

Iterative forecasting and the trap of self-fulfilling prophecies

Carl Boettiger^{a,1}

^aDepartment of Environmental Science, Policy, and Management, University of California, 130 Mulford Hall Berkeley, CA 94720-3114, USA

Abstract

Encouraged by decision makers' appetite for future information on topics ranging from elections to pandemics, and enabled by the explosion of data and computational methods, model based forecasts have garnered increasing influence on a breadth of decisions in modern society. Using a classic example from fisheries management, I demonstrate that selecting the model or models that produce the most accurate and precise forecast can lead to decidedly worse outcomes. This creates a cycle of self-fulfilling prophecy, in which the manager not only over-exploits the fishery leading to declining biomass and declining yields, but also becomes increasingly convinced that these actions are not over-exploitation but in fact consistent with the best models and data for sustainable management.

Keywords: forecasting, adaptive management, stochasticity, uncertainty, optimal control

- abbreviated running title: "The trap of self-fulfilling prophecies"
- Authorship statement: *This is a single-author paper*
- data accessibility statement: *All simulation data generated for analyses here, along with code required for the analysis, is available in the Zenodo Data archive, <https://doi.org/10.5281/zenodo.4660621>*
- number of words in the abstract: 116
- number of words in main text: 4998
- number of cited references: 33
- number of tables & figures: 4 Figures

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

Global change issues are complex and outcomes are difficult to predict (Clark *et al.* 2001). To guide decisions in an uncertain world, researchers and decision makers may consider a range of alternative plausible models to better reflect what we do and do not know about the processes involved (Polasky *et al.* 2011). Forecasts or predictions from possible models can indicate what outcomes are most likely to result under what decisions or actions. This has made model-based forecasts a cornerstone for scientifically based decision making. By comparing outcomes predicted by a model to future observations, a decision maker can not only *plan* for the uncertainty, but also *learn* which models are most trustworthy. The value of iterative learning has long been reflected in the theory of adaptive management (Walters & Hilborn 1978) as well as in actual adaptive management practices such as Management Strategy Evaluation (MSE) (Punt *et al.* 2016) used in fisheries, and is a central tenet of a rapidly growing interest in ecological forecasting (Dietze *et al.* 2018). But, do iterative learning approaches always lead to better decisions?

In this paper, I demonstrate that the model that makes the best prediction is not necessarily the model that makes the best policy. I show that iterative learning or iterative forecasting can be counter-productive to management outcomes: that in some cases, a manager would be better to ignore new information which could reduce uncertainty over available models. Put another way, the value of information (VOI, as measured by the expected utility given that information minus the utility without it; see Howard (1966); Katz *et al.* (1987)), can actually be negative. This is not a consequence of imperfect information (e.g. observation error) or the result of model over-fitting. The issue is an intrinsic consequence of the fact that many models can lead to the same decisions, and no generic solution to the problem exists. Worse, whenever this situation arises, the decision-maker becomes trapped by a form of self-fulfilling prophecy: the model(s) which the decision-makers have learned to favor lead to decisions and outcomes that closely match model expectations. Consequently, the decision-maker can become ever more firmly entrenched in the belief that their new models are accurate and the mediocre outcomes they experience are as good as it gets, while confidently rejecting earlier models which would have led to much better outcomes. As this example will show, by better understanding what aspects of a model are most and least important to shaping a particular decision, we can at least anticipate the circumstances that make this problem more likely.

Iterative learning approaches such as adaptive management or iterative forecasting (Dietze *et al.* 2018) are particularly compelling because they address a different reason why seemingly-accurate models can lead to poor outcomes: the problem of overfitting. When models are estimated from the data they seek to describe, the model may overfit the data, making the data seem even more likely than it would be under the true process (Burnham & Anderson 1998). Over-fit models may underestimate uncertainty and make poor predictions about future outcomes (Ginzburg & Jensen 2004). Because any iterative learning compares models to future outcomes, over-fitting those outcomes is impossible. This concept was rigorously formalized by Gneiting & Raftery (2007)’s proof of “proper” scoring rules. By definition, under a proper scoring rule, no probabilistic prediction $Q(x)$ can score better, on average, than that of the underlying process, $P(x)$. Gneiting & Raftery (2007) further shows which techniques for evaluating forecast predictions do and do not satisfy this condition, i.e. which cannot be over-fit. Adaptive management is based on equally secure footing because model predictions are again compared to future outcomes not used to fit the model. Instead of scoring and selecting among alternative models, adaptive management considers probabilities over all models (Walters & Hilborn 1978). These two approaches are particularly compelling because they can both reduce uncertainty and avoid over-fitting. My example is not meant to undercut the importance of these approaches. Previous work has long acknowledged the panoply of ways in which iterative learning or other model-based decision making can go astray due to conflicting incentives, implementation errors, or lack of resources for monitoring and updating (e.g. Ludwig *et al.* 1993). Here, the problem is more fundamental.

I illustrate this problem using an example from fisheries management. Fisheries are a significant economic and conservation concern worldwide and their management remains an important debate (e.g. Worm *et al.* 2006, 2009; Costello *et al.* 2016). Moreover, their management has been a proving grounds for theoretical and practical decision-making issues which are widely applicable in other areas of ecology and conservation (Ludwig *et al.* 1993; Lande *et al.* 1994), and one that has long wrestled with issues of uncertainty in the context of management decisions (e.g. Clark 1973; Reed 1979; Walters 1981; Ludwig & Walters 1982). While modern fisheries management frequently relies on complex models which may contain over 100 parameters to reflect the specific age or stage structure of a specific fish stock (Ricard *et al.* 2011; Ricard *et al.*

2018), fisheries management and research has always appreciated the central insights offered by simpler one-dimensional models which continue to underpin both the theory and practice. For example, the strategies of constant mortality (Schaefer 1954) and constant escapement (Clark 1973; Reed 1979) are built on the same class of models considered here, which also continue to underpin global analyses of stock-rebuilding (e.g. Costello *et al.* 2016; Memarzadeh *et al.* 2019) and the foundations of adaptive management (e.g. Walters & Hilborn 1978; Smith & Walters 1981; Walters 1981). While more complex models are often required for specific applications, such models both share features with and are guided by theory built on much more general, simpler models (Levins 1966), from the Maximum Sustainable Yield of Schaefer (1954) or the epidemiological reproductive number, R_0 of Kermack & McKendrick (1927).

Ecological Models

The sustainable harvest decision problem 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 $X_t + 1$,

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

A manager seeks each year to select the harvest quota H_t which will maximize the sum of the utility derived from such a harvest and such a state, $U(X_t, H_t)$, over all time, subject to discount rate δ (Clark 1973):

$$\sum_{t=0}^{t=\infty} U(X_t, H_t) \delta^t \quad (2)$$

For simplicity and comparison with prior work, we will assume a fixed price of fish p with no additional cost on additional harvest, $U(X_t, H_t) = p \min(H_t, X_t)$ (noting that realized harvest cannot exceed the stock size). Without loss of generality we will set the price $p = 1$ and modest discount $\delta = 0.99$.

We further imagine that the function f is not known precisely. We restrict ourselves to two simple candidate models f_1 and f_2 . Both share the same underlying structure of logistic recruitment (known as the Gordon-Schaefer model in fisheries context owing to groundbreaking work independently by Schaefer (1954) and Gordon & Press (1954)), differing only in their choice of certain parameters:

$$f_i(Y) = Y + r_i Y \left(1 - \frac{Y}{K_i} \right) * \xi_i(\sigma) \quad (3)$$

Let us assume $\xi_i(\sigma)$ represents log-normal random noise with a mean of unity and log-standard-deviation σ_i for each model. 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 dimensionless units). Having both the larger growth rate and the larger carrying capacity, model 1 is clearly the more optimistic of the two choices.

Selecting between model 1 and model 2 can thus be considered the simplest illustration of the model uncertainty problem. This is a subset of the more general problem of selecting model parameters, assuming a logistic growth, which itself is a subset of estimating the best structural form (e.g. Ricker, Beverton-Holt, etc). There is no need to consider these more complicated versions of the model uncertainty problem here, since they all inherit the same issue. Reducing the model selection problem to these two models simplifies the presentation and will aid intuition at no loss of generality.

The only additional assumption we will need is that the “true” model is not among the suite of models under consideration. Mathematical models are, at best, approximations of the underlying processes. Ecological processes are much too complex to ever be modeled exactly. For illustrative purposes, we will thus assume the “true” process to be given by Model 3, which is unknown to the decision-maker, but similar enough to at least one of the candidate models might be considered a reasonable approximation:

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

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

The task of deciding whether Model 1 or model 2 would be the better choice for decision making is thus perhaps the simplest example of the much studied issue of model uncertainty that we can pose. As in any real world scenario, neither model is the true model, but nevertheless this model set contains a good enough approximation of the true model to make good decisions. However, any of the well-developed approaches for decision-making under model uncertainty will prefer model 2 over model 1, despite the fact that the optimal policy under model 2 leads to much worse outcomes ecologically and economically.

Methods for Managing Under Model Uncertainty

I will use this example to illustrate two alternative approaches for iterative learning over model uncertainty: iterative forecasting and adaptive management. The central difference in the approaches is that iterative forecasting is premised on the ability to score the predictions of alternative models. Iterative forecasting is silent on the issue of what to do with those scores, this is left up to the decision-maker. Adaptive management approaches, by contrast, explicitly seek to integrate probabilities over all candidate models to reach a decision. I consider each in turn.

Statistical approaches: Forecasting under “Proper” Scoring Rules

Like many decision problems, the task of setting a sustainable harvest quota appears to hinge on having an accurate forecast: if we can predict to what size the fish stock will increase next year, $X_t + 1$, and we know the current stock, X_t , then we can sustainably harvest $X_{t+1} - X_t$ without decreasing the biomass over the long term. Selecting a model based on forecast skill is also justifiable on theoretical grounds, since it reduces the risk of overfitting by comparing model predictions to later observations that were not used to estimate the model (Gneiting & Katzfuss 2014).

I illustrate the process of model selection by strictly proper scoring rules using two scenarios. In Scenario A (passive observation) the fish stock is unharvested and allowed to recover towards carrying capacity (as simulated under our “true” model, Model 3) while comparing the observed stock size in each subsequent time step to the distribution predicted under model 1 and model 2 respectively [Fig 1]. The mean, μ_t and variance, σ_t of the forecast are compared against the true observation x_t using a proper scoring rule given by Gneiting & Raftery (2007),

$$S(x_t|\mu_t, \sigma_t) = -(\mu_t - x_t)^2/\sigma_t^2 - \log(\sigma_t) \quad (5)$$

for each prediction over 100 replicate simulations of 100 time steps each [Fig 1].

[Figure 1 about here.]

In Scenario B (actively harvest), I have first solved for the optimal management strategy for both model 1 and model 2 [Fig 1b]. For small noise and concave functions with linear reward structure this can be done analytically (see proof in Reed 1979), or solved more generally by stochastic dynamic programming (see review by Marescot *et al.* 2013, details in Appendix). Under this scenario, replicate simulations of the stock are harvested at each time step using the optimal quota dictated by either model 1 and model 2. The resulting stock sizes in the time-step following this harvest are once again compared to the probabilities predicted by each model using Eq (5). Model 2 unequivocally outperforms model 1 in both scenarios of passive observation and active harvest.

Despite the clearly superior predictive accuracy of model 2 in both scenarios, the outcomes from management under model 2 are substantially worse. We can assess such outcomes in less abstract terms than forecasting skill, such as economic value (in dollars) or the ecological value (unharvested biomass). In our simple formulation of the decision problem, the “utility” the manager seeks to maximize is simply the economic value (net present value: the discounted sum of all profits from future harvests, Eq (2)) of harvested fish. This formulation ignores any utility provided by fish that are not harvested, beyond their contribution to future potential harvests. While it is possible to include such contributions directly in the utility function being optimized (e.g. Halpern *et al.* 2013), even without doing so, model 1 maintains both a higher unharvested biomass and also leads to higher economic returns throughout [Fig 2].

[Figure 2 about here.]

In both scenarios, the careful comparison through proper scoring rules has led us to select the worse-performing model. Crucially, a manager operating under this selection would have little indication that their model was flawed: both future stock sizes and expected harvest yields consistently match model predictions. Had we been able to include Model 3 in our forecast comparisons, it would equal or outperform the forecasting skill of both model 1 and model 2 (as guaranteed by the theorem of Gneiting & Raftery (2007)), while also matching or out-performing their economic utility (as guaranteed by the theorem of Reed (1979)). In practice, we never have access to the generating model, so it is reasonable to expect model selection to determine the best approximation. As we see here, the best approximation for forecasting future states does not in fact lead to better outcomes.

One obvious limitation in this comparison is that scenario B treats each model as fixed over the entire course of the simulation. It would be possible to generate forecasts for the next n steps, but these forecasts would also be conditional on the management action (i.e. harvest quota) selected at each step. To evaluate two-step-ahead predictions we must consider each possible action under each possible state predicted by the step-ahead forecast, weighted by probability of the state given the model and the probability of the model. This rapidly expanding set of possibilities is addressed by adaptive management for sequential decision problems (e.g. Smith & Walters 1981), which I employ in the next section.

Decision-Theoretic Approaches

Any adaptive management strategy updates posterior distributions over model uncertainty (Ludwig & Walters 1982; Punt *et al.* 2016). Unfortunately, any such adaptive updating leads to worse outcomes than the equivalent non-adaptive strategy, in which model uncertainty is held fixed. I illustrate the application of a passive adaptive management strategy to this simple example, following classic examples for parameter (Ludwig & Walters 1982) or structural (Smith & Walters 1981) model uncertainty. Passive adaptive management for a simple sequential decision problem is straightforward to implement over a discrete set of states and actions using dynamic programming with iterative updates (Smith & Walters 1981, example code in Appendix). To demonstrate that the behavior is not driven by failure to explore sufficiently, (which might be addressed by an active adaptive management), I will assign initial probability that model 2 is true at 1%. After a single iteration of learning, these probabilities are completely reversed, with the manager deciding that model 2 is almost certainly correct [Fig 3A]. I compare these results to As before, this results in a management practice with much worse ecological and economic outcomes than would have been realized by a manager who stubbornly clung to model 1 without updating, which achieves a net present value that only 31% that expected under management using model 1 alone [Fig 2].

[Figure 3 about here.]

Increasing the space of possible models to cover a whole plausible range of parameters r and K does little to resolve this problem [Fig 3B]. Iterative updates again quickly dismiss the parameter values assumed by model 1, though with more options to choose from, this probability is spread over a range of seemingly plausible candidate models instead of a single alternative model (see Appendix). While the adaptive management of additional actions and observations slowly narrow this subset of plausible models, decisions based on this uncertainty prevent fish stocks from recovering fully, and realize lower harvests as a result. Note that learning under either adaptive management approach (using two models or 42), the decision-maker becomes ever-increasingly convinced that they are using the right model or models. Future stock sizes fall consistently in the range predicted by the model(s), and consistently outside the range predicted by model 1. Consequently, each iteration the managers are only more firmly convinced that they are maintaining the fish stock near the biomass that supports maximum sustainable yield, when in fact they are sustaining a harvest regime that is preventing recovery of the stock to the much higher productivity regime which would have been achieved under model 1.

Discussion

Given this simple decision problem in which one of the two models leads to effectively optimal ecological and economic outcomes, current approaches invariably choose the other. Moreover, a decision maker employing an adaptive management or iterative forecasting assessment such as those considered here would have no way of realizing that the outcomes they experienced were in fact sub-optimal. In both approaches, the manager quickly concludes that model 1 is entirely implausible, while finding that model 2 is remarkably accurate at predicting future values. This problem is not addressed by re-estimating parameters (equivalently, considering a larger suite of models spanning a wide range of possible K and r values): which improves only marginally over the outcome of considering only two models [Fig 3B]. Both iterative forecast evaluation and the adaptive management approaches to reducing model uncertainty lead to management becoming trapped in a self-fulfilling prophecy: holding the fish population at an unproductively low biomass suggested by model 2 while becoming ever more convinced that model is suitable. Without access to the true model, none of our methods for model choice or reducing model uncertainty can break this trap. This should not be mistaken as a critique of adaptive management or iterative forecasting specifically – other model choice approaches such as goodness-of-fit, information criteria or cross-validation would all prefer model 2 as well.

The reason for model 1's seemingly contradictory ability to make good decisions but bad forecasts becomes obvious once we compare both curves to that of the underlying model, model 3. Plotting the growth rate functions of each model, [Fig 4A], it is hardly surprising that no method exists which would not prefer the closely overlapping model 2 to the no-where-close model 1 as the better approximation of model 3. Nevertheless, decisions based on model 1 are nearly indistinguishable from those based on the true model [Fig 4B], while model 2 leads to over-harvesting. The explanation comes from noticing that the stock size corresponding to the maximum growth rate under model 1 (the peak of the curve) falls at almost exactly the same stock size as that of the peak growth rate for model 3. Meanwhile, the peak of model 2 occurs at a substantially lower stock size. While the optimal control solution appears to depend only on a step-ahead forecast accuracy (indeed, the SDP solution method used here takes only step-ahead forecast probabilities as input, (Marescot *et al.* 2013)), mathematical analysis showed long ago (Reed 1979) that the optimal solution for this problem depends only on keeping the biomass at the value responsible for the maximum growth rate. This realization is quite general: for most decision problems, simple models will exist under which the optimal decision is the same as it would be for the true model, even when that simple model is wrong in most other ways.

Situations similar to this example, in which a model can lead to the optimal decision while producing a poor forecast, are likely to be widespread. For example, marine fisheries typically restrict a Management Strategy Evaluation [MSE; Punt *et al.* (2016)] to constant mortality, (defined as harvest per unit biomass, $F = H/B$). Constant mortality is technically only optimal in a deterministic model (Clark 1973); but is easier to solve for than the SDP methods for constant escapement shown here, which in practice is used mostly in salmon fisheries and the resource economics literature Costello *et al.* (2016). While Model 1 used here will not lead to a good constant-mortality policy, it is just as easy to construct a model which gives optimal decisions and bad forecasts. In the example shown above, Model 1 achieves optimal results by choosing K such that $K/2$ matches the peak growth rate under the true model [Reed (1979)]. Under constant mortality, we instead choose a logistic model with r such that $r/4$ matches the optimal constant mortality under the true model (Schaefer 1954). Everything else about our choice of Model 1 can be far from the true model, and yet it will give optimal results, just as above. Examples in which decisions depend a subset or ratio of model parameters can be found everywhere. In Dee *et al.* (2017), the optimal number of non-critical species to protect in order to conserve critical species is given by a ratio of model parameters. In the SIR model of Kermack & McKendrick (1927) the basic reproductive number that determines if an infection spreads is the ratio of contact rate over the recovery rate, $R_0 = \beta/\gamma$. In general, there will be many models that can lead to the same decision, and thus the possibility of a similar scenario to the one seen here. Likewise there will always be infinitely many models like Model 2, which can fit very close to the data [Fig 4] while leading to a very different decision policy.

[Figure 4 about here.]

Note that the mechanism shown here has nothing to do with the much more familiar issue of over-fitting, in which a better-fitting model will also lead to worse outcomes (Burnham & Anderson 1998). In fact, a model which has been over-fit will gradually be rejected by either the iterative forecasting or adaptive management approaches shown here, as these approaches continually confront the models with new data to which they had not been previously fit. The ability to avoid overfitting is one of the greatest appeals of any iterative management strategy. In the example here, both model 1 and model 2 have the same structural complexity. The better decision performance of model 1 is a consequence of capturing the key attribute needed for a good decision, which is perhaps surprisingly an all-together different criteria than fit. Note that other methods of model selection not considered here, including goodness-of-fit metrics such as r^2 or information criteria (Burnham & Anderson 1998), will all prefer Model 2 over Model 1 for the same reason.

This simple example can be resolved by including a model that closely approximates both the location of the peak and the shape of the curve of our “true” model. Given sufficient data, either adaptive management or iterative forecasting will prefer such a model (as guaranteed by Gneiting & Raftery (2007)’s theorem), while having the correct peak location means the model will also lead to the optimal policy in this case (as guaranteed by Reed (1979)’s theorem). Yet the larger problem remains unresolved. In most realistic management scenarios, we do not have the benefit of a theorem similar to Reed’s to identify what features are most critical to approximate well and what are not. For instance, the theorem no longer holds if stock estimates are uncertain (Memarzadeh & Boettiger 2019). Absent the guidance of such results, iterative learning over even relatively rich possible candidate models can move us towards worse outcomes than not learning at all.

That potential for long-term deterioration in management outcomes due to adaptive learning has not been widely acknowledged before. Certainly the limitations of any modeling approach to improve management outcomes due to the human context of competing interests, imperfect enforcement, etc., is readily appreciated (Ludwig *et al.* 1993). The problem here goes deeper, to the limits of models and learning from data even in a toy example free from other real world challenges. Previous work has acknowledged that iterative learning may not always be feasible and that it may not always be beneficial (Polasky *et al.* 2011), but has failed to recognize the possibility that it can potentially be detrimental. For example, active adaptive management is premised on the observation that future observations may be too uninformative to distinguish between alternatives (Walters & Hilborn 1978). Others have also noted that even big improvements in predictive accuracy could have negligible improvement on the decision outcomes, which motivates quantifying the Value of Information (VOI) (Katz *et al.* 1987). But in both cases, the worst outcome is wasted effort. In the scenario considered here, the VOI is negative: any learning over the proposed model uncertainty leads to lower expected net utility than not learning.

Worse, absent the counter-factual of non-adaptive management to compare to, this deterioration in outcomes is invisible. All the information available to our manager reinforces the conviction that this deteriorated ecological state is in fact the natural, sustainably managed equilibrium of our fish population. By responding in any way to future observations to reduce model uncertainty, our manager is caught in a trap of self-fulfilling prophecy: believing ever more firmly in models that make more accurate predictions while driving decisions that prevent the fish stock from recovering. The only way to avoid this trap is to *forego learning* with each new observation, to avoid iterative updates. It may seem that the adaptive management approach fails because posterior probabilities are updated according to Bayes rule. Like iterative forecasting, this favors models which better predict the data, despite the fact that under the adaptive management approach, the overall optimization is conditioned on actual utility and not model fit. However, this issue is not easily avoided. For example, so-called greedy optimization techniques that favor actions with higher immediate reward do even worse in this context, since such an algorithm would obviously maximize the immediate harvest and therefore collapse the stock. Only by comparing the net utility derived from management under model 1 for many iterations to the net utility derived under model 2 after many iterations can we determine that model 1 leads to better outcomes. This raises several questions. Do such methods of learning and reducing model uncertainty ever lead us towards worse long-term outcomes in the real world? If so, How would we know, and what should we do about it?

Do models which give nearly optimal performance while at the same time making wildly wrong predictions really exist? It is tempting to argue that model 1 is a kind of unicorn, a mythical creature existing only in

theory. But that is a difficult assertion to prove. Examples of serious forecasts that widely fail to predict future observations abound wherever forecasts are common, from elections to economics to environmental change (e.g. Tetlock & Gardner 2015). If there is no shortage of bad forecasts, then could any of them be useful? The premise that a model need not perfectly capture reality to be useful is at the very heart of modeling. It is also important to note that in many cases, any models being used in decision-making are not necessarily being subjected to the rigorous evaluation and updating steps proposed by adaptive management (Walters & Hilborn 1978) or iterative forecasting (Dietze *et al.* 2018). This makes it more likely that such models could persist in practice until now. While iterations that revise these models will no doubt improve decision outcomes in many cases, it is worth bearing in mind from the example here that such improvement is not guaranteed.

If these unicorns do exist in real world management, then what do we do about them? I believe this is an open question, but that our first step must be to recognize it as such. We have seen that the problem cannot be resolved by more data alone, and is not the result of over-fitting. Introducing more models into our candidate set is only helpful to the extent that such models can capture features important to the decision, which as we have seen can be quite different from the feature which matter most to prediction. We have also seen how not updating our uncertainty estimates, or doing so less frequently, can reveal a unicorn model before it is discounted by further iterations. Approaches such as iterative forecasting or adaptive management that can reduce model uncertainty over time remain promising and important techniques, and essential strategies to avoid the very real problem of model over-fitting. But their very success in selecting models which give the most accurate predictions can also be trap. How that is best done in general is an open question.

Acknowledgements

The author acknowledges support from NSF CAREER Award #1942280 and helpful discussions with Melissa Chapman, Jeremy Fox, and anonymous reviewers.

References

- Burnham, K.P. & Anderson, D.R. (1998). Practical Use of the Information-Theoretic Approach. In: *Model Selection and Inference*. Springer New York, New York, NY, pp. 75–117.
- Clark, C.W. (1973). Profit maximization and the extinction of animal species. *Journal of Political Economy*, 81, 950–961.
- Clark, J.S., Carpenter, S.R., Barber, M., Collins, S., Dobson, A., Foley, J.A., *et al.* (2001). Ecological Forecasts: An Emerging Imperative. *Science*, 293, 657–660.
- Costello, C., Ovando, D., Clavelle, T., Strauss, C.K., Hilborn, R., Melnychuk, M.C., *et al.* (2016). Global fishery prospects under contrasting management regimes. *Proceedings of the National Academy of Sciences*, 113, 5125–5129.
- Dee, L.E., De Lara, M., Costello, C. & Gaines, S.D. (2017). To what extent can ecosystem services motivate protecting biodiversity? *Ecology Letters*, 20, 935–946.
- Dietze, M.C., Fox, A., Beck-Johnson, L.M., Betancourt, J.L., Hooten, M.B., Jarnevich, C.S., *et al.* (2018). Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proceedings of the National Academy of Sciences*, 115, 1424–1432.
- Ginzburg, L.R. & Jensen, C.X.J. (2004). Rules of thumb for judging ecological theories. *Trends in Ecology & Evolution*, 19, 121–126.
- Gneiting, T. & Katzfuss, M. (2014). Probabilistic Forecasting. *Annual Review of Statistics and Its Application*, 1, 125–151.
- Gneiting, T. & Raftery, A.E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102, 359–378.
- Gordon, H.S. & Press, C. (1954). The Economic Theory of a Common-Property Resource: The Fishery. *Journal of Political Economy*, 62, 124–142.
- Halpern, B.S., Klein, C.J., Brown, C.J., Beger, M., Grantham, H.S., Mangubhai, S., *et al.* (2013). Achieving the triple bottom line in the face of inherent trade-offs among social equity, economic return, and conservation. *Proceedings of the National Academy of Sciences*, 110, 6229–34.
- Howard, R. (1966). Information Value Theory. *IEEE Transactions on Systems Science and Cybernetics*, 2, 22–26.
- Katz, R.W., Brown, B.G. & Murphy, A.H. (1987). Decision-analytic assessment of the economic value of weather forecasts: The following/planting problem. *Journal of Forecasting*, 6, 77–89.
- Kermack, W.O. & McKendrick, A.G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115, 700–721.
- Lande, R., Engen, S. & Saether, B.-E. (1994). Optimal harvesting, economic discounting and extinction risk in fluctuating populations. *Nature*, 372, 88–90.
- Levins, R. (1966). The strategy of model building in population biology. *American Scientist*, 54, 421–431.
- Ludwig, D., Hilborn, R. & Walters, C. (1993). Uncertainty, Resource Exploitation, and Conservation: Lessons from History. *Science*, 260, 17–36.
- Ludwig, D. & Walters, C.J. (1982). Optimal harvesting with imprecise parameter estimates. *Ecological Modelling*, 14, 273–292.
- Marescot, L., Chapron, G., Chadès, I., Fackler, P.L., Duchamp, C., Marboutin, E., *et al.* (2013). Complex decisions made simple: A primer on stochastic dynamic programming. *Methods in Ecology and Evolution*, 4, 872–884.
- Memarzadeh, M. & Boettiger, C. (2019). Resolving the measurement uncertainty paradox in ecological management. *The American Naturalist*, 193, 645–660.
- Memarzadeh, M., Britten, G.L., Worm, B. & Boettiger, C. (2019). Rebuilding global fisheries under uncertainty. *Proceedings of the National Academy of Sciences*, 116, 15985–15990.
- Polasky, S., Carpenter, S.R., Folke, C. & Keeler, B. (2011). Decision-making under great uncertainty: environmental management in an era of global change. *Trends in Ecology & Evolution*, 26, 398–404.
- Punt, A.E., Butterworth, D.S., Moor, C.L. de, De Oliveira, J.A.A. & Haddon, M. (2016). Management strategy evaluation: Best practices. *Fish and Fisheries*, 17, 303–334.

374 Reed, W.J. (1979). Optimal escapement levels in stochastic and deterministic harvesting models. *Journal of*
375 *Environmental Economics and Management*, 6, 350–363.

376 Ricard, D., Minto, C., Jensen, O.P. & Baum, J.K. (2011). Examining the knowledge base and status of
377 commercially exploited marine species with the RAM Legacy Stock Assessment Database. *Fish and*
378 *Fisheries*, 13, 380–398.

379 Ricard, D., Minto, C., Jensen, O.P. & Baum, J.K. (2018). RAM Legacy Stock Assessment Database v4.44.

380 Schaefer, M.B. (1954). Some aspects of the dynamics of populations important to the management of the
381 commercial marine fisheries. *Bulletin of the Inter-American Tropical Tuna Commission*, 1, 27–56.

382 Smith, A.D.M. & Walters, C.J. (1981). Adaptive Management of Stock–Recruitment Systems. *Canadian*
383 *Journal of Fisheries and Aquatic Sciences*, 38, 690–703.

384 Tetlock, P.E. & Gardner, D. (2015). *Superforecasting: The art and science of prediction*. First edition.
385 Crown Publishers, New York.

386 Walters, C.J. (1981). Optimum Escapements in the Face of Alternative Recruitment Hypotheses. *Canadian*
387 *Journal of Fisheries and Aquatic Sciences*, 38, 678–689.

388 Walters, C.J. & Hilborn, R. (1978). Ecological Optimization and Adaptive Management. *Annual Review of*
389 *Ecology and Systematics*, 9, 157–188.

390 Worm, B., Barbier, E.B., Beaumont, N., Duffy, J.E., Folke, C., Halpern, B.S., *et al.* (2006). Impacts of
391 biodiversity loss on ocean ecosystem services. *Science (New York, N.Y.)*, 314, 787–90.

392 Worm, B., Hilborn, R., Baum, J.K., Branch, T.A., Collie, J.S., Costello, C., *et al.* (2009). Rebuilding global
393 fisheries. *Science (New York, N.Y.)*, 325, 578–85.

394 List of Figures

395	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.	12
396			
397			
398			
399			
400			
401			
402			
403	2	Ecological and economic performance of each forecast. Harvest quotas derived from model 1 result in a significantly higher fish stock size than under Model 2 (panel A). Economic returns under model 1 are also substantially higher (panel B)	13
404			
405			
406	3	Adaptive management under model uncertainty. Solid lines trace the trajectories of the state (fish stock, circles) and action (harvest quota, triangles), under adaptive management (learning). Dotted lines trace the corresponding trajectories if iterative learning is omitted, leaving the prior belief fixed throughout the simulation (planning). Color indicates the belief that model 1 is correct (blue), with an initial prior belief of 99%. Panel A: Management over the two candidate models, Model 1 and Model 2. Within a single iteration of adaptive management, the belief over models switches from a prior belief that heavily favored model 1 to a posterior that favors model 2 with near certainty. Future iterations reinforce the belief in model 2, resulting in both depressed harvests and low stock sizes (solid lines). If no iterative learning updates are performed, stock sizes and realized harvests (and thus economic profit) are both higher. Panel B: given 42 candidate models over a broad range of parameter values, adaptive management quickly reduces the probability of model 1, and substantially underperforms management without learning (dotted lines). While outcomes improve marginally relative to the two-model case (figure A) they remain significantly worse than had no iterative learning been included.	14
407			
408			
409			
410			
411			
412			
413			
414			
415			
416			
417			
418			
419			
420			
421	4	Panel A: Population growth curves of each model. The positive equilibrium of each model occurs where the curve crosses the horizontal axis. Note that while Model 2 is a better approximation to the truth (Model 3), Model 1 better approximates the stock size which leads to maximum growth. Panel B: The optimal control policy under Model 1 is nearly identical to that under the true Model 3, while the optimal policy under Model 2 suppresses stock to a much lower escapement level.	15
422			
423			
424			
425			
426			

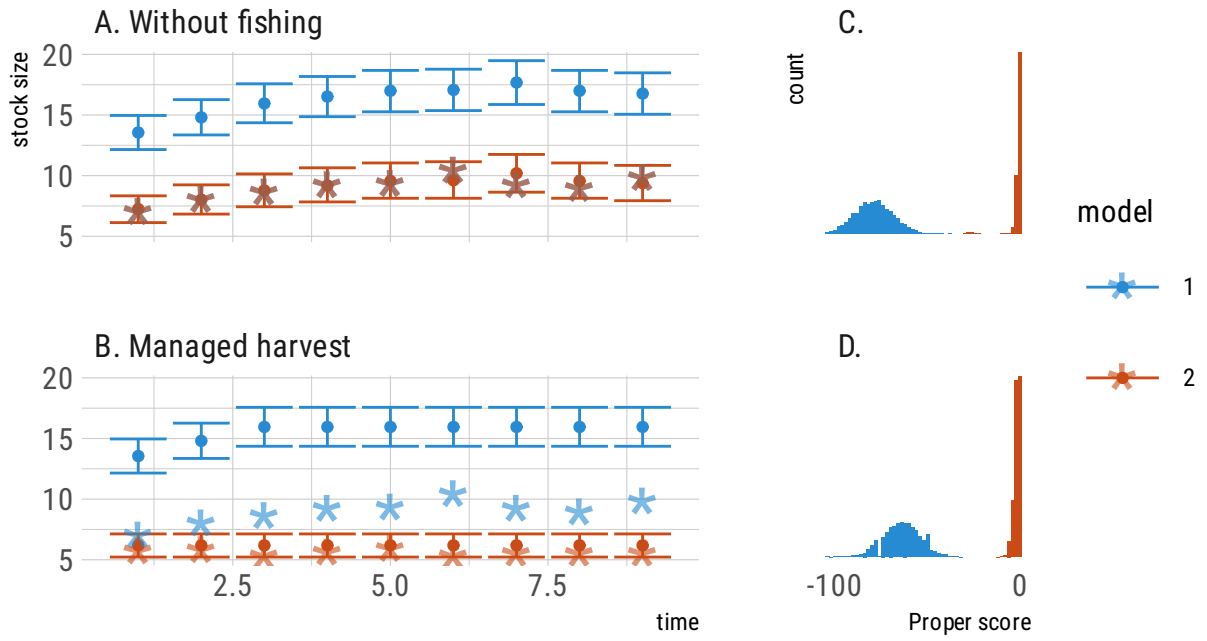


Figure 1: Forecast performance of each model. Panels A, B: Step ahead predictions of stock size under unfished (A) and fished (B) scenarios. Error bars indicating the 95% confidence intervals around each prediction, while stars denote the observed value in that year. Because the models make different 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.

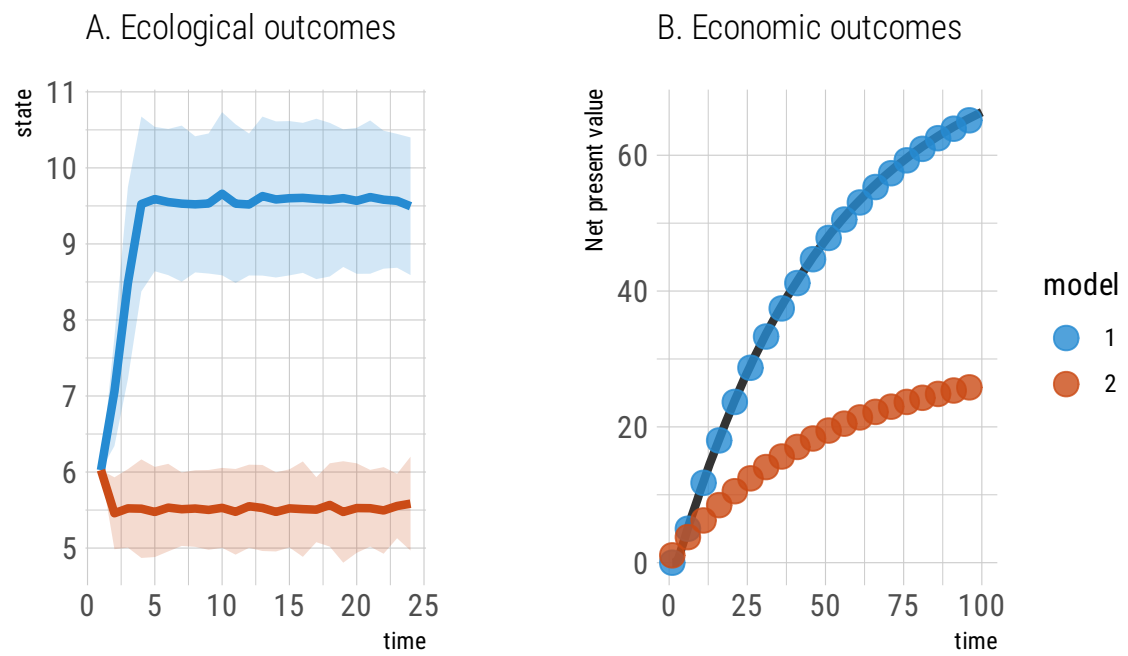


Figure 2: Ecological and economic performance of each forecast. Harvest quotas derived from model 1 result in a significantly higher fish stock size than under Model 2 (panel A). Economic returns under model 1 are also substantially higher (panel B)

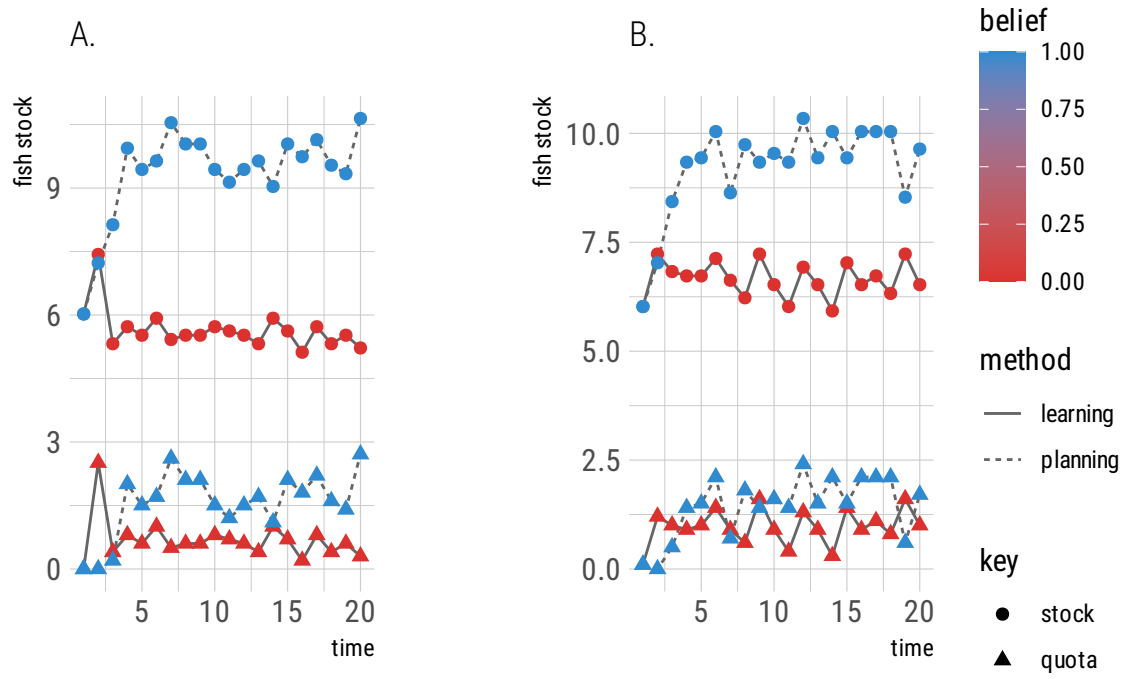


Figure 3: Adaptive management under model uncertainty. Solid lines trace the trajectories of the state (fish stock, circles) and action (harvest quota, triangles), under adaptive management (learning). Dotted lines trace the corresponding trajectories if iterative learning is omitted, leaving the prior belief fixed throughout the simulation (planning). Color indicates the belief that model 1 is correct (blue), with an initial prior belief of 99%. Panel A: Management over the two candidate models, Model 1 and Model 2. Within a single iteration of adaptive management, the belief over models switches from a prior belief that heavily favored model 1 to a posterior that favors model 2 with near certainty. Future iterations reinforce the belief in model 2, resulting in both depressed harvests and low stock sizes (solid lines). If no iterative learning updates are performed, stock sizes and realized harvests (and thus economic profit) are both higher. Panel B: given 42 candidate models over a broad range of parameter values, adaptive management quickly reduces the probability of model 1, and substantially underperforms management without learning (dotted lines). While outcomes improve marginally relative to the two-model case (figure A) they remain significantly worse than had no iterative learning been included.

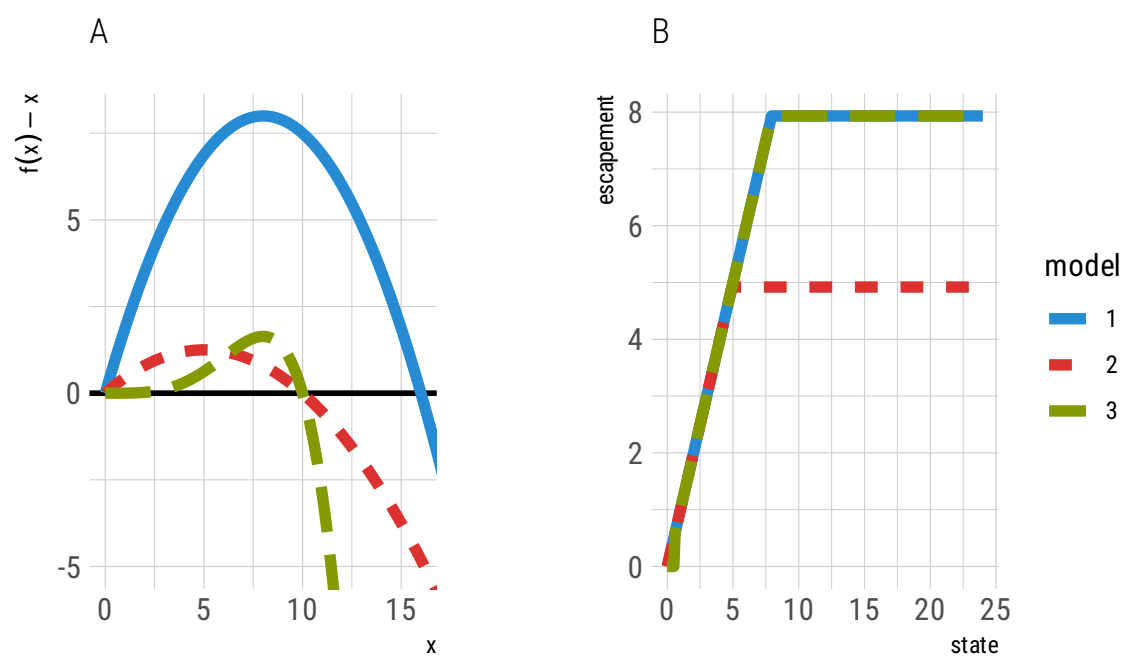


Figure 4: Panel A: Population growth curves of each model. The positive equilibrium of each model occurs where the curve crosses the horizontal axis. Note that while Model 2 is a better approximation to the truth (Model 3), Model 1 better approximates the stock size which leads to maximum growth. Panel B: The optimal control policy under Model 1 is nearly identical to that under the true Model 3, while the optimal policy under Model 2 suppresses stock to a much lower escapement level.