Integrating Time-series of Community Monitoring Data

CHRISTY V. PATTENGILL-SEMMENS[2]*, BRICE X. SEMMENS[1], ELIZABETH E. HOLMES[1], ERIC J. WARD[1], and BEN I. RUTTENBERG[3]
[1]*NOAA Northwest Fisheries Science Center, 2725 Montlake Blvd. East , Seattle, Washington 98112-2097 USA.*
[2]*Reef Environmental Education Foundation (REEF), Key Largo, Florida, USA. *Christy@reef.org.*
[3]*NOAA Southeast Fisheries Science Center, 75 Virgina Beach Dr., Room 207 Miami, Florida 33149-1003 USA.*

ABSTRACT

Assessing population trends, evaluating management actions, and identifying community responses to anthropogenic impacts all require an accurate time-series of populations. In practice, such data are often scarce or of varying quality due to the limited resources of managing agencies. In such situations, analyses that integrating multiple data sources (e.g. agency monitoring programs, citizen science observations, fisheries catch records) can yield dramatic improvements in the estimation of population trajectories. To do so effectively, however, such integrative models must account for differences in observation errors across data sources. We used multivariate state space models (MSSMs) to assess the population trajectories of reef fish species from the Florida Keys National Marine Sanctuary based on data from 1) point count surveys conducted through academic institutions and 2) citizen-science monitoring surveys conducted by volunteer Scuba divers. By developing competing models and applying information theory, we demonstrate how MSSMs can be used to compare and integrate multiple monitoring time series, and ultimately improve estimates of the true states of populations though time. Additionally, we demonstrate that by combining multiple time series, it is possible to recover method-specific observation error estimates even for very short time series of data.

KEY WORDS: State space model, volunteer, REEF, coral reef fish, Florida Keys

Integración de Una Serie de Datos de un Muestreo a Nivel de la Comunidad

PALABRAS CLAVE: Integración, peces de arrecife, REEF, Cayos de Florida

Série de Temps de Intégration de la Communauté Surveillant des Données

MOTS CLÉS: Temps de intégracion, REEF, Florida Keys, communauté surveillant

OVERVIEW

The goal of this project was to develop an analysis technique that produces a more accurate estimate of fish population trends by simultaneously evaluating data collected from multiple monitoring programs. By merging multiple observation time series, we can improve our inference regarding the status and trends of reef fishes. We used a Multivariate Auto-regressive State Space (MARSS) Model to model changes in the abundance of reef fish species as stochastic processes through time, based on observations from both the Reef Environmental Education Foundation (REEF) Fish Survey Project and the Southeast Fisheries Science Center Reef Fish Long-term Monitoring Program.

All time-series monitoring data are susceptible to two sources of error: observation error and process error (population stochasticity). The variability due to observation error is difficult to estimate because it arises from changes in detectability and is a complex (often unknown) function of the environment. MARSS modeling provides a method for analyzing time-series data with both process and observation error. Additionally, these models provide a framework for dealing with multiple observation time series and missing observations. MARSS models simultaneously estimate observation error for multiple data sets as well as the process error associated with each species. These models generate estimates of the "true" unobserved abundances through time based on the observed data and provides quantitative estimates of observation error that can provide guidance for the distribution of monitoring efforts in the future (Figure 1).

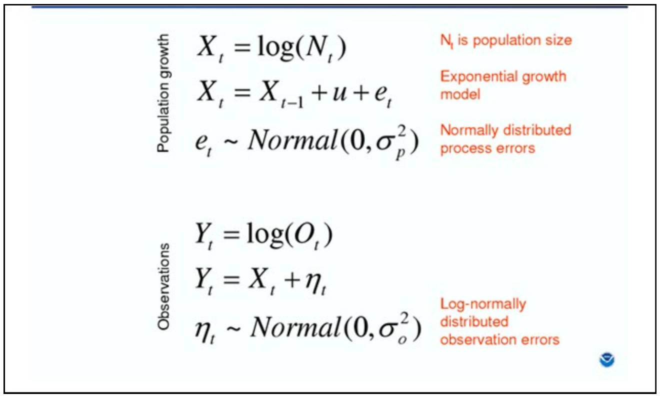


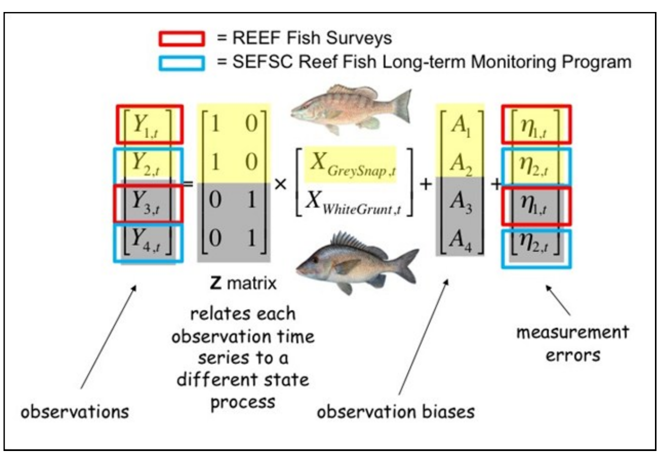
Figure 1. A simple exponential growth model with both process and observation error.

METHODS

As a proof of concept, we used data collected over 17 years at Molasses Reef (Florida Keys, US) by two fish community monitoring programs – the Reef Environmental Education Foundation (REEF) Fish Survey Project and the Southeast Fisheries Science Center Long-term Fish Monitoring Program. This involved simultaneously modeling trends in abundance of 10 species as an auto-regressive stochastic process. The model assumed that process errors are unique to each species and observation errors are unique to each survey method (but shared across species) (Figure 2). In other words, it was assumed that both survey methods are reporting information on the same unobserved states through time. All analyses were carried out in the R open source software environment using the MARSS Package (freely available on CRAN; Holmes, E.E., E.J. Ward. 2010. Analysis of multivariate time-series using the MARSS package. CRAN.)

The REEF Fish Survey Project coordinates volunteer divers and snorkelers to collect fish sighting information during recreational dives. The project started in 1993 in Florida and has expanded to include all coastal areas of North and Central America, the Caribbean, Hawaii, and the South Pacific. REEF volunteers use the Roving Diver Technique (RDT; Schmitt and Sullivan 1996). The method is a non-point survey and allows the surveyors swim freely around the dive site. The diver record all fish species positively identified and estimates categorical abundance for each species: Single (S) -1; Few (F) - 2 - 10; Many (M) - 11 - 100; Abundant (A) - > 100. All species data, along with survey time, depth, temperature and other environmental information are transferred to REEF via online data entry. Quality control programs are run prior to entry into REEF's database.

The Reef Visual Count (RVC) method is a modified version of the stationary point count method (Bohnsack and Bannerot 1986). Data are collected as part of the Southeast Fisheries Science Center Reef Fish Long-term Monitoring Program. Divers sample 15 m diameter circular plots, recording all fish observed. The cylinder extends from bottom to the limits of vertical visibility (usually the surface). The diver lists all species within the field of view. When no new species are noted, divers rotate in one direction for five minutes. Several complete rotations are usually made for each plot. After the initial five minutes, data are collected on the abundance and minimum, mean, and maximum lengths for each species sighted. Depth, bottom substrate composition, estimated benthic percentage cover, and vertical relief characteristics of the seafloor are also recorded for each plot.



$$\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \\ Y_{3,t} \\ Y_{4,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} X_{GreySnap,t} \\ X_{WhiteGrunt,t} \end{bmatrix} + \begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{bmatrix} + \begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \eta_{1,t} \\ \eta_{2,t} \end{bmatrix}$$

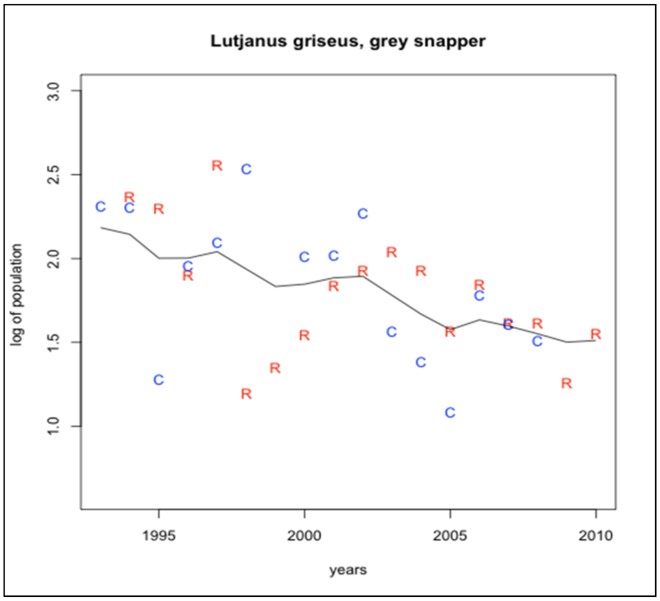
Figure 2. A MARSS model in matrix form, showing an example of two species and two data sources.

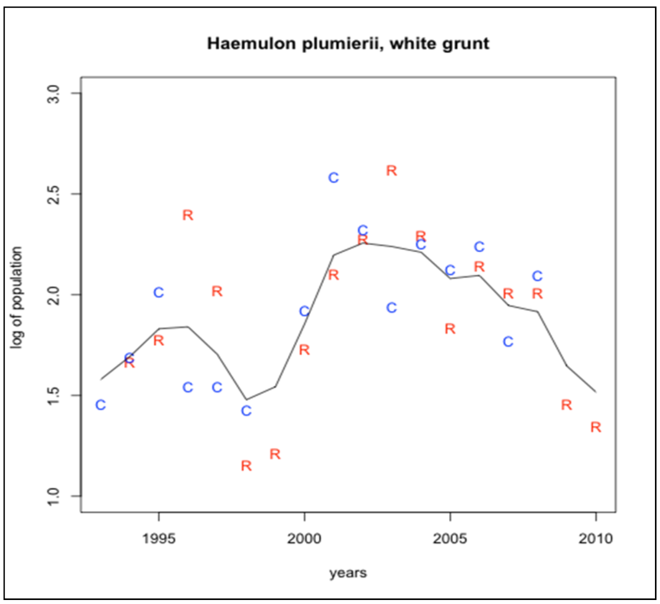
OVERVIEW OF RESULTS

Monitoring data from 1993 - 2010 (REEF) and 1993 - 2008 (RVC) were used in the model. A total of 1,741 REEF surveys and 227 RVC surveys were included. We simultaneously modeled trends in abundance for each of 10 species on Molasses Reef (Florida Keys, US) as an autoregressive stochastic process. Species included in the analysis were: sergeant major, yellowtail snapper, yellowtail damselfish, rock beauty, bar jack, hogfish, redband parrotfish, white grunt, graysby, and gray snapper. The trend for each species was modeled using the average annual abundance score based on REEF surveys and the average annual density based on RVC surveys (log10 + 1 transformed). The model results were plotted to show a population trend that incorporates both observation and process errors. Two representative species plots are given below (Figures 3a and 3b).

Observation errors for both the REEF Fish Survey data and the Southeast Fisheries Science Center Reef Fish Long-term Monitoring Program were similar (variance of 0.07 and 0.06, respectively). The slightly larger observation error associated with the REEF data likely stems both from the method and the variable level of effort across years based on volunteer effort. Nonetheless, the approximate equivalency of the estimated observation errors between two methods suggests that the REEF program continues to yield a valuable observation time series of information separate from but comparable to well designed long term monitoring programs.

On average, species process variance was 0.01 with the a high of 0.05 (white grunt, Figure 3b), suggesting the error in the time series data from both methods was dominated by uncertainty in observations. For this reason, the multivariate approach employed here affords a considerable advantage in estimating the true states of fish populations through time – in other words, by combining two separate but corrupted time series data sets, we can accurately estimate observation errors and simultaneously recover dramatically improved estimates of reef fish population trajectories.





Figures 3a and 3b. Multivariate model results for 2 species of reef fish on Molasses Reef (FL), based on two fish monitoring programs. 'R' is average annual abundance score based on REEF surveys. 'C' is average annual density based on RVC surveys (log10+1 transformed). The black line represents the model results, showing an estimated population trend that incorporates both observation and process errors.

LITERATURE CITED

Bohnsack, J.A. and S.P. Bannerot. 1986. A stationary visual census technique for quantitatively assessing community structure of coral reef fishes. U.S. Dept. of Commerce, NOAA Technical Report NMFS 41:1-15.

Holmes, E.E. and E.J. Ward. 2010. Analysis of multivariate time-series using the MARSS package. CRAN.

Schmitt, R.F. and K.M. Sullivan. 1996. Analysis of a volunteer method for collecting fish presence and abundance data in the Florida Keys. *Bulletin of Marine Science* **59**(2):404-416.