The Design of Sampling Transects for Characterizing Water Quality in Estuaries

A. D. Jassby[a], B. E. Cole[b] and J. E. Cloern[b]

[a]Division of Environmental Studies, University of California, Davis, CA 95616, U.S.A.
[b]U.S. Geological Survey, Water Resources Division, 345 Middlefield Road, Menlo Park, CA 94025, U.S.A.

The high spatial variability of estuaries poses a challenge for characterizing estuarine water quality. This problem was examined by conducting monthly high-resolution transects for several water quality variables (chlorophyll a, suspended particulate matter and salinity) in San Francisco Bay (California, U.S.A.). Using these data, six different ways of choosing station locations along a transect, in order to estimate mean conditions, were compared. In addition, 11 approaches to estimating the variance of the transect mean when stations are equally spaced were compared, and the relationship between variance of the estimated transect mean and number of stations was determined. The results provide guidelines for sampling along the axis of an estuary: (1) choose as many equally-spaced stations as practical; (2) estimate the variance of the mean $\bar{y}$ by var $(\bar{y}) = (1/10n^2)\sum_{j=2}^{n} (y_j - y_{j-1})^2$, where $y_1, \ldots, y_n$ are the measurements at the n stations; and (3) attain the desired precision by adjusting the number of stations according to $\mathrm{var}(\bar{y}) \propto 1/n^2$. The inverse power of 2 in the last step is a consequence of the underlying spatial correlation structure in San Francisco Bay; more studies of spatial structure at other estuaries are needed to determine the generality of this relationship.   © 1997 Academic Press Limited

Keywords: sampling design; spatial structure; water quality; chlorophyll distribution; suspended particulate matter; salinity; geostatistics; San Francisco Bay

Introduction

Estuaries present unusual difficulties in characterizing the spatial distributions of the properties that collectively define water quality — nutrients, dissolved gases, trace contaminants, suspended sediments, salinity and plankton populations. Large-scale patterns of spatial variability include the longitudinal salinity gradient along the continuum between the estuarine drainage basin and the coastal ocean. Superimposed onto this trend are sources of smaller-scale spatial variability, including distributed point sources; features of water circulation such as fronts, eddies or convergences that create localized turbidity maxima (e.g. Peterson et al., 1975); patchiness resulting from irregularities in bottom topography (e.g. Powell et al., 1986); and biologically-mediated spatial differences in processes such as primary production and biogeochemical transformations of reactive constituents (e.g. Jassby et al., 1993; Cloern, 1996). Many of these sources of spatial variability are unique to or amplified for estuaries.

At the same time, by virtue of the large human populations often associated with estuaries, anthropogenic impacts on water quality are strong and the need for characterizing ambient conditions and temporal

trends in these conditions is correspondingly urgent. The variability inherent in estuaries implies that a greater sampling effort is often necessary to describe water quality adequately, compared to other aquatic systems. The question of how to sample the spatial extent of estuaries most efficiently arises naturally, whether the objective is to describe current conditions or temporal trends in these conditions. Historically, most station configurations in estuaries, and arguably in most other aquatic ecosystems as well, have been chosen on the basis of surface physiographic features or by a cursory knowledge of spatial heterogeneity. These configurations may very well turn out to be near-optimal in some useful sense, but there is no way to tell without a more objective approach.

This paper considers the general question: how should samples be taken in an estuary or subembayment so that regional properties (e.g. mean concentration or mean population abundance) can be compared from one time period to another or from one subregion to another? Although this is perhaps the simplest form of trend detection (the underlying goal of most monitoring and assessment programmes), it is a significant issue for several reasons. First, for certain important water quality variables, the regional (estuary-wide) or subregional (subembayment) mean

provides an informative scalar index of ambient conditions. We want to know, for example, if the trophic state of an estuary (as indexed by chlorophyll a) is exhibiting a positive temporal trend, or if a trace contaminant is higher in one subembayment than another. Second, use of the mean enables one to connect empirical observations in the estuary to a large body of results from sampling theory and geostatistics. This connection supports the aim of providing a conceptual framework for both understanding the observations and generalizing them to other water bodies. Finally, the regional and subregional means provide an important way of communicating estuarine conditions to the public and environmental managers, precisely because the mean is so simple and widely understood.

The answer to the question posed here depends on the (usually unknown) spatial structure of the water quality measurements of interest. As spatial structure differs among the different components of water quality (Powell et al., 1989), the present work followed the lead of previous estuarine studies (Madden & Day, 1992; Childers et al., 1994) in choosing three separate but complementary water quality indicators: salinity, suspended particulate matter (SPM) and chlorophyll a. Salinity is a conservative tracer of mixing along the river–ocean continuum and therefore a surrogate for longitudinal processes. Suspended particulate matter, strongly affected by rapid exchange between the water column and bottom sediments, is a surrogate for vertical processes. Chlorophyll a, a measure of phytoplankton biomass and a representative non-conservative constituent that quickly responds to spatially-variable sources and sinks, often reflects lateral processes (Huzzey et al., 1990). As the spatial structures of these different components change with time, the measurements were repeated at monthly intervals. The sampling programme was conducted over an annual period in San Francisco Bay, a complex estuarine system that exhibits all modes of spatial-temporal variability expected in shallow coastal ecosystems influenced by tidal, wind, riverine and anthropogenic effects (Cloern & Nichols, 1985).

Site description

The San Francisco Estuary or ' Bay-Delta ' consists of a landward, tidal freshwater region known as the Delta and a seaward region known as San Francisco Bay (Figure 1). The Delta is a highly dissected region of channels and islands where the Sacramento, San Joaquin and other rivers coalesce and narrow as they flow westward. The outflow from the Delta passes through a narrow notch in the Coast Range into a series of subembayments, and ultimately through a narrow deep trough- -the Golden Gate—into the Pacific Ocean. Four major subembayments are usually recognized: South, Central, San Pablo and Suisun Bays. Together they constitute San Francisco Bay, the largest coastal embayment on the Pacific coast of the United States. Ninety percent of the freshwater input into the Bay flows through the Delta from regional drainage; the remainder is supplied by local tributaries. The drainage basin of the estuary encompasses 40% of California's land area. River inputs are highly seasonal, consisting of rainfall during autumn and winter, and snowmelt during spring and early summer. In addition to this dependence on climate, flow is affected by a series of upstream reservoirs that are managed for agriculture, power, flood control and repulsion of salinity intrusions. A large portion of the flow reaching the Delta is diverted, mostly for agricultural purposes, before it can reach the Bay.

Water quality problems in the Bay-Delta are multiple, complex and linked in various ways. A major underlying issue is management of freshwater inflow, which affects estuarine population abundances both directly, through transport, and indirectly, through effects on salinity and other variables (Jassby et al., 1995). Contaminants include sediments and metals introduced from mining operations, domestic sewage, persistent and toxic trace substances from industrial discharge and urban runoff, and biocides in agricultural drainage (Davis et al., 1991). Occasional high chlorophyll concentrations and threats of harmful algal blooms are also of concern (Jassby et al., 1994). Several large monitoring efforts are in place with the goals of assessing existing water quality, determining trends in trace contaminants and population abundances, and exploring the underlying causal processes. The size of these programmes, the social importance of the water quality problems and the extreme variability of the estuary all demand a closer and more objective examination of the sampling effort.

General approach

This paper considers here only the longitudinal variability along the central channel that connects the seaward and landward domains of the San Francisco Bay system. By using variables that together reflect all three spatial dimensions, however, these observations in the estuarine channel encompass processes occurring upstream, in adjacent marshes and lateral shoals, due to point source discharges, and within the local
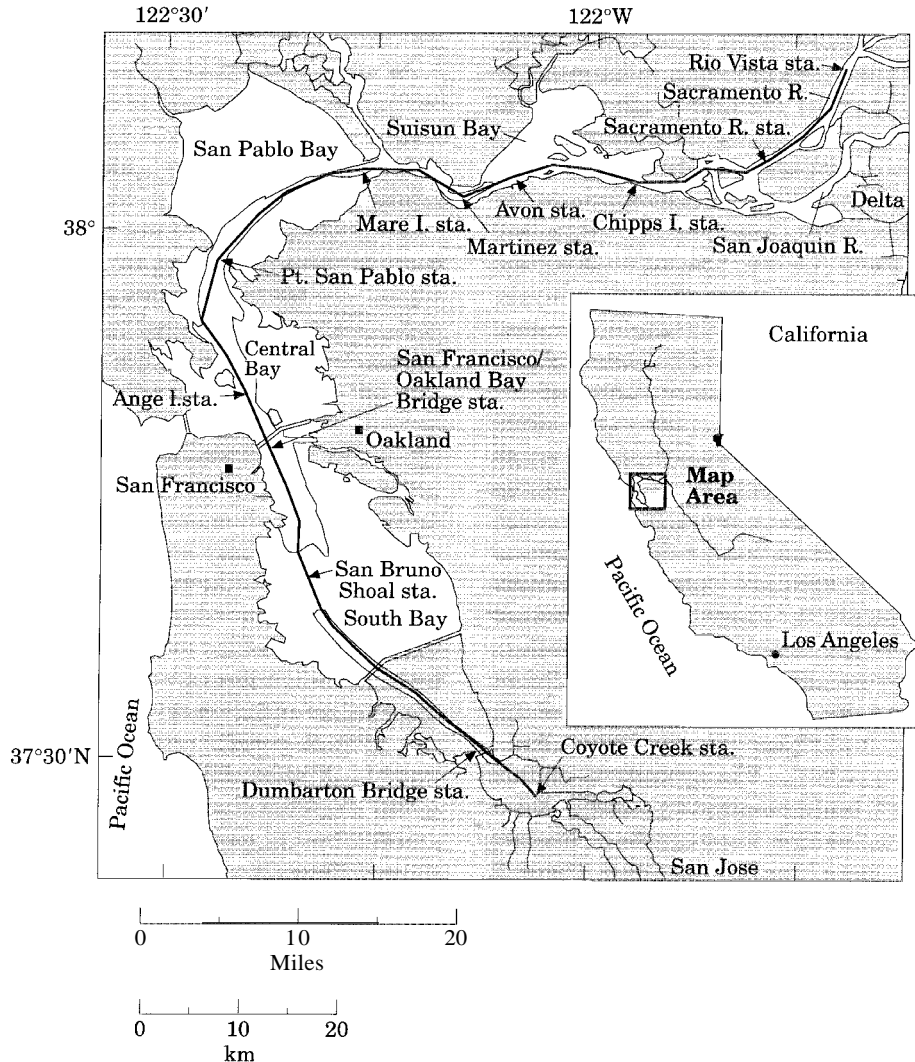
FIGURE 1. San Francisco Bay. The MIDAS cruise track is shown as a solid line along the axis of the estuary.

water column proper. The specific goal is to choose a minimal number of sampling locations along the transect from which one can estimate a scalar index of conditions (in this case, the mean) with sufficient precision (i.e. with sufficiently low variance) that useful comparisons can be made among different time periods.

Deciding on a station array requires consideration of three linked issues, addressed here in sequence:

(1) What kind of sampling design should be adopted (e.g. random, systematic or stratified)?

(2) How can the precision (variance) of the transect mean be estimated?

(3 For a prescribed level of precision, how many samples (stations along a transect) are required?

Answers to these questions require knowledge about the underlying distribution of the parent popu-

lation of all possible samples. The authors' approach is empirical and is not driven by theoretical assumptions about the underlying distribution. Although it is impossible to sample the entire population of water quality measurements in an estuary, a surrogate parent population can be acquired by collecting a large number (thousands) of samples as closely-spaced sensor measurements made while a ship profiles an axial transect. A modified version of the integrated software-instrument package MIDAS (Multiple Interface Data Acquisition System; Walser et al., 1992; see also Madden & Day, 1992) was used to collect and store measurements from flow-through water quality sensors and a Global Positioning System (GPS) navigation system. Subsampling from the high-resolution MIDAS transect data was then used to address the issues listed above.

We begin with a consideration of the spatial sampling design (Issue #1). In simple random sampling, each station is selected randomly and independently in space along the transect line. Although simple in concept and obviously unbiased, random sampling has an important drawback in situations where the spatial correlation is high; if two stations are randomly chosen too close together, then they will have similar values and one of them is, to a certain extent, wasted.

Systematic sampling, i.e. equidistant spacing of stations along the transect line, avoids this problem and therefore yields a more precise estimate of the spatial mean in many situations (Murthy & Rao, 1988). A more precise estimate of the spatial mean implies, in turn, that temporal trends of a given size can be detected in fewer years or, alternately, that smaller temporal trends can be detected in any given time interval. Systematic spatial designs are also more convenient to implement. Systematic samples suffer, however, from a serious drawback in that unbiased estimates of precision are unavailable, and approximations based on assumptions about the nature of the underlying population must be utilized (Bellhouse, 1988). The precision cannot be estimated based on the sampling design alone because a systematic sample is essentially a random sample of size one; once the first station is selected, the locations of the others are completely specified as well. Furthermore, systematic sampling does not always yield the most precise estimates; the relative performance of different designs depends on the structure of the underlying population (Cochran, 1977).

Stratified sampling refers to, in this case, dividing a relatively heterogeneous estuary into more homogeneous subdomains and then carrying out either a random or systematic programme of sampling independently within each subdomain (stratum). Insofar as the within-subdomain variability is reduced relative to the between-subdomain variability, stratification can lead to a more precise estimate of the mean than either simple random or systematic sampling (Cochran, 1977). The strong spatial correlation characteristic of estuaries (Powell et al., 1986) suggests that stratification of sampling into spatially contiguous subregions might be appropriate. In order to choose the strata in a consistent way, a novel method is employed here; the machinery of tree-based modelling (Clark & Pregibon, 1992).

The MIDAS transect data enable one to evaluate the relative performance of these different sampling designs. In particular, simple random sampling, systematic sampling and their stratified counterparts, stratified random and stratified systematic sampling, are compared.

As suggested above, if systematic sampling turns out to give the most precise estimate of the underlying mean, one must decide how the variance of the estimate can best be calculated from the low-resolution station arrays commonly encountered in practice (Issue #2). Many different estimators have been proposed, most of which are based on an assumed model of population behaviour and so are appropriate only when the model truly represents the population. Whether or not a single tractable model can be applied to estuarine data in general is not known. At different times and locations, transects appear to be dominated by noise, linear or higher-order trends, persistence (spatial autocorrelation) or, most often, a complex combination of these basic patterns. When only low-resolution samples are available, there is little hope of identifying a suitable model. The authors' intention is, therefore, to assess the robustness of the different estimators for use when high-resolution data are not accessible, given that the appropriate model may be temporally sensitive. Subsets of these methods have been compared for demographic (Wolter, 1984) and stereological (Mattfeldt, 1989) data but their relative performance cannot be extrapolated to natural ecosystems, which can exhibit quite different population structures. Again, the MIDAS data enable a direct assessment of the different variance estimators by providing high-resolution descriptions of the underlying populations in an estuary.

Given a sampling design and a way to estimate the resulting precision, how does one choose an appropriate number of stations (Issue #3)? The use of some criterion of performance or objective function is an essential step in completing this phase of design, but the criterion depends on the overall monitoring objective (e.g. to describe ambient conditions, assess compliance with standards, detect trends or determine causal mechanisms) and the costs and uncertainties associated with different designs. Rather than linking this analysis to a specific objective function, it was asked how the ability to reproduce the underlying data, as measured by the variance of the estimated mean, depends on the number of stations. This relation is simple enough to calculate with a high-resolution data set in hand, but of specific interest is what can be said when the only data available are from sparser transects (i.e. the usual kind of data collected in monitoring programmes). Therefore, one should look for generalities in the relation that may be characteristic of the underlying spatial structure in an estuary, and can be used to guide sample size when only low-resolution data are available.
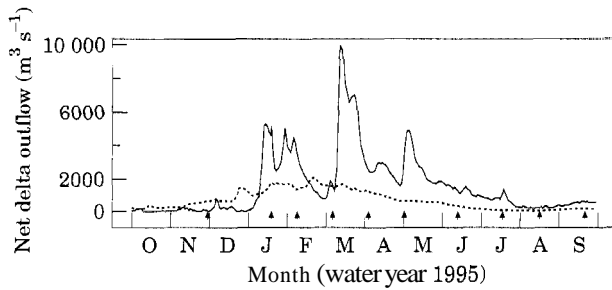
FIGURE 2. Freshwater outflow from the Sacramento-San Joaquin Delta into San Francisco Bay. ——, water year 1995 (1 October 1994–30 September 1995); ···, average for the water years 1956–95; ↑, cruise dates.

Methods

Data collection

Ten cruises were conducted from November 1994 through September 1995 at approximately neap tide, collecting data along a 150 km transect in the main channel of San Francisco Bay from the landward end of South Bay through Central Bay to Rio Vista on the Sacramento River (Figure 1). The timing of the cruises captured the broad range of freshwater flow conditions experienced over the year (Figure 2), as well as major events such as the spring bloom in South Bay and summer estuarine turbidity maximum in the northern Bay. Peak flows during the 1995 water year were among the highest of the last 40 years; the range of flow conditions encountered during these transects is therefore unusually large compared to an ' average water year ' (Figure 2).

Fourteen hydrographic, meteorological and navigational variables were measured using the Multiple Interface Data Acquisition System (MIDAS) on board the RV *Polaris.* The ship's location was determined with a Trimble NavTrac XL Global Positioning System. Water for the hydrographic parameters was pumped from a through-hull fitting located at the bow of the ship at a depth of approximately 2 m. The pumped water sample was directed through an array of sensors for continuous analysis. The hydrographic variables were measures of salinity, temperature, chlorophyll fluorescence and turbidity. Salinity was derived from measures of conductivity and temperature made using a Sea-Bird Electronics SBE-21 thermosalinograph. Temperature was measured with a Sea-Bird Electronics SBE-3 temperature probe located at the bow of the ship in the pumped sample stream, chlorophyll fluorescence with a Turner Designs Model 10 flow-through fluorometer, and turbidity with a Turner Designs Model 10 flow-through nephelometer.

Discrete water samples for chlorophyll *a* and SPM were collected at 10–18 selected stations during the ongoing recording of fluorescence and nephelometry data signals (Edmunds *et al.,* 1995). Samples for chlorophyll were filtered onto a Gelman A/E glass fibre filter and immediately frozen. The air-dried filter was ground in 90% acetone within 1 week of collection. After extraction for 18–24 h at − 10 °C, absorbances of the extracts were determined on a Hewlett-Packard 8452A diode array spectrophotometer. Chlorophyll *a* values were calculated using Lorenzen's (1967) equations. Samples for SPM were filtered onto preweighed, 0·4-μm pore size, polycarbonate membrane filters and then air dried. The filters were reweighed and the concentration of SPM calculated after a correction was made for salt on the filters. The fluorescence and nephelometry signals were calibrated separately for each cruise using the discrete values collected during that cruise. For the first three cruises and the July cruise, the authors were unable to obtain significant regressions of the MIDAS fluorescence signal on discrete chlorophyll values. Fluorescence data were not used in the analyses.

The MIDAS data acquisition system records the data at a sampling interval of approximately 6 s. The ship's speed over ground varied with the tides but was generally about 5 m s^{-1} (10 knots), which resulted in a spatial sampling interval of approximately 30 m. During each transect from South Bay to Rio Vista, approximately 5000 measures of each parameter were collected. The average distance between successive data points in the raw database was approximately 30 m, but the actual distances were variable because of changing ship speed. The implementation of tree-based regression used here is sensitive to the data density (data points per km of transect), and changes in this density over the course of the transect can bias the analysis. In order to equalize the data density over the transect, a subset of the data was formed by marking the transect at 100-m intervals measured along the transect from the starting point, and selecting the single data record closest to each marker.

Horizontal stratification of the estuary

Stratification calculations. To compare the different sampling designs, the estuary must first be stratified. Tree-based modelling or regression operates by successively splitting a dataset (transect) into increasingly homogeneous subsets or strata until some stopping rule comes into effect. In this case, each split is chosen to maximize the difference between the variance of the ' parent ' stratum and the sum of the variances of the

two 'children' strata. As different transects and variables result in different splits, a further step is to extract from the combined collection of splits those regions where they tend to cluster and collectively support the placement of a boundary. Tree-based modelling therefore serves more as a guide to the location of strata boundaries rather than an exact specification of these boundaries.

Trees were 'grown' using the algorithms of S-PLUS (Clark & Pregibon, 1992; Statistical Sciences, 1994). Each transect was successively split along the transect path in order to maximize the quantity AD:

$$\Delta D = \sum_i (y_i - \mu)^2 - \left(\sum_{i \in L} (y_i - \mu_L)^2 + \sum_{i \in R} (y_i - \mu_R)^2 \right) \quad (1)$$

where y_i is the ith observation in the parent stratum; μ is the mean value of the parent stratum; L and R are the sets of indices defining the left-hand and right-hand children strata, respectively; and μ_L and μ_R are the mean values in the two respective children strata. The splitting process continued until none of the resulting strata could account for more than 10% of the original variance. Strata were then compared among variables for the same transect and among transects for the same variable. The value of 10% was chosen because smaller values resulted in strata that were probably dependent on tidal stage. For example, the splits computed for two successive transects on 18 and 19 January 1995 (beginning at the same time of day but at opposite ends of the Bay) essentially coincided if 10% was used as a cutoff; on the other hand, splits that resulted in strata accounting for less than 10% of the original variance did not coincide.

Sample allocation among strata. In order to test the efficacy of a stratification scheme, one must also decide how to allocate samples among the various strata. This can be done in several different ways. The simplest method is proportional allocation, in which the number of samples in any stratum is directly proportional to the stratum size. Stratum size in the case of a MIDAS transect is simply the length along the transect between stratum boundaries.

In contrast to proportional allocation, the most efficient or *optimal* allocation of stations among strata takes into account stratum variability and sampling costs, in addition to stratum size. For any stratum i of a transect, the number of stations that minimizes the variance of the estimated mean for a given total cost is:

$$n_i \propto \frac{W_i S_i}{\sqrt{c_i}} \quad (2)$$

where W_i is the stratum size, S_i is the stratum standard deviation and c_i is the cost per sample in that stratum (Cochran, 1977). If the cost of sampling a station is constant throughout the estuary, then Equation 2 implies that the density of stations within a stratum is simply proportional to the standard deviation of the transect variable.

Due to the potential discrepancy in optimal allocations for different variables (due to different values of S_i), the efficacy of a *compromise* allocation among strata was also examined. The compromise was effected by minimizing the average over all three variables of the proportional increase in variance over optimal allocation (Chatterjee, 1967). It can be shown that the resulting sample sizes are:

$$n_i = n \sqrt{\sum_{j=1}^{3} n_{ij}^2} \Big/ \sum_{i=1}^{h} \sqrt{\sum_{j=1}^{3} n_{ij}^2} \quad (3)$$

where n_{ij} is the optimum sample size in stratum i for variable j.

Sampling design

The variance of the mean was calculated for several different practical sampling strategies, and compared to simple random sampling. The authors' approach was to regard the MIDAS transect data as the underlying population. The variance of the simple random sampling estimate is then given by:

$$\text{var}(\bar{y}_{ran}) = (1 - f) \frac{S^2}{n} \quad (4)$$

where S is the population standard deviation, n is the sample size, and $f = n/N$ is the sampling fraction with N the population size (Cochran, 1977). In practice, n will be much less than 75 and N is usually around 1500, so the true sampling fraction is much less than 5%. As a rule of thumb, the finite population correction $(1-f)$ can be ignored when $f < 0.05$ (Barnett, 1991), and it will be ignored in what follows.

For stratified random sampling, the variance of the estimated mean depends on the sample allocation strategy. In the case of proportional allocation, the variance is:

$$\text{var}(\bar{y}_{st}) = \frac{1}{n} \sum_{i=1}^{h} W_i S_i^2 \quad (5)$$

where W_i is the relative size and S_i is the standard deviation of the ith stratum, and there are h strata. For optimal allocation, the corresponding calculation is:

$$\text{var}(\bar{y}_{st}) = \frac{1}{n} \sum_{i=1}^{h} W_i^2 S_i^2 \qquad (6)$$

The variance for a compromise allocation can be determined from the general result for stratified samples, which can be expressed in the form:

$$\text{var}(\bar{y}_{st}) = \frac{1}{n} \sum_{i=1}^{h} \frac{W_i^2 S_i^2}{\alpha_i} \qquad (7)$$

where $n_i = na_i$.

The traditional station configuration in San Francisco Bay has been an approximately systematic one, i.e. with equal distances between adjacent stations. Assuming N is a multiple of n, there are $m = N/n$ possible systematic samples. When N was not a multiple, the method of circular systematic sampling due to Lahiri (Bellhouse, 1988) was used. The variance of the systematic estimate is then simply:

$$\text{var}(\bar{y}_{sy}) = \frac{1}{m} \sum_{k=1}^{m} (Y_k - \bar{Y})^2 \qquad (8)$$

where $\bar{Y}_k$ is the mean of the kth potential systematic sample and $\bar{Y}$ is the transect mean.

Finally, the performance of stratified sampling with proportional allocation was investigated again, but with systematic rather than simple random sampling within strata. Variances within strata are then given by Equation 8, but the variance of the estimated overall transect mean is:

$$\text{var}(\bar{y}_{stsy}) = \sum_{i=1}^{h} W_i^2 \text{var}(\bar{y}_i)^2 \qquad (9)$$

where $\text{var}(\bar{y}_i)$ is the variance for the estimated mean of stratum i.

Each sampling strategy was compared to simple random sampling by calculating the percent decrease in variance $100(1 - V/V_{ran})$, where V_{ran} is the variance given by Equation 4 and V is the variance due to one of the other strategies. For stratified random sampling, n drops out and the comparisons are independent of sample size. For systematic and stratified systematic, however, the ratio depends on n and so the results for three sample sizes (10, 20 and 40), typical of the range for transects in estuarine research and monitoring programmes and specifically covering the range used in San Francisco Bay, were examined.

Variance estimators

A variety of estimators that have been proposed for systematic sampling and that are simple to compute were examined (Table 1). The first estimator SRS is simply the variance of the simple random sampling estimate (Equation 4). Estimator *MURT1* considers the systematic sample as a stratified random sample with two samples from each of $n/2$ strata; *MURT2* is similar but based on successive differences (Murthy & Rao, 1988). The next three estimators are based on higher-order differences; *WOLT1* and *WOLT2* attempt to account for trends and *WOLT3* for autocorrelation (Wolter, 1984). The estimator KOOP consists of a pseudo-replication in which the sample is split into two systematic subsamples (Koop, 1971). The next three estimators also attempt to take into account the spatial correlation structure of the population. Estimator COCH is an asymptotic result due to Cochran (1946) and assumes an autoregressive process of order one; estimators CHEV (Yates, 1960; Chevrou, 1976) and *GUND* (Gunderson & Jensen, 1987) are based on regionalized variable theory (Mattfeldt, 1989). CHEV was developed specifically for error estimation in linear systematic arrays, while *GUND* is based on a quadratic approximation to the variogram. The final estimator *MAHA* is based on two independent systematic samples of size $n/2$, a technique known as the method of interpenetrating subsamples (Mahalanobis, 1946).

The estimators for sample sizes of 10, 20 and 40 stations were compared. For a given sample size, hydrographic variable and transect, each estimator was applied to all possible systematic samples, and the performance of an estimator was summarized with the mean square error MSE:

$$MSE(v) = E[(v - \text{var}(\bar{y}))^2] \qquad (10)$$

Results for all transects, for a given sample size and variable, were then averaged and ranked.

Sample size

The relation between variance and sample size was demonstrated empirically by computing $\text{var}(\bar{y})$ (Equation 8) from all possible systematic subsamples for a range of sample sizes, specifically 5 through 50. This range encompasses the number of fixed stations likely to be encountered in practice. The calculation was repeated for each transect and variable. The relation was assumed to be of the form:

$$\text{var}(\bar{y}) \propto (1/n)^\alpha \qquad (11)$$

and estimates of a were extracted using the Golub-Pereyra algorithm for partially linear models (Bates & Chambers, 1992).

TABLE 1. Estimators of variance for systematic sampling

Estimator	Description
1 *SRS*	s^2/n, where $s^2 = \Sigma_{j=1}^{n}(y_j - \bar{y})^2/(n-1)$
2 *MURT1*	$(1/n)\Sigma_{j=1}^{n/2}a_{2j}^2/n$, where $a_j = y_i - y_{j-1}$
3 *MURT2*	$(1/n)\Sigma_{j=2}^{n}a_j^2/2(n-1)$, where $a_j = y_j - y_{j-1}$
4 *WOLT1*	$(1/n)\Sigma_{j=3}^{n}b_j^2/6(n-2)$, where $b_j = y_j - 2y_{j-1} + y_{j-2}$
5 *WOLT2*	$(1/n)\Sigma_{j=5}^{n}c_j^2/3\cdot5(n-4)$, where $c_j = y_j/2 - y_{j-1} + y_{j-2} - y_{j-3} + y_{j-4}/2$
6 *WOLT3*	$(1/n)\Sigma_{j=9}^{n}d_j^2/7\cdot5(n-8)$, where $d_j = y_j/2 - y_{j-1} + \ldots - y_{j-7} + y_{j-8}/2$
7 *KOOP*	$(1/4)\{(2/n)\Sigma_{j\text{ even}}y_i - (2/n)\Sigma_{j\text{ odd}}y_j\}^2$
8 *COCH*	$(s^2/n)\{1 + (2/\ln \rho) + 2/(\rho^{-1} - 1)\}$, where $\rho = \Sigma_{i=2}^{n}(y - \bar{y})(y_{i-1} - \bar{y})/\{s^2(n-1)\}$
9 *CHEV*	$(1/n)\Sigma_{j=2}^{n}a_j^2/10n$, where $a_j = y_j - y_{j-1}$
10 GUND	$(1/n)\{3\Sigma_{j=1}^{n}y_j^2 + \Sigma_{j=3}^{n}y_j y_{j-2} - 4\Sigma_{j=2}^{m}y_j y_{j-1}\}/12$
11 *MAHA*	$(1/2)\Sigma_{i=1}^{2}(\bar{y}_i - \bar{y})^2$, where $\bar{y}_i$ are means of independent samples of size $n/2$

The sample is denoted by $y_,$, $j = 1, 2, \ldots, n$.

The relation between variance and sample size is also a relation between variance and interstation distance. In order to portray how variance changes with spatial scale, as opposed to sample size, $\text{var}(\bar{y})$ (Equation 8) was computed from all possible systematic subsamples for station separations of 1–64 km.

Results

Horizontal stratification of the estuary

The transect data are essentially one-dimensional and so are best portrayed as a function of distance measured along the transect. In order to simplify the presentation, the main points regarding stratification are illustrated with results from two cruises (Figures 3 and 4). The first feature to note is that the strata chosen can differ among variables for the same cruise. For example, chlorophyll *a* exhibits near-homogeneity between the Bay Bridge and Sacramento River on 4 April 1995, while four strata have been attributed to SPM in this same region (Figure *3*). Secondly, and in a similar vein, the strata can differ among cruises for the same variable. For example, five strata are required to describe chlorophyll *a* on 4 April 1995 with eight are required on 21 September 1995
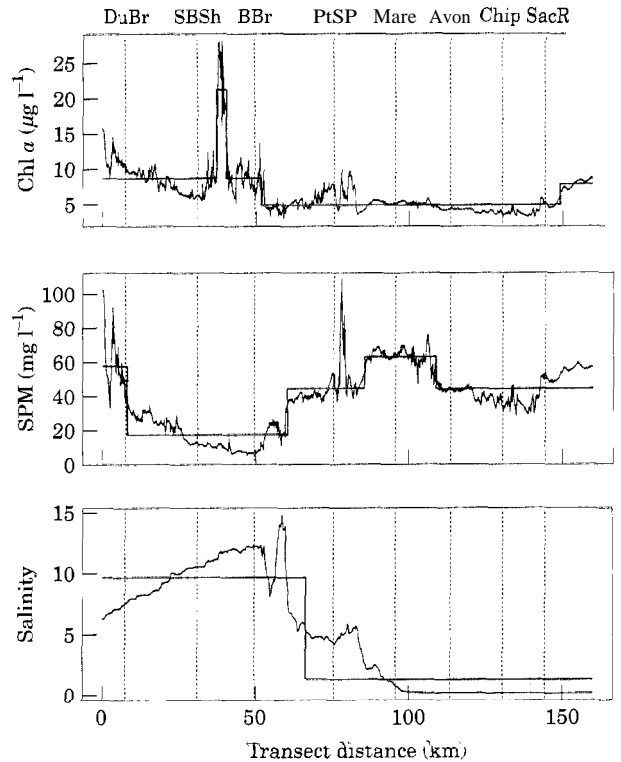

FIGURE 3. MIDAS transects with tree-based model superimposed for 4 April 1995. DuBr, Dumbarton Bridge; SBSh, San Bruno Shoal; BBr, Bay Bridge; PtSP, Pt. San Pablo; Mare, Mare I.; Avon, Avon Pier; Chip, Chipps I.; SacR, first Sacramento R. Station; SPM, suspended particulate matter.
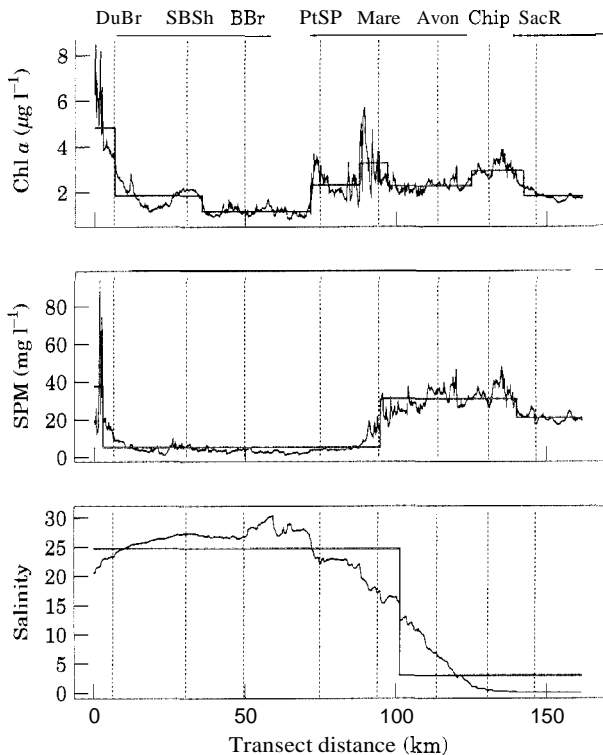
FIGURE 4. MIDAS transects with tree-based model super-
imposed for 21 September 1995. DuBr, Dumbarton Bridge;
SBSh, San Bruno Shoal; BBr, Bay Bridge; PtSP, Pt. San
Pablo; Mare, Mare I.; Avon, Avon Pier; Chip, Chipps I.;
SacR, first Sacramento R. Station; SPM, suspended particu-
late matter.

(Figure 4). Finally, strata boundaries do not necess-
arily demarcate homogeneous regions, but may occur
in the middle of a strong spatial gradient, such as the
one for salinity on 21 September 1995. These
three features illustrate that an estuarine model of
variable-independent, temporally stable and homo-
geneous subdomains can be a poor approximation.
The subdomains defined by tree-based regression,
which turn out to be the appropriate ones for stratified
sampling (see below), change among variables and
seasons. Furthermore, the frequent presence of large
gradients over much of the estuary contradicts the
very notion of homogeneous subdomains.

Although the model of homogeneous subdomains
may be in some sense a poor one for this estuary, the
model need not fit perfectly in order for improvements
in estimating the overall mean. The only require-
ment is that the typical within-stratum variance
is sufficiently small compared to the variance of the
within-stratum means (Barnett, 1991). Further-
more, indicating the positions of splits but not their
importance does not fully characterize the results
and may bias one's view of the efficacy of strati-

fication. In order to test its efficacy more object-
ively, one needs to decide on a compromise stratifi-
cation that summarizes the commonalities among
the tree-based regressions for individual cruises and
variables.

For each variable, the positions of the splits for all
cruises were superimposed on the cruise track (Figure
5). Each split of a stratum is represented by a square,
the area of which is proportional to the total variance
represented by that stratum. Major splits for chloro-
phyll a are situated near the Dumbarton Bridge, the
San Bruno Shoal and Angel Island. For SPM, the
major boundaries are in the vicinity of the Dumbarton
Bridge, Angel Island and in northern San Pablo Bay.
Salinity is stratified most strongly near Angel Island,
northern San Pablo Bay and Martinez. Most of these
locations coincide with important physiographic and
hydrological features (Figure 1). The Dumbarton
Bridge marks a significant constriction in southern
South Bay; the San Bruno Shoal is a large shallow
expanse that is also a hydrodynamic and biological
boundary (Powell et al., 1986); Angel Island marks
the southern boundary of the river-dominated portion
of the estuary, where the flow from the Sacramento
and San Joaquin Rivers turns westward and exits
through the Golden Gate; and Martinez marks the
upstream boundary of the Carquinez Strait, a narrow
constriction in the northern Bay. A number of major
boundaries cluster toward the northern end of San
Pablo Bay, but do not clearly demarcate any single
position. The 'lability' in this region is due to the
strong gradients often present in San Pablo Bay.
Rather than situating a boundary at some location in
the centre of these splits that has no physiographic or
hydrodynamic significance, the authors chose to
locate a boundary at Mare Island, which marks the
northern end of these splits and the seaward boundary
of the Carquinez Strait.

In this way, all the boundaries have a physiographic
or hydrodynamic significance. The exact boundaries
of the six strata are defined in Table 2. Note that
stratum size (length) changes slightly among transects
because of small differences in the actual course taken
by the ship. The means and standard deviations of
water quality variables for each stratum and transect
are summarized in Table 3.

Proportional allocation of stations is the same for all
variables, and simply mirrors stratum size (Table 4).
Optimal allocation, on the other hand, differs greatly
among variables for most strata, indicating that there
is no general optimal allocation. The compromise
allocation resembles the proportional allocation
although, based on their covariance among strata, it is
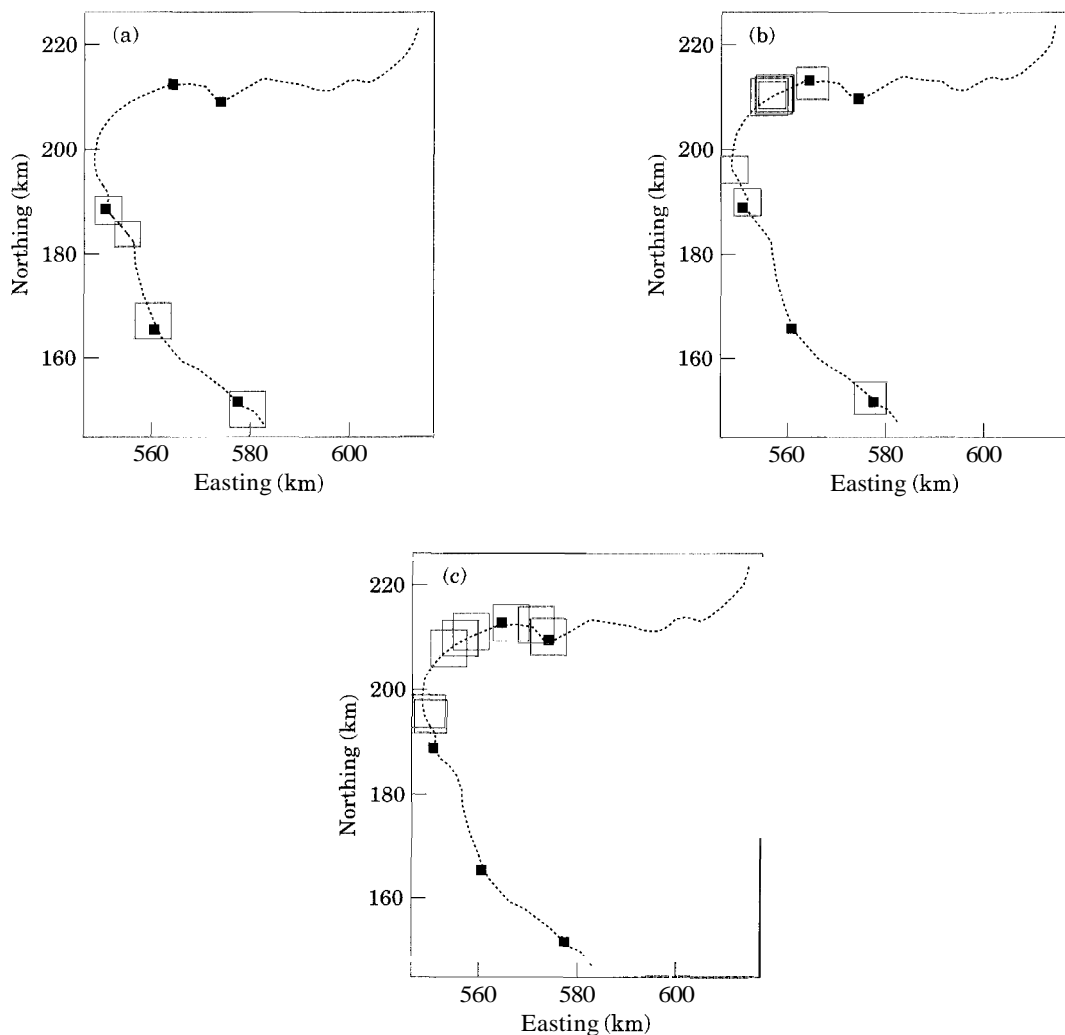most similar to the optimal allocation for chlorophyll.

FIGURE 5. Summary plot of the split locations for all transects. The area of the square representing each split is proportional to the decrease in deviance due to the split. (a) Chlorophyll *a;* (b) suspended particulate matter; (c) salinity.

Sampling design

In the case of stratified random sampling, proportional allocation showed large increases in precision compared to simple random sampling, largest in the case of salinity but substantial for all variables (Table 5). Optimal allocation exhibited further increases in precision in all cases, eliminating 23–45% of the remaining variance. The compromise allocation,

TABLE 2. Definition of a stratification scheme for the MIDAS transects in San Francisco Bay

Stratum no.	Description	Size ± SD (km)	Northing (km)	Easting (km)
1	South of Dumbarton Br.	6·9 ± 0·3	<151·4	
2	Dumbarton Br. to San Bruno Shoal	23.3 ± 0.6	151·4–165.3	
3	San Bruno Shoal to Angel I.	28.7 ● 0.7	165.3–188.8	
4	Angel I. to Mare I.	37.3 ± 2.2	≥188·8	<564·6
5	Mare I. to Martinez	13.1 ± 1.1	≥188.8	564·6–574·5
6	East of Martinez	51.8 ± 1.7	≥188.8	≥574.5

Locations are specified in terms of UTM coordinates.

TABLE. *3.* Mean ± standard deviation of water quality variables within each stratum of Table 2

	Stratum number					
	1	2	3	4	5	6
29 November 1994						
Salinity	28·5 ± 0·5	30·4 ± 0·4	30·3 ± 0·1	27·4 ± 2·4	19·9 ± 2·2	6·31 ± 5·26
SPM	62·8 ± 10·6	24·9 ± 10·8	6·75 ± 1·57	20·3 ± 13·3	55·7 ± 7·0	37·5 ± 14·3
Chl *a*	1·82 ± 0·08	2·08 ± 0·07	2·02 ± 0·08	1·70 ± 0·10	1·46 ± 0·04	1·53 ± 0·05
18 January 1995						
Salinity	19·7 ± 1·3	19·6 ± 1·7	14·8 ± 1·8	6·01 ± 4·76	0·252 ± 0·141	0·0775 ± 0·0179
SPM	27·2 ± 14·1	1·19 ± 2·78	1·53 ± 2·26	55·7 ± 52·5	157 ± 4	152 ± 18
Chl *a*	1·46 ± 0·04	1·53 ± 0·01	1·54 ± 0·01	1·42 ± 0·13	1·16 ± 0·02	1·19 ± 0·06
7 February 1995						
Salinity	15·5 ± 0·2	15·4 ± 0·6	13·4 ± 1·3	5·8 ± 2·2	1·18 ± 0·96	0·0795 ± 0·0054
SPM	4·57 ± 0·96	5·34 ± 0·68	4·91 ± 2·43	36·5 ± 19·0	97·9 ± 13·8	96·3 ± 11·8
Chl *a*	1·44 ± 0·01	1·45 ± 0·02	1·51 ± 0·02	1·41 ± 0·08	1·14 ± 0·06	1·15 ± 0·04
7 March 1995						
Salinity	15·1 ± 0·3	17·5 ± 1·0	20·1 ± 1·1	9·54 ± 3·37	1·54 ± 1·15	0·101 ± 0·021
SPM	47·6 ± 3·8	28·5 ± 9·9	9·72 ± 2·96	20·7 ± 10·1	40·6 ± 3·7	25·9 ± 7·8
Chl *a*	38·1 ± 2·0	28·1 ± 6·5	10·5 ± 5·6	5·09 ± 1·46	4·87 ± 0·28	3·61 ± 0·60
4 April 1995						
Salinity	6·82 ± 0·34	9·03 ± 1·00	11·5 ± 1·2	4·3 ± 2·1	0·211 ± 0·162	0·108 ± 0·011
SPM	59 ± 16	23 ± 7	10·8 ± 5·4	48·7 ± 12·5	62·5 ± 5·6	43·5 ± 8·3
Chl *a*	11·5 ± 1·5	8·12 ± 1·42	8·74 ± 5·16	5·56 ± 1·23	5·12 ± 0·28	4·861·56
2 May 1995						
Salinity	14 ± 0	16·2 ± 1·4	18·9 ± 1·7	6·4 ± 3·6	0·121 ± 0·032	0·0785 ± 0·0137
SPM	95·3 ± 37·2	34·4 ± 13·4	18·5 ± 10·2	70·9 ± 21·5	53·4 ± 14·4	22·1 ± 4·7
Chl **a**	13·6 ± 6·6	3·41 ± 2·05	5·28 ± 4·03	15·2 ± 4·2	7·05 ± 2·24	1·57 ± 0·80
13 June 1995						
Salinity	13·6 ± 1·3	17·5 ± 0·9	21·6 ± 1·5	15 ± 6	2·58 ± 1·35	0·146 ± 0·171
SPM	634 ± 145	85·7 ± 101	14·9 ± 28·9	29·1 ± 25·7	167 ± 73	79 ± 86
Chl *a*	6·97 ± 1·92	2·63 ± 0·529	2·18 ± 0·12	2·27 ± 0·11	2·72 ± 0·30	2·51 ± 0·39
18 July 1995						
Salinity	20·4 ± 0·4	22·2 ± 0·7	24·8 ± 0·8	13·7 ± 5·0	2·42 ± 1·57	0·107 ± 0·113
SPM	19·1 ± 5·2	5·38 ± 4·07	2·54 ± 1·01	19·4 ± 14·6	76 ± 10	30·3 ± 18·7
Chl *a*	2·46 ± 0·09	2·27 ± 0·13	2·22 ± 0·16	2·34 ± 0·24	3·4 ± 0·3	2·67 ± 0·43
16 August 1995						
Salinity	22·5 ± 0·3	24·1 ± 0·7	27·1 ± 1·1	20·9 ± 4·0	10·1 ± 2·2	1·54 ± 2·03
SPM	32·7 ± 5·5	10·5 ± 6·3	3·67 ± 1·08	20·7 ± 7·1	25·1 ± 3·8	32·6 ± 11·9
Chl *a*	6·31 ± 1·2	1·23 ± 1·10	0·9 ± 1·6	8·65 ± 2·29	5·4 ± 1·2	5·62 ± 2·41
21 September 1995						
Salinity	22·6 ± 0·9	26·2 ± 1·0	27·8 ± 1·0	24·2 ± 3·6	14·7 ± 2·1	2·14 ± 3·17
SPM	26·3 ± 19·4	5·78 ± 2·21	4·48 ± 1·05	5·76 ± 3·91	25·7 ± 4·8	29 ± 7
Chl *a*	4·94 ± 1·35	1·92 ± 0·52	1·44 ± 0·36	2·18 ± 1·05	2·38 ± 0·30	2·4 ± 30·53

SPM, suspended particulate matter; Chl *a,* chlorophyll *a.*

however, showed only a modest improvement over proportional allocation for salinity, and was slightly worse in the case of chlorophyll *a* and SPM.

Systematic sampling performed better than stratified random sampling with as few as 10 stations, although the large standard deviations imply that results were highly transect dependent ('Table 5). As the sample size increased, the precision of systematic sampling, increased and exceeded even stratified random sampling with optimal allocation using only 20 stations. Stratified systematic sampling was slightly better than simple systematic in the case of salinity but worse in the case of SPM, regardless of sample size. For chlorophyll, stratified systematic was sometimes better, sometimes worse.

Variance estimators

The top three estimators, COCH, *CHEV* and GUND, were among those which attempted to account for spatial autocorrelation (Table 6). Estimator *CHEV* had an average ranking of 1·6, COCH 2.6 and GUND

TABLE 4. Sample sizes within strata expressed as a percentage of the total number of samples

Stratum no.	Proportional allocation ± SD	Optimal Allocation ± SD			Compromise allocation ± SD
		Chlorophyll a	SPM	Salinity	
1	4·3 ± 0·2	8·6 ± 8·0	6·0 ± 4·9	1·4 ± 0·9	6·5 ± 4·2
2	14·5 ± 0·5	17·0 ± 11·5	11·2 ± 7·4	9·2 ± 5·0	13·1 ± 5·1
3	17·8 ± 0·6	23·5 ± 15·9	5·8 ± 4·4	14·7 ± 8·0	15·6 ± 8·1
4	23·1 ± 0·9	22·3 ± 12·7	32·1 ± 14·6	54·1 ± 15·9	34·1 ± 5·0
5	8·1 ± 0·5	3·9 ± 2·5	6·5 ± 2·9	5·6 ± 3·5	5·1 ± 1·4
6	32·2 ± 1·0	24·7 ± 13·9	38·4 ± 12·8	15·0 ± 23·7	25·6 ± 13·4

SPM, suspended particulate matter.
The standard deviations represent variation among transects.

4.0. The next best estimators were KOOP and *MAHA*, which base their estimates on two subsamples of equal size. The estimator SRS, which treats the sample as if it were a simple random sample, is highly inefficient; it came in last in every instance.

Sample size

Fits of the inverse power relationship separately to each variable and transect (Equation 11) resulted in a narrow range of a averaging 1.9 ± 0.1 (SE) (Figure 6). No significant effects of either variables or transects were found. For theoretical reasons discussed below, the ability of an inverse square curve to fit all data simultaneously was examined. The resulting fits each appeared to be a satisfactory description of the data (Figure 7).

The relative standard error of the median transect is almost always less than 10% when stations are spaced up to 8 km apart (Figure 8).

Discussion

Horizontal stratification *of* the estuary

Horizontal stratification of an estuary, i.e. division of the estuary into subdomains, can be motivated by many different goals:

(1) The need for precise estimates of estuary-wide statistics such as the overall mean was the authors' primary motivation. As discussed above, if the estuary can be divided into subdomains that are relatively homogeneous compared to the between-subdomain variability, then estimates of the overall mean will be more precise than for a simple random sample. These

TABLE 5. Percent decrease in var($\bar{y}$) compared to simple random sampling for different types of sampling strategies

Sampling type	Chlorophyll a ± SD	SPM ± SD	Salinity ± SD
Stratified random			
Proportional	65·4 ± 19·9	73·1 ± 10·1	93·4 ± 2·0
Optimal	75·1 ± 15·6	79·3 ± 9·5	96·4 ± 1·1
Compromise	63·5 ± 23·8	72·4 ± 12·0	94·6 ± 2·2
n = 10			
Systematic	70·8 ± 24·0	86·6 ± 15·3	95·7 ± 1·1
Stratified systematic	66·6 ± 32·7	82·8 ± 9·0	96·9 ± 1·2
n = 20			
Systematic	78·3 ± 11·2	93·1 ± 6·1	97·4 ± 1·3
Stratified systematic	84·1 ± 9·2	91·0 ± 4·6	98·4 ± 0·6
n = 40			
Systematic	86·7 ± 10·3	95·2 ± 5·6	99·2 ± 0·4
Stratified systematic	91·8 ± 4·2	94·4 ± 3·0	99·2 ± 0·3

SPM, suspended particulate matter.
The standard deviations represent variation among transects.

TABLE 6. Ranking of estimators in Table 1 based on their *MSE* (Equation 10) for MIDAS transect data

Estimator	Chl *a*			SPM			Salinity		
	$n=10$	$n=20$	$n=40$	$n=10$	$n=20$	$n=40$	$n=10$	$n=20$	$n=40$
SRS	11	11	11	11	11	11	11	11	11
MURT1	9	9	8	9	9	9	9	9	9
MURT2	10	10	10	10	10	10	10	10	10
WOLT1	7	6	6	8	8	7	5	4	5
WOLT2	6	7	7	7	6	6	8	3	6
WOLT3	8	5	9	6	7	8	7	5	7
KOOP	1	4	3	2	4	4	4	8	8
COCH	4	3	2	4	2	3	1	1	3
CHEV	3	1	1	1	1	1	2	2	2
GUND	2	2	4	5	5	5	3	6	4
MAHA	5	8	5	3	3	2	6	7	1

n, size of systematic sample.
SPM, suspended particulate matter; Chl a, chlorophyll a.

results show that stratification is very effective in improving precision over simple random sampling. Stratified random sampling, however, is inferior to simple systematic sampling with as few as 10 samples. Moreover, stratified systematic sampling offers no real improvements over simple systematic sampling. The authors' conclusion that horizontal stratification is ineffective refers specifically to this context.

(2) Administrative convenience can be a valid reason when, for example, different sampling methods are required for different habitats of an estuary (e.g. shoals *vs* channels).

FIGURE 6. Histogram of values for the inverse power a in Equation 11. Each individual value is the result of fitting Equation 11 to the data for a single variable (salinity, suspended particulate matter or chlorophyll a and transect.

(3) Stratification may also proceed along political boundaries, particularly when the issue is one of compliance with government regulations.

(4) Division into subdomains can also be motivated by the need to understand underlying causal mechanisms, in which case one might want to stratify on the basis of covariability of different spatial locations in time. In fact, previous research on the San Francisco Bay-Delta has clearly shown how different (overlapping) spatial subdomains can be identified with separate causal mechanisms through the use of rotated principal component analysis, a regionalization procedure common in meteorology (Jassby & Powell, 1994; Cloern & Jassby, 1995).

Tree-based regression is one of many approaches to the problem of grouping objects (in this case, locations) into subgroups according to their similarity. Legendre (1987) has reviewed a number of these other techniques, some similar to tree-based regression, that respect spatial contiguity, i.e. that give weight to proximity in space as well as to similarity in magnitude. Several features of tree-based regression attracted us originally. First, by operating through a binary recursive partitioning, it automatically preserves spatial contiguity within subdomains. Second, although not so much a consideration in this study, it can be applied to higher-dimensional data. Finally, it is easily shown that the criterion used by tree-based regression to choose splits (Equation 1) is equivalent to maximizing:
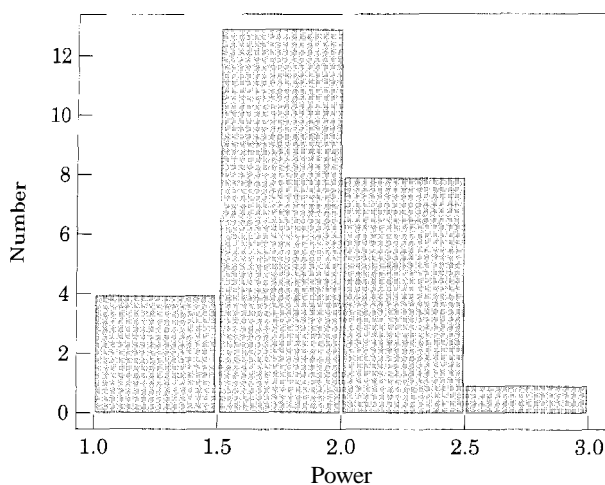
$$\Delta V = \frac{1}{n}[W_L(\mu_L-\mu)^2 + W_R(\mu_R-\mu)^2] \qquad (12)$$
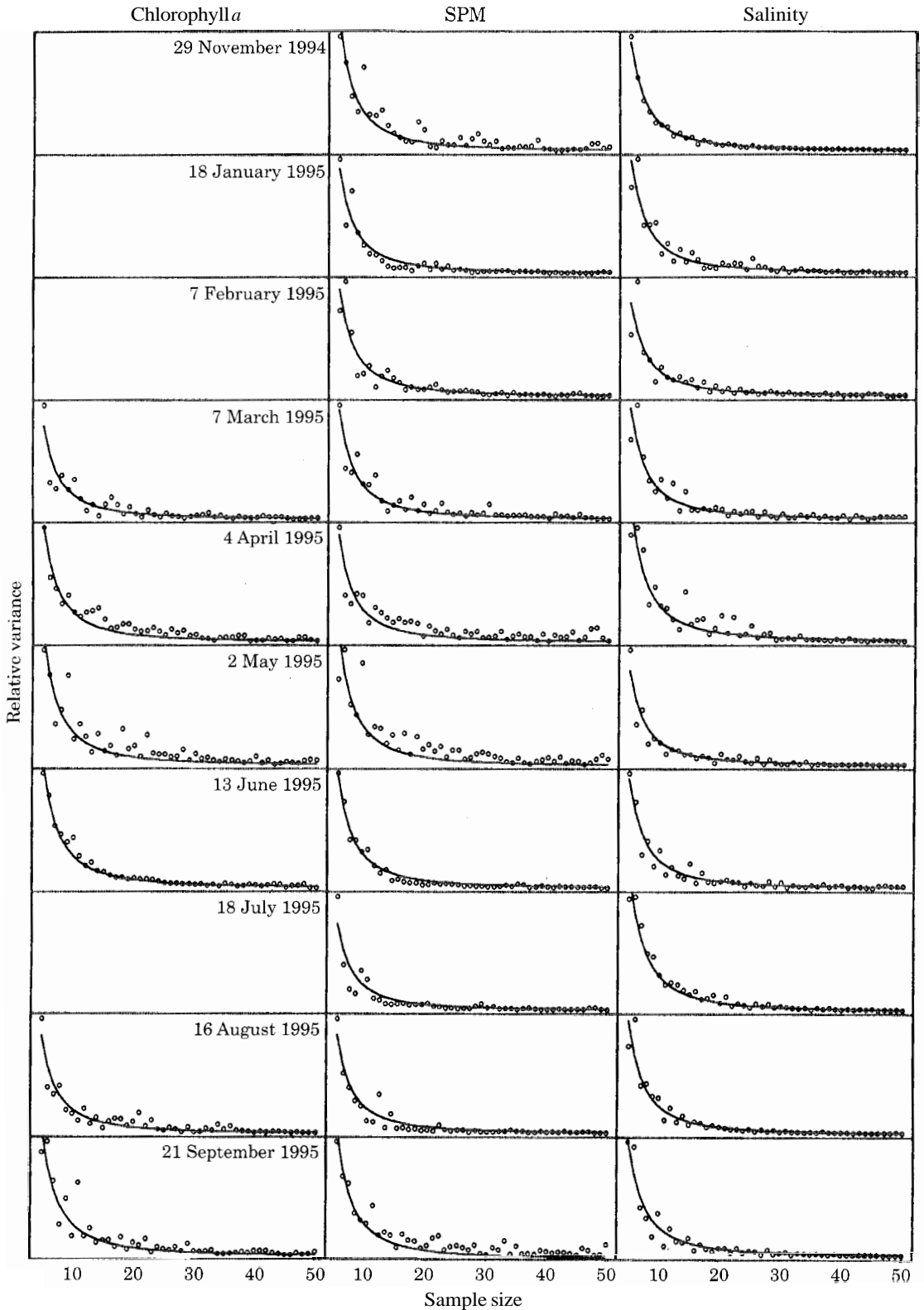
FIGURE 7. Variance (relative to maximum value) plotted against systematic sample size for each cruise (1–10) and variable. The lines are fitted inverse square curves. SPM, suspended particulate matter.
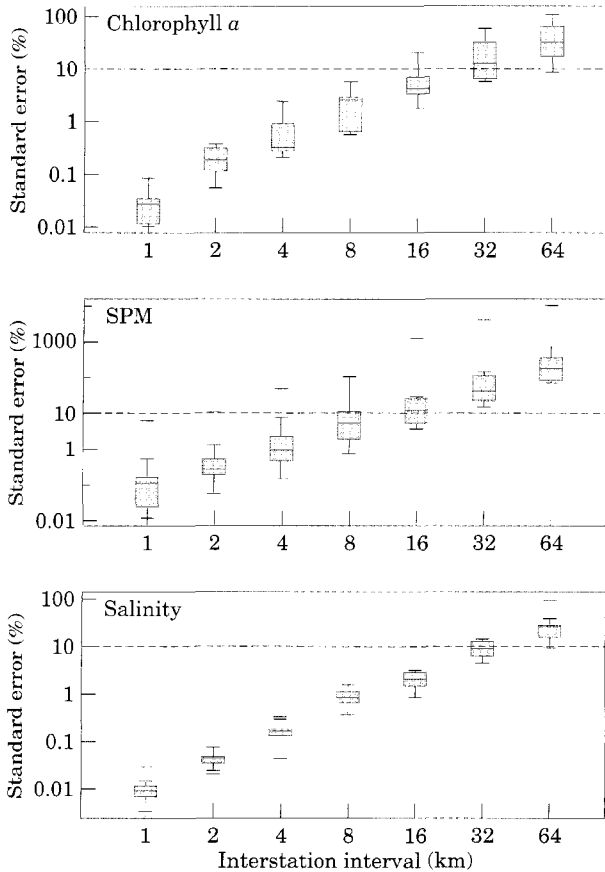
FIGURE 8. Relative standard error of the estimated transect mean from systematic samples as a function of the interstation distance. The boxplot for each distance represents the variability among transects. The interior horizontal line marks the median; the lower and upper box boundaries mark the first and third quartiles, respectively; the vertical lines extend to all points within 1·5 times the interquartile distance; more extreme points are shown by horizontal lines standing alone. SPM, suspended particulate matter.

which is the difference in the variances for simple random sampling and stratified random sampling with proportional allocation (Cochran, 1977). In other words, at each iteration, tree-based regression chooses the split that maximizes the benefits of stratified sampling.

The performance of tree-based regression may sometimes appear disappointing, specifically in the presence of strong gradients (e.g. the salinity panel in Figure 4); boundaries are laid down at apparently arbitrary locations on the gradient that have no distinguishing features. This behaviour, however, reflects the fact that the estuary often does not fit the implicit model of comprising homogeneous subdomains. Similar behaviour would be found with other techniques that partition to minimize the within-subdomain variability. In fact, tree-based regression was actually very

effective in guiding the authors' choice of stratum boundaries, considering that the resulting stratified sampling estimate decreased the variance of the estimated mean by 73 to 97% (Table 5). Note, however, that tree-based regression may not be appropriate for identifying subdomains in other contexts; as pointed out above, it fails to isolate transitional subdomains, tending to split them instead.

Sampling design

Despite the efficacy of stratified sampling, systematic sampling almost always yields a higher precision, regardless of the method of sample allocation for stratified sampling (Table 5). In theory, the relative performance of the different sampling designs depends on the properties of the underlying population (Cochran, 1977; Murthy & Rao, 1988). If the population is completely randomly arranged, systematic sampling is no better than simple random sampling. For a population dominated by a linear trend, stratified random sampling is the most efficient. For a population varying periodically in space, performance of the systematic sample depends on the interstation interval: if the sampling interval is divisible by the wavelength, estimates will be highly inefficient; on the other hand, if the sampling interval is an odd multiple of half the wavelength, estimates will be highly efficient. For populations with just serial correlation, the results depend on the nature of the spatial covariance structure. For example, Hajek (1959) extended the earlier results of Cochran (1946) to show that, in the case of stationary populations, systematic sampling minimizes the variance of the sample mean as long as the spatial correlation function is positive, decreasing and convex.

The study transects do not fall clearly into any of these ideal categories. Many of the features of the large-scale variability are clearly related to large-scale structural aspects of the estuary basin, such as the transition from the narrow Carquinez Strait to open San Pablo Bay in the vicinity of Mare Island or the shallow expanse of the San Bruno Shoal in the southern Bay (Figures 3 and 4). The large-scale variability, therefore, is most properly treated as a 'deterministic' spatial trend. In more confined reaches or on smaller scales, many of the aforementioned special cases may apply. For example, a linear trend in fluorescence occurs on the 10-km scale between San Bruno Shoal and the Dumbarton Bridge (Figure 3), while stationary time series models incorporating serial correlation appear to be appropriate for this series on the scale 1 km and smaller. These results, therefore, demonstrate the robustness of systematic sampling for a range of spatial variability types found in estuaries.

Note that a stratified systematic design offers only modest improvements at best, and sometimes even worse precision than unstratified systematic sampling (Table 5). As the systematic samples for different strata are chosen independently, stations from different strata may fall very close together near boundaries between strata and provide redundant information, also a failing of both the simple and stratified random design. The improvements are too modest to warrant the additional complications of stratified systematic sampling.

Variance estimators

The variance estimators with the lowest MSE (Equation 10)—COCH, CHEV and GUNS--are those specifically devised to account for spatial autocorrelation in the population (Table 1). Estimator CHEV is recommended on the basis of its empirical performance; interestingly, it was designed specifically for linear systematic samples (Chevrou, 1976) and the results here attest to the success of that design. Estimator COCH would also be a good choice. It is one devised by Cochran (1946) for stationary populations with an autoregressive structure of order one, equivalent to an exponential correlogram or variogram. As discussed below, the correlograms commonly encountered in these data are in fact exponential. Wolter (1984) observed that this estimator had remarkably good properties for artificial populations that are dominated by linear trends or autocorrelation. On the other hand, it was found that treating the systematic sample as if it were a simple random sample (SRS) leads to a poor estimate of the variance. Note also that *MURT1* and *MURT2* turned out to be relatively inefficient for these estuarine data, in contrast to their superior performance for demographic data (Wolter, 1984).

Sample size

The relation between variance of the mean and sample size is well described by an inverse power law with an average power of 1.9 ± 0.1 and more than 80% of the cases occur in the range 1·5–2·5 (Figure 6). It can be shown that the power is related to the nature of the variogram (Simard et al., 1992). Let n systematic samples be taken along the transect distance T, so $L=T/n$ is the distance between samples. Suppose for distances h up to L, the variogram has the fonn:

$$\gamma(h) = ch^b, \text{ with } 0 \leq b < 2$$

where c is a constant. Then one can derive for the one-dimensional case (namely, a transect) that (D. Marcotte, pers. comm.):

$$\text{var}(\bar{y}) \propto \frac{1}{n^{b+1}}$$

Now in the case of exponential, linear or spherical variograms (Isaaks & Srivastava, 1989), provided that the range is much larger than L, $b=1$ and an inverse square law would be expected. For Gaussian variograms, which have parabolic as opposed to linear behaviour near the origin, $b=2$, and for a nugget effect, $b=0$, so that a different power law would hold in these cases. The present finding of an inverse square law is therefore in agreement with past studies of spatial correlation in San Francisco Bay, which have demonstrated the presence of exponential spatial correlation (Powell et al., 1986).

How widely can the inverse square relationship be applied to other systems? Based on the above arguments, this question can be rephrased by asking how representative are variograms with linear behaviour near the origin (i.e. linear, exponential and spherical variograms). A nugget effect has been observed for water quality variables in both estuarine (Legendre & Trousellier, 1988; Legendre et al., 1989; Simard et al., 1992) and coastal waters (Denman & Freeland, 1985; Yoder et al., 1987). The shapes of the variograms, however, were usually exponential or at least compatible with an exponential shape where the resolution was too poor to be certain. Furthermore, the nugget effect may represent sampling error and not be an inherent feature such as high short scale variability (Isaaks & Srivastava, 1989); with more precise measurement techniques, the nugget effect could weaken or disappear. Nonetheless, the evidence from estuaries on variogram shape is sparse. Parabolic behaviour has been observed, moreover, in at least one other tidal estuary; North Inlet, South Carolina (Childers et al., 1994). Based on the existing evidence, an inverse square law cannot therefore be assumed for other estuaries, and the actual power could lie between 1 and 3. There is clearly a need here for expanding the empirical knowledge of estuarine spatial autocorrelation. Theoretical investigation of the link between the variogram or correlogram and underlying physical and biological processes could also help resolve the generality of any sampling design.

Where high-resolution data such as the MIDAS data are not available, it may be possible to take a geostatistical approach both to the variance estimate and the relation between precision and sample size. Based on a model of the underlying spatial autocorrelation, i.e. the variogram, kriging methodology

provides optimal point or global estimates, including the precision of the estimates. It has been widely used in ecology (Rossi *et al.*, 1992) and sometimes applied to estuaries (Simard *et al.*, 1992). It can also be used to generate an empirical relation between variance and sample size (Burgess *et al.*, 1981; Oliver & Webster, 1991). A number of important conditions must be satisfied, however, and as many as 200 stations may be required to properly define the variogram for kriging (Oliver & Webster, 1991), much larger than the number of stations typically constituting an estuarine sampling programme. With too few points, it may be impossible to resolve behaviour near the origin and variance estimates can then be unreliable (Thioulouse *et al.*, 1993). In that case, the variance estimator suggested by this study provides a useful alternative, although how to scale the variance for different sample sizes will remain uncertain without knowledge of the variogram shape.

Despite the central importance of the mean for both theoretical and practical reasons, as pointed out in the Introduction, it is not the most relevant statistic for many water quality variables. In the case of pollutant indices such as fecal coliforms, for example, the proportion of the population exceeding some specific level is the characteristic of interest. A global mean that falls within sanitary guidelines may disguise locally important water quality problems. Further work should, therefore, consider not only how to generalize these results regarding regional means to other estuaries, but also the need for similar analyses on other population statistics such as quantiles.

Concluding remarks

The results of this study provide guidelines for estimating estuary-wide means with low-resolution data from fixed stations. Given any desired precision for the estimate, the first step is to take systematic samples with as many stations as practical. Next, the estimator *CHEV* or COCH (Table 1) is used to calculate the variance of each sample, which will of course vary somewhat from transect to transect even when the number of stations is constant (Figure 8). Finally, the desired station number is determined from the typical or characteristic variance found in the previous step, the target variance, and the inverse power relation between variance and sample size. At present, the actual value for the power must come from prior knowledge of the variogram or correlogram shape or, if enough stations are used, by calculating the variogram from the initial array of stations. Further research may reveal some general rules for deducing variogram shape or this power from features of estuarine dynamics that can be observed with fewer stations.

References

Barnett, V. 1991 *Sample Survey: Principles and Methods.* Edward Arnold, London, 173 pp.

Bates, D. M. & Chambers, J. M. 1992 Nonlinear models. In *Statistical Models* in *S* (Chambers, J. M. & Hastie, T. J., eds). Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, California, pp. 421–454.

Bellhouse, D. R. 1988 Systematic sampling. In *Handbook of Statistics*, vol. 6 (Krishnaiah, P. R. & Rao, C. R., eds). Elsevier Science Publishers, Amsterdam, pp. 125–145.

Burgess, T. M., Webster, R. & McBratney, A. B. 1981 Optimal interpolation and isarithmic mapping of soil properties. IV. Sampling strategy. *Journal of Soil Science* **32,** 643–659.

Chatterjee, S. 1967 A note on optimum stratification. *Skandinavisk Aktuarietidskrift* 50, 40–44.

Chevrou, R. B. 1976 Précisions des mesures de superficie estimee par grille de points ou intersections de parallèles. *Annales des Sciences Forestières* **33,** 257–269.

Childers, D. L., Sklar, F. H. & Hutchinson, S. E. 1994 Statistical treatment and comparative analysis of scale-dependent aquatic transect data in estuarine landscapes. *Landscape Ecology* **9,** 127–141.

Clark, L. A. & Pregibon, D. 1992 Tree-based models. In *Statistical Models in S* (Chambers, J. M. & Hastie, T. J., eds). Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, California, pp. 377–419.

Cloern, J. E. 1996 Phytoplankton bloom dynamics in coastal ecosystems a review with some general lessons from sustained investigation of San Francisco Bay (California, USA). *Reviews of Geophysics* **34,** 127–168.

Cloern, J. E. & Jassby, A. D. 1995 Year-to-year fluctuation of the spring phytoplankton bloom in South San Francisco Bay: an example of ecological variability at the land-sea interface. In *Ecological Time Series* (Steele, J. H. & Powell, T. M., eds). Chapman and Hall, New York, pp. 139–149.

Cloern, J. E. & Nichols, F. H. 1985 Temporal dynamics of an estuary: San Francisco Bay. *Hydrobiologia* **129**, 1–237.

Cochran, W. G. 1946 Relative accuracy of systematic and stratified random samples for a certain class of populations. *Annals of Mathematical Statistics* **17**, 164–177.

Cochran W. G. 1977 *Sampling Techniques,* 3rd ed. John Wiley & Sons, New York, 428 pp.

Davis J. A., Gunther, A. J., Richardson, B. J., O'Connor, J. M., Spies, R. B., Wyatt, E., Larson, E. and A. Chan Meiorin. 1991 *Status and Trends Report on Pollutants in the San Francisco Estuary*. San Francisco Estuary Project, Oakland, California, 240 pp.

Denman, K. L. & Freeland, H. J. 1985 Correlation scales, objective mapping and a statistical test of geostrophy over the continental shelf. *Journal of Marine Research* **43**, 517–539.

Edmunds J. L., Cole, B. E., Cloern, J. E., Caffrey, J. M. & Jassby, A. D. 1995 *Studies of the San Francisco Bay, California, Estuarine Ecosystem: Pilot Regional Monitoring Results, 1994.* Open-File Report 95–378. U.S. Geological Survey, Menlo Park, California, 436 pp.

Gunderson, H. J. G. & Jensen, E. B. 1987 The efficiency of systematic sampling in stereology and its prediction. *Journal of Microscopy* **147**, 229–263.

Hájek, J. 1959 Optimum strategy and other problems in probability sampling. *Casopis Pro Pestovani Matematiky* **84**, 387–420.

Huzzey, L. M., Cloern, J. E. & Powell, T. M. 1990 Episodic changes in lateral transport and phytoplankton distribution in south San Francisco Bay. *Limnology and Oceanography* **35**, 472–478.

Isaaks, E. H. & Srivastava, R. M. . 1989 *An Introduction to Applied Geostatistics*. Oxford University Press, New York.

Jassby, A. D. & Powell, T. M. 1994 Hydrodynamic influences on interannual chlorophyll variability in an estuary: upper San Francisco Bay-Delta (California, U.S.A.). *Estuarine, Coastal and Shelf Science* **39**, 595–618.

Jassby, A. D., Cloern, J. E. & Powell, T. M. 1993 Organic carbon sources and sinks in San Francisco Bay: variability induced by river flow. *Marine Ecology Progress Series* **95**, 39–54.

Jassby, A. D., Cloern, J. E., Caffrey, J. M., Cole, B. E. & Rudek, J. 1994 San Francisco Bay/Delta Regional Monitoring Program: plankton and water quality pilot study 1993. In *San Francisco Estuary Regional Monitoring Program for Trace Substances: 1993 Annual Report.* San Francisco Estuary Institute, Richmond, California, pp. 117–128.

Jassby, A. D., Kimmerer, W. J., Monismith, S. G. *et al.* 1995 Isohaline position as a habitat indicator for estuarine populations. *Ecological Applications* **5**, 272–289.

Koop, J. C. 1971 On splitting a systematic sample for variance estimation. *Annals of Mathematical Statistics* **42**, 1084–1087.

Legendre, P. 1987 Constrained clustering. In *Developments in Numerical Ecology* (Legendre, P. & Legendre, L., eds). Springer-Verlag, Berlin, pp. 289–307.

Legendre, P. & Troussellier, M. 1988 Aquatic heterotrophic bacteria: Modeling in the presence of spatial autocorrelation. *Limnology and Oceanography* **33**, 1055–1067.

Legendre, P., Troussellier, M., Jarry, V. & Fortin, M.-J. F. 1989 Design for simultaneous sampling of ecological variables: from concepts to numerical solutions. *Oikos* **55**, 30–42.

Lorenzen, C. J. 1967 Determination of chlorophyll and pheo-pigments: Spectrophotometric equations. *Limnology and Oceanography* **12**, 343–346.

Madden, C. J. & Day, J. W. Jr. 1992 An instrument system for high speed mapping of chlorophyll a and physico-chemical variables in surface waters. *Estuaries* **15**, 421–427.

Mahalanobis, P. C. 1946 Recent experiments in statistical sampling in the Indian Statistical Institute. *Journal of the Royal Statistical Society, Series A* **109**, 325–370.

Mattfeldt, T. 1989 The accuracy of one-dimensional systematic sampling. *Journal of Microscopy* **153**, 301–313.

Murthy, M. N. & Rao, T. J. 1988 Systematic sampling with illustrative examples. In *Handbook of Statistics*, vol. 6 (Krishnaiah, P. R. & Rao, C. R., eds). Elsevier Science, Amsterdam, pp. 147–185.

Oliver, M. A. & Webster, R. 1991 How geostatistics can help you. *Soil Use and Management* **7**, 206–217.

Peterson, D. H., Conomos, T. J., Broenkow, W. W. & Doherty, P. C. 1975 Location of the non-tidal current null zone in northern San Francsico Bay. *Estuarine and Coastal Marine Science* **3**, 1–11.

Powell, T. M., Cloern, J. E. & Walters, R. A. 1986 Phytoplankton spatial distribution in South San Francisco Bay: mesoscale and small-scale variability. In *Estuarine Variability* (Wolfe, D. A., ed.). Academic Press, London, pp. 369–383.

Powell, T. M., Cloem, J. E. & Huzzey, L. M. 1989 Spatial and temporal variability in South San Francisco Bay (USA). I. Horizontal distributions of salinity, suspended sediments, and phytoplankton biomass and productivity. *Estuarine, Coastal and Shelf Science* **28**, 583–597.

Rossi, R. E., Mulla, D. J., Journel, A. G. & Franz, E. H. 1992 Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecological Monographs* **62**, 277–314.

Simard, Y., Legendre, P. & Lavoie, G. M. D. 1992 Mapping, estimating biomass, and optimizing sampling programs for spatially autocorrelated data: case study of the northern shrimp (*Pandalus borealis*). *Canadian Journal of Fisheries and Aquatic Sciences* **49**, 32–45.

Statistical Sciences 1995 *S-PLUS, Version 3.3 for Windows.* Statistical Sciences, Seattle, Washington.

Thioulouse, J., Royet, J. P., Ploye, H. & Houllier, F. 1993 Evaluation of the precision of systematic sampling: nugget effect and covariogram modelling. *Journal of Microscopy* **172**, 249–256.

Walser, W. E. Jr., Hughes, R. G. & Rabalais, S. C. 1992 Multiple interface data acquisition speeds at-sea research. *Sea Technology* **33**, 29–34.

Wolter, K. M. 1984 An investigation of some estimators of variance for systematic sampling. *Journal of the American Statistical Association* **79**, 781–790.

Yates, F. 1960 *Sanzpling Methodsfor Censuses and Surveys*, 3rd ed. Charles Griffin and Company, London, 440 pp.

Yoder, J. A., McClain, C. R., Blanton, J. O. & Oey, L. Y. 1987 Spatial scales in CZCS-chlorophyll imagery of the southeastern U.S. continental shelf. *Limnology and Oceanography* **32**, 929–941.