Bull. Environm. Contam. Toxicol. 27, 660–670 (1981)

Statistical Models for Estimating Seawater Metal Concentrations from Metal Concentrations in Mussels (*Mytilus edulis*)

J. D. Popham and J. M. D'Auria

Department of Chemistry, Simon Fraser University, Burnaby, British Columbia, Canada V5A 1S6

Direct measurements of trace metal concentrations in seawater for pollution monitoring reouire extensive and clean sample preparatory techniques, are time consuming, and require operation of the analytical apparatus at its limits of detection (i.e. ppb). Further, such measurements provide only estimates of an instantaneous seawater concentration which in turn can vary significantly over time, especially near sites releasing industrial effluents. Recent studies such as those by PENTREATH (1973), SCHULZ-BALDES (1973) and PHILLIPS (1976a) have shown that the mussel, *Mytilus edulis*, accumulates and concentrates (at the ppm level) trace metals, over time in proportion to the concentration of metals in seawater. As a result, many studies of estuarine pollution have used mussels as approximate indicators of the loadings of trace metals in the seawater (SCHULZ-BALDES 1973, PHILLIPS 1976a, DAVIES AND PIRIE 1978).

In principal then, mussels can provide a quantitative record of a metal polluting an estuarine environment over a reasonable period of time, eliminating the need and expense of instantaneous measurements, provided statistical models relating these quantities have been developed. It would appear though that little research in this direction has been attempted. In a laboratory study SCHULZ-BALDES (1974) derived a model for estimating lead concentrations in seawater knowing lead concentrations in the mussels, but the applicability of this approach to a multiparameter situation, i.e. the field has been questioned (WALDICHUK 1979). Recently POPHAM & D'AURIA (1981a) have demonstrated that it is possible to develop meaningful relationships between a widely varying instantaneous variable (seawater metal concentration) and a time-averaged variable (mussel metal concentration) in a field-based study, subjected to many varying factors. Because of the small sample size only bivariate regression models were developed and thus up to only 20% of the observed variance of the metal concentration in the seawater could be accounted for. In the present study over 700 mussels along with 25 seawater samples were collected in the southwestern coastal waters of S.W. British Columbia and used to develop multiple regression models. This report presents the regression coefficients of these models for estimating the copper, zinc and lead concentrations in seawater based upon the metal concentrations in mussels along with their accuracy and reliability.

MATERIALS AND METHODS

Details of collection and analyses of water samples and mussels are

0007-4861/81/0027-0660 $02.20

given in POPHAM & D'AURIA (1981a,b). In summary, over a twelve-month period, mussels were systematically collected near the outfall of a storm water drainage system (labelled Seabus) and at another site (Rocky Point) further up the estuary. Samples were also collected on an occasional basis in other waters of British Columbia. Where possible, mussels and water samples were collected from floating wharves to avoid the effect of zonation or depth affecting the trace metal concentrations in the mussels (GRAHAM 1971, PHILLIPS 1976a) and to prevent the collection of sedimentary material derived from the suspension in the water column by wave action. Mussels were depurated for 24 h in clean seawater before their meats were removed from the shells after measuring their length, then weighed (wet weight), lyophilised, and weighed again (dry weight) before pulversising. 40 mg of this homogenised powder were pressed into pellets and concentrations of iron, copper, zinc, bromine, lead and strontium determined by X-ray energy spectroscopy (XES) as previously described (STUMP *et al* 1979). Water samples were collected in acid-cleaned polyethylene bottles opened 10 cm below the water surface. The concentrations of the same trace elements as suspended particulate material (SPM) were determined by analysing the residue left on a 0.22 μm Millipore filter through which had been passed 1 L of the water sample. Soluble trace metal concentrations (Sol) were determined by analysing an aliquot of Chelex-100 which was added to filtered seawater (POPHAM & D'AURIA 1981a).

The parameter used to express the effect of season was a circular transformation of the day of collection (January 1, 1979 representing day 1):

$$\text{Day}' = \sin(\text{Day} - 364) + \cos(\text{Day} - 364)$$

All data except Day′ were log-transformed (base 10) prior to statistical analyses. Occasionally the concentration of lead in the samples were non-detectable and assigned a value of zero. Hence for computational ease 1 was added to all values of lead concentrations. The methodology for the model development follows that given in DRAPER & SMITH (1966). The stepwise regression program P2R of the BMDP statistical package (DIXON & BROWN 1979) was used to develop the models for estimating copper, zinc and lead concentrations in the seawater. Stepwise regression was used because many of the variates are highly correlated (shell length and dry weight of the meats, for example) and hence would be redundent (POPHAM & D'AURIA 1981a,b). Stepwise regression is a method by which only those variates which make a significant contribution to the regression model are added to it. The criterion for addition and deletion for the models in this study were 3.5 and 3.49 respectively for the F ratio testing the significance of the regression model. The program P2R permits the option of forcing entry of the variables. This option was used to ensure that concentrations of copper, zinc, and lead in the mussels were entered as variables in the models even if the f-value to enter was lower than those of other variables. This was done with the hope of making the models more universal. In one instance for example, a preliminary development of the model for copper was such that one model for predicting copper levels in seawater just required the knowledge of lead concentration in the mussel and its day of collection.

The following variates were also used as potential estimators of trace metal concentrations in the seawater: concentrations of Fe, Br, Sr, and water content (1-dry weight/wet weight) in the mussel; shell length and dry weight of the meats (Wt); and season of collection (Day'). Variates estimating mussel size were entered on account of previous studies showing that trace metals in concentrations in mussels can be dependent upon their size (BOYDEN 1977). Other studies (PHILLIPS 1976, SIMPSON 1979, POPHAM & D'AURIA 1981a,b) have shown that trace metal concentrations in mussels can also be influenced by seasonal effects. Previously COSSA *et al.* (1979) suggested that accurate estimates of pollution in seawater may be more easily obtained using sexually immature mussels. While it was not possible to introduce age of the mussels as a variable, it was possible to develop models based on trace element concentrations in what were presumed to be immature mussels by using only those specimens whose shell length was less than 25 mm. Conversely models were also developed using only those mussels longer than 25 mm.

RESULTS AND DISCUSSION

A summary of the statistics for the six models (for each estimating SPM and Sol of the copper, zinc and lead concentrations in the seawater) are listed in Table 1. A variance covariance matrix (Table 2) and the mean values of the variates are also given so that confidence limits around an estimated variate can be calculated. The models for estimating the concentration of copper, zinc and lead as SPM and as soluble metal ions (Sol) in $\mu g \cdot \ell^{-1}$ are respectively:

$$\log[Cu_{SPM}] = -0.4523 - 0.2171 \log[Cu] - 0.3076 [Day'] + 1.5212 \log[Fe] + 0.4566 \log[Weight] - 0.5682 \log[Sr] - 0.108 \log[1+Pb]$$

$$\log[Cu_{Sol}] = 1.9218 + 0.8192 \log[Cu] - 0.5334 \log[Fe] + 0.1012 [Day'] - 0.3691 \log[Zu]$$

$$\log[Zn_{SPM}] = -0.434 + 0.4862 \log[Zn] - 0.1395 \log[1+Pb] + 0.341 \log[Fe] + 5.6458 \log[Water]$$

$$\log[Zn_{Sol}] = 0.2396 + 0.2001 \log[Zn] + 7.389 \log[Water] + 0.6587 \log[Fe] - 0.0804 [Day'] + 0.5458 \log[Sr]$$

$$\log[Pb_{SPM}+1] = -1.0487 + 0.1116 \log[1+Pb] + 0.7097 \log[Fe]$$

$$\log[Pb_{Sol}+1] = 1.1435 + 0.1353 \log[1+Pb] - 11.453 \log[Water]$$

Figure 1 shows the agreement between the measured data (instantaneous measurement) and the estimated concentrations (averaged value). The figure shows that the models are biased such that high concentrations of the metals measured in the seawater are underestimated by the models while low measured values are overestimated.

TABLE 1. Summary statistics of multiple regression analysis for estimating indicated variate in seawater against variates measured in mussels.

Seawater Variate	Variable Entered	Multiple R^2	Standard Error	F-to-Enter	Regression Coefficient	Mean
$[Cu_{SPM}]$	Y-intercept		0.4685		-0.4523	
	Cu	0.0194	0.4662	2.06	-0.2171	1.3517
	Day'	0.1175	0.4444	11.44	-0.3076	0.0972
	Fe	0.2388	0.4147	16.26	1.5212	2.7835
	Wt	0.3343	0.3898	14.50	0.4566	-1.5355
	Sr	0.3366	0.3821	5.09	-0.5682	2.4506
	P+1	0.0897	0.3769	3.75	-0.1080	1.2594
$[Cu_{Sol}]$	Y-intercept		0.3281		1.9218	
	Cu	0.3974	0.2559	65.58	0.8192	1.3517
	Fe	0.4939	0.2356	19.65	-0.5334	2.7835
	Day'	0.5391	0.2260	9.99	0.1012	0.0972
	Zn	0.5800	0.2168	9.84	-0.3691	2.8371
$[Zn_{SPM}]$	Y-intercept				-0.4340	
	Zn	0.1138	0.3091	13.99	0.4862	2.8371
	Pb+1	0.1526	0.2871	4.95	-0.1395	1.2594
	Fe	0.1944	0.2813	5.55	0.3410	2.7835
	Water	0.2237	0.2774	4.01	5.6458	-0.0784
$[Zn_{Sol}]$	Y-intercept		0.2706		0.2396	
	Zn	0.2967	0.2280	43.87	0.2001	2.8317
	Water	0.3663	0.2175	11.31	7.3898	-0.0784
	Fe	0.4209	0.2089	9.61	0.6587	2.7835
	Day'	0.5010	0.1949	16.22	-0.0804	0.0972
	Sr	0.5458	0.1869	9.87	-0.3849	2.4506
$[Pb_{SPM}]$	Y-intercept		0.5403		-1.0487	
	Pb+1	0.1479	0.5011	18.05	0.1116	1.2594
	Fe	0.2211	0.4814	9.69	0.7097	2.7835
$[Pb_{Sol}]$	Y-intercept		0.4535		1.1435	
	Pb+1	0.4111	0.4154	21.16	0.1353	1.2594
	Water	0.4762	0.4026	7.69	11.4501	-0.0784

TABLE 2. Variance-covariance matrix values from regression analyses.

Seawater Variate		Variance-Covariance Matrix					
$[Cu_{SPM}]$		Cu	Day'	Fe	Wt	Sr	Pb+1
	Cu	0.1548					
	Day'	0.0854	0.8835				
	Fe	0.0613	0.1202	0.0630			
	Wt	-0.0437	0.0187	-0.0287	0.0999		
	Sr	0.0229	0.0673	0.0239	-0.1358	0.0334	
	Pb+1	0.2419	0.1035	0.1362	-0.0093	0.0431	0.9030
$[Cu_{Sol}]$		Cu	Fe	Day'	Zn		
	Cu	0.1548					
	Fe	0.0613	0.0630				
	Day'	0.0854	0.1202	0.8835			
	Zn	0.0580	0.0450	0.1062	0.0676		
$[Zn_{SPM}]$		Zn	Pb+1	Fe	Water		
	Zn	0.0676					
	Pb+1	0.1925	0.9030				
	Fe	0.0450	0.1362	0.0630			
	Water	0.0015	0.0048	0.0008	0.0001		
$[Zn_{Sol}]$		Zn	H_2O	Fe	Day'	Sr	
	Zn	0.0676					
	Water	0.0011	0.0001				
	Fe	0.0450	0.0008	0.0630			
	Day'	0.1062	0.0016	0.1202	0.8835		
	Sr	0.0113	0.0004	0.0673	0.0673	0.0334	
$[Pb_{SPM}]$		Pb+1	Fe				
	Pb+1	0.9030					
	Fe	0.1362	0.0630				
$[Pb_{Sol}]$		Pb+1	H_2O				
	Pb+1	0.9030					
	H_2O	0.0048	0.0001				

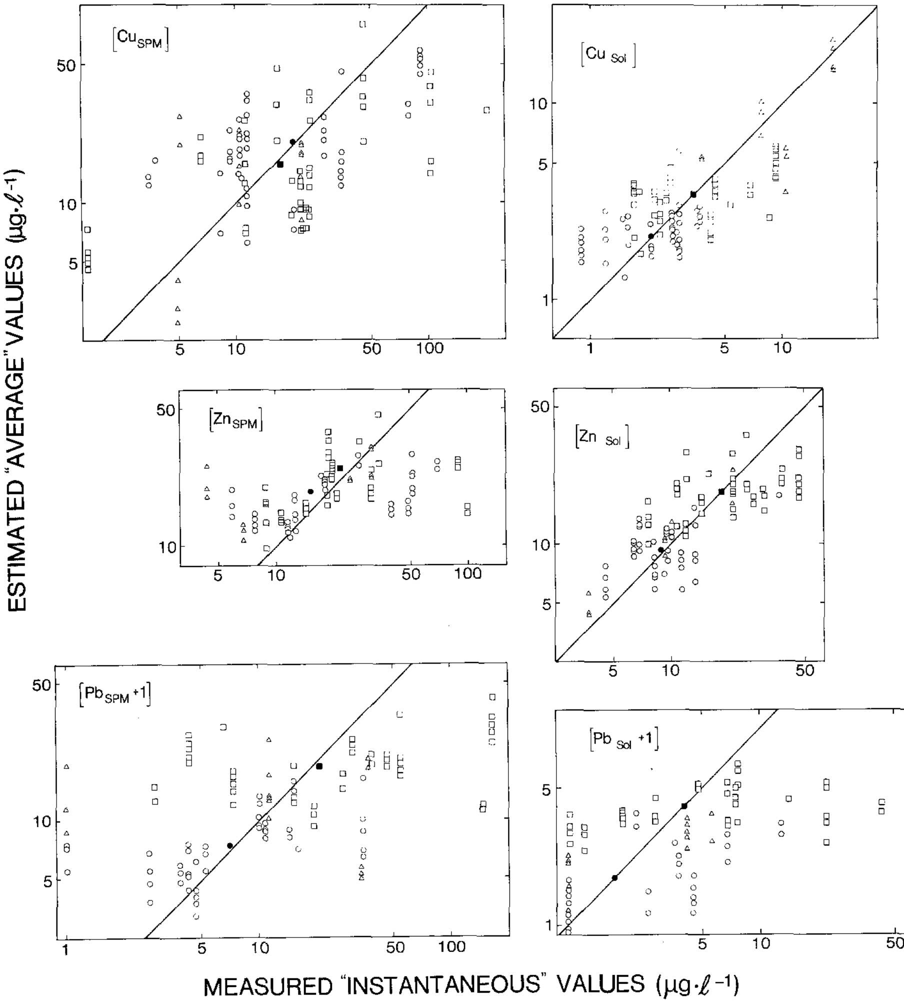

Figure 1

As indicated these figures display the measured values for the total concentration of elemental Cu, Zn, and Pb as either suspended particulate matter (SPM) or soluble ions (Sol) as a function of the values estimated by the statistical models discussed in the text. The solid lines represent perfect agreement. The data points are for mussels taken from Rocky Point (○), Seabus (□) and other locations in B.C. (Δ) with (●) and (■) indicating average values for the RP and SB samples, respectively.

An explanation for the bias is presented later. Of the two size ranges of the mussels tested better fits between the measured and estimated data were obtained with the models in which only small mussels were used rather than the large mussels.

Models developed for estimating the concentration of the soluble component of the metals were able to account for a higher proportion of the variances (had higher multiple R2 values; see COOLEY & LOHNES 1971) than the ones for estimating the concentration of the suspended particulate component.

While it is hazardous to evaluate the importance of the variables in the models on account of the correlations among these variables (COOLEY & LOHNES 1971), some comment on the selection of some of the variates seems necessary.

A weakness associated with the model developed for estimating copper as SPM is that the regression coefficient is negative implying that the concentration of copper decreases in the mussels with increasing copper concentrations in the seawater! Apparently what is occurring is an interaction among the variables such that when lead concentrations are high there is a correspondingly high concentration of copper in the mussel regardless of whether or not high concentrations of copper are actually in the seawater. Presumably the same explanation applies with respect to a copper-zinc interaction in the model developed for estimating the concentration of soluble copper.

Five of the six models include iron as an estimating variable and yet iron concentrations in the sea water are not significantly different at Seabus and Rocky Pt., the sites from where the majority of the samples were collected (POPHAM & D'AURIA 1981a).

The next most common variables were Day′ (indicating seasonal effects), and water content in the mussel, both occurring in three of the six models. Since Day′ was entered in both models estimating seawater copper concentrations, it would appear that copper concentrations, either in the mussels or in the seawater, vary according to season as previously found by SIMPSON (1979).

A possible explanation for mussel water concentration being entered in the models is that mussels from Seabus not only had high concentrations of zinc and lead but they also had high concentrations of water compared with mussels collected from other areas.

The question can be raised as to whether better estimations of the metal concentrations could be made. As previously mentioned the models are biased, presumably because mussels have long time constants with respect to trace metal accumulation (PENTREATH 1973, SCHULZ-BALDES 1974). While long time constants are advantageous in determining time-integrated estimates of the "average" concentration of the trace metals in the seawater they preclude accurate estimates of "instantaneous" concentrations in the seawater. As a result, deviations above the mean values in the seawater will be underestimated and vice versa. Evidence to support the suggestion that the model estimates the mean value of the concentrations are as follows. If the average trace metal concentration of the samples collected from Seabus and Rocky Pt. are plotted against the average values for the estimators the two points are found to lie very close to a diagonal line representing perfect agreement between the measured data and estimated values. Hence it seems likely that the models can be improved by entering some easily measured value or values reflecting instantaneous deviations from the mean. Thus the models would consist of two components; one component estimating the "average" concentration of the metal in the seawater and the other component measuring the "deviation from the average"; namely

$$\text{Estimated [Metal]} = a + b1(\text{"average estimate"}) + b2(\text{"deviations from average"}) .$$

A number of variates, such as measurements of turbidity or salinity could be used to provide a parameter reflecting deviations due to environmental effects. Unfortunately at the time of the collections none of these data were measured. (As an aside, one purpose of the model building is to discover new variables that should be considered for helping to uncover the nature of the relationships between variables. In this case the procedure fulfilled its expectations very well.) However, one set of data which can be used to support this contention is the use of the seawater concentration of iron as SPM. There is no significant difference between iron concentrations between Rocky Pt. and Seabus (POPHAM & D'AURIA 1981a), but the concentrations do vary over a wide range and can be correlated with the amount of rainfall over a 96 h period prior to the time of water sampling ($r = 0.51$). Therefore in order to test this hypothesis a preliminary model, in which SPM iron in the seawater was used as the variate to represent instantaneous fluctuations due to such conditions as surface runoff, was developed for Pb_{SPM} concentrations: Table 3 and Fig. 2 show the results of this approach. A comparison of Fig. 2 with Fig. 1 shows an improvement in the model with the plots of the estimates vs the measured data being rotated anticlockwise and lying more closely on the diagonal of perfect fit.* A better improvement can be made if all variates are treated separately, but it then becomes impossible to separate out the estimate of the "average soluble lead concentration" from the estimate of the "deviation from the average" due to environmental factors.

CONCLUSIONS

The results of this study have shown that field studies can be used to develop statistical models for estimating the concentration of trace metals in seawater knowing the concentration of metals in the mussels. Although only about 50% of the variance of the soluble

---

*This attempt to improve the model by introducing a variable reflecting instantaneous changes increased by over 14% to the proportion of the variance of suspended particulate lead accounted for by the model.

metal and about 20% of the metal as SPM can be accounted for by the models there is no reason to doubt that better models can be obtained in the future if parameters which reflect instantaneous effects are included. Such parameters could be measurements of the freshwater input as reflected by salinity and turbidity measured with sechi disk readings. Obviously one of the important variables to consider is salinity as both DAVENPORT (1977) and PHILLIPS (1976) have shown that this variable affects copper and zinc uptake. These parameters could help solve the problem encountered in this study in which a constant bias is seen in the results: low concentrations of metals in the seawater are overestimated and high values are underestimated. Nevertheless the models presented in this paper can be considered as a first step in estimating time-integrated concentrations of trace metals using easily measured parameters in mussels collected in a natural environment.

TABLE 3. Summary statistics of multiple regression analysis for estimating the concentration of lead as suspended particulate matter in seawater using variates measured in mussels and the concentration of iron as suspended particulate matter in seawater.

Model: Estimated $[Pb_{SPM}+1] = Y + b_1[\text{Average } Pb_{SPM}+1] + b_2$ ("deviation from average)

Variable Entered	Multiple R^2	Standard Error	F-to-Enter	Regression Coefficients	Means
Y-intercept		0.5403		-1.6996	
Estimated average $[Pb_{SPM}+1]$	0.2211	0.4791	29.52	0.7195	3.0645
$[Fe_{SPM}]$	0.3671	0.4340	23.76	0.6523	1.0673

Variance-Covariance Matrix

	"Average" $[Pb_{SPM}+1]$	$[Fe_{SPM}]$
"Average"$[Pb_{SPM}+1]$	0.0645	
$[Fe_{SPM}]$	0.0278	0.1121

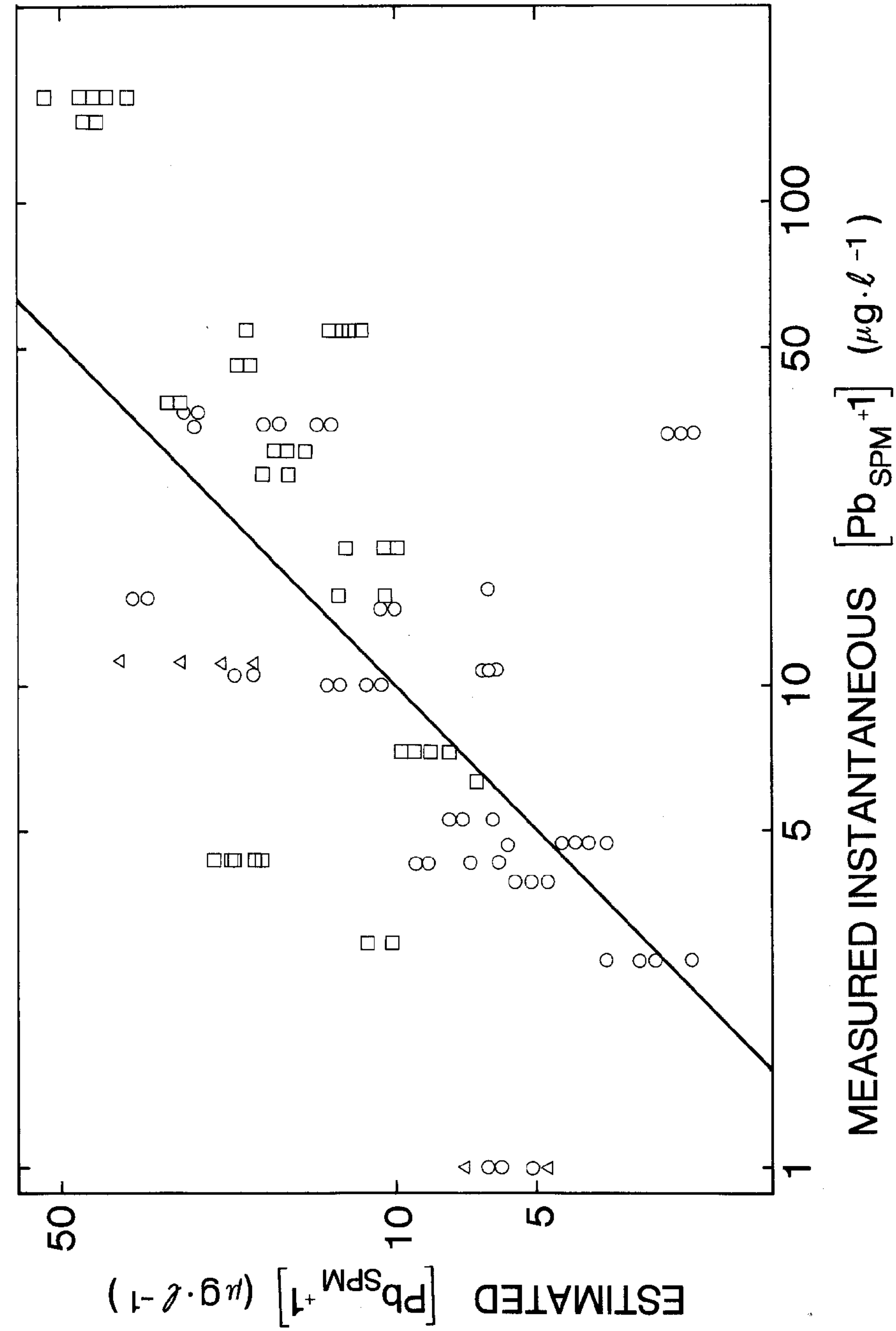

Figure 2

Similar to Fig. 1 with the inclusion of an additional term $[Fe_{SPM}]$ in the statistical model estimating the lead concentration as SPM (see text).

REFERENCES

BOYDEN, C.R.: J. mar. biol. Ass. U.K. 57, 675 (1977).

COSSA, D., E. BOURGET, & J. PIUZE: Mar. Pollut. Bull. 10, 174 (1979).

COOLEY, W.W., & P.R. LOHNES: Multivariate Data Analysis. New York: John Wiley & Sons 1971.

DAVENPORT, J.: J. mar. biol. Ass. U.K. 57, 63 (1977).

DAVIES, I.M., & J.M. PIRIE: Mar. Pollut. Bull. 9, 128 (1978).

DIXON, W.J., & M.B. BROWN: Biomedical Computer Progress, P-Series, Berkeley, U. Calif. Press.

DRAPER, N.R., & H. SMITH: Applied Regression Analysis. New York: John Wiley & Sons 1966.

GRAHAM, D.: Veliger, 14, 365 (1971).

PENTREATH, R.J.: J. mar. biol. Ass. U.K. 53, 127 (1973).

PHILLIPS, D.J.H.: Mar. Biol. 58, 59 (1976a).

———:Mar. Biol. 58, 71 (1976b).

POPHAM, J.D., & J.M. D'AURIA: Archives Environ. Toxicol. (1981a), (in press).

———: Archives Environ. Toxicol. (1981b), (submitted).

SCHULZ-BALDES, M.: Mar. Biol. 21, 98 (1973).

———: Mar. Biol. 25, 177 (1974).

SIMPSON, R.D.: Mar. Pollut. Bull. 10, 74 (1979).

STUMP, I.G., J. KEARNEY, J.M. D'AURIA & J.D. POPHAM: Mar. Pollut. Bull. 10, 270 (1979).

WALDICHUK, M.: Phil. Trans. R. Soc. Lond. B. 286, 899 (1979).

Accepted August 22, 1981