THE 2 PPT BOTTOM ISOHALINE AS AN INDEX OF CONDITIONS FOR BIOLOGICAL RESOURCES

Alan D. Jassby
Division of Environmental Studies
University of California, Davis

398

June 17, 1992

Introduction

The purpose of this report is to portray the empirical relationships between several biological resources in the San Francisco estuary and the position of the 2 ppt bottom isohaline, denoted here by X_2. As is widely understood, statistical relationships are not proof of causal connections, and it is not the intention of this report to suggest that X_2 itself controls biological resources in the estuary. Rather, the particular hypothesis investigated here is that X_2 can serve as an index of those factors that *do* underly the variability in biological resources, such as freshwater inflows. This hypothesis is of interest because of the well-defined nature of X_2, as well as its relative ease of measurement compared to such variables as net delta outflow.

The associations between biological resources (e.g., longfin smelt abundance) and environmental factors such as X_2 are expressed here using *generalized linear models*, which, as the name implies, are flexible extensions of classical linear models.[1] We use only the simplest generalized linear models in this report, with the goal of determining whether or not some significant statistical relationship exists between biological resources and X_2. These models no doubt can be refined for other purposes, such as forecasting or determining standards.

Methods

If Y is the response variable, the Z_i are predictor variables, and $E(Y)=\mu$, a generalized linear model takes the following form:

$$g(\mu) = \alpha + \sum_{i=1}^{p} \beta_i Z_i$$

$$\text{var } Y = \phi V(\mu)$$

where α and the β_i are constants. V is a *variance* function describing how the variance of Y depends on the mean and ϕ is the *dispersion parameter*; g is a *link* function describing how the mean depends on the linear combination of predictors. If V=1 and g=I, the identity function, the model is equivalent to a classical linear model. The models are estimated by maximum-likelihood, using an iteratively reweighted least-squares algorithm.

The response variables were chosen so that populations at a number of trophic levels would be represented. The predictor variables were determined as follows: For each biological resource, a specialist was consulted to recommend the averaging period over which X_2 was likely to be related to the resource. In the case of longfin smelt, for example, the average of X_2 for the period February-May was used. These periods were chosen on the basis of the biology of the resource in question, not by trying to optimize some statistic. The variables used and observations available for each model are summarized in Table 1. Where noted, the observation corresponding to 1983 flows was eliminated. In these cases, a significant portion of the population may have been seaward of the sampling stations, causing an underestimate of the annual abundance.

---

[1] P. McCullagh and J.A. Nelder, *Generalized Linear Models*, 2nd. ed. (London: Chapman and Hall, 1989).

Table 1. Response variables, associated predictor variables (X_2 averaged over some period), and observations available.

Response	X_2 period	Observations
Phytoplankton primary production plus river load of algal-derived POC in Suisun Bay, annual total (Gg yr^{-1})[2]	Jan-Dec	75-89
Neomysis mercedis, Mar-Nov abundance index (no.)	Mar-Nov	72-88[a]
Crangon franciscorum juveniles, annual abundance index (no.)	Mar-May	80-90
molluscs in Grizzly Bay, annual abundance (no. m^{-2})	3-yr average Jan-Dec	81-90
Striped bass survival index, egg to 38 mm	Apr-Jul	69-82, 84-91
Starry flounder year-class 1+, annual abundance index (no.)	previous year Mar-Jun	80-91[a]
Longfin smelt YOY plus adults, annual abundance index (no.)	Jan-Jun	68-73, 75-78, 80-91[a]

[a]1983 intentionally omitted

Because of the small amount of data available (number of observations n≤22), models requiring estimation of more than 2 parameters (aside from the intercept) were not considered. For each resource, two models were estimated: (1) X_2 alone, averaged over some suitable period, and (2) a natural spline in X_2 with 1 interior knot (2 degrees of freedom).[3] If more than one model was "well-behaved" (coefficients significantly different from zero [$p<.05$] and residuals consistent with model assumptions), the final model was selected on the basis of the *AIC* statistic.[4]

Results

The results for each biological resource are summarized in Table 2 and in Figs. 1-7. The models are well-behaved in that all coefficients are significantly different from 0 ($p<.05$) and the residuals are in conformity with model assumptions, although the number of points is generally too small for a powerful test of these assumptions.

[2]See Appendix for the rationale underlying the use of this variable.

[3]Generally speaking, a natural spline is superior to a polynomial for representing nonlinearities; with the few degrees of freedom permitted here, however, the difference between the two may be unimportant.

[4]T.J. Hastie and R.J. Tibshirani, *Generalized Additive Models* (London: Chapman and Hall, 1990).

Table 2. Summary of relationships between response variables Y and X_2: n, number of observations; g, link function; V, variance function; df, degrees of freedom for X_2 in model (1=linear, 2=natural spline with 1 interior knot); r, multiple correlation between Y and the value fitted from the model.

Y	n	g	V	df	r
phytoplankton	15	I	1	1	.85
Neomysis	16	I	μ	1	.79
Crangon	11	I	1	1	.91
molluscs	10	I	μ	2	.80
striped bass	22	I	1	1	.62
starry flounder	11	log	μ	2	.92
longfin smelt	21	log	μ	1	.86

Discussion

The data demonstrate that simple and statistically significant relationships exist between X_2 and biological populations at many trophic levels. Moreover, the supply of energy to the base of the food web, as represented by phytoplankton carbon, is also associated with X_2. Except for mollusc density in Grizzly Bay, all responses show a decline as X_2 increases, i.e., as flows decrease. Molluscs exhibit an increase at extreme values of X_2, whether high or low.[5]

X_2 is clearly a viable candidate for indexing estuarine conditions. In addition to its well-defined and measurable nature, it has a pervasive and clear relationship with many estuarine biological properties. Relationships exist between X_2 and an important component of the food web base in Suisun Bay (phytoplankton POC), primary and secondary zooplankton consumers (*Neomysis* and *Crangon*), a major group of benthic consumers in Suisun Bay (molluscs), bottom-foraging fish (starry flounder), and both survival (striped bass) and abundance (longfin smelt) of fish that feed in the water column. The statistical significance of these simple relationships and the wide variety of trophic levels involved reflects the ability of X_2 to act as a surrogate for net delta outflow and other hydrodynamic variables.

Although these empirical relationships offer evidence that X_2 summarizes estuarine conditions for many properties of interest, they are not adequate for forecasting in their present form. X_2 represents the influence of net delta outflow in these relationships, as well as partial influences of those variables that are correlated with net delta outflows, such as diversions and food supply. Because such correlations may very well break down in the future as water management changes, the effects of these other variables should be included explicitly in

[5]Although the causal mechanisms underlying these associations are not at issue here, the distinctive response of the mollusc community deserves a few comments. Persistent high values of X_2 (persistent low flows) permits the colonization of Suisun Bay by marine benthic macroinvertebrates. In times past, the main colonizing species was *Mya arenaria*, but this role has been usurped by the invader *Potamocorbula amurensis* during the current drought. In a similar manner, persistent low values of X_2 (persistent high flows) leads to colonization by freshwater benthic macroinvertebrates, particularly *Corbicula fluminea*. The net effect of these "high-density" colonizations from both the seaward and landward directions under persistent low or high flows, respectively, is a minimum in mollusc density at intermediate values of X_2 (Fig. 4).

predictive models, if possible. This is a challenging problem, because the relatively small amount of data restricts the number of variables and nonlinearities that can be incorporated rationally into predictive models.

Appendix.

The organic matter supply for Suisun Bay is dominated by loading from the Sacramento-San Joaquin rivers, at least for 1980.[6] Because of the lack of data for other years, as well as uncertainty regarding the availability of this organic matter for the food web, it is not possible to characterize the year-to-year changes in the total food supply for Suisun Bay. Data do exist, however, for examining interannual changes in an important part of this food supply, namely, the particulate organic carbon (POC) contributed by phytoplankton. Phytoplankton POC is supplied both by primary productivity within Suisun Bay, as well as by the input of phytoplankton and phytoplankton-derived detritus from upstream of Suisun Bay. The phytoplankton POC from upstream sources may, in fact, represent, much of the utilizable organic matter supplied by river inflows.

Two main factors underlying interannual variability in primary productivity *within* Suisun Bay have been implicated. The first is the effect of river inflow on residual circulations that retain phytoplankton biomasss in the upper estuary. The second is consumption by benthic herbivores, which is indirectly influenced by river inflows: during low flow periods of at least 16 consecutive months, estuarine benthic macroinvertebrates colonize Suisun Bay, leading to a large increase in their biomass and grazing impact. Together, these factors result in highest productivity at intermediate flows.[7]

Data also exist for an assessment of year-to-year fluctuations in the phytoplankton-associated particulate organic carbon carried *into* Susiun Bay. Riverine loading of algal-derived pigments was determined for 1975-1989 by summing flow-weighted chlorophyll *a* and phaeophytin *a* loads from the Sacramento and San Joaquin Rivers at Emmaton and Jersey Point, respectively. Pigment loads were then converted to POC using a mean carbon:pigment ratio of 40.

Together, these two different series (primary productivity and algal-derived POC load) constitute the phytoplankton-associated organic carbon supplied to the food web of Suisun Bay for 1975-1989.

---

[6]B. Herbold, A.D. Jassby, and P.M. Moyle, *Status and Trends Report on Aquatic Resources in the San Francisco Estuary* (San Francisco: San Francisco Estuary Project, U.S. Environmental Protection Agency, 1992)

[7]A. Alpine and J.E. Cloern, "Trophic Interactions and Direct Physical Effects Control Phytoplankton Biomass and Production in an Estuary," *Limnology and Oceanography* 37 (1992): in press.

70 km, when it is 12%. Since this would provide little overall improvement to the model, and since we will use monthly or seasonal average data, we did not pursue this further.

Filling in the gaps The gaps in the data record were filled in by using the autoregressive model stepwise to predict each value in turn. However, this generally resulted in relatively large jumps in X_2 from the last value in each gap to the next value, which had been measured. In principle each X_2 value should be equally correlated to the succeeding value as to the preceding value. Therefore we forecast the data in the gaps in both directions. Each value was then calculated as a weighted mean of the two calculated forecasts, the weighting factor being the time in days between the individual date and the previous or following known value. This approach caused the values to vary smoothly from their values at either end of the gaps.

Figures 5 through 10 show the values of the interpolated data, the values filled in as described above, and the values determined by the autoregressive model starting at the beginning of the entire time series. The latter regression reproduces general patterns well, but does not appear as useful in filling in gaps as our 2-way method described above. Of a total of 8827 days of data, 1033 days or 12% were missing (Figure 11), mostly because high flows pushed X_2 downstream of the most seaward monitoring station.

Monthly and seasonal means Once the filled-in daily values of X_2 had been obtained, we were able to calculate values by month and by season. Monthly values (Figure 12) varied with flow as has been observed before, and the regression line was close to that obtained using Phil Williams' data. The best regression to predict monthly X_2 was structurally the same as that for daily values:

$$X_2(t) = 122.2 + 0.3278\, X_2(t-1) - 17.65\, \mathrm{LOG}\,[Q_{out}(t)],$$

where t is now in months. The R^2 value was 0.96 and the standard error of the estimate was 2.30 km. The standard error of the estimate for predicting the entire time series from flow and the first value was 2.45 km. This regression has a tight fit with no obvious outliers or departure from linearity (Figure 13). If the lag term in X_2 is removed by assuming steady state and setting $X_2(t) = X_2(t-1)$, the slope of the relationship with log (Q_{out}) is -26.3.

Seasonal mean values (Figure 14) show somewhat lower variability in summer (July-September) than the other seasons. Except during droughts, when X_2 has been high in all seasons, winter values generally fall below 70 km and autumn values below 80 km, while summer values tend to exceed 80 km most of the time.

Comparison with previous estimates The previous estimates based on regressions with flow produce a higher value under low-flow conditions (Figure 15). This is consistent with what Richard Denton observed in attempting to predict salinity at fixed stations from flow: the calculated flow values are underestimates at low flows in summer because the estimate of net consumption in the delta at those times has been reduced.

Comparison of the grab sample data with the current best estimates reveals a large amount of scatter and apparent bias at values of X_2 above about 80 km (Figure 16). Below 80 km the residual is not significantly different from 0 (t-test), while at or above 80 km it averages 2.9±0.6 km (mean ± 95% CL of the mean). The scatter may be largely due to the effects of tidal variation, while the bias could be due to the common practice of sampling while attempting to follow the high tide upstream.

Conclusions The current X_2 estimates provide a value that is easily determined from the data and whose residuals appear well-behaved. The differences between the best estimates and other values (i.e. previous estimates and grab sample data) are explainable. Thus this data set is probably adequate for setting a salinity standard.

The model could be improved in several ways that are probably not warranted now. First, the small but significant tidal effect could be included. Second, it is evident from Figures 5-10 that some hysteresis may exist in the signal: that is, the response to a declining flow differs from that to an increasing flow. Furthermore, there appear to be differences in response at high and low flows, to the extent that these are observable with the present data set.

Figure captions and figures

Fig. 1. Supply of particulate organic carbon (POC) to Suisun Bay from phytoplankton production and riverine loading of algal-derived particulate matter, compared to annual average X_2.

Fig. 2. March-November mean abundance index of *Neomysis mercedis* in the San Francisco estuary.

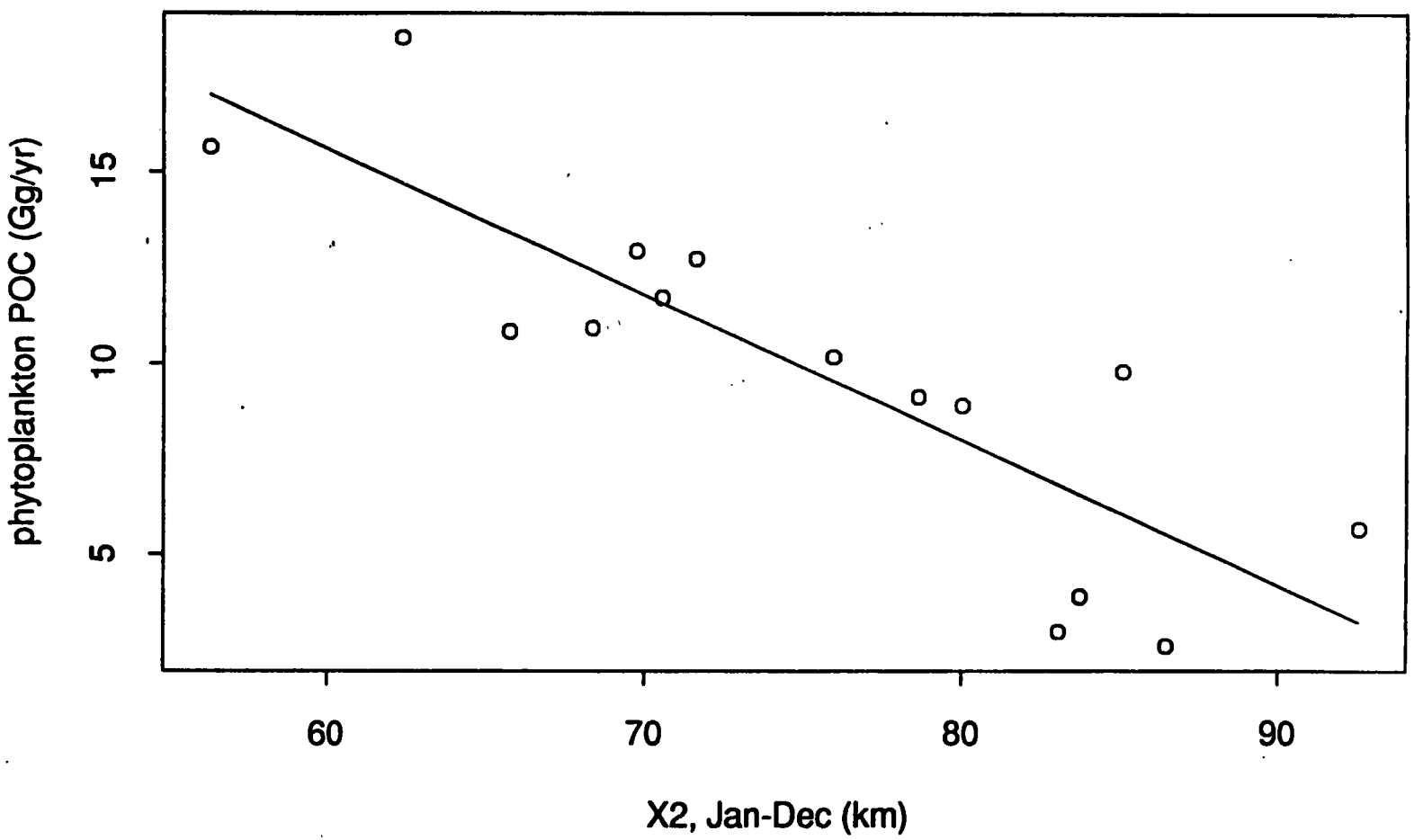
Fig. 3. Annual mean abundance index of *Crangon franciscorum* juveniles in the San Francisco estuary.

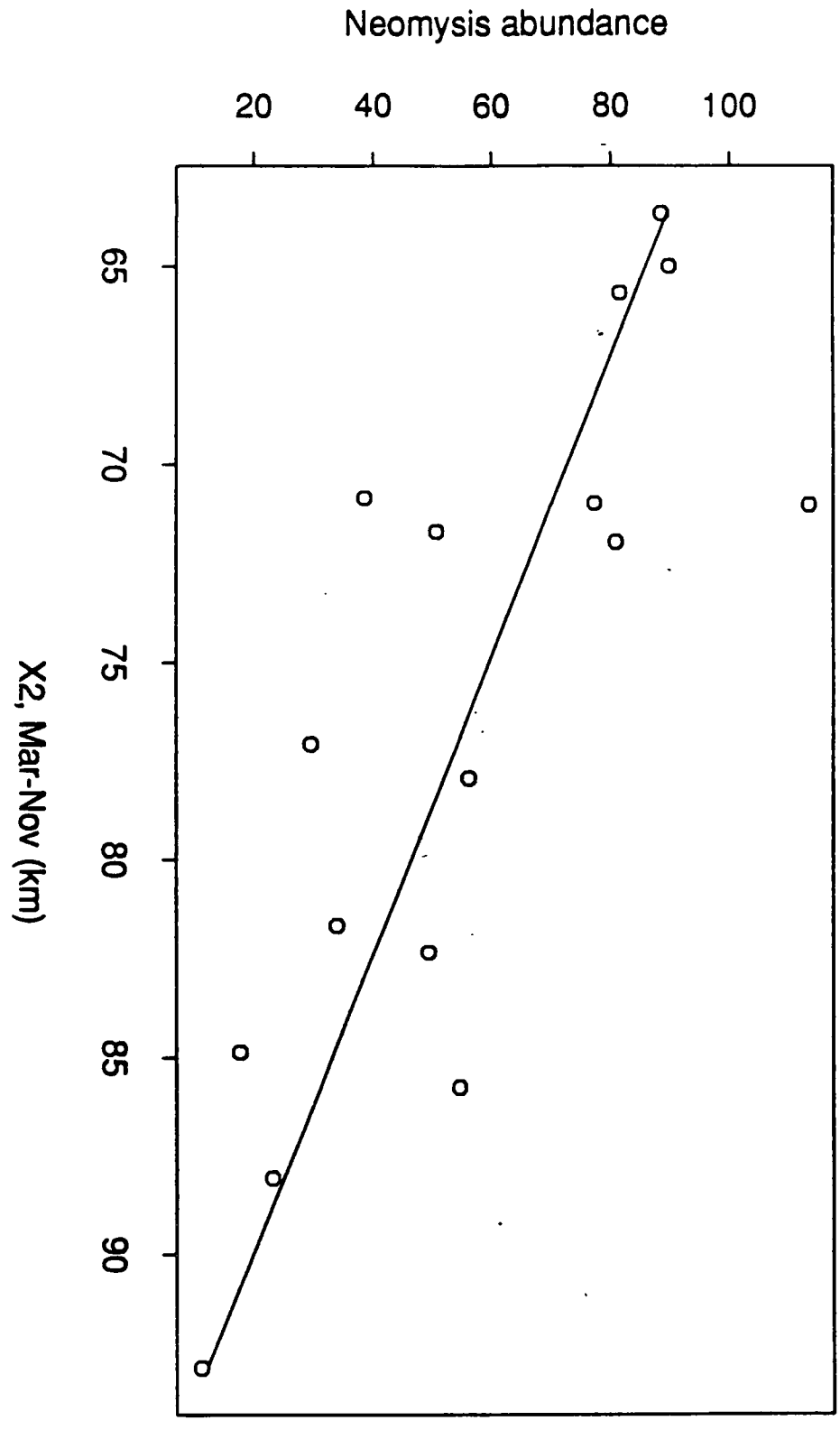
Fig. 4. Annual mean abundance of molluscs in Grizzly Bay, compared to X_2 averaged over the current and previous two years.

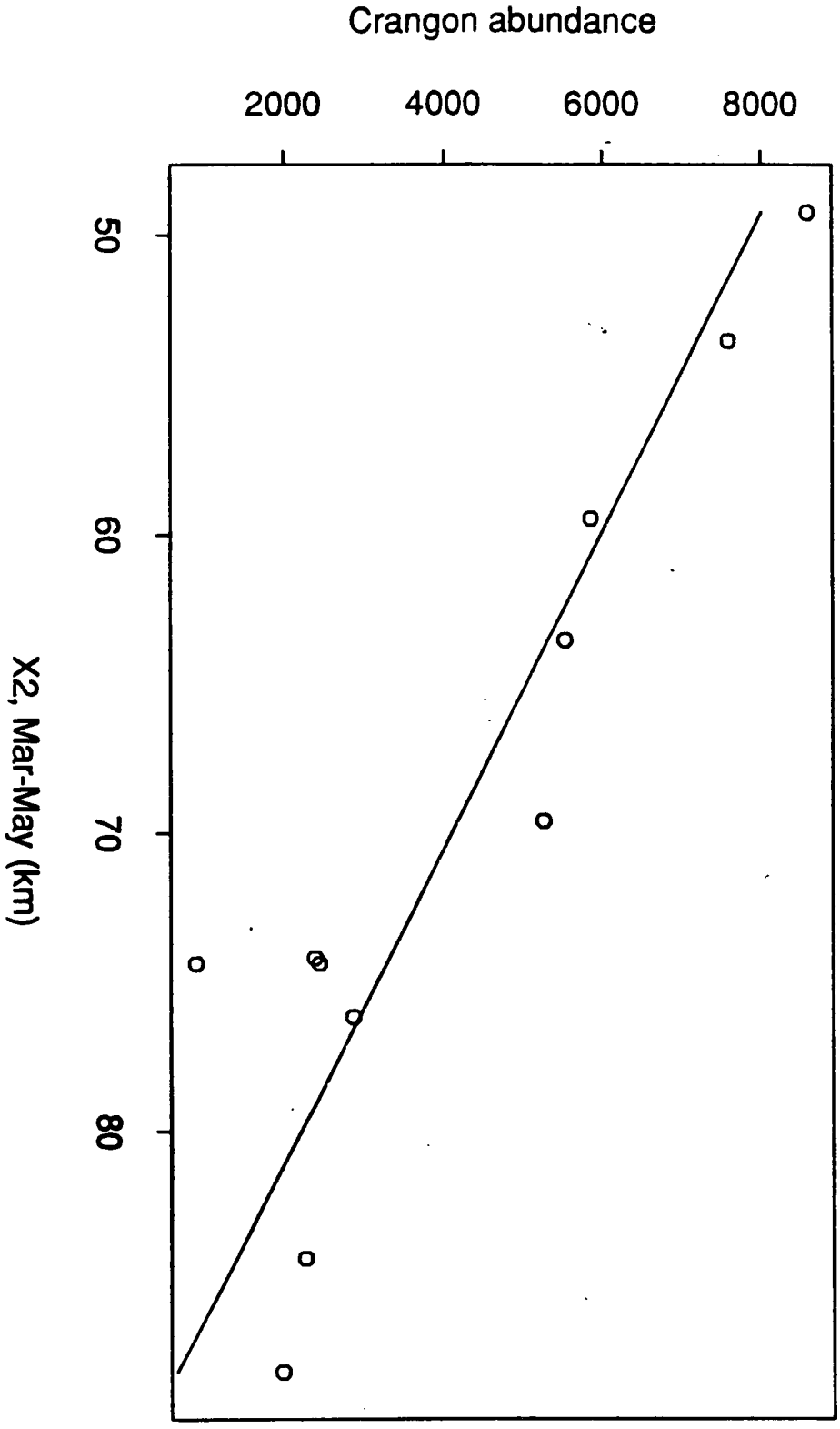
Fig. 5. Survival of striped bass (*Morone saxatilus*) from egg to adult (Peterson egg index to 38 mm index) in the San Francisco estuary.

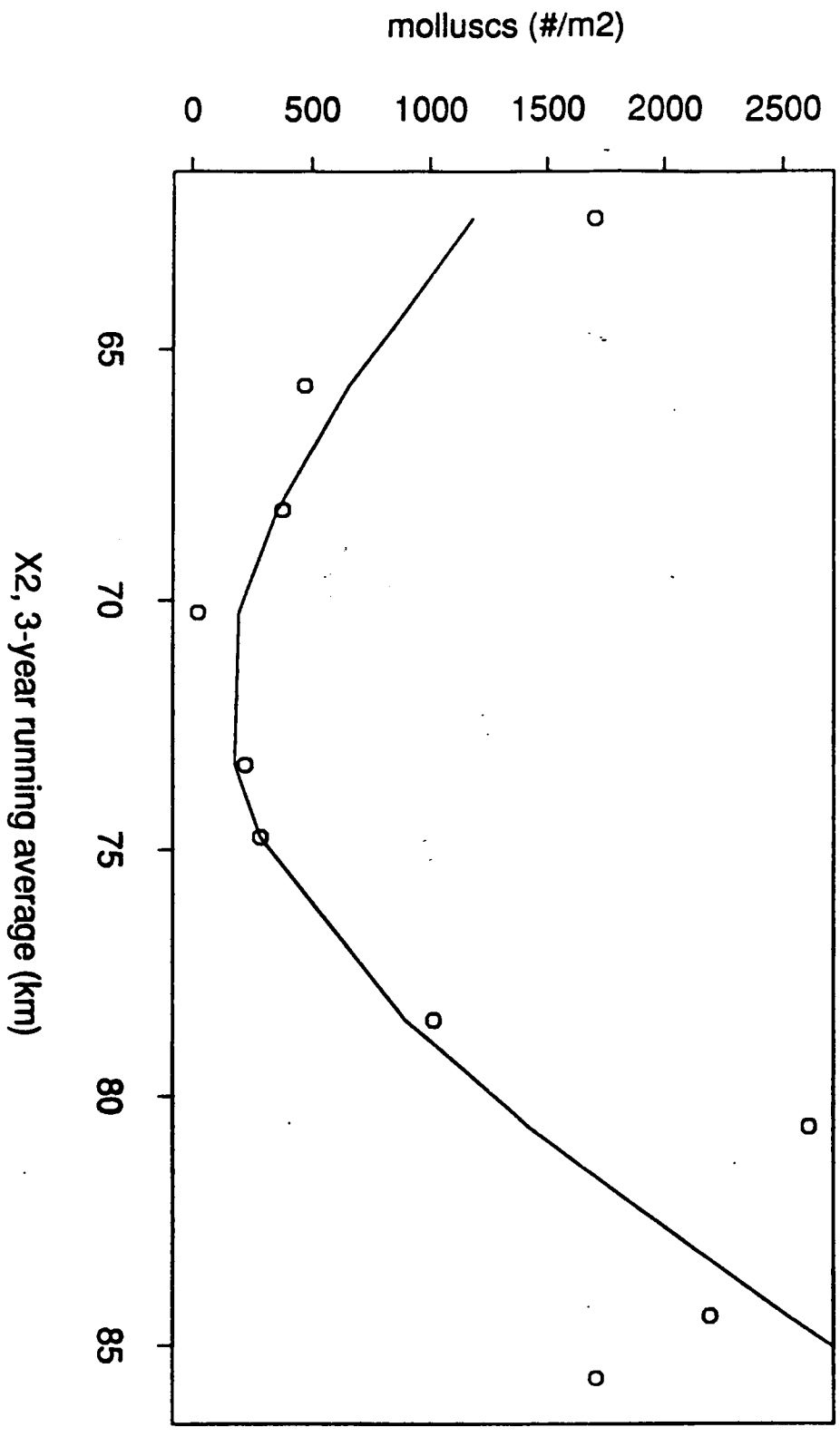
Fig. 6. Annual mean abundance index of starry flounder (*Platichthys stellatus*) year-class 1+ in the San Francsico estuary.

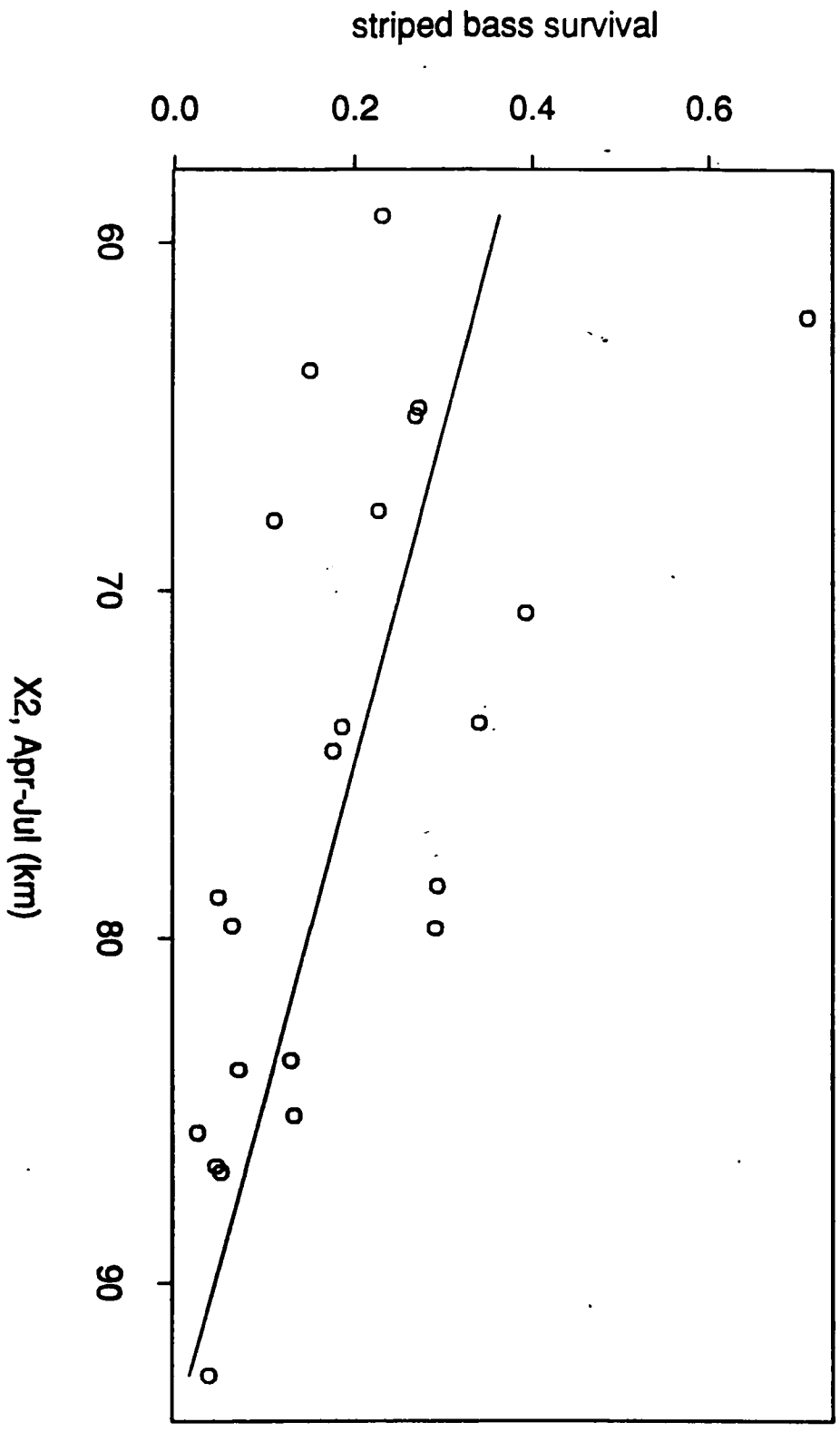
Fig. 7. Annual mean abundance index of longfin smelt (*Spirinchus thaleichthys*) YOY plus adults in the San Francsico estuary.

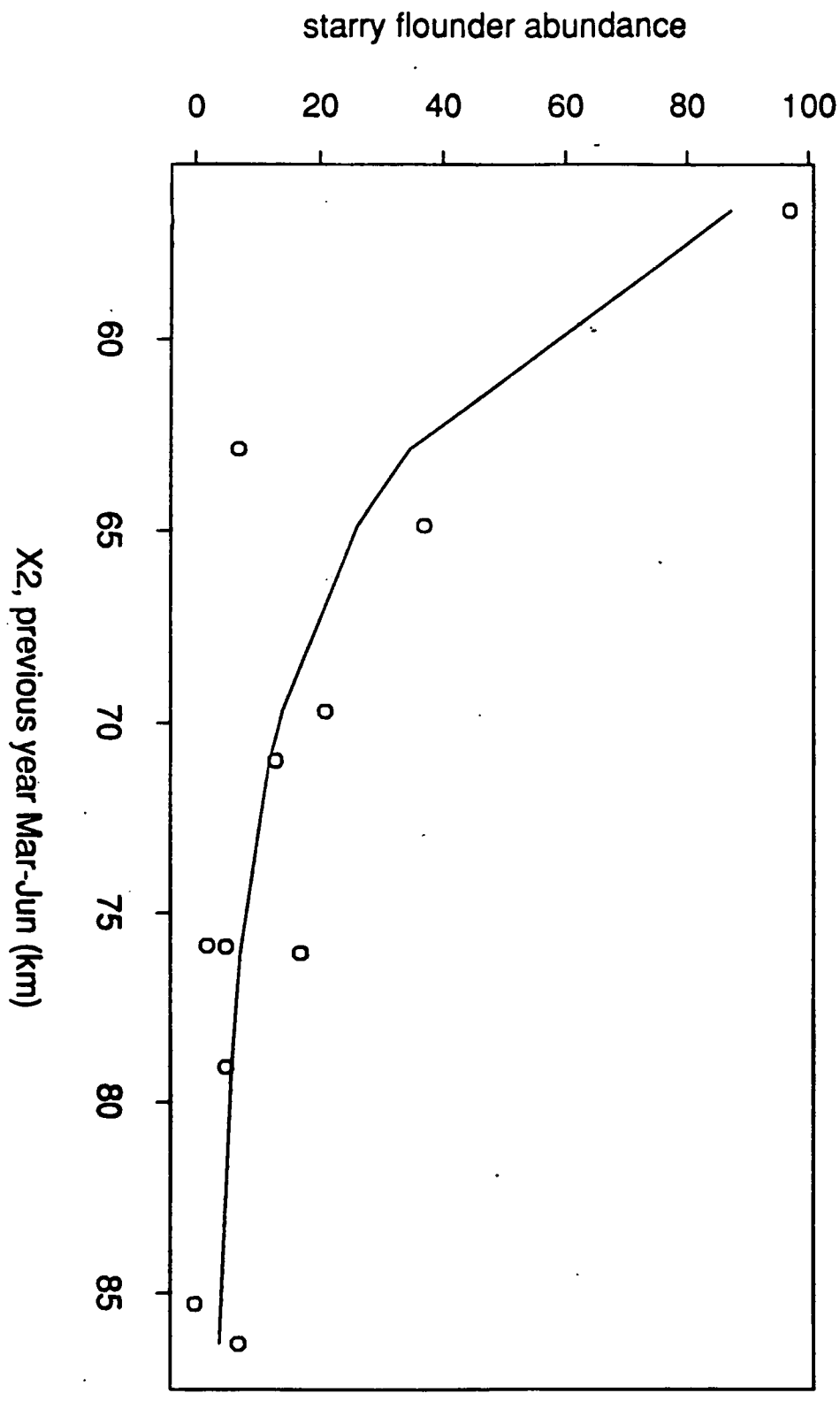

phytoplankton POC (Gg/yr)
5
10
15
60
70
80
90
X2, Jan-Dec (km)

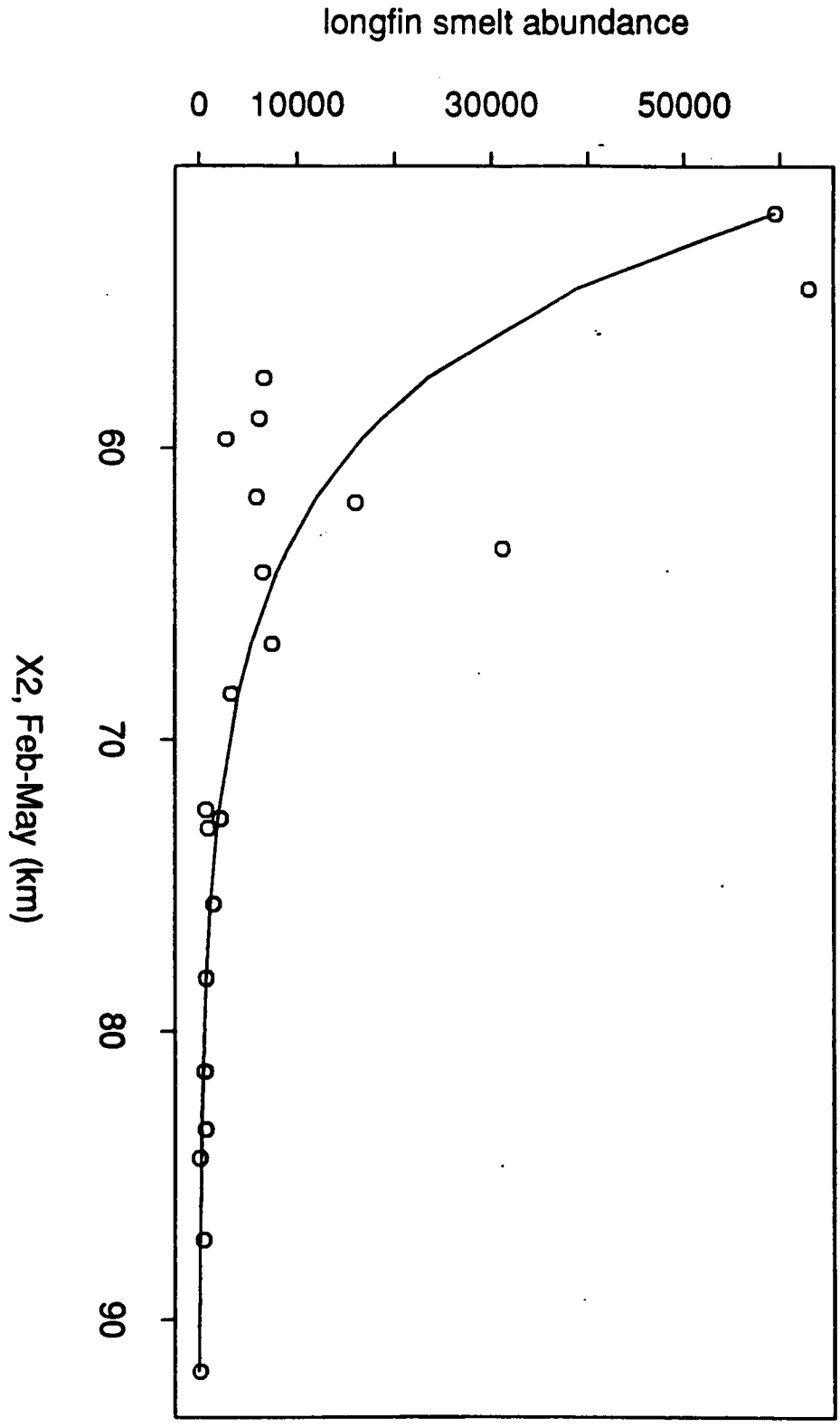

Neomysis abundance
20
40
60
80
100
X2, Mar-Nov (km)
65
70
75
80
85
90


Crangon abundance
2000
4000
6000
8000
X2, Mar-May (km)
50
60
70
80


molluscs (#/m2)
0
500
1000
1500
2000
2500
X2, 3-year running average (km)
65
70
75
80
85


striped bass survival
0.0
0.2
0.4
0.6
X2, Apr-Jul (km)
60
70
80
90


starry flounder abundance
0
20
40
60
80
100
X2, previous year Mar-Jun (km)
60
65
70
75
80
85


longfin smelt abundance
0
10000
30000
50000
X2, Feb-May (km)
60
70
80
90