Bad forecast, good decision: what makes a model relevant for determining policy

Carl Boettiger[a]

[a] *ESPM Department, University of California, 130 Mulford Hall Berkeley, CA 94720-3114, USA*

Abstract

With the rapid expansion of available data, it is frequently suggested that *forecasting performance* is the ultimate test of a scientific model. Yet the model that makes the most accurate forecasts is not always the model that yields the best decisions. This is no accident of statistics, even formal criteria such as strictly proper scoring rules cannot determine the best model for decision-making because they do not consider the context of the decision process itself. I illustrate this paradox using a classic example of ecological decision making in fisheries management. This example also reveals how an understanding of the decision problem can both explain the poor management outcomes of model which produces the best forecast, and also see how to avoid making such a mistake. As forecasting becomes an increasing attractive and viable possibility for assessing ecological models and applying them to real-world decisions, I hope this illustration serves as a reminder that modeling can not be done in a vacuum, but must reflect the context in which the model will be applied.

Keywords: transients, optimal control, adaptive management, stochasticity, uncertainty, ecological management

A primary purpose of statistical analyses and ecological modeling is to make forecasts for the future.[1,2] Accurate forecasts are important both in assessing the accuracy of our models and understanding of natural processes, they also underpin policy and decision making.[3] Yet the model that leads to the best decisions is not always the model that makes the most accurate forecasts, as I illustrate here. Ecological processes are intrinsically complex, so that even our best models can only ever be approximations of underlying processes. Even strictly proper scoring rules for probabilistic forecasts[4] will not always select the best model to guide decision-making. Surprisingly, this can even happen when a decision is derived directly from a complex optimization routine of a probabilistic predictive models.[5] Here, I use a classic, well-understood example from fisheries management[6–8] to illustrate both the paradox of how a model with the worst forecast provides the best decision outcomes, as well as show how we can avoid selecting models that are poorly suited for management by considering the management context more explicitly. These results underscore that in choosing the best model for decision-making, it can be more important to capture a single key feature of the process than it is to make the most accurate prediction about future states.

Fisheries are a significant economic and conservation concern world wide and their management remains an important debate.[9–11] Moreover, their management has been both a proving grounds for theoretical and practical decision-making issues which are widely applicable in other areas of ecology and conservation.[5,12,13] The decision-making problem is characterized by the need for a manager to set an acceptable harvest quota H_t each year given some stock assessment estimate of the current stock size (population abundance) of the species in question.[7] Such a decision problem appears to hinge on an accurate forecast: if we can predict to what size the stock will increase next year, $X_t + 1$, knowing the current stock, X_t, then we can safely harvest $X_{t+1} - X_t$. Overestimating or underestimating such recruitment will result in over-harvesting or under-harvesting, respectively. Thus it may seem natural that our first step would be to select the model

Email address: cboettig@berkeley.edu (Carl Boettiger)

that makes the most accurate forecast of next year's stock, X_{t+1}. I illustrate how we do this using strictly proper scoring criteria[4] for a set of candidate models, and show that it leads to worse decisions.

Ecological Models

For simplicity, I will focus on the classic case case of a single-species model whose population is observed annually without error in a stochastic but stationary environment without age structure.[8,11,13] These are not necessary assumptions – in fact, the more complex the models become, the easier it is find examples in which the best forecast does not produce the best decision. Rather, using a simple model merely reflects the famous compromise of Richard Levins[14] in choosing generality over precision. More precisely, the decision problem in question can be stated as follows: The fish stock is observed to be in state X_t at time t, and is then subjected to some harvest H_t before recruiting new fish, subject to stochastic environmental noise ξ_t, to bring the stock to $X_t + 1$,

$$X_{t+1} = f(X_t - H_t, \xi_t) \tag{1}$$

Further we imagine that the function f is not known precisely, and so we will rely on an evaluation of forecasting skill across a set of candidate models to determine which one to use to manage the fishery. Again for simplicity, we will restrict ourselves to two simple candidate models f_1 and f_2. Both share the same underlying structure of logistic recruitment:

$$f_i(Y) = Y + r_i Y \left(1 - \frac{Y}{K_i}\right) \tag{2}$$

Model 1 is given by $r_1 = 2$, $K_1 = 16$, $\sigma_1 = 0.05$, Model 2 by $r_2 = 0.5$, $K_2 = 10$, $\sigma = 0.075$ (in non-dimensionalized units). Having both the larger growth rate and the larger carrying capacity, Model 1 is clearly the more optimistic of the two choices.

Mathematical models are, at best, approximations of the underlying processes. Ecological processes are much too complex to ever be modeled exactly down to the last atom. For illustrative purposes, we will thus assume the "true" process to be given by a third model, which is unknown to the decision-maker:

$$f_3(Y) = Y + r_3 Y^4 \left(1 - \frac{Y}{K_3}\right) \tag{3}$$

with $r_3 = 0.002$, $K_3 = 10$ and $\sigma_3 = 0.05$.

Results

Forecast performance

The one-step-ahead prediction performance of each model in a simulation of an un-fished environment show consistently better performance of Model 2 (Fig. 1A). Model 1 predictions appear far too optimistic, with the true value falling well below the 95% confidence intervals. In contrast, all observed values fall easily within the confidence intervals produced by model 2.

Predictive performance of the un-fished population does not give us the full picture, since it reflects predictive accuracy only in the region of the true carrying capacity, while an actively harvested stock will be at a lower size. The model that predicts the equilibrium size may not be the one that best forecasts stock recovery. This comparison does not reflect the influence of any decisions that might be made based on the model forecast. To address these concerns, we consider a second scenario where our fishery is managed according to the optimal harvest predicted by each model in turn. Each year the model produces both a forecast and a decision about the harvest quota. (The mechanics of determining a harvest quota given the model follow standard methods for Markov Decision Process, which depend on step-ahead predictions,[5] see appendix for details.) We then implement that harvest and compare the observed stock size the following year to that which the model has predicted, (Fig. 1B). Again we observe that the observations under model

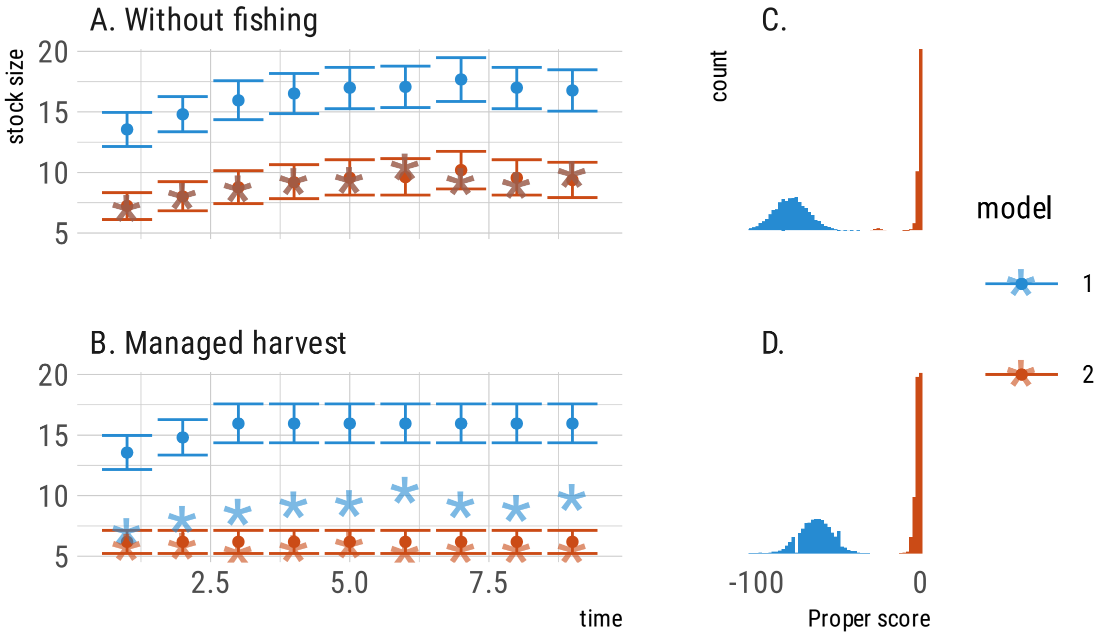

Figure 1: Forecast performance of each model. Panels A, B: Step ahead predictions of stock size under unfished (A) and fished (B) scenarios. Error bars indicating the 95% confidence intervals around each prediction, while stars denote the observed value in that year. Because the models make different decisions each year in the fished scenario, the observed stock size in year 2, 3, etc under the management of model 1 (blue stars) is different from that under model 2 (red stars). Panels C, D: corresponding distribution of proper scores across all predictions (100 replicates of 100 timesteps). Higher scores are better, confirming that model 2 makes the better forecasts.

1 consistently fall well outside of the 95% confidence intervals it predicts, while under model 2, stock sizes consistently fall within the predicted intervals. Once again, model 2 shows a higher forecast accuracy while model 1 appears consistently over-optimistic.

Interest in assessing probabilistic forecasts has led to the development of rigorous methods of scoring.[2,15] A scoring rule is "proper" if it has the convenient property that no other prediction can achieve a higher score, on average, than we would if we used the true distribution.[4] The distribution of proper scores across 100 replicate simulations for both un-fished and actively managed scenarios (Fig. 1C, 1D respectively) show consistently higher scores of model 2, using a proper scoring rule (Eq (27) of Ref. 4).

Ecological and economic performance

Given this evidence, model 2 clearly provides the more accurate forecast and we would no doubt conclude that model 2 was thus a better approximation of the true model and thus the better choice to inform decision making about harvest quotas. Yet if we revisit our experiment of managing the fishery under each model in turn, and focus not on *predictive accuracy* but on *ecological* and *economic* outcomes, it quickly becomes clear that model 1 gives much better results (Fig. 2A). For comparison, we have also included the results of optimal management given the true model. Despite its optimistic predictions, model 1 does not result in over-fishing, but holds the stock near the same level as the optimal management strategy. In contrast, model 2 suppresses the stock to a much lower level. The over-fishing in model 2 is not economically efficient either (Fig. 2B). The net present value of the fishery, as calculated as the cumulative, discounted value of the harvest (assuming a fixed unit price for fish with negligible cost for harvest, see appendix) under the fishing regime of model 1 falls precisely along that of the optimal solution, while the value derived under model 2 is consistently lower.

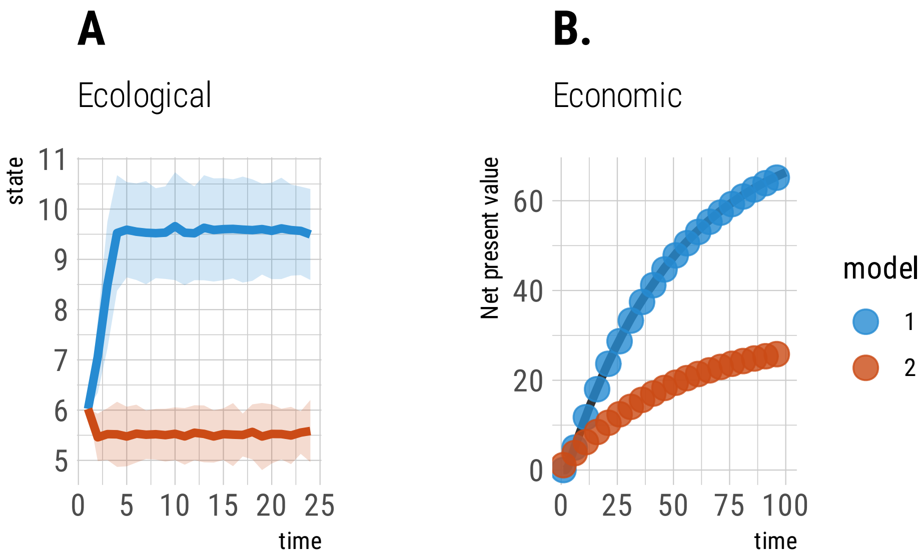

Figure 2: Ecological and economic performance of each forecast. Harvest quotas derived from model 1 result in a significantly higher stock size than under model 2 (panel A). Despite this less agressive harvest policy, economic returns under model 1 are substantially higher (panel B).

Discussion

This paradox in performance of forecasting vs performance in decision making can be easily resolved by considering the context of the decision problem more closely. Comparing plots of the functional form of our two logistic-curve models, compared to the functional form of the "true" model used to drive the simulations (Fig. 3A), it is clear to see that model 2 does indeed lie closer to the true model throughout the state space, agreeing precisely with the true carrying capacity (where both functions cross zero with negative slope). However, the peak of model 3 very nearly matches the peak of model 1. The optimal decision literature, dating back to the 1950s,[6] demonstrates that the Maximum Sustainable Yield (MSY) is maintained by harvesting a stock down to the size at which it achieves its maximum growth rate, i.e. 50% of the un-fished equilibrium size for a symmetric growth model ($K/2$). Model 1, while being very wrong about both the growth rate and the un-fished equilibrium, is nevertheless nearly perfect in estimating the stock size at which maximum growth rate is achieved, and this gives nearly optimal decisions (Fig 3B) despite its terrible forecasts.

Thus, each year our model 1 managers are again chagrined to see the stock size estimates come in far below their rosy predictions, but nevertheless manage to set a nearly optimal quota by comparing the observed stock size to the model's predicted optimal escapement level.[8] Meanwhile, model 2 managers could only congratulate themselves that each year's observations fall neatly within their predicted interval, unaware that the they were over-exploiting the fishery by both economic and ecological metrics. If we had access to model 3, we would no doubt find that it outperformed model 2 in forecast accuracy as well as ecological and economic performance. But in real ecological decision making, we never know the true model – we will always be comparing among approximations. Within fisheries, even in today's parameter-rich age-structured models, recruitment approximations with symmetric growth functions (Logistic, Ricker, Beverton-Holt, etc) still dominate.[16,17]

This issue is by no means unique to fisheries. Throughout resource management and conservation, and no doubt other fields, decisions about which model to use are guided by which model best fits available data.[13] Increasingly, these are joined by calls to assess *forecast accuracy*[1,3,18] as the ultimate test of a model. Yet as this example illustrates, such metrics, no matter how rigorously defined, may select entirely the wrong model for the task at hand. A decision maker has other objectives than prediction accuracy, and approaches which ignore these considerations do so at their peril. This example has also shown that once

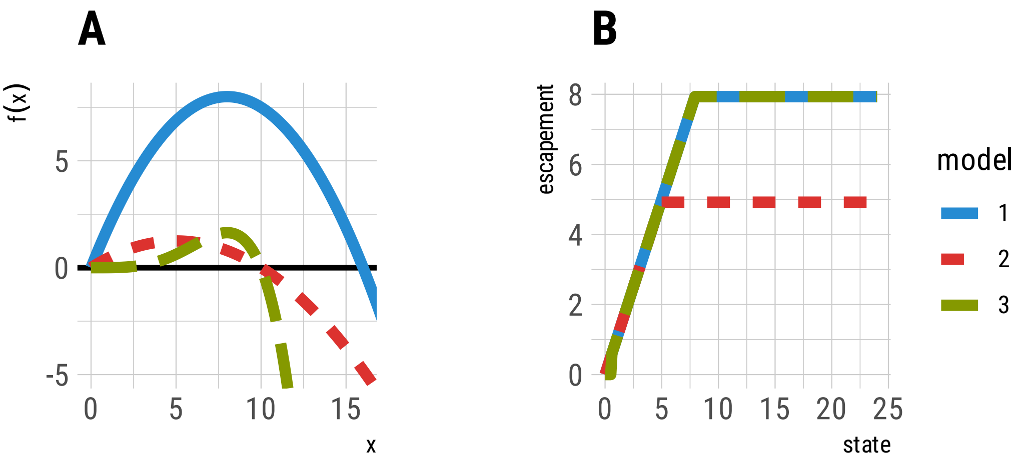

Figure 3: Panel A: Population recruitment curves for each model, compared to that of the true model. Panel B: The computed optimal policy of each model, derived by SDP, expressed in terms of the target escapement (population size remaining after harvest) for each possible stock size. Model 1 overharvests consistently, while the target escapement under model 2 is nearly identical to that of model 3.

we are managing with the wrong model, no amount of comparing predictions from that model to actual outcomes will guarantee we discover our mistake. Despite its consistently good predictions, model 1 is in fact over-fishing to a dangerous level.

Because we will never know the "true" model, we must never forget that our choice of models must reflect the context for which those models will be used.[14] Model 2 would indeed be a better choice than model 1 if our objective was to determine the natural size of our fish stock in the absence of fishing. Only when we focus on the outcomes we actually care about – in this case, economic and ecological performance – can we see which model is best for decision-making. Model 1, despite its many mistakes, is right about one key feature: the biomass for peak growth – and that is enough to guarantee nearly optimal performance. This conclusion should also be reassuring to both modelers and decision makers, for it reminds us that effective models need be perfect or even all that close in every aspect, as long as they capture the key features of the decision context. Decision theory[7,8,19] and research into the socio-ecological models[20] helps us better understand that context. Adaptive management approaches[21] can apply that theory to compare management outcomes between models directly. It is not true that we need good forecasts to make good decisions.

Methods

Stochastic transition matrices are defined for models 1-3 on a discrete grid of 240 possible states spaced uniformly from 0 to 24. A discrete action space enumerating possible harvest quotas is set to the same grid. The utility of a harvest quota H_t given a population state X_t is given by $U(X_t, H_t) = \min(X_t, H_t)$ (i.e. a fixed price for realized harvest).A modest discount of $\gamma = 0.99$ allows comparisons to approaches that ignore[6] or include[7,8] discounting; results are not sensitive to this choice. The optimal policy for each model is determined by stochastic dynamic programming.[5] Details of the implementation, including fully reproducible R code, have been included in the appendix.

Acknowledgements

This work was supported in part by NSF CAREER (#1942280) and computational resources from NSF's XSEDE Jetstream (DEB160003) and Chameleon cloud platforms.

References

1. Clark, J. S. *et al.* Ecological Forecasts: An Emerging Imperative. *Science* **293**, 657–660 (2001).
2. Gneiting, T. & Katzfuss, M. Probabilistic Forecasting. *Annual Review of Statistics and Its Application* **1**, 125–151 (2014).
3. Dietze, M. C. *et al.* Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proceedings of the National Academy of Sciences* **115**, 1424–1432 (2018).
4. Gneiting, T. & Raftery, A. E. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association* **102**, 359–378 (2007).
5. Marescot, L. *et al.* Complex decisions made simple: A primer on stochastic dynamic programming. *Methods in Ecology and Evolution* **4**, 872–884 (2013).
6. Schaefer, M. B. Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. *Bulletin of the Inter-American Tropical Tuna Commission* **1**, 27–56 (1954).
7. Clark, C. W. Profit maximization and the extinction of animal species. *Journal of Political Economy* **81**, 950–961 (1973).
8. Reed, W. J. Optimal escapement levels in stochastic and deterministic harvesting models. *Journal of Environmental Economics and Management* **6**, 350–363 (1979).
9. Worm, B. *et al.* Impacts of biodiversity loss on ocean ecosystem services. *Science (New York, N.Y.)* **314**, 787–90 (2006).
10. Worm, B. *et al.* Rebuilding global fisheries. *Science (New York, N.Y.)* **325**, 578–85 (2009).
11. Costello, C. *et al.* Global fishery prospects under contrasting management regimes. *Proceedings of the National Academy of Sciences* **113**, 5125–5129 (2016).
12. Mangel, M. Decision and control in uncertain resource systems. 255 (1985).
13. Clark, C. W. *Mathematical Bioeconomics: The Optimal Management of Renewable Resources, 2nd Edition.* 400 (Wiley-Interscience, 1990).
14. Levins, R. The strategy of model building in population biology. *American Scientist* **54**, 421–431 (1966).
15. Raftery, A. E. Use and communication of probabilistic forecasts: Use and Communication of Probabilistic Forecasts. *Statistical Analysis and Data Mining: The ASA Data Science Journal* **9**, 397–410 (2016).
16. Ricard, D., Minto, C., Jensen, O. P. & Baum, J. K. Examining the knowledge base and status of commercially exploited marine species with the RAM Legacy Stock Assessment Database: The RAM Legacy Stock Assessment Database. *Fish and Fisheries* **13**, 380–398 (2012).
17. Database, R. L. S. A. RAM Legacy Stock Assessment Database v4.44. (2018) doi:10.5281/ZENODO.2542919.
18. White, E. P. *et al.* Developing an automated iterative near-term forecasting system for an ecological study. *Methods in Ecology and Evolution* **10**, 332–344 (2019).
19. Memarzadeh, M. & Boettiger, C. Resolving the measurement uncertainty paradox in ecological management. *The American Naturalist* **193**, 645–660 (2019).
20. Tallis, H. & Kareiva, P. Shaping global environmental decisions using socio-ecological models. *Trends in Ecology & Evolution* **21**, 562–568 (2006).
21. Walters, C. J. & Hilborn, R. Ecological Optimization and Adaptive Management. *Annual Review of Ecology and Systematics* **9**, 157–188 (1978).