Review

Outstanding Challenges in the Transferability of Ecological Models


Predictive models are central to many scientific disciplines and vital for informing management in a rapidly changing world. However, limited understanding of the accuracy and precision of models transferred to novel conditions (their ‘transferability’) undermines confidence in their predictions. Here, 50 experts identified priority knowledge gaps which, if filled, will most improve model transfers. These are summarized into six technical and six fundamental challenges, which underline the combined need to intensify research on the determinants of ecological predictability, including species traits and data quality, and develop best practices for transferring models. Of high importance is the identification of a widely applicable set of transferability metrics, with appropriate tools to quantify the sources and impacts of prediction uncertainty under novel conditions.

Predicting the Unknown

Predictions facilitate the formulation of quantitative, testable hypotheses that can be refined and validated empirically [1]. Predictive models have thus become ubiquitous in numerous scientific disciplines, including ecology [2], where they provide means for mapping species distributions, explaining population trends, or quantifying the risks of biological invasions and disease outbreaks (e.g., [3,4]). The practical value of predictive models in supporting policy and decision making has therefore grown rapidly (Box 1) [5]. With that has come an increasing desire to predict (see Glossary) the state of ecological features (e.g., species, habitats) and our likely impacts upon them [5], prompting a shift from explanatory models to anticipatory predictions [2]. However, in many situations, severe data deficiencies preclude the development of specific models, and the collection of new data can be prohibitively costly or simply impossible [6]. It is in this context that interest in transferable models (i.e., those that can be legitimately projected beyond the spatial and temporal bounds of their underlying data [7]) has grown.

Transferred models must balance the tradeoff between estimation and prediction bias and variance (homogenization versus nontransferability, sensu [8]). Ultimately, models that can

Highlights

Models transferred to novel conditions could provide predictions in data-poor scenarios, contributing to more informed management decisions.

The determinants of ecological predictability are, however, still insufficiently understood.

Predictions from transferred ecological models are affected by species’ traits, sampling biases, biotic interactions, nonstationarity, and the degree of environmental dissimilarity between reference and target systems.

We synthesize six technical and six fundamental challenges that, if resolved, will catalyze practical and conceptual advances in model transfers.

We propose that the most immediate obstacle to improving understanding lies in the absence of a widely applicable set of metrics for assessing transferability, and that encouraging the development of models grounded in well-established mechanisms offers the most immediate way of improving transferability.
Box 1. Why Transfer Models in the First Place?

Ecological models are extensively and increasingly used in support of environmental policy and decision making [77]. The process of transferring models typically stems from the need to support resource management in the face of pervasive data deficiencies, limited research funding, and accelerating global change [6]. Spatial transfers have been used to guide the design of protected areas, search for species on the brink of extinction, inform species relocations or reintroductions, outline hotspots of invasive pests, design field sampling campaigns, and assist the regulation of human activities (e.g., [78,79]). For instance, projections of Asian tiger mosquito (Aedes albopictus) distribution models onto all continents helped identify areas at greatest risk of invasion, with important implications for human health [81]. Temporal transfers have largely been applied to forecast species’ responses to climate warming, retrospectively describe pristine population states, characterize evolutionary patterns of speciation, quantify the repercussions of land use changes, or estimate future ecosystem dynamics (e.g., [88,72,82]). Despite being difficult to quantify, the societal and economic gains from transferring models can be substantial, and are most readily illustrated by the mitigation of costs associated with invasive species [83]. For instance, the establishment of the zebra mussel (Dreissena polymorpha) in the Great Lakes region of North America has led to $20–100 million in annual mitigation expenditure, with additional, unquantified nonmarket costs ensuing from the loss of biodiversity and ecosystem services [8]. Transferred models accurately predicted the establishment of the zebra mussel 5 years before it was actually discovered in the region, however model predictions were not used to take preventative action, illustrating that developing a transferable model is only the start of the road to informing decision makers (see Outstanding Questions). Ultimately, the widespread need to make proactive management decisions in data-poor situations drives the need to improve our understanding of model transferability. This goal fundamentally requires better transferability metrics and estimates of prediction uncertainty, which can assist in selecting the most consistent and effective management options while avoiding unanticipated outcomes [84].

simultaneously achieve high accuracy and precision, even when predicting into novel contexts, will provide maximum utility for decision making [9]. To date, however, tests of transferability across taxa and geographic locations have failed to demonstrate consistent patterns (Figure 1), and a general approach to developing transferable models remains elusive (but see [6,10]). Here, we outline challenges that, if addressed, will improve the harmonization, uptake, and application of model transfers in ecology. We argue that moving the field of model transferability forward requires a two-pronged approach focused on: (i) investing in fundamental research aimed at enhancing predictability, and (ii) establishing technical standards for assessing transferability.

Defining the Challenges

We first identified challenges using a modified Delphi technique [11] (see the supplementary information online), and then divided them into those that reflected conceptual obstacles (“fundamental challenges”), and those related to best practices (“technical challenges”). Acknowledging significant overlap and linkages between these challenges (Figure 2), we explore each separately below. Attempts to understand and enhance transferability face many of the same hurdles as ecological modeling generally (e.g., data quality, stochasticity), and adhering to best practice recommendations (e.g., [12,13]) is thus imperative. We do not focus on these well-established standards, but concentrate on the additional challenges posed by transferring models. Whilst spatial transferability studies retain prominence in the literature (and thus in this manuscript), this is not an indication of relative importance, but rather a reflection of the inherent difficulties in evaluating models transferred through time. Our review of published studies is not exhaustive, and the online supplementary information provides additional literature relevant to each challenge.

Fundamental Challenges

Is Model Transferability Trait- or Taxon-Specific?

Knowing whether models are more transferable for some taxonomic groups would be useful to increase confidence in predictions and prioritize resources for model development (Box 1).
Evidence indicates discrepancies in model performance among taxa with divergent life-history traits, and populations with different age structures and sex ratios (e.g., [14]). Meta-analyses demonstrate that body size and trophic position are strong indicators of ecological predictability [15], with some studies also indicating greater hurdles in building transferable models for wide-ranging organisms with broad environmental niches than for narrow-ranging specialists [16]. For example, model transfers for butterflies were less accurate in species with long flight seasons [17]. By contrast, models of vascular plants with higher dispersal ability exhibited better transferability than those built for endemics with limited dispersal capacity [18]. Developing transferable models for species with greater behavioral or adaptive plasticity might also be more difficult, regardless of spatial range size [8]. Subsetting movement and observational data by behavioral state (e.g., foraging versus breeding) or group composition (e.g., presence of mother–young pairs) prior to model calibration might improve model performance and transferability.

Which Response Variables Make Models More or Less Transferable?

The superior information content inherent to abundance data should facilitate greater transferability than models of occurrence built from presence–absence or presence-only data, so that models of abundance might better project the ecological impacts of global change [19]. While this has been shown for some birds [19], fitting abundance models remains difficult for most taxa [20], not least because counting individuals is more challenging than recording presence–absence (despite issues caused by imperfect detectability). Accordingly, interest has grown in comparing the predictions obtained from occurrence and abundance models, and testing the reliability of the former as a surrogate for the latter [21]. In general, stronger correlations between abundance and occurrence are expected for rare organisms. However, the strength of this relationship can be nonlinear, species-specific, and conditional on spatial behavior, social organization, life-history strategies, population density, resource availability, and biotic interactions [22]. Most studies have also applied model transfers to single species. Community- and ecosystem-level models that fit shared environmental responses for multiple species simultaneously could achieve higher transferability [23], but this potential has been inconsistently demonstrated. Integrated models that unite presence-only and presence–absence data [24], and those that combine occupancy probabilities (e.g., derived from regional monitoring) with density-given-occupancy (e.g., derived from telemetry), offer further promise [25]. The former provide more accurate predictions than models based on a single data type, whereas the latter can account for suitable but unoccupied habitats.

To What Extent Does Data Quality Influence Model Transferability?

More accurate and/or precise data should result in better transfers on theoretical grounds, with evidence showing that the accuracy of species records can be more important for transferability than their spatial extent [26]. Data of unverifiable quality (e.g., anecdotal reports of easily misidentified species) should therefore be avoided, even if available over broader geographical areas. Model transfers can be further hampered by imperfect detectability, spatial and temporal biases in data collection, insufficient sample sizes, the omission of known drivers, or the use of proxy variables [27]. Additionally, species’ characteristics such as range size can impact positional accuracy, leading to erroneous predictions if analyses are conducted at scales corresponding with those of the original locational errors [28]. The magnitude of these effects is ultimately unclear, and data quality therefore represents a substantial source of uncertainty [29]. Simulation studies based on virtual species with known reference information represent a critical resource in tackling this knowledge gap.
How Can Sampling Be Optimized to Maximize Model Transferability?
Samples encompassing the full range of environmental conditions and their possible combinations should avoid incomplete niche characterization and improve transferability [Box 2]. However, data are often collected opportunistically and pooled during analysis, such that model building ought to account for unevenly sampling in environmental space (e.g., by including random effects, or through explicit balancing methods that capture the intensity and distribution of sampling effort [30]). Importantly, data resolution influences model fit, prediction, and by extension, transferability. For example, poorly resolved predictors might not capture important aspects of a species’ ecology, and relate only indirectly to observed patterns of occurrence and biogeography [31,32]. Where possible, the scale(s) over which the processes of interest operate should therefore drive predictor choice, with sensitivity tests advisable [31]. As habitat availability, and thus perceived preference, also often link to scale [33], models will be sensitive to the extent of the study region, especially for fragmented habitats and steep environmental gradients [8]. As such, combining geographically and environmentally distinct regions ought to increase model transferability [34]. Temporal replication in sampling can also help by capturing natural variability and stochastic processes, as well as alleviating imperfect detectability and false negative rates. When resources are limited, sampling should ideally focus on designs that address existing data limitations and maximize information gain.

How Does Model Complexity Influence Model Transferability?
Excessively complex models risk overfitting training data and can erroneously attribute patterns to sampling or environmental noise [35], leading to predictions that are biased or too specific to the reference system to be transferable [36]. Greater transferability is thus generally expected in parsimonious models with smooth univariate response curves and few predictors [37]. However, while simple models have been shown to lead to better transferability, they can also yield misleading predictions when transferred to new contexts, implying that simplicity is not always beneficial [38,39]. Ultimately, simple and complex models serve different purposes [40], and in some instances, a preference for accurate and precise predictions over ecological interpretability might be justifiable, making complex models more appropriate [41]. Complex models are also not necessarily more arduous to interpret, and can be valuable for discovering hidden, unexpected patterns [40]. Additionally, they could be useful in exploring nonlinear and dynamic associations of species with indirect predictors across landscapes, seasons, or years [40], to help better accommodate nonstationarity. That said, as complexity grows, so do potential predictor combinations and the likelihood of mismatch between reference and target conditions, which can result in incorrect interpolation and extrapolation [42]. Species’ life-history traits, physiology, or behavior can also influence complexity, such that choosing an optimally complex model requires identifying the most sensible predictors and datasets relative to a given study objective. Novel indices of complexity that emphasize the structural properties of the input data might help [43], as could standardized metrics of predictive performance.

Are There Spatial and Temporal Limits to Extrapolation in Model Transfers?
While prediction error is expected to increase with ‘distance’ (e.g., km, days) from reference conditions [1], model transferability appears little related to geographic (and temporal) separation between reference systems and target systems (Figure 1). Instead, environmental dissimilarity is what matters most for successful transfers, for which spatio-temporal distances might only occasionally be good surrogates. The minimum level of similarity required to support transferable models, however, remains unknown. Some authors caution against seeking inference beyond one-tenth of the sampled covariate range, yet this rule of
thumb [44] does not translate into practical and comprehensible guidelines for model end-users (e.g., spatial planners, resource managers). Another solution could lie in the ‘forecast horizon’, which defines the point beyond which sufficiently useful predictions can no longer be made in any given dimension (e.g., space, time, phylogeny, environment) [45]. Calculating this horizon requires choosing a measure of prediction quality (i.e., a function of accuracy and precision), and a proficiency threshold for ‘acceptable’ predictions [45]. Both choices can be framed in decision theory and informed through stakeholder participation, making the forecast horizon a flexible and policy-relevant instrument for assessing and communicating ecological predictability.

Technical Challenges

How Can Non-analog Conditions Be Accounted for When Transferring Models?

Transferring models into non-analogue environments brings numerous and well-documented perils [46], but the predictive performance of models transferred into novel conditions is rarely tested explicitly [47]. Different techniques to account for non-analog conditions will likely be required depending on the degree of environmental dissimilarity (i.e., novel conditions just beyond those observed versus those that are extremely dissimilar). Several tools are available to visualize regions whose characteristics depart from the initial covariate range (e.g., [42,48]), and these can help assess the potential impacts of non-analog conditions on predictive performance. However, these tools cannot predict species’ responses to novel conditions, which can be particularly unexpected if environmental change imposes selection pressures that disrupt biotic interactions and cause communities to evolve [49]. Further development of these tools for future transfers, and their application in examining the outcomes of historical transfers, will improve our understanding on how non-analog conditions can be accounted for when transferring models.

How Can Nonstationarity and Interactions Be Incorporated in Model Transfers?

Successful transfers rely on the inherent premise that species–environment relationships are stationary at the calibration site and remain so beyond it. However, species’ responses to the environment are rarely static, and can vary nonlinearly with resource availability, species ontology, and population density [50]. Species–environment relationships are therefore context-specific, and habitat occupation ultimately depends on relative habitat availability [33]. Moreover, anthropogenic activities can strongly influence species’ distribution and abundance patterns, and are themselves variable [51]. Disentangling their effects from environmentally driven covariance is difficult, especially when histories of human exposure are unknown, or the magnitude of impacts unobservable. Recent studies have also reconciled transferability with strong evidence for the role of biotic interactions in shaping species’ ranges at large spatial scales [52], offering a blueprint for determining when biotic information can support predictions under unobserved conditions. Methods that incorporate functional responses have now progressed to combine data from different regions and use nonstationary model coefficients, enabling enhanced transferability [8,53]. We expect further improvements in knowledge will be made by encouraging the development of models grounded in well-described mechanisms (Box 3).

Do Specific Modeling Approaches Result in Better Transferability?

Studies have benchmarked the predictive capacity and transferability of existing algorithms under a range of parameterization scenarios, with mixed results (e.g., [54,55]). Random forests and boosted regression trees, two data-driven approaches that are relatively immune to overfitting and can handle predictor interactions, can demonstrate high performance in unsampled areas (e.g., [56]). MaxEnt, another machine learning method, has been ranked

**Glossary**

**Anticipatory predictions:** predictions arising from extrapolating the state of a system (ecological or otherwise) either into the future (forecasts), under uncertainty around model parameters (projections), or within systems likely to be impacted by human action (scenarios).

**Biotic interactions:** interactions between organisms, such as predation, competition, facilitation, parasitism, and symbiosis.

**Correlative model:** model fitted to data and relating species occurrence or abundance at known times and locations to sets of environmental (biotic and abiotic) factors. The aim of a correlative model is to describe the conditions prescrib ing a species’ range, thereby generating a quantitative estimate of its geographical distribution.

**Cross-validation:** process of partitioning a dataset into complementary subsets, developing the model on one (i.e., training set) and validating it on the other(s) (i.e., the validation set). Cross-validation is most commonly used to estimate predictive performance; a single final model is often fitted to the full dataset.

**Explanatory predictions:** testable expectations about individual systems, outcomes, or properties, derived from scientific theory. The aim of explanatory predictions is to construct and/or corroborate hypotheses, and establish explanations for the mechanisms underpinning the functioning of natural systems.

**Extrapolation:** process of making predictions to covariate values that are outside the range, correlation structure, or value combinations of those in the training data. Can be spatial, temporal, environmental, or any combinations thereof.

**Fundamental niche:** full set of conditions and resources an organism is capable of exploiting to maintain populations in the absence of biotic interactions, dispersal limitations, habitat degradation, or immigration subsidy.

**Mechanistic model:** model representing causal processes underlying relationships between components of the studied system. Usually developed based on
the most transferable in some studies (e.g., [57]). Generalized linear and additive models have also been identified as robust choices for extrapolation (e.g., [37]), despite potential for generating unrealistic predictions outside the training scope. However, different approaches to model tuning and data treatment contribute to heterogeneity in performance [58], making the suitability of any given technique largely case-specific. A ‘silver bullet’ algorithm that is best under all circumstances is therefore highly unlikely, and other factors, such as species’ characteristics, can sometimes matter more than model choice [59]. Model averaging can avoid overreliance on a single technique by providing a weighted average of competing model predictions [60], and techniques that enable model coefficients to fluctuate in response to changes in habitat and resource availability [53] should improve transferability [8]. In recent years, dynamic models capable of tracking the temporal aspects of a species’ behavior and distribution, and joint species distribution models designed to simultaneously account for the co-occurrence of multiple species, have also gained traction. Although still in early stages of development, preliminary findings indicate potential for improved predictive performance [61]. Mechanistic models that harness prior biological knowledge within a given system (Box 3) could also enhance transferability, yet remain mostly undertested [62,63].

How Should Uncertainty Be Quantified, Propagated, and Communicated When Transferring a Model?

Uncertainty arises from many sources [64], including: sampling methodology, species vagrancy, data quality, environmental stochasticity, initial conditions, species identification, model specification, predictor choice, algorithm selection, and parameter estimation [7,45,57]. Improving predictability, and thus decision making (Box 1) [65], requires understanding the origins, propagation pathways, and ramifications of uncertainty, including its spatial and temporal patterns [64]. Model uncertainty is grounded in model assumptions, which underpin the choice of model algorithm, structure, and parameterization [65]. Uncertainty also varies spatially across a species’ predicted habitat [66], spreads through the multiple phases of model development (e.g., in hierarchical, multitarget models), and has multiplicative effects, such that its magnitude remains generally underappreciated [64]. These are significant challenges, which possibly explain the scarcity of attempts to account jointly for multiple types of variation (but see [29,66]). For this reason, clear protocols for measuring, accounting for, and reporting on uncertainty remain largely lacking. The latter often relates to the model’s intended purpose, such that quantifying parameter uncertainty might be a priority when seeking inference about a given predictor, but prediction uncertainty will gain importance when the primary objective is model transfer. Model averaging can help, though it is important to choose a model averaging method that adequately preserves the uncertainty of the combined prediction [64]. Recent advances in hierarchical modeling allow error estimates to propagate through various submodels within one ‘integrated statistical pipeline’, and could offer a solution in some cases.

How Can We Best Transfer Models through Time and Evaluate Them in Temporally Dynamic Systems?

All ecological systems exhibit temporal variability, whether predictable (e.g., tides), systematic (e.g., gradual climate warming), or random (e.g., cyclones). Constructing models using the full span (diurnal, seasonal, phenological, and annual) of conditions under which they will likely be applied can address this variation, although distinguishing erroneous predictions from temporally stochastic events in model validations remains a challenge. Time series of environmental variation could help diagnose anomalous conditions falling outside the baseline characteristics of reference and target systems. Studies suggest that some models can project more reliably over centuries [67] than shorter [68] or longer [69] time scales. A fundamental issue for
Figure 1. Snapshot of Some Predictive Model Transfers Published in the Ecological Literature. It is commonly assumed that transferability degrades away from the area of calibration. However, geographic separation is often a poor predictor of environmental similarity, such that even models transferred over short distances can yield erroneous predictions if conditions between reference and target sites substantially differ. This paradox is reflected in the contrasting performance of models projected over a range of distances. Here, we showcase studies selected to capture a broad range of taxa (e.g., birds, mammals, plants), ecosystems (terrestrial, freshwater, marine), locations (e.g., Australia, North America, Eurasia), and transfer distances (tens to thousands of km). Colors indicate whether transfers were considered successful (green, unbroken lines; 1–9) or not (dark red, dashed lines; 16–20), as reported by the authors and irrespective of the statistical methods chosen to build the models or the metrics used to evaluate them. Dual colors indicate scenarios in which the quality of transfers varied as a function of modeling algorithms (10–11), space (12), or species (13–15). Line thickness is proportional to the number of modeled species. Reference and target systems are shown as filled and open circles, respectively. Note that, for clarity, not all individual model transfers are portrayed for each study. Photographs depict model organisms and include: (1) Eurasian badger, Meles meles; (2) smooth crotalaria, Crotalaria pallida; (3) Norwegian lobster, Nephrops norvegicus; (4) garlic mustard, Alliaria petiolata; (5) invasive seaweed, Caulerpa cylindracea; (6) blueside sleeper, Lutjanus kasmira; (7) spiny water flea, Bythotrephes longimanus; (8) Bengal florican, Houbaropsis bengalensis; (9) northern pike, Esox lucius; (10) marbled murrelet, Brachyramphus marmoratus; (11) yellow-billed cuckoo, Coccyzus americanus; (12) blue whale, Balaenoptera musculus; (13) bronze dung beetle, Onitis ales; (14) daisy fleabane, Erigeron annuus; (15) rainbow darter, Etheostoma caeruleum; (16) koala, Phascolarctos cinereus; (17) Asian tiger mosquito, Aedes albopictus; (18) grey petrel, Procellaria cinerea; (19) black-backed woodpecker, Picoides arcticus; and (20) common toad, Bufo bufo. References and additional details are given in the supplementary material online.

forecasting is that temporal transfers are often impossible to validate because future events are unknown. One solution is to evaluate predictions of past events (i.e., hindcasting) based on independent historical (e.g., harvest and museum records) or paleoecological datasets, although spatio-temporal, collector’s, and taphonomic biases will complicate model calibration and validation [70]. However, for many species of management interest, such records remain unavailable or undermined by issues of spatial or temporal bias, mismatching resolutions between past and present data, and error propagation [71]. Sampling the response variable across its range of habitat variability offers an alternative. This strategy embodies the principle of “space-for-time substitution”, which assumes that spatial heterogeneity across multiple
contemporaneous sites at different positions along an environmental gradient can approximate temporal variability [72]. Such would be the case, for example, for areas subject to temperature regimes similar to those anticipated in the future, noting it will not be appropriate for species occupying small ranges or those not well-represented in the fossil record.

How Should Transferability Be Assessed?
Assessments of transferability demand appropriate diagnostics of prediction accuracy and precision [73], yet there is still little consensus on which metrics are most appropriate [6, 74]. All else being equal, true validation is possible only with independent data, which are often
Box 2. Ecological Niches in Model Transfers
Transferability should be greater in models fitted to observations that document all dimensions of, and constraints placed upon, the fundamental niche. However, most datasets fall short of meeting this requirement, because organisms do not always occupy all suitable habitats (or conversely occupy unsuitable ones), either as a result of dispersal barriers, gregarious behavior, anthropogenic disturbances, biotic exclusion (e.g., competition, parasitism), or simply because those habitats do not currently exist [61]. These constraints apply not only to the fundamental niche but also equally to the realized niche (i.e., subset of habitats and resources accessible to a given species), meaning that the latter is often restricted in comparison with the former, and that even a perfect understanding of the fundamental niche alone does not make for correct predictions [88]. In practice, failure to fully represent the fundamental niche might lead to truncated response curves that yield unrealistic predictions [38,74], and models that disregard information on absences (e.g., presence-only and environmental envelope models) have been criticized accordingly. While fundamental niches can be expected to stay constant over timescales relevant to management (i.e., daily to decadal), realized niches will typically vary both spatially and temporally. This complicates model transfers, particularly when the realized niche becomes a direct function of habitat selection behavior as it relates to resource availability or physiological tolerance limits [53]. The selection of environmental predictors also impacts the degree to which both fundamental and realized niches can be captured. Emphasis should thus be placed on more direct, functional predictors to foster improved model transfers. Understanding the relationship between niche types can help determine when transfers are more likely to succeed or fail (Box 3), and might be facilitated by jointly modeling target species with their competitors, predators, or facilitators [41]; by coupling distribution and population dynamics models; or by incorporating complex eco-evolutionary factors into model formulations [49] (but likely at the expense of higher data requirements [86]). While mechanistic models (Box 3) are well suited to delimiting species’ fundamental niches [87], to date their application remains limited to a few, well-studied taxa for which physiological parameters are documented in detail [4].

Box 3. Correlative Versus Mechanistic Models
Correlative and mechanistic modeling are two contrasting modeling philosophies that respectively emphasize patterns versus processes [2,87]. Correlative models draw statistical linkages between response variables (e.g., species occurrence) and features of the environment (i.e., biotic and abiotic predictors), but have been criticized for failing to explicitly capture the underlying processes (e.g., dispersal ability) that affect said response variable [41]. By contrast, mechanistic or process-based models are built around explicit descriptions of biological mechanisms and parameters that have a clear a priori interpretation. If formulated appropriately (e.g., experiment-based parameterizations of species’ responses to environmental conditions [82,88]), some mechanistic models can be expected to achieve greater realism, with potential for higher transferability [88]. However, mechanistic models suffer from the same issues of nonstationarity as correlative models, and are thus not immune to potentially inaccurate extrapolation (Box 2). The limited availability of experimental data also remains a major constraint, and it is thus uncertain if mechanistic models can live up to their promise of providing more accurate forecasts of species’ range shifts under climate change [62,86]. Indeed, a few studies have found mechanistic and correlative models to perform equally well [63,90]. Mechanistic model implementation also comes at the cost of increased data and computational requirements, limiting their wider use. Although a useful methodological dichotomy, the distinction between correlative and mechanistic models is usually blurred in practice [91] because models within each class rely to some degree on parameterization against observed data, and most ecological mechanisms are actually empirical, rather than theoretical. This need not be detrimental, as it allows a progressive transition from the phenomenological extreme of regression models towards the process-based extreme of mechanistic models. In reality, the approach undertaken will often be dictated by the study context (e.g., availability of prior knowledge). In recent years, arguments have been made for blending correlative and mechanistic approaches, by using mechanistic knowledge as a benchmark for validating correlative models [92], by using mechanistic variables as direct inputs to correlative models [89], or by combining the respective predictions of each model class [93]. Irrespective of the approach chosen, explicitly considering the underlying mechanisms that affect the response is important, and developing a thorough rationale for selecting environmental predictors is crucial to ensure that they are functionally, ecologically, and physiologically meaningful and therefore support transferability [73,94].

unavailable (e.g., a region not yet invaded by a pest) or insufficient (e.g., small sample size). In the absence of validation data for a target site, transferability can only be estimated by contrasting predictions with existing expert knowledge or simulations, and, where feasible, benchmarking performance by projecting models into multiple alternative data-rich scenarios [6]. Cross-validation can also provide a reasonable approximation of independence, so long as it can be structured to mimic prediction conditions and minimize correlations (e.g., by deliberately choosing cross-validation folds to emulate extrapolation) [37]. Ultimately, consistent assessments of transferability will require unified and widely applicable standard metrics.
that enable direct comparisons among studies, systems, and taxa [6]. Instrumental to this are novel approaches to model evaluation and validation (e.g., [75]) that are generally independent of model choice and response variable type.

Concluding Remarks
Predictions remain a major frontier in ecology [1,45], not least because they are most pressingly needed where we lack sufficient ecological information (Box 1). This leads to a catch-22, where the absence of knowledge encourages the search for transferable models but also impedes their evaluation. Concerted efforts to increase both data quality and data availability are therefore crucial to enhancing the practice of model transfers in ecology [2,76]. Ideally, data should be: unbiased, with explicit coverage of important gradients, high-frequency, long-term, and real-time, so as to maximize opportunities for anticipatory predictions that can be validated with minimal delay [76]. Alternatively, model transfers into novel systems can provide a platform against which data can later be benchmarked once available. Whilst remote sensing, increasingly used in distribution modeling, has the potential to fulfill many of these data ideals, care must be taken to match scales of data to the phenomena that the models are attempting to quantify. Indeed, models that are built with a thorough consideration of ecological processes and the scales at which they operate, even if they are not actually mechanistic models, should have a greater chance of being transferable. Ultimately, the fastest way to enhance predictions is to use them as tools for learning [9] (Figure 2). This necessitates meticulous monitoring of predictive performance, and importantly, rigorous documentation of failures to transfer [20] (Box 4). Quantifying transferability also requires clarity and coherence, yet assessments of model predictions have rarely been harmonized [1]. Without widely applicable transferability metrics that summarize different aspects of predictive success, comparisons between studies will retain little meaning [6]. Indeed, “How should transferability be assessed?” emerged as the knowledge gap of highest priority during our discussion. Filling this gap appears essential, not only to demonstrate greater levels of transparency in model applications (Box 1), but also

Box 4. Why Can Model Transfers Fail?
Failures to transfer occur for many reasons [96]. Arguably the most obvious is that models tightly fitted to calibration data often do not extrapolate well to novel data [38]. Predictive models also often assume that organisms are at quasi-equilibrium with their environment, such that occupancy or abundance data reflect site suitability. However, biological interactions, disturbance regimes, habitat loss and human impacts (e.g., harvesting), stochastic mortality, or dispersal constraints can prevent species from persisting in or accessing favorable habitats, potentially leading to biased representation of environmental conditions (i.e., a failure to sample the fundamental niche) (Box 2). Species can exhibit immediate responses to one or several components of global change, even though disruptions to networks of biotic interactions can slow down or hasten evolutionary adaptations [96], and population dynamics can lag behind the trend of global change drivers [97]. Nonstationarity can also undermine transferability, because species–habitat relationships vary in complexity, strength, and direction across different ecosystems. The effects of population density on apparent habitat preferences can compromise transferability if increases in population density force individuals into suboptimal areas [50], although modeling the dependencies of habitat coefficients on population density offers a potential solution [8,30]. Numerous datasets additionally suffer from sampling biases as well as spatial and temporal autocorrelation, leading to underestimations of heterogeneity among environmental gradients or populations, which cause problems for fitting and validating models [98]. Where possible, statistical methods for dealing with spatial and temporal correlation should be employed to mitigate these issues [99]. Further bias in predictions can arise from local factors that remain undetected due to the coarse resolution at which most models are calibrated [100]. Mismatched scales between reference and target systems (e.g., temporal range, sampling year, transect size) and the omission of important predictors (e.g., fishing pressure, habitat structure) are among other explanations for models transferring poorly [20,27]. Lastly, failures to transfer can simply ensue from inadvertent stochastic events in the evaluation data, rather than from poor transferability per se (i.e., a model might correctly predict the presence of a species, but the validation data do not record the presence due to some stochastic process). Clearly, advancing the application of model transfers in ecology requires increased understanding of the processes and conditions that affect transferability, which will be aided by encouraging researchers to publish the results of unsuccessful model transfers [20].
because transferring models beyond the environments in which they were initially built weakens their credibility and defensibility. Confidence in model predictions will therefore remain limited until we can determine how well models actually perform on independent datasets [64]. Rather counterintuitively, better transferability might not necessarily equate to better decisions if the uncertainties associated with model predictions are not suitably measured, reported, and communicated to end-users and policy makers [13]. So far, a comprehensive treatment of uncertainty and its sources has been too complex and laborious to achieve [2], although significant advances are being made towards this goal. Whilst substantial challenges lie ahead on the road to realizing the full potential of transferable models, the prospective gains are great. As Houlanah et al. [1] note, ‘transferability is critical to [scientific] understanding because understanding without transferability is [...] ephemeral and transient’.

Author Contributions
A.M.M.S., K.L.Y., and P.J.B. conceived the study. K.L.Y., A.M.M.S., P.J.B., M.J.C., and K.M. organized and delivered the conference workshop. All authors formulated challenges and voted on the assembled list. K.L.Y., P.J.B., and A.M.M.S. compiled the data and led the writing of the manuscript. K.L.Y., P.J.B., A.M.M.S., M.J.C., K.M., and C.R. led working subgroups. All authors contributed to the writing of individual sections of the manuscript and provided comments on drafts.

Acknowledgements
A.M.M.S. was supported by the Australian Research Council (grant: DE170100841) and an IOMRC (UWA/AiMS/CSIRO) collaborative Postdoctoral Fellowship. P.J.B. received support from the Australian Government’s National Environmental Science Programme (NESP), C.M. was supported by the Australian Research Council (grant: DE140100701). D.Z. was supported by the Swiss National Science Foundation (grant: PZ00P3_168136/1) and the German Science Foundation (grant: Zu 361/1-1). J.E. was supported by the Australian Research Council Centre of Excellence for Environmental Decisions (CE1101000104). S.P. was supported by the USDA (grant: 17-8130-0570-CA) and DEFRA. P.N.H., J.J.R., and L.M. were supported by a US Navy Cooperative Agreement (N62470-15-2-8003). A.R.J. was supported by The Spencer Gulf Ecosystem Development Initiative and the Goyder Institute for Water Research (project number: CA-16-04). A.M.B. was supported by FCT and FEDER/COMPETE 2020 (project: IF/00266/2013/CP1168/CT0001). C.F.D. was supported by the German Science Foundation (grant: DO 786/10-1). We thank all additional participants in the August 2016 IMCO4 symposium on ‘Increasing the utility of predictive models: Understanding model transferability’, including: Bapu Vatla, Chris Golden, Mohd Qurban, Kerry Howell, Sara Maxwell, Telmo Morato, Robin Anderson, Pierre Pepin, Stephanie Sardelise. Our thanks also go to Nikolaus Zimmermann for contributing initial questions and providing thoughtful input in the early stages of manuscript preparation. We also gratefully acknowledge the constructive comments provided by three anonymous reviewers on a previous version of this manuscript.

Supplemental Information
Supplemental information associated with this article can be found, in the online version, at https://doi.org/10.1016/j.tree.2018.08.001.

References