### Transferring Biodiversity Models for Conservation: Opportunities and Challenges

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Methods in Ecology and Evolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>Draft</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Review</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>n/a</td>
</tr>
</tbody>
</table>
| Complete List of Authors: | Sequeira, Ana; The University of Western Australia, School of Biological Sciences and The Oceans Institute  
Bouchet, Phillipe; University of Western Australia, Centre for Marine Futures and School of Biological Sciences  
Yates, Katherine; University of Salford, School of Environment and Life Sciences  
Mengersen, Kerrie; Queensland University of Technology, School of Mathematics and Statistics; ARC Centre of Excellence for Mathematics and Statistics  
Caley, M. Julian; Queensland University of Technology; ARC Centre of Excellence for Mathematics and Statistics |
| Keywords: | Statistics, Applied Ecology, Conservation, Modelling (Community Ecology) < Community Ecology |

**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.
Transferring Biodiversity Models for Conservation: Opportunities and Challenges

AMM Sequeira¹, PJ Bouchet², KL Yates³, K Mengersen⁴, MJ Caley⁵

Running title: Roadmap for improved model transfers in ecology
Word count: 8514

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

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

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

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

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

⁶Australian 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 very large, 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). 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). 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 (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), two situations can occur: (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 systems 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. We suggest therefore that the form of the transferability test being performed always should 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). 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,
cautions 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 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). 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. 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. 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. 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. Another common assumption is space-
for-time substitutability, 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). 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) 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. 2016).

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), 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). 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. 2016), 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 (Sequeira et al. 2016). 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. 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 distributions of different species groups. SSIM provides an overall
measure of similarity between two sets of spatial observations (i.e. maps) ranging between -1
and 1, and could be used to test model predictive power through comparison of predictions
obtained from a model developed for the target system. SSIM also provides information
about where dissimilarities might be greater by comparing means, variances, and covariances,
which can provide information about where and why a model fails 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 (REF?). 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 in 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 which 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 the 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. 2011), even though predictions obtained under previously unobserved conditions are prone to both statistical and ecological error (Dormann 2007). Despite 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 despite known risks (Fitzpatrick and Hargrove 2009). Because the consideration of such risks is central to effective decision-making, the development of techniques that can diagnose or address the consequences of extrapolation when transferring models is a research priority (Araujo and Guisan 2006).

Various strategies have been suggested for dealing with extrapolation including 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. If predictions are not made beyond $1/10^{th}$ of the sampled parameter range, extrapolation might be considered negligible (Dormann 2007). Extrapolation effects can also be mitigated 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. Examples include the Multivariate Environmental Similarity Surface (MESS) within MaxEnt, which relies only on a rectilinear (univariate) detection of extrapolation. Other techniques such as ‘extrapolation detection’, ‘environmental overlap mask’, ‘envelope uncertainty maps’, ‘prediction uncertainty assessments using residual variation’, and ‘mobility-oriented parity’ are also capable of highlighting different types of environmental novelty arising from mathematical and combinatorial extrapolation.

Ultimately, the ‘monsters’ that lurk in the waters of extrapolation will be less ferocious if the uncertainties underlying model projections are not ignored. 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. 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. 2016). 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-studies comparison 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 sections 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 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 the strategy to use 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 that can be used to calculate common metrics, including R2 (as a minimum).
   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 our understanding increases, our recommendations may evolve but the need for standardisation will remain. Lastly, because model transferability is relevant in other fields, such as transport (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


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.


lessons from the land present challenges and opportunities. Global Ecology and
Biogeography 20:7889-7802.

Schadt, S., E. Revilla, T. Wiegand, F. Knauer, P. Kaczensky, U. Breitenmoser, L. Bufka, J.
suitability of central European landscapes for the reintroduction of Eurasian lynx.

Predicting current and future global distributions of whale sharks. Global Change


of predictive fish distribution models in two coastal systems. Estuarine, Coastal and
Shelf Science 83:90-96.

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


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).
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.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Effect</th>
<th>Explanation</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field sampling</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Larger than known geographic extent</td>
<td>+</td>
<td>To guarantee that the tails of a species' response curve are adequately captured.</td>
<td>[5, 76]</td>
</tr>
<tr>
<td>Central position within the species range</td>
<td>+</td>
<td>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.</td>
<td>[65, 87]</td>
</tr>
<tr>
<td>Large sample size</td>
<td>+</td>
<td>To capture variability in habitat conditions and to allow more precise estimates of model parameters.</td>
<td>[61, 88, 93]</td>
</tr>
<tr>
<td>Systematic vs. random design</td>
<td>+/-</td>
<td>Systematic sampling may lead to wider environmental coverage and reduce autocorrelation. However, random sampling has desirable statistical properties and avoids spatial bias.</td>
<td>[23, 40]</td>
</tr>
<tr>
<td>Uneven or incomplete coverage</td>
<td>-</td>
<td>As biased sampling may add noise to the data, constrain the range of habitat variability being captured, and introduce taxonomic and environmental biases.</td>
<td>[6, 48, 81]</td>
</tr>
<tr>
<td>Imperfect detectability</td>
<td>-</td>
<td>May lead to biased samples and misleading inferences of species-environment relationships.</td>
<td>[18, 95]</td>
</tr>
<tr>
<td>Model building, calibration and validation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model training in more than one region</td>
<td>+</td>
<td>Multi-region models place less emphasis on locally dominant environmental features making them more structurally similar and transferable.</td>
<td>[4, 32, 80]</td>
</tr>
<tr>
<td>Correct choice of calibration areas</td>
<td>+</td>
<td>May reduce the rate of false positives.</td>
<td>[21]</td>
</tr>
<tr>
<td>Validation with independent data</td>
<td>+</td>
<td>Allows for objective assessment of predictions to novel conditions and of the relative importance and nature of uncertainty sources.</td>
<td>[20, 28, 32, 82]</td>
</tr>
<tr>
<td>Accounting for true absences</td>
<td>+</td>
<td>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]</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Choice of model algorithm</td>
<td>+/-</td>
<td>No superior method has yet been identified (refer to text for details). [22]</td>
<td></td>
</tr>
<tr>
<td>Model complexity</td>
<td>+/-</td>
<td>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]</td>
<td></td>
</tr>
<tr>
<td>Direction of model transferability test</td>
<td>+/-</td>
<td>Can depend on asymmetries in environments or be species-specific. [43, 65, 73]</td>
<td></td>
</tr>
<tr>
<td>Calibration and projection at different scales (local vs. regional)</td>
<td>+/-</td>
<td>Regional model stability tends to be site and taxon-specific while adjacent areas are more likely to be similar. [7, 52, 59, 71, 77]</td>
<td></td>
</tr>
<tr>
<td>Accounting for autocorrelation</td>
<td>+/-</td>
<td>Reducing autocorrelation may be necessary, but the added autocorrelation term may fail to account for some biological structure affecting transferability. [14, 35, 69, 74]</td>
<td></td>
</tr>
<tr>
<td>Multi-model inference and model averaging</td>
<td>+/-</td>
<td>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]</td>
<td></td>
</tr>
<tr>
<td>Model misspecification</td>
<td>-</td>
<td>Erroneous choices of error distributions may limit the effectiveness of model transferability. [47]</td>
<td></td>
</tr>
<tr>
<td>Overfitting</td>
<td>-</td>
<td>Can lead to weak correlations among variables and incorrect inferences. [65, 85]</td>
<td></td>
</tr>
</tbody>
</table>

**Data and predictor variables**

<table>
<thead>
<tr>
<th>Use of direct, functional and ecologically relevant predictors</th>
<th>+</th>
<th>To allow estimation of ecologically sound predictor–response relationships. [16, 32, 65, 73, 80]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large training dataset</td>
<td>+</td>
<td>To minimising the variance in parameter estimates. [72, 84]</td>
</tr>
<tr>
<td>Including biotic interactions</td>
<td>+</td>
<td>To better describe species distributions and range margins at broad spatial extents, thereby maximising model transferability. [2, 8, 31, 36, 41, 62, 79, 89]</td>
</tr>
<tr>
<td>Higher quality and resolution of predictors</td>
<td>+</td>
<td>To capture relevant ecological processes and detect local variations more effectively. [3, 5, 38, 80]</td>
</tr>
<tr>
<td>Adequate geographical projection</td>
<td>+</td>
<td>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)</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>---</td>
<td>---------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Use of independent remote sensing data</td>
<td>+</td>
<td>To capture complex spatial features where data are and measure environmental variability directly indicating limiting environmental conditions beyond climatic suitability.</td>
</tr>
<tr>
<td>Inclusion of climatic extremes</td>
<td>+</td>
<td>Because they affect demographic processes including growth, regeneration and mortality in some organisms.</td>
</tr>
<tr>
<td>Contingent absences</td>
<td>+</td>
<td>To improve descriptive and predictive potential of models.</td>
</tr>
<tr>
<td>Use of surrogate predictors</td>
<td>-</td>
<td>May fail to express the true habitat requirements of the species, particularly if relationships vary spatially and temporally.</td>
</tr>
<tr>
<td>Multicollinearity</td>
<td>-</td>
<td>Can lead to the misidentification of the most relevant predictors and biases in estimated coefficients and their standard errors.</td>
</tr>
<tr>
<td>Type of scenario in climate projections</td>
<td>-</td>
<td>May lead to over-confidence in single maps of future distributions.</td>
</tr>
<tr>
<td>Niche shifts</td>
<td>-</td>
<td>Can compromise model predictions.</td>
</tr>
<tr>
<td>Species</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abundance data</td>
<td>+</td>
<td>Can improve detection of the relative suitability of habitats, particularly for species with low prevalence.</td>
</tr>
<tr>
<td>Assemblage composition</td>
<td>+</td>
<td>Enables the detection of general trends that are more immune to the idiosyncrasies of individual taxa.</td>
</tr>
<tr>
<td>Free movement in space</td>
<td>+</td>
<td>May provide better representation of suitable habitats.</td>
</tr>
<tr>
<td>Ecological traits</td>
<td>+/-</td>
<td>Are a species-specific characteristic but can result in greater model transferability.</td>
</tr>
<tr>
<td>Endemism</td>
<td>+/-</td>
<td>Should result in better species-habitat relationships but some studies have found better transferability of models based on non-endemic species.</td>
</tr>
<tr>
<td><strong>Prevalence and rarity</strong></td>
<td>+/-</td>
<td>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.</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-----</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Eurytopic species</strong></td>
<td>-</td>
<td>May persist in a wide array of conditions that are not easily defined by data, independent variables or model design.</td>
</tr>
<tr>
<td><strong>Dispersal ability</strong></td>
<td>-</td>
<td>Is a source of autocorrelation likely to influence the accuracy of projections across time and space in changing environments.</td>
</tr>
<tr>
<td><strong>Behavioural plasticity</strong></td>
<td>-</td>
<td>Model projections may be misleading if the capacity of a species to adapt to new conditions is ignored.</td>
</tr>
<tr>
<td><strong>Ecotypes</strong></td>
<td>-</td>
<td>Being selectively adapted to localised environmental conditions, model transferability may not be applicable.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Environment</strong></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-analogue conditions</strong></td>
<td>-</td>
<td>May result in the truncation of species response curves under unobserved conditions rendering extrapolating risky.</td>
<td>[9, 24, 29, 50, 56, 66, 76, 77, 91, 96]</td>
</tr>
<tr>
<td><strong>Non-stationarity</strong></td>
<td>-</td>
<td>Will impede model transferability due to inconsistent relationships across space or time.</td>
<td>[43, 86]</td>
</tr>
<tr>
<td><strong>History of human use</strong></td>
<td>-</td>
<td>May lead to inconsistent model transferability due to different disturbance regimes, management practices, or histories of human use between regions/times.</td>
<td>[11, 19, 51, 53]</td>
</tr>
</tbody>
</table>
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; \(\text{cov}\): covariance; \(\sigma\): 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).

<table>
<thead>
<tr>
<th>Metric</th>
<th>Calculation / Definition</th>
<th>References (see Appendix S2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Presence-only data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*Modified Hellinger Distance (I similarity statistic)</td>
<td>(I = 1 - \frac{1}{2} \sqrt{\sum \left(\sqrt{P_{x_1}} - \sqrt{P_{x_2}}\right)^2} )</td>
<td>[6, 14, 15, 16]</td>
</tr>
<tr>
<td>*Schoener's D</td>
<td>(D = 1 - \frac{1}{2} \sum</td>
<td>P_{x_1} - P_{x_2}</td>
</tr>
<tr>
<td><strong>Presence-absence data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy (overall)</td>
<td>(ACC = \frac{TP + TN}{N} \times 100)</td>
<td>[1]</td>
</tr>
<tr>
<td>Area under the curve (AUC) (or concordance c-index)</td>
<td>(AUC = \sum_{i=2}^{SE_i + SE_{i-1}} \frac{SE_i + SE_{i-1}}{2} x</td>
<td>SP_{i-1} - SP_i</td>
</tr>
<tr>
<td>Clinical term</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>Coefficient of determination</td>
<td>( R^2 = 1 - \frac{\text{Residual sum of squares}}{\text{Total sum of squares}} ) [8–10]</td>
<td></td>
</tr>
<tr>
<td><strong>Cohen's Kappa</strong></td>
<td>( K = \frac{(TP + TN) - (TP + FN)(TP + FP) + (FP + TN)(FN + TN)}{N - (TP + FN)(TP + FP) + (FP + TN)(FN + TN)} ) [2, 11–12]</td>
<td></td>
</tr>
<tr>
<td><strong>Correct classification rate</strong></td>
<td>( CCR = \frac{TP + TN}{N} ) [2, 4, 9]</td>
<td></td>
</tr>
<tr>
<td><strong>Extrapolative transferability index</strong></td>
<td>( \text{EX} - \text{TR} = \frac{AUC_{\text{Extrapolated}}}{AUC_{\text{Internal}}} ) [12]</td>
<td></td>
</tr>
<tr>
<td><strong>Minimal predicted area</strong></td>
<td>( MPA = \frac{\Pi \text{ above specified threshold}}{N} ) [13]</td>
<td></td>
</tr>
<tr>
<td><strong>Percentage of FN or FP</strong></td>
<td>- [5]</td>
<td></td>
</tr>
<tr>
<td><em>Point biserial (Pearson correlation)</em></td>
<td>( \rho = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} ) [10]</td>
<td></td>
</tr>
<tr>
<td><strong>Sensitivity</strong></td>
<td>( SE = \frac{TP}{TP + FN} ) [1, 4, 18, 19]</td>
<td></td>
</tr>
<tr>
<td><strong>Spearman Rank</strong></td>
<td>( \rho = \frac{\text{cov}(x_{\text{ranked}}, y_{\text{ranked}})}{\sigma_{x_{\text{ranked}}} \sigma_{y_{\text{ranked}}}} ) [10, 20–22]</td>
<td></td>
</tr>
<tr>
<td><strong>Specificity</strong></td>
<td>( SP = \frac{TN}{FP + TN} ) [4]</td>
<td></td>
</tr>
<tr>
<td><strong>Transferability index</strong></td>
<td>( \text{TRI} = \frac{1}{2} \left( \left( 1 - \frac{</td>
<td>AUC_{A \rightarrow A} - AUC_{A \rightarrow B}</td>
</tr>
<tr>
<td><strong>Transferability test (one-sided ( \chi^2 ) test)</strong></td>
<td>( \text{TR} = \frac{\sqrt{N} (TP \cdot TN - FP \cdot FN)}{\sqrt{(TP + FP)(FN + TN)(TP + FN)(FP + TN)}} ) [23, 24]</td>
<td></td>
</tr>
<tr>
<td>Parameter</td>
<td>Formula</td>
<td>Reference</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>True skill statistic</td>
<td>$TSS = \frac{TP \cdot TN - FP \cdot FN}{(TP + FN)(FP + TN)}$</td>
<td>[25]</td>
</tr>
<tr>
<td><strong>Abundance data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coefficient of determination</td>
<td>$R^2 = 1 - \frac{\text{Residual sum of squares}}{\text{Total sum of squares}}$</td>
<td>[8 – 10]</td>
</tr>
<tr>
<td>(or Calibration curve; CU)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spearman Rank</td>
<td>$\rho = \frac{\text{cov}(x_{\text{ranked}}, y_{\text{ranked}})}{\sigma_{x_{\text{ranked}}} \sigma_{y_{\text{ranked}}}}$</td>
<td>[10, 20 – 22]</td>
</tr>
<tr>
<td>Absolute difference (or root-</td>
<td>$\text{Dif} = \left</td>
<td>y_{\text{target}} - y_{\text{reference}} \right</td>
</tr>
<tr>
<td>mean square difference)</td>
<td>$\text{RMS} = \sqrt{\frac{1}{n}(x_1^2 + x_2^2 + \cdots + x_n^2)}$</td>
<td>[26]</td>
</tr>
</tbody>
</table>
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 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 adapted from Sequeira et al. (2016). Here we show that local means and variances between the predictions obtained for NR from the transferred GBR and the NR predictions derived from the NR model are similar (SIM = 0.992; SIV = 0.691). The pattern of spatial covariance is also positively correlated 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
Figure 2
Figure 3

Model transfer

From: REFERENCE

\[ x_{1:n}^i \rightarrow y_{1:n}^i \]
\[ y_{1:n}^i = f(x_{1:n}^i) \]

\* Same degree of similarity (e.g. in terms of relationship with covariates) should be expected between locations where model transferability is tested to ensure successful predictions.

To: TARGET(S)

Data for the response variable (y) available at sufficient sample size (m)

Yes

All predictors (p) available at sufficient (m) sample size

Yes

ACTION: Return to reference model, remove predictors unavailable for target location (x, y) and repeat procedure.

No

No

ACTION: Return to reference model, remove predictors unavailable for target location so that p = q.

No

No

Sufficient sample size (m) for reduced number of predictors (q)

Yes

DATA SPARSE | DATA DEFICIENT | DATA RICH

\* 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 data from reference and target systems

\* 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

EXAMPLE

German–Czech border  Slovenia

Single-region model  Combined model

Figure 4
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.