

MULTIPLE REGRESSIONAL ANALYSIS FOR A MODEL OF ZOOPLANKTON PRODUCTION IN WATERS OFF THE WEST COAST OF INDIA

S. S. SHASTRI AND M. J. GEORGE

National Institute of Oceanography, Dona Paula, Goa-403 004, India

An attempt was made for the prediction of biomass of zooplankton by analysing the IIOE data collected from the region of the west coast of India between latitudes 5° to 20°N, and longitudes 70° to 75° E. For fitting the multiple regression curve of biomass against temperature, salinity and oxygen, the corresponding mean values of all these parameters for the entire column of water were taken. The pooled data for May were taken as the conditions representing SW monsoon period, while December was taken as the month for NE monsoon. Lines of regression for May and December were fitted by two methods. They were of the form:

$$\hat{y} = b_0 + b_1X_1 + b_2X_2 + b_3X_3 \dots \dots \dots \text{model 1}$$

$$\log \hat{y} = b'_0 + b'_1X_1 + b'_2X_2 + b'_3X_3 \dots \dots \dots \text{model 2}$$

where \hat{y} = estimated volume (in cc), X_1 = temperature (°c), X_2 = salinity (‰) and X_3 = oxygen (ml/l) and b_i , b'_i where $i = 1$ to 4 the coefficients. These coefficients were estimated.

The biomass of the area for the two periods has been estimated. Although the individual correlation coefficients between the volume and the three parameters were low, the multiple correlation coefficients and the first and second order correlation coefficients have shown improvements. In the case of model 1, 14 and 77 percent in the variation of biomass is due to the combined effect of these parameters. The explained variations show further improvement in the case of model 2. The remaining percentages probably refer to variation due to other parameters which have not been considered.

INTRODUCTION

The zooplankton biomass of the Arabian Sea and the Bay of Bengal are given in the International Indian Ocean Expedition (IIOE) plankton Atlases (IOBC, 1968). Making use of the data on biomass and environmental parameters collected from the Arabian Sea during IIOE (IOBC, 1969; 1971) it was attempted to fit a multiple regression model for the production of zooplankton. In order to get a fitting model the information on all the parameters which may have influence on the secondary production must be taken into account. Obviously to get a proper picture of the production at the second trophic level, the basic need is the information on primary productivity which is the first trophic level. Although some scattered information on primary production has been collected later on, no data on this is available along with the data collected during IIOE. Hence in the present contribution the data on the available environmental parameters like dissolved oxygen, salinity and temperature have been used to work out a model of the regime of the zooplankton production.

MATERIAL AND METHODS

The region selected for this study is the eastern Arabian Sea in between 5° to 20° and 70° to 75° E (Fig. 1.). The two months *viz.* May and December

were considered representing the SW monsoon (April 16 to October 15) and NE monsoon (October 16 to April 15) periods respectively and all the data available during these months have been pooled for the study.

Two different methods were used to find out the relationship between dependent variate 'y', the biomass and the three independent variates, viz. temperature (X_1), salinity (X_2) and dissolved oxygen (X_3). The two estimates of 'y' may be represented by the models:

$$\text{Model 1: } \hat{y} = \sum_{i=0}^3 B_i X_i + E$$

$$\text{Model 2: } \log \hat{y} = \sum_{i=0}^3 B'_i X_i + E'$$

where $X_0 = 1$; B_0, B'_0 are the intercepts when variables are zero;

B_1 or B'_1 is the normal change in biomass associated with unit change in average temperature that has the same salinity and oxygen;

B_2 or B'_2 and B_3 or B'_3 can be interpreted on the same line;

E or E' are the usual residual errors with the usual assumptions, i. e. Mean=0, Variance=constant etc.

Since the estimation procedure of different coefficients is the same in both the cases, it is explained in detail for the first model only.

The four parameters, namely, B_0, B_1, B_2 and B_3 were estimated by the corresponding statistic b_0, b_1, b_2 and b_3 by making use of the least square technique. The normal equations required for the same were obtained by minimising error sum of squares. Utilizing these normal equations and making use of inversion of matrix technique the elements of C - matrix were obtained. The regression coefficients b's can be obtained with the help of the elements of C-matrix. Thus

$$b_i = \frac{\sum_{j=1}^3 [C_{ii} \sum x_j y]}{\sum_{j=1}^3 [C_{ii} \sum x_j^2]} \quad (\text{for } j=1 \text{ to } 3 \text{ and the summation in bracket}$$

over the n observations) will give the different values of b's. Substituting these values of b's and putting $x_i = X_i - \bar{X}_i$ for $i = 1$ to 3 will give the regression line of biomass on temperature, salinity and oxygen, viz.

$$\hat{y} = \sum_{i=0}^3 b_i X_i \quad \text{where } X_0 = 1$$

The regression coefficients b's were tested by usual 't' test. For studying the effect of the parameters on biomass, zero order, first order and second order correlation coefficients were obtained. The tests of significance of second order correlation coefficients and multiple correlation coefficients were carried through the usual 't' and 'F' test.

RESULTS

The fitted curves of regression for the two periods were as follows:

Model of Zooplankton Production

Model 1 :

$$\text{May } \hat{y} = 102.27 + 1.53 X_1 - 3.27 X_2 - 4.99 X_3$$

$$\text{December } \hat{y} = 617.04 - 2.10 X_1 - 16.13 X_2 + 3.98 X_3$$

Model 2:

$$\text{May } \log \hat{y} = -4.22 + 0.19 X_1 + 0.08 X_2 - 0.82 X_3$$

$$\text{December } \log \hat{y} = 45.62 - 0.21 X_1 - 1.16 X_2 + 0.44 X_3$$

The respective regressional coefficients and the results of the tests of significance are shown in Table 1. It is clear from the table that the second model gives the better fit.

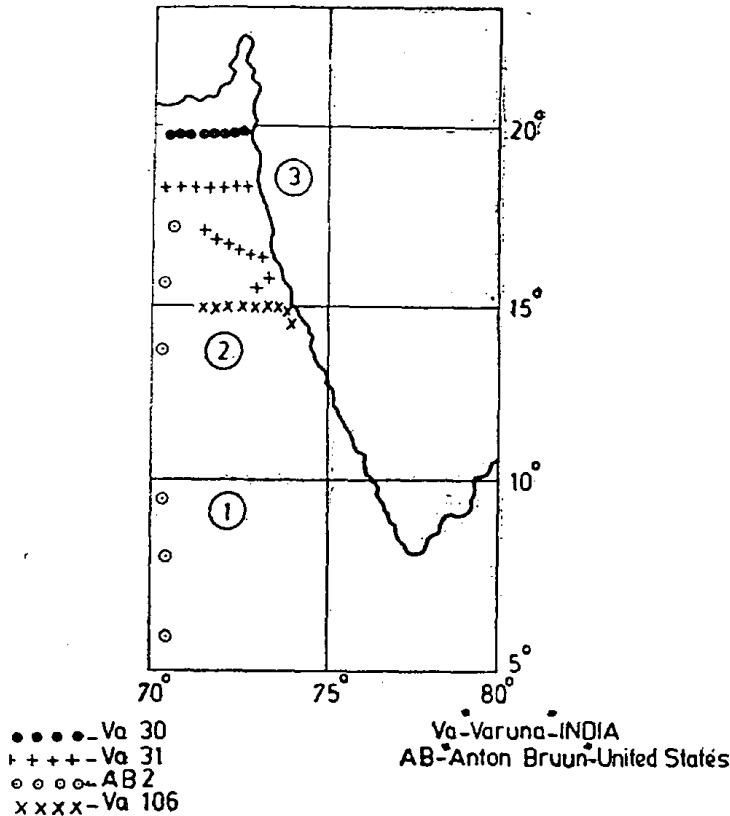


Fig. 1. Stations from which standard zooplankton samples were taken for the present study

The 'Analysis of variance' tables were prepared for testing the regressional fit and are presented in Table 2. The low values of F ratio in the case of model 1 indicates the 'poor' fitting of multiple regressional line. The use of logarithmic transformation in model 2 has sufficiently improved the situation and the F ratios were found out to be highly significant. This transformation leads to a close fit.

By making use of respective analysis of variance tables, multiple correlation coefficients were obtained. These coefficients along with their results of tests of significance are shown in Table 3. It is evident from this table that the explained variations in biomass due to the combined effect of the three para-

TABLE 1. DIFFERENT REGRESSIONAL COEFFICIENTS AND THE RESULTS OF THEIR TESTS OF SIGNIFICANCE

Model	Period	Regressional coefficients			Test of significance		
		b1	b2	b3	b1	b2	b3
1	May	1.53	-3.27	-4.99	*	NS	*
	Dec.	-2.10	-16.13	3.98	*	**	NS
2	May	0.19	0.08	-0.82	***	NS	***
	Dec.	-0.21	-1.16	0.44	***	***	**

*** Significant at 1% level

** Significant at 5% level

* Significant at 10% level.

NS-Not significant even at 10% level

TABLE 2. ANALYSIS OF VARIANCE TABLE*

Month	Source	d.f.	Model 1			Model 2		
			s.s.	m.s.s.	F ratio	s.s.	m.s.s	F ratio
May	Regression	3	157.63	54.54	1.26	3.86	1.29	12.9***
	Deviation	24	999.90	41.60		2.49	0.10	
	Total	27	1157.53			6.35		
Dec.	Regression	3	54.79	18.26	4.54*	0.36	0.12	12.0*
	Deviation	4	16.09	4.02		0.04	0.02	
	Total	7	70.88			0.40		

*Notations as in Table 1.

TABLE 3. MULTIPLE CORRELATION COEFFICIENTS AND THE RESULTS OF THEIR RESPECTIVE TESTS OF SIGNIFICANCE*

Model	Period	Multiple correlation coefficients (R)		Test of significance of R.
1.	May	0.37	(14 %)	NS
	Dec.	0.88	(77 %)	*
2.	May	0.78	(61 %)	***
	Dec.	0.95	(90 %)	**

*Notations as in Table 1

Model of Zooplankton Production

TABLE 4. ZERO, FIRST AND SECOND ORDER CORRELATION COEFFICIENTS AND THE RESULTS OF THEIR TESTS OF SIGNIFICANCE*

Model	Month	Biomass and	Zero order	First order		Second order
				after eliminating the effect of		
1.	May	Temperature	0.09 ^{NS}	Salinity	0.09 ^{NS}	Salinity and Oxygen 0.34* ^{NS}
		Salinity	0.10 ^{NS}	Temperature	0.09 ^{NS}	Temperature and Oxygen 0.13 ^{NS}
		Oxygen	0.11 ^{NS}	Temperature	0.34* ^{NS}	Temperature and Salinity 0.35*
	Dec.	Temperature	0.38 ^{NS}	Salinity	0.69* ^{NS}	Salinity and Oxygen 0.73*
		Salinity	0.41 ^{NS}	Temperature	0.70* ^{NS}	Temperature and Oxygen 0.82* ^{NS}
		Oxygen	0.34 ^{NS}	Temperature	0.07 ^{NS}	Temperature and Salinity 0.67 ^{NS}
2.	May	Temperature	0.13 ^{NS}	Salinity	0.12 ^{NS}	Salinity and Oxygen 0.67*** ^{NS}
		Salinity	0.11 ^{NS}	Temperature	0.10 ^{NS}	Temperature and Oxygen 0.07 ^{NS}
		Oxygen	0.54***	Temperature	0.78***	Temperature and Salinity 0.78***
	Dec.	Temperature	0.48 ^{NS}	Salinity	0.69* ^{NS}	Salinity and Oxygen 0.87** ^{NS}
		Salinity	0.26 ^{NS}	Temperature	0.61 ^{NS}	Temperature and Oxygen 0.55 ^{NS}
		Oxygen	0.36 ^{NS}	Temperature	0.31 ^{NS}	Temperature and Salinity 0.81*

* Notations as in Table 1.

meters considered has increased from 14.0 and 77.0 percent for May and December in model 1 to 61.0 and 90.0 percent for the respective periods in model 2. The results of the tests of significance of R also indicated that the estimates of R obtained through transformation in model 2 are more precise than those obtained in model 1.

The different partial correlation coefficients of zero, first and second order between biomass and the three parameters in the case of first model and between logarithmic biomass and the same three parameters in the case of model 2 were worked out. The result of their respective tests of significance and their mod values are given in Table 4. It shows that there is an improvement in the relationship from zero order to second order correlation coefficients. The first order correlation coefficients give us an idea for assessing the factor that helps in bringing the improvement.

TABLE 5. PARAMETERS AFFECTING CORRELATIONS

Model	Month	Relationship between Biomass and	Probable parameter
1.	May	Temperature Salinity Oxygen	Oxygen Oxygen Temperature
	Dec.	Temperature Salinity Oxygen	Salinity Temperature and Oxygen Salinity
2.	May	Temperature Salinity Oxygen	Oxygen Oxygen Temperature
	Dec.	Temperature Salinity Oxygen	Salinity Temperature and Oxygen Salinity

The parameters that probably may affect the improvement in the relationship between biomass (or logarithmic biomass in the case of model 2) and each of the three parameters are shown in Table 5. It is interesting to note that the probable parameters affecting the improvement in relationship remains the same for a particular month, irrespective of the fact whether logarithmic transformation for biomass is used or not.

The regressional curve in model 2 by making use of logarithmic transformation for biomass is definitely a better fit. Again among the two multiple regressional curves fitted for May and December in model 2, the curve for December seems to be closer, in which 90.0 percent variations in logarithmic value of biomass were due to the combined effect of temperature, salinity and oxygen. The remaining percentages probably refer to the variation due to other parameters which have not been considered. It is evident from the best possible fitted curve that for predicting biomass the factors, viz. salinity, temperature and dissolved oxygen, in order of preference, are important.

Model of Zooplankton Production

ACKNOWLEDGEMENTS

The authors are grateful to Dr. S. Z. Qasim, Director, N. I. O., Goa for constant encouragement.

REFERENCES

- IOBC. 1968. Maps on total zooplankton biomass in the Arabian Sea and the Bay of Bengal. In: Panikkar, N. K., (Ed.). *International Indian Ocean Expedition Plankton Atlas*, 1, (1 & 2), NIO, CSIR, New Delhi.
- IOBC. 1969. *Handbook to the International zooplankton collections. 1. Station list.* 129 pp. IOBC Natl. Inst. Oceanogr., CSIR, Cochin, India.
- IOBC. 1971. *Handbook to the International zooplankton collections. II. Environmental data.* IOBC, Natl. Inst. Oceanogr. CSIR, Cochin, India.