Title: Bad forecast, good decision: predictive accuracy is not everything

Authors: Carl Boettiger[1]*

Affiliations:

[1] Department of Environmental Science, Policy, and Management; University of California, Berkeley

*Correspondence to: cboettig@berkeley.edu

Abstract: Model-based forecasts have been enabled by recent explosion of data and computational methods and spurred on by decision-makers appetite for forecasts in everything from elections to pandemic response. It is taken for granted that the model which makes the most accurate forecast, accounting for uncertainty, will also be the best model to inform decision-making. Using a classic example from fisheries management, I demonstrate that selecting the model that produces the best forecast can lead to decidedly worse outcomes. This situation can arise whenever the models are only approximations of the reality, underscoring the risk of evaluating models only by prediction accuracy and not decision outcomes.

One Sentence Summary: A model that makes decisively bad forecasts can lead to much better decisions than the model with more accurate predictions.

Main Text:

A primary purpose of statistical analyses and modeling is to make forecasts for the future.[1,2] Accurate forecasts are important not only in assessing our understanding of natural processes, they also underpin policy and decision making.[3] Strictly proper scoring rules provide mathematical guarantees that they will, on average, select the true model over any other,[4] paving the way for ever more complex and more accurate forecasting models. Yet the model that leads to the best decisions is not always the model that makes the most accurate forecasts, as I illustrate here. Surprisingly, this can even happen when a decision is derived directly from a complex optimization routine of a probabilistic predictive models.[5] Reality is complex; even our best models can only ever be approximations of underlying processes. 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 decision-making 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 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.

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},

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

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:

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

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:

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

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

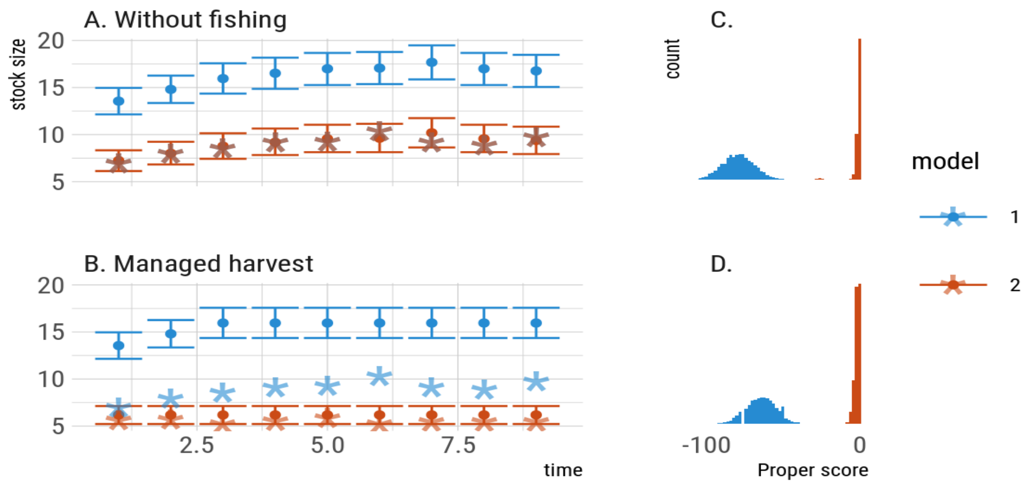

Figure 1: *Forecast performance of each model. Panels A, B: Step ahead predictions of stock size under unfished (A) and fished (B) scenarios. Error bars indicating the 95% confidence intervals around each prediction, while stars denote the observed value in that year. Because the models make different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

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 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).

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.

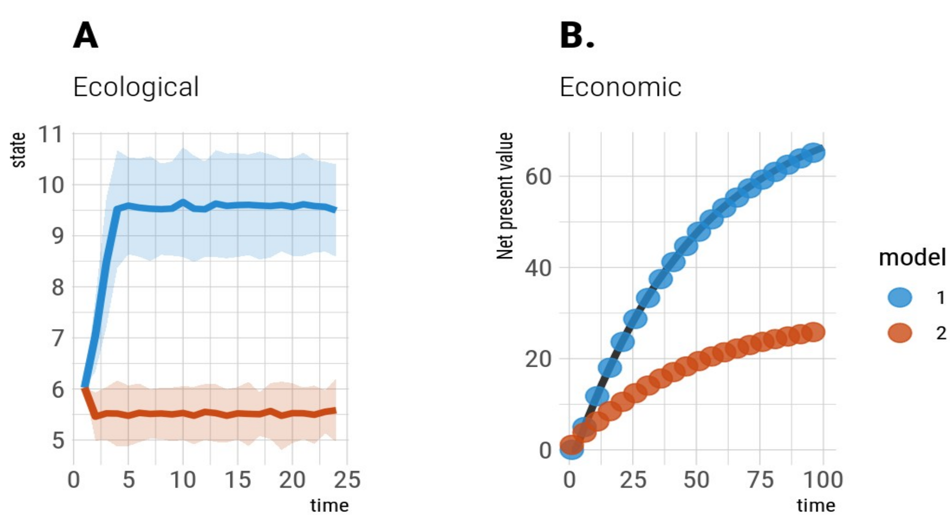

Figure 2: *Ecological and economic performance of each forecast. Harvest quotas derived from Model 1 result in a significantly higher fish stock size than under Model 2 (panel A). Economic returns under Model 1 are also substantially higher (panel B).*

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 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.

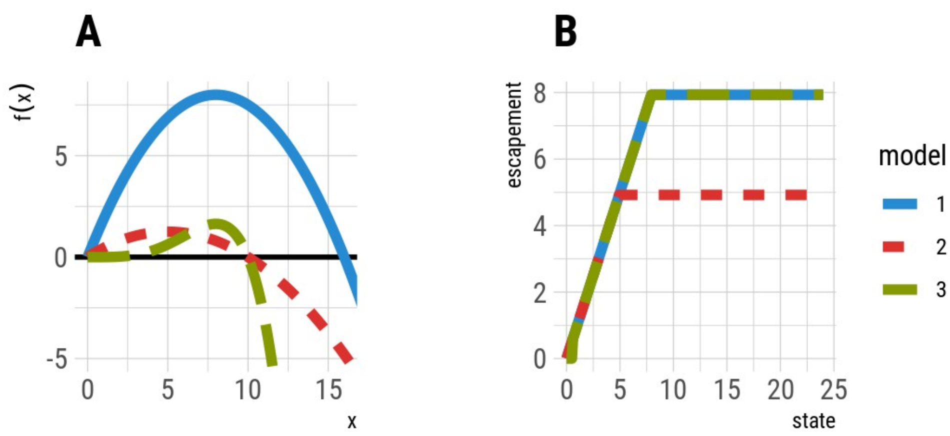

Figure 3: *Panel A: Population recruitment curves for each model, compared to that of the true Model. 3 Model 2 more closely approximates the true Model 3, but note that maximum value Model 1 and Model 3 occur at nearly the same value for the state, x. Panel B: The computed optimal policy of each model, derived by SDP, expressed in terms of the target escapement (population size remaining after harvest) for each possible stock size. Model 1 over-harvests consistently, while the target escapement under Model 2 is nearly identical to that of the true Model 3.*

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.

References and Notes:

1. J. S. Clark, S. R. Carpenter, M. Barber, S. Collins, A. Dobson, J. A. Foley, D. M. Lodge, M. Pascual, R. Pielke, W. Pizer, C. Pringle, W. V. Reid, K. A. Rose, O. Sala, W. H. Schlesinger, D. H. Wall, D. Wear, Ecological Forecasts: An Emerging Imperative. *Science*. **293**, 657–660 (2001).
2. T. Gneiting, M. Katzfuss, Probabilistic Forecasting. *Annual Review of Statistics and Its Application*. **1**, 125–151 (2014).
3. M. C. Dietze, A. Fox, L. M. Beck-Johnson, J. L. Betancourt, M. B. Hooten, C. S. Jarnevich, T. H. Keitt, M. A. Kenney, C. M. Laney, L. G. Larsen, H. W. Loescher, C. K. Lunch, B. C. Pijanowski, J. T. Randerson, E. K. Read, A. T. Tredennick, R. Vargas, K. C. Weathers, E. P. White, Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proceedings of the National Academy of Sciences*. **115**, 1424–1432 (2018).
4. T. Gneiting, A. E. Raftery, Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*. **102**, 359–378 (2007).
5. L. Marescot, G. Chapron, I. Chadès, P. L. Fackler, C. Duchamp, E. Marboutin, O. Gimenez, Complex decisions made simple: A primer on stochastic dynamic programming. *Methods in Ecology and Evolution*. **4**, 872–884 (2013).
6. M. B. Schaefer, 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. C. W. Clark, Profit maximization and the extinction of animal species. *Journal of Political Economy*. **81**, 950–961 (1973).
8. W. J. Reed, Optimal escapement levels in stochastic and deterministic harvesting models. *Journal of Environmental Economics and Management*. **6**, 350–363 (1979).
9. B. Worm, E. B. Barbier, N. Beaumont, J. E. Duffy, C. Folke, B. S. Halpern, J. B. C. Jackson, H. K. Lotze, F. Micheli, S. R. Palumbi, E. Sala, K. a Selkoe, J. J. Stachowicz, R. Watson, Impacts of biodiversity loss on ocean ecosystem services. *Science (New York, N.Y.)*. **314**, 787–90 (2006).
10. B. Worm, R. Hilborn, J. K. Baum, T. A. Branch, J. S. Collie, C. Costello, M. J. Fogarty, E. a Fulton, J. a Hutchings, S. Jennings, O. P. Jensen, H. K. Lotze, P. M. Mace, T. R. McClanahan, C. Minto, S. R. Palumbi, A. M. Parma, D. Ricard, A. a Rosenberg, R. Watson, D. Zeller, Rebuilding global fisheries. *Science (New York, N.Y.)*. **325**, 578–85 (2009).
11. C. Costello, D. Ovando, T. Clavelle, C. K. Strauss, R. Hilborn, M. C. Melnychuk, T. A. Branch, S. D. Gaines, C. S. Szuwalski, R. B. Cabral, D. N. Rader, A. Leland, Global fishery prospects under contrasting management regimes. *Proceedings of the National Academy of Sciences*. **113**, 5125–5129 (2016).
12. M. Mangel, Decision and control in uncertain resource systems, 255 (1985).
13. C. W. Clark, *Mathematical Bioeconomics: The Optimal Management of Renewable Resources, 2nd Edition* (Wiley-Interscience, 1990).
14. R. Levins, The strategy of model building in population biology. *American Scientist*. **54**, 421–431 (1966).
15. A. E. Raftery, 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. D. Ricard, C. Minto, O. P. Jensen, J. K. Baum, 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. RAM Legacy Stock Assessment Database v4.44 (2018),, doi:10.5281/ZENODO.2542919.
18. E. P. White, G. M. Yenni, S. D. Taylor, E. M. Christensen, E. K. Bledsoe, J. L. Simonis, S. K. M. Ernest, Developing an automated iterative near-term forecasting system for an ecological study. *Methods in Ecology and Evolution*. **10**, 332–344 (2019).
19. M. Memarzadeh, C. Boettiger, Resolving the measurement uncertainty paradox in ecological management. *The American Naturalist*. **193**, 645–660 (2019).
20. H. Tallis, P. Kareiva, Shaping global environmental decisions using socio-ecological models. *Trends in Ecology & Evolution*. **21**, 562–568 (2006).
21. C. J. Walters, R. Hilborn, Ecological Optimization and Adaptive Management. *Annual Review of Ecology and Systematics*. **9**, 157–188 (1978).

Acknowledgments: Acknowledgments follow the references and notes but are not numbered. Acknowledgments should be gathered into a paragraph after the final numbered reference. This section should start by acknowledging non-author contributions and then should provide information under the following headings: **Funding:** This work was supported in part by NSF CAREER (#1942280) and computational resources from NSF's XSEDE Jetstream (DEB160003) and Chameleon cloud platforms; **Competing interests:** Author declares no competing interests. **Data and materials availability:** Code for fully reproducing the analysis can be found in the appendix, and associated GitHub repository: <https://github.com/cboettig/bad-forecast-good-decision>.

List Of Supplementary Materials: See the appendix for annotated code and detailed description of methods involved.

Appendix for: "Bad Forecast, Good Decision"

Carl Boettiger

2020-11-19

Here we provide annotated code necessary to completely reproduce all of the analysis presented in the main paper. This analysis is run in R (R Core Team 2019) uses `MDPtoolbox` (Chades et al. 2017) for solving Markov Decision Processes (MDP) using stochastic dynamic programming functionality, `expm` for matrix exponentials (Goulet et al. 2020), and a few custom MDP functions provided by our package, `mdplearning` (Carl Boettiger 2018). We will also use `tidyverse` packages for basic manipulation and plotting (Wickham et al. 2019). This file is also available as an RMarkdown document (Xie, Allaire, and Grolemund 2018) at https://github.com/cboettig/bad-forecast-good-decision.

```
library(tidyverse)
library(MDPtoolbox)
library(expm)
# remotes::install_github("boettiger-lab/mdplearning")
library(mdplearning)
```

Our decision problem is defined as follows: A manager seeks to 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:

$$\sum_{t=0}^{t=\infty} U(X_t, H_t)\delta^t \tag{1}$$

While in principle the utility could reflect many things, including the cost of fishing, market responses to supply and demand, the value of recreational fishing, the intrinsic value fish left in the sea (see Halpern et al. 2013), for simplicity we will assume utility is merely a linear function of the harvest quota set by the manager, i.e. a fixed price p per kilogram of fish harvested: $U(X_t, H_t) = p \min(H_t, X_t)$ (noting that realized harvest cannot exceed the available stock). As units are considered already non-dimensionalized in this example, without loss of generality we will set $p = 1$. This problem is already well studied, and it is worth noting that even under such a pessimistic assumption, the optimal strategy still seeks to sustain the fish population indefinitely; as Clark (1973) shows for the deterministic function f and Reed (1979) extended to the stochastic case. Given the function f with known parameters, it is straight forward to determine the optimal harvest policy by stochastic dynamic programming (SDP, see Mangel (1985); Marescot et al. (2013)).

To solve the decision problem using SDP, we define the state space and the action space on a discrete grid of 240 points. The maximum state is set well above the largest carrying capacity used in the models, which limits the influence of boundary effects introduced by the transformation to a discrete, finite grid. Available harvest actions match the state space, effectively allowing any harvest level to be possible.

```
states <- seq(0, 24, length.out = 240)
actions <- states
obs <- states
```

The utility (reward) of an action is set to the (realized) harvest (e.g. a fixed price per unit fish, with no cost applied to harvest effort). Future utility is discounted by fixed factor of 0.99. Classic maximum sustainable yield models (Schaefer 1954) ignore discounting, while modern economic optimization models insist on it, so a small discount conforms to the latter while reasonably approximating the former. The qualitative conclusion is not sensitive to the discounting rate.

```r
reward_fn <- function(x, h) pmin(x, h)
discount <- 0.99
```

Alternate reward functions with varying price and cost structures are possible but do not qualitatively impact the conclusions. This reward function is a limiting case of any more complex reward, and corresponds both to classic work that does not model utility explicitly (e.g. MSY theory, Schaefer 1954), as well as explicit assumptions typically made in more recent models (Reed 1979). Moreover, using a simple reward function makes it clear that model that performs best is not doing so merely because of particular features baked into the a carefully chosen reward rule.

Ecological Models

The manager chooses between two different logistic growth models, each of which can be thought of as an approximation to the underlying "true" population model:

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

Model 1 has $r_1 = 2$, $K = 16$, and $\sigma_1 = 0.05$. Model 2 has $r_2 = 0.5, K_2 = 10$, and $\sigma_2 = 0.075$.

Meanwhile, the true population growth rate is simulated using a function with non-linear per-capita growth:

$$f_i(Y) = Y + Y^4 r_i\left(1 - \frac{Y}{K_i}\right) \tag{3}$$

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

```r
# K is at twice max of f3; 8 * K_3 / 5
f1 <- function(x, h = 0, r = 2, K = 10 * 8 / 5) {
  s <- pmax(x - h, 0)
  s + s * (r * (1 - s / K))
}
f2 <- function(x, h = 0, r = 0.5, K = 10) {
  s <- pmax(x - h, 0)
  s + s * (r * (1 - s / K))
}

# max is at 4 * K / 5
f3 <- function(x, h = 0, r = .002, K = 10) {
  s <- pmax(x - h, 0)
  s + s^4 * r * (1 - s / K)
}

## gather models together, indicate true model
sigma_g <- 0.05
models <- list("1" = f1, "2" = f2, "3" = f3)
model_sigmas <- c(sigma_g, 1.5 * sigma_g, sigma_g)
true_model <- "3"
```

On a discrete grid of possible states and actions, we can define the growth rate of a given state X_t subject to harvest H_t, $f(X_t, H_t)$ as set of matrices. Each matrix i gives the transition probabilities for any current state to any future state, given that action i is taken.

```r
transition_matrices <- function(f, states, actions, sigma_g) {
  n_s <- length(states)
  n_a <- length(actions)
  transition <- array(0, dim = c(n_s, n_s, n_a))
  for (k in 1:n_s) {
    for (i in 1:n_a) {
      nextpop <- f(states[k], actions[i])
      if (nextpop <= 0) {
        transition[k, , i] <- c(1, rep(0, n_s - 1))
      } else if (sigma_g > 0) {
        x <- dlnorm(states, log(nextpop), sdlog = sigma_g)
        if (sum(x) == 0) { ## nextpop is computationally zero
          transition[k, , i] <- c(1, rep(0, n_s - 1))
        } else {
          x <- x / sum(x) # normalize evenly
          transition[k, , i] <- x
        }
      }
    }
  }
  transition
}
```

This follows the standard setup for standard stochastic dynamic programming, see Marescot et al. (2013). Having defined a function to compute the transition matrix, we can use it to create matrices corresponding to each of the three models:

```r
transitions <- lapply(
  seq_along(models),
  function(i) {
    transition_matrices(
      models[[i]],
      states,
      actions,
      model_sigmas[[i]]
    )
  }
)
names(transitions) <- c("1", "2", "3")
```

Likewise, a corresponding matrix defining the rewards associated with each state X and each harvest action H can also be defined.

```r
## Compute reward matrix (shared across all models)
n_s <- length(states)
n_a <- length(actions)
reward <- array(0, dim = c(n_s, n_a))
for (k in 1:n_s) {
  for (i in 1:n_a) {
    reward[k, i] <- reward_fn(states[k], actions[i])
  }
}
```

Optimal control solutions

We use value iteration to solve the stochastic dynamic program (Marescot et al. 2013; Chades et al. 2017) for each model. This determines the optimal harvest policy for each possible state, given each model. Because this step is the most computationally intensive routine, we cache the results using memosization conditioned on the transition matrices (Wickham et al. 2017). Running this code with alternate transition matrices automatically invalidates that cache, reducing the risk of loading spurious results.

```
mdp <- memoise::memoise(mdp_value_iteration,
  cache = memoise::cache_filesystem("cache/")
)

policies <-
  map_dfr(transitions,
    function(P) {
      soln <- mdp(P, reward,
        discount = discount,
        epsilon = 0.01, max_iter = 1000, V0 = rep(0, dim(P)[[1]])
      )
      escapement <- states - actions[soln$policy]
      tibble(states, policy = soln$policy, escapement)
    },
    .id = "model"
  )
```

Simulations and step-ahead forecasts

We simulate fishing dynamics under the optimal policy for each model, using a simple helper function from the `mdplearning` package. Because growth dynamics are stochastic, we perform 100 simulations of each model from identical starting condition to ensure results are not the result of chance alone.

```
library(mdplearning)
Tmax <- 100
x0 <- which.min(abs(states - 6))
reps <- 100
set.seed(12345)

## Simulate each policy reps times, with `3` as the true model:
simulate_policy <- function(i, policy) {
  mdp_planning(transitions[[true_model]], reward, discount,
    policy = policy, x0 = x0, Tmax = Tmax
  ) %>%
    select(value, state_index = state, time, action_index = action) %>%
    mutate(state = states[state_index])
}

sims <-
  map_dfr(names(transitions),
    function(m) {
      policy <- policies %>%
        filter(model == m) %>%
        pull(policy)
      map_dfr(1:reps, simulate_policy, policy = policy, .id = "reps")
    },
```

```
    .id = "model"
  )
```

Using the transition matrices directly, we can examine what each model would have forecast the future stock size to be in the following year when no fishing occurs (note that for each model, we use the transition matrix that corresponds to 'no fishing', `model[[state_index, ,1]]`) (Fig 1a, main text).

The transition matrices give the full (discretized) probability distribution, from which we can easily calculate both the expected value and the 95% confidence interval.

```
stepahead_unfished <- sims
stepahead_unfished$state_index <- rep(sims$state_index[sims$model == "1"], 3)

stepahead_unfished <- stepahead_unfished %>%
  filter(model != "3") %>%
  mutate(next_state = dplyr::lead(state_index), model = as.integer(model)) %>%
  rowwise() %>%
  mutate(
    expected = transitions[[model]][state_index, , 1] %*% states,
    var = transitions[[model]][state_index, , 1] %*% states^2 - expected^2,
    low = states[max(which(cumsum(transitions[[model]][state_index, , 1]) < 0.025))],
    high = states[min(which(cumsum(transitions[[model]][state_index, , 1]) > 0.975))],
    true = states[next_state]
  )
```

We also look at the forecast each model makes when implementing the corresponding optimal harvest:

```
stepahead_fished <- sims %>%
  filter(model != "3") %>%
  mutate(next_state = dplyr::lead(state_index), model = as.integer(model)) %>%
  rowwise() %>%
  mutate(
    prob = transitions[[model]][state_index, next_state, action_index],
    expected = transitions[[model]][state_index, , action_index] %*% states,
    var = transitions[[model]][state_index, , action_index] %*% states^2 - expected^2,
    low = states[max(which(cumsum(transitions[[model]][state_index, , action_index]) < 0.025))],
    high = states[min(which(cumsum(transitions[[model]][state_index, , action_index]) > 0.975))],
    true = states[next_state]
  ) %>%
  select(time, model, true, expected, low, high, var, prob, reps)
```

Proper scores

It is straight forward to apply the proper scoring formula of Gneiting and Raftery (2007) based on the first two moments of the distribution to score the respective forecasts under both the unfished and actively managed scenario sfor each model:

```
# Gneiting & Raferty (2007), eq27
scoring_fn <- function(x, mu, sigma) {
  -(mu - x)^2 / sigma^2 - log(sigma)
}

stepahead_unfished <- stepahead_unfished %>%
  mutate(
    sd = sqrt(var),
```

```
    score = scoring_fn(expected, true, sd)
  )

stepahead_fished <- stepahead_fished %>%
  mutate(
    sd = sqrt(var),
    score = scoring_fn(expected, true, sd)
  )
```

```
predictions <-
  stepahead_unfished %>%
  select(time, model, reps, expected, low, high, true, score) %>%
  mutate(scenario = "A_unfished") %>%
  bind_rows(stepahead_fished %>%
    select(time, model, reps, expected, low, high, true, score) %>%
    mutate(scenario = "B_fished")) %>%
  mutate(model = as.character(model))
```

Plotting

Having computed a single data frame of simulation data under both optimally fished (according to each model) and unfished scenarios for each model, along with the proper scores for the step-ahead forecasts a la Gneiting and Raftery (2007), we have all the results necessary to generate the figures presented in the main paper.

```
fig1ab <- predictions %>%
  filter(reps == "2", time < 10) %>%
  ggplot(aes(time, col = model, fill = model)) +
  geom_point(aes(y = expected)) +
  geom_errorbar(aes(ymin = low, ymax = high)) +
  geom_point(aes(y = true),
    pch = "*", size = 12,
    alpha = 0.6
  ) +
  facet_wrap(~scenario, ncol = 1, labeller = as_labeller(c(
    A_unfished = "A. Without fishing",
    B_fished = "B. Managed harvest"
  ))) +
  ylab("stock size")
```

Likewise we can plot the data on proper scores associated with each prediction of each model:

```
fig1cd <- predictions %>%
  ggplot(aes(x = score, group = model, fill = model)) +
  geom_histogram(binwidth = 2, show.legend = FALSE) +
  coord_cartesian(xlim = c(-100, 1), ylim = c(0, 4000)) +
  xlab("Proper score") +
  facet_wrap(~scenario, ncol = 1, labeller = as_labeller(c(
    A_unfished = "C.",
    B_fished = "D."
  ))) +
  scale_x_continuous(breaks = c(-100, 0)) +
  theme(
    axis.text.y = element_blank(),
```

```
    plot.margin = margin(0, 0, 0, 0, "cm"),
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank()
  )
```

Forecast ecological and economic performance

The ecological performance of the model can easily be visualized by plotting the results of each simulation.

```
fig2a <-
  sims %>%
  group_by(model, time) %>%
  summarise(mean_state = mean(state), sd = sd(state), .groups = "drop") %>%
  filter(time < 25, model != "3") %>%
  ggplot(aes(time, mean_state)) +
  geom_line(aes(col = model), lwd = 1.5, show.legend = FALSE) +
  geom_ribbon(aes(
    ymin = mean_state - 2 * sd,
    ymax = mean_state + 2 * sd,
    fill = model
  ),
  alpha = 0.2, show.legend = FALSE
  ) +
  ylab("state") +
  ggtitle("A") +
  labs(subtitle = "Ecological")
```

To plot the economic value over time, we must sum up the discounted values at each time step, and then average over replicate simulations of each model:

```
##  Net Present Value accumulates over time
npv_df <- sims %>%
  group_by(model, reps) %>%
  mutate(npv = cumsum(value * discount^time)) %>%
  group_by(time, model) %>%
  summarise(mean_npv = mean(npv), .groups = "drop") %>%
  arrange(model, time)

optimal <- select(filter(npv_df, model == "3"), time, mean_npv)

fig2b <-
  npv_df %>%
  filter(model != "3", time %in% seq(1, 100, by = 5)) %>%
  ggplot(aes(time, mean_npv)) +
  geom_line(data = optimal, lwd = 1.5, col = "grey20") +
  geom_point(aes(col = model), size = 4, alpha = 0.8) +
  ylab("Net present value") +
  xlab("time") +
  ggtitle("B.") +
  labs(subtitle = "Economic")
```

Plotting the model functions themselves are shown in Fig 3a:

```
d <-
  map_dfc(models, function(f) f(states) - states) %>%
  mutate(state = states)
```

```r
fig3a <-
  d %>%
  pivot_longer(names(models), "model") %>%
  ggplot(aes(state, value, col = model, lty = model)) +
  geom_hline(aes(yintercept = 0), lwd = 1) +
  geom_line(lwd = 2, show.legend = FALSE) +
  coord_cartesian(ylim = c(-5, 8), xlim = c(0, 16)) +
  ylab(bquote(f(x))) +
  xlab("x")
```

We can easily plot the optimal policy derived from each model, as shown in Fig3b.

```r
fig3b <- policies %>%
  ggplot(aes(states, escapement, col = model, lty = model)) +
  geom_line(lwd = 2) +
  xlab("state")
```

References

Carl Boettiger, Milad Memarzadeh. 2018. *Mdplearning: Bayesian Learning Algorithms for Markov Decision Processes* (version 0.1.0). Zenodo. https://doi.org/10.5281/zenodo.1161519.

Chades, Iadine, Guillaume Chapron, Marie-Josee Cros, Frederick Garcia, and Regis Sabbadin. 2017. *MDP-toolbox: Markov Decision Processes Toolbox*.

Clark, Colin W. 1973. "Profit Maximization and the Extinction of Animal Species." *Journal of Political Economy* 81 (4): 950–61. https://doi.org/10.1086/260090.

Gneiting, Tilmann, and Adrian E Raftery. 2007. "Strictly Proper Scoring Rules, Prediction, and Estimation." *Journal of the American Statistical Association* 102 (477): 359–78. https://doi.org/10.1198/016214506000001437.

Goulet, Vincent, Christophe Dutang, Martin Maechler, David Firth, Marina Shapira, and Michael Stadelmann. 2020. *Expm: Matrix Exponential, Log, 'Etc'*. http://R-Forge.R-project.org/projects/expm/.

Halpern, Benjamin S, Carissa J Klein, Christopher J Brown, Maria Beger, Hedley S Grantham, Sangeeta Mangubhai, Mary Ruckelshaus, 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 (15): 6229–34. https://doi.org/10.1073/pnas.1217689110.

Mangel, Marc. 1985. "Decision and control in uncertain resource systems," 255. http://dl.acm.org/citation.cfm?id=537497.

Marescot, Lucile, Guillaume Chapron, Iadine Chadès, Paul L. Fackler, Christophe Duchamp, Eric Marboutin, and Olivier Gimenez. 2013. "Complex Decisions Made Simple: A Primer on Stochastic Dynamic Programming." *Methods in Ecology and Evolution* 4 (9): 872–84. https://doi.org/10.1111/2041-210X.12082.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing* (version 3.6.2). Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Reed, William J. 1979. "Optimal escapement levels in stochastic and deterministic harvesting models." *Journal of Environmental Economics and Management* 6 (4): 350–63. https://doi.org/10.1016/0095-0696(79)90014-7.

Schaefer, Milner 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 (2): 27–56. https://doi.org/10.1007/BF02464432.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Jim Hester, Kirill Müller, and Daniel Cook. 2017. *Memoise: Memoisation of Functions*. https://github.com/hadley/memoise.

Xie, Yihui, J. J. Allaire, and Garrett Grolemund. 2018. *R Markdown: The Definitive Guide*. Boca Raton, Florida: Chapman; Hall/CRC. https://bookdown.org/yihui/rmarkdown.