The Forecast Trap

[a]*Department 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 (measured by statistical scores) can lead to decidedly worse outcomes (measured by real-world objectives). This can create a forecast trap, in which the outcomes such as fish biomass or economic yield decline while the manager becomes increasingly convinced that these actions are consistent with the best models and data available. The forecast trap is not unique to this example, but possible whenever (1) the optimal management policy is not unique to the generative process, and (2) the generative process is not in our candidate set of models.

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

- abbreviated running title: "The Forecast Trap"
- 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: 149
- number of words in main text: 4967
- number of cited references: 33
- number of tables & figures: 4 Figures

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 better prediction (rigorously defined as a strictly proper score, Gneiting & Raftery (2007)) is not necessarily the model that makes the better policy (rigorously defined in terms of utility, e.g. expected net present value, Clark (1990)). I show that our best methods for learning about model structure or parameters by repeatedly comparing forecasts to observations can be counter-productive. 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. When VOI is negative, the decision-maker may become trapped into accepting mediocre outcomes derived from a model that makes accurate forecasts, even when a less accurate model that would generate better outcomes is available. In our example, the manager will decide the fishery in question simply has low productivity, because such a model yields better predictions, rather than realizing that the low productivity observed is in fact a consequence of the over-harvesting. This disconcerting situation can arise whenever two conditions are met: (1) the optimal management policy is not unique to the generative process, and (2) the generative process is not included in candidate set of models. These conditions do not guarantee the trap will occur, only the circumstances in which it cannot be ruled out entirely.

The forecast trap is not the only mechanism by which some model-choice methods lead to worse outcomes. Previous work has long acknowledged the panoply of ways in which 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). Another widely recognized problem is that of over-fitting (Burnham & Anderson 1998), in which the model that best fits historical data fails to best predict future data (Ginzburg & Jensen 2004). Under such circumstances, it is easy to see how an over-fit model would also lead to bad outcomes. However, over-fitting plays no role in the forecast trap, where model predictions are assessed only using probabilistic forecasts, and not observations which had previously been used to fit the models. Formally, these scores satisfy the 'proper scoring' rule of Gneiting & Raftery (2007), which proves no other probabilistic prediction $Q(x)$ will have a better expected score than that of the generative process $P(x)$. Gneiting & Raftery (2007)'s proof of proper scoring has since become a critical tool to avoid over-fitting when choosing models to make decisions, but as I illustrate, will not prevent the forecast trap.

From Predictive Models to Decision Policies

How do we translate a model-based forecast into a decision? It is impossible to discuss outcomes associated with a forecast without first agreeing on this process. In practice, decision-makers may use a forecast in a wide variety of ways in selecting a course of action, including ways which may run counter to the stated objectives of management (Ludwig *et al.* 1993). In principle at least, the field of decision theory provides a formal mechanism for determining the optimal strategy given a model forecast. For instance, a wide range of ecological conservation and management problems can be expressed as a Markov Decision Process (MDP) problems (Marescot *et al.* 2013). Existing computer algorithms such as stochastic dynamic programming (SDP) take a probabilistic model *forecast* (more precisely, the probability $P(x_{t+1}|x_t, a_t)$ of the system being in state x_{t+1} in the next iteration given that it was previously in state x_t and the manager selected action a_t) and the *desired management objective* (i.e. the maximize the expected biomass of species protected or the expected dollar profit of a fishery (see Clark 1990; Halpern *et al.* 2013)) as input, and return the *decision* *policy* which maximizes that objective (Marescot *et al.* 2013). This provides a principled way to associate a decision policy with any given forecast model.

Two features of this approach are worth emphasizing. First, the resulting decision is derived directly from the forecast model and the desired objective. The SDP algorithm is a reasonable description of the approach any ideal manager would use – considering all possible outcomes from all possible sequences of actions and selecting the best sequence. For complex models this process is too laborious even for a computer, and is often simplified by considering only a selection of predetermined policies (as in Management Strategy Evaluation, MSE, Punt *et al.* (2016)), or scenarios (as in scenario analysis, Polasky *et al.* (2011)). Such shortcuts are often necessary for complex real-world models, but open additional room for error: the policy we derive from a given forecast may perform poorly not because the model forecast was at fault, but because of those simplifying assumptions about possible policies. To ensure that the forecast trap is not a result of such assumptions about possible policies, we will consider a problem simple enough to solve directly with SDP. This leads to the second point: the resulting decision policy is optimal, so long as the forecast model is correct. In this way, the SDP merely stands in for a mathematically precise way in which forecasts are turned into decisions. Recognizing that the SDP-derived policy (A) comes directly from the forecast model, and (B) gives the optimal policy for said forecast, seems to suggest that the whatever model makes the better forecast will surely also lead to better outcomes (as measured in terms of whatever utility we have chosen to maximize). While this intuition is no doubt *often* accurate, our purpose here is to demonstrate that it is by no means *guaranteed*: it is also possible for the model which makes the better forecast to lead to worse outcomes.

Ecological Models

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 methods such as iterative forecasting (Dietze *et al.* 2018) and adaptive management (Walters & Hilborn 1978) can be *applied* to real-world using empirical data, we can only *evaluate* their potential in hypothetical examples when the true model is known, e.g. through numerical simulation. That approach allows us to compare both predictions and outcomes across implementations in independent identical replicate worlds. As noted above, we will assume our underlying model simple enough to solve by SDP, ensuring any poor outcomes from a given forecast are not merely an artifact of an imperfect decision process. Simple models also have the virtue in being accessible to closed form analysis, which, as we shall see, can give greater insight into when and why this forecast trap arises. That insight will in turn will allow us to examine if the same problem is likely to arise under more complex 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 $X_t + 1$,

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

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):

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

We will assume we have been given a fixed price of fish $p = 1$ with no additional cost on additional harvest, $U(X_t, H_t) = p \min(H_t, X_t)$ modest discount $\delta = 0.99$.

Let us assume that our set of candidate models are simply the possible parameterizations of a stochastic version of the classic Gordon-Schaefer model (Gordon & Press 1954; Schaefer 1954):

$$f_i(Y_{t+1}) = Y_t + r_i Y_t \left(1 - \frac{Y_t}{K_i}\right) * \xi_t(\sigma) \tag{3}$$

Where Y is the population size after harvest, $Y_t = X_t - H_t$ and $\xi_t(\sigma)$ represents log-normal random noise with a mean of unity and log-standard-deviation σ_i.

Before this or any other model can generate a forecast, we must first come up with some parameter estimates. Because the model includes (log-normal) stochastic growth, no amount of data will make any parameter combination impossible, though certain parameter values are more likely than others. Remember too that parameter estimates may be derived in other ways than than model fitting, especially when parameters are amenable to biological interpretation. We will consider the whole range of possible parameters in a moment, but for simplicity, let us begin by focusing in around two of the most interesting regions of that are already included within that larger parameter space of all possible values for r, K, and σ. Let us take "Model 1" as being given by $r_1 = 2$, $K_1 = 16$, $\sigma_1 = 0.05$, "Model 2" by $r_2 = 0.5$, $K_2 = 10$, $\sigma_2 = 0.075$. We can imagine our comparison of these two models as a microcosm of the larger comparison between all possible paremeterizations.

Ecologists will rightly scoff at the simplicity of these models – the real world is much more complicated. So it is important to bear in mind that these are not models that seek to approximate the stock dynamics of real world fisheries, only to approximate whatever "true model" we are using to drive the simulation. In recognition of the fact that real world is always more complex than even our best ecological models, we will assume a "true model" for the simulations that is not in the Gordon-Schaefer class (i.e. our candidate models will never contain the true model), but is not so rich that a Gordon-Schaefer curve would seem a hopelessly poor approximation.

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:

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

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

Certainly, the challenge of choosing which model to base a decision policy on in the real world is much harder than this binary choice between two models, and yet it is sufficient to illustrate the trap. We will see later why making the models much more complex does not guarantee that the task becomes easier or that the trap may be ruled out.

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 using the forecast-matrices of both model 1 and model 2 [Fig 1b] using SDP [Marescot *et al.* (2013); code in the Appendix]. Replicate simulations of the stock are harvested at each time step using the optimal quota dictated by either model's forecasts, according to the SDP. The resulting stock sizes in the subsequent timestep are scored against the forecast probabilities of 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 better approximation. As we see here, the better 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 management simulation. In reality, managers will typically re-estimate model parameters after each subsequent observation. And rather than consider each model/parameter combination in isolation, managers will generate forecasts which reflect the current uncertainty as to which model or parameter values are most likely. Re-estimating model parameters results in adjusting those probabilities. This approach is characterized 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, in this case, 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, (active adaptive management), I will assign initial probability that model 2 is true at 1%. After a single iteration of adaptive learning, these probabilities are completely reversed, with the manager deciding that model 2 is almost certainly correct [Fig 3A]. 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.]

So far we have considered only two alternative combinations of the parameters r, K and σ. This simplifies the calculations, because each unique parameter value combination requires a new run of the SDP algorithm to determine the optimal policy from the corresponding forecast. 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 better ecological and economic outcomes, current approaches invariably choose the wrong one. Moreover, despite continuing to collect new observations, the decision maker has no way of realizing their mistake. The manager is trapped into believing whichever model produces the better forecast, even when this results in decidedly worse objective outcomes. Re-estimating parameters with as new observations accumulate only reinforces the problem [Fig 3A], and introducing a larger suite of models, such as our wide range of r and K values, does not escape this trap either [Fig 3B]. Other model choice approaches such as goodness-of-fit, information criteria or cross-validation would all prefer model 2 as well. Only by including the true model in our set of candidates can we be certain that forecast-based methods will converge on optimal outcomes.

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. Looking at plots of the growth rate curves for each model [Fig 4A], it is hardly surprising that all model selection approaches prefer the closely overlapping curve of model 2 to the no-where-close curve of model 1 as the better approximation of model 3. Nevertheless, the decision policy derived from model 1 forecasts is indistinguishable from that based on the true model [Fig 4B], while the policy derived from model 2 forecasts lead to over-harvesting. Being closest to the true model's forecast skill never guarantees that we are closest to the true model's optimal policy.

[Figure 4 about here.]

Perhaps this should not be surprising: ecologists have long observed that all models are wrong and the choice of better model depends on the the modeling goals (Levins 1966; Walters & Hilborn 1978; Ludwig *et al.* 1993; Getz *et al.* 2018). And we are clearly considering different goals: forecast skill (a unitless statistical measure) vs policy outcomes (be they measured in dollars or fish in the ocean). Yet the result is surprising all the same. The forecast isn't just some other arbitrary modeling objective; it is a central input into the decision making process, both in the real world (Clark *et al.* 2001; Dietze *et al.* 2018) and in our idealized decision-making algorithm, SDP (Marescot *et al.* 2013). Nor can we say the same model can never be best at both goals – obviously the 'true' model always optimizes both objectives.[1] It is natural to assume from this that the candidate with the closest forecast will also be the one with the closest policy. The example presented here proves this is by no means guaranteed, but also begs the question – how common is this forecast trap?

It may seem reasonable to expect that the forecast trap would be rare in real world situations: the chance that the candidate set of models would include anything coming close to the optimal policy of the (unknown) true model without also providing a good forecast seems like it ought to be vanishingly small. In our example, we were only able to capture the optimal policy with a Gordon-Schafer model because it turned out the optimal policy boiled down to a very simple rule. Surely this does not happen in the more complex models of the real world? Surprisingly, more complex models offer no such guarantee, while real-world constraints make this situation *more* likely, not less.

How can the very different forecasts from model 1 and model 3 could produce exactly the same optimal management policy (Fig 4B) under the SDP algorithm? Analytic solutions offer more insight as to when and why very different forecasts can generate the identical policy. Such a solution was first provided by Reed (1979), who demonstrated the optimal policy in the case considered here would be a so-called "bang-bang" policy. Intuitively one can think of this as maintaining the biomass at the most productive size: the maximum population growth rate (position of the peak of the growth curves in Fig 4A), though this is only precisely true without discounting ($\delta = 1$): the optimal stock size $\hat{x}$ is the solution to $f(\hat{x}) = \hat{x}/\delta$ when stochasticity is sufficiently small (Reed 1979). Thus, all models in which the peak growth rate occurs at the same stock size will have the same optimal policy. These are not merely bad models getting lucky – all such models correctly capture the crucial feature relevant to the decision. In more complex models, such features are more difficult or impossible to identify analytically; but just because we cannot intuit the optimal policy does not mean it is uniquely complex.

Do more complex models lead to more complex control rules? The theorems of Reed (1979), while quite general, say nothing about structured models or those with predator-prey or competitive interactions. Yet recent mathematical breakthroughs such as Holden & Conrad (2015), Hening *et al.* (2019), and Hening (2021) have finally been able to extend Reed's theorem to such cases more generally. As with Reed's result, these recent proofs make it clear that the optimal strategy for managing these more complex models is not unique, but will be shared by many much simpler models. So while the true population dynamics may be given by very complex non-linear functions of the interacting species, there will be simple two species models and even un-coupled, one-species models which would lead to the identical optimal policy. It may be more likely that our candidate set contains a model which matches the optimal policy than that it contains a model which matches the generative process.

Real world considerations actually make this situation more likely, not less likely. Managers cannot resort to arbitrarily complex policies, regardless of how complex their models. Policy adjustments are costly (Boettiger *et al.* 2016) and the space of available actions is usually far more limited than the space of available states. In some ecological decision-making contexts, a manager may only be able to select between a handful of alternative actions. Such constraints make it much more likely that the optimal policy will be replicated by a much simpler model. Most well-managed marine fisheries are constrained to constant or simple piecewise-linear harvest control rules (Punt 2010). This ensures that an infinite number of possible

[1] With enough data from enough of the state space, an SDP algorithm using a Gaussian Process prior (Boettiger *et al.* 2015), which spans the "true model" given by Eq 4, will escape the forecast trap, as guaranteed by Gneiting & Raftery (2007)'s theorem. However, real ecological systems are much more complex than Eq 4, and not so easily spanned by mathematical models.

models will share the optimal solution with the true model.

Because even complex models frequently have simple control rules or else we are constrained to simple control rules as a practical matter, it is much more likely that any set of candidate models includes parameterizations that reproduce the true optimal policy than that they reproduce the optimal forecast. When data is initially limited, those simple models that could generate optimal outcomes may have non-negligible probabilities associated with them. This sets the stage for the forecast trap, adjustments with additional data or comparisons against alternative models that produce more accurate forecasts erodes those probabilities in favor of models which lead to less desirable outcomes.

The forecast trap may become a more acute issue as the simple, process-based models that have historically underpinned ecology and conservation policy are challenged by more accurate forecasts from statistical and machine learning tools. Buoyed by the rapid expansion of available data and computational power, complex and increasingly opaque models are becoming more common (Desjardins-Proulx *et al.* 2019). Evaluating such models based on forecast skill will not only reduce concerns about over-fitting, but will make a compelling illustration of their viability as a tool for informing policy. In many cases, these more accurate forecasts may very well prove invaluable in delivering better (or less bad) real-world outcomes. But this paper is a reminder that such outcomes are by no means an inevitable consequence of better forecasts. These more accurate predictions are still not the true model, and it so it is always possible to improve predictive accuracy while less accurately reflecting the unknown key features that really drive the policy decision. I hope this simple and intuitive example will provide a ready reminder as to why the model that produces the best forecast will not always produce the best decision.

Acknowledgements

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

References

Boettiger, C., Bode, M., Sanchirico, J.N., LaRiviere, J., Hastings, A. & Armsworth, P.R. (2016). Optimal management of a stochastically varying population when policy adjustment is costly. *Ecological Applications*, 26, 808–817.

Boettiger, C., Mangel, M. & Munch, S. (2015). Avoiding tipping points in fisheries management through Gaussian process dynamic programming. *Proceedings of the Royal Society B: Biological Sciences*, 282, 20141631–20141631.

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, C.W. (1990). *Mathematical Bioeconomics: The Optimal Management of Renewable Resources, 2nd Edition*. Wiley-Interscience.

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.

Desjardins-Proulx, P., Poisot, T. & Gravel, D. (2019). Artificial Intelligence for Ecological and Evolutionary Synthesis. *Frontiers in Ecology and Evolution*, 7, 402.

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.

Getz, W.M., Marshall, C.R., Carlson, C.J., Giuggioli, L., Ryan, S.J., Romañach, S.S., *et al.* (2018). Making ecological models adequate. *Ecology Letters*, 21, 153–166.

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.

Hening, A. (2021). Coexistence, Extinction, and Optimal Harvesting in Discrete-Time Stochastic Population Models. *Journal of Nonlinear Science*, 31, 1.

Hening, A., Nguyen, D.H., Ungureanu, S.C. & Wong, T.K. (2019). Asymptotic harvesting of populations in random environments. *Journal of Mathematical Biology*, 78, 293–329.

Holden, M.H. & Conrad, J.M. (2015). Optimal escapement in stage-structured fisheries with environmental stochasticity. *Mathematical Biosciences*, 269, 76–85.

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 fallowing/planting problem. *Journal of Forecasting*, 6, 77–89.

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.

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. (2010). Harvest control rules and fisheries management. In: *Handbook of Marine Fisheries Conservation and Management* . (eds. Grafton, RQ, Hilborn, R, Squires, D, Tait, M & Williams, M). Oxford University Press.

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.

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

Schaefer, M.B. (1954). 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.

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

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

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

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

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

List of Figures

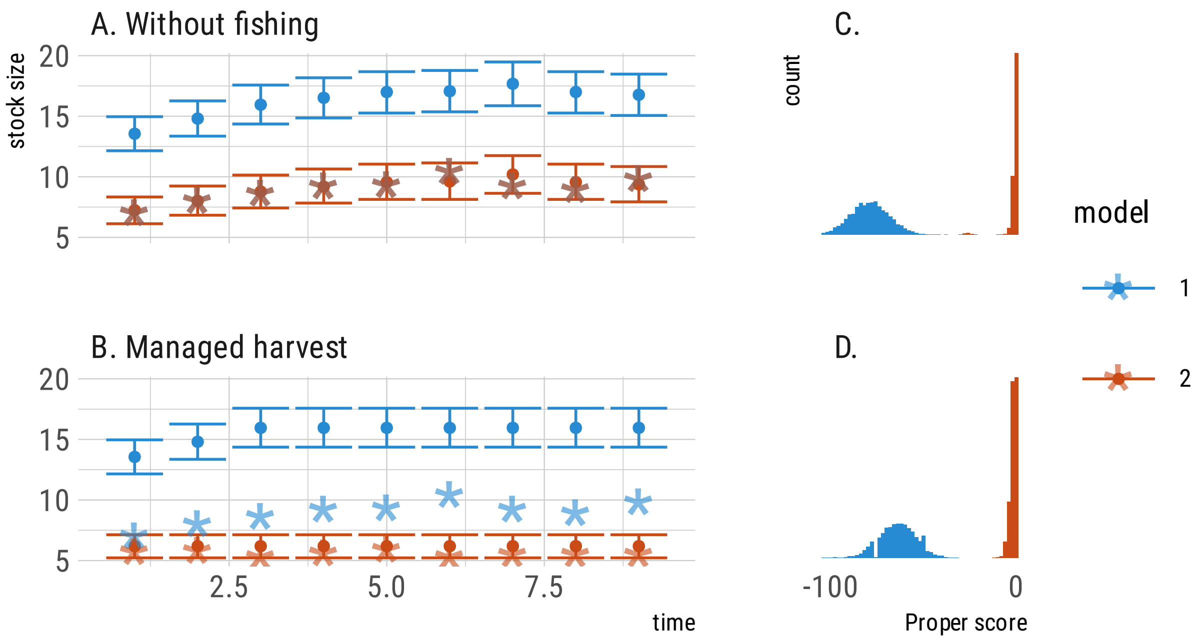

Figure 1: Forecast performance of each model. Panels A, B: Step ahead predictions of stock size under unfished (A) and fished (B) scenarios. Error bars indicating the 95% confidence intervals around each prediction, while stars denote the observed value in that year. Because the models make different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

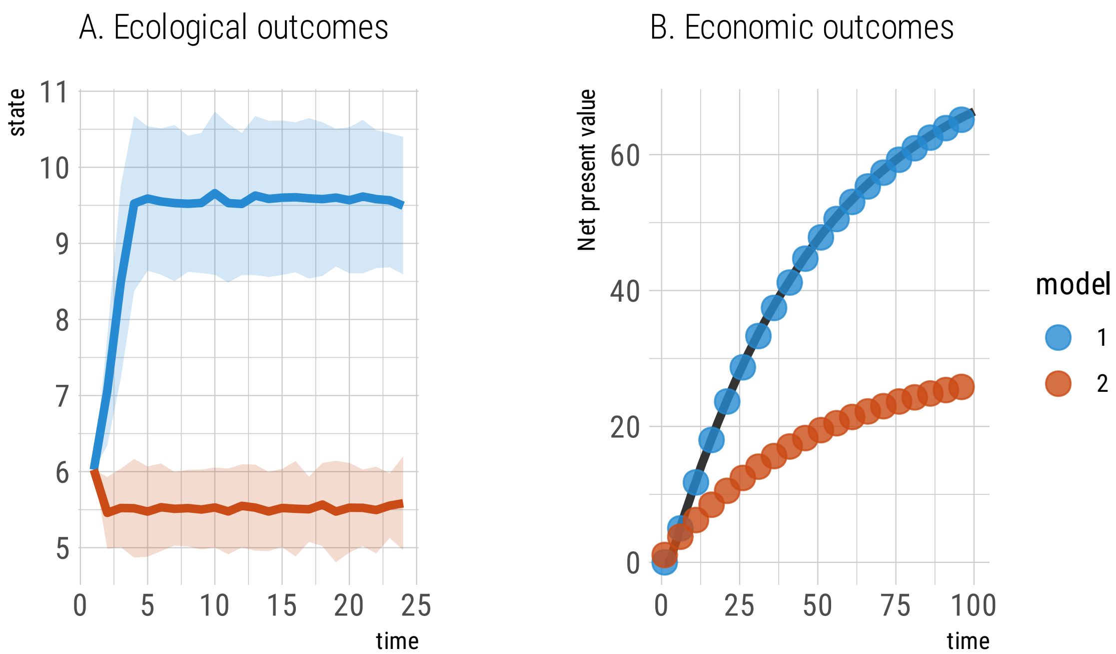

Figure 2: Ecological and economic performance of each forecast. Harvest quotas derived from model 1 result in a significantly higher fish stock size than under Model 2 (panel A). Economic returns under model 1 are also substantially higher (panel B)

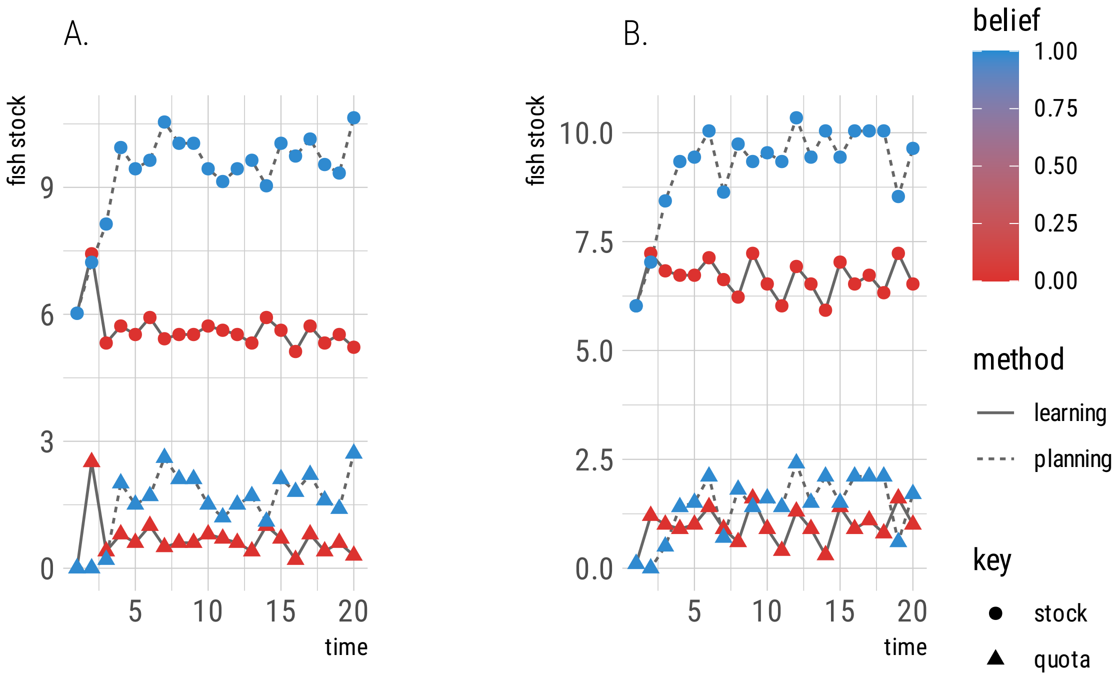

Figure 3: Adaptive management under model uncertainty. Solid lines trace the trajectories of the state (fish stock, circles) and action (harvest quota, triangles), under adaptive management (learning). Dotted lines trace the corresponding trajectories if iterative learning is omitted, leaving the prior 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.

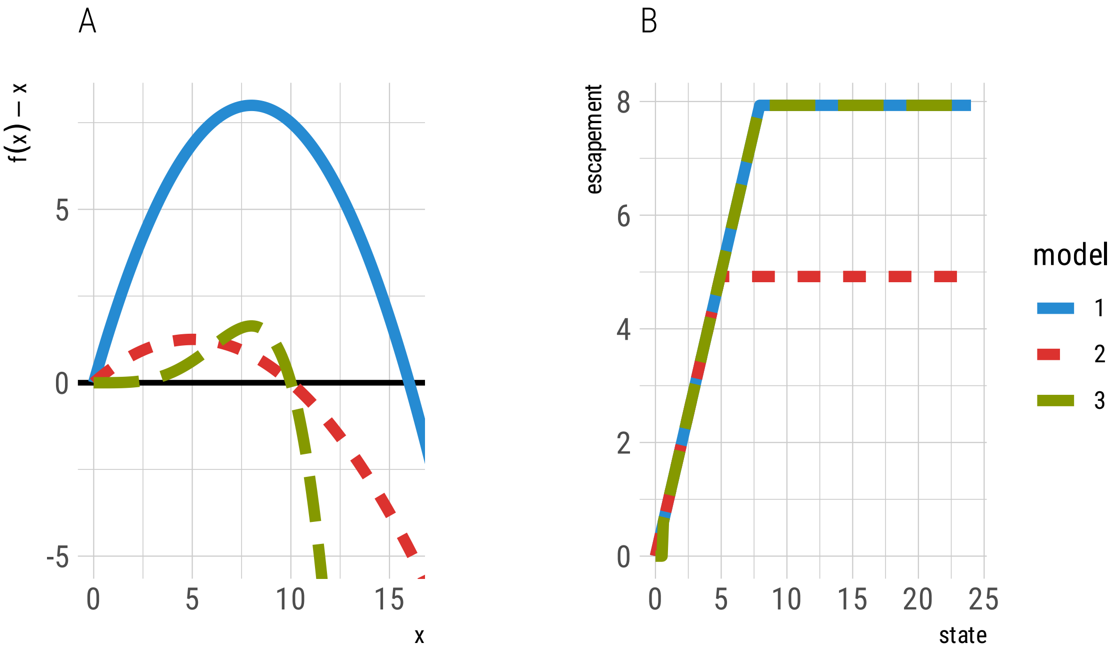

Figure 4: Panel A: Population growth curves of each model. The positive equilibrium of each model occurs where the curve crosses the horizontal axis. Note that while Model 2 is a better approximation to the truth (Model 3), Model 1 better approximates the stock size which leads to maximum growth. Panel 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.