Biological Sciences, University of Queensland, St. Lucia, Queensland, Australia

[5]School of Mathematical Sciences, Queensland University of Technology, Brisbane, Queensland, Australia

[6]Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS)

Corresponding author:

IOMRC building Room 2.23; Cnr Fairway and Service Road 4 (M470)
The University of Western Australia
35 Stirling Highway, Crawley; Western Australia 6009

Phone: +61 8 6488 2219

Email: ana.sequeira@uwa.edu.au

Abstract

1. After decades of extensive surveying, knowledge of the global distribution of species still remains inadequate. In the short to medium term, such knowledge is unlikely to improve greatly given the often prohibitive costs of surveying and the typically limited resources available.
2. By forecasting biodiversity patterns in time and space, predictive models can help fill critical knowledge gaps and prioritize research to support better conservation and management.
3. The ability of a model to predict biodiversity metrics in novel environments is termed ‘transferability’, and models with high transferability will be the most useful in this context.
4. Despite their potential broad utility, little guidance exists on what confers high transferability to biodiversity models.
5. We synthesise recent advances in biodiversity model transfers to facilitate increased understanding of what underpins successful model transferability, demonstrating that a consistent approach has so far been lacking but is essential for achieving high levels of repeatability, transparency, and accountability of model transfers.
6. We provide a set of guidelines to support efficient learning and the improvement of model transferability.

Keywords:

Statistics, Applied ecology, Biodiversity, Conservation, Modelling, Extrapolation, Forecasting

Model Transferability as a Tool for Conservation and Management

Effective conservation and management of biodiversity requires a robust understanding of the distribution and status of wildlife species. However, our existing knowledge of biological systems is frequently inadequate and field surveys too costly to generate the data necessary for addressing many immediate management needs (Margules and Pressey 2000). As a result, spatial planning is often constrained to proceed in the absence of relevant ecological information (Possingham et al. 2007). In these cases, indirect methods for estimating biodiversity patterns, such as predictive models, could be of great utility for decision-making. Indeed, correlative models relating ecological metrics to environmental and spatial predictors (henceforth biodiversity models) are now commonplace (see e.g., Franklin 2010 for a summary of methods used and a framework for spatial prediction of species occurrence) and play a critical role in supporting management and conservation efforts worldwide (Margules and Pressey 2000, Robinson et al. 2011). Such models are increasingly sought where data are scarce or non-existent, for example, when predicting the distributions of threatened or invasive species into unsampled locations, or under future climates (Austin and Meyers 1996, Elith et al. 2010, Duque-Lazo et al. 2016). While the literature on biodiversity models is now extensive, little guidance exists on how to maximise the utility of transferring these models.

The ability of a model developed for a specific site and/or time to predict biodiversity in a different time or place defines its transferability (Fig. 1), which can be tested for specific taxa or across taxa. Transferability therefore, encompasses the statistical concept of extrapolation which is embedded in the broader topic of model validation. Transferability has the potential to leverage legacy investments in sampling and increase the efficiency of management and conservation actions. However, whilst the development of biodiversity models and model validation has proceeded rapidly in the last decade (Fig. 2) (Elith et al. 2006, Elith and Leathwick 2009), their transferability remains poorly understood (Robinson et al. 2011), particularly in marine systems (e.g., Lauria et al. 2015, Sequeira et al. 2016*b*). Moreover, transferability assessments have not yet been standardised, leading to disparate interpretations that preclude comparisons of relative performance among model transfers. To address these knowledge gaps, we synthesise the emerging field of model transferability with the aim of fostering consistent and transparent model transfers in ecological studies. We begin by defining multiple scenarios under which model transfers can take place, reviewing lessons learnt, and describing modelling approaches applied to date. We then clarify critical, yet commonly ignored assumptions, discussing perceived constraints and limitations of model transferability, and highlighting the diversity of assessment metrics currently available. Building on these observations, we describe how the standardisation of transferability assessments among studies will catalyse the widespread application of model transfers in marine and terrestrial environments. We make practical recommendations on ways to achieve such consistency and advocate for future research that would improve the transferability of biodiversity models.

From Reference to Target Systems

A model must be developed for a reference system (location and/or point in time) where data are available at a number of sampling locations, before being transferred. This reference model is calibrated within a range of values for each predictor, and then projected, or transferred, to a target system (some other location or time). Model transferability can take many forms depending on the predictors available for the target system, their range of values, as well as those of the response variable (Fig. 1). The most straightforward situations are those where both response and predictor variables are within the same ranges in both the reference and target systems (e.g., Mannocci et al. 2015) (Fig 1a, External 1; and Fig 1b, blue ellipse). In this scenario, the environmental space is the same in both systems and the realised niche in the target system is fully contained in the model created for the reference system.

Good transferability can therefore be expected, assuming reasonable overlap in the combinations of response and predictor variables. When the range of predictors in the target system is within that used for model calibration in the reference system, but the response variable is not (e.g., when observed abundances are different in the target and reference systems) (e.g., Thogmartin et al. 2006, Owens et al. 2013), two situations can arise: (i) poor transferability should be expected where the relationship between response and predictor variables differs (Fig 1b, External 2), and (ii) when this relationship is similar (Fig 1b, External 3), transferred predictions will always be biased. If this bias is relatively consistent (e.g., always lower or higher), then a simple correction to the resulting predictions could allow for good transferability. For example, the probability of occurrence of a species in the target system might always be 20 % lower than in the reference system, in which case a simple correction to account for this bias could result in a shift from poor to good model transferability. Where sample sizes are sufficiently large, the predictive performance of the transferred model can be robustly tested, yielding potential insights into how model transferability is affected by novel conditions. When predictor values are within the range of those used in the reference model (Fig 1a, External 5), models transferred from an information-rich area may successfully predict species distributions and help inform new sampling designs in spite of data scarcity in the target system. However, when the range of predictor values in the target system extends beyond that encountered during model building for the reference system, model transfers will involve some mathematical extrapolation (Fig 1b, External 4 and Fig 1a, External 6 and 7). If the range of predictors in the target system is entirely outside that used in the reference model (Fig 1d, External 7), transferability tests would typically only serve to capture the potential distributions of species under novel environmental conditions. Despite the many possible different scenarios, confidence in the resulting predictions derived from transferred models is paramount and will depend strongly on the identification of opportunities for robust model validation. Therefore the form of the transferability test being performed should always be explicitly described in model transferability studies, including a summary of the relationships between predictor and response variables for data from reference and target systems.

The amount of data available in each system will also be a key determinant of the model transfer protocols chosen in any given situation, and three different scenarios can be considered. First, where sufficient information exists for both response and predictor variables in the reference and target systems (data-rich), it is possible to directly test model transferability through empirical validation (Barbosa et al. 2009, Sequeira et al. 2016*b*). Such scenarios are useful to better understand the factors affecting model transfers (Fig. 3) and inform model transfer endeavours elsewhere. Second, when data for both response and predictor variables at the target system are insufficient (data-deficient), testing model transferability is only possible if alternative data sources for predictor variables are sought and/or if assumptions are made about the similarity of the covariate space and relationships between the response and predictor variables (Fig. 3) (Schadt et al. 2002). Third, when information is lacking for either response or predictor variables in the target system (Fig. 3) (data-sparse), several courses of action are possible (e.g., Sequeira et al. 2014). Where data for the target system are available for predictor variables but not response variables, the same set of predictor variables (p) need to be considered in both the reference and target systems, and it is preferable that the range of each predictor in both systems is similar, as detailed above. If the range of predictors in the target system exceeds that in the reference system, caution is advised (see extrapolation section). Examples and recommendations for each of the three scenarios presented are included in Fig 3.

Choice of Modelling Algorithms Can Affect Transferability

The transferability of predictive models can be dataset, question, and algorithm-specific (Elith et al. 2006, Duque-Lazo et al. 2016). It is often thought that model simplicity and parsimony should be preferred over complexity. Simpler models are easier to interpret and also return smoother response curves that are relatively immune to overfitting and may therefore be more widely applicable (Vaughan and Ormerod 2005, Lauria et al. 2015). However, defining complexity of biodiversity models is not straightforward as it depends on the interplay between the complexity of the underlying processes and the amount of data available to build the model. For these reasons, past comparisons of model performance as a function of complexity may have been unfair (García-Callejas and Araújo 2015). New complexity indices are being studied (García-Callejas and Araújo 2015, Moreno-Amat et al. 2015, Bell and Schlaepfer 2016), but further testing of their influence on the interpretation of model transferability is needed before clear advice can be given.

Similarly, best practice regarding algorithm choice for testing model transferability is, so far, not settled. For example, previous studies have reported better transferability in generalised additive or linear models (GAMs or GLMs), maximum entropy models (MaxEnt) and boosted regression trees (BRTs) compared to random forests (RFs) (Dobrowski et al. 2011, Heikkinen et al. 2012). GAMs have also been seen as robust when predictions are needed beyond the range of sampled values (i.e., when extrapolating) (Fronzek et al. 2011), despite being unconstrained outside the bounds of the observations used during model calibration (unlike RFs, MaxEnt and BRTs). Similarly, in the history of machine learning studies, no clear preference for a specific method has been provided despite multiple comparisons (Fielding 1999, Hsieh 2009). Without consensus on the best algorithms or methods to develop more transferable models (Araujo and Guisan 2006, Randin et al. 2006), three different approaches to multi-model comparisons have been suggested: (*i*) adopt a single modelling procedure regarded as the most suitable to the research problem being addressed; (*ii*) implement a suite of algorithms and proceed with the best performing one; or (*iii*) combine several approaches to generate "ensemble" forecasts. While it is hard to provide general guidance, ensemble approaches have been suggested to offer superior predictive power (Marmion et al. 2009, Grenouillet et al. 2011).

All of the approaches described above relate to correlative models. Although beyond the scope of this review, process-based mechanistic models are also worth considering in the context of transferability because they can explicitly incorporate the dynamics of biological processes thought to limit species' ranges (Buckley et al. 2010) such as dispersal, connectivity, and trophic interactions. Mechanistic models, however, tend to be more challenging to construct and test, owing to greater demands on computational resources and data needed for calibration and validation. Consequently, the transfer of mechanistic models in predictive ecology remains rare despite their considerable promise in strengthening and guiding conservation practice under climate change and biological invasions (Kearney et al. 2010, Cuddington et al. 2013).

Common Assumptions Underlying Model Transfers

Irrespective of the approach used, model transferability hinges on a number of assumptions that are commonly overlooked (Mesgaran et al. 2014, Werkowska et al. 2017) and may explain some difficulties in obtaining good model transferability. For example, species are often assumed to be at equilibrium with their environment and present in all suitable sites (Araujo and Pearson 2005). However, such equilibrium is taxon-dependent and inversely related to a taxon's ability to track dynamic conditions, and should therefore be explicitly considered beforehand (Araujo and Pearson 2005). The set of conditions under which species can persist (niche, sensu Hutchinson 1957) is also commonly assumed to be stationary in both space and time (Wiens and Graham 2005). A discussion on the reasoning behind this assumption should be clearly specified in any transfer test as it may not hold if existing biotic or abiotic interactions break down, shift, or new ones form (Broennimann et al. 2007, Martinez-Freiria et al. 2016). Moreover, increasing evidence for rapid niche shifts (Broennimann et al. 2007) may render static models invalid for predicting into space or time (Forester et al. 2013). Integrating information from mechanistic models can improve trust in the resulting predictions in this context (Elith et al. 2010). Species occurrences are also commonly assumed to be solely driven by environmental forces, but realised distributions can also reflect current/past disturbances and human uses, and so historical events should be considered when interpreting transferability results (Beans et al. 2012, Davis et al. 2014). Another common assumption is space-for-time substitutability (Blois et al. 2013), under which unobservable past or future temporal trends are inferred from contemporary spatial models of different aged locations. However, conflicting predictions generated by different models (Kharouba et al. 2009b) challenge the general validity of this assumption, and the current best advice is, whenever possible, to use models that have been successful at hindcasting (Kharouba et al. 2009a).

Overall, the transferability of biodiversity models can be affected by a wide range of factors associated with the choice of study design, model algorithm, or target taxa, as well as with the type, quantity and quality of input data, and the characteristics of the environment (Table 1, and references within). To understand what confers transferability on biodiversity models, more will need to be learned about the consequences of violating assumptions. This highlights the need to encourage best practice when testing model transferability by explicitly checking the validity of assumptions wherever possible (Werkowska et al. 2017) or explicitly reporting why they cannot be checked and if they are likely to affect the results of the transferred model. Results of these checks need to be reported as they will be crucial for understanding and improving model transferability.

Common Methods for Assessing Model Transferability

Model accuracy generally decreases when a model is transferred (Kharouba et al. 2009b, Dobrowski et al. 2011, Torres et al. 2015). Assessments of model results beyond where and when data were available (external evaluations) are commonly more variable than within the same location or time (internal evaluations) (Dobrowski et al. 2011). Furthermore, better fit during internal evaluation has not always resulted in greater model transferability (Heikkinen et al. 2012, Sequeira et al. 2016*b*) and the reported ability of transferred models to predict well has varied widely (Fielding and Haworth 1995, Randin et al. 2006, Martin et al. 2012, Sequeira et al. 2016b).

Part of this variability in transferability across studies may stem from the range of different metrics used during evaluation (Wang and Jackson 2014), with at least 17 different metrics reported to date (Table 2). Metrics derived from confusion matrices, such as sensitivity, area under the curve (AUC)/receiver operating characteristic (ROC), Cohen's kappa statistic, or the c-index (Vaughan and Ormerod 2005), which are based on the number of true and false positives and negatives, have been favoured for models of presence/absence or presence-only data (Randin et al. 2006). ROC has however, been considered weak (Peterson et al. 2007), but see Phillips (2008). Spearman rank and coefficients of determination (R^2) from linear regressions are commonly used for models using count data (Lauria et al. 2015) to assess the relationship between observed and predicted values. Compound performance metrics have also been proposed and applied, including the AUC-based transferability index TRI (Table 2) (Randin et al. 2006), which compares internal and external evaluations made in both directions. Reciprocal model transferability is, however, not always possible or warranted (Sequeira et al. 2016b), and being AUC-based, TRI only applies to models of presence-absence data. Furthermore, indices focusing on the evaluation of model fit only provide partial assessments of transferability; they do not evaluate maps of transferred predictions (Randin et al. 2006). To assess these predictions, other metrics have been proposed, including Kulczynski's coefficient (Legendre and Legendre 2012), Schoener's D (Warren et al. 2010), and the I similarity statistic (or Modified Hellinger Distance) (Warren et al. 2008). In face of such heterogeneity, standardised assessments of results across studies, modelling techniques, datasets and environmental settings are needed to improve the utility of model transfers for conservation. Such standardization will facilitate direct comparisons across studies and promote their transparent interpretation, thereby strengthening our ability to meaningfully address key knowledge gaps in model transferability.

Transferability assessments can also be model-focused, instead of data-focused, in which case the onus is on finding the most general model (Wenger and Olden 2012). So, when testing model transferability, a thorough assessment of model generality should also be performed (Wenger and Olden 2012). As a general principle, only models demonstrating good fit and evaluation for the reference system should be transferred, even if they do not always result in good transferability. For example, better transferability is expected of models developed for large, heterogeneous ecosystems where sampling effort has been sufficient to capture the full range of environmental conditions available (sensu Sequeira et al. 2016*b*). Therefore, when transferring models it is important to clearly justify the choice of a particular model for any particular situation.

A Standardised Roadmap for Model Transferability Studies

When quantifying transferability, three aspects of model performance are commonly compared (adapted from Randin et al. 2006). These include the fit to the set of observations on which the model was built in the reference system, evaluation against observed data in the reference system not used for model training, and the model predictive power to the target system by comparison with observations or to the results of models developed in the new location or time.

Comparisons of model fit between reference and target systems are relevant for transferability assessments (see Fig. 3 for options when model fit in the target system is poor) but are not universally applicable because comparable results are only expected in special cases, such as when covariance structures are similar in both systems. By contrast, focusing on model evaluation and comparison of prediction results should allow meaningful insights to be obtained on the relevance of distinct modelling approaches relative to available datasets and study sites. Even though the choice of metric for comparing predictive results might vary between studies (Table 2), most metrics can be derived from a small set of basic statistics. These include confusion matrices (TP, FP, TN and FN; refer to Table 2) and comparisons of grid-cell values, such as predicted probabilities from a presence-only model or predicted abundances or differences between observed and predicted values. If these basic statistics are reported alongside the results for the metric chosen in each study, they can then be used to derive most other metrics commonly used. They can also be used to calculate additional descriptive statistics (e.g., mean, standard deviation, covariance, collinearity or correlation) for datasets resulting from any model type. Such reporting will facilitate comparisons across studies (Koricheva et al. 2013), and improve our ability to provide clearer guidance in the future. Moreover, to further expedite comparisons, we suggest as a minimum that one metric, the coefficient of determination (R^2), should always be reported.

In relation to model predictive power, the assessment of predictions from transferred models should also estimate uncertainty and spatial dependencies among grid-cells. This can be achieved using the structural similarity index (SSIM) (Jones et al. 2016), which was designed to compare distribution maps of different species groups, and to assess the similarity in the two sets of spatial data (i.e. maps). SSIM provides an overall measure of similarity ranging between -1 and 1, and could, therefore, be used in the context of model transferability to test the predictive power of the transferred model through comparison with predictions obtained from a model developed for the target system. SSIM also compares the mean, variance, and covariance of both datasets, and provides information on where dissimilarities might be greatest(?). Such information will be relevant for understanding where and why a model may fail to transfer well (see example of application of SSIM to model transferability in Fig. 4).

These ways of standardising transferability assessments are appropriate for data-rich scenarios. However, models are usually transferred because data are lacking for the target system, making direct validation challenging. Where sufficient predictor data are available, model transferability can still be tested and result in informative predictions (Fig. 3; data-sparse scenarios). Trust in transferred predictions, however, can be increased by selecting models that have already been successfully transferred to other analogous target system(s), in a similar way as hindcasting is used to increase confidence in climate change predictions. Resulting predictions should then be used to inform new survey designs for the target system, and validation should be contemplated as these data become available. Such procedures can provide transparent assessments prior to decision-making, and allow for adaptive learning in support of the best possible management outcomes at each step. The 'holy grail' of model transferability is, clearly, to achieve reliable predictions into unsampled space/time, an exercise that involves some kind of extrapolation (mathematical, geographical or combinatorial, i.e., derived from novel combinations of covariates within the univariate range). Indeed, overcoming challenges associated with predicting into unknown space/time is becoming urgent in face of accelerating environmental change.

The Uncharted Waters of Mathematical Extrapolation: "Here Be Monsters"?

More and more models are being projected into non-homologous environments (Elith et al. 2011b), even though predictions obtained under previously unobserved conditions are prone to both statistical and ecological error (Dormann 2007), and despite the problems associated with extrapolation being long-known (Perrin 1904). Regardless of such potential errors, the immediate need for solutions to wildlife management problems within new temporal or spatial domains often overrides caution in extrapolation, with forecasts frequently treated as reliable (Fitzpatrick and Hargrove 2009, Jimenez-Valverde et al. 2011). Because attention to extrapolation risks is central to effective decision-making, the development of techniques that can diagnose or address the consequences of extrapolation when transferring models has been considered a research priority (Araujo and Guisan 2006).

Various strategies have been suggested for dealing with extrapolation. These include avoidance, mitigation, and explicit description. The 'avoidance' strategy is based on the subjective rule of thumb that extrapolation errors increase with distance from the envelope of the observed data. It has been suggested that, if predictions are not made beyond 1/10th of the sampled parameter range, extrapolation might be considered negligible (Dormann 2007). Extrapolation mitigation can be achieved by minimising the likelihood of encountering novel combinations of environmental conditions, e.g., by modelling the complete range of a species given its dispersal abilities and any major biogeographical transitions when projecting future distributions (Thuiller et al. 2004). Methodologies have also been proposed to explicitly highlight extrapolated predictions, e.g. by discriminating the areas where the data used most resemble the characteristics of the reference system or by explicitly defining prediction intervals (which highlight the probability that a new result will lie within certain values) (e.g., Kampichler and Sierdsema 2017). Examples include the Multivariate Environmental Similarity Surface (MESS) within MaxEnt (Elith et al. 2011a), which relies on a rectilinear (univariate) detection of extrapolation. Other techniques to highlight extrapolated predictions include ‘extrapolation detection’ (Mesgaran et al. 2014), ‘environmental overlap mask’ (Owens et al. 2013), ‘envelope uncertainty maps’ (Platts et al. 2008), ‘prediction uncertainty assessments using residual variation’ (Rödder and Engler 2012), and ‘mobility-oriented parity’ (Zurell et al. 2012) are also capable of highlighting different types of environmental novelty arising from mathematical and combinatorial extrapolation (Capinha and Pateiro-López 2014, Conn et al. 2015b). ‘Explicit description’ has therefore been suggested by many researchers as a way to deal with extrapolation, echoing the solution presented by Steel *et al.* (2013) of ‘being honest’. Others have also highlighted the issues of extrapolation as sometimes being more philosophical than statistical (Conn et al. 2015a).

Ultimately, the ‘monsters’ that lurk in the waters of extrapolation will be less ferocious if the uncertainties underlying model projections are not ignored. And here, we focus on Just as mean values should be reported with their associated confidence intervals, there is mounting support for reporting of extrapolated results in prediction maps (Fitzpatrick and Hargrove 2009, Mesgaran et al. 2014). Because excluding areas as uninhabitable is currently impossible for most taxa, particularly data-poor ones (Fitzpatrick and Hargrove 2009), validation of results will need to be prioritized and comparisons across models and modelling techniques (sensu Fielding 1999, Hsieh 2009) will continue to be helpful in the context of model transferability. We suggest that standardisation in the presentation of extrapolated results by explicitly using one of the three strategies suggested above would assist our general understanding of model transferability. Importantly, wherever possible, extrapolation results should be tested to improve the performance of transferred models through adaptive learning (Dormann 2007).

Towards Better Practice in Model Transferability

A number of rules of good practice for testing model transferability have already been put forward in the literature and are summarised in Table 1. Key principles highlighted relate to the consideration of species characteristics, chosen predictors, and modelling approaches (Werkowska et al. 2017). For example, it has been suggested that only models of species whose physiological constraints will remain similar throughout their ranges should be used for testing (Pearman et al. 2008). While such constraints might be difficult to establish, one possible way to promote best practise would be to compare forecasts from transferred models only for species for which hindcasting has demonstrated good predictive performance (Kharouba et al. 2009b). Predictions of species distributions in different areas should also be tested (Lauria et al. 2015) including those separated by thousands of km (Sundblad et al. 2009, Sequeira et al. 2016a). Suggestions that individual species traits and functional groups should be well known prior to testing transferability have also been put forward, as well as the previous assessment of the effects of population demographics and conservation status. Furthermore, competition among species should also be considered, especially when such interactions are dynamic.

The careful choice of predictors has been advised due to potential effects of using proxy data and scale dependencies (see Table 1 for details). In particular, the discriminative ability and effect sizes of the predictor variables should be assessed before models are transferred. Also, predictors included in the models, their estimated coefficients, and resulting predictions should all be fully reported. Exploration of other traits (e.g., fine scale traits such as molecular data) that might be more correlated with specific geographical features, has also been encouraged.

Best practice in modelling includes careful checks of model assumptions, and reporting on residuals to allow better interpretation of the results. The use of multiple modelling approaches should also be considered to test appropriateness for different applications. This will be useful to reduce or quantify uncertainty, acknowledging though that models resulting in good internal evaluation (and interpolation) will not always transfer well. The effects of model complexity should also be explored, as they are highly dependent on the system and on the data at hand. Finally, we argue here that best practice should include standardisation of the presentation of results to facilitate cross-study comparisons and allow for general improvements in model transferability and in our understanding of what affects or hinders the transferability of biodiversity models. We therefore suggest that model transferability studies specifically include a summary section where each of the optional steps are clearly stated (Box 1).

Conclusion

The future of model transferability and its utility for conservation and management will rely on the careful consideration of each of the above recommendations, together with the presentation of standardised results from transferability assessments. To promote advances in model transferability, we propose that ten key points should be addressed. While the focus of this manuscript is on correlative models, most of these guidelines will also be applicable to mechanistic models:

1- Define the situation according to the scheme presented in Fig 1

2- Summarise the relationships between predictor and response variables in the reference and target systems

3- Justify assumptions and check their validity where possible

4- Decide on a strategy to deal with extrapolation: avoidance, mitigation or discrimination

5- Check the list of suggestions for best practise included in Table 1 and address the relevant points for the study

6- Decide on the modelling approach; using the procedure most suitable for the study, implementing a suite of algorithms and selecting the best performing one, or using an ensemble approach

7- Develop reference models following Fig. 3, depending on the amount of data available for the study

8- Check model fit and assess model performance focusing on:

  a. Evaluation by reporting the basic statistics (e.g., observed vs predicted comparisons or true and false positives and negatives, for continuous or categorical variables, respectively) that can be used to calculate common metrics, including, as a minimum, R2, which can be considered equivalent to the normalised chi-squared (i.e., the ratio of the observed chi-squared to the maximum possible chi-squared for categorical data)(Rosenberg 2010) and provides a score between 0 and 1 for the overall assessment.

  b. Predictive power by using the SSIM where applicable (i.e., data-rich scenarios)

9- Clearly indicate where predictions are the result of extrapolation

10- Report a minimum standardized set of results to facilitate comparisons between studies (see Box 1).

Transferring predictive biodiversity models has the potential to become a time- and cost-effective tool for management and conservation, provided it is done in a consistent way that guarantees transparency, comparability and confidence in model outputs. The information we synthesise here (Table 1 and Fig. 3) can assist in dealing with different scenarios where model transfers may be helpful. The careful consideration of the issues we outline will improve the performance of model transfers in ecological studies and management practice, and more importantly, will help us learn what confers transferability on biodiversity models. As understanding increases, these recommendations may evolve but the need for standardisation will remain. Lastly, because model transferability is relevant in other fields, such as transport (Ibeas et al. 2012, Rashidi et al. 2013), health and economics (Drummond et al. 2009), and urban pollution (Patton et al. 2015), cross-fertilization among these communities of practice has the potential to further enhance our confidence in transferring models, their resulting predictions, and their application to real-world situations.

References

Araujo, M. B., and A. Guisan. 2006. Five (or so) challenges for species distribution modelling. Journal of Biogeography **33**:1677-1688.

Araujo, M. B., and R. G. Pearson. 2005. Equilibrium of species' distributions with climate. Ecography **28**:693-695.

Austin, M. P., and J. A. Meyers. 1996. Current approaches to modelling the environmental niche of eucalypts: Implication for management of forest biodiversity. Forest Ecology and Management **85**:95-106.

Barbosa, A. M., R. Real, and J. M. Vargas. 2009. Transferability of environmental favourability models in geographic space: the case of the Iberian desman (Galemys pyrenaicus) in Portugal and Spain. Ecological Modelling **220**:747-754.

Beans, C. M., F. F. Kilkenny, and L. F. Galloway. 2012. Climate suitability and human influences combined explain the range expansion of an invasive horticultural plant. Biological Invasions **14**:2067-2078.

Bell, D. M., and D. R. Schlaepfer. 2016. On the dangers of model complexity without ecological justification in species distribution modeling. Ecological Modelling **330**:50-59.

Blois, J. L., J. W. Williams, M. C. Fitzpatrick, S. T. Jackson, and S. Ferrier. 2013. Space can substitute for time in predicting climate-change effects on biodiversity. Proceedings of the National Academy of Sciences of the United States of America **110**:9374-9379.

Broennimann, O., U. A. Treier, H. Müller-Schärer, W. Thuiller, A. Peterson, and A. Guisan. 2007. Evidence of climatic niche shift during biological invasion. Ecology Letters **10**:701-709.

Buckley, L. B., M. C. Urban, M. J. Angilletta, L. G. Crozier, L. J. Rissler, and M. W. Sears. 2010. Can mechanism inform species’ distribution models? Ecology Letters **13**:1041-1054.

Capinha, C., and B. Pateiro-López. 2014. Predicting species distributions in new areas or time periods with alpha-shapes. Ecological Informatics **24**:231-237.

Conn, P. B., D. S. Johnson, and P. L. Boveng. 2015a. On Extrapolating Past the Range of Observed Data When Making Statistical Predictions in Ecology. PLoS ONE **10**.

Conn, P. B., D. S. Johnson, and P. L. Boveng. 2015b. On Extrapolating Past the Range of Observed Data When Making Statistical Predictions in Ecology. PLoS ONE **10**:e0141416.

Cuddington, K., M. J. Fortin, L. R. Gerber, A. Hastings, A. Liebhold, M. O'Connor, and C. Ray. 2013. Process-based models are required to manage ecological systems in a changing world. Ecosphere **4**.

Davis, A. Y., N. Malas, and E. S. Minor. 2014. Substitutable habitats? The biophysical and anthropogenic drivers of an exotic bird's distribution. Biological Invasions **16**:415-427.

Dobrowski, S. Z., J. H. Thorne, J. A. Greenberg, H. D. Safford, A. R. Mynsberge, S. M. Crimmins, and A. K. Swanson. 2011. Modeling plant ranges over 75 years of climate change in California, USA: temporal transferability and species traits. Ecological Monographs **81**:241-257.

Dormann, C. F. 2007. Promising the future? Global change projections of species distributions. Basic and Applied Ecology **8**:387-397.

Drummond, M., M. Barbieri, J. Cook, H. A. Glick, J. Lis, F. Malik, S. D. Reed, F. Rutten, M. Sculpher, and J. Severens. 2009. Transferability of Economic Evaluations Across Jurisdictions: ISPOR Good Research Practices Task Force Report. Value in Health **12**:409-418.

Duque-Lazo, J., H. van Gils, T. Groen, and R. Navarro-Cerrillo. 2016. Transferability of species distribution models: The case of Phytophthora cinnamomi in Southwest Spain and Southwest Australia. Ecological Modelling **320**:62-70.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. Overton, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberón, S. Williams, M. S. Wisz, and N. E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography **29**:129-151.

Elith, J., M. Kearney, and S. Phillips. 2010. The art of modelling range-shifting species. Methods in Ecology and Evolution **1**:330-342.

Elith, J., and J. R. Leathwick. 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution and Systematics **40**:677-697.

Elith, J., S. J. Phillips, T. Hastie, M. Dudík, Y. E. Chee, and C. J. Yates. 2011a. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions **17**:43–57.

Elith, J., S. J. Phillips, T. Hastie, M. Dudík, Y. E. Chee, and C. J. Yates. 2011b. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions **17**:43-57.

Fielding, A. H. 1999. Machine learning methods for ecological applications. Springer Science + Business Media, New York.

Fielding, A. H., and P. F. Haworth. 1995. Testing the generality of bird-habitat models. Conservation Biology **9**:1466-1481.

Fitzpatrick, M. C., and W. W. Hargrove. 2009. The projection of species distribution models and the problem of non-analog climate. Biodiversity and Conservation **18**:2255-2261.

Forester, B. R., E. G. DeChaine, and A. G. Bunn. 2013. Integrating ensemble species distribution modelling and statistical phylogeography to inform projections of climate change impacts on species distributions. Diversity and Distributions **19**:1480-1495.

Franklin, J. 2010. Mapping species distributions - spatial inference and prediction. Cambridge University Press, New York, United States of America.

Fronzek, S., T. Carter, and M. Luoto. 2011. Evaluating sources of uncertainty in modelling the impact of probabilistic climate change on sub-arctic palsa mires. Natural Hazards and Earth System Sciences **11**:2981-2995.

García-Callejas, D., and M. B. Araújo. 2015. The effects of model and data complexity on predictions from species distributions models. Ecological Modelling.

Grenouillet, G., L. Buisson, N. Casajus, and S. Lek. 2011. Ensemble modelling of species distribution: the effects of geographical and environmental ranges. Ecography **34**:9-17.

Heikkinen, R. K., M. Marmion, and M. Luoto. 2012. Does the interpolation accuracy of species distribution models come at the expense of transferability? Ecography **35**:276-288.

Hsieh, W. W. 2009. Machine learning methods in environmental sciences: neural networks and kernels. Cambridge University Press, Cambridge

Hutchinson, G. E. 1957. Concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology **22**:415-427.

Ibeas, A., J. L. Moura, A. Nuzzolo, and A. Comi. 2012. Urban freight transport demand: transferability of survey results analysis and models. Proceedings of Ewgt 2012 - 15th Meeting of the Euro Working Group on Transportation **54**:1068-1079.

Jimenez-Valverde, A., A. T. Peterson, J. Soberon, J. M. Overton, P. Aragon, and J. M. Lobo. 2011. Use of niche models in invasive species risk assessments. Biological Invasions **13**:2785-2797.

Kampichler, C., and H. Sierdsema. 2017. On the usefulness of prediction intervals for local species distribution model forecasts. Ecological Informatics **dx.doi.org/10.1016/j.ecoinf.2017.07.003**.

Kearney, M. R., B. A. Wintle, and W. P. Porter. 2010. Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. Conservation Letters **3**:203-213.

Kharouba, H. M., A. C. Algar, and J. T. Kerr. 2009a. Historically calibrated predictions of butterfly species' range shift using global change as a pseudo-experiment. Ecology **90**:2213-2222.

Kharouba, H. M., A. C. Algar, and J. T. Kerr. 2009b. Historically calibrated predictions of butterfly species' range shift using global change as a pseudo-experiment. Ecology **90**:2213-2222.

Koricheva, J., J. Gurevitch, and K. Mengersen. 2013. Handbook of meta-analysis in ecology and evolution. Princeton University Press.

Lauria, V., A. M. Power, C. Lordan, A. Weetman, and M. P. Johnson. 2015. Spatial transferability of habitat suitability models of Nephrops norvegicus among fished areas in the Northeast Atlantic: sufficiently stable for marine resource conservation? PLoS ONE **10**:e0117006.

Legendre, P., and L. Legendre. 2012. Numerical ecology. 3rd English edition. Elsevier, Amsterdam, The Netherlands.

Mannocci, L., P. Monestiez, S. Jerôme, and V. Ridoux. 2015. Extrapolating cetacean densities beyond surveyed regions: habitat-based predictions in the circumtropical belt. Journal of Biogeography **42**:1267-1280.

Margules, C. R., and R. L. Pressey. 2000. Systematic conservation planning. Nature **405**:243-253.

Marmion, M., M. Parviainen, M. Luoto, R. K. Heikkinen, and W. Thuiller. 2009. Evaluation of consensus methods in predictive species distribution modelling. Diversity and Distributions **15**:59-69.

Martin, J., E. Revilla, P.-Y. Quenette, J. Naves, D. Allainé, and J. E. Swenson. 2012. Brown bear habitat suitability in the Pyrenees: transferability across sites and linking scales to make the most of scarce data. Journal of Applied Ecology **49**:621-631.

Martinez-Freiria, F., P. Tarroso, H. Rebelo, and J. C. Brito. 2016. Contemporary niche contraction affects climate change predictions for elephants and giraffes. Diversity and Distributions **22**:432-444.

Mesgaran, M. B., R. D. Cousens, and B. L. Webber. 2014. Here be dragons: A tool for quantifying novelty due to covariate range and correlation change when projecting species distribution models. Diversity and Distributions **20**:1147-1159.

Moreno-Amat, E., R. G. Mateo, D. Nieto-Lugilde, N. Morueta-Holme, J.-C. Svenning, and I. García-Amorena. 2015. Impact of model complexity on cross-temporal transferability in Maxent species distribution models: An assessment using paleobotanical data. Ecological Modelling **312**.

Owens, H. L., L. P. Campbell, L. L. Dornak, E. E. Saupe, N. Barve, J. Soberón, K. Ingenloff, A. Lira-Noriega, C. M. Hensz, C. E. Myers, and A. T. Peterson. 2013. Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. Ecological Modelling **263**:10-18.

Patton, A. P., W. Zamore, E. N. Naumova, J. I. Levy, D. Brugge, and J. L. Durant. 2015. Transferability and Generalizability of Regression Models of Ultrafine Particles in Urban Neighborhoods in the Boston Area. Environmental Science & Technology **49**:6051-6060.

Pearman, P. B., C. F. Randin, O. Broennimann, P. Vittoz, W. O. v. d. Knaap, R. Engler, G. L. Lay, N. E. Zimmermann, and A. Guisan. 2008. Prediction of plant species distributions across six millennia. Ecology Letters **11**:357-369.

Perrin, E. 1904. On some dangers of extrapolation. Biometrika **3**:99-103.

Peterson, A. T., M. Papes, and M. Eaton. 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. Ecography **30**:550-560.

Phillips, S. J. 2008. Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson et al. (2007). Ecography **31**:272-278.

Platts, P. J., C. J. McClean, J. C. Lovett, and R. Marchant. 2008. Predicting tree distributions in an East African biodiversity hotspot: model selection, data bias and envelope uncertainty. Ecological Modelling **218**:121-134.

Possingham, H. P., H. Grantham, and C. Rondinini. 2007. How can you conserve species that haven't been found? Journal of Biogeography **34**:758-759.

Randin, C. F., T. Dirnböck, S. Dullinger, N. E. Zimmermann, M. Zappa, and A. Guisan. 2006. Are niche-based species distribution models transferable in space? Journal of Biogeography **33**:1689-1703.

Rashidi, T. H., J. Auld, and A. Mohammadian. 2013. Effectiveness of Bayesian Updating Attributes in Data Transferability Applications. Transportation Research Record:1-9.

Robinson, L. M., J. Elith, A. J. Hobday, R. G. Pearson, B. E. Kendall, H. P. Possingham, and A. J. Richardson. 2011. Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. Global Ecology and Biogeography **20**:7889-7802.

Rödder, D., and J. O. Engler. 2012. Disentangling interpolation and extrapolation uncertainties in species distribution models: A novel visualization technique for the spatial variation of predictor variable collinearity. Biodiversity Informatics **8**.

Rosenberg, M. S. 2010. A Generalized Formula for Converting Chi-Square Tests to Effect Sizes for Meta-Analysis. PLoS ONE **5**.

Schadt, S., E. Revilla, T. Wiegand, F. Knauer, P. Kaczensky, U. Breitenmoser, L. Bufka, J. Cerveny, P. Koubek, T. Huber, C. Stanisa, and L. Trepl. 2002. Assessing the suitability of central European landscapes for the reintroduction of Eurasian lynx. Journal of Applied Ecology **39**:189-203.

Sequeira, A. M., C. Mellin, H. M. Lozano-Montes, M. A. Vanderklift, R. C. Babcock, M. D. Haywood, J. J. Meeuwig, and M. J. Caley. 2016a. Transferability of predictive models of coral reef fish species richness. Journal of Applied Ecology **53**:64-72.

Sequeira, A. M. M., C. Mellin, D. A. Fordham, M. G. Meekan, and C. J. A. Bradshaw. 2014. Predicting current and future global distributions of whale sharks. Global Change Biology **20**:778–789.

Sequeira, A. M. M., C. Mellin, H. M. Lozano-Montes, M. A. Vanderklift, R. C. Babcock, M. Haywood, J. J. Meeuwig, and M. J. Caley. 2016b. Transferability of predictive models of coral reef fish species richness. Journal of Applied Ecology **53**:64-72.

Steel, E. A., M. C. Kennedy, P. G. Cunningham, and J. S. Stanovick. 2013. Applied statistics in ecology: common pitfalls and simple solutions. Ecosphere **4**.

Sundblad, G., M. Härmä, A. Lappalainen, L. Urho, and U. Bergström. 2009. Transferability of predictive fish distribution models in two coastal systems. Estuarine, Coastal and Shelf Science **83**:90-96.

Thogmartin, W. E., M. G. Knutson, and J. R. Sauer. 2006. Predicting regional abundance of rare grassland birds with a hierarchical spatial count model. Condor **108**:25-46.

Thuiller, W., L. Brotons, M. B. Araujo, and S. Lavorel. 2004. Effects of restricting environmental range of data to project current and future species distributions. Ecography **27**:165-172.

Torres, L. G., P. J. H. Sutton, D. R. Thompson, K. Delord, H. Weimerskirch, P. M. Sagar, E. Sommer, B. J. Dilley, P. G. Ryan, and R. A. Phillips. 2015. Poor transferability of species distribution models for a pelagic predator, the Grey Petrel, indicates contrasting habitat preferences across ocean basins. PLoS ONE **10**:e0120014.

Vaughan, I., and S. Ormerod. 2005. The continuing challenges of testing species distribution models. Ecology **42**:720-730.

Wang, L., and D. A. Jackson. 2014. Shaping up model transferability and generality of species distribution modeling for predicting invasions: implications from a study on Bythotrephes longimanus. Biological Invasions **16**:2079-2103.

Warren, D. L., R. E. Glor, and M. Turelli. 2008. Environmental Niche Equivalency Versus Conservatism: Quantitative Approaches to Niche Evolution. Evolution **62**:2868-2883.

Warren, D. L., R. E. Glor, and M. Turelli. 2010. ENMTools: a toolbox for comparative studies of environmental niche models. Ecography **33**:607-611.

Wenger, S. J., and J. D. Olden. 2012. Assessing transferability of ecological models: an underappreciated aspect of statistical validation. Methods in Ecology and Evolution **3**:260-267.

Werkowska, W., A. L. Márquez, R. Real, and P. Acevedo. 2017. A practical overview of transferability in species distribution modeling. Environmental Reviews **25**:127-133.

Wiens, J. J., and C. H. Graham. 2005. Niche conservatism: integrating evolution, ecology, and conservation biology. Annual Review of Ecology, Evolution, and Systematics:519-539.

Zurell, D., J. Elith, and B. Schröder. 2012. Predicting to new environments: tools for visualizing model behaviour and impacts on mapped distributions. Diversity and Distributions **18**:628-634.

Acknowledgements

AMMS was supported by an IOMRC (UWA/AIMS/CSIRO) collaborative Postdoctoral Fellowship and by ARC grant DE170100841. PJB received support from the Marine Biodiversity Hub within the Australian Government’s National Environmental Science Programme. Thanks to H. Lozano-Montes, J. Meeuwig and E. Peterson for early discussions on the relevance of a synthesis on the transferability of biodiversity models.

Data accessibility

NA.

Box 1. Summary section needed for standardisation of the presentation of results from model transfer studies

1. Type of transferability (as per Fig. 1)
2. Summary of range of responses and predictors (e.g., boxplot)
3. Assumptions made (equilibrium, stationarity, environment versus human drivers, space-for-time substitutability)
4. Models and algorithms used
5. Data scenario: rich, sparse, poor
6. Summary of model fit and performance (including R^2 as a minimum, and SSIM where applicable).

Tables

Table 1: Summary of features reported so far to affect model transferability.

Some features have been reported to clearly improve (green; +) or impair (red; -) model transferability. Others have yielded conflicting results in empirical studies, and their influence is therefore unclear (grey; +/-). References are provided in the Supplementary Information Appendix S1.

Feature	Effect	Explanation	Ref.
Field sampling			
Larger than known geographic extent	+	To guarantee that the tails of a species' response curve are adequately captured.	[5, 76]
Central position within the species range	+	As species thrive in a more varied array of habitats at the centre of their distribution and are more restricted to specialized habitats towards the margins.	[65, 87]
Large sample size	+	To capture variability in habitat conditions and to allow more precise estimates of model parameters.	[61, 88, 93]
Systematic vs. random design	+/-	Systematic sampling may lead to wider environmental coverage and reduce autocorrelation. However, random sampling has desirable statistical properties and avoids spatial bias.	[23, 40]
Uneven or incomplete coverage	-	As biased sampling may add noise to the data, constrain the range of habitat variability being captured, and introduce taxonomic and environmental biases.	[6, 48, 81]
Imperfect detectability	-	May lead to biased samples and misleading inferences of species-environment relationships.	[18, 95]
Model building, calibration and validation			
Model training in more than one region	+	Multi-region models place less emphasis on locally dominant environmental features making them more structurally similar and transferable.	[4, 32, 80]
Correct choice of calibration areas	+	May reduce the rate of false positives.	[21]
Validation with independent data	+	Allows for objective assessment of predictions to novel conditions and of the relative importance and nature of uncertainty sources.	[20, 28, 32, 82]

Accounting for true absences	+	True absences contain valuable information on the conditions that are unfavourable for a given species, and support the modelling of the realized, rather than potential, distributions.	[44]
Choice of model algorithm	+/-	No superior method has yet been identified (refer to text for details).	[22]
Model complexity	+/-	Complex models may overfit data while simpler models often yield better transferability, but more needs to be learned.	[30, 39, 47, 53, 55, 57, 83]
Direction of model transferability test	+/-	Can depend on asymmetries in environments or be species-specific.	[43, 65, 73]
Calibration and projection at different scales (local vs. regional)	+/-	Regional model stability tends to be site and taxon-specific while adjacent areas are more likely to be similar.	[7, 52, 59, 71, 77]
Accounting for autocorrelation	+/-	Reducing autocorrelation may be necessary, but the added autocorrelation term may fail to account for some biological structure affecting transferability.	[14, 35, 69, 74]
Multi-model inference and model averaging	+/-	Consensus forecasts are generally seen as more robust, though recent studies suggest that their extrapolative performance may be species- and context-specific.	[10, 15, 68]
Model misspecification	-	Erroneous choices of error distributions may limit the effectiveness of model transferability.	[47]
Overfitting	-	Can lead to weak correlations among variables and incorrect inferences.	[65, 85]
Data and predictor variables			
Use of direct, functional and ecologically relevant predictors	+	To allow estimation of ecologically sound predictor–response relationships.	[16, 32, 65, 73, 80]
Large training dataset	+	To minimising the variance in parameter estimates.	[72, 84]
Including biotic interactions	+	To better describe species distributions and range margins at broad spatial extents, thereby maximising model transferability.	[2, 8, 31, 36, 41, 62, 79, 89]
Higher quality and resolution of predictors	+	To capture relevant ecological processes and detect local variations more effectively.	[3, 5, 38, 80]

Adequate geographical projection	+	To minimise biases associated with grid-cell sizes (i.e., increased apparent sample size at high latitudes as a degree of longitude decreases in length from the equator to the poles)	[12]
Use of independent remote sensing data	+	To capture complex spatial features where data are and measure environmental variability directly indicating limiting environmental conditions beyond climatic suitability.	[13, 17]
Inclusion of climatic extremes	+	Because they affect demographic processes including growth, regeneration and mortality in some organisms.	[94]
Contingent absences	+	To improve descriptive and predictive potential of models.	[5]
Use of surrogate predictors	-	May fail to express the true habitat requirements of the species, particularly if relationships vary spatially and temporally.	[16, 32, 65, 73, 80]
Multicollinearity	-	Can lead to the misidentification of the most relevant predictors and biases in estimated coefficients and their standard errors.	[33, 45, 78]
Type of scenario in climate projections	-	May lead to over-confidence in single maps of future distributions.	[27, 64]
Niche shifts	-	Can compromise model predictions.	[60]
Species			
Abundance data	+	Can improve detection of the relative suitability of habitats, particularly for species with low prevalence.	[42]
Assemblage composition	+	Enables the detection of general trends that are more immune to the idiosyncrasies of individual taxa.	[46, 63, 71]
Free movement in space	+	May provide better representation of suitable habitats.	[43]
Ecological traits	+/-	Are a species-specific characteristic but can result in greater model transferability.	[1, 20, 26, 37, 54, 65, 75, 92]
Endemism	+/-	Should result in better species-habitat relationships but some studies have found better transferability of models based on non-endemic species.	[20, 43]

Prevalence and rarity	+/-	Common species yield more false positive predictions, whilst more false negatives are expected for rarer taxa, suggesting that intermediate levels of prevalence may be optimal for transferability.	[20, 28, 34, 49, 65, 67]
Eurytopic species	-	May persist in a wide array of conditions that are not easily defined by data, independent variables or model design.	[25, 70, 72]
Dispersal ability	-	Is a source of autocorrelation likely to influence the accuracy of projections across time and space in changing environments.	[14, 20, 90]
Behavioural plasticity	-	Model projections may be misleading if the capacity of a species to adapt to new conditions is ignored.	[58]
Ecotypes	-	Being selectively adapted to localised environmental conditions, model transferability may not be applicable.	[65]
Environment			
Non-analogue conditions	-	May result in the truncation of species response curves under unobserved conditions rendering extrapolating risky.	[9, 24, 29, 50, 56, 66, 76, 77, 91, 96]
Non-stationarity	-	Will impede model transferability due to inconsistent relationships across space or time.	[43, 86]
History of human use	-	May lead to inconsistent model transferability due to different disturbance regimes, management practices, or histories of human use between regions/times.	[11, 19, 51, 53]

Table 2: Metrics used for assessment of model transferability (i.e., external evaluation)

All metrics are used for presence-absence or presence-only (PO) data and result in a score between 0 and 1 (or 0 and 100 %), with the exception of the c-index, which varies between 0.5 and 1. P: probability distribution with x_1 and x_2 representing the response variable predicted by the two models being compared; *cov*: covariance; σ: standard deviation; A and B represent distinct spatial regions; TP, FP, FN and TN refer to each position in a confusion matrix including true positives, false positive, false negative and true negatives, respectively; N: sum of all positives and negatives. Examples of the use of each metric in a model transferability context are given in the last column. For AUC, '*i*' represents a range of cut-off values used to build the plot of sensitivity (SE) versus 1 – specificity (SP).

Metric	Calculation / Definition	References (see Appendix S2)
Presence-only data		
*Modified Hellinger Distance (I similarity statistic)	$I = 1 - \frac{1}{2}\sqrt{\sum_{i}\left(\sqrt{P_{X1}} - \sqrt{P_{X2}}\right)^2}$	[6, 14, 15, 16]
*Schoener's D	$\mathrm{D} = 1 - \frac{1}{2}\sum_{i}\lvert P_{X1} - P_{X2}\rvert$	[15, 17]
Presence-absence data		
Accuracy (overall)	$ACC = \frac{TP + TN}{N}\ x\ 100$	[1]
Area under the curve (AUC) (or concordance c- index)	$AUC = \sum_{i=2}\frac{SE_i + SE_{i-1}}{2}\ x\lvert SP_{i-1} - SP_i\rvert$	[2 – 7]

Coefficient of determination (or Calibration curve; CU)	$R^2 = 1 - \frac{Residual\ sum\ of\ squares}{Total\ sum\ of\ squares}$	[8 – 10]
Cohen's Kappa	$K = \frac{(TP+TN) - \frac{(TP+FN)(TP+FP)+(FP+TN)(FN+TN)}{N}}{N - \frac{(TP+FN)(TP+FP)+(FP+TN)(FN+TN)}{N}}$	[2, 11 – 12]
Correct classification rate	$CCR = \frac{TP+TN}{N}$	[2, 4, 9]
Extrapolative transferability index	$\mathrm{EX-TR} = \frac{AUC_{Extrapolated}}{AUC_{Internal}}$	[12]
Minimal predicted area	$MPA = \frac{Pixels\ above\ specified\ threshold}{N}$	[13]
Percentage of FN or FP	-	[5]
*Point biserial (Pearson correlation)	$\rho = \frac{cov(x,y)}{\sigma_x \sigma_y}$	[10]
Sensitivity	$SE = \frac{TP}{TP+FP}$	[1, 4, 18, 19]
Spearman Rank	$\rho = \frac{cov(x_{ranked}, y_{ranked})}{\sigma_{x_{ranked}} \sigma_{y_{ranked}}}$	[10, 20 – 22]
Specificity	$SP = \frac{TN}{FP+TN}$	[4]
Transferability index	$\mathrm{TRI} = \frac{\frac{1}{2}\left(\left(1 - \frac{\lvert AUC_{A\to A} - AUC_{A\to B}\rvert}{0.5}\right) + \left(1 - \frac{\lvert AUC_{B\to B} - AUC_{B\to A}\rvert}{0.5}\right)\right)}{1 + \left\lvert \left\lvert \frac{AUC_{A\to A} - AUC_{A\to B}}{0.5} \right\rvert - \left\lvert \frac{AUC_{B\to B} - AUC_{B\to A}}{0.5} \right\rvert \right\rvert}$	[11]
Transferability test (one-sided χ^2 test)	$\mathrm{TR} = \frac{\sqrt{\mathrm{N}}\,(\mathrm{TP\ TN} - \mathrm{FP\ FN})}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{FN}+\mathrm{TN})(\mathrm{TP}+\mathrm{FN})(\mathrm{FP}+\mathrm{TN})}}$	[23, 24]

True skill statistic	$\mathrm{TSS} = \frac{TP\,TN - FP\,FN}{(\mathrm{TP} + \mathrm{FN})(\mathrm{FP} + \mathrm{TN})}$	[25]
Abundance data		
Coefficient of determination (or Calibration curve; CU)	$R^2 = 1 - \frac{Residual\ sum\ of\ squares}{Total\ sum\ of\ squares}$	[8 – 10]
Spearman Rank	$\rho = \frac{cov(x_{ranked}, y_{ranked})}{\sigma_{x_{ranked}} \sigma_{y_{ranked}}}$	[10, 20 – 22]
Absolute difference (or root-mean square difference)	$\mathrm{Dif} = \left\lvert y_{target} - y_{reference} \right\rvert$ $\left(\mathrm{RMS} = \sqrt{\frac{1}{n}(x_1^2 + x_2^2 + \cdots + x_n^2)} \right)$	[26]

Figure captions

Figure 1. Examples of range of conditions where model transferability might be tested. ‘Internal’ refers to conditions modelled in the reference system, while ‘External’ to model transfers to a target system (adapted from Mesgaran et al. 2014). The slope of each ellipse reflects the nature of the relationship between the associated variables in the x and y axes. In a) we consider scenarios in which the range of response and predictor variables varies in both the reference and target systems. For example, the range of predictor and response variables might overlap (External 1), the external range of predictors might fall, at least partially, within that of internal (i.e., could be External 2 on ‘a’) but the range of the response variable differ and the relationship might also different (External 2), the range of predictors might overlap in both systems but not the range of responses however response predictor relationship (slope) is similar (External 3), and the range of predictors might differ in both systems but the range of responses is same as is the relationship (slope) between response and predictor variables (External 4). In b) we consider examples concerning the range of values for two environmental predictors used in each system and showing situations where: the external range of predictors lies within the range of internal predictors (External 5), there is some overlap in predictor range (External 6), and no overlap in predictors (External 7).

Figure 2. Temporal trends in published papers related to biodiversity models (lighter colours) and model transferability (darker colours). Bars indicate the cumulative number of peer-reviewed journal articles listed on the ISI Web of Science (webofknowledge.com). Search details included in Supplementary Information.

Figure 3: Summary of possible scenarios and recommended actions when testing model transferability. Transferability refers to the predictive performance of a model developed at a reference location and used to make predictions at a different target location. The transfer possibilities and the actions that can be undertaken depend on the availability of data for the response and predictor variables, and occur or ‘data rich’ situation. For example, in a ‘data sparse’ situation, home ranges for Eurasian Lynx were predicted in Germany based on data from the Swiss Jura Mountains (action 1) and then validated results using telemetry data from Czech Republic and Slovenia (action 2) (image adapted from Schadt et al. (2002), and photo credit to National Geographic). The future geographical distribution of whales sharks (‘data deficient’ situation) was predicted based on existing knowledge for their occurrence and using modelled sea surface temperature for 2070 (action 1) while assuming that relationships between response and predictors will be maintained (action 2) (image adapted from Sequeira et al. (2014), and photo credit to Brian J. Skerry; National Geographic). The distributions of the Iberian desman were predicted comparing results from a reference and target model with validation data (actions 1 and 2, ‘data rich’ situation) before producing a general model combining data from both reference and target locations (action 3) (image adapted from Barbosa et al. (2009) and photo credit to Réseau Education Pyrénées Vivantes).

Figure 4: Application of the SSIM approach in a model transferability context, using the prediction results of a reference model developed for Ningaloo Reef (NR; on the west coast of Australia), and those from a model transferred from the Great Barrier Reef (on the east coast Australia) to NR; adapted from Sequeira *et al.* (2016a). Here we applied the SSIM index to both sets of prediction maps and results show that local means and variances are similar (SIM = 0.992; SIV = 0.691: maps to the left) with resulting maps showing mostly red and pink color indicating high similarity. When using SSIM to explore the pattern of spatial

covariance, we obtained a positive correlation (> 0) in both maps (SIP = 0.122). The overall SSIM obtained thus reflects the combination of the three components (0.992 x 0. 691 x 0.122 = 0.080) resulting in only a slightly positive similarity between the transferred and the native maps for NR.

Figure 1

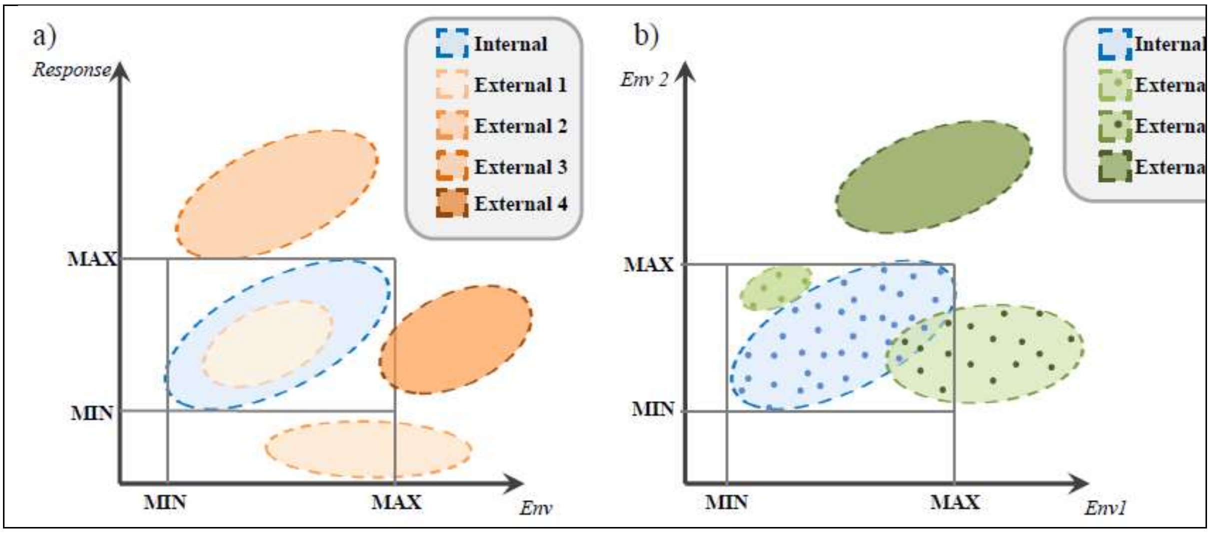

a)
Response
Env
MIN
MAX
MIN
MAX
Internal
External 1
External 2
External 3
External 4
b)
Env 2
Env1
MIN
MAX
MIN
MAX
Internal
Externa
Externa
Externa

Figure 2

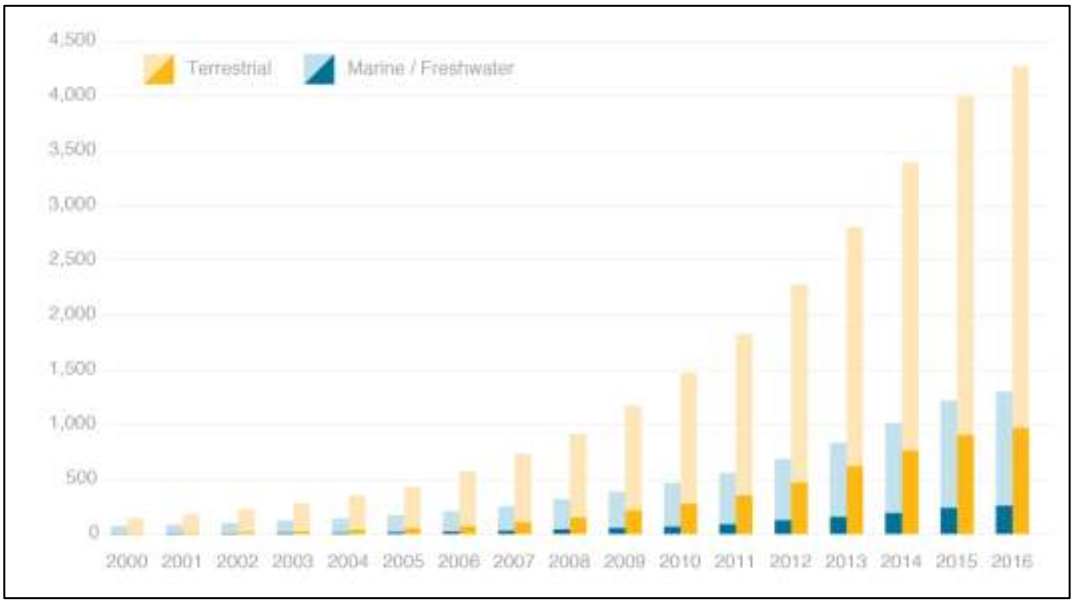
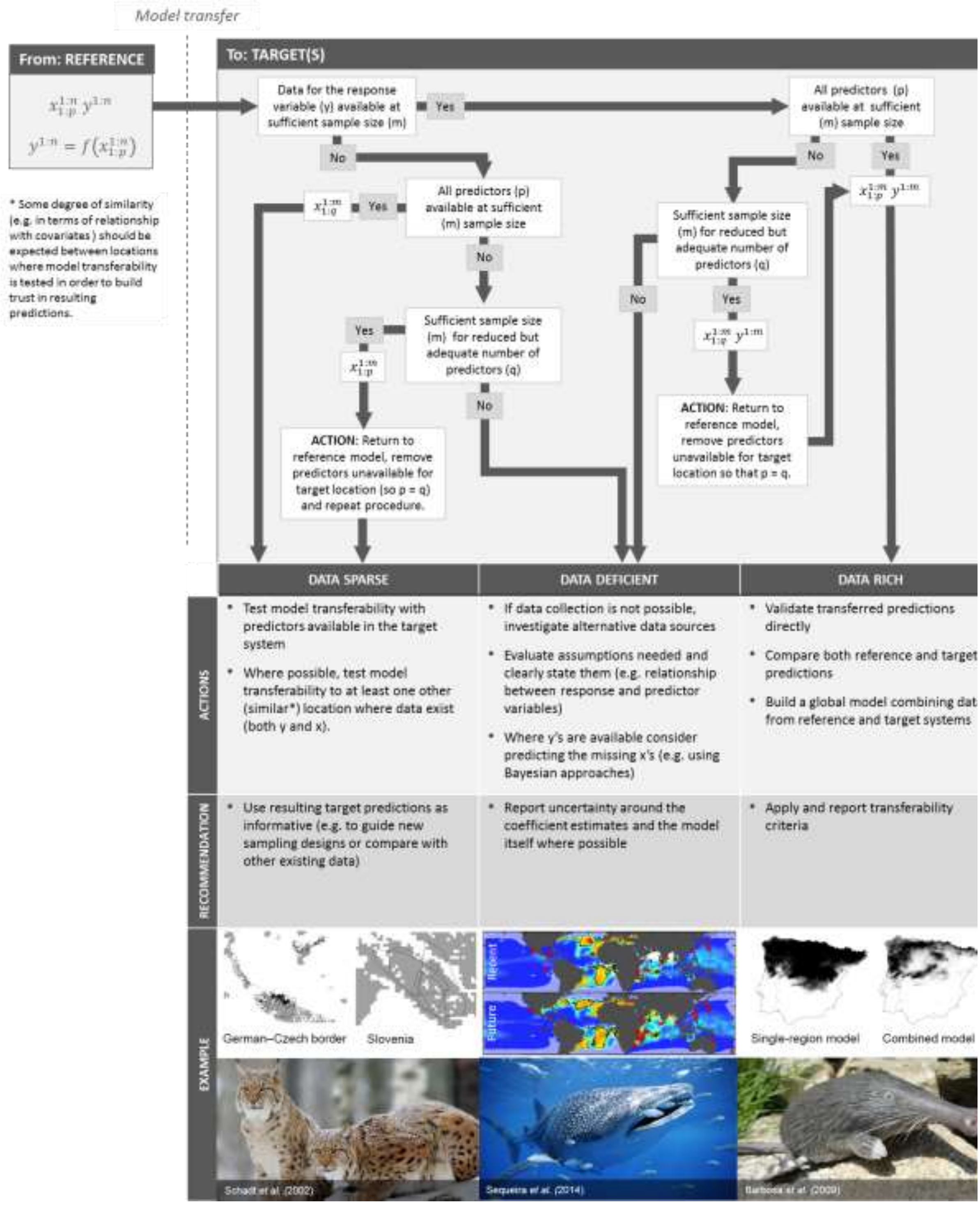

4,500
4,000
3,500
3,000
2,500
2,000
1,500
1,000
500
0
Terrestrial
Marine / Freshwater
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016

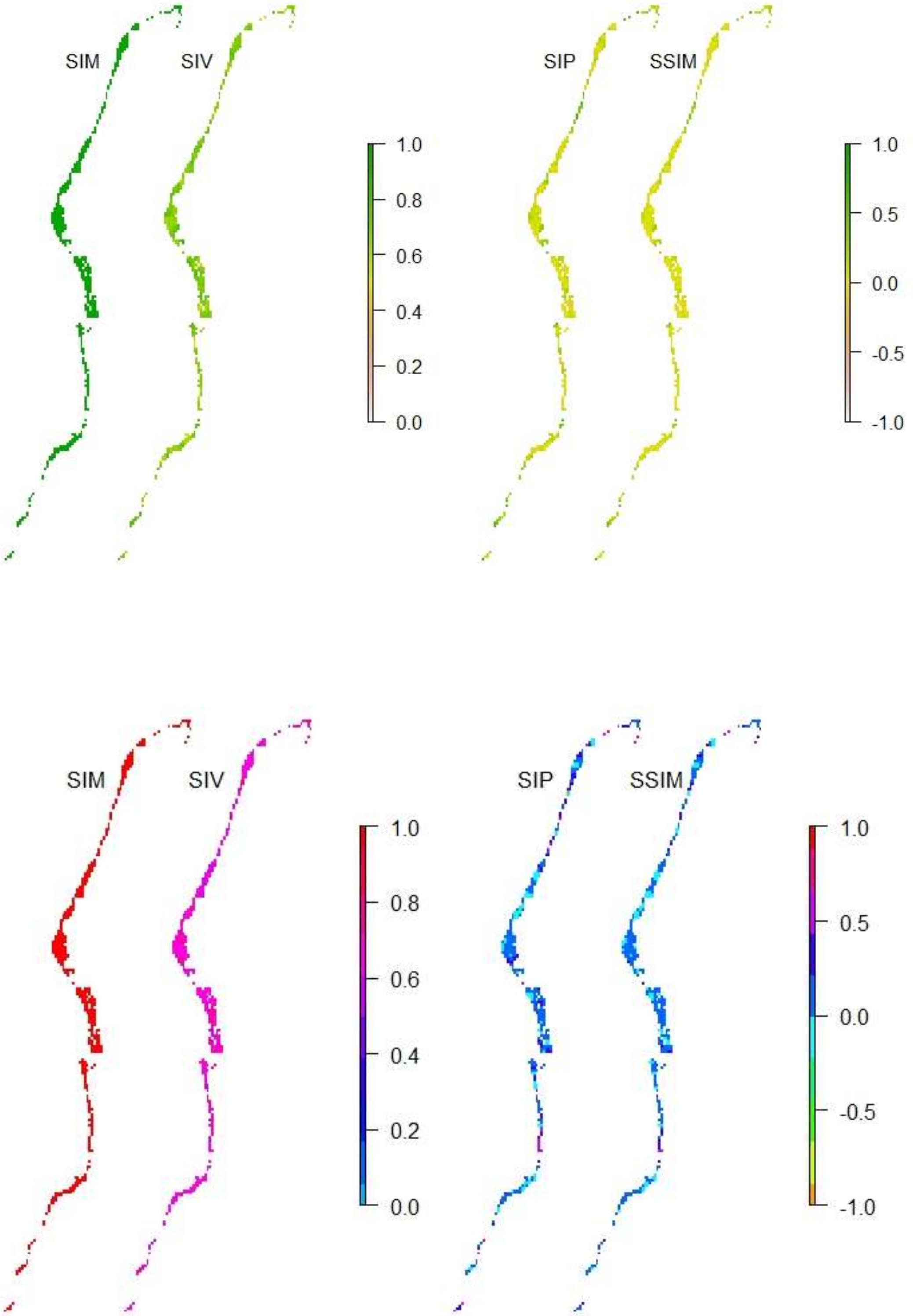
Figure 3



* Some degree of similarity (e.g. in terms of relationship with covariates) should be expected between locations where model transferability is tested in order to build trust in resulting predictions.

	DATA SPARSE	DATA DEFICIENT	DATA RICH
ACTIONS	• Test model transferability with predictors available in the target system • Where possible, test model transferability to at least one other (similar*) location where data exist (both y and x).	• If data collection is not possible, investigate alternative data sources • Evaluate assumptions needed and clearly state them (e.g. relationship between response and predictor variables) • Where y's are available consider predicting the missing x's (e.g. using Bayesian approaches)	• Validate transferred predictions directly • Compare both reference and target predictions • Build a global model combining dat from reference and target systems
RECOMMENDATION	• Use resulting target predictions as informative (e.g. to guide new sampling designs or compare with other existing data)	• Report uncertainty around the coefficient estimates and the model itself where possible	• Apply and report transferability criteria

Figure 4


SIM
SIV
1.0
0.8
0.6
0.4
0.2
0.0
SIP
SSIM
1.0
0.5
0.0
-0.5
-1.0
SIM
SIV
1.0
0.8
0.6
0.4
0.2
0.0
SIP
SSIM
1.0
0.5
0.0
-0.5
-1.0

Glossary

Biodiversity models: family of predictive statistical models commonly known as “species distribution models”, “environmental niche models”, “resource selection functions”, “habitat suitability” or “distribution models” or “climate envelope” or “climate matching models”, and including those operating at a community level (e.g., species richness and total abundance as response variables).

Geographical extrapolation: prediction made for an area beyond the location where calibration data were collected, but where predictor values remain within the range of observed values.

Interpolation: prediction made within the range of parameter values used during model calibration, where the relationships between the response and predictor variables are known (e.g., prediction for the area within two sampled points).

Mathematical extrapolation: prediction made beyond the range of parameter values used during model calibration and beyond known relationships between the response and predictor variables.

Model generality (also known as “generalizability”): a model’s ability to capture species-environment interactions, and therefore, be applied across a large number of systems or conditions (model-focused concept).

Model transferability: a model’s ability to yield accurate and reliable predictions of a given response variable in a new context, be in another geographical area, time period, or both (Elith and Leathwick 2009) (data-focused concept) - commonly and interchangeably referred to in the literature as “hindcasting”, “forecasting” or “back-casting”, “generality” or “generalizability”, “interpolation” or “extrapolation”, and also “projection”.

Predictor: independent variable used to predict the value of a dependent/response variable.

Target: system (location or time) to where a model is transferred.