



---

Year: 2017

---

## The practice of prediction: What can ecologists learn from applied, ecology-related fields?

Pennekamp, Frank ; Adamson, Matthew W ; Petchey, Owen L ; Poggiale, Jean-Christophe ; Aguiar, Maíra ; Kooi, Bob W ; Botkin, Daniel B ; DeAngelis, Donald L

**Abstract:** The pervasive influence of human induced global environmental change affects biodiversity across the globe, and there is great uncertainty as to how the biosphere will react on short and longer time scales. To adapt to what the future holds and to manage the impacts of global change, scientists need to predict the expected effects with some confidence and communicate these predictions to policy makers. However, recent reviews found that we currently lack a clear understanding of how predictable ecology is, with views seeing it as mostly unpredictable to potentially predictable, at least over short time frames. However, in applied, ecology-related fields predictions are more commonly formulated and reported, as well as evaluated in hindsight, potentially allowing one to define baselines of predictive proficiency in these fields. We searched the literature for representative case studies in these fields and collected information about modeling approaches, target variables of prediction, predictive proficiency achieved, as well as the availability of data to parameterize predictive models. We find that some fields such as epidemiology achieve high predictive proficiency, but even in the more predictive fields proficiency is evaluated in different ways. Both phenomenological and mechanistic approaches are used in most fields, but differences are often small, with no clear superiority of one approach over the other. Data availability is limiting in most fields, with long-term studies being rare and detailed data for parameterizing mechanistic models being in short supply. We suggest that ecologists adopt a more rigorous approach to report and assess predictive proficiency, and embrace the challenges of real world decision making to strengthen the practice of prediction in ecology.

DOI: <https://doi.org/10.1016/j.ecocom.2016.12.005>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-134675>

Journal Article

Accepted Version



The following work is licensed under a Creative Commons: Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.

Originally published at:

Pennekamp, Frank; Adamson, Matthew W; Petchey, Owen L; Poggiale, Jean-Christophe; Aguiar, Maíra; Kooi, Bob W; Botkin, Daniel B; DeAngelis, Donald L (2017). The practice of prediction: What can ecologists learn from applied, ecology-related fields? *Ecological Complexity*, 32:156-167.

DOI: <https://doi.org/10.1016/j.ecocom.2016.12.005>

1 The practice of prediction: What can  
2 ecologists learn from applied, ecology-  
3 related fields?

4 Frank Pennekamp, Matthew W. Adamson, Owen L. Petchey, Jean-Christophe  
5 Poggiale, Maira Aguiar, Bob W. Kooi, Daniel B. Botkin, Donald L. DeAngelis

6 Affiliations:

7 FP: University of Zurich, Institute of Evolutionary Biology and Environmental Studies,  
8 Winterthurerstrasse 190, 8057 Zurich, Switzerland

9 MWA: University of Osnabrück, Institute of Environmental Systems Research, Barbarastrasse  
10 12, 49076 Osnabrück, Germany

11 OLP: University of Zurich, Institute of Evolutionary Biology and Environmental Studies,  
12 Winterthurerstrasse 190, 8057 Zurich, Switzerland

13 JCP: Aix Marseille Université, CNRS/INSU, Mediterranean Institute of Oceanography (MIO) UM  
14 110, 13288, Marseille, France

15 MA: Centro de Matemática e Aplicações Fundamentais da Universidade de Lisboa, Avenida  
16 Prof. Gama Pinto 2, 1649-003 Lisboa, Portugal.

17 BK: VU University, Faculty of Earth and Life Sciences, De Boelelaan 1087, 1081 HV  
18 Amsterdam, The Netherlands

19 DB: Department of Ecology, Evolution, and Marine Biology, University of California,  
20 Santa Barbara, California 93106 USA

21 DLD: U. S. Geological Survey, Wetland and Aquatic Research Center, Gainesville, FL,  
22 32653 USA

23

24 **Corresponding author:** [Frank.Pennekamp@ieu.uzh.ch](mailto:Frank.Pennekamp@ieu.uzh.ch)

25

26 **Abstract**

27 The pervasive influence of human induced global environmental change affects biodiversity  
28 across the globe, and there is great uncertainty as to how the biosphere will react on short and  
29 longer time scales. To adapt to what the future holds and to manage the impacts of global change,  
30 scientists need to predict the expected effects with some confidence and communicate these  
31 predictions to policy makers. However, recent reviews found that we currently lack a clear  
32 understanding of how predictable ecology is, with views seeing it as mostly unpredictable to  
33 potentially predictable, at least over short time frames. However, in applied, ecology-related fields  
34 predictions are more commonly formulated and reported, as well as evaluated in hindsight,  
35 potentially allowing one to define baselines of predictive proficiency in these fields. We searched  
36 the literature for representative case studies in these fields and collected information about  
37 modeling approaches, target variables of prediction, predictive proficiency achieved, as well as  
38 the availability of data to parameterize predictive models. We find that some fields such as  
39 epidemiology achieve high predictive proficiency, but even in the more predictive fields proficiency  
40 is evaluated in different ways. Both phenomenological and mechanistic approaches are used in  
41 most fields, but differences are often small, with no clear superiority of one approach over the  
42 other. Data availability is limiting in most fields, with long-term studies being rare and detailed  
43 data for parameterizing mechanistic models being in short supply. We suggest that ecologists  
44 adopt a more rigorous approach to report and assess predictive proficiency, and embrace the  
45 challenges of real world decision making to strengthen the practice of prediction in ecology.

46

47 **Key words: predictive proficiency, forecast, hindcast, forecast horizon**

48

# 49 1 Introduction

50 Accurate predictions about the consequences of environmental change for natural populations,  
51 communities, and ecosystems would be valuable to inform conservation, management and  
52 adaptation strategies (Clark et al., 2001). This is even more evident when considering the current  
53 speed and magnitude of environmental change, for instance climate change, which has spurred  
54 scientific disciplines such as climatology to invest considerable effort in predicting the future  
55 (IPCC, 2014).

56

57 Ecology has a long history of using *explanatory* prediction to test hypotheses and theories (Peters,  
58 1991; Reserits and Bernardo, 1998). The purpose of *anticipatory* prediction, in contrast, is to  
59 provide useful information about the future state of a system (Mouquet et al., 2015). As such it is  
60 unimportant how anticipatory predictions are made (mechanistic versus phenomenological  
61 models), so long as they are useful. A culture of *anticipatory* predictions is only beginning to  
62 develop, and opinion about the success of such an enterprise is divided (Petchey et al., 2015).  
63 Some believe that medium- to long-term predictions in ecology are impossible due to factors such  
64 as model and parameter uncertainty, system complexity and non-ergodicity (i.e., not having the  
65 same behavior averaged over time as over all the system's states), or long-term transients  
66 (Planque, 2016), making predictions “computationally irreducible” (Beckage et al., 2011). Others  
67 show that mechanistic models are able to make precise, accurate, and reliable predictions about  
68 a variety of state variables of complex ecosystems (Purves et al., 2008). General and specific  
69 statements about the ability to make useful anticipatory predictions about ecological variables  
70 could be facilitated by the considerations below (Petchey et al., 2015).

71

72 First, one should not ask whether ecology is predictable or not, but about the predictive proficiency  
73 for a given response and a given time frame. It may be easy to predict that a 50% increase in a  
74 forest fragmentation index in certain locations will result in some bird species going locally extinct  
75 within the next 100 years. It would, however, be harder to predict the percentage of bird species  
76 that would become extinct, and still harder to predict exactly which bird species would become  
77 extinct. So ‘what is being predicted’ needs to be specified carefully, as well as the time frame of  
78 prediction (Petchey et al., 2015).

79

80 Second, coherence about how to measure predictive ability is desirable, yet there are many  
81 metrics available, some of which are redundant, whereas others measure distinct features of  
82 predictive ability (Olsen et al., 2016). Petchey et al. (2015) proposed that coherence and  
83 generality could be achieved by the ecological forecast horizon (EFH). The EFH is a quantitative  
84 tool to assess the predictive proficiency when observations are compared (e.g. using  $R^2$ ) to a  
85 particular model of the system. The forecast horizon is the time into the future for which forecasts  
86 can be made within a given predictive proficiency domain. Use of the EFH makes both time frame  
87 and predictive proficiency explicit.

88

89 Third, a view of past and current predictive ability, and a vision for the future would be useful  
90 (Figure 1). In weather forecasting, predictive proficiency has continuously improved since the  
91 1980's from about 80% to better than 95% in 2013 for forecasts three days ahead, while weekly  
92 forecasts improved from about 40% to 70% (Bauer et al., 2015). Some of the success in improving  
93 predictions is related to the meticulous monitoring of predictive success. Hence, knowing and  
94 critically evaluating predictive proficiency is essential, as it allows evaluation of our progress and  
95 enables identification of areas with deficient predictive proficiency.

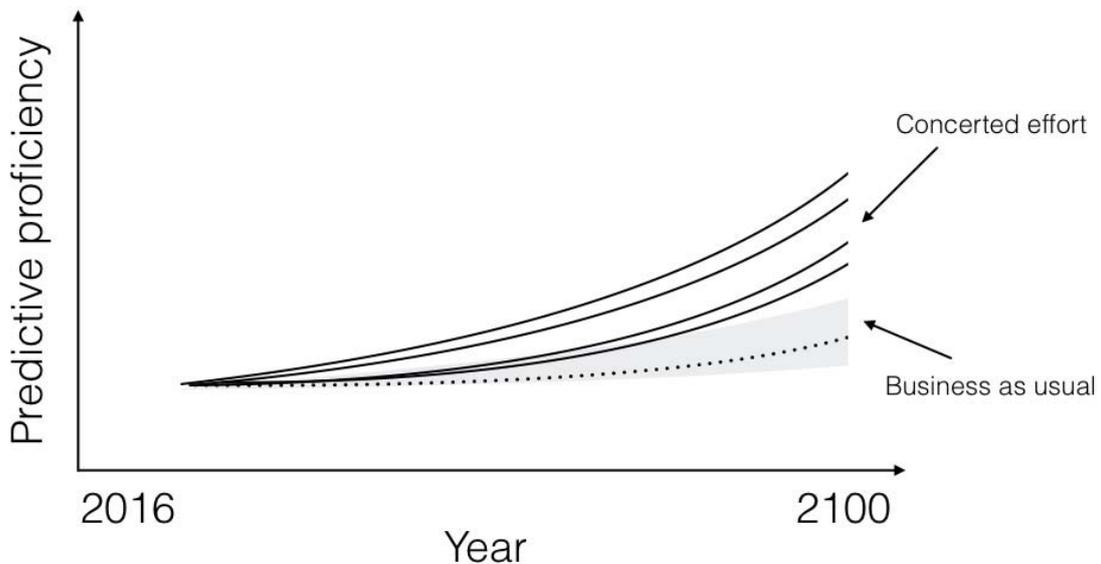
96

97 Fourth, ecologists need to understand where advances in predictive ability are most easily  
98 achieved, and what is required to make such advances. For example, one major difference  
99 between ecology and fields such as weather forecasting is the availability of data to check  
100 predictions. Ecological studies are often conducted over a given time frame (e.g., a thesis or  
101 research grant) and may be short compared to the relevant time scale of the study system (e.g.,  
102 population dynamics of a particular animal or plant species). The vast majority of datasets in  
103 ecology fall into the category of short-term independent studies (Mouquet et al., 2015).  
104 Furthermore, datasets are often not collected with the specific purpose of making anticipatory  
105 predictions (Mouquet et al., 2015). This currently limits our ability to check the predictive success  
106 of particular forecasting techniques and to define the baseline of predictive success in ecology.

107

108 While ecology in general is only beginning to develop the practice of prediction, related fields such  
109 as fisheries science that have to provide quantitative predictions to government agencies, may  
110 have already developed standardized reporting rules and rigorous means for assessing predictive  
111 proficiency from which ecologists can generally learn. We therefore selected fields and  
112 phenomena such as fisheries, epidemiology, eutrophication and algal blooms, ecotoxicology,  
113 forestry, and marine and terrestrial biogeochemistry and searched for representative case

114 studies. Importantly, these fields often deal with similar kinds and levels complexity. Given the  
115 vast literature in each field, our overview is necessarily incomplete; hence we informally (i.e.,  
116 through discussion rather than quantitative analysis) review representative case studies. Our goal  
117 is to derive some insights as to why and when predictions succeed in these fields and produce  
118 some suggestions as to how to strengthen the practice of prediction in ecology.  
119



120  
121 Figure 1. Scenarios of how the ability to predict ecological dynamics may evolve in the future. Business as  
122 usual (shaded region) involves relatively sparse and uncoordinated efforts in ecological forecasting, and  
123 would result in no or slow increase in predictive ability, with occasional breakthroughs (not illustrated).  
124 Concerted effort is another scenario to transform ecological science into being primarily concerned with  
125 and coordinated to improve anticipatory predictions. The resulting increase in predictive ability is uncertain  
126 (hence multiple different lines). One scenario of limited advances in predictive ability despite increased  
127 efforts (dotted line) could result from there being hard limits to ecological predictability (e.g., computational  
128 irreducibility). Other scenarios (solid lines) showing faster increases in predictive ability, could result from  
129 advances in data availability and modeling, for example.

130

## 131 **2 Predictions in ecology-related fields**

132 In this section we give an overview of fields, in no particular order, in which policy relevant  
133 predictions are made. To facilitate comparisons across fields, we use a common template to  
134 describe the predictive practice. In each subsection we first describe why prediction is important  
135 for the field and what type of predictions are made. We then discuss the predictive proficiency  
136 obtained and the types of models used in the representative case studies. Finally, we assess the  
137 importance of data availability and quality in the field, and highlight particular strengths and  
138 challenges for the practice of prediction (summarized in Table 1).

139 Predictive models span a range of techniques, from simple extrapolation, to time series modelling  
140 using statistical or machine learning type models that can capture linear and non-linear patterns,  
141 to process-based models (e.g. individual-based models or population models based on first  
142 principles) that include biological mechanisms and environmental dependencies. Here we follow  
143 the rough separation of models into mechanistic (e.g. individual-based models) versus  
144 phenomenological models (including extrapolation, statistical and machine learning approaches)  
145 introduced by Mouquet et al. (2015). Whereas the latter are powerful at capturing patterns in the  
146 data, they do not capture explicit mechanisms and hence may predict poorly out of the range of  
147 data (Evans et al., 2013). On the other hand, process-based models are expected to work better  
148 under novel conditions, provided the key mechanisms are correctly included. Approaches also  
149 differ in terms of the data required for parameterization. Process-based models tend to be more  
150 demanding in terms of the data required, whereas phenomenological approaches often are  
151 applied directly to the state variable (e.g. time series analysis of population sizes).

### 152 **2.1 Fisheries**

153 Anticipatory predictions in fishery science are needed to inform management and conservation  
154 as fish stocks are depleted on a global scale. According to Garcia and Graininger (2005) 52% of  
155 stocks are fully exploited and 16% overfished. Fish stock assessments provide guidelines for  
156 sustainable management of focal fish species, and are based on catch data, scientific surveys  
157 and biological information about the species. Important predicted quantities are the total  
158 population size or biomass to inform maximum sustainable yield, the age structure of a stock, and  
159 its fecundity.

160

161 Since the 1950's very simple population dynamic models have been used to make forecasts of  
162 abundance based on the single-species formalism of Beverton and Holt (1957). These models  
163 usually consider only age or stage structure, with multiple cohorts. Stock-recruitment relationships  
164 such as the Ricker or Beverton-Holt curve are used to predict the recruitment of a harvested  
165 population using information on its spawning biomass. Cury and colleagues (2014) have found  
166 low explanatory power of the stock-recruitment relationship, explaining only 8.8% of the variation  
167 in a global dataset of stock-recruitment curves, even though it is still widely used in fish stock  
168 assessments. They suggested that a better consideration of density-independent factors (e.g.  
169 species interactions and temperature) may increase the amount of variation explained.

170

171 Moving beyond simple linear relationships, non-linear time series analysis has gained traction in  
172 fishery sciences. One promising technique is empirical dynamic modeling (EDM) as introduced  
173 by Sugihara (1994). Recent papers show application of the method to forecast population  
174 dynamics (Glaser et al., 2014), and unravel environmental dependencies of population dynamics  
175 (Deyle et al., 2013; Hsieh et al., 2005). The methodology can deal with non-linearity and readily  
176 produces forecasts from time series of abundance for which relatively long-term records exist in  
177 fisheries. Glaser et al. (2014) found that about 70% of 206 time series of fish and marine  
178 invertebrates had significant predictable structure, but that the average predictive proficiency  
179 (using the correlation coefficient  $\rho$ ) was only about 0.39. They concluded that short-term (e.g., 1  
180 year forecasts) are feasible, whereas mid to long-term forecasts (2-10 years) are not reliable yet.  
181 Francis et al. (2005) used a more traditional generalized additive modeling approach to model  
182 relative abundance and occurrence of individual species. They report that models performed  
183 reasonably well ( $\rho > 0.5$ ) on only 4 out of 12 fish species; however, presence was predicted with  
184 greater success (8 out of 12). These authors also reported an independent assessment, using a  
185 different dataset collected later, in which 5 out of 8 species were predicted with a  $\rho$  larger than  
186 0.5.

187

188 It has been suggested that predictions of fish stocks could be improved by a more ecosystem-  
189 focused fishery management that includes interactions between species and human exploitation  
190 (Travis et al., 2014). Olsen et al. (2016) performed a rigorous assessment of the Northeast U.S.  
191 Atlantic marine ecosystem model using 40-year hindcasts and 10-year forecasts using different  
192 metrics of predictive proficiency (including correlation and different error measures). Twenty-two  
193 ecosystem indicators were evaluated, including emergent properties of the model (e.g., average  
194 trophic position). Model performance had large variation across species, but several of the target

195 species showed correlations greater than 0.5 in both hindcast and forecast skill. Root mean  
196 squared error for the normalized biomass data ranged between 0.1 and 0.8.

197

198 Data for fish assessments is often based on fish landings (sold fish biomass) or portside  
199 assessments of catch. Whereas the former has the advantage of capturing the effects of a major  
200 human intervention, they can be confounded by complex human behavior, including under-  
201 reporting of catch (Glaser et al., 2014). Time series of abundance can be of considerable length  
202 (> 50 years) in fisheries, facilitating data-driven approaches, however, detailed data to  
203 parameterize more process-based models is not so readily available (Travers et al., 2007). More  
204 explicit consideration of biotic interactions and environmental drivers would probably lead to better  
205 predictions (Brander, 2003; Travis et al., 2014). A general shift towards an ecosystem-based  
206 fisheries management approach, based on trophic web models with many components and  
207 across levels of organization, can be observed (e.g. Travers-Trolet et al., 2014). These models  
208 still suffer from over-simplifications, such as describing interactions in marine communities as  
209 largely unidirectional trophic relationships instead of fully embracing their complexity (Travis et  
210 al., 2014).

211

212 The use of forecasting centered methods such as EDM, as well as the use of ecosystem models,  
213 provide a clear baseline for the level of predictive proficiency in fisheries. Forecasting proficiency  
214 with of rhos about 0.4 – 0.5 are obtained with different methodologies on population and  
215 ecosystem level.

## 216 **2.2 Epidemiology**

217 The public health implications of infectious diseases renders epidemiology an important field for  
218 frequent and policy relevant predictions. Important targets of prediction are: the likelihood of an  
219 outbreak of a disease (Woolhouse, 2011), the timing and amplitude and shape of an epidemic  
220 (Grenfell et al., 2002), and the outcome of specific interventions (Grenfell and Keeling, 2008).

221

222 The SIR model developed by Kermack and McKendrick (1927), in which the population is divided  
223 into susceptible, infective and recovered individuals, forms the basis of many epidemiological  
224 models. The crucial parameter of the model,  $R_0$ , predicts the spread of the epidemic, i.e., when  
225  $R_0 > 1$ , the infection will spread in a population. A recent example of epidemiological prediction is  
226 the occurrence of dengue during the 2014 FIFA World Cup in Brazil. Lowe et al. (2014) predicted

227 a high risk of acquiring dengue in Brazil, providing probabilistic forecasts of risk based on seasonal  
228 climate forecasts. Aguiar et al. (2015) also analyzed the epidemiological data, taking into  
229 consideration population densities in the twelve host cities, and the seasonality of mosquitoes. In  
230 contrast, they predicted a low infection rate during the World Cup due to the incorporation of  
231 different drivers, which was confirmed later (M. Aguiar et al., 2015). Another predictive success  
232 in epidemiology was the foot-and-mouth-disease outbreak in Great Britain in 2001, which yielded  
233 detailed insights into the disease dynamics and a high quality data set (Grenfell and Keeling,  
234 2008; Woolhouse, 2011). A statistical model (logistic regression) and individual-based models  
235 were used in hindsight to check their ability to predict which farms were at risk during the 2001  
236 outbreak (Woolhouse, 2011). The statistical model was found to predict the risk of becoming  
237 infected slightly better than the dynamic model (Woolhouse, 2011). Measles dynamics are  
238 emblematic cases of successful prediction due to characteristic recurrent cycles in large  
239 population centers and occasional breakouts in smaller communities (Bjørnstad et al., 2002). The  
240 authors predicted cases of measles by fitting a time-series SIR model (TSIR, Grenfell et al., 2002)  
241 to data from 60 UK communities and reported impressive short-term (two weeks ahead) mean  $R^2$   
242 of 0.85 by comparing predicted to observed cases. The  $R^2$  ranged from 0.98 – 0.92 for large cities,  
243 and was still reasonably high (0.74) for small communities. Extensions of measles modeling to  
244 small communities that have highly stochastic dynamics still achieved  $R^2$  of 0.86 to 0.55, with 5  
245 out of 6 communities scoring higher than 0.73 (Caudron et al., 2015). A recent review gives a  
246 comprehensive account of the predictability of influenza outbreaks, comparing time series  
247 modeling, individual-based, compartmental and metapopulation models (Nsoesie et al., 2014).  
248 Three studies predicted the magnitude of influenza activity and reported correlation coefficients  
249 between 0.58 and 0.94 (Nsoesie et al., 2014), whereas mechanistic approaches were evaluated  
250 in terms of observations falling within the confidence intervals of the model. The use of different  
251 measures hence hampers direct comparisons between the predictive proficiency among  
252 approaches.

253

254 Epidemiology has excellent long-term records of disease-incidence through space and time. This  
255 can be used for validation and testing models, including the influence of measures such as  
256 vaccination, quarantining and vector control. Epidemiology also benefits from new data sources,  
257 such as internet search queries. A correlation between predicted and observed influenza cases  
258 of 0.96 was obtained (Ginsberg et al., 2009), highlighting the potential of these techniques in  
259 improving forecasts and decreasing time delays. As these models rely on correlations between  
260 search patterns and disease symptoms, input data have to be reviewed carefully to protect

261 against false alerts (Woolhouse, 2011). However, issues related to data quality and availability  
262 are not unique to new predictive approaches in epidemiology, but apply equally to more traditional  
263 approaches. Aguiar et al. (2014) describe a situation where datasets based on different  
264 interpretations of official documents created divergent predictions of disease dynamics. Another  
265 problem is that at an international level some countries may be unwilling to share the data with  
266 the World Health Organization (Woolhouse, 2011), exacerbating the genuine lack of data for  
267 many other diseases.

268

269 Epidemiology shows impressive examples of forecast proficiency with  $R^2$  ranging above 0.9 and  
270 even above 0.7 in more difficult settings (e.g. smaller communities, larger influence of  
271 stochasticity). Nevertheless, the review by Nsoesie et al. (2014) shows that even when targets of  
272 prediction are well defined, the use of different measures of forecast proficiency can hamper  
273 conclusions regarding the state of the art of prediction.

## 274 **2.4 Eutrophication and algal blooms**

275 Eutrophication models predict the availability of key nutrients (nitrogen, phosphate),  
276 phytoplankton biomass (i.e. chlorophyll a concentration) or oxygen availability in aquatic systems.  
277 Excessive nutrients in aquatic system can lead to phytoplankton blooms, especially cyanobacteria  
278 (Conley et al., 2009). As some of the organisms can harm humans and ecosystems,  
279 environmental monitoring aims to predict harmful algal blooms.

280

281 Early studies of N and P fertilization were successful at predicting that fertilization with P would  
282 lead to dominance by N-fixing phytoplankton, but not which species would be dominant  
283 (Schindler, 1977). Modern eutrophication models include a significant, relatively well understood  
284 hydro-dynamical component. Nevertheless, this results in eutrophication models being  
285 computationally expensive, requiring substantial data to calibrate, often to the detriment of  
286 biological detail (Robson, 2014a). Arhonditsis and Brett (2004) compiled a list of 153 mechanistic  
287 aquatic biogeochemical modelling studies. Models could generally reproduce the temperature  
288 and dissolved oxygen time series well, with  $R^2$  values of 0.93 and 0.7, respectively, and median  
289 relative error <10%. In contrast, state variables relevant to eutrophication—nutrients and  
290 phytoplankton—were only moderately predictable, with  $R^2$  values ranging from 0.4 to 0.6, and  
291 median relative errors of around 40%. The models predicted the dynamics of bacteria and  
292 zooplankton even less well. Trolle et al. (2014) compared the ability of three of the most widely

293 used aquatic ecosystem models—DYRESM-CAEDYM (Hamilton and Schladow, 1997); PCLake  
294 (Janse, 1997) and PROTECH (Elliott et al., 2010) to predict chlorophyll a concentrations in lakes.  
295 No model performed particularly well when predicting the day-to-day chlorophyll concentrations  
296 (max  $R^2$  0.33, minimum relative error 103%), but predictions of monthly means were reasonable  
297 (max  $R^2$  0.62, minimum relative error 77%).

298  
299 Jacobs et al. (2014) modelled the presence, abundance and potential virulence of *Vibrio vulnificus*  
300 in marine surface waters. They used a logistic regression model based on the output of  
301 ChesROMS, a three-dimensional model that simulated the circulation and physical properties of  
302 the estuary (sea surface height, temperature, salinity, density and velocity). The model achieved  
303 82% classification success in the training data based on a set of environmental variables. The  
304 same variables were also useful in predicting abundance (low versus high) with concordance of  
305 92% ( $R^2$  of frequency of occurrence on validation high: 0.94 - 0.98). Froehlich et al. (2013)  
306 modelled the presence of the whole *Vibrio* genus in estuarine waters. They used a mechanistic  
307 model based on hydrodynamics, growth and death rates and a statistical model (multiple linear  
308 regression) based on environmental parameters (temperature and salinity) to predict *Vibrio*. This  
309 allows direct comparison of predictive abilities of statistical and mechanistic models. The  
310 mechanistic model was based on the Environmental Fluid Dynamics Code (Hamrick, 1992) and  
311 predicted physical environmental parameters such as salinity and temperature very well ( $R^2$  86.6  
312 and 97.1, respectively). The mechanistic model achieved 63% of explained variation for log-  
313 transformed abundances, timing and magnitude of the peak abundance were mostly well  
314 predicted. The statistical model explained 48% of variation in abundance.

315  
316 Mechanistic eutrophication models require many physical inputs, including hydrodynamic data,  
317 weather conditions, nutrient influx and outflow, as well as water quality parameters and ecological  
318 inputs. Data availability on the biological components is likely to increase as new monitoring  
319 schemes based on environmental genetic data becomes more widely used (Paerl et al., 2003).  
320 This may enable real-time microbiological assessments of aquatic systems in the future.

321  
322 Eutrophication models tend to make good predictions when they 'have strong physical drivers'  
323 (Robson, 2014b). Both Froehlich et al. (2013) and Arhonditsis and Brett (2004) show that physical  
324 properties of the environment are often well captured, whereas the biological layer has  
325 considerably lower predictive proficiency. In the N and P fertilization studies, both the behavior of  
326 nutrients and the behavior of the phytoplankton community as a whole were fairly law-like.

327 However, exactly which N-fixing species would dominate was not so predictable, because  
328 detailed knowledge about the individual species was lacking (Schindler, 1977). Nevertheless,  
329 reported levels of  $R^2$  for phytoplankton dynamics (range 0.3-0.8) indicate low predictive  
330 proficiency even for the biotic components. The ensemble approach used by Trolle et al. (2014)  
331 can improve proficiency and indicate prediction uncertainty, and is commonly used to compare  
332 climate (Murphy et al., 2004) and meteorological models (Houtekamer et al., 1996; Tracton and  
333 Kalnay, 1993).

## 334 **2.5 Ecotoxicology**

335 Ecotoxicology aims to predict the movement of toxicants in the environment, their uptake and bio-  
336 concentration in organisms, and the resulting population level effects. Contaminant fate models  
337 describe the fate and distribution of contaminants in the aquatic system. Important processes are  
338 transport (flow and dispersion), degradation, volatilization, sorption, sedimentation and  
339 resuspension.

340  
341 The predictive ability of several models simulating the bioconcentration of organic chemicals by  
342 fish has been reviewed by Barber (2003). Based on the properties of various organic chemicals,  
343 these models successfully predicted the bioconcentration of chemicals in tissues through time.  
344 Progress in extrapolating the effects of toxicants on individuals to the population level is being  
345 made through individual-based models. Dynamic Energy Budget (DEB) theory extrapolates the  
346 effects of toxicants measured at the individual level to the population level. DEB is based on first  
347 principles in bioenergetics and uses a common model structure for all species. Martin et al. (2013)  
348 present the general approach for animals, and give an example using *Daphnia magna* exposed  
349 to an herbicide. The model captured the density dynamics and changes in the size structure  
350 without fitting or calibration at the population level. It therefore successfully extrapolated to  
351 environmental conditions not included in the parameterization process.

352  
353 Eco-toxicological data are mostly collected in controlled laboratory studies where lethal effects of  
354 toxicants on individuals are measured. Detailed physiological models of toxicant concentration in  
355 individuals hence exist, but the challenge is to predict the effects on higher levels of organization.  
356 Individual-based models and DEB theory in particular show promise for predicting across levels  
357 of organization and take advantage of the rich data sources available on the individual level. In  
358 addition, accidents can provide important and realistic situations to predict contaminant spill in

359 the environment; e.g., accidental discharge of nitrobenzene in 2005 in the Songhua River, China  
360 (Lei et al., 2008). Ecotoxicology, with its strong foundation in physiology relies a lot on mechanistic  
361 modelling, rather than more phenomenological approaches. However, it has been suggested that  
362 machine learning has considerable promise in detecting individual level biomarkers based on  
363 gene expression profiles (Vandersteen, 2011).

## 364 **2.6 Forestry**

365 Two primary targets of prediction in forestry are forest succession and forest productivity.  
366 Foresters need predictions to take long-term decisions regarding the forest composition in the  
367 face of climate change, invasion by non-native trees, and forest fragmentation, while at the same  
368 time managing forests for production. Forests also play an important role in the carbon cycle and  
369 hence in the response of the global climate system to carbon dioxide emissions (Purves and  
370 Pacala, 2008).

371

372 Individual-based forest gap models have been used to predict forest succession, composition,  
373 and effects of environmental changes on forests from their inception in the early 1970's with the  
374 development of JABOWA (Botkin et al., 1972). They use data on individual trees modified by  
375 environmental conditions, including growth, competition through local interactions (shading), and  
376 reproduction. Such models have successfully reproduced the species composition of old-growth,  
377 semi-natural forests (Purves et al., 2008). Ngugi and Botkin (2011) used the Ecosystem Dynamics  
378 Simulator (EDS), based on JABOWA-II (Botkin, 1993) in projecting growth dynamics of mature  
379 remnant Australian brigalow forest communities and the recovery of brigalow thickets. The model  
380 was parameterized for 34 tree and shrub species and tested with independent long-term  
381 measurements. It closely approximated actual development trajectories of mature forests and  
382 regrowth thickets. Changes in species composition in remnant forests were projected with a 10%  
383 error. Basal area values observed in all remnant plots ranged from 6 to 29 m<sup>2</sup> ha<sup>-1</sup> and EDS  
384 projections between 1966 and 2005 (39 years) explained 89.3 (+/- 1.8)% of the observed basal  
385 area of the plots.

386

387 Individual-based forest models are often quite complex and hence analytically intractable.  
388 Recently the perfect-plasticity approximation (PPA) was developed as a model of forest dynamics  
389 (Purves et al., 2008). It is based on individual tree parameters, including allometry, growth, and  
390 mortality. For eight common species in the US, timing and magnitude of basal area dynamics and

391 ecological succession on different soil types were found to be accurate, and predictions for the  
392 diameter distribution of 100-year-old stands had qualitatively correct shape (Purves et al., 2008).

393

394 A serious obstacle for testing predictions of forest succession is that this process can take  
395 centuries to reach its final state. For that reason, 'space-for-time' substitutions have been used  
396 (Pickett, 1989). Clebsch and Busing (1989) empirically measured forests after 63 years of  
397 agricultural abandonment. Forest composition at 63 years was used as starting condition for a  
398 forest gap succession simulator (FORET). The prediction (300 years ahead) matched the state  
399 of a nearby old-growth forest.

400

401 Phenomenological approaches have been used to model the geographic distribution of 30  
402 different tree species, based on environmental data, in Switzerland (Guisan et al., 2007). Ten  
403 different methods (ranging from GLMs to GAMs, MaxEnt and regression trees) were compared  
404 using the area-under-the-curve metric (AUC >0.9 means good skill, <0.7 poor skill). Predictive  
405 proficiency varied among species more than among modeling technique, so that most of the  
406 species' distributions were predicted reasonably well (AUC > 0.7) with at least one of the  
407 techniques.

408

409 It is not surprising that the economic importance of forests has led to abundant data, from  
410 individual tree growth to the dynamics of forest stands. Individual-based models require detailed  
411 information such as light transmission and seed dispersal kernels for parameterization, but have  
412 proven to be successful at local scales (Purves et al., 2008). However, currently available data  
413 from long-term forest monitoring programs is often still insufficient to implement parameter-rich,  
414 process-based models (Evans and Moustakas, 2016). New developments such as the PPA may  
415 circumvent some of these limitations, as they require less data to upscale in a computationally  
416 efficient manner. Forest inventory data, where sample plots are measured on a regular basis  
417 (every 5-10 years), are becoming increasingly available and can be used as input for PPA (Purves  
418 et al., 2008).

419

420 Forestry can be considered a quite successful predictive science. Predictive proficiency is  
421 assessed by the ability to capture patterns (e.g. size distributions, growth dynamics) rather than  
422 metrics (e.g.  $R^2$ ). The need for long-term predictions has led to original approaches (e.g. space-  
423 for-time) to test models. Modelling approaches that aim to predict across levels of organization

424 (individual growth), stand dynamics (population), compositional changes (community), as well as  
425 ecosystem properties are needed, and IBMs have shown some promise in achieving this goal.

426

## 427 **2.7 Terrestrial and marine biogeochemistry**

428 Biogeochemistry encompasses linked physical, chemical, geological, and biological processes in  
429 the environment at all scales (Schlesinger, 1991). Biogeochemical models are key components  
430 in predictions of climate change and in understanding the feedbacks with the biosphere. They  
431 therefore have very important implications for global policy. For instance, global vegetation  
432 models are important components of climate change models (Purves and Pacala, 2008).

433 An important goal of terrestrial biogeochemical models is to predict carbon storage as a function  
434 of increasing CO<sub>2</sub> in the atmosphere. Predictions of four global terrestrial ecosystem models  
435 regarding the terrestrial carbon storage from 1920 to 1992 were compared (McGuire et al., 2001).  
436 These models, which have a spatial resolution of 0.5°, have been calibrated and tested on small  
437 scales. In the tests, atmospheric CO<sub>2</sub>, climate, and cropland extent were used as inputs. Among  
438 other predictions, three of the four models predicted net release of terrestrial carbon up to 1958,  
439 and all four predicted net uptake after 1958. At local spatial scales, the CENTURY model is  
440 representative of terrestrial biogeochemistry models, which are based on relationships between  
441 climate, human management (fire, grazing), soil properties, plant productivity, and decomposition  
442 (Parton et al., 1993). CENTURY is a general ecosystem level model that simulates plant  
443 production, soil water fluxes, soil organic matter dynamics and nutrient cycling for grassland,  
444 forest, savanna and agroecosystems (Parton et al., 1993). It has been tested using observations  
445 from many temperate and tropical grasslands around the world. The results show that soil C and  
446 N levels can be simulated to within +/- 25% of the observed values (100 and 75% of the time,  
447 respectively) for a diverse set of soils.

448

449 Models in marine biogeochemistry have been developed to understand and predict biogenic  
450 cycles (carbon, nitrogen, phosphorus, silica, etc.) over broad temporal and spatial scales. They  
451 were also used to understand the drivers of spatio-temporal variation in primary production. Najjar  
452 et al. (2007) compared twelve models predicting global primary production, sea surface  
453 concentration of dissolved organic carbon and seasonal oxygen fluxes. The results agreed with  
454 empirical data, but predictions were very sensitive to the circulation and to the mixing layer depth.

455 Carr et al. (2006) made predictions of global primary production on the basis of satellite data, and  
456 the predictions varied within a factor of 2. For comparison, 24 biogeochemical models were used  
457 in the same regions, with the result that the predictions encompassed the same ranges as the  
458 empirical estimates, and shared the property of a strong divergence in the Austral Ocean, in the  
459 polar and subpolar regions and in eutrophic regions. A follow-up study compared 21 ocean color  
460 models and 9 biogeochemical models in their ability to predict primary production in the tropical  
461 Pacific region from 1983-1996 (Friedrichs et al., 2009). Models varied widely in predictive  
462 proficiency, but this was not related to model type or complexity. Saba et al. (2010), used 36  
463 models (22 ocean color models using spectral analysis to estimate water constituents, including  
464 chlorophyll-*a*) and 14 biogeochemical models to calculate the primary production in two regions  
465 from 1989 to 2004. 90% of these models underestimated the primary production, with the bias of  
466 the biogeochemical models being twice that of the ocean color models, and only 2% of the  
467 biogeochemical models were able to reproduce the primary production increase observed in  
468 these stations (2% per year), indicating it may take time for marine biogeochemical simulation  
469 models to catch up with more direct observational indicators such as ocean color.

470

471 The broad spatio-temporal scales of biogeochemical cycles make reliable measurements difficult.  
472 Primary sources of data could be reconstructions of biogeochemical cycles preserved in paleo-  
473 records. Remote sensing techniques have promise in improving data availability for global  
474 biogeochemical models and increasingly provide data for more local scales (Asner and Vitousek,  
475 2005).

476

477 Biogeochemical models are often based on first principles (laws of thermodynamics and  
478 chemistry), and hence some predictions can be made with confidence on these aspects. The  
479 biotic component of these models is dominated by plants and decomposers, which allows  
480 reasonable predictions to be made about what to expect when environmental change occurs, or  
481 when different ecosystems are studied. Nevertheless, the complexity of the global cycles and the  
482 feedbacks between abiotic and biotic processes pose great challenges to accurate prediction.

483

## 3 Discussion

484

485 Our review of the practice of prediction in ecology-related fields showed that some fields achieve  
486 relatively high predictive proficiency. Both phenomenological and mechanistic approaches are  
487 used in most fields, but when direct comparisons are made, differences are often small, with no  
488 clear superiority of one approach over the other. Comparisons are hampered by proficiency being  
489 evaluated in many different ways: different metrics are used within and among fields, and  
490 predictive proficiency is judged differently for phenomenological and mechanistic models. Data  
491 availability is limiting in most fields, with long-term studies being rare, and detailed data for  
492 parameterization of mechanistic models being in short supply. Learning from these examples, we  
493 give suggestions as to how we may improve the practice of prediction in ecology, summarized in  
494 the “forecasting loop” (Figure 2).

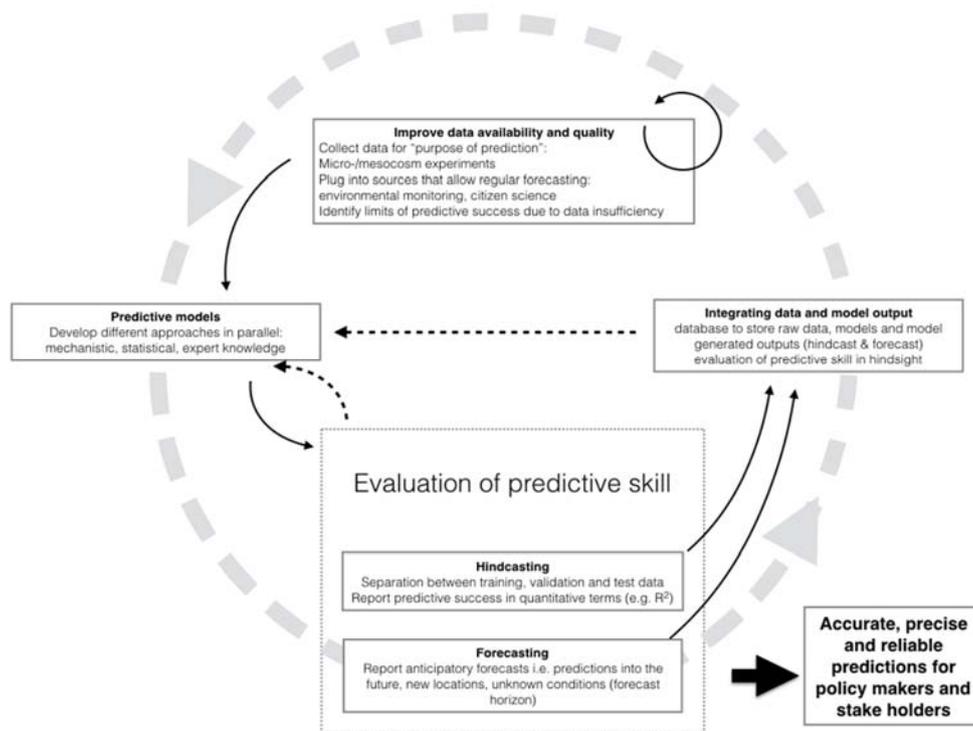
495

### 496 **The practice of prediction in ecology-related fields**

497 We found that truly anticipatory predictions are not very common in applied fields, even though  
498 some are reported to government agencies (e.g. epidemiology, fisheries), with hindcasts more  
499 commonly used to assess predictive proficiency. Epidemiology is one of the most predictive fields,  
500 with high predictive proficiency achieved ( $R^2 > 0.9$ ). Whenever new pathogens arise (e.g. Zika  
501 virus) anticipatory predictions are made, evaluated in real-time, and also hindsight (e.g.,  
502 evaluation of models used during the foot and mouth epidemic in the UK in 2001). Another  
503 indication for the level of sophistication obtained in epidemiology is the integration of evolutionary  
504 processes in predictive models (Gandon et al., 2016).

505

506 Fisheries also report yearly forecasts to government agencies; however, we are not aware of  
507 rigorous tests of truly anticipatory predictions. Nevertheless, hindcasting using a  
508 phenomenological approach provides a baseline of average predictive proficiency ( $\rho$  0.4) in  
509 fisheries for different time frames (Glaser et al., 2014). Lower proficiency may be due to  
510 uncertainty about the abundances, or time lags between forecasts and the reporting of stock  
511 assessments (Brander, 2003).



512

513 Figure 2. The “forecasting loop” to improve the practice of prediction in ecology. Improved data availability  
 514 and quality will help to develop predictive models (in the widest sense) that are not limited by insufficient  
 515 data quality. Both hindcasting and forecasting skills of these predictive models will be tested rigorously to  
 516 quantify the current state of predictive proficiency. Such evaluations are facilitated by making data, models  
 517 and model output available in a common database. Importantly, improvements in predictive proficiency are  
 518 expected to be greater if forecasts are made and reported on a regular basis (indicated by the large loop).  
 519 Inner loops illustrate that regular and long-term data are important and provide opportunity to evaluate  
 520 forecasts. The goal is to provide accurate, precise and reliable predictions to policy makers and stake  
 521 holders to inform about and adapt to recent challenges such as global environmental change.

522

523 Generally, we found that models dominated by basic physical and chemical processes are often  
 524 better at predicting than models that do not have strong drivers. Epidemics with strong seasonal  
 525 forcing and eutrophication are good examples. Eutrophication models show high  $R^2$  for  
 526 environmental properties, but considerably lower skill for the biotic component. This indicates that  
 527 the properties of the biota are still not sufficiently well understood, and/or that biological processes  
 528 involve complexities inherently more difficult to model. Forestry predictions are successful when  
 529 a lot is known about the effects of environmental conditions on the survival and growth of  
 530 individuals (especially shade tolerance), and their life-cycle characteristics. The relative  
 531 importance of physical and biotic drivers hence may contribute to the predictability of a variable.

532 Whereas physical processes often arise from a single mechanism (or a small number thereof),  
533 biological processes often arise from a large, complex system of interacting sub-processes. While  
534 these sub-processes may be individually mathematically and experimentally tractable, the  
535 complete system rarely is. Consequently, models of biological processes are typically coarse-  
536 grained approximations of these complex systems with a resulting increase in uncertainty. By  
537 contrast, physical processes can be much more completely characterised and thus are more  
538 predictable.

539

540 We divided models into phenomenological and mechanistic models. It is often argued that  
541 mechanistic models are superior under changing conditions if they capture the correct  
542 mechanisms (Evans et al., 2013; Stillman et al., 2015). A few studies compared mechanistic to  
543 phenomenological approaches, with phenomenological models often being equal on average  
544 (inferior performance in eutrophication models, but slightly better predictions in some  
545 epidemiology studies). For the time being it seems that no approach is clearly superior in terms  
546 of predictive proficiency, but this may be because challenging and novel predictions are rarely  
547 formulated. Therefore, we should be agnostic about the approach and rigorously assess the  
548 accuracy and precision of our predictions. Comparing mechanistic and phenomenological  
549 approaches, we also found that predictions are seldom evaluated with a set of standard metrics,  
550 which can reveal complementary aspects of predictive proficiency. Mechanistic models are more  
551 often judged by qualitative assessments (e.g. do observations fall within the model confidence  
552 intervals), whereas phenomenological approaches usually rely on single metrics such as  $R^2$  or  
553 correlation coefficients. Meta-analyses of predictive proficiency would be facilitated by either using  
554 a standard set of evaluation metrics, including deviations between predictions and observations  
555 (e.g. RMSE) and the range of predicted values (e.g. the specificity of predictions).

556

557 Some technical issues regarding the application of complex simulation models to prediction  
558 should also be mentioned. The accuracy of a prediction is not only contingent on the  
559 parameterization of mechanistic equations capturing relevant processes (*parameter sensitivity*),  
560 but even small, purely quantitative, errors in parameterization can lead to inaccuracy of  
561 predictions if the system is *structurally sensitive* (Adamson and Morozov, 2014; Cordoleani et al.,  
562 2011; Wood and Thomas, 1999).

563 The relevant outputs of mechanistic models are also usually the asymptotic dynamics. For that  
564 reason, simulations usually include an initial period of considerable length to get rid of transient  
565 effects of arbitrary initial conditions. However, long-term transients have been found in spatio-

566 temporal predator-prey models that last far longer than the effects of initial conditions (Banerjee  
567 and Petrovskii, 2010; Rodrigues et al., 2011). Therefore, more attention should be paid to  
568 transient behaviors, to determine whether they are spurious or, possibly in some cases, important  
569 parts of predicted behavior (Hastings, 2004, 2001).

## 570 **The role of data in ecological prediction**

571 The quality and availability of data is essential for developing a predictive science, and better data  
572 archiving practices will improve access to data (Mouquet et al., 2015). High quality data are  
573 needed to parameterize models and test explanatory predictions, which help us to understand  
574 systems. Low quality data compromise our ability to test models rigorously, as it is unknown  
575 whether predictions are incorrect due to inappropriate data or to poorly specified models (i.e., due  
576 to model or parameter uncertainty). Simulated data can be very valuable for discovering whether  
577 modeling approaches are able to recover the parameters that generated the data after adding  
578 moderate levels of noise and are commonly used to show the usefulness of the modeling  
579 technique (e.g. Pascual and Kareiva, 1996). However, data collectors need to know the noise  
580 levels where modeling approaches will fail to recover signals from data: this will help design  
581 measurement and sampling schemes guaranteeing appropriate data for predictive models. Other  
582 important data properties are sampling frequency, duration of time series, and spatial replicates.  
583 More dialogue between modelers and practitioners in terms of data collection is clearly needed  
584 and could foster the collection of data specifically for the purpose of prediction. Micro- and  
585 mesocosms are widely used and very well suited tools to generate time-series of population,  
586 community and ecosystem dynamics (Altermatt et al., 2015; Fraser and Keddy, 1997; Resetarits  
587 and Bernardo, 1998). Because the study organisms are fast-growing and small, they are  
588 amenable to frequent monitoring. Experimental systems can furthermore be manipulated to study  
589 the ability of models to capture press or pulse perturbations.

590  
591 Data availability is as crucial as data quality to foster the practice of prediction. Ideally, data should  
592 be long-term and real-time, such that anticipatory predictions can be made and checked with the  
593 smallest delay. This seems in reach for epidemiology; however, the majority of studies reviewed  
594 evaluated predictive proficiency by hindcasts. An alternative to collecting data for the purpose of  
595 prediction is to integrate data from governmental environmental monitoring schemes such as the  
596 long-term ecological research network (LTER, <https://www.lternet.edu/>) (Niu et al., 2014) or the  
597 collection of phenology data (monitoring by national meteorological agencies). Another source of  
598 regular data could be the use of volunteer-based monitoring schemes (for instance for birds or

599 butterflies). These citizen science projects have the advantage of often covering large geographic  
600 areas (national to continental) and are designed to run for decades. However, careful  
601 assessments are needed to ensure data quality and comparability (Isaac et al., 2014).  
602 Nevertheless, making frequent predictions (e.g. abundance trends), and checking their  
603 precision/accuracy in hindsight, would allow comparison of different model types in terms of  
604 predictive success, as well as definition of an ecology-wide baseline of predictive success.

## 605 **Ensuring representative predictive proficiency of models and towards** 606 **a gold standard of prediction**

607 The majority of the predictive models reviewed here make hindcasts: cross-validations where a  
608 set of data is partitioned into a training set, to which the model is calibrated, and a test data set  
609 which the model aims to predict. Because of that they fall short of anticipatory predictions, for  
610 several reasons. One major issue with hindcasts is that the modelers have access to the test  
611 data. Therefore, test and training data sets are unlikely to be independent, as the training data  
612 set is unlikely to be chosen if it is not representative of the time series as a whole. In addition,  
613 investigators have unlimited attempts to predict (Franks, 2009), usually without indicating how  
614 many times it has failed. Complete failures or low success of models to predict a test data set will  
615 even go unpublished, limiting our ability to determine a representative measure of predictive  
616 success.

617 The predominance of hindcasts may cause several biases in the predictive power of ecological  
618 models, both in terms of the general predictive ability of a field and the predictive proficiency of  
619 certain model types relative to one another. They may be misleading because a model with  
620 enough degrees of freedom is likely to perform well through 'overfitting', regardless of how well it  
621 represents the scenario it is predicting. In this case, the predictive ability of complex mechanistic  
622 models may be overstated, because they have greater numbers of unconstrained parameters,  
623 and are therefore often underspecified given that data in ecology are hard to come by. The  
624 predictive ability of sufficiently flexible phenomenological/statistical models may be overstated as  
625 well, because of their focus on reproducing observations instead of incorporating mechanisms  
626 (Wenger and Olden, 2012). On the other hand, the predictive ability of simple mechanistic or more  
627 constraining phenomenological models could be underestimated. Tools to deal with over-fitting  
628 such as Akaike Information Criterion and procedures to systematically simplify complex IBMs

629 (pattern oriented modelling) exists and can help to tackle these issues (Burnham and Anderson,  
630 2002; Grimm and Railsback, 2012).

631 For these reasons, we should improve the way in which predictions are evaluated in ecology. This  
632 does not entail a complete rejection of hindcasting, but an attempt to mitigate the problems  
633 associated with retrospective predictions playing ‘too safe’. First, test and training data should be  
634 kept as independent as possible, potentially keeping the test data inaccessible to the researchers  
635 making the prediction (e.g. by using a database). The modelers can later submit predictions that  
636 are compared against the test data by an independent party. To keep test and training data  
637 independent, they should naturally be kept separated in time, but should also ideally be obtained  
638 from different locations and scenarios in order to properly test the model’s general applicability  
639 beyond its calibration data—also known as transferability (Wenger and Olden, 2012). This is the  
640 rationale behind the use of ‘space-for-time’ substitution in forest succession modeling, in which  
641 predictive models are calibrated for new forests, run for time periods of hundreds of years, and  
642 subsequently tested for their ability to predict nearby old growth forests. In contrast with  
643 hindcasting, anticipatory predictive studies automatically guarantee that modelers don’t have  
644 access to test data beforehand. The genuine anticipatory prediction of a different situation to  
645 which the model has been calibrated also guarantees independence of the test and training data,  
646 and is therefore the “gold standard” of prediction.

## 647 **How should we report predictions?**

648 Anticipatory predictions are rarely formulated and even less frequently checked in hindsight, even  
649 when predictions are reported to government agencies on a regular basis. This is surprising, as  
650 predictions could easily be checked, as soon as new data becomes available. Hence, we suggest  
651 to make predictions in the first place so we have something to compare with when new data  
652 become available. A good example is Glaser et al. (2014), who used hindcasts to test the  
653 predictive proficiency of their model but also provide an anticipatory prediction for the next year,  
654 for which data was not yet available. A rigorous assessment of proficiency would require us to  
655 collect the predictions (ideally for a number of steps into the future) in a database with specific  
656 information about the model and data used. Whereas databases with population dynamic data  
657 (e.g., Global Population Dynamics Database, GPDB) and stock assessments (RAM legacy  
658 database) are available, we are not aware of databases that store model predictions and thereby  
659 allow the quantification of predictive proficiency. One could perform model inter-comparisons to  
660 evaluate their proficiency or rely on ensemble forecasts to study the consistency of different model

661 types (e.g. statistical versus mechanistic). This may help to quantitatively disentangle whether  
662 certain ecological levels of organization, processes or organism properties are related to  
663 predictive success.

664 A major difficulty in comparing predictive proficiency across fields was the great diversity in how  
665 predictive proficiency was reported. A diversity of measures was used across studies, including  
666 correlation coefficients between predictions and observations, different measures of error (the  
667 absolute or relative difference between predictions and observations) as well as whether  
668 observations fell within the confidence bounds of predictions. This is true even in more predictive  
669 fields such as epidemiology (e.g. Nsoesie et al., 2014), which has a good practice of reporting  
670 and communicating predictions but in which no single metric is consistently reported. Whereas  
671 there may have been good reasons to choose one measure over the other in specific studies, we  
672 would advocate either the use of a single metric for reporting predictive proficiency in ecology, or  
673 the use of several complementary metrics that capture different aspects of predictive success.  
674 For time series, several recent studies (Garland et al., 2014; Ward et al., 2014) used the mean  
675 absolute scaled error (Hyndman and Koehler, 2006) which facilitates assessment of forecast  
676 accuracy within and among time series and hence facilitate comparisons. On the other hand,  
677 Olsen et al. (2016) report different metrics of predictive success to assess the predictive ability of  
678 a large ecosystem model and conclude that only the combination of metrics captures the different  
679 aspects of predictive proficiency. Ideally, predictions would be stored in a database, such that  
680 several metrics could be calculated across a variety of predictions and observations.

## 681 **The practice of prediction and evidence-based policy**

682 Although most models reviewed were quantitative, predictions do not need to rely on  
683 mathematical models to be useful. Evidence-based policy requires interactions among various  
684 groups and can take evidence from various sources. Policy makers and practitioners make  
685 decisions about how to achieve a particular outcome, or to solve a particular problem; e.g. how  
686 to control the outbreak of a particular disease, or how to reverse the decline in abundance of a  
687 threatened species. Predictions are intrinsic in any decision, in that one is predicting that the  
688 course of action decided upon will have the desired outcome.

689  
690 In the case of controlling the 2001 foot and mouth disease outbreak in the UK, scientists quickly  
691 developed mathematical models, which were used to make forecasts about the effects of  
692 alternate vaccination and culling policies (Ferguson et al., 2001). When mathematical models are

693 unavailable, or cannot be developed in time, decisions can be informed by other sources of  
694 evidence, including expert opinion. Expert opinion can be defined as a prediction made by a  
695 specialist with extensive experience and expertise relevant to the problem at hand. Unfortunately,  
696 expert opinion must be treated with great care, due to the thoroughly demonstrated cognitive  
697 biases that greatly reduce its predictive value (Sutherland and Burgman, 2015; Tetlock, 2006).  
698 One solution to the frailty of expert opinion is to train experts in the practices associated with  
699 accurate forecasting.

700  
701 Groups and individuals making forecasts should attempt to first collate and apply relevant  
702 objective evidence to the problem at hand (the *outside view*). Evidence from comparable  
703 situations, and from relevant empirical studies, can be used as a baseline. For example, during  
704 the 2001 foot and mouth outbreak, comparisons with previous outbreak, particularly the large  
705 outbreak in 1967, and past interventions, were made. Only after application of the *outside view*,  
706 predictions are modified according to particular characteristics of the specific problem at hand  
707 (the *inside view*). As for quantitative models, assessment of predictive proficiency is an essential  
708 feedback to experts, especially when operating in groups that share evidence on which forecasts  
709 are based. Assembling empirical evidence (outside view) in advance of needing it for a particular  
710 policy decision has been termed “solution scanning” and explicitly involves decision makers  
711 (Sutherland et al., 2014).

712  
713 All of the previous points share similarities with systematic reviews in medicine and evidence-  
714 based conservation, which require careful assessment by experts to compile the evidence,  
715 assess potential confounding factors, and make these available in forms of databases that are  
716 frequently updated (<http://www.cebc.bangor.ac.uk/ebconservation.php>). It also shows how  
717 important it is to still keep humans “in the loop” in terms of checking model predictions. Indeed,  
718 even in fields like meteorology, which showed impressive gains in predictive proficiency, model  
719 predictions are still cross-checked by experts for errors (Doswell, 2004).

720

## 721 **Making predictions relevant beyond predictive proficiency**

722 While we argued here for rigorous assessment of predictive proficiency, we acknowledge that  
723 improving predictive proficiency requires resources and may only be justified if there is a higher  
724 payback in terms of better informed policy. Analytical frameworks to assess the value of  
725 information are available from decision theory and applied in fields such as health economics

726 (Claxton et al., 2002) or conservation biology (Canessa et al., 2015). Such analysis requires clear  
727 specification of the uncertainty of alternative decisions under a suite of hypotheses (scenarios  
728 describing what the future may look like) to evaluate the costs of certain actions (Canessa et al.,  
729 2015). Based on this information, scientists can decide whether it is worthwhile to collect further  
730 information, and if so prioritize where reduced uncertainty will yield the highest pay backs.  
731 Coupling predictive models with socio-economic models may also help to account for economic  
732 constraints and also better understand when sociological factors limit the adoption of evidence  
733 based policy (Sutherland and Freckleton, 2012).

734 Another important aspect to consider for scientists is that counterintuitively, better predictive  
735 proficiency does not necessarily lead to better decisions (Pielke and Conant, 2003). This is  
736 because science is not directly translated into decisions, but is only part of the decision making  
737 process, together with communication and the multiple constraints (i.e. societal and economic)  
738 that need to be balanced (Pielke and Conant, 2003). Communication of the inherent uncertainty  
739 of ecological predictions to policy makers is essential, as well as considering the needs of policy  
740 makers in terms of ecological evidence (Sutherland and Freckleton, 2012). One major factor for  
741 the adoption of ecological prediction is the experience (exposure to and ability to assess the  
742 quality of predictions) that policy makers have with ecological predictions (Pielke and Conant,  
743 2003). Only predictions considered useful for decisions will be incorporated in the decision making  
744 process. These points are essential to make ecological predictions more relevant to policy makers  
745 and may be as important as improving predictive proficiency itself.

## 746 **Conclusions**

747 Global environmental change poses many threats to natural ecosystems and global biodiversity.  
748 Hence, there is a pressing need for anticipatory predictions, which will help to foresee, manage  
749 and adapt to the effects of global change (Mouquet et al., 2015; Petchey et al., 2015). Ecologists  
750 have come a long way towards making their science more quantitative and have developed the  
751 habit of testing theories using explanatory predictions. Now there is an urgent need to follow the  
752 example of other fields to develop a rigorous practice of prediction to inform policy makers and  
753 the public. More anticipatory predictions, as well as critical evaluation of predictive proficiency,  
754 are needed in ecology to define the baseline of predictive proficiency and we propose various  
755 ways how to foster such a practice of prediction summarized by the forecasting loop. We believe  
756 that such concerted actions by ecologists may lead to larger gains in predictive proficiency in the  
757 long run, and will lead to more accurate and precise predictions to inform policy makers and stake  
758 holders.

759 **Acknowledgements**

760 We thank Jennifer Dunne, Alan Hastings and Andrew Morozov for the opportunity to participate  
761 in the current topic workshop on “Predictability, Uncertainty and Sensitivity in Ecology:  
762 Mathematical challenges and ecological applications” hosted by the Mathematical Biosciences  
763 Institute (MBI), Columbus, Ohio 26<sup>th</sup> to 30<sup>th</sup> of October 2015. We thank Henri Laurie, Thomas M.  
764 Massie, David Nerini and Paulo Tilles and the other workshop participants for discussion and  
765 stimulating talks at the MBI, Columbus, Ohio as well as three reviewers whose comments have  
766 led to a much improved manuscript. FP and OLP were financially supported by Swiss National  
767 Science Foundation Grant 31003A\_159498. DLD was supported by the USGS’s Greater  
768 Everglades Priority Ecosystem Research program. MA was funded by DENFREE (grant 282378)  
769 and supported by Fundação para a Ciência e a Tecnologia (grant UID/MAT/04561/2013).

770

# References

- 772 Adamson, M.W., Morozov, A.Y., 2014. Defining and detecting structural sensitivity in biological  
773 models: developing a new framework. *J. Math. Biol.* 69, 1815–1848.  
774 doi:10.1007/s00285-014-0753-3
- 775 Aguiar, M., Coelho, G.E., Rocha, F., Mateus, L., Pessanha, J.E.M., Stollenwerk, N., 2015.  
776 Dengue transmission during the 2014 FIFA World Cup in Brazil. *Infect. Dis.* 15, 765–  
777 766.
- 778 Aguiar, M., Paul, R., Sakuntabhai, A., N, S., 2014. Are we modelling the correct dataset?  
779 Minimizing false predictions for dengue fever in Thailand. *Epidemiol Infect* 142, 2447–  
780 2459.
- 781 Aguiar, M., Rocha, F., Pessanha, J.E.M., Mateus, L., Stollenwerk, N., 2015. Carnival or football,  
782 is there a real risk for acquiring dengue fever in Brazil during holidays seasons? *Sci.*  
783 *Rep.* 5, 8462. doi:10.1038/srep08462
- 784 Altermatt, F., Fronhofer, E.A., Garnier, A., Giometto, A., Hammes, F., Klecka, J., Legrand, D.,  
785 Mächler, E., Massie, T.M., Pennekamp, F., Plebani, M., Pontarp, M., Schtickzelle, N.,  
786 Thuillier, V., Petchey, O.L., 2015. Big answers from small worlds: a user's guide for  
787 protist microcosms as a model system in ecology and evolution. *Methods Ecol. Evol.* 6,  
788 218–231. doi:10.1111/2041-210X.12312
- 789 Arhonditsis, G.B., Brett, M.T., others, 2004. Evaluation of the current state of mechanistic  
790 aquatic biogeochemical modeling. *Mar. Ecol. Prog. Ser.* 271, 13–26.
- 791 Asner, G.P., Vitousek, P.M., 2005. Remote analysis of biological invasion and biogeochemical  
792 change. *Proc. Natl. Acad. Sci. U. S. A.* 102, 4383–4386. doi:10.1073/pnas.0500823102
- 793 Banerjee, M., Petrovskii, S., 2010. Self-organised spatial patterns and chaos in a ratio-  
794 dependent predator–prey system. *Theor. Ecol.* 4, 37–53. doi:10.1007/s12080-010-0073-  
795 1
- 796 Barber, M.C., 2003. A review and comparison of models for predicting dynamic chemical  
797 bioconcentration in fish. *Environ. Toxicol. Chem.* 22, 1963–1992. doi:10.1897/02-468
- 798 Bauer, P., Thorpe, A., Brunet, G., 2015. The quiet revolution of numerical weather prediction.  
799 *Nature* 525, 47–55. doi:10.1038/nature14956
- 800 Beckage, B., Gross, L.J., Kauffman, S., 2011. The limits to prediction in ecological systems.  
801 *Ecosphere* 2, art125. doi:10.1890/ES11-00211.1
- 802 Beverton, R.J.H., Holt, S.J., 1957. *On the Dynamics of Exploited Fish Populations*. Springer  
803 Science & Business Media.
- 804 Bjørnstad, O.N., Finkenstädt, B.F., Grenfell, B.T., 2002. Dynamics of Measles Epidemics:  
805 Estimating Scaling of Transmission Rates Using a Time Series Sir Model. *Ecol. Monogr.*  
806 72, 169–184. doi:10.1890/0012-9615
- 807 Botkin, D.B., 1993. *Forest Dynamics: An Ecological Model*. Oxford University Press.
- 808 Botkin, D.B., Janak, J.F., Wallis, J.R., 1972. Some Ecological Consequences of a Computer  
809 Model of Forest Growth. *J. Ecol.* 60, 849–872. doi:10.2307/2258570
- 810 Brander, K.M., 2003. What kinds of fish stock predictions do we need and what kinds of  
811 information will help us make better predictions? *Sci. Mar.* 67, 21–33.
- 812 Burnham, K.P., Anderson, D.R., 2002. *Model Selection and Multi-Model Inference: A Practical  
813 Information-Theoretic Approach*. Springer, New York.
- 814 Canessa, S., Guillera-Aroita, G., Lahoz-Monfort, J.J., Southwell, D.M., Armstrong, D.P.,  
815 Chadès, I., Lacy, R.C., Converse, S.J., 2015. When do we need more data? A primer on  
816 calculating the value of information for applied ecologists. *Methods Ecol. Evol.* 6, 1219–  
817 1228. doi:10.1111/2041-210X.12423

818 Carr, M.-E., Friedrichs, M.A.M., Schmeltz, M., Noguchi Aita, M., Antoine, D., Arrigo, K.R.,  
819 Asanuma, I., Aumont, O., Barber, R., Behrenfeld, M., Bidigare, R., Buitenhuis, E.T.,  
820 Campbell, J., Ciotti, A., Dierssen, H., Dowell, M., Dunne, J., Esaias, W., Gentili, B.,  
821 Gregg, W., Groom, S., Hoepffner, N., Ishizaka, J., Kameda, T., Le Quéré, C., Lohrenz,  
822 S., Marra, J., Mélin, F., Moore, K., Morel, A., Reddy, T.E., Ryan, J., Scardi, M., Smyth,  
823 T., Turpie, K., Tilstone, G., Waters, K., Yamanaka, Y., 2006. A comparison of global  
824 estimates of marine primary production from ocean color. *Deep Sea Res. Part II Top.*  
825 *Stud. Oceanogr.* 53, 741–770. doi:10.1016/j.dsr2.2006.01.028

826 Caudron, Q., Mahmud, A.S., Metcalf, C.J.E., Gottfreðsson, M., Viboud, C., Cliff, A.D., Grenfell,  
827 B.T., 2015. Predictability in a highly stochastic system: final size of measles epidemics in  
828 small populations. *J. R. Soc. Interface* 12, 20141125. doi:10.1098/rsif.2014.1125

829 Clark, J.S., Carpenter, S.R., Barber, M., Collins, S., Dobson, A., Foley, J.A., Lodge, D.M.,  
830 Pascual, M., Jr, R.P., Pizer, W., Pringle, C., Reid, W.V., Rose, K.A., Sala, O.,  
831 Schlesinger, W.H., Wall, D.H., Wear, D., 2001. Ecological Forecasts: An Emerging  
832 Imperative. *Science* 293, 657–660. doi:10.1126/science.293.5530.657

833 Claxton, K., Sculpher, M., Drummond, M., 2002. A rational framework for decision making by  
834 the National Institute For Clinical Excellence (NICE). *The Lancet* 360, 711–715.  
835 doi:10.1016/S0140-6736(02)09832-X

836 Clebsch, E.E.C., Busing, R.T., 1989. Secondary Succession, Gap Dynamics, and Community  
837 Structure in a Southern Appalachian Cove Forest. *Ecology* 70, 728–735.  
838 doi:10.2307/1940223

839 Conley, D.J., Paerl, H.W., Howarth, R.W., Boesch, D.F., Seitzinger, S.P., Karl, E., Karl, E.,  
840 Lancelot, C., Gene, E., Gene, E., 2009. Controlling eutrophication: nitrogen and  
841 phosphorus. *Science* 123, 1014–1015.

842 Cordoleani, F., Flora, C., Nerini, D., David, N., Gauduchon, M., Mathias, G., Morozov, A.,  
843 Andrew, M., Poggiale, J.-C., Jean-Christophe, P., 2011. Structural sensitivity of  
844 biological models revisited. *J. Theor. Biol.* 283, 82–91. doi:10.1016/j.jtbi.2011.05.021

845 Cury, P., Fromentin, J.M., Bonhommeau, S., 2014. Resolving Hjort's Dilemma : How is  
846 recruitment related to spawning stock biomass in marine fish? *Oceanography* 27, 42–47.

847 Deyle, E.R., Fogarty, M., Hsieh, C., Kaufman, L., MacCall, A.D., Munch, S.B., Perretti, C.T., Ye,  
848 H., Sugihara, G., 2013. Predicting climate effects on Pacific sardine. *Proc. Natl. Acad.*  
849 *Sci.* 110, 6430–6435. doi:10.1073/pnas.1215506110

850 Doswell, C.A., 2004. Weather Forecasting by Humans—Heuristics and Decision Making.  
851 *Weather Forecast.* 19, 1115–1126. doi:10.1175/WAF-821.1

852 Elliott, J.A., Irish, A.E., Reynolds, C.S., 2010. Modelling Phytoplankton Dynamics in Fresh  
853 Waters: Affirmation of the PROTECH Approach to Simulation. *Freshw. Rev.* 3, 75–96.  
854 doi:10.1608/FRJ-3.1.4

855 Evans, M.R., Bithell, M., Cornell, S.J., Dall, S.R.X., Díaz, S., Emmott, S., Ernande, B., Grimm,  
856 V., Hodgson, D.J., Lewis, S.L., Mace, G.M., Morecroft, M., Moustakas, A., Murphy, E.,  
857 Newbold, T., Norris, K.J., Petchey, O., Smith, M., Travis, J.M.J., Benton, T.G., 2013.  
858 Predictive systems ecology. *Proc. R. Soc. B Biol. Sci.* 280, 20131452.  
859 doi:10.1098/rspb.2013.1452

860 Evans, M.R., Moustakas, A., 2016. A comparison between data requirements and availability for  
861 calibrating predictive ecological models for lowland UK woodlands: learning new tricks  
862 from old trees. *Ecol. Evol.* 6, 4812–4822. doi:10.1002/ece3.2217

863 Ferguson, N.M., Donnelly, C.A., Anderson, R.M., 2001. The foot-and-mouth epidemic in Great  
864 Britain: pattern of spread and impact of interventions. *Science* 292, 1155–60.  
865 doi:10.1126/science.1061020

866 Francis, M.P., Morrison, M.A., Leathwick, J., Walsh, C., Middleton, C., 2005. Predictive models  
867 of small fish presence and abundance in northern New Zealand harbours. *Estuar. Coast.*  
868 *Shelf Sci.* 64, 419–435. doi:10.1016/j.ecss.2005.03.007

869 Franks, P.J.S., 2009. Planktonic ecosystem models: perplexing parameterizations and a failure  
870 to fail. *J. Plankton Res.* 31, 1299–1306. doi:10.1093/plankt/fbp069

871 Fraser, L.H., Keddy, P., 1997. The role of experimental microcosms in ecological research.  
872 *Trends Ecol. Evol.* 12, 478–481.

873 Friedrichs, M.A.M., Carr, M.-E., Barber, R.T., Scardi, M., Antoine, D., Armstrong, R.A.,  
874 Asanuma, I., Behrenfeld, M.J., Buitenhuis, E.T., Chai, F., Christian, J.R., Ciotti, A.M.,  
875 Doney, S.C., Dowell, M., Dunne, J., Gentili, B., Gregg, W., Hoepffner, N., Ishizaka, J.,  
876 Kameda, T., Lima, I., Marra, J., Mélin, F., Moore, J.K., Morel, A., O'Malley, R.T.,  
877 O'Reilly, J., Saba, V.S., Schmeltz, M., Smyth, T.J., Tjiputra, J., Waters, K., Westberry,  
878 T.K., Winguth, A., 2009. Assessing the uncertainties of model estimates of primary  
879 productivity in the tropical Pacific Ocean. *J. Mar. Syst., Skill assessment for coupled  
880 biological/physical models of marine systems* 76, 113–133.  
881 doi:10.1016/j.jmarsys.2008.05.010

882 Froelich, B., Bowen, J., Gonzalez, R., Snedeker, A., Noble, R., 2013. Mechanistic and statistical  
883 models of total *Vibrio* abundance in the Neuse River Estuary. *Water Res.* 47, 5783–  
884 5793.

885 Gandon, S., Day, T., Metcalf, C.J.E., Grenfell, B.T., 2016. Forecasting Epidemiological and  
886 Evolutionary Dynamics of Infectious Diseases. *Trends Ecol. Evol.* 31, 776–788.  
887 doi:10.1016/j.tree.2016.07.010

888 Garcia, S.M., Grainger, R.J.R., 2005. Gloom and doom? The future of marine capture fisheries.  
889 *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 21–46. doi:10.1098/rstb.2004.1580

890 Garland, J., James, R., Bradley, E., 2014. Model-free quantification of time-series predictability.  
891 *Phys. Rev. E* 90, 52910.

892 Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., Brilliant, L., 2009.  
893 Detecting influenza epidemics using search engine query data. *Nature* 457, 1012–1014.  
894 doi:10.1038/nature07634

895 Glaser, S.M., Fogarty, M.J., Liu, H., Altman, I., Hsieh, C.-H., Kaufman, L., MacCall, A.D.,  
896 Rosenberg, A.A., Ye, H., Sugihara, G., 2014. Complex dynamics may limit prediction in  
897 marine fisheries. *Fish Fish.* 15, 616–633. doi:10.1111/faf.12037

898 Glaser, S.M., Ye, H., Sugihara, G., 2014. A nonlinear, low data requirement model for producing  
899 spatially explicit fishery forecasts. *Fish. Oceanogr.* 23, 45–53. doi:10.1111/fog.12042

900 Grenfell, B., Keeling, M., 2008. Dynamics of infectious disease. *Theor. Ecol. 3rd Ed Oxf. Univ.*  
901 *Press Oxf.* 132–147.

902 Grenfell, B.T., Bjørnstad, O.N., Finkenstädt, B.F., 2002. Dynamics of Measles Epidemics:  
903 Scaling Noise, Determinism, and Predictability with the TSIR Model. *Ecol. Monogr.* 72,  
904 185–202. doi:10.2307/3100024

905 Grimm, V., Railsback, S.F., 2012. Pattern-oriented modelling: a “multi-scope” for predictive  
906 systems ecology. *Philos. Trans. R. Soc. B Biol. Sci.* 367, 298–310.  
907 doi:10.1098/rstb.2011.0180

908 Guisan, A., Zimmermann, N.E., Elith, J., Graham, C.H., Phillips, S., Peterson, A.T., 2007.  
909 WHAT MATTERS FOR PREDICTING THE OCCURRENCES OF TREES:  
910 TECHNIQUES, DATA, OR SPECIES CHARACTERISTICS? *Ecol. Monogr.* 77, 615–630.

911 Hamilton, D.P., Schladow, S.G., 1997. Prediction of water quality in lakes and reservoirs. Part I  
912 — Model description. *Ecol. Model.* 96, 91–110. doi:10.1016/S0304-3800(96)00062-2

913 Hamrick, J.M., 1992. A three-dimensional environmental fluid dynamics computer code:  
914 Theoretical and computational aspects. Virginia Institute of Marine Science, College of  
915 William and Mary.

916 Hastings, A., 2004. Transients: the key to long-term ecological understanding? *Trends Ecol.*  
917 *Evol.* 19, 39–45. doi:10.1016/j.tree.2003.09.007

918 Hastings, A., 2001. Transient dynamics and persistence of ecological systems. *Ecol. Lett.* 4,  
919 215–220. doi:10.1046/j.1461-0248.2001.00220.x

920 Houtekamer, P.L., Lefavre, L., Derome, J., Ritchie, H., Mitchell, H.L., 1996. A System  
921 Simulation Approach to Ensemble Prediction. *Mon. Weather Rev.* 124, 1225–1242.  
922 doi:10.1175/1520-0493

923 Hsieh, C., Glaser, S.M., Lucas, A.J., Sugihara, G., 2005. Distinguishing random environmental  
924 fluctuations from ecological catastrophes for the North Pacific Ocean. *Nature* 435, 336–  
925 340. doi:10.1038/nature03553

926 Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. *Int. J.*  
927 *Forecast.* 22, 679–688. doi:10.1016/j.ijforecast.2006.03.001

928 IPCC, 2014. Summary for Policymakers, in: Field, C.B., Barros, V.R., Dokken, D.J., Mach, K.J.,  
929 Mastrandrea, M.D., Bilir, T.E., Chatterjee, M., Ebi, K.L., Estrada, Y.O., Genova, R.C.,  
930 Girma, B., Kissel, E.S., Levy, A.N., MacCracken, S., Mastrandrea, P.R., White, L.L.  
931 (Eds.), *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and*  
932 *Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the*  
933 *Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge,  
934 United Kingdom, and New York, NY, USA, pp. 1–32.

935 Isaac, N.J.B., van Strien, A.J., August, T.A., de Zeeuw, M.P., Roy, D.B., 2014. Statistics for  
936 citizen science: extracting signals of change from noisy ecological data. *Methods Ecol.*  
937 *Evol.* 5, 1052–1060. doi:10.1111/2041-210X.12254

938 Jacobs, J.M., Rhodes, M., Brown, C.W., Hood, R.R., Leight, A., Long, W., Wood, R., 2014.  
939 Modeling and forecasting the distribution of *Vibrio vulnificus* in Chesapeake Bay. *J. Appl.*  
940 *Microbiol.* 117, 1312–1327. doi:10.1111/jam.12624

941 Janse, J.H., 1997. A model of nutrient dynamics in shallow lakes in relation to multiple stable  
942 states. *Hydrobiologia* 342–343, 1–8. doi:10.1023/A:1017018812215

943 Kermack, W.O., McKendrick, 1927. A contribution to the mathematical theory of epidemics.  
944 *Proc R Soc Lond* 115, 700–720.

945 Lei, B., Huang, S., Qiao, M., Li, T., Wang, Z., 2008. Prediction of the environmental fate and  
946 aquatic ecological impact of nitrobenzene in the Songhua River using the modified  
947 AQUATOX model. *J. Environ. Sci.* 20, 769–777. doi:10.1016/S1001-0742(08)62125-7

948 Martin, B.T., Jager, T., Nisbet, R.M., Preuss, T.G., Hammers-Wirtz, M., Grimm, V., 2013.  
949 Extrapolating ecotoxicological effects from individuals to populations: a generic approach  
950 based on Dynamic Energy Budget theory and individual-based modeling. *Ecotoxicology*  
951 22, 574–583. doi:10.1007/s10646-013-1049-x

952 McGuire, A.D., Sitch, S., Clein, J.S., Dargaville, R., Esser, G., Foley, J., Heimann, M., Joos, F.,  
953 Kaplan, J., Kicklighter, D.W., others, 2001. Carbon balance of the terrestrial biosphere in  
954 the twentieth century: Analyses of CO<sub>2</sub>, climate and land use effects with four process-  
955 based ecosystem models. *Glob. Biogeochem. Cycles* 15, 183–206.

956 Mouquet, N., Lagadeuc, Y., Devictor, V., Doyen, L., Duputié, A., Eveillard, D., Faure, D.,  
957 Garnier, E., Gimenez, O., Huneman, P., Jabot, F., Jarne, P., Joly, D., Julliard, R., Kéfi,  
958 S., Kergoat, G.J., Lavorel, S., Le Gall, L., Meslin, L., Morand, S., Morin, X., Morlon, H.,  
959 Pinay, G., Pradel, R., Schurr, F.M., Thuiller, W., Loreau, M., 2015. Predictive ecology in  
960 a changing world. *J. Appl. Ecol.* 52, 1293–1310. doi:10.1111/1365-2664.12482

961 Murphy, J.M., Sexton, D.M.H., Barnett, D.N., Jones, G.S., Webb, M.J., Collins, M., Stainforth,  
962 D.A., 2004. Quantification of modelling uncertainties in a large ensemble of climate  
963 change simulations. *Nature* 430, 768–772. doi:10.1038/nature02771

964 Najjar, R.G., Jin, X., Louanchi, F., Aumont, O., Caldeira, K., Doney, S.C., Dutay, J.C., Follows,  
965 M., Gruber, N., Joos, F., Lindsay, K., Maier Reimer, E., Matear, R.J., Matsumoto, K.,  
966 Monfray, P., Mouchet, A., Orr, J.C., Plattner, G.K., Sarmiento, J.L., Schlitzer, R., Slater,  
967 R.D., Weirig, M.F., Yamanaka, Y., Yool, A., 2007. Impact of circulation on export  
968 production, dissolved organic matter, and dissolved oxygen in the ocean: Results from  
969 Phase II of the Ocean Carbon-cycle Model Intercomparison Project (OCMIP-2). *Glob.*  
970 *Biogeochem. Cycles* 21. doi:10.1029/2006GB002857

971 Ngugi, M.R., Johnson, R.W., McDonald, W.J.F., 2011. Restoration of ecosystems for  
972 biodiversity and carbon sequestration: Simulating growth dynamics of brigalow  
973 vegetation communities in Australia. *Ecol. Model.* 222, 785–794.  
974 doi:10.1016/j.ecolmodel.2010.09.022

975 Niu, S., Luo, Y., Dietze, M.C., Keenan, T.F., Shi, Z., Li, J., Iii, F.S.C., 2014. The role of data  
976 assimilation in predictive ecology. *Ecosphere* 5, art65. doi:10.1890/ES13-00273.1

977 Nsoesie, E.O., Brownstein, J.S., Ramakrishnan, N., Marathe, M.V., 2014. A systematic review  
978 of studies on forecasting the dynamics of influenza outbreaks. *Influenza Other Respir.*  
979 *Viruses* 8, 309–316. doi:10.1111/irv.12226

980 Olsen, E., Fay, G., Gaichas, S., Gamble, R., Lucey, S., Link, J.S., 2016. Ecosystem Model Skill  
981 Assessment. *Yes We Can! PLoS ONE* 11, e0146467. doi:10.1371/journal.pone.0146467

982 Paerl, H.W., Dyble, J., Moisaner, P.H., Noble, R.T., Piehler, M.F., Pinckney, J.L., Steppe, T.F.,  
983 Twomey, L., Valdes, L.M., 2003. Microbial indicators of aquatic ecosystem change:  
984 current applications to eutrophication studies. *FEMS Microbiol. Ecol.* 46, 233–246.  
985 doi:10.1016/S0168-6496(03)00200-9

986 Parton, W.J., Scurlock, J.M.O., Ojima, D.S., Gilmanov, T.G., Scholes, R.J., Schimel, D.S.,  
987 Kirchner, T., Menaut, J.-C., Seastedt, T., Garcia Moya, E., Kamnalrut, A., Kinyamario,  
988 J.I., 1993. Observations and modeling of biomass and soil organic matter dynamics for  
989 the grassland biome worldwide. *Glob. Biogeochem. Cycles* 7, 785–809.  
990 doi:10.1029/93GB02042

991 Pascual, M.A., Kareiva, P., 1996. Predicting the Outcome of Competition Using Experimental  
992 Data: Maximum Likelihood and Bayesian Approaches. *Ecology* 77, 337–349.  
993 doi:10.2307/2265613

994 Petchey, O.L., Pontarp, M., Massie, T.M., Kéfi, S., Ozgul, A., Weilenmann, M., Palamara, G.M.,  
995 Altermatt, F., Matthews, B., Levine, J.M., Childs, D.Z., McGill, B.J., Schaepman, M.E.,  
996 Schmid, B., Spaak, P., Beckerman, A.P., Pennekamp, F., Pearse, I.S., 2015. The  
997 ecological forecast horizon, and examples of its uses and determinants. *Ecol. Lett.* 18,  
998 597–611. doi:10.1111/ele.12443

999 Peters, R.H., 1991. *A Critique for Ecology*. Cambridge University Press.

1000 Pickett, S.T.A., 1989. Space-for-Time Substitution as an Alternative to Long-Term Studies, in:  
1001 Likens, G.E. (Ed.), *Long-Term Studies in Ecology*. Springer New York, pp. 110–135.  
1002 doi:10.1007/978-1-4615-7358-6\_5

1003 Pielke, R.A., Conant, R.T., 2003. Best Practices in Prediction for Decision-Making: Lessons  
1004 from the Atmospheric and Earth Sciences. *Ecology* 84, 1351–1358.

1005 Planque, B., 2016. Projecting the future state of marine ecosystems, “la grande illusion”? *ICES*  
1006 *J. Mar. Sci. J. Cons.* 73, 204–208. doi:10.1093/icesjms/fsv155

1007 Purves, D., Pacala, S., 2008. Predictive Models of Forest Dynamics. *Science* 320, 1452–1453.  
1008 doi:10.1126/science.1155359

1009 Purves, D.W., Lichstein, J.W., Strigul, N., Pacala, S.W., 2008. Predicting and understanding  
1010 forest dynamics using a simple tractable model. *Proc. Natl. Acad. Sci.* 105, 17018–  
1011 17022.

1012 Resetarits, W.J., Bernardo, J., 1998. *Experimental Ecology: Issues and Perspectives*. Oxford  
1013 University Press.

1014 Robson, B.J., 2014a. State of the art in modelling of phosphorus in aquatic systems: Review,  
1015 criticisms and commentary. *Environ. Model. Softw.* 61, 339–359.  
1016 doi:10.1016/j.envsoft.2014.01.012

1017 Robson, B.J., 2014b. When do aquatic systems models provide useful predictions, what is  
1018 changing, and what is next? *Environ. Model. Softw.* 61, 287–296.  
1019 doi:10.1016/j.envsoft.2014.01.009

1020 Rodrigues, L.A.D., Mistro, D.C., Petrovskii, S., 2011. Pattern formation in a space- and time-  
1021 discrete predator–prey system with a strong Allee effect. *Theor. Ecol.* 5, 341–362.  
1022 doi:10.1007/s12080-011-0139-8

1023 Saba, V.S., Friedrichs, M.A.M., Carr, M.-E., Antoine, D., Armstrong, R.A., Asanuma, I., Aumont,  
1024 O., Bates, N.R., Behrenfeld, M.J., Bennington, V., Bopp, L., Bruggeman, J., Buitenhuis,  
1025 E.T., Church, M.J., Ciotti, A.M., Doney, S.C., Dowell, M., Dunne, J., Dutkiewicz, S.,  
1026 Gregg, W., Hoepffner, N., Hyde, K.J.W., Ishizaka, J., Kameda, T., Karl, D.M., Lima, I.,  
1027 Lomas, M.W., Marra, J., McKinley, G.A., Mélin, F., Moore, J.K., Morel, A., O'Reilly, J.,  
1028 Salihoglu, B., Scardi, M., Smyth, T.J., Tang, S., Tjiputra, J., Uitz, J., Vichi, M., Waters,  
1029 K., Westberry, T.K., Yool, A., 2010. Challenges of modeling depth-integrated marine  
1030 primary productivity over multiple decades: A case study at BATS and HOT. *Glob.*  
1031 *Biogeochem. Cycles* 24, GB3020. doi:10.1029/2009GB003655

1032 Schindler, D.W., 1977. Evolution of Phosphorus Limitation in Lakes. *Science* 195, 260–262.

1033 Schlesinger, W.H., 1991. *Biogeochemistry: An Analysis of Global Change*, 3rd Edition.  
1034 Academic Press, Amsterdam ; Boston.

1035 Stillman, R.A., Railsback, S.F., Giske, J., Berger, U., Grimm, V., 2015. Making Predictions in a  
1036 Changing World: The Benefits of Individual-Based Ecology. *BioScience* 65, 140–150.  
1037 doi:10.1093/biosci/biu192

1038 Sugihara, G., 1994. Nonlinear Forecasting for the Classification of Natural Time Series. *Philos.*  
1039 *Trans. Phys. Sci. Eng.* 348, 477–495.

1040 Sutherland, W.J., Burgman, M.A., 2015. Use experts wisely. *Nature* 526, 317–318.

1041 Sutherland, W.J., Freckleton, R.P., 2012. Making predictive ecology more relevant to policy  
1042 makers and practitioners. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 322–330.  
1043 doi:10.1098/rstb.2011.0181

1044 Sutherland, W.J., Gardner, T., Bogich, T.L., Bradbury, R.B., Clothier, B., Jonsson, M., Kapos,  
1045 V., Lane, S.N., Möller, I., Schroeder, M., Spalding, M., Spencer, T., White, P.C.L., Dicks,  
1046 L.V., 2014. Solution scanning as a key policy tool: Identifying management interventions  
1047 to help maintain and enhance regulating ecosystem services. *Ecol. Soc.* 19, art3-art3.  
1048 doi:10.5751/ES-06082-190203

1049 Tetlock, P.E., 2006. *Expert Political Judgement: How good is it? How can we know?* Princeton  
1050 University Press, Princeton.

1051 Tracton, M.S., Kalnay, E., 1993. Operational Ensemble Prediction at the National  
1052 Meteorological Center: Practical Aspects. *Weather Forecast.* 8, 379–398.  
1053 doi:10.1175/1520-0434(1993)008<0379:OEPATN>2.0.CO;2

1054 Travers, M., Shin, Y.-J., Jennings, S., Cury, P., 2007. Towards end-to-end models for  
1055 investigating the effects of climate and fishing in marine ecosystems. *Prog. Oceanogr.*  
1056 75, 751–770. doi:10.1016/j.pocean.2007.08.001

1057 Travers-Trolet, M., Shin, Y.J., Shannon, L.J., Moloney, C.L., Field, J.G., 2014. Combined  
1058 Fishing and Climate Forcing in the Southern Benguela Upwelling Ecosystem: An End-to-  
1059 End Modelling Approach Reveals Dampened Effects. *PLOS ONE*.

1060 Travis, J., Coleman, F.C., Auster, P.J., Cury, P.M., Estes, J.A., Orensanz, J., Peterson, C.H.,  
1061 Power, M.E., Steneck, R.S., Wootton, J.T., 2014. Integrating the invisible fabric of nature  
1062 into fisheries management. *Proc. Natl. Acad. Sci.* 111, 581–584.  
1063 doi:10.1073/pnas.1305853111

1064 Trolle, D., Elliott, J.A., Mooij, W.M., Janse, J.H., Bolding, K., Hamilton, D.P., Jeppesen, E.,  
1065 2014. Advancing projections of phytoplankton responses to climate change through  
1066 ensemble modelling. *Environ. Model. Softw.* 61, 371–379.  
1067 doi:10.1016/j.envsoft.2014.01.032

1068 Vandersteen, W., 2011. Detecting gene expression profiles associated with environmental  
1069 stressors within an ecological context. *Mol. Ecol.* 20, 1322–1323. doi:10.1111/j.1365-  
1070 294X.2011.05052.x

1071 Ward, E.J., Holmes, E.E., Thorson, J.T., Collen, B., 2014. Complexity is costly: a meta-analysis  
1072 of parametric and non-parametric methods for short-term population forecasting. *Oikos*  
1073 123, 652–661. doi:10.1111/j.1600-0706.2014.00916.x  
1074 Wenger, S.J., Olden, J.D., 2012. Assessing transferability of ecological models: an  
1075 underappreciated aspect of statistical validation. *Methods Ecol. Evol.* 3, 260–267.  
1076 doi:10.1111/j.2041-210X.2011.00170.x  
1077 Wood, S.N., Thomas, M.B., 1999. Super-sensitivity to structure in biological models. *Proc. R.*  
1078 *Soc. Lond. B Biol. Sci.* 266, 565–570. doi:10.1098/rspb.1999.0673  
1079 Woolhouse, M., 2011. How to make predictions about future infectious disease risks. *Philos.*  
1080 *Trans. R. Soc. B Biol. Sci.* 366, 2045–2054. doi:10.1098/rstb.2010.0387  
1081

1082 Table 1: Overview of the ecology-related fields that were reviewed in terms of responses, model types used for predictions, measures  
 1083 of predictive success as well as strengths and challenges to achieve better predictive proficiency. Abbreviations used: RE = relative  
 1084 error, AE = absolute error, R2 = coefficient of determination, ROC = receiver operating characteristic curve, AUC = area under the  
 1085 ROC curve.

1086

field	targets of prediction (variables to be predicted)	models used (e.g. phenomenological, mechanistic)	measures of predictive success (e.g. R <sup>2</sup> , correlation coefficient, RE or AE)	strengths	challenges
Fisheries	abundance and landings, recruitment, biomass, relative abundance and occurrence, maximum sustainable yield	Empirical dynamic modeling, generalized additive models, single age or stage structured population models, EBFM model	correlation coefficient, R <sup>2</sup> , ROC curves, multiple regressions, observations within predicted boundaries	regular data collection over large spatio-temporal scales	Delay between data collection and assessments, often only short-term forecasts, insufficient data about species interactions
Epidemiology	establishment of novel pathogens, peak and timing of epidemics, effect of interventions such as vaccination, quarantining, culling, vector control, disease outbreak and invasion thresholds	SIR and extensions such as TSIR, stochastic-, statistical-, eco- epidemiological-models, individual- based model	R <sup>2</sup> , likelihood, time difference in epidemics peak, disease outbreak timing	High quality data sets covering vast spatio-temporal scales; new technologies such as mobile phone data and search engine queries can be used for real-time prediction	Inclusion of social network data, complex system models, individual oriented disease control modeling, inclusion of immune system models
Eutrophication and algal blooms	phytoplankton biomass, nitrate levels, phosphate levels	mechanistic (often with several simple statistical components), statistical models	R <sup>2</sup> and RE	relatively well-understood physical and chemical components, predictions of physical variables is good, predictive ability is high when there are strong physical drivers	Biological processes less well understood, difficult to get enough data to parameterize complex models, there's a widespread practice of bolting together models without questioning the suitability of the underlying functional forms
Ecotoxicology	Contaminant fate in environment, mortality, chemical uptake in organisms	Mechanistic models, individual-based models	Statistical comparisons with empirical data, both in the laboratory and field	Ability to use hydrologic models for contaminant fate, large amount of laboratory data for contaminant uptake and effects of toxicants on individuals	Huge numbers of new chemicals entering the environment for which few data on effects are available
Forestry	Forest succession and primary production	Individual-based models, Perfect- Plasticity Approximation, species distribution models	Statistical comparisons with species composition and production, R <sup>2</sup> , AUC	Ability to test against 'space for time' for forest succession and plantation data for wood production	Increasing the tree species for which adequate data exist to parameterize IBMs, upscaling to global level
Biogeochemistry	Nutrient flows, plant growth and nutrient uptake, plant turnover and decomposition, primary production (amount and variability) of marine systems, amount of C trapped in the ocean	Compartment models, structured in elements (C, N, P, Si, Fe)	Statistical comparisons with empirical data at various spatial and temporal scales	Models with physico-chemical components for which data exist; newly available data (satellite, AUV, ...)	Plant dynamics and decomposition processes for new systems may not be well known; some physical models (e.g. marine hydrology) still have high uncertainty; marine biogeochemical models are sensitive to herbivory, which is often uncertain

1087