Abstract. Models are increasingly being used for prediction in ecological research. The ability to generate accurate and robust predictions is necessary to help respond to ecosystem change and to further scientific research. Successful predictive models are typically accurate, reliable, and transparent regarding their assumptions and expectations, indicating high predictive capacity, robustness, and clarity in their objectives and standards. Research on improving these properties is becoming more common, but often individual research projects are focused on a single aspect of the modelling process and are typically disseminated only within the field where the research originated. The goal of this review is to synthesize research from various disciplines and topics to provide a coherent framework for developing efficient predictive models. Our framework summarizes the process of creating predictive models into three main stages: (1) Framing the Question; (2) Model-Building and Testing; and (3) Uncertainty Evaluation with proposed strategies associated with each stage to help produce more successful predictive models. The key strategies identified within our framework form specific guidelines, providing a new perspective to help researchers make predictive modelling more accurate, reliable, and transparent.

Key words: adaptive modelling; ecological forecasting; ecological prediction; model selection; uncertainty quantification; uncertainty reduction.

INTRODUCTION

In the Anthropocene, systems are being altered worldwide at an accelerating rate. Preparing for the challenges arising from these changes requires models that can disentangle complex dynamics, predict the consequences of environmental changes, guide proactive management decisions, and account for key known uncertainties (Clark et al. 2001, Urban et al. 2016, Dietze et al. 2018). Models designed for prediction, hereafter referred to as “predictive models,” produce quantifiable and testable hypotheses, which, when evaluated against new observations, may improve our knowledge, refine our hypotheses, and lead to better predictions in the future (Dietze 2017a).

Research on how to improve ecological prediction is an ongoing pursuit (Clark et al. 2001, Carpenter 2002, Evans 2012, Schindler and Hilborn 2015, Dietze et al. 2018). Efficient predictive modelling requires maximizing: accuracy (a measure of how closely a model’s outputs resemble the “true” value), reliability (a measure of a model’s precision), and transparency (the explicit provision of model choices, their assumptions, the steps in the modelling process, and the
expectations for the outputs), hereafter referred to as ART. We highlight accuracy as it can act as an indicator of successful science (Evans 2012), reliability as it plays a key role in improving planning and decision-making (Clark et al. 2001), and transparency as it is both a prerequisite for reproducibility and it ensures that model strengths and limitations are understood by the scientific community, practitioners, and stakeholders (Addison et al. 2013).

A rich literature exists on the development and use of predictive models, but the key building blocks of modelling (e.g., the philosophy of prediction, model construction, model selection, and uncertainty identification) are often discussed in isolation (e.g., Aho et al. 2016 for model selection; Walker et al. 2003 for uncertainty identification). Moreover, despite prediction being a necessary component of many fields (e.g., ecology, epidemiology, engineering, hydrology, and computer science), relatively few ideas are disseminated among them (Saltelli et al. 2008). The goal of this review is to synthesize ideas about modelling practices from different topics and disciplines to provide simple guidelines for how to make predictive modelling ART. Our primary audience is those new to predictive modelling, but given the breadth of topics covered, our hope is that even experienced modellers will gain insight into new techniques that will benefit their research. For our guidelines, we created a simple coherent structure, the Predictive Modelling Framework, that summarizes the process of predictive modelling in three key stages (Fig. 1): (1) Framing the Question; (2) Model-Building and Testing; and (3) Uncertainty Evaluation. For each stage, we provide key goals and strategies for achieving these goals, to help make predictive modelling ART (cf. Table 1). While aspects of our guidelines overlap with other recent recommendations (e.g., Urban et al. 2016, Dietze et al. 2018, Garcia-Diaz et al. 2019), our unique framework along with the breadth of strategies discussed provides a complementary perspective to the existing literature.

STAGE A: FRAMING THE QUESTION

First, the scope and constraints of the question need to be defined based on knowledge derived from the biological system.

Key goals

1. Goal A.1: Explicitly define the purpose of the predictive model
2. Goal A.2: Establish expected prediction properties

The first step to building an appropriate predictive model is to establish the context and the frame of the question. We identify two goals at this stage to help make predictive modelling ART: (A.1) explicitly define the purpose of the predictive model; and (A.2) establish expected prediction properties. Explicitly defining the purpose refers to specifying the intentions of a predictive model, which could be, for example, to inform policy management (Taylor and Hastings 2004, Lowe et al. 2014) or to test a theory (e.g., Bailly et al. 2014, Borregaard et al. 2017) in either a general or specific system. Establishing expected prediction properties refers to establishing a priori the characteristics and expected accuracy and/or precision levels of predictions (Røskaft et al. 2017). For example, we may specify if predictions will be qualitative or quantitative, and if quantitative, whether predictions will be regarded as first approximations or precise values. We may also state whether they will be deterministic or probabilistic and what level of accuracy is acceptable given the model purpose. Additionally, if the predictions are probabilistic, we may state the amount of uncertainty expected given our purpose and knowledge about the model, data, and system.

Strategies to achieve goals

The definition of a clear model purpose is often omitted during model creation (Milner-Gulland and Shea 2017). This is unfortunate because a clear purpose helps avoid unnecessary controversies that may just stem from misunderstandings regarding what a model is and is not supposed to do (Grimm et al. 2010). To explicitly define a model’s purpose, we first propose choosing a general model purpose: to test theory (explanatory predictions), based on hypothesis testing, or to make predictions about a nonobserved time, space, or species (anticipatory predictions), based on our understanding of the system’s functioning (Mouquet et al. 2015). Then, the model characteristics and available data help to refine the specific
The purpose of the model. For example, given a question about a species’ future niche, the type of data available (e.g., presence/absence, abundance, physiological, and/or dispersal) in concert with the model choice (e.g., pattern-based species distribution model or a mechanistic range expansion model) will inform what the model can and cannot capture. In this case, the data and model inform whether the purpose is to generate short- or long-term predictions and whether the resulting predictions should be process- or pattern-based: emerging from explicitly modelling the underlying mechanisms driving patterns (e.g., species distributions arising from movement) or based on analysis and extrapolation of patterns (e.g., future species distributions from current distributions) without explicitly incorporating underlying mechanisms. We advise being flexible as the availability of new data may allow for updates or expansions to the model purpose.

To achieve the second goal (A.2, establish expected prediction properties), we propose using information from the model purpose and the characteristics of the data and model to inform expectations. For example, given an explicit purpose of predicting species’ ranges under climate change with presence data and a maximum entropy species distribution model...
Table 1. Guidelines for predictive modelling as outlined in the text: the three main sections, specified by the columns, and three subsections, the rows.

| Predictive Modelling Framework | (A) Framing the Question | (B) Model-Building and Testing | (C) Uncertainty Evaluation |
|-------------------------------|--------------------------|-------------------------------|---------------------------|
| (1) Key goals                | • GOAL A.1: Explicitly define the purpose of the predictive model | • GOAL B.1: Choose the most appropriate model type(s) | • GOAL C.1: Identify uncertainties across multiple dimensions |
|                               | • GOAL A.2: Establish expected prediction properties | • GOAL B.2: Create an appropriate suite of candidate models (using the chosen model type(s)) | • GOAL C.2: Account for known uncertainties |
| (2) Conceptual and technical strategies to achieve goals | • Make a philosophical decision whether the model purpose is anticipatory or explanatory (Mouquet et al. 2015) | • Use model purpose and data qualities to inform model type | • GOAL C.3: Reduce uncertainties where feasible |
|                               | • Utilize model and data characteristics to inform the model’s purpose | • Choose specialized model types to handle difficult data quantities (e.g., hierarchical state-space models for small data; machine learning for big data) | |
|                               | • Use model and data characteristics in conjunction with the purpose to inform the expected prediction properties | • Adopt specialized techniques for parameterization and variable selection (e.g., Bayesian inference or dimensionality reduction) to handle small or big data during candidate model construction | |
|                               | • Iteratively refine the model purpose and prediction properties as new information becomes available | • Utilize a multi-model approach and an adaptive modelling approach (e.g., Urban et al. 2016) to build candidate models | |
|                               | • Clearly state the purpose and the expected prediction properties in the research article | • Perform in-sample and out-of-sample assessment using information-theoretic approaches/cross-validation and validation to select appropriate final models | |
| (3) How achieving goals makes predictive modelling “ART” | • Defining the explicit model purpose prevents misinterpretation of the model’s goals and limitations (Grimm et al. 2010) and can lead to more informed model choices | • Choosing model types that complement the research question and the data available leads to predictions that address the question | • Identifying uncertainties across multiple dimensions communicates the source and quantity of uncertainty, providing directions on how future improvements can make predictive models more reliable |
|                               | • A priori establishment of expected prediction properties establishes a baseline for model evaluations, thus increasing transparency, and allows for mismatches between desired and observed properties to be addressed, thus improving accuracy and reliability | • Creating a suite of appropriate candidate models using multiple models, appropriate parameterizations, and suitable variables communicates uncertainties and helps reduce error. Iteratively updating models typically leads to declines in parameter uncertainty (Dietze 2017a) and informs or updates new hypotheses leading to better model structures | • Accounting for uncertainty quantifies and apportions uncertainty to appropriate inputs, highlighting uncertainties that dominate a system |
|                               | • Selecting the appropriate model in final selection improves accuracy and reliability as predictive capacity and precision are selected for when balancing residual error with parameter uncertainty (Dietze 2017b) | • Selecting the appropriate model type(s) to handle difficult data quantities (e.g., hierarchical state-space models for small data; machine learning for big data) | • Reducing uncertainties increases the signal-to-noise ratio and thus model precision |

Notes: The columns are the stages of the Predictive Modelling Framework (A, B, and C in Fig. 1), and rows are (1) the key goals for increasing accuracy, reliability, and/or transparency (ART); (2) the recommended conceptual and technical strategies to help meet goals; and (3) an overview regarding how achieving the goals can help make predictive modelling ART.
(Maxent), a first-order approximation is likely the most appropriate expected property (Carlson et al. 2017). As more data become available and/or the purpose of the model changes, the prediction expectations should typically be updated as well. For example, a model that is parameterized using broad interspecific relationships may subsequently be updated with species-specific data to improve first-order coarse-level predictions to fine-scaled species-specific ones (Grimm et al. 2006, Molnár et al. 2017).

Achieving the first goal (A.1, define the model purpose) also helps to achieve the second goal (A.2, establish expected prediction properties). For example, selecting how generalizable a model should be, which is part of a prediction’s purpose, will inform the level of predictive accuracy deemed acceptable, an aspect of its expectations (Fig. 2). If a model purpose is to test a generalizable theory, we may tolerate coarser predictions (i.e., a greater discrepancy between predicted and observed values) than if the model’s aim is to generate system-specific predictions. Similarly, if our purpose is to predict farther into the future (e.g., into the next decade), our expectations may include greater uncertainty than if our purpose is to generate short-term predictions (e.g., into the next month) (see the forecast horizon in Stage C: Uncertainty Evaluation).

Lastly, to define the model purpose and to establish expected prediction properties, we need to ensure that both are clearly stated in the research. Ensuring they are stated or restated prior to presenting model results helps to frame the results and prevent miscommunications.

**Stage B: Model-Building and Testing**

Model-Building and Testing constitutes the second stage of developing predictive models and involves choosing a model type, selecting functional forms and variables, estimating parameters, and comparing and assessing models with available data.

**Key goals**

1. Goal B.1: Choose the most appropriate model type(s)
2. Goal B.2: Create an appropriate suite of candidate models
3. Goal B.3: Select the most appropriate final model(s)

An appropriate model is one that produces predictions that address the proposed research question, accounts for applicable variables and/or processes, and demonstrates accuracy and precision when compared to independent data. We identify three sequential goals in Model-Building and Testing to help make predictive modelling ART: (B.1) choose the most appropriate model type(s); (B.2) create an appropriate suite of candidate models; and (B.3) select the most appropriate final model(s). By model type, we mean broad mathematical, statistical, or algorithmic approaches designed to address the general purpose of the study (e.g., SIR models for disease transmission, dynamic energy budget models, correlative species distribution models, and ecophysical models). By candidate models, we mean a set of models, sometimes of different functional forms, that contain various processes and/or variables of interest that often act as competing hypotheses. By final models, we mean models chosen at the end of Model-Building and Testing from the set of candidate models that are deemed the “best” given the available data and the model’s purpose. For example, given a purpose to predict distributions of species with limited input data, an appropriate model type may be a correlative species distribution model (SDM) (an approach designed to describe patterns, as opposed to mechanisms, between species occurrences and environmental data [Dormann et al. 2011]). Candidate models would consist of a set of correlative SDMs constructed using different combinations of variables such as climate and vegetation indices, hypothesized to drive species occupancy. Final models would be those candidate models selected from the set that demonstrate a balance between parsimony and predictive power when tested against available data.

**Strategies to achieve goals**

**Strategies to choose the most appropriate model type(s).**—Selecting the most appropriate model type requires a consideration of the strengths, limitations, and assumptions of different model types. At the broadest level, models can be classified as phenomenological (synonyms: correlative,
pattern-based) or mechanistic (synonym: process-based); deterministic or stochastic; dynamic (i.e., accounts for time-dependent changes in the state of a system) or static (i.e., describes a state at a fixed point in time); continuous or discrete; and/or heterogeneously or homogeneously structured (e.g., with respect to space and age) (Haefner 1996). These broad categorizations, paired with the model purpose and the data, will inform the selection of a specific model type. Given a specific topic, potential appropriate model types can be found through a thorough search of the relevant literature. Here, we illustrate the nonlinear process of selecting an appropriate model type for the purpose of predicting a species’ distribution. First, using the literature (Briscoe et al. 2019), we isolate seven SDM classes ranging from correlative SDMs to individual-based SDMs that fulfill this general purpose. The strengths and limitations of each approach, along with their data requirements (see Briscoe et al. 2019 for this overview), inform the selection of the specific SDM. Correlative SDMs typically have more readily available data that are easier to collect than other types of SDMs (Briscoe et al. 2019), and may be useful when processes are unknown or the parameters involved in these processes cannot be estimated from available data (Dormann et al. 2011). However, if species’ movement or biotic interactions drive species’ future distribution, correlative SDMs may overpredict the distribution of species’ presence if they fail to capture dispersal constraints or those key interactions (Uribe-Rivera et al. 2017). If instead dispersal is expected to be a key process and occurrence data are available, a process-based SDM such as an occupancy dynamics model may be the most appropriate choice, as they estimate probabilities of local extinction and colonization from occupancy data (Briscoe et al. 2019). Alternatively, if species’ vital rates are known to be linked to the environment, mechanistic niche models (ecophysiological models) can offer improved predictive capacity (Pomara et al. 2014) and therefore may be the most appropriate type, particularly when long-term transferability is desired (Briscoe et al. 2019). For any specific topic, weighing the benefits and drawbacks of various model types and considering the data available lead to a more informed selection of a model type.

The choice of an appropriate model type may not only be informed by data quality or type but also by the amount of data available. Many systems and species remain understudied with little to no data for the development of predictive models, while in others, citizen science and technological advances have resulted in massive influxes of data that are too big for classical modelling approaches to handle. For small data, hierarchical models (structures comprised of modular hierarchical units such as metapopulations, populations, and individuals) are useful for parameterization as they allow data-rich model parts to inform data-limited parts. In particular, hierarchical state–space models (a type of hierarchical model comprised of a process and data model) help overcome missing information by using partial data to reconstruct underlying states and parameters in a hierarchical framework (Kindsvater et al. 2018). For example, hierarchical state–space models can allow multiple parameters to be estimated for data-limited taxa by using information from data-abundant taxa.

Fig. 2. Model generalizability vs. prediction accuracy: Model generalizability typically trades off with average prediction accuracy. As generalizable models are typically expected to have lower average accuracy than highly system-specific models, the range of “acceptable” average accuracy (blue region) increases with increasing generalizability. The “select model” threshold plateaus at high generalizability as below a certain average accuracy all models should be rejected. Note the exact shape of the threshold curve is subjective (and is context-dependent) but should generally be monotonic and approaching a horizontal asymptote.
with a shared ancestry (Dick et al. 2017). For big data, machine learning techniques, defined as efficient and accurate prediction algorithms (Mohri et al. 2012), are potentially appropriate model types as they excel at modelling highly dimensional, nonlinear data with complex interactions (Olden et al. 2008). These algorithms avoid important constraints inherent to many traditional statistical models (e.g., a priori specification of interactions, error distributions, and functional forms) allowing greater flexibility when handling large data (Thomas et al. 2018). Additionally, as machine learning techniques can be integrated with frequentist or Bayesian methods, non-probabilistic or probabilistic models are both viable options when selecting a machine learning model type. Examples of machine learning algorithms include decision trees, boosted regressions, generative adversarial networks, and classical and deep learning neural networks (Hastie et al. 2009).

Strategies for creating an appropriate suite of candidate models.—Appropriate suites of candidate models are sets of models, each representing various processes and variables, that are properly parameterized and easily updated. In general, we suggest adopting multi-model and adaptive modelling frameworks (Urban et al. 2016) to facilitate model improvements and, when needed, adopting specialized methods such as Bayesian parameter estimation and dimensionality reduction (Torrecilla and Romo 2018) to handle small and large data quantities, respectively. Multi-model frameworks consist of either multiple independent models or ensemble models. Multiple independent models are separate models representing different hypotheses or relationships that each produce a single prediction. Conversely, ensemble models are weighted combinations of separate models (that may differ only by a parameter value or variable) that produce one general prediction. While ensemble models have been typically shown to outperform multiple independent models (e.g., Breiner et al. 2015, Abrahms et al. 2019), their benefits typically increase as the covariance of the individual model predictions and the mean bias of the individual models decrease (Dormann et al. 2018). When adopting a multi-model framework, models should be comparable (Dormann et al. 2018) and a statistical or process-based null model (a base model incorporating basic applicable biological processes) should be included as a simple baseline against which candidate models may be compared and evaluated (Haefner 1996).

The second broad strategy for creating candidate models, adopting the adaptive modelling framework (as outlined by Urban et al. [2016]), treats modelling as an iterative process of revision and testing, focused on making models amenable to new data that become available. While the adaptive modelling framework applies to the entire Predictive Modelling Framework, we emphasize it here so that we set a precedent for future updates and model comparisons. By embracing these adaptive strategies and viewing modelling as a dynamic process, researchers ensure that candidate models are data-driven and applicable to the system under consideration (Restif et al. 2012). The implementation of an adaptive modelling framework is optimized when candidate models are built to determine future data collection priorities and when highly sensitive and uncertain parameters are identified such that resources can be allocated toward improving their estimates.

An appropriate suite of candidate models also implicitly assumes that models are appropriately parameterized and contain suitable variables. If issues pertaining to sparse or big data are not addressed when choosing a model type (as outlined in the previous section), specialized techniques such as Bayesian inference and dimensionality reduction can help create well-fitted appropriate models. While not limited to small data, Bayesian inference, a method that utilizes Bayes’ theorem to incorporate prior beliefs, can assist with parameterization in data-limited cases by allowing expert opinion or previous analyses to inform current models (Hilborn and Mangel 1997, Dietze et al. 2018). For example, Bayesian inference has been shown to improve model fitting in cases where species have low intensity or missing occurrence data (e.g., Jasper et al. 2018, Outhwaite et al. 2018). Note that if expert elicitation is adopted to inform analyses, proper elicitation procedures must be employed to correct for potential biases stemming from expert overconfidence (Speirs-Bridge et al. 2010).

For big data, dimensionality reduction, a process that reduces the number of random variables using feature selection or extraction
(Torrecilla and Romo 2018), is one potential option to help improve big data usability. While feature selection isolates a subset of the original variables, feature extraction projects data onto a lower dimension, creating new variables that are combinations of the originals. Feature selection is particularly useful when a subset of variables is needed from combined datasets (e.g., Fassnacht et al. 2014) or from datasets containing complex variable interactions (e.g., Asakura et al. 2018). For example, feature selection algorithms have been shown to optimize the performance of ecological niche models by identifying variables with high relevancy and low redundancy from large sets of possible climatic, topographic, and anthropogenic variables (e.g., Tracy et al. 2018). In contrast, feature extraction is valuable when original variables do not need to be retained, often the case with remote sensing data (e.g., Gholizadeh et al. 2018) or large-scale environmental data (e.g., Macintyre et al. 2018). Examples of feature extraction methods include linear techniques such as principal component analysis (PCA) and multidimensional scaling (MDS) and nonlinear methods such as isometric feature mapping (Isomap) (Mahecha et al. 2007).

Strategies for selecting the most appropriate final model(s).—For the final goal of Model-Building and Testing (selecting the most appropriate final model(s)), we propose using in-sample and out-of-sample assessments to determine the best models. In-sample assessment measures model fit using data from model-building and calibration, whereas out-of-sample assessment measures a model’s predictive capabilities against semi- or fully independent data (Fig. 3). Semi-independent data are temporally, spatially, or otherwise distinct from data used in model fitting (Wenger and Olden 2012), whereas fully independent data meet these criteria and additionally differ from model calibration data in terms of observers, measurement tools, or sampling design. Both in-sample and out-of-sample

Fig. 3. Final model selection: Selecting the most appropriate final model(s) requires two assessments: (1) in-sample assessment, where models are evaluated using data from model fitting (in-sample data) and (2) out-of-sample assessment, where models are evaluated against data that are independent of in-sample data in terms of time, space, etc. (out-of-sample data). In-sample assessment eliminates the worst fitting models (rejected) with the remaining set (selected) evaluated by out-of-sample assessment. Note that the best model determined by in-sample assessment may not be the best in out-of-sample assessment. Given that out-of-sample assessment measures transferability, typically models selected here (final selected) are deemed the most appropriate final model(s).
assessments aim to select models with high predictive capacity.

The first assessment, in-sample assessment, typically involves either information-theoretic approaches or cross-validation (Akaike 1973, Stone 1977). Information-theoretic approaches include, for example, the well-known Akaike information criterion (AIC) and versions thereof (e.g., AICc for small data), the deviance information criterion (DIC) for Bayesian models (Spiegelhalter et al. 2002), and the Watanabe-Akaike information criterion (WAIC) (Watanabe 2010) for Bayesian models and singular models (e.g., hierarchical and machine learning model types). These approaches estimate the quality of a statistical model by balancing complexity and model fit (Burnham and Anderson 2002). The other approach, cross-validation, randomly splits the data into at least two subsets, fitting models to one subset (or more) and evaluating their accuracy using the remaining subsets. For an overview of cross-validation approaches, see Krstajic et al. (2014). In addition to helping select final independent models, information-theoretic and cross-validation techniques are also key for selecting final weights for ensemble models (Dormann et al. 2018).

In ecological research, information-theoretic approaches are commonly adopted to perform in-sample assessment (Hooten and Hobbs 2015). However, as they do not provide a direct measure of predictive accuracy, the best-performing model or models may only be the best of a poor selection (Taper et al. 2008). In contrast, cross-validation can provide a direct estimate of prediction error for in-sample data and, due to it being a nonparametric method, includes fewer assumptions about the true underlying model (Gelman et al. 2014). However, cross-validation may be unreliable when sample sizes are small (Isaksson et al. 2008) and information-theoretic approaches are typically simpler to implement as they are often already built into statistical packages. As such, we suggest using information-theoretic approaches when their underlying assumptions are fulfilled (that models have the same dependent variables and that realized error distributions conform to theoretical expectations). Conversely, we suggest using cross-validation when candidate models differ in their dependent variables or when the realized error distribution does not match theoretical expectations. In many cases, both techniques will yield similar results as information-theoretic approaches such as AIC, DIC, and WAIC can be viewed as asymptotic approximations of different versions of cross-validation (Gelman et al. 2014). However, both approaches may select for more complex models when uncertainties are not accounted for (Dietze 2017a), reinforcing the need for out-of-sample assessment and uncertainty quantification when selecting final models (the latter covered in Stage C: Uncertainty Evaluation).

Following in-sample assessment, out-of-sample assessment evaluates the predictive abilities of candidate models by testing model predictions against fully or semi-independent data. Relying only on in-sample assessments is suboptimal (Mosteller and Tukey 1977) as top-performing models from in-sample assessment may have worse transferability than lower ranked models (Wenger and Olden 2012) (Fig. 3). While out-of-sample assessment may be optimized with fully independent data, often this type of data is not available in ecology (Urban et al. 2016, Dietze et al. 2018). Nonrandom cross-validation techniques, which use semi-independent data (e.g., splitting a time series systematically into distinct, nonoverlapping time frames [see Wenger and Olden 2012]), may then be a sufficient substitute if data are extrapolated in terms of times, locations, and/or conditions of interest. Additional details for conducting in-sample and out-of-sample assessment techniques may be found in Johnson and Omland (2004), Link and Sauer (2016), Vehtari et al. (2017).

STAGE C: UNCERTAINTY EVALUATION

The final stage in the Predictive Modelling Framework is Uncertainty Evaluation. Uncertainty, the amount of incomplete knowledge about a value (van Oijen 2017), is found in all parts of modelling, including the context, model structure, parameters, inputs, and outputs. Uncertainty Evaluation is the stage where uncertainties are qualified, quantified, and minimized.

Key goals

1. GOAL C.1: Identify uncertainties across multiple dimensions
2. GOAL C.2: Account for known uncertainties
3. GOAL C.3: Reduce uncertainties where feasible

Uncertainty is not limited to poorly constructed models as even the best parameterized, most realistic models contain uncertainties (Schindler and Hilborn 2015). However, sources of uncertainty are often not identified (Beale and Lennon 2012) or remain unquantified in final predictions (Clark et al. 2001, Milner-Gulland and Shea 2017). To help researchers account for uncertainty in their predictions and make predictive modelling more ART, we discuss three sequential goals: (C.1) identify uncertainties across multiple dimensions, such as level (the degree of uncertainty) and source (whether uncertainties arise from a lack of knowledge or are a consequence of stochastic elements); (C.2) account for known uncertainties by quantifying their nature and magnitude; and (C.3) reduce uncertainties where feasible by increasing the signal-to-noise ratio in models.

**Strategies to achieve goals**

**Strategies to identify uncertainties across multiple dimensions.**—To identify uncertainties, we propose adopting frameworks that include more than a single dimension. The uncertainty matrix (Walker et al. 2003) is one such framework designed to identify, assess, and prioritize uncertainties across three key dimensions: (1) level, representing a metric that qualifies uncertainty on a scale from determinism to total ignorance; (2) nature, classifying uncertainty as variability (the intrinsic natural variability of a system) or epistemic (the degree of human knowledge); and (3) location, identifying where uncertainty manifests itself in the modelling process (the model’s context, its structure, its technical implementation, its inputs, or its parameter estimates) (cf. Fig. 4) (Walker et al. 2003, Maier et al. 2016).

A main purpose of identifying uncertainties is to assist in uncertainty quantification and reduction. Many uncertainty typologies are not comprehensive, only classifying uncertainties across a single dimension or a subset of them (e.g., Roy and Oberkampf 2011, Beale and Lennon 2012, Uusitalo et al. 2015), and thus provide less direction on how to account for and reduce uncertainties. For example, identifying input uncertainty solely along the location dimension may make it difficult to distinguish whether sensitivity analysis or scenario analysis would be the more appropriate tool to quantify input sensitivities. While sensitivity analysis explores the probable space, or the current conditions, and is useful for capturing statistical-level uncertainty, scenario analysis explores the plausible space, or multiple plausible conditions that may happen, and is useful for capturing higher, scenario-level uncertainty (van der Sluijs et al. 2003, Maier et al. 2016). The uncertainty matrix and similar multi-dimensional frameworks help inform these decisions. For example, if we identify that uncertainty is located in the inputs (e.g., uncertainty in initial conditions, input data, and driving forces) and we classify its level as statistical uncertainty, we may opt to use a sensitivity analysis to perform a restricted exploration of the input space (i.e., test a limited subset of possible input values). If, instead, we classify the level of the input uncertainty to be scenario uncertainty, we could opt to perform a scenario analysis to capture a greater range of uncertainty. For a list of uncertainty quantification and reduction techniques that correspond to the dimensions of the uncertainty matrix, see van der Sluijs et al. (2003).

**Strategies to account for known uncertainties.**—Following identification, uncertainty can be accounted for by measuring prediction uncertainty (i.e., the output uncertainty corresponding to forecasting into non-analogue conditions [Beale and Lennon 2012]) and by quantifying the uncertainties and sensitivities that give rise to this prediction uncertainty (Dietze et al. 2018). Uncertainty quantification is a two-step process where first, we quantify the uncertainties in initial conditions, drivers, parameters, processes, and model structures, and second, we propagate their variabilities through the model (or ensemble) into the final prediction (Smith 2013). A simple example of uncertainty quantification and propagation is the confidence and prediction intervals of linear regression (Dietze 2017a). Often with uncertainty quantification, we perform uncertainty analysis, which is the process of attributing uncertainties from the final prediction to their appropriate source (Dietze 2017b).
Following the location dimension in the uncertainty matrix (cf. Fig. 4), we propose strategies that address the uncertainties within context, input, model, parameter, and output. First, context uncertainty, the uncertainty that manifests in the assumptions underlying the modelling process (e.g., input and model structure). Level is the degree of uncertainty and contains three categories: Statistical uncertainty, the lowest uncertainty, assumes the persistence of current conditions (probable conditions) and can be described in statistical terms; scenario uncertainty, medium-level uncertainty, assumes a range of likely conditions (plausible conditions) with no known probability of each occurring; and recognized ignorance, the highest level of known uncertainty, assumes many possible conditions without knowing functional relationships or statistical properties (possible conditions) (Walker et al. 2003, Maier et al. 2016). Nature, the last dimension of the matrix, classifies uncertainty as either due to a lack of human knowledge or due to natural variability inherent in the system. The matrix is primarily a qualitative tool where users can check boxes or provide an ordinal ranking of low, medium, or high in each cell. The output represents the total uncertainty caused by the uncertainties in the above locations (context, model, inputs, and parameters) that are propagated through the model.

Following the location dimension in the uncertainty matrix (cf. Fig. 4), we propose strategies that address the uncertainties within context, input, model, parameter, and output. First, context uncertainty, the uncertainty that manifests in the assumptions underlying the modelling process (Warmink et al. 2011), often arises due to an unclear purpose or an inappropriate scope for the analyses (Thompson and Warmink 2017). For example, context uncertainty could arise due to indecision regarding the appropriate spatiotemporal scale for a problem. To account for this type of uncertainty, we propose establishing a clear model purpose to help outline context uncertainties, as discussed in Section A, and to incorporate reasonable alternative framings to ensure a range of context uncertainties is included in the final prediction (Walker et al. 2003).

Second, parameter uncertainty, unlike context uncertainty, arises from the imprecision involved in the parameterization of different biological, physical, and chemical processes in the model (Bonan and Doney 2018). In practice, when accounting for parameter uncertainty, both the degree of its uncertainty and the system's

| Location         | Level          | Nature          |
|------------------|----------------|-----------------|
|                  | Statistical    | Scenario        | Recognized | Epistemic  | Variability |
|                  | Uncertainty    | Uncertainty     | Ignorance   | Uncertainty| Uncertainty|
|                  | Probable       | Plausible       | Possible   | Lack of    | Inherent    |
|                  | Conditions     | Conditions      | Conditions  | human knowledge | stochasticity|

Fig. 4. Uncertainty matrix: The uncertainty matrix, adapted from Walker et al. (2003), is a tool for identifying uncertainties using three dimensions: location, level, and nature. Location is where uncertainty is found within the modelling process (e.g., input and model structure). Level is the degree of uncertainty and contains three categories: Statistical uncertainty, the lowest uncertainty, assumes the persistence of current conditions (probable conditions) and can be described in statistical terms; scenario uncertainty, medium-level uncertainty, assumes a range of likely conditions (plausible conditions) with no known probability of each occurring; and recognized ignorance, the highest level of known uncertainty, assumes many possible conditions without knowing functional relationships or statistical properties (possible conditions) (Walker et al. 2003, Maier et al. 2016). Nature, the last dimension of the matrix, classifies uncertainty as either due to a lack of human knowledge or due to natural variability inherent in the system. The matrix is primarily a qualitative tool where users can check boxes or provide an ordinal ranking of low, medium, or high in each cell. The output represents the total uncertainty caused by the uncertainties in the above locations (context, model, inputs, and parameters) that are propagated through the model.
sensitivity to that parameter influence a model’s predictive performance. High parameter uncertainty combined with high sensitivity produces high prediction uncertainty, whereas low parameter uncertainty combined with low sensitivity produces low uncertainty (Fig. 5). By taking a product of the uncertainty and sensitivity for each parameter, each parameter’s relative contribution to the overall parameter uncertainty can be determined (Dietze 2017a). Various strategies are available to quantify parameter sensitivity and uncertainty such as local and global sensitivity analyses (e.g., Benke et al. 2008, Pianosi et al. 2016), generalized likelihood uncertainty estimation (GLUE) (e.g., Sathyamoorthy et al. 2014), Bayesian inference approaches (e.g., Chen et al. 2017), analytical approaches such as Kalman filters (e.g., Massoud et al. 2018), and Monte Carlo simulations (e.g., Zhang et al. 2015). When data are limited, expert opinion can also be adopted to help estimate uncertainty (Uusitalo et al. 2015), particularly within a Bayesian context (O’Hagan 2012, 2019).

Third, model structure uncertainty may arise from unwittingly excluding variables of influence, utilizing surrogate variables, and/or approximating functional forms (Ascough et al. 2008), whereas the fourth location uncertainty, input uncertainty, may arise from a lack of knowledge about, or inherent stochasticity in, initial conditions, forcing variables, and drivers (Walker et al. 2003). Both input and model structure uncertainties may have equal or more influence on predictions than parameter uncertainty (Lindenschmidt 2006, Dietze 2017b). Ensemble models, previously discussed in Section B, may help account for model structure uncertainty when ensembles are constructed with multiple, preferably independent, model structures (Wintle et al. 2003) and parameter and/or input

---

Fig. 5. The role of parameter uncertainty and sensitivity on output uncertainty: Model output uncertainty is determined by parameter uncertainty (beige distributions on x-axis) and parameter sensitivity (purple and pink lines). High parameter uncertainty (dark beige distribution), combined with high sensitivity (dotted purple line), produces the highest model output uncertainty, whereas low parameter uncertainty (light beige distribution) combined with low sensitivity (solid pink line) produces the lowest output uncertainty. Combinations of low uncertainty/high sensitivity (light beige distribution with the solid purple line) and high uncertainty/low sensitivity (dark beige distribution with dotted pink line) result in intermediate levels of model output uncertainty. Schematic adapted from Lal (W. Lal, unpublished report) and Loucks and van Beek (2017).
uncertainties when ensembles include different initial conditions, boundary conditions, and parameter estimates (Vrugt et al. 2008, Zhang et al. 2016). For more details on constructing ensemble models, see Dormann et al. (2018).

Finally, context, parameter, input, and model structure uncertainties amalgamate into output uncertainty (also referred to as prediction uncertainty). As prediction uncertainty increases with greater extrapolation (e.g., across time or space), uncertainty evaluations should include a measure of the speed at which that occurs (Dietze et al. 2018). A forecast horizon (Petchey et al. 2015), sometimes referred to as a forecast limit, quantifies how far the predictions of a model transfer across a dimension (e.g., time, space, phylogeny) before prediction uncertainty becomes unacceptably large. Included in this technique are a forecast proficiency, a measure of how good a forecast is, and a forecast proficiency threshold, a cutoff for when predictions are deemed not acceptable (Fig. 6A). The forecast proficiency threshold may be established when the desired properties of a prediction are established (cf. Stage A: Framing the Question), and is usually a value predetermined by the output of a null model (e.g., Massoud et al. 2018) but it can be set by researchers or stakeholders. The forecast horizon is the point at which the average

![Fig. 6. (A) Forecast horizon: The forecast horizon is the length of the interval between the simulation’s start and when the forecast proficiency intersects the forecast proficiency threshold. The forecast proficiency measures the quality of a forecast, and it declines across space, time, or any other dimension for which predictions are being generated. The inclusion of uncertainty (e.g., parameter uncertainty) produces a forecast horizon distribution, indicated by the yellow bands. The average forecast horizon measures the average interval across which predictions are deemed acceptable after accounting for sources of uncertainty. (B) Prediction uncertainty over time: Prediction uncertainty increases as we extrapolate across dimensions such as space or time but may differ in its level depending on the uncertainties of our context, models, parameters, and inputs. The prediction uncertainty can then be classified as one of three levels: probable (statistical uncertainty), plausible (scenario uncertainty), or possible (recognized ignorance) (Walker et al. 2003) (cf. Fig. 4). Statistical uncertainty is the lowest level of uncertainty, recognized ignorance is the greatest, and scenario uncertainty falls between. Here, scenarios are snapshots of plausible alternative conditions. Hence, when performing scenario analysis, we seek to include all, or at a minimum an appropriate range of conditions to capture what we deem plausible conditions (i.e., scenario uncertainty). Panel A adapted from Petchey et al. (2015); panel B adapted from Maier et al. (2016).]
forecast proficiency drops below the forecast proficiency threshold. The forecast horizon can, for example, be calculated by simulating dynamics of a system and comparing it to model outputs when uncertainties or stochasticity is added, or it can be a measure of declining predictability (e.g., using $R^2$) when models are fit to data and tested against out-of-sample data (see Petchey et al. [2015] for more details). Including different parameters, inputs, and models in these analyses yields a distribution of possible forecast horizons rather than a single value, providing a measure of prediction uncertainty. The forecast horizon is extendable to multiple dimensions and may be used to identify dimensions that are larger contributors to prediction uncertainty (e.g., Gavish et al. 2018).

High prediction uncertainty typically leads to shorter than desired forecast horizons. Prediction uncertainty is typically high when the level of uncertainty in contexts, parameters, models, or inputs falls into scenario uncertainty or recognized ignorance in the level dimension (Fig. 4). Here, to account for scenario uncertainty, projections based on alternative scenarios (sometimes referred to as “what-ifs”) may be adopted to explore various plausible scenarios (Fig. 6B). Alternative scenarios, which consist of different environmental, social, technological, or economic conditions, may take the form of alternative model formulations, input data, or both (Walker et al. 2003). These inputs scenarios and/or model structure scenarios are then incorporated into models or ensemble models to provide projections. In ecology, scenario analysis is typically adopted to project potential impacts of human-induced factors such as climate change (e.g., Wenger et al. 2013), harvesting (e.g., de-Miguel et al. 2014), land-use change (e.g., Visconti et al. 2011), and disease interventions (e.g., Trauer et al. 2016). To properly account for scenario uncertainty in model outputs, more than a single model or scenario should be considered, and optimally, all plausible scenario–model combinations should be explored (Suggitt et al. 2017). When the degree of uncertainty moves toward recognized ignorance, the highest degree of uncertainty classified in level, there is not enough knowledge to make informed predictions or projections. In such cases, models may still be used in an exploratory manner, for example, to test the likely resiliency of a system under different possible conditions (Maier et al. 2016), but care must be taken to state model limitations properly (Section A).

Strategies to reduce uncertainties where feasible.—After known uncertainties have been identified and accounted for, next we reduce them where feasible. Using the uncertainty matrix, we can separate reducible from irreducible uncertainties by distinguishing epistemic uncertainty, uncertainty due to a lack of knowledge, from natural variability, inherent variability such as environmental stochasticity (cf. Fig. 4). Epistemic uncertainty is theoretically resolvable by increasing our system knowledge, whereas natural variability cannot be eliminated (Marotzke 2018). For reducible uncertainties, the reduction is often based on increasing the signal-to-noise ratio (increasing our ability to detect a signal), which can be accomplished by (1) collecting more and/or better data, (2) improving the hypotheses embedded in our models, and (3) improving (or selecting better) techniques for extracting and assimilating information from available data (Liu and Gupta 2007). Note that many strategies proposed in previous sections are already designed to help reduce uncertainties. For example, the adaptive modelling framework outlined in Model-Building and Testing is a strategy that encourages the inclusion of additional experiments and data to iteratively and systematically improve predictive models and reduce uncertainties—thus addressing points 1 and 2 above. There is no guarantee that collecting more data will increase the signal-to-noise ratio but iteratively incorporating new data into a model typically leads to declines in parameter uncertainty (Dietze 2017a) and may lead to new or revised hypotheses in the form of different model structures. To reduce uncertainty in already available data, we can adopt better techniques to extract and assimilate information (point 3 above): techniques such as dimensionality reduction and Bayesian inference, as referenced in Stage B: Model-Building and Testing. Both techniques reduce uncertainty: dimensionality reduction by reducing the number of uncertain variables used as inputs (Mahecha et al. 2007) and Bayesian models by incorporating
previous posterior distributions as priors when new information becomes available (Dietze 2017a). Finally, uncertainty analysis, which attributes the uncertainty in the response variable to its different inputs, can assist in uncertainty reduction efforts by directing resources toward the parameters and processes that contribute most to a model’s uncertainty (Dietze 2017a). Once uncertainty has been minimized, uncertainty identification, quantification, and propagation may be performed again.

HOW ACHIEVING GOALS MAKES PREDICTIVE MODELLING “ART”

The three stages of the predictive modelling process (Framing the Question, Model-Building and Testing, and Uncertainty Evaluation) and their corresponding goals provide a framework for generating and communicating predictions. By meeting these goals through implementing the strategies proposed above, we can make predictive modelling more accurate, reliable, and transparent.

In all stages of the Predictive Modelling Framework, achieving the proposed goals may lead to improved transparency and hence reproducibility. In Framing the Question, by defining the model purpose, we help prevent misinterpretations of the model’s goals and limitations (Grimm et al. 2010) and we highlight why aspects of reality are included while others are ignored (Grimm et al. 2006). Similarly, by establishing the predictions’ expected properties, we set clear expectations for a model’s performance, helping to define what is a “good prediction.” In Model-Building and Testing, by creating appropriate suites of candidate models, we produce a range of predictions (Lovenduski and Bonan 2017) showcasing prediction uncertainty, and by explaining and justifying decisions regarding model type, model structures, selected variables, model assumptions, and expectations, we encourage reproducibility. In Uncertainty Evaluation, by identifying uncertainties, the source and amount of uncertainty present in our models are communicated, and by accounting for uncertainties, we quantify and apportion uncertainty to appropriate inputs, highlighting which uncertainties dominate the system. Particularly, by adopting a forecast horizon, we create a measure of prediction uncertainty enabling a proper means to assess the accuracy and precision of a model (Gavish et al. 2018).

In addition to increasing transparency, the goals outlined throughout our framework also lead to improvements in accuracy and reliability. After defining the model purpose and establishing the expected properties of the predictions, we can improve accuracy and reliability by addressing factors that cause discrepancies between the desired and observed prediction properties, which are often caused by either a lack of knowledge, limited data availability (Alexandridis et al. 2017), or an inappropriate model being utilized (Carlson et al. 2017). During Model-Building and Testing, by choosing the most appropriate model types, we help ensure the research question and available data are complementary, such that the predictions generated fit the purpose of the study. Next, by creating appropriate suites of candidate models, we capture a range of different hypotheses and processes, potentially leading to reductions in error. Then, by selecting the most appropriate final model, we calculate predictive performance on independent data and typically select models exhibiting higher accuracy and lower uncertainty. In the final stage, Uncertainty Evaluation, by reducing uncertainties through increased data collection, improved knowledge of key processes, and enhanced techniques for information extraction, we can better detect signals in the data, thereby increasing precision.

CONCLUSIONS

Constructing a predictive model from scratch may be a daunting task. However, models on average outperform human judgment for outcomes related to environmental decision-making (Czaika and Selin 2017, Holden and Ellner 2019). Therefore, it is imperative to incorporate models in both applied and theoretical research. It is our hope that the Predictive Modelling Framework outlined here and its corresponding goals will help guide researchers through the process of building effective predictive models. By achieving the key goals outlined in Framing the Question, Model-Building and Testing, and Uncertainty Evaluation, we increase the accuracy, reliability, and transparency of the overall
process of making ecological predictions. We emphasize that not all goals need always be completed to obtain objectives, but taking steps toward formalizing our thinking, even if only within subsets of the framework, will improve how we build and use predictive models. Ultimately, by making predictive modelling more ART, we create better forecasts, foster improved scientific communication, and produce overall better science.

ACKNOWLEDGMENTS

The authors thank Carina Firkowski, Andrew Chin, Emily Chenery, Stephanie Penk, Juan Vargas Soto, Alexander Nascou, Luke Rogers, and Cedric B. Hunter for their helpful comments and insightful discussions. Funding was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant to Péter K. Molnár (#06301) and Marie-Josée Fortin (#5134), NSERC Canada Research Chair (CRC) to Marie-Josée Fortin, Canadian Foundation for Innovation (CFI) John R. Evans Leader Funds (#35341) to Péter K. Molnár, and Ministry of Research, Innovation, and Science (MRIS) Ontario Research Funds to Péter K. Molnár. Korryn Bodner, Péter K. Molnár, and Marie-Josée Fortin conceived of the paper; Korryn Bodner wrote the paper; and all co-authors contributed to paper review and revision.

LITERATURE CITED

Abrahms, B., et al. 2019. Dynamic ensemble models to predict distributions and anthropogenic risk exposure for highly mobile species. Diversity and Distributions 25:1182–1193.

Addison, P. F. E., L. Rumpff, S. S. Bau, J. M. Carey, Y. E. Chee, F. C. Jarrad, M. F. McBride, and M. A. Burgman. 2013. Practical solutions for making models indispensable in conservation decision-making. Diversity and Distributions 19:490–502.

Aho, K., D. Derryberry, and T. Peterson. 2016. A graphical framework for model selection criteria and significance tests (refutation, confirmation and ecology). Methods in Ecology and Evolution 8:47–56.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. Pages 267–281 in B. N. Petrov and F. Csáki, editors. Second International Symposium on Information Theory. Akadémiai Kiadó, Budapest, Hungary.

Alexandridis, N., J. M. Dambacher, F. Jean, N. Desroy, and C. Bacher. 2017. Qualitative modelling of functional relationships in marine benthic communities. Ecological Modelling 360:300–312.

Asakura, T., K. Sakata, Y. Date, and J. Kikuchi. 2018. Regional feature extraction of various fishes based on chemical and microbial variable selection using machine learning. Analytical Methods 10:2160–2168.

Ascough, J. C. II, H. R. Maier, J. K. Ravalico, and M. W. Strudley. 2008. Future research challenges for incorporation of uncertainty in environmental and ecological decision-making. Ecological Modelling 219:383–399.

Bailly, D., F. A. S. Cassemi, C. S. Agostinho, E. E. Marques, and A. A. Agostinho. 2014. The metabolic theory of ecology convincingly explains the latitudinal diversity gradient of Neotropical freshwater fish. Ecology 95:553–562.

Beale, C. M., and J. J. Lennon. 2012. Incorporating uncertainty in predictive species distribution modelling. Philosophical Transactions of the Royal Society B: Biological Sciences 367:247–258.

Benke, K. K., K. E. Lowell, and A. J. Hamilton. 2008. Parameter uncertainty, sensitivity analysis and prediction error in a water-balance hydrological model. Mathematical and Computer Modelling 47:1134–1149.

Bonan, G. B., and S. C. Doney. 2018. Climate, ecosystems, and planetary futures: the challenge to predict life in Earth system models. Science 359: eaam8328.

Borregaard, M. K., et al. 2017. Oceanic island biogeography through the lens of the general dynamic model: assessment and prospect. Biological Reviews 92:830–853.

Breiner, F. T., A. Guisan, A. Bergamini, and M. P. Nobis. 2015. Overcoming limitations of modelling rare species by using ensembles of small models. Methods in Ecology and Evolution 6:1210–1218.

Briscoe, N. J., et al. 2019. Forecasting species range dynamics with process-explicit models: matching methods to applications. Ecology Letters 22:1940–1956.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Second edition. Springer-Verlag, New York, New York, USA.

Carlson, C. J., et al. 2017. Parasite biodiversity faces extinction and redistribution in a changing climate. Science Advances 3:e1602422.

Carpenter, S. R. 2002. Ecological futures: building an ecology of the long now. Ecology 83:2069–2083.

Chen, D., Y. Xue, D. Fink, S. Chen, and C. P. Gomes. 2017. Deep multi-species embedding. preprint arXiv:1609.09353.
Clark, J. S., et al. 2001. Ecological forecasts: an emerging imperative. Science 293:657–660.

Czaika, E., and N. E. Selin. 2017. Model use in sustainability policy making: an experimental study. Environmental Modelling and Software 98:54–62.

De-Miguel, S., J. A. Bonet, T. Pukkala, and J. Martínez de Aragón. 2014. Impact of forest management intensity on landscape-level mushroom productivity: a regional model-based scenario analysis. Forest Ecology and Management 330:218–227.

Dick, E. J., S. Beyer, M. Mangel, and S. Ralston. 2017. A meta-analysis of fecundity in rockfishes (genus Sebastes). Fisheries Research 187:73–85.

Dietze, M. C. 2017a. Prediction in ecology: a first-principles framework. Ecological Applications 27:2048–2060.

Dietze, M. C. 2017b. Ecological forecasting. First edition. Princeton University Press, Princeton, New Jersey, USA.

Dietze, M. C., et al. 2018. Iterative near-term ecological forecasting: needs, opportunities, and challenges. Proceedings of the National Academy of Sciences of the United States of America 115:1424–1432.

Dormann, C. F., et al. 2011. Correlation and process in species distribution models: bridging a dichotomy. Journal of Biogeography 39:2119–2131.

Dormann, C. F., et al. 2018. Model averaging in ecology: a review of Bayesian, information-theoretic, and tactical approaches for predictive inference. Ecological Monographs 88:485–504.

Evans, M. R. 2012. Modelling ecological systems in a changing world. Philosophical Transactions of the Royal Society B 367:181–190.

Fassnacht, F. E., H. Latifi, A. Ghosh, P. K. Joshi, and B. Koch. 2014. Assessing the potential of hyperspectral imagery to map bark beetle-induced tree mortality. Remote Sensing of Environment 140:533–548.

García-Díaz, P., T. A. A. Prowse, D. P. Anderson, M. Lurgi, R. N. Binny, and P. Cassey. 2019. A concise guide to developing and using quantitative models in conservation management. Conservation Science and Practice 1:e11.

Gavish, Y., J. O’Connell, and T. G. Benton. 2018. Quantifying and modelling decay in forecast proficiency indicates the limits of transferability in land-cover classification. Methods in Ecology and Evolution 9:235–244.

Gelman, A., J. Hwang, and A. Vehtari. 2014. Understanding predictive information criteria for Bayesian models. Statistics and Computing 24:997–1016.

Gholizadeh, H., J. A. Gamon, A. I. Zygelbaum, R. Wang, A. K. Schweiger, and J. Cavender-Bares. 2018. Remote sensing of biodiversity: Soil correction and data dimension reduction methods improve assessment of α-diversity (species richness) in prairie ecosystems. Remote Sensing of Environment 206:240–253.

Grimm, V., et al. 2006. A standard protocol for describing individual-based and agent-based models. Ecological Modelling 198:115–126.

Grimm, V., U. Berger, D. L. DeAngelis, J. G. Polhill, J. Giske, and S. F. Railsback. 2010. The ODD protocol: a review and first update. Ecological Modelling 221:2760–2768.

Haefner, J. W. 1996. Modeling biological systems: principles and applications. Chapman and Hall, New York, New York, USA.

Hastie, T., R. Tibshirani, and J. H. Friedman. 2009. The elements of statistical learning: data mining, inference, and prediction. Second edition. Springer, New York, New York, USA.

Hilborn, R., and M. Mangel. 1997. The ecological detective: confronting models with data. Princeton University Press, Princeton, New Jersey, USA.

Holden, M. H., and S. P. Ellner. 2019. Human judgment vs. quantitative models for the management of ecological resources. Ecological Applications 26:1553–1565.

Hooten, M. B., and N. T. Hobbs. 2015. A guide to Bayesian model selection for ecologists. Ecological Monographs 85:3–28.

Isaksson, A., M. Wallman, H. Göransson, and M. G. Gustafsson. 2008. Cross-validation and bootstrapping are unreliable in small sample classification. Pattern Recognition Letters 29:1960–1965.

Jasper, J. R., M. Short, C. Shelden, and W. S. Grant. 2018. Hierarchical Bayesian estimation of unobserved salmon passage through weirs. Canadian Journal of Fisheries and Aquatic Sciences 75:1151–1159.

Johnson, J. B., and K. S. Omland. 2004. Model selection in ecology and evolution. Trends in Ecology and Evolution 19:101–108.

Kindsvater, H. K., N. K. Dulvy, C. Horswill, M. J. JuanJordá, M. Mangel, and J. Matthiopoulos. 2018. Overcoming the data crisis in biodiversity conservation. Trends in Ecology and Evolution 33:676–688.

Krstaic, D., L. J. Buturovic, D. E. Leahy, and S. Thomas. 2014. Cross-validation pitfalls when selecting and assessing regression and classification models. Journal of Cheminformatics 6:10.

Lindenschmidt, K. 2006. Parameter, data input and structural uncertainty propagating through coupled models in a river water quality modelling system. International Congress on Environmental Modelling and Software, Burlington, Vermont, USA.
Link, W. A., and J. R. Sauer. 2016. Bayesian cross-validation for model evaluation and selection, with application to the North American Breeding Bird Survey. Ecology 97:1746–1758.

Liu, Y., and H. V. Gupta. 2007. Uncertainty in hydrologic modeling: toward an integrated data assimilation framework. Water Resources Research 43: W07401.

Loucks, D. P., and E. van Beek. 2017. System sensitivity and uncertainty analysis. Pages 331–374 in Water resource systems planning and management. Springer, Cham, Switzerland.

Lovenduski, N. S., and G. B. Bonan. 2017. Reducing uncertainty in projections of terrestrial carbon uptake. Environmental Research Letters 12:044020.

Lowe, R., et al. 2014. Dengue outlook for the World Cup in Brazil: an early warning model framework driven by real-time seasonal climate forecasts. Lancet Infectious Diseases 14:619–626.

Macintyre, P. D., A. Van Niekerk, M. P. Dobrowolski, J. L. Tsakalos, and L. Mucina. 2018. Impact of ecological redundancy on the performance of machine learning classifiers in vegetation mapping. Ecology and Evolution 8:6728–6737.

Mahecha, M. D., A. Martinez, G. Lischied, and E. Beck. 2007. Nonlinear dimensionality reduction: alternative ordination approaches for extracting and visualizing biodiversity patterns in tropical montane forest vegetation data. Ecological Informatics 2:138–149.

Maier, H. R., J. H. A. Guillaume, H. van Delden, G. A. Riddell, M. Haasnoot, and J. H. Kwakkel. 2016. An uncertain future, deep uncertainty, scenarios, robustness and adaptation: How do they fit together? Environmental Modelling and Software 81:154–164.

Marotzke, J. 2018. Quantifying the irreducible uncertainty in near-term climate projections. Wiley Interdisciplinary Reviews: Climate Change 10:e656.

Massoud, E. C., J. Huisman, E. Benincá, M. C. Dietze, W. Bouten, and J. A. Vrugt. 2018. Probing the limits of predictability: data assimilation of chaotic dynamics in complex food webs. Ecology Letters 21:93–103.

Milner-Gulland, E. J., and K. Shea. 2017. Embracing uncertainty in applied ecology. Journal of Applied Ecology 54:2063–2068.

Mohri, M., A. Rostamizadeh, and A. Talwalkar. 2012. Foundations of machine learning. MIT Press, Cambridge, Massachusetts, USA.

Molnár, P. K., J. P. Skrabulis, K. A. Altman, and T. R. Raffel. 2017. Thermal performance curves and the metabolic theory of ecology: a practical guide to models and experiments for parasitologists. Journal of Parasitology 103:423–439.

Mosteller, F., and J. Tukey. 1977. Data analysis and regression: a second course in statistics. In Addison-Wesley series in behavioral science: Quantitative methods. First edition. Addison-Wesley, Reading, Massachusetts, USA.

Mouquet, N., et al. 2015. Predictive ecology in a changing world. Journal of Applied Ecology 52:1293–1310.

O’Hagan, A. 2012. Probabilistic uncertainty specification: overview, elaboration techniques and their application to a mechanistic model of carbon flux. Environmental Modelling and Software 36:35–48.

O’Hagan, A. 2019. Expert knowledge elicitation: subjective but scientific. American Statistician 73:69–81.

Olden, J. D., J. J. Lawler, and N. L. Poff. 2008. Machine learning methods without tears: a primer for ecologists. Quarterly Review of Biology 83:171–193.

Outhwaite, C. L., R. E. Chandler, G. D. Powney, B. Col- len, R. D. Gregory, and N. J. B. Isaac. 2018. Prior specification in Bayesian occupancy modelling improves analysis of species occurrence data. Ecological Indicators 93:333–343.

Petchey, O. L., et al. 2015. The ecological forecast horizon, and examples of its uses and determinants. Ecology Letters 18:597–611.

Pionosi, F., K. Beven, J. Freer, J. W. Hall, J. Rougier, D. B. Stephenson, and T. Wagener. 2016. Sensitivity analysis of environmental models: a systematic review with practical workflow. Environmental Modelling and Software 79:214–232.

Pomara, L. Y., O. E. Ledee, K. J. Martin, and B. Zucker- berg. 2014. Demographic consequences of climate change and land cover help explain a history of extirpations and range contraction in a declining snake species. Global Change Biology 20:2087–2099.

Refsgaard, J. C., J. P. van der Sluijs, A. L. Højberg, and P. A. Vanrolleghem. 2007. Uncertainty in the environmental modelling process: a framework and guidance. Environmental Modelling and Software 22:1543–1556.

Restif, O., et al. 2012. Model-guided fieldwork: practical guidelines for multidisciplinary research on wildlife ecological and epidemiological dynamics. Ecology Letters 15:1083–1094.

Roy, C. J., and W. L. Oberkampf. 2011. A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing. Computer Methods in Applied Mechanics and Engineering 200:2131–2144.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. 2008. Global sensitivity analysis: the primer. John Wiley & Sons, New York, New York, USA.
Sathyamoorthy, S., R. M. Vogel, S. C. Chapra, and C. A. Ramsburg. 2014. Uncertainty and sensitivity analyses using GLUE when modeling inhibition and pharmaceutical cometabolism during nitrification. Environmental Modelling and Software 60:219–227.

Schindler, D. E., and R. Hilborn. 2015. Prediction, precaution, and policy under global change. Science 347:953–954.

Smith, R. C. 2013. Uncertainty quantification: theory, implementation, and applications. SIAM Computational Science & Engineering Series, Philadelphia, Pennsylvania, USA.

Speirs-Bridge, A., F. Fidler, M. McBride, L. Flander, G. Cumming, and M. Burgman. 2010. Reducing overconfidence in the interval judgments of experts. Risk Analysis 30:512–523.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society Series B 64:583–639.

Stone, M. 1977. An asymptotic equivalence of choice of model by cross-validation and Akaike’s criterion. Journal of the Royal Statistical Society Series B 39:44–47.

Sugitt, A. J., et al. 2017. Conducting robust ecological analyses with climate data. Oikos 126:1533–1541.

Taper, M. L., D. F. Staples, and B. B. Shepard. 2008. Model structure adequacy analysis: selecting models on the basis of their ability to answer scientific questions. Synthese 163:357–370.

Taylor, M. L., and A. Hastings. 2004. Finding optimal control strategies for invasive species: a density-structured model for Spartina alterniflora. Journal of Applied Ecology 41:1049–1057.

Thomas, M. K., S. Fontana, M. Reyes, M. Kehoe, and F. Pomati. 2018. The predictability of a lake phytoplankton community, over time-scales of hours to years. Ecology Letters 21:619–628.

Thompson, M., and J. J. Warmink. 2017. Natural hazard modeling and uncertainty analysis. Pages 11–19 in K. Riley, P. Webley, and M. Thompson, editors. Natural hazard uncertainty assessment: modeling and decision support. First edition. American Geophysical Union, Washington, D.C., USA.

Torrecilla, J. L., and J. Romo. 2018. Data learning from big data. Statistics and Probability Letters 136:15–19.

Tracy, J. L., A. Trabucco, A. M. Lawing, J. T. Giermakowski, M. Thakerian, G. M. Drus, and R. N. Coulson. 2018. Random subset feature selection for ecological niche models of wildfire activity in western North America. Ecological Modelling 383:52–68.

Trauer, J. M., J. T. Denholm, S. Waseem, R. Ragonnet, and E. S. McBryde. 2016. Scenario analysis for programmatic tuberculosis control in Western Province, Papua New Guinea. American Journal of Epidemiology 183:1138–1148.

Urban, M. C., et al. 2016. Improving the forecast for biodiversity under climate change. Science 353: aad8466.

Uribe-Rivera, D. E., C. Soto-Azat, A. Valenzuela-Sánchez, G. Bizama, J. A. Simonetti, and P. Pliscoff. 2017. Dispersal and extrapolation on the accuracy of temporal predictions from distribution models for the Darwin’s frog. Ecological Applications 27:1633–1645.

Usaitis, L., A. Lehikoinen, I. Helle, and K. Myrberg. 2015. An overview of methods to evaluate uncertainty of deterministic models in decision support. Environmental Modelling and Software 63:24–31.

van der Sluijs, J. P., P. H. M. Janssen, A. C. Petersen, P. Kloprogge, J. S. Risbey, W. Tuinstra, and J. R. Ravetz. 2003. RIVM/MNP guidance for uncertainty assessment and communication: tool catalogue for uncertainty assessment. Volume 3. Copernicus Institute & RIVM, Utrecht, The Netherlands.

van Oijen, M. 2017. Bayesian methods for quantifying and reducing uncertainty and error in forest models. Current Forestry Reports 3:269–280.

Vebtaris, A., A. Gelman, and J. Gabry. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing 27:1413–1432.

Visconti, P., R. L. Pressley, D. Giorgini, L. Maiorano, M. Bakkenes, L. Boitani, R. Alkemade, A. Falcucci, F. Chiozza, and C. Rondinini. 2011. Future hotspots of terrestrial mammal loss. Philosophical Transactions of the Royal Society B 366:2693–2702.

Vrugt, J. A., C. G. H. Diks, and M. P. Clark. 2008. Ensemble Bayesian model averaging using Markov Chain Monte Carlo sampling. Environmental Fluid Mechanics 8:579–595.

Walker, W. E., P. Harremoës, J. Rotmans, J. P. van der Sluijs, M. B. A. van Asselt, P. Janssen, and M. P. K. von Krauss. 2003. Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. Integrated Assessment 4:5–17.

Warmink, J. J., H. Van der Klis, M. J. Booij, and S. J. M. H. Hulscher. 2011. Identification and quantification of uncertainties in a hydrodynamic river model using expert opinions. Water Resources Management 25:601–622.

Watanabe, S. 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. Journal of Machine Learning Research 11:3571–3594.
Wenger, S. J., and J. D. Olden. 2012. Assessing transferability of ecological models: an underappreciated aspect of statistical validation. Methods in Ecology and Evolution 3:260–267.

Wenger, S. J., N. A. Som, D. C. Dauwalter, D. J. Isaak, H. M. Neville, C. H. Luce, J. B. Dunham, M. K. Young, K. D. Fausch, and B. E. Rieman. 2013. Probabilistic accounting of uncertainty in forecasts of species distributions under climate change. Global Change Biology 19:3343–3354.

Wintle, B. A., M. A. McCarthy, C. T. Volinsky, and R. P. Kavanagh. 2003. The use of Bayesian model averaging to better represent uncertainty in ecological models. Conservation Biology 17:1579–1590.

Zhang, C., Y. Chen, and Y. Ren. 2015. Assessing uncertainty of a multispecies size-spectrum model resulting from process and observation errors. ICES Journal of Marine Science 72:2223–2233.

Zhang, J. L., Y. P. Li, G. H. Huang, C. X. Wang, and G. H. Cheng. 2016. Evaluation of uncertainties in input data and parameters of a hydrological model using a Bayesian framework: a case study of a snowmelt–precipitation-driven watershed. Journal of Hydrometeorology 17:2333–2350.