Episodic Fluctuations in Larval Supply

Paul A. Dixon,[1]* Maria J. Milicich,[2]† George Sugihara[1]

The lack of a clear relationship between spawning output and recruitment success continues to confound attempts to understand and manage temporally variable fish populations. This relationship for a common reef fish is shown to be obscured by nonlinear processes in operation during the larval phase. Nonlinear responses of larval fish to their noisy physical environment may offer a general explanation for the erratic, often episodic, replenishment of open marine populations.

Understanding the nature of feedback control between the size of a population and its successful replenishment is fundamental to the understanding of the mechanics of population change. Fisheries assessors in particular use various models (for example, Beverton-Holt) to try to predict recruitment levels from reproductive stock size (*1*). Such fitted curves, however, have been plagued by extreme variance in recruitment, and explaining this bivariate scatter has remained a largely unrealized goal for the last 100 years. Some researchers have attributed the unexplained variability to sampling error and the misuse of analytical techniques (*2*), while others have suggested that unpredictable and high levels of mortality during an intervening larval phase will necessarily obscure the stock-recruitment relationship (*3*). Because an understanding of the fate of larval fishes has remained elusive, it has proven difficult to evaluate the relative contributions of these sources of error. Here, we offer a new approach to investigating processes at work during the larval phase.

For many benthic marine species, observations of rates of larval supply have been recorded for considerable lengths of time (*4*). Such records are typically erratic, dominated by a small number of high-magnitude events (*5*). Attempts to relate larval numbers to concurrently measured environmental variables using linear techniques—that is, regression analysis—have been largely unsuccessful. Episodes of strong larval supply have often been termed random and thought to be unpredictable with available oceanographic and meteorological data. However, we provide evidence that pulses in larval supply may arise from a nonlinear response of larval fishes to a relatively small number of key physical forcing variables.

We focused on a population of *Pomacentrus amboinensis*, a common damselfish species found on the Great Barrier Reef, Australia. Daily observations of spawning output (visual census of eggs released from nests) and recruitment (counts of recently settled juveniles) have been recorded for 2 years, and larval supply (mature larvae collected with light traps) for 3 years, at Lizard Island. As is characteristic of fisheries problems, density-dependent feedback between stock size and subsequent recruitment accounts for only ~10% of the variation in the relationship between these two variables at the scale of a single reef (*6*).

We applied techniques from nonlinear time series analysis to determine whether the fluctuations in the larval supply time series are indicative of the action of nonlinear processes or are random features of the data. The approach was to construct a series of algorithms, ranging from global/linear to local/nonlinear, to forecast the data based on lagged-coordinate embeddings (*7*). The underlying philosophy is that, if nonlinear algorithms are found to outperform their linear counterparts, this may be taken as evidence that the processes that generated the time series were themselves nonlinear (*8*). Alternatively, if the underlying processes were linear or stochastic, linear models should provide better forecasts, as they use global information to generate predictions. For forecasting algorithms, we used the method of S-maps (*9*). S-maps generate forecasts for each point in the embedded time series by first excluding that point (the predictee), and then fitting a linear surface to the remaining data (the predictors). This technique is similar to linear kernel regression (*10*) in that the degree of nonlinearity in these mappings is given by the extent to which the contribution of predictor points to a fitted surface is a function of their distance from the predictee. This weighting procedure is controlled by a tuning parameter theta, where theta equal to zero gives a global linear map, and increasing values of theta give

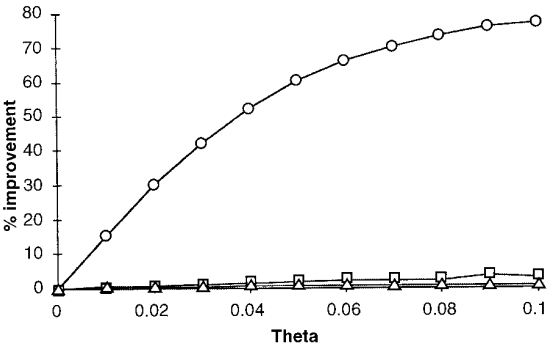


Fig. 1. Results of S-map analysis for reproductive output (squares), larval supply (circles), and recruitment (triangles) of *P. amboinensis*. The ordinate gives the percent increase in correlation coefficient between predicted and observed values for nonlinear versus linear models. Predictability for spawning output and recruitment remains constant for across the spectrum of linear to nonlinear algorithms, whereas for larval supply, predictability increases significantly as the forecasting algorithms become more nonlinear (theta increases).

Table 1. Summary of model statistics. Percent illumination was calculated by multiplying the proportion of a full moon for each night by the elongation (phase angle) for that night. Wind data were taken from an automatic recording station (Willis Island) situated in the same weather stream as Lizard Island. Cross-reef wind components were calculated by weighting wind speed according to the angle perpendicular to the reef.

Variables		Model type	Rho	Error
Percent illumination	19 day lag	Linear (in-sample)	0.33	99.4
		Linear (out-of-sample)	0.29	100.5
		Nonlinear (n = 70)	0.32	101.4
Percent illumination	19 day lag	Linear (in-sample)	0.38	106.1
Plus cross-shelf wind speed	1 day	Linear (out-of-sample)	0.34	107.7
		Nonlinear (n = 70)	0.45	91.9
Percent illumination	19 day lag	Linear (in-sample)	0.42	112.4
Plus cross-shelf wind speed	1 day	Linear (out-of-sample)	0.37	114.6
Plus daily wind speed	16 day lag	Nonlinear (n = 30)	0.55	91.8
		Nonlinear S-map (θ = 0.8)	0.53	90.4

[1]Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093, USA. [2]Griffith University, Nathan Brisbane, Australia.

*To whom correspondence should be addressed. E-mail: paul@complex.ucsd.edu

†Present address: M_2 Environmental Ltd., 3D #24, Tung, Shan Terrace, Happy Valley, Hong Kong.

increasingly local and nonlinear algorithms (*11*).

Results of the S-map analysis are given in Fig. 1. Forecastability in both egg production and recruitment remained essentially constant across the spectrum from linear to nonlinear models; daily dynamics of spawning and recruitment were well described by a simple linear autoregressive model (*12*). However, for the larval supply series, prediction improved significantly ($P < 0.05$, Z test) as the model was tuned toward nonlinear solutions. This suggested that, in addition to purely linear or stochastic events, there was a nonlinear component of larval phase processes that was obscuring the spawning-recruitment relationship.

To test this, we examined three distinct physical-biological interactions: a lunar or semilunar entrainment of egg release from the nest, the transport environment experienced by mature larvae returning to the reef, and the effect of turbulence levels on young, first-feeding larvae (*13*). We chose, respectively, the following physical variables as representative of these processes: nighttime irradiance lagged 19 days (corresponding to average larval life), cross-shelf wind speed lagged one day, and average daily wind speed at a range of time lags. Model generation involved fitting linear surfaces; model nonlinearity was controlled by adjusting the neighborhood size to which the surfaces were fit (*14*). To increase the number of nonzero points in the time series, we considered all pomacentrid species in this analysis (*15*). Table 1 gives summary statistics for all the multivariate models constructed, and representative model forecasts are given in Fig. 2.

We began by modeling lunar entrainment, a deterministic component of larval supply. There was little improvement in model performance when nonlinear models were considered in comparison with linear regression. We then added cross-shelf wind speed, a stochastic component, into the model. In this case, nonlinear maps offered some increase in forecasting performance (Table 1) (*16*).

Choosing a physical variable to reflect events occurring early in larval life was more difficult. Such effects are plausible, given the suggestion that retention mechanisms may work to keep larvae local to their native reef (*17*), and otolith studies describing the age of these fish have been reported (*6*). However, because not all larvae are precisely the same age, and further because both the time required for yolk-sac resorption and the time to inevitable mortality in the absence of successful feeding are unknown, we proceeded conservatively by considering a wide range of lags (from 1 to 30 days) for daily averaged wind speed. Viewed linearly, there was no significant difference in model performance between any of these lags.

We then tuned all 30 lags toward nonlinear model solutions, creating a rho-lag-neighborhood surface (Fig. 3A). The slight improvement in correlation coefficient at intermediate lags, which was evident in linear models, increased significantly (Z test; $P < 0.05$) as the models were made more nonlinear. Notably, only models coinciding with the age of the larvae offered improvement over the bivariate nonlinear model. Otolith analysis of these fishes has indicated that they are from 16 to 23 days old, with a mean age of 19 days; the strongest effects we identified ranged from 15 to 17 days and centered at 16 days, 3 days after release from the nest.

To extract the relationship between larval supply and each of the three forcing variables in this model, we employed a neighborhood

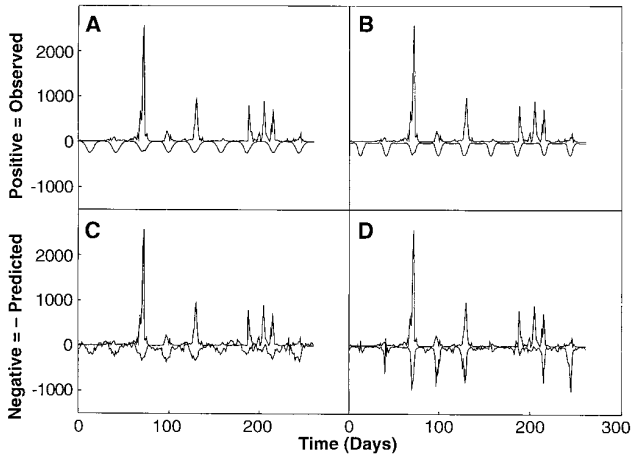
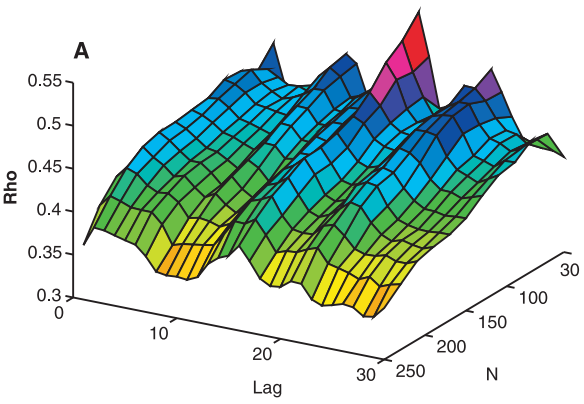


Fig. 2. Predicted versus observed values for four physical models considered (Table 1). Positive axis gives observed larval number; negative axis is (–1 × predicted value). (**A**) Linear, univariate model using lunar phase with optimal lag 19 days). (**B**) As in (A), but with a nonlinear model operating on lunar phase. (**C**) Linear, trivariate model. (**D**) As in (C), but with a nonlinear model. In addition to superior prediction statistics, the nonlinear trivariate model (D) best captures the erratic appearance of the series.



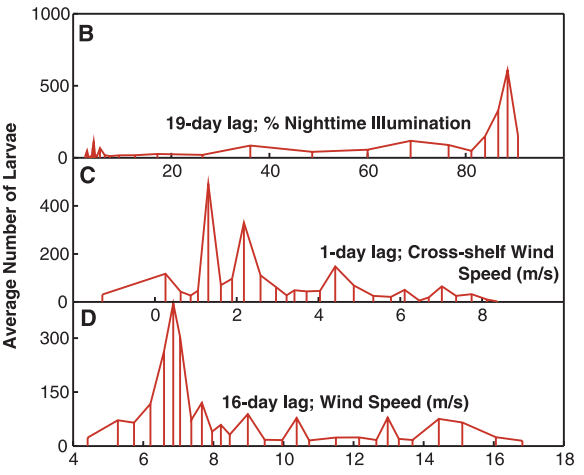


Fig. 3. (**A**) Rho-neighborhood size-wind stress lag surface. The slight increase in rho at intermediate lags strengthens as the models become increasingly nonlinear. The best model, with a 16-day lag, corresponds to 3 days after the average date of larval release from the nest. (**B**) Response of larvae to 19-day lagged nighttime irradiance. Maximum larval abundance is associated with release on the full moon. (**C**) Larval response to recent cross-shelf wind speed. Larvae are positively associated with weak onshore winds (positive values), which may enable mature larvae to promote their return to the reef. (**D**) Larval number as a function of 16-day lagged average daily wind stress. Note the dome-shaped response to wind stress, centered around 4 to 8 m/s.

averaging procedure (*18*). For nighttime irradiance (Fig. 3B), maximum larval abundance was associated with a full moon 19 days prior. For cross-shelf wind speeds (Fig. 3C), larval abundance was positively associated with weak onshore winds. Recent findings have demonstrated that mature reef fish larvae are strong swimmers (*19*), and we hypothesize that lower wind speeds may promote easier orientation and increase directional swimming in larvae as they approach benthic habitats for the first time. The relationship between larval supply and 16-day lagged wind speed is shown in Fig. 3D. Recent hypotheses have proposed a dome-shaped relationship between larval feeding and turbulence levels (*20*). Our analysis supports this, and we suggest that wind effects are indeed important to early larval survival. In this case, the optimal window for early larval survival appears to have been approximately 4 to 8 meters per second.

It seems clear that, in this system, variability in larval supply originates from both deterministic (lunar entrainment) and stochastic processes (wind stress). The key elements of nonlinearity appear to be the response of the larvae to these stochastic variables and the sequential nature of processes operating at different times in larval life. Such sequential action serves to couple physical variables, causing their total impact on larval supply to be multiplicative. In other systems, one may expect the most relevant physical variables and lags, and the responses to those variables, to vary. However, it is likely that these key elements of nonlinearity will remain. We therefore offer a dynamical explanation for the episodic nature of larval supply: the nonlinear amplification by biological processes of stochastic physical forcing.

References and Notes

1. B. J. Rothschild, *Dynamics of Marine Fish Populations* (Harvard Univ. Press, Cambridge, MA, 1986).
2. R. Hilborn and C. Walters, *Quantitative Fisheries Stock Assessment: Choice, Dynamics, and Uncertainty* (Chapman & Hall, London, 1992); N. Caputi, *Can. J. Fish. Aquat. Sci.* **45**, 178 (1988).
3. D. Cushing, *Climate and Fisheries* (Academic Press, London, 1982); R. Lasker, *Marine Fish Larvae* (Univ. of Washington Press, Seattle, WA, 1981).
4. J. M. Shenker *et al.*, *Mar. Ecol. Prog. Ser.* **98**, 31 (1993); M. J. Milicich and P. J. Doherty, *ibid.* **110**, 121 (1994); S. Sponaugle and R. K. Cowen, *ibid.* **133**, 13 (1996).
5. P. J. Doherty and D. M. Williams, *Oceanogr. Mar. Biol. Annu. Rev.* **26**, 487 (1988).
6. M. G. Meekan, M. J. Milicich, P. J. Doherty, *Mar. Ecol. Prog. Ser.* **93**, 217 (1993).
7. F. Takens, in *Dynamical Systems and Turbulence*, D. A. Rand and L. S. Young, Eds. (Springer, New York, 1981), pp. 366–381.
8. G. Sugihara and R. M. May, *Nature* **344**, 734 (1990).
9. G. Sugihara, *Philos. Trans. R. Soc. Lond. Ser. A* **348**, 477 (1994).
10. M. P. Wand and M. C. Jones, *Kernel Smoothing* (Chapman and Hall, London, 1995); Q. Yao and H. Tong, *Stat. Sin.* **4**, 51 (1994).
11. For an embedded time series $X_t \in R^{m+1}$, where the constant term in Eq. 2 below is given by $X_t(0) \equiv 1$, and the time series value Tp steps forward is $X_{t+Tp}(1) = Y_t$, forecasts at T_p are given by

$$\hat{Y}_t = \sum_{j=0}^{m} C_t(j)\, X_t(j) \qquad (1)$$

For each predictee, X_t, we used singular value decomposition to solve for C by using the rest of the data set as follows

$$B = AC \qquad (2)$$

where

$$B_i = w(\|X_i - X_t\|)Y_i,\ A_{ij} = w(\|X_i - X_t\|)\,X_i(j)$$

and

$$w(d) = e^{-\theta d/\bar{d}}$$

12. Best forecasts were given by a three-term autoregressive model (AR3); however, other AR models performed similarly.
13. M. J. Milicich, *Mar. Ecol. Prog. Ser.* **110**, 135 (1994); ———, M. G. Meekan, P. J. Doherty, *ibid.* **86**, 153 (1992).
14. Neighborhood regression, S-maps, and linear kernel regression are all related in the use of a weighting function to control the contribution of points to local linear surface: a step function, an exponential, and (typically) a probability density function with area one such as the normal, respectively. In many situations, kernel regression has been shown to outperform simple neighborhood regression, because the contribution of points to the forecast diminishes as a smooth function of distance. S-maps also have this property, however, and in this case, give very similar results (Table 1).
15. Data on *P. amboinensis* larval supply contained 35 days with no individuals sampled. Decreasing the taxonomic resolution to the family level reduced this to 29 days. S-map analysis gave similar results (P. A. Dixon, M. J. Milicich, G. Sugihara, data not shown) when applied to all pomacentrid species. Also, nonlinear physical models performed comparably when constructed for *P. amboinensis* alone.
16. The specific choices of both of these variables were not critical. Because of the deterministic nature of lunar phase, many choices of lags for nighttime irradiance performed similarly (although the best lag coincided with the mean age of the larvae). Other lag choices for cross-shelf wind speed also gave comparable results.
17. J. M. Leis, *Mar. Biol.* **90**, 505 (1986).
18. For each point in the larval supply series, we identified its 30 nearest neighbors in model space, and calculated the average value of each physical variable and larval supply of those neighbors. We then sorted the averages for the physical variables, binned them into groups of ten, and plotted the values of the bins against the corresponding binned values of larval supply.
19. I. C. Stobutzki and D. R. Bellwood, *J. Exp. Mar. Biol. Ecol.* **175**, 275 (1994).
20. W. C. Legett and E. DeBlois, *Neth. J. Sea Res.* **32**, 119 (1994).

21. We are most grateful to the following: R. Alford, L. Bersier, R. Brown, M. Casdagli, H. Choat, G. Jones, H. Hastings, A. Hobday, T. Hughes, J. Lough, R. Penner, and G. Russ. Supported by endowment funds from the John Dove Isaacs Chair in Natural Philosophy, the Office of Naval Research, and Deutsche Bank NA (P.A.D. and G.S.); and by Griffith University, Lizard Island Research Station (Australian Museum); Australian Institute of Marine Science; and James Cook University (M.J.M.).

18 August 1998; accepted 2 February 1999

Exon Shuffling by L1 Retrotransposition

John V. Moran,*† Ralph J. DeBerardinis, Haig H. Kazazian Jr.†

Long interspersed nuclear elements (LINE-1s or L1s) are the most abundant retrotransposons in the human genome, and they serve as major sources of reverse transcriptase activity. Engineered L1s retrotranspose at high frequency in cultured human cells. Here it is shown that L1s insert into transcribed genes and retrotranspose sequences derived from their 3′ flanks to new genomic locations. Thus, retrotransposition-competent L1s provide a vehicle to mobilize non-L1 sequences, such as exons or promoters, into existing genes and may represent a general mechanism for the evolution of new genes.

The human genome is littered with noncoding DNA, often disparaged as "junk DNA." Much "junk DNA" results from the reverse transcription of cellular RNAs and insertion of the cDNAs into new genomic locations by retrotransposition. L1s make up about 15% of human DNA (*1*). The majority of L1s cannot retrotranspose, but an estimated 30 to 60 full-length L1s remain retrotransposition-competent (*2*). These L1s contain a 5′ untranslated region (UTR), two nonoverlapping open reading frames (ORF1 and ORF2), and a 3′ UTR that ends in a polyadenylic acid [poly(A)] tail (*3*, *4*). ORF1 encodes an RNA-binding protein (*5*), whereas ORF2 encodes an endonuclease (EN) activity (*6*), a reverse transcriptase (RT) activity (*7*), and a cysteine-rich (C) domain of unknown function (*8*) (Fig. 1A). L1 retrotransposition is ongoing because recent insertions have caused diseases in humans and mice (*4*). L1s also are thought to mobilize Alus and processed pseudogenes, which make up another 10% of human DNA (*4*, *9*). Thus, either directly or through the promiscuous mobilization of cellular RNAs, L1s may be evolutionarily responsible for one-fourth of human DNA (*1*).

Department of Genetics, University of Pennsylvania School of Medicine, Philadelphia, PA, 19104–6145 USA.

*Present address, Departments of Human Genetics and Internal Medicine, University of Michigan Medical School, Ann Arbor, MI, 48109–0650 USA.
†To whom correspondence should be addressed. E-mail: moranj@umich.edu or kazazian@mail.med.upenn.edu