



A nonlinear, low data requirement model for producing spatially explicit fishery forecasts

SARAH M. GLASER,^{1,*} HAO YE² AND
GEORGE SUGIHARA²

¹Virginia Institute of Marine Science, The College of William & Mary, PO Box 1346, Gloucester Point, VA, 23062, U.S.A.

²Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive MC 0202, La Jolla, CA, 92093-0202, U.S.A.

ABSTRACT

Spatial variability can confound accurate estimates of catch per unit effort (CPUE), especially in highly migratory species. The incorporation of spatial structure into fishery stock assessment models should ultimately improve forecasts of stock biomass. Here, we describe a nonlinear time series model for producing spatially explicit forecasts of CPUE that does not require ancillary environmental or demographic data, or specification of a model functional form. We demonstrate this method using spatially resolved ($1^\circ \times 1^\circ$ cells) CPUE time series of North Pacific albacore in the California Current System. The spatial model is highly significant ($P < 0.00001$) and outperforms two spatial null models. We then create a spatial forecast map for years beyond the range of data. Such approaches can guide spatial management of resources and provide a complement to more data-intensive, highly parameterized population dynamics and ecosystem models currently in use.

Key words: California Current System, catch per unit effort, data-poor fisheries, nonlinear forecasting, North Pacific albacore, spatial distribution model

INTRODUCTION

Spatial variability in local population abundance and distribution is driven by differences in resources, physiological tolerances, predation and mortality, and habitat degradation (Sale, 1977; Werner *et al.*, 1983; Rose and Leggett, 1990; Dunning *et al.*, 1992; Wiens

et al., 1993; Fahrig and Merriam, 1994). In exploited populations, uneven harvesting may confound further the spatial variability resulting from environmental or biological interactions (Pelletier *et al.*, 2011). Ignoring such variability can hamper identification of key ecological relationships: links between variables can be masked if the scale of analysis is not resolved highly enough. In fisheries, failure to account for spatial variability is one reason catch per unit effort (CPUE) may not be proportional to real abundance (Vignaux, 1996; Walters, 2003). Ecosystem-based approaches to managing commercially harvested fish populations have therefore stressed the importance of including spatial structure in stock assessment models (Booth, 2000; Apostolaki *et al.*, 2002; Hobday and Hartmann, 2006).

A host of challenges plagues spatially explicit population models (SEPM), especially when considering highly mobile, difficult-to-observe species. First, when linking population estimates to environmental variables, one must decide which variables to include. The choice may be constrained artificially by the availability of overlapping variables at an appropriate scale. If few environmental variables are available, they may be included in a model less for their relevance and more for their availability. If many environmental variables are available, biological meaning may be sacrificed if the inclusion of additional environmental indices improves model fit (especially when processes are not orthogonal). Secondly, when the spatial scale of population and environmental variables do not match, the modeler must choose to interpolate from low resolution to higher resolution, or aggregate from high resolution to lower resolution. Either choice may distort patterns. Thirdly, as models become more structurally complex and require more parameter estimation, error propagation becomes a significant concern (Ruckelshaus *et al.*, 1997). Finally, most SEPMs assume the population under study has linear temporal dynamics and therefore the relationship (parameter estimate) between the population variable and environmental variable is stable through time. However, not only do many exploited populations contain nonlinear (e.g., chaotic) dynamics (Hsieh *et al.*, 2005), but the relationships between biology

*Correspondence. e-mail: sglaser@vims.edu

Received 19 September 2012

Revised version accepted 31 May 2013

and physics can also change unexpectedly. Populations with strong density-dependent controls, in particular, cast doubt on the appropriateness of linear methods (Royer and Fromentin, 2006).

In a comprehensive review of SEPMs, Planque *et al.* (2011) argue that one constraint of species distribution models (SDM) is the assumption that the environment is the overriding factor determining distribution. To that end, SDMs usually exclude behavioral considerations such as learning, site fidelity, foraging strategies, and reproductive behaviors. Consequently, SDMs are capable of predicting the potential occurrence of a species given environmental envelopes, but they often fall short of predicting the actual occurrence of a population. A good environmental model that ignores fishing pressure will fail if the population learns to avoid heavily fished areas (Bakun, 2001) or if local depletion affects population dynamics. Likewise, a model that correctly parameterizes predator–prey interactions for some period may unexpectedly fail if predators switch in response to changes in relative prey abundance. Ultimately, the success of a SEPM hinges on how well it addresses structural uncertainty in model functional form (Charles, 1998).

The aim of this paper is to produce short-term forecasts of fishery CPUE by introducing a new application of a well-developed nonlinear forecasting model that has several advantages over traditional SEPMs. First, the model uses only time series of CPUE and does not require additional demographic data such as migration rates or mortality estimates. Secondly, the forecasting model contains only two free parameters that are user-provided; although the effects of external variables on the population are implicit in our approach (elaborated upon in the Methods section), they are not included explicitly. We contend this avoids the inclusion of incorrect environmental variables in the model. Thirdly, the model does not rely on assumptions of equilibrium, stability or linear dynamics and is not constrained by assumptions of the functional form of a population dynamics model; rather, model structure accounts for non-equilibrium and complex nonlinear dynamics that can be found in fisheries systems (Hsieh *et al.*, 2005; Glaser *et al.*, 2011; Liu *et al.*, 2012). These latter two points address an important criticism of SEPMs: a mismatch in scale of measured variables, or lack of spatially resolved ancillary data, may cause SEPMs to exacerbate error compared with non-spatial population models (Conroy *et al.*, 1995; Wennergren *et al.*, 1995; Ruckelshaus *et al.*, 1997).

Here, we apply a nonlinear forecasting model to spatially explicit CPUE data for the North Pacific

commercial albacore fishery in the California Current System (CCS). Albacore migrate annually through the CCS where they are caught by troll vessels; in 2010, the commercial albacore surface fishery based in Washington, Oregon and California was valued at US \$29.5 million, ex-vessel (PFMC, 2011). Juvenile albacore (ages 0–4, 50–90 cm fork length) migrate into CCS waters in the late spring and reside there until late fall. Some juveniles overwinter at the southern end of the CCS along the continental shelf (Childers *et al.*, 2011), whereas some migrate across the North Pacific to the Kuroshio system (Kimura *et al.*, 1997). Juvenile albacore aggregate near oceanic convergence zones characterized by thermal boundaries and high primary productivity (Laurs and Lynn, 1977; Polovina *et al.*, 2001).

MATERIALS AND METHODS

Data

Data were commercial catch (number of fish) and effort (number of boat days spent fishing) collected from voluntarily-submitted logbooks from the United States commercial troll fishery for juvenile North Pacific albacore. These data were provided to the authors by the Southwest Fisheries Science Center, National Marine Fisheries Service at a scale of $1^\circ \times 1^\circ$ spatial cells and cover the period 1966–2005 in monthly bins (ISC, 2006). Nominal annual CPUE values were calculated by summing catch and effort from June to October (the residence period of albacore in the CCS) for each spatial cell (Walters, 2003), then dividing catch by effort. Data were analyzed as normalized (mean = 0, standard deviation = 1) first differences ($\Delta x = x_{t+1} - x_t$). The scaling achieved through normalization allowed the pooling of time series (described below), and analysis on first differences addressed issues of non-stationarity. First differencing also provides a convenient interpretation of results: Δx represents the increase or decrease in CPUE relative to x_t , a metric more valuable to the fishery than a CPUE index.

We limited analysis to cells that contained at least 20 annual values, resulting in 126 unique spatial cells ranging from northern Baja, Mexico, to Vancouver Island, Canada, and extending from the shore to nearly 1000 km offshore at the widest range. All data considered, the temporal and spatial extent was substantial. However, for any given year, data coverage differed both in number of cells containing non-zero effort and the geographic extent of coverage. The commercial albacore fishery was centered on San Diego prior to the mid-1980s but,

with the relocation of canneries, the fleet has since shifted northward.

Nonlinear forecasting model (S-map)

The forecasting model described herein contains three key features: (1) the influence of external variables on time series dynamics is intrinsic to the approach; (2) model selection is done using out-of-sample forecasts and thus avoids overfitting; (3) our model accommodates time series with either linear or nonlinear dynamics and therefore strictly dominates methods that rely on assumptions of linearity.

In a dynamic system, of which fisheries are emblematic, the system state space is defined by driving variables (also called dimensions) and their relationships to each other (Packard *et al.*, 1980). For example, in a system governing albacore CPUE, dynamics may be a function of anchovy (prey) abundance (Zainuddin *et al.*, 2006; Glaser, 2011), sea surface temperature (Laurs and Lynn, 1977; Lehodey *et al.*, 2003) and fishing pressure. A point in state space is therefore defined by the ordered quadruple (albacore CPUE, anchovy, temperature, fishing). If we have many observations of these variables, we can reconstruct the full model and, theoretically, make predictions about future behavior. Unfortunately, there are many challenges to reconstructing state spaces from ecological data. For example, a mis-match of the scale of data collection between biological and physical variables is common, or the relevant variables may not have been measured. Thus, the true shape of the function defining the state space often is unobserved.

Takens (1981) offered a solution to this dilemma: the dynamics of a D -dimensional system can be reconstructed through the information contained in the lags of one time series. In our example, the influences of anchovy, temperature and fishing (or another latent driver) on albacore at time t (x_t) can be approximated by x_{t-1} , x_{t-2} and x_{t-3} . The dynamic effects of anchovy, temperature and fishing are thus contained in the historical observations of albacore CPUE. Then, vectors of CPUE with lagged coordinates (x_{t-3} , x_{t-2} , x_{t-1} , x_t) can be used to reconstruct the state space; this is known as an embedding or state space reconstruction. According to the Whitney embedding theorem (Whitney, 1936), a system of dimensionality D can be embedded in E dimensions where $D \leq E < 2D+1$. Note that there is not necessarily a 1 : 1 correspondence between E and 'real' dimensionality. The embedding dimension E describes the operational dimensionality of a system, or the minimum number of dimensions needed to obtain a certain level of model forecasting skill. This point is critical because, although external

variables such as temperature are not explicitly included in our model, the effects of external variables are modeled implicitly by the approach. In short, state space reconstruction techniques use the dynamic structure of data as a proxy for the influence of external variables (Takens, 1981).

A state space reconstruction model, sequential locally weighted global linear maps (S-map; Sugihara, 1994), was used to forecast CPUE in individual spatial cells. This nonlinear regression model has been developed in detail elsewhere (Sugihara, 1994; Dixon *et al.*, 1999; Hsieh *et al.*, 2005; Glaser *et al.*, 2011). The model uses short-term forecasts in two ways: (1) to produce estimates of future values outside the range of data availability (i.e., a true forecast) and (2) as a means of estimating model skill by comparing out-of-sample forecasts to observed data.

Time series were deconstructed into vectors of length E comprising consecutive time lags and divided into a set of library vectors $\{\mathbf{x}_{\text{lib}}\}$, from which the model was built, and prediction vectors $\{\mathbf{x}_{\text{pred}}\}$, on which the model was tested. In this application, $\{\mathbf{x}_{\text{lib}}\}$ included all spatial cells except one, and the excluded cell was $\{\mathbf{x}_{\text{pred}}\}$. We forecast, one cell at a time, each of 126 cells using data pooled from the other 125 cells. The model also withheld from the library any values contemporary with the prediction vector being analyzed. For example, for a prediction vector containing the years 2000–2003, all library vectors containing 2000–2003 were excluded from the model. This prevented overestimation of model skill through spatial autocorrelation.

Let us rename $\{\mathbf{x}_{\text{lib}}\} \equiv \{\mathbf{x}_i\}$ and $\{\mathbf{x}_{\text{pred}}\} \equiv \{\mathbf{x}_t\}$. First, a time series of length T was decomposed into a set of vectors of length E ; the coordinates of these vectors defined the lagged-coordinate embedding. Secondly, for a given time step t , the prediction vectors \mathbf{x}_t were defined as $\langle x_t, x_{t-1}, \dots, x_{t-E+1} \rangle$. The Euclidean distances between \mathbf{x}_t and $\{\mathbf{x}_i\}$ were calculated. Thirdly, \mathbf{x}_t was projected ahead in coordinate space t_p time steps (here, $t_p = 1$). The forward trajectories of the vectors in $\{\mathbf{x}_i\}$ were defined by the points $\{y_i\}$, and the forecast (\hat{y}_t) for \mathbf{x}_t was calculated as an exponentially weighted function of $\{y_i\}$ (equation below). The process was repeated for all t , ultimately producing $T - E - t_p + 1$ forecasts for a given cell. In addition to forecasting cells with observed data, we produced forecasts for 2006 for all cells that contained data during 2002–2005.

The S-map model includes a nonlinear tuning parameter, θ . When $\theta = 0$, the S-map weights all library vectors equally and the forecast represents output from a global linear model [an autoregressive

model of order E , AR(E)] with identical coefficients defining the projections for all \mathbf{x}_t in one time series. When θ is tuned above 0, the library vectors neighboring \mathbf{x}_t are more highly weighted for forecasting and coefficients are recalculated for each \mathbf{x}_t in the time series. If the global linear model outperforms the localized nonlinear model, forecasts are best approximated by all available information (similar to using an average value). If the local nonlinear model outperforms, the system is state-dependent and library vectors nearby \mathbf{x}_t in state space at time t are expected to diverge over time, losing their predictive value. Thus, S-map measures the tendency for nearby trajectories in an attractor to diverge, a characteristic of nonlinear systems.

Specifically, again defining the projections for library vectors $\{\mathbf{x}_i\}$ as $\{y_i\}$ and setting $x_t(0) \equiv 1$, the forecast \hat{y}_t for the prediction vector \mathbf{x}_t is calculated as

$$\hat{y}_t = \sum_{j=0}^E c_t(j)x_t(j) \quad (1)$$

c is solved by singular value decomposition as $\mathbf{b} = \mathbf{A}c$, and

$$b(i) = w(\|\mathbf{x}_i - \mathbf{x}_t\|)y_i \quad (2)$$

$$A(i, j) = w(\|\mathbf{x}_i - \mathbf{x}_t\|)x_i(j) \quad (3)$$

For the matrix \mathbf{A} , j corresponds to the E time lags in each library vector. The weighting function

$$w(d) = e^{-\frac{\theta d}{\bar{d}}} \quad (4)$$

defines the model as global linear or local nonlinear, where d is the Euclidean distance between one library vector \mathbf{x}_i and the prediction vector \mathbf{x}_t , and \bar{d} is the average distance between all $\{\mathbf{x}_i\}$ and \mathbf{x}_t . We begin with $\theta = 0$ and iterate the model for increasing values of θ . The θ values explored here were $\{0, 0.001, 0.005, 0.01, 0.025, 0.05, 0.075, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1, 2, 5, 7, 10\}$.

Model forecast skill (ρ , the Pearson correlation between observed values and forecasts) and model error (mean absolute error, MAE) were used for model selection based on tuning θ . If ρ was highest and MAE lowest for $\theta = 0$, the dynamics in the time series can be classified as linear. If ρ was highest and MAE lowest for $\theta > 0$, the dynamics in the time series can be classified as nonlinear. We therefore tested the null hypothesis that our data were best described by a linear AR model.

We have thus far neglected the important choice of system dimensionality, E , which is an input to the S-map model (Sugihara and May, 1990). E defines the number of consecutive time lags to include in library vectors; it also relates to the complexity of the system

and the number of exogenous variables, as described above. Previous analysis of these data (Glaser *et al.*, 2011) found $E = 4$ and $E = 5$ to best describe the data. We chose $E = 4$ to maximize the number of forecasts possible.

Spatial null models

Our measures of model skill, ρ and MAE, describe how well the S-map model performs given a null hypothesis of no correlation between observations and model forecasts. We used the improvement in MAE as the model is tuned above $\theta = 0$ to test the hypothesis that S-map forecasts outperform an equivalent autoregressive model. To determine whether forecasts merely reflected strong spatial pattern or temporal autocorrelation in the data, we generated two null models based on MAE (which is not sensitive to sample size like ρ). Our null models tested the following hypotheses: (1) forecasts reflect long-term geographic affinity for a given cell and (2) forecasts reflect short-term autocorrelation within a cell. We calculated the test statistic MAE using the real data, computed the distribution of MAE under the null hypotheses, and compared the test statistic to the null distribution to obtain a Monte Carlo P-value.

To examine the first hypothesis, we tested whether the S-map forecast model outperformed a model of mean CPUE in a given cell, addressing the possibility that spatial patterns in CPUE remain constant in time. The MAE test statistic was

$$\text{MAE}_T = \frac{\sum_i \sum_j |x_{ij,\text{forecast}} - x_{ij,\text{observed}}|}{N} \quad (5)$$

from forecasts and observations of CPUE in cell i and year j , given N pairs of forecasts and observations. The null model was calculated as

$$\frac{\text{MAE}_{0,1} \sum_i \sum_j |x_{i,\text{mean}} - x_{ij,\text{observed}}|}{N} \quad (6)$$

The null distribution was generated using a spatial bootstrap in which cells were randomly replaced and MAE recalculated (500 bootstraps with 10% replacement). For this null model, CPUE values were converted from normalized first differences (on which forecasts were made) to raw (nominal) values; such back-transformation was necessary because the normalized values had, by definition, a mean of 0.

The second null model tested whether the S-map forecast model outperformed a model of CPUE values observed the prior year in a given cell, addressing short-term temporal stationarity. The test statistic was

the same as that calculated in the first test above. The null model was

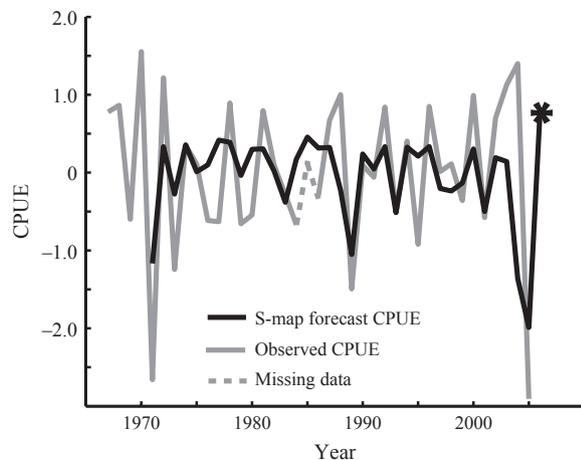
$$\text{MAE}_{0,2} = \frac{\sum_i \sum_j |x_{ij-1, \text{observed}} - x_{ij, \text{observed}}|}{N} \quad (7)$$

The null distribution was generated from 500 spatial bootstraps as in null model 1. For both tests, the P -value was the fraction of bootstrap samples for which the S-map MAE was higher.

RESULTS

Forecasts were produced for 1971–2006: forecasts for 1967–1970 were not possible because the model requires 4 yr (for $E = 4$) at the beginning of the time series to create vectors. Altogether, 1884 forecasts of CPUE were produced, and overall model forecast skill was highly significant ($\rho = 0.57$, $P < 0.00001$). Figure 1 illustrates a forecasted time series from one spatial cell, chosen for both its complete temporal coverage and average forecast skill ($\rho = 0.58$). For any given year, the number of spatial forecasts varied with the availability of CPUE data (minimum 31 cells yr^{-1} , maximum 84 cells yr^{-1} , median 53 cells yr^{-1}). Although cells were screened to contain at least 20 yr of data, not all cells contained 20 yr of consecutive data. Using $E = 4$, library and forecast vectors required at least 5 yr of consecutive data points (four points for the model vector, and one for the observation of the

Figure 1. Catch per unit effort time series (first-differenced and normalized) for $1^\circ \times 1^\circ$ cell at 45°N , 127°W . The S-map forecast (black line) was significantly correlated with observed data (gray line; $r = 0.58$, $P < 0.001$). The asterisk denotes a future forecast for 2006. The dashed gray line is a hypothetical example illustrating a forecast into a year of missing data.



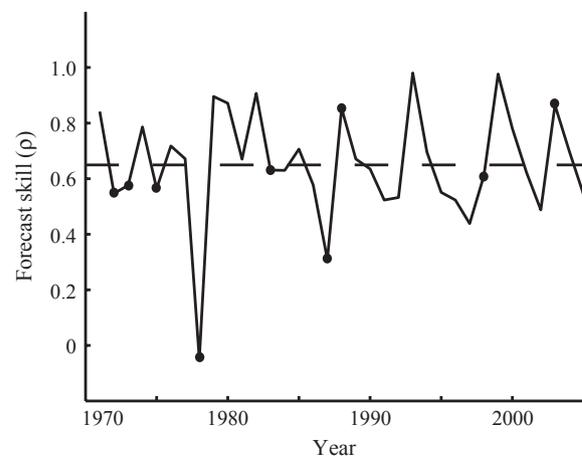
point to be forecast). For any given cell, the number of annual forecasts (of 35 possible yr) was also highly variable (minimum 1 yr cell^{-1} , maximum 35 yr cell^{-1} , median 11 yr cell^{-1}). For the statistics that follow, we limit reported results to cells that contained at least six forecasts (106 cells).

Forecast skill at the annual scale varied significantly, and 9 of 35 yr did not produce forecasts significantly related to observations (Fig. 2). The spatial forecast skill achieved was generally high (Fig. 3): 51 of 106 cells were modeled individually at $P < 0.05$, and 71 cells were significant at $P < 0.1$. Most cells (95 of 106) were best modeled using $\theta > 0$ (Table 1, average $\theta = 3.63$), consistent with previous findings that albacore CPUE data contain a weak nonlinear dynamic signal (Glaser *et al.*, 2011). In all cases, the nonlinear forecasting model outperformed the spatial null models (Table 2). Finally, the spatial model built with S-map produced a map of forecasts for 2006 (Fig. 4).

DISCUSSION

We demonstrate that spatially explicit CPUE data for juvenile North Pacific albacore in the California Current System can be forecast 1 yr ahead with significant skill using a low-dimensional nonlinear model. Importantly, this model does not require the addition of demographic data or environmental correlates and illustrates a viable approach for fisheries that lack auxiliary data. The potential uses for short-term forecasts of spatially explicit CPUE data are numerous

Figure 2. Model-achieved forecast skill in a given year over 126 spatial cells (black line). Dashed line shows average forecast skill over all years. Circles denote years that did not produce statistically significant forecasts ($P > 0.05$), mostly due to small sample sizes.



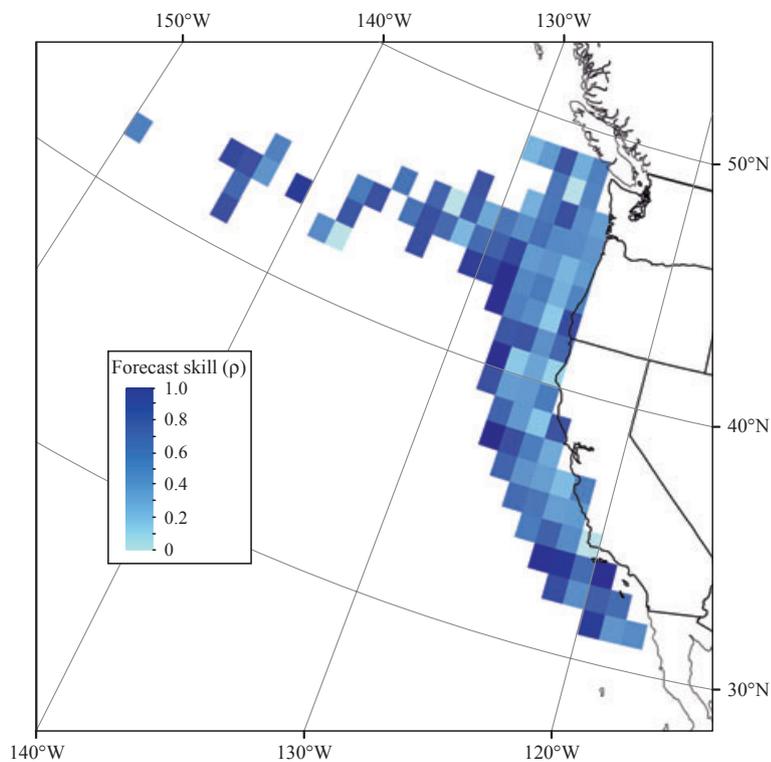


Figure 3. Model forecast skill (ρ) for North Pacific albacore CPUE during 1967–2005 for 126 spatial cells. ρ is the correlation coefficient between observed and forecast CPUE time series in a given cell.

Table 1. Distribution of θ values (nonlinear tuning parameter) corresponding to lowest model error (MAE) for 106 spatial cells.

θ	No. of cells	Proportion
0	11	0.10
0.2	2	0.02
0.4	4	0.04
0.5	1	0.01
0.75	4	0.04
1.0	7	0.07
2.0	32	0.30
5.0	23	0.22
7.0	9	0.08
10.0	13	0.12

(Babcock *et al.*, 2005). First, changes in distribution of CPUE could inform evaluation of potential habitat and climate change impacts (Zainuddin *et al.*, 2006; Singh and Milner-Gulland, 2011). Secondly, identification of CPUE hotspots can guide efforts to reduce species interactions; in the case of marine fisheries, managers could reduce by-catch of non-targeted species (Hobday and Hartmann, 2006). Thirdly, spatial forecasts provide an objective way to incorporate flexible boundaries into nature reserves or provide guidance for other forms of spatial resource management

Table 2. Comparison of the S-map forecasting model to two spatial null models.

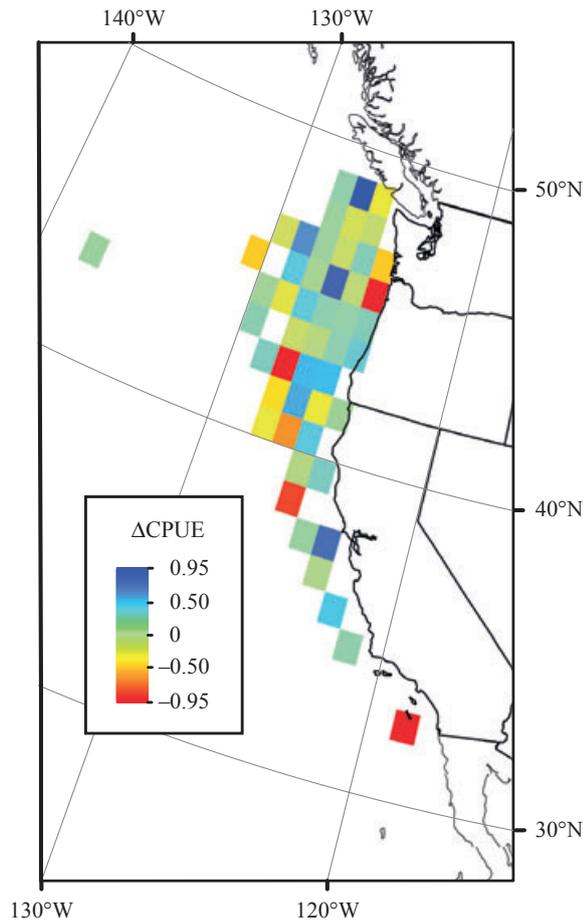
Null model	Test MAE	Null MAE	<i>P</i> -value
1. Mean cell CPUE	36.04	$37.6 \pm 0.56^\dagger$	<0.005
2. CPUE at time $t-1$	0.80	$1.2 \pm 0.02^\dagger$	<0.001

[†]Null distribution generated by spatial bootstraps in which cells were replaced (10%) and resampled 500 times. Null statistic is the mean \pm SD bootstrap MAE.

(Myers *et al.*, 2000; Apostolaki *et al.*, 2002; Singh and Milner-Gulland, 2011). Fourthly, most current generation stock assessment models produce both short-term and long-term forecasts of stock abundance or recruitment. These methods provide alternative short-term forecasts that can be compared directly to those produced by more traditional models (Booth, 2000; Brander, 2003). Fifthly, the flexibility of S-map to forecast missing years in spatial cells that contain some data may help overcome one of Walters' (2003) 'catch rate fantasies': ignoring unfished strata (elaborated below). Finally, industry should be interested in short-term forecasts to target fishing effort and potentially reduce costs.

Our low data approach avoids error propagation inherent to many SEPMs (Doak and Mills, 1994;

Figure 4. Forecasts of Δ CPUE for 2006 made from time series spanning 1967–2005 for North Pacific albacore caught by the U.S. commercial surface fleet. Δ CPUE is the relative change in CPUE from the prior year.



Conroy *et al.*, 1995; Beissinger and Westphal, 1998). Depending on the complexity of the model (i.e., the number of estimated parameters) and the quality of data available, error propagation may pose a significant challenge (Doak and Mills, 1994; Ruckelshaus *et al.*, 1997, 1999; but see Mooij and DeAngelis, 1999, 2003). Specifically, uncertainty in three types of parameterizations (misclassification of habitat, dispersal or migration rates, and survival rates) produces significant forecast error. The lack of data available for mapping habitat patches is considered another serious shortcoming for the validity of SEPMs, and these problems are magnified when habitat preferences are also unknown (Wennergren *et al.*, 1995). One solution is to construct models that require less environmental and life history data (Beissinger and Westphal, 1998), something our model achieves. The most accurate SEPMs will contain few parameters, matching model

complexity to data availability (Ruckelshaus *et al.*, 1997). Although methods for decreasing error in SEPM models exist (Wiegand *et al.*, 2004), they retain a need for numerous parameters and may be difficult for populations that are not easily observed.

Used in conjunction, Figures 2 and 3 provide context for assigning credibility, or some measure of uncertainty, to model skill for given years and cells, respectively. This provides guidance in selecting regions for which forecasts are most reliable, and may inform future sampling protocols to reduce uncertainty about forecasts for a given region. Temporally, if a year or set of years is poorly modeled, one might give less weight to those periods when linking CPUE to environmental variables. Ultimately, the goal should be to understand when and where estimates of catch rates provide a strong dynamic signal over the underlying observation noise. If noise swamps the dynamic signal (Sugihara, 1994), model skill will be correspondingly low and forecasts unreliable. While some years and cells were not modeled in a way that produced statistically significant forecasts, a coherent dynamic signal was detected in the vast majority of cells.

Null models addressed alternatives that would invalidate both the usefulness of this approach at producing forecasts and detection of a signal in the data that can be attributed to a dynamic mechanism (Gotelli and Graves, 1996). The first null model addressed the possibility that forecasts are trivially a function of spatial consistency in patterns of CPUE; that is, if a given cell is regularly a hotspot for fishing, either due to behavior of albacore or behavior of fishers, the ability to forecast that individual cell becomes less meaningful when interpreted as an expression of the dynamics of the fishery. Such an outcome is possible: Kirchner and McAllister (2001) found spatial structure of orange roughy CPUE was more accurately predicted by mean CPUE values than by environmental variables. In such a case, fishers would simply return to the same cell year after year. We demonstrate pattern in these data not explained by such spatial affinity.

The second null model addressed the question of short-term stationarity in the time series at the scale of a cell. CPUE data are known to contain short-term autocorrelation due to biological patterns (e.g., consistency in migration routes and habitat preference) or socio-economic patterns (e.g., the tendency for the same boats to fish in consecutive years). If CPUE in year t was best predicted by CPUE at time $t - 1$, the forecast model would not outperform the null model. The rejection of null model two suggests the patterns detected in the dynamic forecasting model are not simply a result of stationary short-term patterns.

We now highlight several unique aspects of the S-map model that distinguish this application from some other spatial regression models. A key advantage of S-map is the flexibility achieved as a phenomenological model, tuning from linear to nonlinear with only one continuous parameter (θ). The distinction is not trivial: in a system with linear dynamics, a linear model should produce skilful predictions far into the future. In a system with nonlinear dynamics, a nonlinear model will produce skilful predictions over short time horizons, but a linear model will forecast poorly (Sugihara, 1994). Because S-map is essentially a series of progressively more nonlinear models nested in an E -order autoregressive model, it applies to systems with both linear and nonlinear dynamics.

Another advantage of state space reconstruction is the leverage gained from pooling data sets, thereby overcoming difficulties posed by short and incomplete time series without the need to manufacture data (Walters, 2003; Hsieh *et al.*, 2008; Glaser *et al.*, 2011). The development of models that can accommodate incomplete data sets has long been a challenge (Doak and Mills, 1994). If gaps in time series are filled via linear interpolation or with mean values, the resulting dynamics are more likely to reflect a linear process, regardless of whether that process is correct. S-map searches simultaneously over library vectors from all available cells to produce forecasts for the withheld cell, and can produce forecasts for years in the withheld cell that lack data (see Fig. 1 for a hypothetical example). S-map disregards library vectors that contain missing data but can forecast missing data because the full library of vectors from other cells manifest the same dynamics (Glaser *et al.*, 2011). To accomplish this, the model does not require modification of the methods outlined herein: forecasts are automatically produced for all prediction vectors of length E . In the case of a missing value in the time period being forecast, the user substitutes a NaN (not a number) value.

Ideally, one would construct a spatial model that produces skillful forecasts of CPUE, contains few parameters so error propagation is minimized, and provides mechanistic understanding of the ecological processes governing system dynamics. We have achieved the first two goals, and extending S-map to a multivariate exploration of fishery state spaces is an important next step (Deyle and Sugihara, 2011). The addition of environmental or biological variables to state space reconstruction [rather than the univariate approach outlined by Takens (1981)] should improve model forecast skill (Dixon *et al.*, 1999) and aid in development and validation of a mechanistic model to

improve forecasts of albacore CPUE in the North Pacific (Zainuddin *et al.*, 2006).

ACKNOWLEDGEMENTS

We sincerely thank the Southwest Fisheries Science Center, especially Paul Crone, Suzy Kohin, John Childers and Al Coan, for providing all data and guidance on its use. We thank M. Maunder, A. MacCall, M. Fogarty, L. Kaufman, E. Deyle, C. Hsieh, C. Hendrix, A. Rosenberg, H. Liu, B. Fissel, M. Cape, S. Martin, B. Carr, E. Klein, L. Storch and I. Altman for discussion of this modeling approach. Finally, the study would not be possible without the cooperation of the U.S. commercial albacore fleet and their willingness to report catch and effort data through logbooks. Funding was provided through NA08OAR4320894 and NA09NMF4720177, grants from the Comparative Analysis of Marine Ecosystem Organization (CAMEO) program, a partnership between the US National Science Foundation and NOAA National Marine Fisheries Service.

REFERENCES

- Apostolaki, P., Milner-Gulland, E.J., McAllister, M.K. and Kirkwood, G.P. (2002) Modelling the effects of establishing a marine reserve for mobile fish species. *Can. J. Fish. Aquat. Sci.* **59**:405–415.
- Babcock, E.A., Pikitch, E.K., McAllister, M.K., Apostolaki, P. and Santora, C. (2005) A perspective on the use of spatialized indicators for ecosystem-based fishery management through spatial zoning. *ICES J. Mar. Sci.* **62**:469–476.
- Bakun, A. (2001) 'School-mix feedback': a different way to think about low frequency variability in large mobile fish populations. *Prog. Oceanogr.* **49**:485–511.
- Beissinger, S.R. and Westphal, M.I. (1998) On the use of demographic models of population viability in endangered species management. *J. Wildl. Manage.* **62**:821–841.
- Booth, A.J. (2000) Incorporating the spatial component of fisheries data into stock assessment models. *ICES J. Mar. Sci.* **57**:858–865.
- Brander, K. (2003) What kinds of fish stock predictions do we need and what kinds of information will help us make better predictions? *Sci. Mar.* **67**(S1):21–33.
- Charles, A.T. (1998) Living with uncertainty in fisheries: analytical methods, management priorities and the Canadian groundfishery experience. *Fish. Res.* **37**:37–50.
- Childers, J., Snyder, S. and Kohin, S. (2011) Migration and behavior of juvenile North Pacific albacore (*Thunnus alalunga*). *Fish. Oceanogr.* **20**:157–173.
- Conroy, M.J., Cohen, Y., James, F.C., Matsinos, Y.G. and Maurer, B.A. (1995) Parameter-estimation, reliability, and model improvement for spatially explicit models of animal populations. *Ecol. Appl.* **5**:17–19.
- Deyle, E.R. and Sugihara, G. (2011) Generalized theorems for nonlinear state space reconstruction. *PLoS ONE* **6**:e18295.
- Dixon, P.A., Milicich, M.J. and Sugihara, G. (1999) Episodic fluctuations in larval supply. *Science* **283**:1528–1530.

- Doak, D.F. and Mills, L.S. (1994) A useful role for theory in conservation. *Ecology* **75**:615–626.
- Dunning, J.B., Danielson, B.J. and Pulliam, H.R. (1992) Ecological processes that affect populations in complex landscapes. *Oikos* **65**:169–175.
- Fahrig, L. and Merriam, G. (1994) Conservation of fragmented populations. *Conserv. Biol.* **8**:50–59.
- Glaser, S.M. (2011) Do albacore exert top-down pressure on northern anchovy? Estimating anchovy mortality as a result of predation by juvenile North Pacific albacore in the California current system. *Fish. Oceanogr.* **20**:242–257.
- Glaser, S.M., Ye, H., Maunder, M.N., MacCall, A.D., Fogarty, M.J. and Sugihara, G. (2011) Detecting and forecasting complex nonlinear dynamics in spatially structured catch-per-unit-effort time series for North Pacific albacore (*Thunnus alalunga*). *Can. J. Fish. Aquat. Sci.* **68**:400–412.
- Gotelli, N.J. and Graves, G.R. (1996) Null Models in Ecology. Washington, DC: Smithsonian Institution Press.
- Hobday, A.J. and Hartmann, K. (2006) Near real-time spatial management based on habitat predictions for a longline bycatch species. *Fish. Manag. Ecol.* **13**:365–380.
- Hsieh, C.H., Glaser, S.M., Lucas, A.J. and Sugihara, G. (2005) Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean. *Nature* **435**:336–340.
- Hsieh, C.H., Anderson, C.N.K. and Sugihara, G. (2008) Extending nonlinear analysis to short ecological time series. *Am. Nat.* **171**:71–80.
- ISC (2006) Report of the ISC – Albacore Working Group Stock Assessment Workshop. Shizuoka-Shi, Japan: National Research Institute of Far Seas Fisheries 76 pp.
- Kimura, S., Munenori, N. and Sugimoto, T. (1997) Migrations of albacore, *Thunnus alalunga*, in the North Pacific Ocean in relation to large oceanic phenomena. *Fish. Oceanogr.* **6**:51–57.
- Kirchner, C.H. and McAllister, M.K. (2001) A new improved method to compute swept area estimates of biomass from commercial catch rate data: application to Namibian orange roughy (*Hoplostethus atlanticus*). *Fish. Res.* **56**:69–88.
- Laur, R.M. and Lynn, R.J. (1977) Seasonal migration of North Pacific albacore, *Thunnus alalunga*, into North American coastal waters – distribution, relative abundance, and association with Transition Zone waters. *Fish. B-NOAA* **75**:795–822.
- Lehodey, P., Chai, F. and Hampton, J. (2003) Modelling climate-related variability of tuna populations from a coupled ocean-biogeochemical-populations dynamics model. *Fish. Oceanogr.* **12**:483–494.
- Liu, H., Fogarty, M.J., Glaser, S.M. et al. (2012) Nonlinear dynamic features and co-predictability of the Georges Bank fish community. *Mar. Ecol. Prog. Ser.* **464**:195–207.
- Moij, W.M. and DeAngelis, D.L. (1999) Error propagation in spatially explicit population models: a reassessment. *Conserv. Biol.* **13**:930–933.
- Moij, W.M. and DeAngelis, D.L. (2003) Uncertainty in spatially explicit animal dispersal models. *Ecol. Appl.* **13**:794–805.
- Myers, N., Mittlemeier, R.A., Mittlemeier, C.G., da Fonseca, G.A.B. and Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature* **403**:853–858.
- Packard, N.H., Crutchfield, J.P., Farmer, J.D. and Shaw, R.S. (1980) Geometry from a time series. *Phys. Rev. Lett.* **45**:712–716.
- Pelletier, D., Leleu, K., Mou-Tham, G., Guillemot, N. and Chabanet, P. (2011) Comparison of visual census and high definition video transects for monitoring coral reef fish assemblages. *Fish. Res.* **107**:84–93.
- PFMC. (2011) Status of the U.S. West Coast Fisheries for Highly Migratory Species through 2010. Stock Assessment and Fishery Evaluation. Portland: Pacific Fishery Management Council, 164 pp.
- Planque, B., Loots, C., Petitgas, P., Lindstrom, U. and Vaz, S. (2011) Understanding what controls the spatial distribution of fish populations using a multi-model approach. *Fish. Oceanogr.* **20**:1–17.
- Polovina, J.J., Howell, E., Kobayashi, D.R. and Seki, M.P. (2001) The transition zone chlorophyll front, a dynamic global feature defining migration and forage habitat for marine resources. *Prog. Oceanogr.* **49**:469–483.
- Rose, G.A. and Leggett, W.C. (1990) The importance of scale to predator-prey spatial correlations – an example of Atlantic fishes. *Ecology* **71**:33–43.
- Royer, F. and Fromentin, J.M. (2006) Recurrent and density-dependent patterns in long-term fluctuations of Atlantic bluefin tuna trap catches. *Mar. Ecol. Prog. Ser.* **319**:237–249.
- Ruckelshaus, M., Hartway, C. and Kareiva, P. (1997) Assessing the data requirements of spatially explicit dispersal models. *Conserv. Biol.* **11**:1298–1306.
- Ruckelshaus, M., Hartway, C. and Kareiva, P. (1999) Dispersal and landscape errors in spatially explicit population models: a reply. *Conserv. Biol.* **13**:1223–1224.
- Sale, P.F. (1977) Maintenance of high diversity in coral-reef fish communities. *Am. Nat.* **111**:337–359.
- Singh, N.J. and Milner-Gulland, E.J. (2011) Conserving a moving target: planning protection for a migratory species as its distribution changes. *J. Appl. Ecol.* **48**:35–46.
- Sugihara, G. (1994) Nonlinear forecasting for the classification of natural time-series. *Philos. Trans. R. Soc. London* **348**:477–495.
- Sugihara, G. and May, R.M. (1990) Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature* **344**:734–741.
- Takens, F. (1981) Detecting strange attractors in turbulence in Dynamical Systems and Turbulence. Springer-Verlag.
- Vignaux, M. (1996) Analysis of spatial structure in fish distribution using commercial catch and effort data from the New Zealand hoki fishery. *Can. J. Fish. Aquat. Sci.* **53**:963–973.
- Walters, C. (2003) Folly and fantasy in the analysis of spatial catch rate data. *Can. J. Fish. Aquat. Sci.* **60**:1433–1436.
- Wennergren, U., Ruckelshaus, M. and Kareiva, P. (1995) The promise and limitations of spatial models in conservation biology. *Oikos* **74**:349–356.
- Werner, E.E., Mittelbach, G.G., Hall, D.J. and Gilliam, J.F. (1983) Experimental tests of optimal habitat use in fish – the role of relative habitat profitability. *Ecology* **64**:1525–1539.
- Whitney, H. (1936) Differentiable manifolds. *Ann. Math.* **37**:645–680.
- Wiegand, T., Revilla, E. and Knauer, F. (2004) Dealing with uncertainty in spatially explicit population models. *Biodivers. Conserv.* **13**:53–78.
- Wiens, J.A., Stenseth, N.C., Vanhorne, B. and Ims, R.A. (1993) Ecological mechanisms and landscape ecology. *Oikos* **66**:369–380.
- Zainuddin, M., Kiyofuji, H., Saitoh, K. and Saitoh, S.I. (2006) Using multi-sensor satellite remote sensing and catch data to detect ocean hot spots for albacore (*Thunnus alalunga*) in the northwestern North Pacific. *Deep-Sea Res. II* **53**:419–431.