

**MODELING PHYTOPLANKTON:
COVARIANCE AND VARIOGRAM MODEL SPECIFICATION
FOR PHYTOPLANKTON LEVELS IN LAKE MICHIGAN**

L.J. WELTY and M.L. STEIN
*Department of Statistics, University of Chicago
5734 South University Avenue, Chicago, IL, USA*

Abstract. Algae and phytoplankton are crucial elements of marine ecosystems and of the global carbon cycle, which engenders widespread interest in better understanding their spatial and temporal variability. In situ fluorometry provides detailed measurements of phytoplankton levels; appropriate statistical models are necessary in order to elicit information about the distribution of phytoplankton biomass from this data. Challenges associated with such a data analysis include covariance model specification for processes in which variation in the vertical and horizontal directions differ greatly. Though the ideas presented here were developed with an eye to understanding phytoplankton dynamics, they may be helpful in developing models for other geophysical and environmental processes measured along vertical and horizontal dimensions.

1. Introduction

Phytoplankton, unicellular algae found primarily in oceans and lakes, are important components of marine ecosystems. They are at the base of the food chain, so insufficient numbers mean that few other species can survive. Excessive algal growth, common in nutrient rich polluted marine waters, may squeeze these same species out. Furthermore, most of the carbon fixation that occurs in the oceans is due to phytoplankton respiration and accounts for a sizable portion of global carbon fixation (Falkowski *b*).

Chlorophyll is found in algae and phytoplankton, and because it may be measured using a variety of methods, it is often tracked as an indicator of algal biomass. Determining chlorophyll content by discrete water samples alone is an accurate but inefficient method for obtaining detailed descriptions of phytoplankton dynamics. A more efficient method for measuring chlorophyll is by chlorophyll fluorescence.

In situ chlorophyll fluorescence measurements capture chlorophyll levels over large spatial scales in real-time. When exposed to blue light near 430

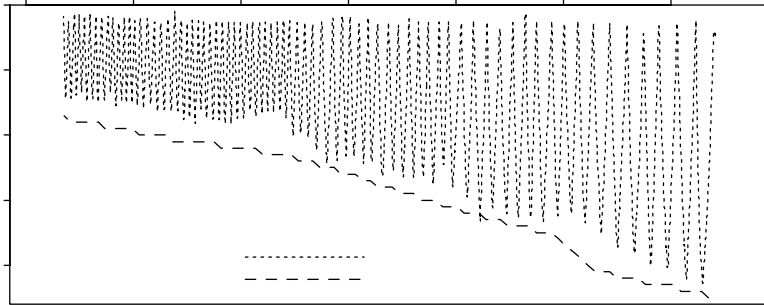


Figure 1. Example of the sawtooth-like collection scheme for chlorophyll fluorescence measurements. Numbers indicate the locations of the first and last measurements for this collection, which was taken at the southern tip of Lake Michigan in mid-March of 2000.

nm, chlorophyll emits red light near 680 nm (Falkowski *a*). The strength of the emission is roughly linearly related to chlorophyll. In situ fluorometers provide a real-time voltage output (approximately one observation per second, for instance), and when towed through lake or ocean waters provide a more complete picture of chlorophyll dynamics than water samples alone. The voltage output is calibrated to chlorophyll level by water samples taken along with fluorescence measurements.

The fluorescence profiles used in this research were obtained in the lower basin of Lake Michigan approximately six times per year from 1998 through 2000 as part of EEGLE: Episodic Events Great Lakes Experiment. Data collections consisted of towing a fluorometer in an undulating fashion from surface to bottom and bottom to surface repeatedly for approximately 25 kilometers along transects extending from shore toward the lake's center (Figure 1).

2. Axial dependence of phytoplankton

Unlike the most commonly used spatial models that assume isotropy, any tenable model for phytoplankton or similar marine measurements must account for the distinctly different processes along the horizontal and vertical axes. Variables affecting phytoplankton (temperature, water mixing, nutrient levels, to name only a few) change differently depending on if one moves roughly parallel to the surface/bottom or toward the surface/bottom.

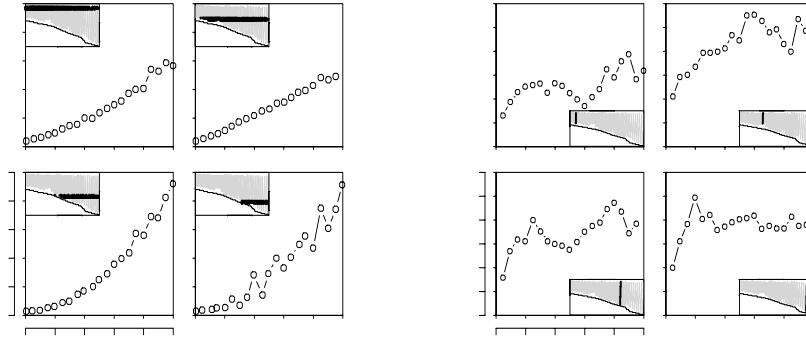


Figure 2. Empirical variograms to investigate the variability in the vertical and horizontal directions for the southern Lake Michigan fluorescence measurements. The horizontal empirical variograms $\hat{\gamma}(h)$ are calculated using $\log(\text{fluorescence})$ for measurements within three meter depth bands, and distance is calculated using difference in horizontal position only. Insets show the three meter depth band used. The vertical empirical variograms $\hat{\gamma}(v)$ are calculated using $\log(\text{fluorescence})$ for measurements on the same run (either a trip of the fluorometer from bottom to surface or from surface to bottom), and distance is calculated using difference in vertical position only. Insets show the run of measurements used for the corresponding empirical variogram.

For example, during the spring and summer months, the surface layer in Lake Michigan is warmed by solar radiation and becomes thermally isolated from the deeper and cooler waters. The temperature differential inhibits water mixing between the two layers, so that the chlorophyll levels may vary significantly through the water column. During this time, a five meter change in vertical position may result in a much greater change in chlorophyll level than a kilometer change in the horizontal.

3. Variability in the horizontal and vertical dimensions

We consider 13,815 fluorescence measurements collected in mid-March of 2000 at the southern tip of Lake Michigan. Temperature measurements taken simultaneously are nearly constant in the vertical, and decrease from 4.5°C to 2.5°C as we move away from shore. The sawtooth-like collection scheme (Figure 1) prevents us from completely separating the horizontal variability from the vertical variability in our measurements, so as an approximate method for investigating the variability in each direction we subdivide the data into regions with small range in depth and regions with small range in distance from shore. We calculate empirical variograms based on distance from shore for observations that are no more than three

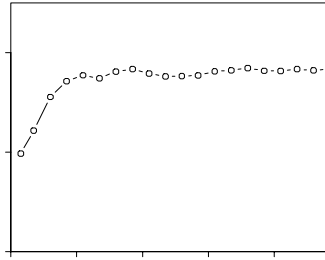


Figure 3. Average across all runs of the empirical variograms (each calculated as $\hat{\gamma}(v)$ in Figure 2) for the vertical dimension. Here $\bar{\gamma}(v) = \frac{1}{146} \sum_{i=1}^{146} \hat{\gamma}_i(v)$, where $\hat{\gamma}_i(v)$ is the empirical variogram based on depth for run i and we have averaged over all 146 runs.

meters apart vertically, as well as empirical variograms based on depth for observations that are in the same run – either a pass up from bottom to surface or a pass down from surface to bottom (Figure 2). In addition, we calculate the average across all runs of the empirical variograms based on depth (Figure 3).

The empirical variograms show that there is considerably greater variability along the horizontal dimensions than the vertical. The process along the horizontal appears nonstationary with no discernible sill, while the process along the vertical appears stationary with a sill near 0.002 (see Figure 3). The horizontal process appears to have range greater than 5 km while the vertical process appears to have range near 1 m. There is no obvious evidence that the horizontal variogram depends on depth or the vertical variogram on distance from shore, so an intrinsic model of order zero may be appropriate.

4. Inappropriateness of traditional models

We consider our log(fluorescence) measurements as observations from a random field $Z(h, v)$ on \mathbb{R}^2 , where h is horizontal distance from shore (measured in kilometers), v is depth (measured in meters), and our specific region of \mathbb{R}^2 is the approximately triangular area bounded by the water surface and lake bottom. For notational simplicity, we do not consider any nugget effects in what follows. (The empirical variograms suggest that our model should include a nugget of approximately 0.0009).

We have observed that $Z(h, v)$ is a badly anisotropic process. Linear transformation of one or more of the dimensions also fails to produce tenable

variogram or covariance models for $Z(h, v)$. If the variogram for $Z(h, v)$ had the form

$$\gamma((h_1, v_1), (h_2, v_2)) = f\left(\sqrt{h^2 + \alpha^2 v^2}\right)$$

where $|h_1 - h_2| = h$, $|v_1 - v_2| = v$, and where f is conditionally negative definite, then $\gamma((h_1, v_1), (h_2, v_1)) = f(h)$ and $\gamma((h_1, v_1), (h_1, v_2)) = f(\alpha v)$. Since f must have the same sill in both directions, this model cannot possibly describe the variograms shown in Figures 2 and 3.

5. Specifying valid models

Before formulating more complex models, we consider what functions make valid covariance and variogram models. Having $K(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$ positive definite is a necessary and sufficient condition for $K(\mathbf{x})$ to be the covariance function of a weakly stationary random field on \mathbb{R}^d . If $K_1(x)$ and $K_2(y)$ are valid covariance functions for $x, y \in \mathbb{R}$, then for $a, b > 0$, $aK_1(x) + bK_2(y)$ and $K_1(x)K_2(y)$ are valid covariance functions on \mathbb{R}^2 .

Analogously, $\gamma(\mathbf{x})$ is a valid variogram model on \mathbb{R}^d if it is conditionally negative definite on \mathbb{R}^d . If $\gamma(\mathbf{x})$ is a valid variogram model, then so is $b\gamma(\mathbf{x})$ for $b > 0$ (Cressie). If $\gamma_1(x)$ and $\gamma_2(y)$ are valid variogram models on \mathbb{R} , then $\gamma(x, y) = \gamma_1(x) + \gamma_2(y)$ is valid as well (this follows directly from the conditional negative definiteness). It is generally *not* the case that $\gamma_1(x)\gamma_2(y)$ will be a valid variogram. As an illustration, suppose W is a stationary Gaussian random field on \mathbb{R}^2 with variogram model $\gamma(x, y) = b\gamma_1(x)\gamma_2(y)$, where b is nonnegative and $\gamma_1(0) = \gamma_2(0) = 0$. Then

$$\begin{aligned} \text{Var}\{W(x, y) - W(x, 0) - W(0, y) + W(0, 0)\} \\ &= 4b\gamma(x, 0) + 4b\gamma(0, y) - 4b\gamma(x, y) \\ &= -4b\gamma_1(x)\gamma_2(y), \end{aligned}$$

which is nonnegative if and only if $b = 0$.

6. Potential models

Given the axial dependence of the fluorescence measurements, it seems natural to consider a tensor product approach with covariance function for $Z(h, v)$ of the form

$$\text{Cov}\{Z(h_1, v_1), Z(h_2, v_2)\} = K_H(h)K_V(v) + K_H(h)$$

where K_H and K_V are the covariance models for the horizontal and vertical directions respectively. Then $\text{Cov}\{Z(h_1, v_1), Z(h_2, v_2)\}$ is positive definite as long as K_V and K_H are. This form allows for most of the variation to

occur in the horizontal dimension as we require, but cannot be well specified when the variation along one dimension (in our case, the horizontal) is nonstationary.

One might also consider the model

$$\gamma((h_1, v_1), (h_2, v_2)) = \gamma_H(h) + \gamma_V(v),$$

which allows for different (possibly nonstationary) variogram models in the horizontal and vertical dimensions. As Chilès and Delfiner discuss, this model treats the process as exactly additive, *i.e.*

$$\begin{aligned} \text{Var}\{Z(h, v) - Z(h, 0) - Z(0, v) + Z(0, 0)\} \\ &= 4\gamma_H(h) + 4\gamma_V(v) - 4[\gamma_H(h) + \gamma_V(v)] \\ &= 0, \end{aligned}$$

taking $\gamma_H(0) = \gamma_V(0) = 0$ (Chilès). It seems unwise to use the above model unless one is quite sure that $Z(h, v)$ is exactly additive.

7. Revised potential models

Modifications to the above approaches do suggest tenable variogram models. First, consider

$$\gamma((h_1, v_1), (h_2, v_2)) = \gamma_H(h) - K_H(h)K_V(v) + K_H(0)K_V(0) \quad (1)$$

where γ_H is a variogram model for the horizontal direction, and K_H and K_V are covariance models for the horizontal and vertical directions respectively (note that the two terms are not equivalent to $\gamma_H(h)\gamma_V(v)$). Taking both K_H and K_V positive definite results in $\gamma((h_1, v_1), (h_2, v_2))$ conditionally negative definite as required. The first term γ_H accounts for the nonstationarity in the horizontal direction, and the remaining terms for the interaction between horizontal and vertical variability.

Another reasonable model would be

$$\gamma((h_1, v_1), (h_2, v_2)) = \gamma_H(h) + \gamma_R\left(\sqrt{h^2 + \alpha^2 v^2}\right). \quad (2)$$

Again the first term accounts for the nonstationarity in the horizontal direction, but here the second term requires that the remaining variability be attributed to geometric anisotropy. This model offers some flexibility and simplicity over (1) in that it does not require specification or existence of covariance functions in the vertical and horizontal directions. Model (1) is however quite different from model (2) in that it treats the process as locally nearly additive (Stein, 2.11).

TABLE I. Variogram parameter estimates for M_T and M_G obtained by maximizing the approximate restricted log likelihood \tilde{R} for the log(fluorescence) observations \mathbf{z} . Estimates for parameters describing the long range horizontal dependence are quite similar, while estimates for parameters describing the local behavior of the process differ markedly. We consider M_G preferable to M_T as it has the larger approximate log likelihood.

Model	\tilde{R}	Parameter Estimates					
M_T	25527.26	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5
		3.93×10^{-4}	3.39×10^{-3}	1.22	1.43×10^{-3}	2.58×10^2	2.75
M_G	25531.66	ϕ_0	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5
		3.49×10^{-4}	3.35×10^{-3}	1.25	1.47×10^{-3}	3.04×10^2	3.29

8. Model comparison for fluorescence measurements

We take (1) to have the form

$$\gamma(h, v) = \theta_0 1_{\{h+v>0\}} + \theta_1 h^{\theta_2} - \theta_3 [1 - \mathcal{M}_\nu(\theta_4 h) \mathcal{M}_\nu(\theta_5 v)]$$

where θ_0 is the nugget effect, $\mathcal{M}_\nu(z) = 2^{1-\nu} z^\nu \mathcal{K}_\nu(z) / \Gamma(\nu)$, and \mathcal{K}_ν is a modified Bessel function (Abramowitz). The indicator function $1_{\{h+v>0\}}$ takes the value one if $(h+v) > 0$ and zero otherwise. This model, which we will denote $M_T(h, v; \boldsymbol{\theta})$ for its tensor product like last term, treats $\gamma_H(h)$ in (1) as a power law variogram and K_H and K_V in (1) as covariance functions from the Matérn class (Stein). We take (2) to have the form

$$\gamma(h, v) = \phi_0 1_{\{h+v>0\}} + \phi_1 h^{\phi_2} - \phi_3 \left[1 - \mathcal{M}_\nu \left(\sqrt{\phi_4^2 h^2 + \phi_5^2 v^2} \right) \right]$$

where again we have added a nugget effect ϕ_0 , $\gamma_H(h)$ in (2) is a power law variogram and γ_R in (2) is a Matérn class variogram. We denote this model by $M_G(h, v; \boldsymbol{\phi})$ for its treatment of the local behavior in the horizontal and vertical as geometrically anisotropic. We note that M_T and M_G are equivalent when $h = 0$ or $v = 0$.

Obtaining parameter estimates for M_T and M_G using exact likelihood methods is intractable given that each evaluation of the likelihood function requires $O(n^3)$ operations, and here $n = 13,815$. Our solution is to employ an approximate likelihood method similar to that proposed by A. Vecchia (Vecchia). For an observation vector $\mathbf{z} = (z_1, z_2, \dots, z_n)$, Vecchia noted that it may be possible to approximate the likelihood

$$L(\boldsymbol{\theta} | \mathbf{z}) = p(z_1 | \boldsymbol{\theta}) \prod_{i=2}^n p(z_i | z_{i-1}, \dots, z_1, \boldsymbol{\theta}),$$

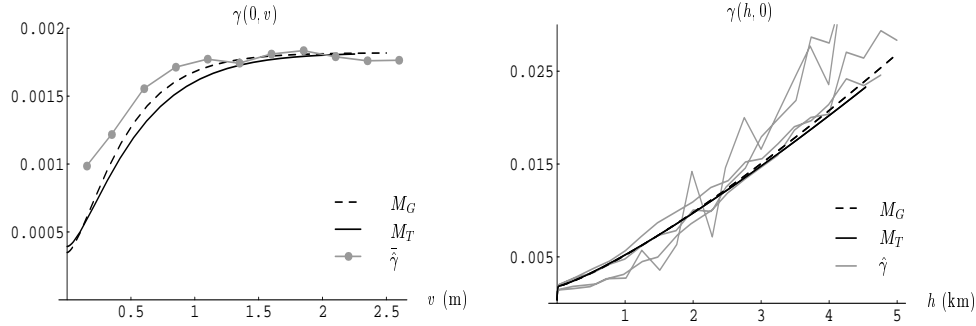


Figure 4. Implied horizontal and vertical variograms for $M_G(h, v; \hat{\phi})$ and $M_T(h, v; \hat{\theta})$ as well as empirical variograms $\hat{\gamma}(v)$ and $\hat{\gamma}(h)$, calculated as in Figures 2 and 3.

where p denotes probability density, by considering the conditional density of z_i on some subset of z_{i-1}, \dots, z_1 and hence reducing computation. Vecchia proposed conditioning on the m points in z_{i-1}, \dots, z_1 nearest to z_i , with m generally much less than $i - 1$ (e.g. $m = 5$). In order to account for long range spatial dependence, we alter Vecchia's scheme to condition on some points in z_{i-1}, \dots, z_1 near z_i as well as some points in z_{i-1}, \dots, z_1 that are much farther away from z_i . Estimating variogram parameters is our primary interest, so we extend Vecchia's idea to approximate the restricted log likelihood (Kitanidis) rather than the log likelihood. Forthcoming work by M. Stein, Z. Chi, and L. Welty details methodology and results for these extensions of Vecchia's work.

We therefore take our log(fluorescence) observations \mathbf{z} to be from a Gaussian random field with covariance structure given by $M_T(h, v; \theta)$ or $M_G(h, v; \phi)$, and let $\tilde{R}_T(\theta|\mathbf{z})$ and $\tilde{R}_G(\phi|\mathbf{z})$ represent the approximate restricted log likelihoods under the respective models. We order our observations $z_1, z_2, \dots, z_{13,815}$ by the order in which they were collected (Figure 1). With appropriate adjustments for small values of i , we select our conditioning subset for z_i to consist of ten nearby previous points as well as ten roughly evenly spaced observations from more distant observations. We maximize \tilde{R} using a conjugate gradient algorithm (Press, 10.6). For initial computational simplicity, we do not maximize over ν , the smoothness parameter for the Matérn covariance function. Based on comparisons of the likelihood for $\nu = 0.5, 1.0$, and 1.5 , as well as the shape of the empirical vertical variogram, we set $\nu = 1.0$. Results are shown in Table I.

That $\tilde{R}_G(\hat{\phi}|\mathbf{z}) > \tilde{R}_T(\hat{\theta}|\mathbf{z})$ suggests that $M_G(h, v; \phi)$ more reasonably describes the dependence structure of the log(fluorescence) measurements. The parameters describing the long range horizontal dependence are nearly the same for $M_G(h, v; \hat{\theta})$ and $M_T(h, v; \hat{\phi})$; the largest discrepancies in es-

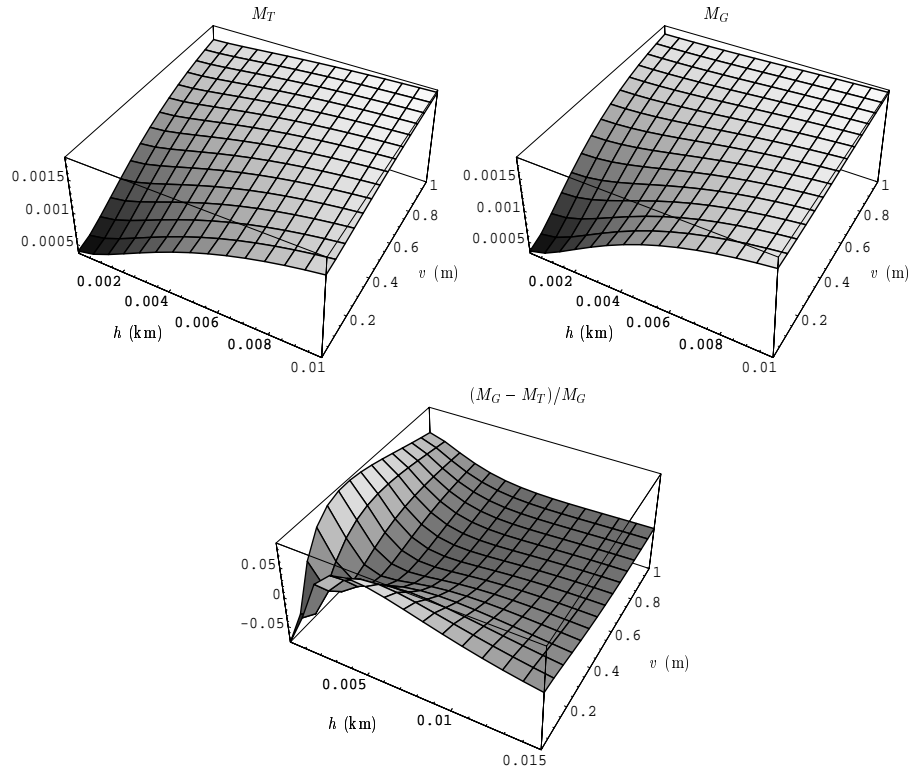


Figure 5. Estimated variograms $M_G(h, v; \hat{\phi})$ and $M_T(h, v; \hat{\theta})$, as well as their percent difference, illustrating the distinct treatments of the short scale horizontal and vertical variation.

estimates are for the nugget effect and fifth and sixth parameters, which describe the local variation of the process as well as the interaction between the horizontal and vertical dependence. One should note however that $1/\hat{\theta}_4 \approx 1/\hat{\phi}_4$ and $1/\hat{\theta}_5 \approx 1/\hat{\phi}_5$, so that the models do give similar estimates for the range of the vertical process and for the range of the local horizontal processes. The models predict nearly indistinguishable variograms in the horizontal direction for long ranges, and similar variograms along the vertical direction (Figure 4).

We note that the average empirical vertical variogram is slightly above either parameteric estimate (Figure 4), but that the difference is not necessarily an indication of model misfit. One possible reason for the discrepancy may be that $\tilde{\gamma}(v)$ contains some horizontal variation (recall we calculated each $\hat{\gamma}(v)$ using points in the same run, which will vary slightly in horizontal coordinate). The difference may also be due to sampling variability in the

empirical variogram. Points in empirical variograms are highly correlated, so empirical variograms contain significantly less information about processes than appearances suggest. For a more detailed discussion of this problem with empirical variograms and the advantages of using maximum likelihood, see (Stein).

Figure 5 shows the models' distinct treatments of the interaction between the local horizontal and vertical variation. Significant differences near the origin illustrate the distinction in modeling the local process as geometrically anisotropic versus as a product form. As one moves away from the origin, $M_G - M_T$ is nearly zero, suggesting little qualitative difference in the models for larger values of h and v .

9. Conclusions

We conclude that models of the form in (2) may more reasonably account for the axial dependence of our log(fluorescence) measurements, the distinctly different variability they exhibit along the axes, and the interactions of the variability along the horizontal and the vertical. The accuracy of interpolated values for chlorophyll fluorescence, which are crucial to calibrating the fluorescence values with water samples and hence to producing fields of predicted chlorophyll levels, may depend significantly on the relevance of our covariance model. In any case, it is important to consider alternative models to the usual isotropic or geometrically anisotropic suspects when data exhibit very different horizontal and vertical variability. One would expect such situations to arise not only for marine data, but also for geological data (where moving parallel to the the earth's surface is much different from tunneling toward its core), or for meteorological data (where variability through a layer of atmosphere may be much different than variability as one moves up through the atmosphere). Comparing models like those we have suggested here may also provide additional understanding of such processes.

Acknowledgements

The chlorophyll fluorescence data were produced as part of the Episodic Events Great Lakes Experiment (EEGLE) Program, supported by the National Oceanic and Atmospheric Administration's Coastal Ocean Program and the National Science Foundation. The first author was supported by a National Science Foundation Graduate Research Participant Fellowship. The second author was supported by the United States Environmental Protection Agency through STAR Cooperative Agreement #R-82940201-0 to the University of Chicago. This research has not been subjected to the

Agency's required peer and policy review and therefore does not necessarily reflect the views of the Agency, and no official endorsement should be inferred. The authors are greatly indebted to and wish to thank Barry Lesht of Argonne National Laboratory, Environmental Research Division, and Tom Johengen and Henry Vanderploeg of Great Lakes Environmental Research Laboratory, NOAA for providing access to and detailed information about the chlorophyll fluorescence data.

References

- Abramowitz, M. and Stegun, I. (1965) *Handbook of Mathematical Functions*, ninth ed. Dover, New York.
- Chilès, J and Delfiner, P. (1999) *Geostatistics*. Wiley, New York.
- Cressie, N. (1993) *Statistics for Spatial Data*. Wiley, New York.
- Falkowski, P.G., Kiefer, D.A. (1985) Chlorophyll *a* fluorescence in phytoplankton: relationship to photosynthesis and biomass, *Journal of Plankton Research*, **7**, pp. 715-731.
- Falkowski, P.G. (1994) The role of phytoplankton photosynthesis in global biogeochemical cycles. *Phytosynthesis Research*, **39**, pp. 235-258.
- Kitanidis, P.K. (1983) Statistical estimation of polynomial generalized covariance functions and hydrologic applications. *Water Resources Research* **19**, pp. 909-921.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P. (1992) *Numerical Recipes in C*. Cambridge University Press, Cambridge.
- Stein, M.L. (1999) *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York.
- Vecchia, A. V. (1988) Estimation and identification for continuous spatial processes, *Journal of the Royal Statistical Society B*, **50**, pp. 297-312.

