Re-analysis of a salp population time-series

Kelly R. Sutherland[1,2,*], Andrew R. Beet[1], Andrew R. Solow[1]

[1]**Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA**

[2]*Present address:* **California Institute of Technology, Pasadena, California 91125, USA**

ABSTRACT: Weekly abundances of 2 salp populations (*Salpa fusiformis* and *Thalia democratica*) were determined for the period from 1967 to 1990 at a coastal station in the Western Mediterranean using a discrete abundance scale. A previous analysis modeled these data as a Markov chain with ordered states and found a significant effect of stratification on the Markov transition probabilities. An alternative analysis combining an underlying population model and a model of the observation process found qualitatively similar results, with population growth rate significantly related to stratification. However, by explicitly separating the population model from the observation model, the alternative approach has advantages including a closer connection to the underlying biology and the capability to predict salp abundance.

KEY WORDS: Salp · Population dynamics · Stratification · Negative binomial distribution

INTRODUCTION

A central problem in marine ecology is understanding the dependence of the dynamics of a population on variations in the physical environment. Increasingly, this understanding is being furthered by the use of formal methods of statistical inference. In some cases, statistical modeling is complicated by the method of enumeration. One example of this is binning observed counts of individuals. Binning has been used in population analyses of zooplankton, phytoplankton and cetaceans (Legendre & Legendre 1998 and references therein). In an interesting example — discussed in more detail below — of the relationship between the abundance of salps and hydrographic variables, Ménard et al. (1993, 1994) were forced to rely on weekly binned counts of salps. To accommodate binning, they treated these data as an ordered categorical time series with Markov dependence and focused their analysis on the dependence of the transition probabilities on hydrography. A disadvantage of this approach is that these transition probabilities are not directly interpretable in terms of meaningful population parameters like the growth rate. The purpose of this paper is to describe an alternative approach aimed directly at the growth rate.

The basic statistical approach is outlined in the Methods section. In the Results section, this method is applied to the data of Ménard et al. (1993, 1994). The Discussion section contains some concluding remarks.

METHODS

Let $\mu(t)$ be the mean abundance or density of a salp population at time t. The basic biological assumption of this paper is that the per capita growth rate follows the model:

$$\frac{d\mu(t)}{dt}\frac{1}{\mu(t)} = \beta x(t) \qquad (1)$$

where $x(t) = X(t) - \overline{X}$ is the anomaly in an environmental variable $X(t)$ at time t from its long-term average $\overline{X}$. Formulating the model in terms of the anomaly $x(t)$ instead of the uncentered variable $X(t)$ is for convenience only. Specifically, the model in Eq. (1) is equivalent to:

$$\frac{d\mu(t)}{dt}\frac{1}{\mu(t)} = \beta X(t) - \beta \overline{X} \qquad (2)$$

Under this model, if $\beta > 0$, then the population exhibits positive or negative growth according to whether the environmental variable is above or below its long-term average. If $\beta < 0$, then these results are reversed. In either case, the population is in steady state when $x(t) = 0$. Under this model, abundance is given by:

$$\mu(t) = \mu(0)\exp\left(\beta\int_0^t x(u)\mathrm{d}u\right) \qquad (3)$$

where $\mu(0)$ is abundance at $t = 0$, and u is a dummy variable of integrtion Interest here centers on the parameter β, which controls the dependence of growth on $x(t)$.

Let the random variable $Y(t)$ be the unobserved number of salps contained in a sample of unit volume collected at time t. If salps were distributed randomly in space, then it would be natural to assume that $Y(t)$ has a Poisson distribution with mean $\mu(t)$. However, the salp life-cycle involves alternating solitary and aggregated stages, giving rise to a patchy spatial distribution, with many samples containing no individuals and a few containing very large numbers. To account for this so-called over-dispersion, we assume that $Y(t)$ has a negative binomial distribution with probability mass function:

$$\text{prob}[Y(t) = y] = \frac{\Gamma(y+\gamma^{-1})}{y!\,\Gamma(\gamma^{-1})}\left[\frac{\gamma\mu(t)}{1+\gamma\mu(t)}\right]^y[1+\gamma\mu(y)]^{-1/\gamma} \quad (4)$$

where Γ is the gamma function and $\gamma > 0$ is called the over-dispersion parameter. Under this model, $Y(t)$ has mean $\mu(t)$ and variance $\mu(t) + \gamma\,\mu(t)^2$. The limiting case $\gamma \to 0$ corresponds to the Poisson distribution. The negative binomial distribution is commonly used as a model for over-dispersed counts (Horne & Schneider 1995). This distribution can be extended to accommodate a surplus of zero counts, leading to the zero-inflated negative binomial distribution. For the data of Ménard et al. (1993, 1994), this did not lead to a significant improvement in fit. This completes the model for the unobserved counts, $Y(t)$. In addition to β, this model includes the unknown parameters $\mu(0)$ and γ.

Turning to the model for the observed binned counts, let (l_k, u_k), $k = 1, 2, \ldots, K$, be the lower and upper bounds of a set of K bins and define the random variable, $Z(t) = k$, if and only if $l_k < Y(t) \le u_k$. That is, $Z(t)$ corresponds to the bin to which a sample collected at time t is assigned. In statistical terminology, $Z(t)$ is an ordered categorical random variable. The probability mass function of $Z(t)$ is:

$$\begin{aligned}\text{prob}[Z(t) = k] &= \text{prob}[l_k < Y(t) \le u_k]\\ &= \sum_{y=l_k+1}^{u_k}\text{prob}[Y(t) = y]\end{aligned} \qquad (5)$$

where $\text{prob}(Y(t) = y)$ is given in Eq. (4).

The data available to Méinard et al. (1993, 1994) and used here for inference about the parameter β consist of a time series $Z(t), Z(t_2), \ldots, Z(t_n)$ of the bin numbers for samples collected at n times $t_1, t_2, \ldots, t_n$. Inference can be based on the likelihood (Azzalini 1996). The likelihood is defined as the joint probability mass of this observed time series regarded as a function of the unknown parameters. It is more convenient to work with the log likelihood, which is given by:

$$\log L[\beta,\mu(0),\gamma] = \sum_{j=1}^n \log\left(\sum_{y=l_{k_j}+1}^{u_{k_j}}\text{prob}[Y(t_j) = y]\right) \qquad (6)$$

where k_j is the observed value of $Z(t_j)$. That is, k_j is the bin to which the sample at time t_j is assigned. It is straightforward to calculate this log likelihood numerically.

As interest here centers on β, inference can be based on the profile log likelihood: $\log L_{\text{prof}}(\beta)$. Briefly, for a fixed value of β, $\log L_{\text{prof}}(\beta)$ is given by $\log L(\beta, \mu(0), \gamma)$, with $\mu(0)$ and γ replaced by their maximum likelihood (ML) estimates and β held at its fixed value. An approximate $1 - \alpha$ confidence interval for β found by inverting the likelihood ratio test is given by the values satisfying:

$$2[\log L_{\text{prof}}(\hat{\beta}) - \log L_{\text{prof}}(\beta)] < \chi_1^2(\alpha) \qquad (7)$$

where $\chi_1^2(\alpha)$ is the upper α-quantile of the chi-square distribution with 1 df.

RESULTS

We applied the outlined method to data of Ménard et al. (1993, 1994) relating to the abundance of salps at a site in the western Mediterranean. Salps are pelagic filter feeders that, under favorable conditions, exhibit very high population growth rates. By virtue of high filtration rates over a broad spectrum of particle sizes, salps are major grazers of phytoplankton and, through the production of fecal pellets are fundamental in the transfer of organic matter from surface waters to depth. Despite their important role as secondary producers and grazers of primary production, relatively little is known about the oceanographic conditions controlling growth and decay of salp blooms.

The data analyzed consist of weekly binned counts for the salp species *Salpa fusiformis* and *Thalia democratica* over the 24 yr period 1967–1990 from a sampling site in the Bay of Villefranche-sur-Mer (43° 41′ 10″ N, 7° 19′ 0″ E) in the western Mediterranean. *Salpa fusiformis* and *T. democratica* are the most common salp species in this region. Because the number of samples collected varied from week to week, the weekly abundance values were formed by dividing the approximate

number of salps in the pooled sample by the number of hauls and assigning this quotient to 1 of 6 bins. The lower and upper bounds of these bins are given in Table 1. Average counts were rounded up to the nearest integer. Further details about the data are available in Ménard et al. (1994). Representative sections of the binned time series for the 2 species for the 250 wk period from March 1984 through December 1988 are shown in Fig. 1a,b. Although only representative sections of the data are shown here, the analysis described below used the entire 24 yr data set.

As an illustration, we examined the effect of stratification, as measured by the difference in seawater density at 50 and 10 m, on the abundance of these species. Stratification is expected to influence abundance through its effect on the delivery of nutrients from deeper waters to surface waters. The weekly time series of the stratification anomaly, $x(t)$, over the representative 250 wk period

Table 1. Salp abundance bins

Abundance bin	Number of individuals in bin
1	0
2	1–3
3	4–17
4	18–80
5	81–350
6	351–27000

is shown in Fig. 1c. This time series has been centered at 0 by subtracting its (non-seasonal) average, so that $x(t)$ represents an anomaly in the stratification index. This has no effect on the estimates and confidence intervals for β reported below.

The method described does not require that $Z(t)$ be observed at every time point. In contrast, $\mu(t)$ in Eq. (3) involves the values of the environmental variable at every time point up to time t. Approximately 13 % of the measurements in the entire time series were missing. These missing values were reconstructed by fitting a periodic function with annual period to the remaining measurements. The fitted function:

$$\hat{x}(t) = -0.65 + 0.10\cos\omega t + 0.93\sin\omega t + 0.40\cos 2\omega t - 0.02\sin 2\omega t + 0.06\cos 3\omega t - 0.09\sin 3\omega t + 0.004\cos 4\omega t - 0.06\sin 4\omega t \qquad (8)$$

is also shown in Fig. 1c, where $\omega = 2\pi / 52$. The fit is not significantly improved by the addition of further harmonic terms in this model (p = 0.61, F test).

The profile log likelihoods for the parameter β for the 2 species are shown in Fig. 2. The source code for constructing these profile log likelihoods is available upon request. For *Salpa fusiformis*, the ML estimate of β is −0.28 with a ~0.95 CI of (−0.32, −0.25). This interval does not include 0, so the null hypothesis that β = 0 can be rejected at the 0.05 significance level. The ML estimates of $\mu(0)$ and γ are 1.08 and 8.80, respectively. The ML esti-
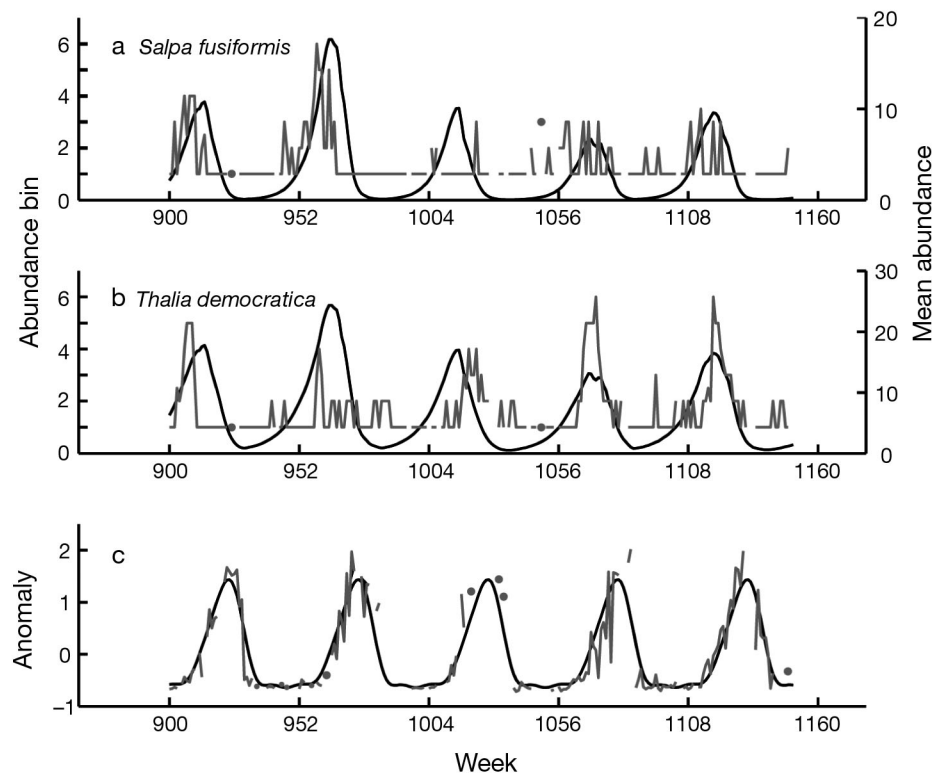

Fig. 1. *Salpa fusiformis, Thalia democratica.* Salp abundance and stratification anomaly. Observed binned abundance counts (gray line, left hand *y*-axis) and modeled abundance (black line, right-hand *y*-axis) of (a) *S. fusiformis* and (b) *T. democratica* with seawater stratification as environmental parameter. (c) Observed and fitted seawater stratification anomaly where anomaly is based on difference in density at 50 and 10 m. Data are a representative section of the 24 yr time series and span the period Mar 1984–Dec 1988
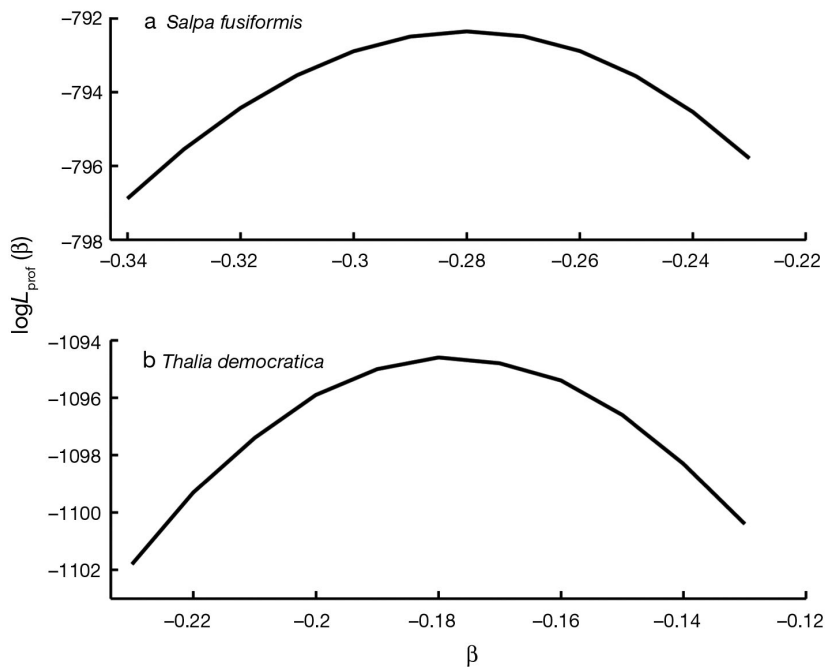
Fig. 2. Profile log likelihood of stratification parameter β for (a) *Salpa fusiformis* and (b) *Thalia democratica*

focuses on the effect of such variables on the dynamics of the underlying population. A disadvantage of the earlier approach is that the dynamics of $Z(t)$ reflect a combination of population growth and variability associated with patchiness. For example, a transition from a higher category at sampling time t_j to a lower category at sampling time t_{j+1} can occur if population growth is positive but a high density patch is sampled by chance at the earlier time but not at the later time. In contrast, the approach proposed here, while accounting for patchiness through the negative binomial model, is aimed directly at inference about the growth parameter β. For this reason, a direct comparison of the 2 approaches is difficult. However, in qualitative terms, both identify a significant effect of stratification — presumably operating through its effect on the supply of nutrients to the surface waters.

The focus of this paper has been on the general statistical problem of making inferences about population dynamics from an ordered categorical time series formed by binning counts of individuals and not on the specific details of the population dynamics of salps. For this reason, the analysis presented here was based on a simple model of population growth. It would be straightforward to increase the complexity of this model, for example, by including other environmental variables or allowing for density dependence.

mate of β for *Thalia democratica* is –0.18, with a ~0.95 CI of (–0.20, –0.15). Again, the null hypothesis that β = 0 can be rejected at the 0.05 significance level. The ML estimates of $\mu(0)$ and γ for this species are 2.18 and 9.64, respectively.

The fitted abundances $\hat{\mu}(t)$ of the 2 species are shown in Fig. 1a,b for the representative sections. These appear to capture the pattern in the time series $Z(t)$ reasonably well. As a rough measure, the sample correlation coefficients between $\hat{\mu}(t)$ and $Z(t)$ are 0.30 for *Salpa fusiformis* and 0.29 for *Thalia democratica*. These correlations should be interpreted with caution in light of the categorical nature of $Z(t)$. In particular, it is difficult to provide a formal assessment of significance.

DISCUSSION

We describe an alternative approach to analyzing the categorically binned data of Ménard et al (1993, 1994). Although both their analysis and the current analysis identified a significant effect of stratification on the salp populations in the Bay of Villefranche-sur-Mer, the 2 approaches, are not directly comparable. Ménard et al. (1993, 1994) focused on the effect of environmental variables on the dynamics — specifically, the transition probabilities — in the ordered categorical time series $Z(t)$. The approach described here

LITERATURE CITED

Azzalini A (1996) Statistical inference based on the likelihood. Chapman & Hall/CRC, Boca Raton, FL

Horne JK, Schneider DC (1995) Spatial variance in ecology. Oikos 74:18–26

Legendre P, Legendre L (1998) Numerical ecology. Elsevier Science BV, Amsterdam

Ménard F, Dallot S, Thomas G (1993) A stochastic-model for ordered categorical time-series: application to planktonic abundance data. Ecol Model 66:101–112

Ménard F, Dallot S, Thomas G, Braconnot JC (1994) Temporal fluctuations of two Mediterranean salp populations from 1967 to 1990: analysis of the influence of environmental variables using a Markov-chain model. Mar Ecol Prog Ser 104:139–152