# The priority of prediction in ecological understanding

**Jeff E. Houlahan, Shawn T. McKinney, T. Michael Anderson and Brian J. McGill**

*J. E. Houlahan (http://orcid.org/0000-0003-3491-353X) (jeffhoul@unb.ca), Dept of Biology, 100 Tucker Park Road, Univ. of New Brunswick, Saint John, NB, E2L 4L5, Canada. – S. T. McKinney, Univ. of Maine, Maine Cooperative Fish and Wildlife Research Unit, Orono, Maine, United States. – T. M. Anderson, Wake Forest University, Dept of Biology, Winston-Salem, NC, USA. – B. J. McGill, School of Biology and Ecology, Mitchell Center for Sustainability Solutions, Univ. of Maine, Orono, ME, USA.*

The objective of science is to understand the natural world; we argue that prediction is the only way to demonstrate scientific understanding, implying that prediction should be a fundamental aspect of all scientific disciplines. Reproducibility is an essential requirement of good science and arises from the ability to develop models that make accurate predictions on new data. Ecology, however, with a few exceptions, has abandoned prediction as a central focus and faces its own crisis of reproducibility. Models are where ecological understanding is stored and they are the source of all predictions – no prediction is possible without a model of the world. Models can be improved in three ways: model variables, functional relationships among dependent and independent variables, and in parameter estimates. Ecologists rarely test to assess whether new models have made advances by identifying new and important variables, elucidating functional relationships, or improving parameter estimates. Without these tests it is difficult to know if we understand more today than we did yesterday. A new commitment to prediction in ecology would lead to, among other things, more mature (i.e. quantitative) hypotheses, prioritization of modeling techniques that are more appropriate for prediction (e.g. using continuous independent variables rather than categorical) and, ultimately, advancement towards a more general understanding of the natural world.

**Synthesis**

Ecology, with a few exceptions, has abandoned prediction and therefore the ability to demonstrate understanding. Here we address how this has inhibited progress in ecology and explore how a renewed focus on prediction would benefit ecologists. The lack of emphasis on prediction has resulted in a discipline that tests qualitative, imprecise hypotheses with little concern for whether the results are generalizable beyond where and when the data were collected. A renewed commitment to prediction would allow ecologists to address critical questions about the generalizability of our results and the progress we are making towards understanding the natural world.

The objective of science is to understand the natural world. Prediction is the only way to demonstrate scientific understanding. These two assertions should guide all scientific disciplines including ecology, but there is little evidence in the recent ecological literature that prediction is playing anything but a peripheral role (Hooten and Hobbs 2015).

Since the Open Science Collaboration (2015) published its report on reproducibility in psychology there has been heightened concern about the reliability of scientific findings in psychology and, more broadly, in science. Reproducibility is, at its essence, about how well a model obtained from one study can be applied to independent data – usually those data will have been collected at a different time or place and maybe even in different ways. Ecology is not immune to this problem and while there is no general panacea, a renewed commitment to prediction would mitigate many of the issues around reproducibility.

In this paper we discuss the fundamental importance of evaluating models based on their predictive abilities.

We do not suggest that this is a novel concept (indeed, it is the essence of the scientific method), but rather that the critical link between prediction and understanding has not been widely acknowledged. Even Rob Peter's paean to prediction 'A critique for ecology' did not make a direct link between prediction and understanding (Peters 1991). The result is that prediction is not at the core of today's ecological research, and further, that this omission acts to retard the growth of ecology. We provide suggestions for raising the status of prediction in ecology and discuss important questions that will need to be addressed to accomplish this. Our hope is that this paper stimulates discussion on the role of prediction in ecology and leads to increased emphasis on predictive testing.

## Prediction and scientific understanding

Prediction is often used as a synonym for forecasting (i.e. predicting the future) but here we define prediction more

broadly – a prediction is any statement about what an unknown quantity or state was, is, or will be, based solely on a putative understanding of how natural systems work. We define scientific understanding as a mechanistic understanding of how natural systems operate.

Predictions arise from theory – a scientist's proposed description of how natural systems work. Without theories and the models that arise from them there are no predictions. Here again, we use a very broad definition of 'model' – a model is any description of how the natural world might work. Models can take many forms – they can be verbal descriptions, physical constructs, logical relationships or mathematical equations, but any potentially useful model must make predictions about some unknown state of the natural world. For example, the statement that large areas contain more species than small areas is a model of how the world works and it allows us to make predictions about the relative number of species in different sized areas.

$$\text{species richness} = 1.43 \times \text{area}^{0.20}$$

is also a model of how the world works with respect to patch size and species richness but one with more information content that makes more precise predictions about the number of species that will be found in a patch of a particular size.

Any scientific claim that research has increased our understanding of the natural world implies that the research has resulted in a model that is a better representation of how natural systems operate than previous models. But how do we demonstrate the claim of increased understanding? Only by making better predictions. That is the foundation on which this paper stands.

One key point to emphasize is that prediction is necessary for 'demonstrating' understanding not 'acquiring' understanding. However, claims of understanding shouldn't be accepted without demonstration. There is no doubt that we can learn how the natural world works from descriptive studies – it would be difficult to make the argument that mapping the human genome has not contributed to scientific understanding – but the understanding from this kind of work can't be demonstrated without making predictions to new data.

## Prediction and ecology

Many disciplines, including remote sensing (Congalton and Green 2009), sound recognition (Jurafsky and Martin 2008), financial forecasting (Zhang 2009), disease prognosis (Taktak and Fisher 2007), and meteorology (LeTreut 1995) recognize the importance of assessing their understanding with prediction. However, there are few examples in ecology of repeated and rigorous tests of ecological models on new data.

A hallmark of ecological research is that we test coarse hypotheses that have relatively low information content. Connell's (1961) paper on marine intertidal zonation is a classic in ecology but what were the take-home messages of Connell and how have they contributed to our understanding of how the world works? The key findings are, 1) the distribution of *Chthamalus* was affected by competition with *Balanus*, 2) predation by starfish reduced the effect of *Balanus* on *Chthamalus* (presumably because starfish preferred to eat *Balanus*), 3) zonation occurred because *Chthamalus* could establish higher on the shoreline than *Balanus* and avoid competition with *Balanus*. These are all qualitative statements with relatively low information content but of greater concern is that we don't know how well these findings predict general patterns for these taxa in intertidal systems let alone providing general understanding about how ecological systems work.

It is not that concern with prediction has been absent in ecology. One of ecology's greatest success stories – modeling the relationship between aquatic nutrient levels and primary productivity – combined theoretical predictive models (Vollenweider and Dillon 1974), large-scale experiments that demonstrated causality (Schindler 1974), and observational data (Sakamoto 1966, Dillon and Rigler 1975) to demonstrate generality. None of this required sophisticated statistical techniques – the Vollenweider models were mathematically simple (requiring only data on loading, lake depth and flushing rate), the models examining the relationship between phosphorus and chlorophyll a were simple linear regressions and the predictions in the whole-lake manipulation were demonstrated with a photograph. Each piece was compelling because they made risky predictions that were confirmed. This understanding has resulted in the development of public policy in many countries around nutrient inputs. More recently, Bahn and McGill (2013) and Wenger and Olden (2012) explicitly address the issue of assessing the transferability of habitat occupancy models. Bahn and McGill demonstrated that performance of models predicting bird abundance or occupancy declined as training and test data became increasingly independent and as the information content of predictions increased. Wenger and Olden (2012) developed models to predict the presence/absence of two invasive fish species (brook and brown trout) in US streams. They spatially partitioned their training and test data so that training and test data were not only independent but from different geographic regions and built models that had very good predictive power on the data the models were built from, but much poorer predictive power for areas from different geographic regions.

The research done on fundamental ecological questions such as 'How does species richness affect productivity?' and 'How does population density affect population fluctuations?' are much more representative of the state of ecological research. These questions have been studied intensively for more than 50 years but there isn't even a clear consensus about how species richness affects productivity (Mittelbach et al. 2001, Whitaker 2010, Adler et al. 2011) or whether population density is an important driver of population fluctuations (Inchausti and Halley 2001) let alone quantitative understanding of these relationships. A classic paper by Tilman and Downing (1994) examined the diversity–productivity relationship and used an experimental design that would have allowed for quantitative predictions, but the authors chose to interpret their results qualitatively. The quantitative predictions that arise from their models have never been tested on independent data.

It has been suggested that prediction matters more in applied than basic research because there is some practical application (Freckleton 2004). A 2012 issue of Philosophical Transaction of the Royal Society (B) – Predictive ecology: systems approaches – did an admirable job of presenting a case for increased emphasis on prediction in ecology (Evans et al. 2012, Evans 2012, Grimm and Railsback 2012) but even here the explicit focus was on the applied value of prediction (Evans 2012, Sutherland and Freckleton 2012). We have often heard bright, accomplished ecologists say that prediction is important for applied questions but relatively less important for basic questions where our primary concern is understanding; as if, prediction and understanding are separate and independent concepts. In fact, prediction should have a central role in all ecological research and be recognised as inextricably intertwined with understanding.

Our recommendations are consistent with a standard interpretation of the scientific method but they are not consistent with how the scientific method is generally practiced in ecology. For example, they fit comfortably into a Popperian framework. Popper insisted that the difference between science and non-science was the existence of falsifiable predictions that follow deductively from a hypothesis. He also asserted that risky predictions containing a lot of information are better than safe predictions containing little information. What often gets lost in the focus on Popperian falsification is that Popper understood very well that the objective of science was NOT to identify all the incorrect hypotheses/models – it was to identify the correct ones. Popper believed that you could demonstrate a lack of understanding with a single falsifying event but he also believed that to convincingly demonstrate understanding one must repeatedly make correct and information-rich predictions. In ecology, Popper's position on falsification via very strong tests has mutated into falsification via the weak test of failing to reject the null and his commitment to repeated tests is routinely ignored. So, we believe Popper would have said that demonstrating understanding is a slow and iterative process that relies exclusively on correct and risky predictions.

## Why do not ecologists predict more?

One key reason is institutional – funding agencies, academic journals and research institutions place a high priority on novelty. Indeed, both NSF and NSERC explicitly identify novelty as a key criterion in funding decisions. There is little room for applying already developed models to new datasets in a research setting that places a high priority on novelty. Similarly, the top journals such as Nature and Science publish only novel research and also explicitly identify novelty as a criterion in evaluating manuscripts. Since academic positions rest primarily on a scientist's ability to publish, to publish in high impact journals and to secure research funding, novelty becomes a critical factor in acquiring and being successful at, a research position. The result is that ecologists have disincentives to test models on new data.

A second reason is that most ecological research is not held to rigorous public scrutiny because it is not seen as being relevant to the general public. When scientific research has important health or economic implications the impetus for demonstrating real understanding is strong – when we get it wrong on cures for cancer the public is aware and unhappy. When we get it wrong on the effect of intra-specific competition on population fluctuations few people care. So, ecological research that addresses practical problems (e.g. harvesting a fish stock sustainably or conserving the right type and amount of habitat to protect a SARA listed species) is more often compelled to show that their models work. That said, ecology's standard for demonstrating understanding should not be set by public perception.

A third reason is that ecologists do not predict because when we do it reveals how little we understand about the natural world. Nature is complex and does not lend itself to accurate and precise predictions. It is widely believed that the dynamics of ecological systems are inherently complex – that the abundances in time and space of living organisms depend on an enormous number of variables and that these relationships are often non-linear – and as a result they are always going to be difficult to predict. This is an understandable but unacceptable reason for not making predictions. It may be that ecologists will rarely be able to make accurate and precise predictions but, if this is true, we need to know it.

## Model selection and parsimony

Models arise from theory and differ in essentially just three ways: 1) variables, 2) functional relationships between independent and dependent variables, and 3) parameter estimates. Model selection involves choosing among models that differ in the variables, functional relationships and parameter estimates.

If prediction were central in ecology model selection would be done very differently. First, model selection would involve out-of sample data (i.e. data that were not used to inform the model being tested). Second, the quality of the best current model would be evaluated based on the difference between predicted values and out-of-sample observations. Common statistics used to assess model fit, such as the coefficient of determination ($R^2$) or root mean squared error (rmse), are often interpreted as estimating the predictive ability of a model, but given the way they are typically implemented, these types of in-sample statistics quantify only the fit of the model to the data that generated that model. These metrics are notoriously bad at estimating how well a model predicts out-of-sample data (Picard and Cook 1984, Bahn and McGill 2013). Traditional approaches to model selection (AIC, p-values, root mean squared error) that are carried out on in-sample data would come prior to prediction evaluation and be used to distill the set of all possible models down to a manageable subset to be tested against out-of-sample data. Third, model transferability (Wenger and Olden 2012, Godsoe et al. 2015) – how well models generalise to new contexts (e.g. different times, places or taxa) – would become a central question in ecology. Transferability is critical to understanding because understanding without transferability is not as valuable. In particular, without temporal transferability, our understanding is ephemeral and transient.

Parsimony holds a hallowed place in model selection. In ecology, simplicity is accorded inherent value (Marquet et al. 2014) when, in fact, there is no logical reason to believe that simple models are closer to the truth than complex models (Pearl 1978). Parsimony often has enormous practical value but claims of inherent value have relatively little support. Simplicity/parsimony has practical value for two primary reasons: 1) when using finite data to construct and test theory, parsimony, used systematically and rigorously, will often increase the probability of approaching the 'true' underlying process (Grunwald 2007). This is because, when data are limited, complex models are more likely to capture idiosyncrasies of the data rather than true underlying processes. However, this problem is entirely due to limitations the data impose on our ability to detect the underlying process, rather than on any inherent value of simple models. 2) When making predictions in an applied context it may be useful to simplify the description of natural processes to lower the costs of acquiring the inputs necessary to make predictions. Both of these are practical constraints – one imposed by limited data and the other by limited resources. There is no inherent value in choosing a simple theory that is far from the truth over a complex one that is close to it.

## Questions that are rarely addressed in ecology but are essential if prediction is the focus or 'Figures we have never seen in ecology'

### 1. How does predictive ability/understanding decay with distance in space or time?
Ideally, models of the natural world predict equally well at any place in space or moment in time but this is unlikely to be true for most models. For example, a model of fish abundance constructed or parameterized using data from lakes in Wisconsin would be expected to have better predictive ability for other nearby lakes in Wisconsin, worse for lakes in Minnesota, and worse yet for lakes in California. We have never seen an ecological study that has quantitatively addressed the effect of spatial, temporal or taxonomic distance on transferability (Fig. 1). Petchey et al. (2015) explicitly discussed the concept of predictive horizons – that is, the spatial, temporal, and taxonomic distances at which models are transferable – but we have not seen the questions addressed empirically in ecology. These questions are fundamental when prediction is at the core of a discipline.
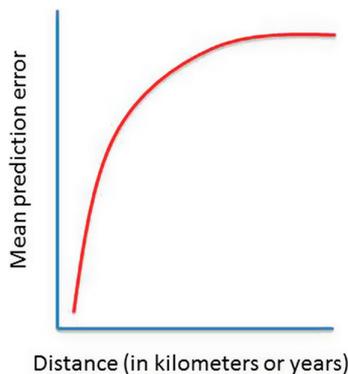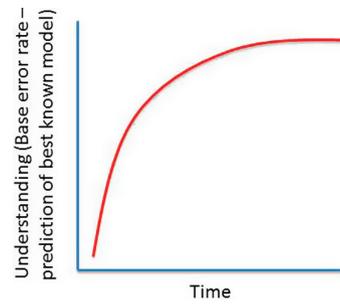


Figure 2. Hypothetical temporal trend in understanding of a single phenomenon (as measured by improved predictive accuracy).

### 2. How well do we understand what causes changes in a response variable Y?
When prediction is at the core of a discipline, predictive ability can be used to quantify understanding. This requires estimating prediction error with full knowledge and with no knowledge, and then measuring where we sit on that continuum. Defining full knowledge and no knowledge are thorny but soluble problems (Fig. 2).

### 3. What is the rate of scientific progress in a particular discipline?
One reasonable way to measure the progress of science is by how our understanding of the natural world improves, and we measure that improved understanding by how well our predictions improve. Conceptually, this makes measuring scientific progress relatively straightforward – a 10% increase in predictive accuracy implies a 10% increase in understanding. In practice, it implies identifying all the phenomena that a discipline attempts to explain and estimating improvement in predictive accuracy across all phenomena (Fig. 3).

### 4. What is the upper limit on model predictive ability?
The upper limit on predictive ability is set by measurement error if we assume that true stochasticity doesn't exist (Fig. 4). Thus, the upper limit will vary depending on the variables in the model and how the variables are being measured. Understanding the upper limits on predictive ability would prevent ecologists from continuing to try and improve models that were already close to maximum predictive ability given the constraints imposed by measurement error.



Figure 1. Hypothetical relationship between spatial or temporal distance and prediction error.
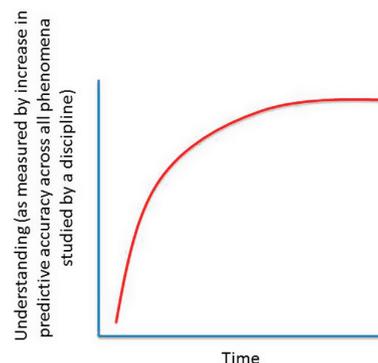


Figure 3. Hypothetical temporal trend in understanding in a discipline (as measured by improved predictive accuracy).
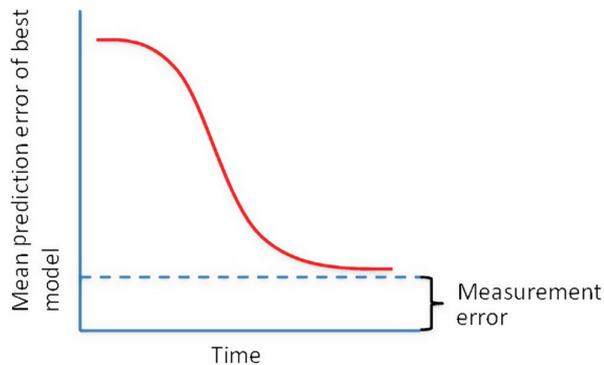
Figure 4. Hypothetical limits that measurement error could set on predictive ability.

## Problems associated with prediction

We see a commitment to prediction as key to accelerating progress in ecology; however, we acknowledge that there are problems associated with emphasizing prediction that deserve discussion.

### Prediction is necessary but not sufficient to prove understanding
Simply because one variable is a good predictor of another doesn't mean that we understand the causal relationship between the two variables. We could make good out-of-sample predictions when the independent variables have no causal effect but are correlated with the true drivers. This is where experiments have their greatest value. However, we point out that experimentation is not the only way to provide evidence of causality. Consider the many theories which have never been tested experimentally, or for which experimentation is impossible, but for which we feel a strong casual mechanistic understanding exists (e.g. planetary orbits, that smoking causes cancer in humans, etc.). The literature on causal modelling is deep (Grace 2006, Pearl 2009a, b) and well beyond the scope of this article; however, linking mechanistic understanding to accurate predictions is an important future challenge facing ecologists (McGill and Nekola 2010). But it must be preceded by an acceptance that prediction is the first step to demonstrating understanding.

### The bias-variance problem
While it may seem counter-intuitive, the model which makes the best predictions on out-of-sample data may not contain all the true drivers or capture the true functional relationships among variables. This is because of the bias-variance tradeoff – all other things being equal, for each additional parameter that has to be estimated, the precision of all parameter estimates declines (Hastie et al. 2009). So, while you may be estimating the right parameters, your estimates of the parameters are increasingly unreliable and potentially the model that has more of the true drivers (albeit poorer estimates of the parameters associated with the variables) will make poorer predictions. This has implications for our argument regarding model parsimony: while we see no inherent value in lower dimensional models, as model dimensionality increases, greater numbers of

samples are required to maintain the reliability of parameter estimates, and therefore, there can be a practical value to parsimony that has nothing to do with any inherent value of simplicity.

### Test-data problem
We have suggested that models should be assessed based on their ability to predict out-of-sample data. However, the size and quality of the test data is a consideration. How much testing is enough? If the training set is much larger or measured with much more precision and accuracy than the test set, should out-of-sample results be preferred? Probably not, but it's not clear where or how to draw that line. While these issues are discussed in other fields (i.e. the study of "big data", Hastie et al. 2009), we see this as an area ripe for research in the ecological sciences.

### The 'gold standard' problem
Can we identify a gold standard for test data? That is, can we develop a library of datasets that will be used only to assess which models are best, a library of datasets that will be used as our yardstick for measuring what we understand about the natural world?

We have identified a handful of problems associated with a greater emphasis on prediction and as prediction increasingly becomes the diagnostic of understanding we suspect the list will grow. We believe the problems identified above are serious but tractable.

## How prediction should change the way we do ecology

### We should develop more mature (i.e. quantitative) hypotheses
One hallmark of an immature discipline is the preponderance of qualitative hypotheses. Ecology has a long and continuing tradition of null hypothesis statistical testing (NHST) and the problems associated with NHST have been well documented (Anderson et al. 2000). However, one weakness that gets relatively little attention is that null hypothesis testing usually implies testing qualitative hypotheses (e.g. when patch size gets larger species richness will increase). Such qualitative models are legitimate and can increase our understanding but only in a limited way because they have relatively low information content. A maturing discipline must move beyond such qualitative coarse predictions to riskier, more quantitative, precise predictions, sensu Popper.

### We should identify modelling techniques that are most appropriate for prediction
For example, we should avoid using categorical independent variables to represent variables that are continuous (Cottingham et al. 2005). For most ecologists, experimental designs fall into one of two categories: predictor variables are categorical or continuous. Ecology has a long tradition of using categorical independent variables in controlled experiments even when the independent variables were clearly continuous variables. However, if the objective of an experiment is to develop a model that can be used to make predictions (and it almost always should be), then you

want a model that includes continuous variables to treat those variables as continuous. If we identify prediction as the diagnostic of understanding, then the only instances we would use categorical variables are when 1) the independent variables are truly categorical, 2) preliminary work is being done to examine if there are any effects at all and so extreme low and extreme high values are chosen to maximize the potential of detecting an effect, or 3) low sample size precludes testing for non-linear effects and so testing at extreme values provides the most accurate and precise linear slope estimate (Steury and Murray 2005).

### We should quantify transferability

Cross-validation should be a fundamental step in model selection. There is still much debate about the preferred method but there are many reasonable alternatives (Arlot and Celisse 2010). But perhaps more important is the assessment of transferability – how well does our understanding in a particular context generalise to novel contexts such as different places or times. Estimating (quantifying) temporal transferability simply involves collecting data from the same sampling points at two or more different times and building the model(s) using data from one time and testing the model(s) against data from the others. Estimating (quantifying) spatial transferability implies collecting data in two or more different regions and building the model(s) using data from one region and testing the model(s) against data from the other region(s).

### We should estimate measurement error

This becomes particularly important when prediction is recognized as the only measure of understanding because measurement error places a hard ceiling on the amount of understanding that we can demonstrate. In theory, perfect understanding would be demonstrated by perfect prediction but, of course, the limit on the accuracy and precision of predictions is imposed by measurement error (Koreisha and Fang 1999). Without good estimates of measurement error it is impossible to know if you are approaching the ceiling on predictive ability. For example, if a model makes 'out-of-sample' predictions that are within 20% of the observed values and measurement error is also ± 20%, it implies that there is no use trying to improve the model.

### We should change how we think about power

Like measurement error, sampling error places a ceiling on predictive performance. Currently, power is a concept that applies strictly to null-hypothesis testing and estimates the probability of detecting an effect of a specified size given sampling error, alpha and sample size. Power, in a predictive context, would estimate the potential precision of predictions (if you have the correct model) given sampling error and sample size. This would allow us to estimate the limit that experimental design places on prediction performance and choose sample sizes based on the desired prediction performance.

### We should clearly identify what kind of science we are doing

Most published manuscripts attempt to make the case that they are contributing to our understanding of how natural systems work but they rarely describe how they have done that let alone provide evidence for the assertion. We have identified seven legitimate avenues towards progress in understanding the natural world. That is, types of science that fit into the prediction framework.

1) Identifying a novel pattern that needs explanation.
2) Developing theoretical models that identify new variables/functional relationships to test.
3) Improving measurement accuracy and precision.
4) Estimating the effects of new variables on predictive ability.
5) Estimating the effects of novel functional relationships on predictive ability.
6) Estimating the effects of new parameter estimates on predictive ability.
7) Using experiments or structural causal models (Pearl 2009a, b) to address mechanism/cause.

These seven types include most approaches published in ecological journals so it does not separate good science from bad, but it does force us to confront the type of science we are using and whether the research is likely to make a substantive contribution to increased understanding. For example, if we have a long list of unexplained patterns is there a need for more descriptive work?

## Conclusion and summary

Our central thesis can be stated in three main points:

1) prediction is the only way to demonstrate understanding and therefore is essential and should be at the core of any scientific discipline,

2) predictions should be validated with independent data from natural systems,

3) precise, quantitative (i.e. risky) predictions are the hallmark of a mature discipline and ecologists should strive to make or test precise quantitative predictions to move our field forward.

No working scientist would dispute that prediction is important. The fundamental difference here is the assertion that only prediction can demonstrate scientific understanding. That without prediction there is no evidence of understanding. That, in practice, without evidence of understanding there is no understanding. Our fundamental concern is with the distance between how important prediction is and how important ecologists perceive it to be; the distance between what we do and what we need to do to make progress in ecology. And finally, it is the distance between the role that ecology could play in people's lives and the role that it does play.

## References

Anderson, D. R. et al. 2000. Null hypothesis testing: problems, prevalence and alternative. – J. Wildl. Manage. 64: 912–923.

Adler, P. B. et al. 2011. Productivity is a poor predictor of plant species richness. – Science 333: 1750–1753.

Arlot, S. and Celisse, A. 2010. A survey of cross-validation procedures for model selection. – Stat. Surv. 4: 40–79.

Bahn, V. and McGill, B. J. 2013. Testing the predictive performance of distribution models. – Oikos 122: 321–331.

Congalton, R. G. and Green, K. 2009. Assessing the accuracy of remotely sensed data – Principles and practices, 2nd edn. – Taylor and Francis/CRC Press.

Connell, J. H. 1961. The influence of interspecific competition and other factors on the distribution of the barnacle, *Chthamalus stellatus*. – Ecology 42: 710–723.

Cottingham, K. L. et al. 2005. Knowing when to draw the line: designing more informative ecological experiments. – Front. Ecol. Environ. 3: 145–152.

Dillon, P. J. and Rigler, F. H. 1975. A simple method for predicting the capacity of a lake for development based on lake trophic status. – J. Fish. Res. Bd Can. 32: 1519–1531.

Evans, M. R. 2012. Modelling ecosystems in a changing world. – Phil. Trans. R. Soc. B 367: 181–190.

Evans, M. R. et al. 2012. Predictive ecology: systems approaches. – Phil. Trans. R. Soc. B 367: 163–169.

Freckleton, R. J. 2004. The problems of prediction and scale in applied ecology: the example of fire as a management tool. – J. Appl. Ecol. 41: 599–603.

Godsoe, W. et al. 2015. Information on biotic interactions improves transferability of distribution models. – Am. Nat. 185: 281–290.

Grace, J. B. 2006. Structural equation modeling and natural systems. – Cambridge Univ. Press.

Grimm, V. and Railsback, S. F. 2012. Pattern-oriented modelling: a 'multi-scope' for predictive systems ecology. – Phil. Trans. R. Soc. B 367: 298–310.

Grunwald, P. D. 2007. The minimum description length principle. – MIT Press.

Hastie, T. et al. 2009. The elements of statistical learning: data mining, inference and prediction, 2nd edn. Springer series in statistics. – Springer.

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

Inchausti, P. and Halley, J. 2001. Investigating long-term ecological variability using the global population dynamics database. – Science 293: 655–657.

Jurafsky, D. and Martin, J. H. 2008. Speech and language processing, 2nd edn. – Pearson Ltd.

Koreisha, S. G. and Fang, Y. 1999. The impacts of measurement errors on ARMA prediction. – J. Forecasting. 18: 95–109.

LeTreut, H. Ed. 1995. Climate sensitivity to radiative perturbations: physical mechanisms and their validation. – Springer.

Marquet, P. A. et al. 2014. On theory in ecology. – BioScience 64: 701–710.

McGill, B. J. and Nekola, J. C. 2010. Mechanisms in macroecology: AWOL or purloined letter? Towards a pragmatic view of mechanism. – Oikos 119: 591–603.

Mittelbach, G. G. et al. 2001. What is the observed relationship between species richness and productivity. – Ecology 82: 2381–2396.

Open Science Collaboration 2015. Estimating the reproducibility of psychological science. – Science 349: 4716.

Pearl, J. 1978. On the connection between the complexity and credibility of inferred models. – Int. J. Gen. Syst. 4: 255–264.

Pearl, J. 2009a. Causality: models, reasoning and inference, 2nd edn. – Cambridge Univ. Press.

Pearl, J. 2009b. Causal inference in statistics: an overview. – Stat. Surv. 3: 96–146.

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

Peters, R. H. 1991. A critique for ecology. – Cambridge Univ. Press.

Picard, R. R. and Cook, R. D. 1984. Cross-validation of regression models. – J. Am. Stat. Ass. 79: 575–583.

Sakamoto, M. 1966. Primary production by phytoplankton community in some Japanes lakes and its dependence of lake depth. – Arch. Hydrobiol. 62: 1–28.

Schindler, D. W. 1974. Eutrophication and recovery in experimental lakes: Implications for lake management. – Science 184: 897–899.

Steury, T. D. and Murray, D. L. 2005. Regression versus ANOVA. – Front. Ecol. Environ. 3: 356–357.

Sutherland, W. J. and Freckleton, R. P. 2012. Making predictive ecology more relevant to policy makers and practitioners. – Phil. Trans. R. Soc. B 367: 322–330.

Taktak, A. F. G. and Fisher, A. C. 2007. Outcome prediction in cancer. – Elsevier Science.

Tilman, D. and Downing, J. A. 1994. Biodiversity and stability in grasslands. – Nature 367: 363–365.

Wenger, S. J. and Olden, J. D. 2012. Assessing transferability of ecological models: an underappreciated aspect of statistical validation. – Meth. Ecol. Evol. 3: 260–267.

Vollenweider, R. A. and Dillon, P. J. 1974. The application of the phosphorus loading concept to eutrophication research. – Natl Res. Counc. Can. NRCC 13690 CCIW, Burlington, ON.

Whitaker, R. J. 2010. Meta-analyses and mega-mistakes: calling time on meta-analysis of the species richness–productivity relationship. – Ecology 91: 2522–2533.

Zhang, M. 2009. Artificial higher order neural networks for economics and business. – Information Science Reference.