Optimization of water column monitoring: statistical treatments

Massachusetts Water Resources Authority

Environmental Quality Department
Technical Report Series No. 95-15

FINAL REPORT

OPTIMIZATION OF WATER COLUMN MONITORING: STATISTICAL TREATMENTS

by

Carlton D. Hunt[1]
Bruce E. Buxton[2]
Allen D. Pate[2]
Darlene E. Wells[2]

[1]Battelle Ocean Sciences
397 Washington Street
Duxbury, MA 02332
(617) 934-0571

[2]Battelle Statistics and Analysis Systems
505 King Avenue
Columbus, OH 43201-2693
(614) 424-4547

Prepared for
Massachusetts Water Resources Authority
Charlestown Navy Yard
100 First Avenue
Boston, MA 02129
(617) 242-6000

Environmental Quality Department Technical Report
Series No. 95-15

Citation:

Hunt, C.D. , B.E. Buxton, A.D. Pate, and D.E. Wells. 1995. **Optimization of water column monitoring: statistical treatments**. MWRA Enviro. Quality Dept. Tech. Rpt. Series No. 95-15. Massachusetts Water Resources Authority, Boston, MA. 106 pp.

EXECUTIVE SUMMARY

The Massachusetts Water Resources Authority (MWRA) Outfall Monitoring Project is being implemented such that post-discharge monitoring can be conducted in a scientifically sound, statistically defensible, cost-effective manner. As described in MWRA (1991), a baseline monitoring program was developed to provide information related to the spatial and temporal variability in the system for areas of concern identified by the regulatory, public and scientific communities. A substantial body of data has been developed between 1989 and 1994. Much of this information has been used to refine the continuing baseline monitoring design and will guide the Phase II post-discharge monitoring program. Statistical treatment of design has been completed for some monitoring areas (*e.g.,* fish and shellfish monitoring; Hillman *et al.,* 1994). However, other areas have not received sufficient statistical treatment such that the most cost-effective, environmentally responsive monitoring plan can be developed. Specifically, the water column monitoring program, which comprises the largest part of the baseline program, has not received focused statistical treatment that addresses the adequacy of the spatial and temporal design of the baseline program. To address this and to begin evaluation of post-discharge water column monitoring design, a series of exploratory statistical analyses was conducted. This report summarizes the exploratory statistical evaluations conducted on selected key water column variables related to biological processes in the water column (dissolved oxygen, chlorophyll *a*, total suspended solids, dissolved inorganic nitrogen) and other variables (salinity and temperature) used to support scientific interpretation of the key biological variables. The exploratory analyses were designed to address the following questions:

> What are the spatial scales of variability in key water column parameters measured in Massachusetts Bay?
>
> Are temporal and spatial scales for the Phase I baseline water column sampling adequate?
>
> What are the changes in the system that the present monitoring design can detect?
>
> Are these detectable changes sufficient to measure change before levels considered to be meaningful are approached?
>
> Are there alternative sampling designs that could be considered without loss of statistical confidence?

The evaluations that were conducted to address these questions included semivariogram analysis, kriging analysis, and reverse power analysis. The results show the following:

Semivariogram and Kriging Analysis — Application of semivariogram analysis and kriging, a geostatistical interpolation technique, to evaluate temporal and spatial variability in chlorophyll *a* concentrations in the nearfield surface (0-5 m depth) waters sampled in 1992 and 1993 suggests that the temporal variability is greater than spatial variability. These analyses also indicate that the detectable difference that can be determined for a single survey is similar to that obtained with 12 surveys when at least 12 nearfield stations are included in the sample design. Detectable differences of about 6.2 μg/L can be observed under either scenario. Improvements in estimation precision by increasing the number of annual surveys from 12 to 17 are negligible. This exploratory analysis suggests that the present nearfield monitoring design may oversample the nearfield area from a spatial perspective.

Evaluation of the spatial scales of variance based on data collected from throughout Massachusetts Bay showed differences in the level of variability between summer and winter conditions. The smallest detectable differences for chlorophyll were 0.27 μg/L in February and 3.7 μg/L in August at spatial scales of 2 km. The detectable difference increased by about twofold at scales of 5 km (relative to the 2 km scale) and showed very little additional increase at scales of 15 km. Reduction of the number of stations sampled from 46 to 23 did not increase the detectable differences that could be measured during August, a period characterized by heterogeneous distributions of surface chlorophyll. However, further reducing the number of stations sampled to 11 caused the detectable difference to increase by about twofold over the current design at the 2 km scale. At the larger scales, decreasing the number of stations had little effect on the detectable differences that could be observed. Spatial scale and reduction in the number of monitoring stations had little effect on detectable differences during February when the surface chlorophyll in the system is relatively low and homogeneous.

Similar results were obtained when the photic zone was included in the estimation of detectable differences, although the relative increase in the detectable differences as stations were removed from the design did not increase to the extent observed for the surface samples.

Power Analysis — The change that could be detected over the entire water column of the nearfield as a percentage of the baseline mean concentrations was estimated to be $\approx$2% for dissolved oxygen, $\approx$25% for chlorophyll *a*, $\approx$45% for DIN, and $\approx$100% for total suspended solids. Seasonal variability in these values was clearly apparent. These results indicate that the present nearfield monitoring design is robust and can detect any level of change that might be considered meaningful in the system.

The power analyses suggest that the temporal and spatial sampling scales of the current water column baseline monitoring program are more than adequate to detect change. The results from these exploratory analyses suggest that the water column sampling program could be reduced both in its spatial coverage (number of stations), particularly in the nearfield, and possibly in the temporal frequency of sampling, without significant loss in the ability to detect differences.

The final Phase II water column monitoring design will, however, require clarification of the level of change that is to be detected, the temporal and spatial scales of concern, and the levels of meaningful change for this system. The statistical results from the exploratory analyses presented in this report suggest that, relative to present conditions, detectable change goals could be set at 10-20% for dissolved oxygen, 100-200% (2-3x) for chlorophyll, 100% for total suspended solids, and 100-200% for dissolved inorganic nitrogen, without loss in the ability to provide adequate warning of unacceptable change. Continued statistical evaluation using techniques such as kriging should be used to develop more information on optimal designs and their cost effectiveness. Continuing this type evaluation will, however, require that design criteria be agreed upon and that specific questions related to the monitoring be developed to guide further statistical analysis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.0 INTRODUCTION

1.1 Water Column Monitoring Program

The Massachusetts Water Resources Authority (MWRA) outfall monitoring program (MWRA, 1991) was developed to address regulatory, public, and scientific concerns related to the relocation of the Deer Island sewage effluent outfall from the mouth of Boston Harbor to Massachusetts Bay. The MWRA and the Outfall Monitoring Task Force (OMTF) that oversees the monitoring have committed to implementation of a scientifically sound, statistically valid, cost-effective monitoring program that can provide warning of unacceptable impact to the receiving waters with sufficient time to implement various investigative and management actions. Meeting this goal requires that the monitoring plan be based on a sound fundamental knowledge of the physical and ecological processes operating within the system, the expected interactions between the system and the discharge, and variability inherent to the system. An inadequate understanding of the ecological and physical processes operating in the Massachusetts Bay system and poorly quantified estimates of spatial and temporal variability in the Massachusetts Bay system compelled those charged with developing the outfall monitoring program to address this lack of understanding through an extensive baseline monitoring program (MWRA, 1991). The resultant baseline monitoring program (Phase I), implemented in 1992, was designed to develop information that could be used to (1) understand the functioning of the system, (2) establish quantitative measures of the spatial and temporal variability in the system for a large number of parameters related to either eutrophication or toxic chemicals, and (3) develop an understanding of the status of the system prior to relocation of the effluent discharge. Using the information developed during the baseline period, a post-discharge monitoring plan (Phase II) was to be developed.

In anticipation of the post-discharge monitoring program, in 1994 the MWRA held a series of workshops that specifically focused on issues central to the design of the monitoring program. These workshops addressed such topics as key monitoring parameters, meaningful change, detectable change, and measurement location and frequency (Hunt and Steinhauer 1994a, b; Hunt *et al.*, 1994). The discussions at these workshops identified key parameters that should be monitored to ascertain unacceptable impact and to provide warning of change in the system. The workshops also reinforced the need to address and determine levels of change that are meaningful for the system, and levels of change that can be detected under the present monitoring design. These issues were addressed for several program elements in reports that discuss the results of the 1993 monitoring. Specifically, detectable and meaningful change in fish and shellfish monitoring were addressed in Hillman *et al.* (1994). Other areas of the program, such as benthic monitoring, are still in the process of defining either the level of change that is meaningful and, even more fundamental, the design criteria and metrics that should be used to describe change.

Intense spatial and temporal monitoring of the water column (Albro *et al.*, 1993) has been ongoing since 1992. This, coupled with a clear understanding of the key water column measurement parameters to be monitored, enables this report to focus on statistical evaluation of these key parameters and to provide recommendations for the Phase II monitoring design. Exploratory statistical evaluations have been undertaken to address areas such as detectable levels of change, variability in the system (spatial and temporal scales), and an initial exploration of spatial and temporal sampling modifications that could result in a more cost-effective monitoring program, while retaining appropriate levels of statistical confidence and power. The issues addressed by these analyses ultimately require that certain elements (*e.g.*, statistical confidence and power) be defined, under which the water column monitoring program will operate. For the analyses that follow, the 95% confidence level and statistical power of 80% are used. These metrics are recommended for use throughout the program.

Finally, monitoring design is directly linked to the questions being asked of the monitoring, to scientific understanding, to levels of meaningful change, and to available resources. This report does not attempt to define meaningful levels of change for the water column of the Massachusetts/Cape Cod Bay system. Definition of this is left to others. This report also does not provide definite conclusions regarding the optimal monitoring design (*e.g.*, the number of stations, samples or frequency of sample collection). Such recommendations require more extensive analysis than could be accommodated under the resources provided and require, as indicated above, a set of clearly stated questions with defined statistical goals. When such information becomes available, it is recommended that the issues addressed here be reevaluated.

1.2 Report Objectives

Achievement of appropriate and optimal design for the water column monitoring program requires an understanding of the temporal and spatial variability in the system. Statistical evaluations designed to provide such information can take on may facets and, if not carefully controlled, can go in many directions. Thus, objectives for this effort were developed; they include the following:

- Develop preliminary information about temporal and spatial scales of variability
- Provide information on detectable change for various parameters
- Provide information useful for development of Phase II monitoring hypotheses
- Develop information useful for the design of Phase II water column monitoring

Narrative and graphical discussion of the variability in the water column data can be found in Kelly *et al.* (1992), Kelly *et al.* (1993a,b,c), Kelly *et al.* (1994 a,b,c,d,e,f,g), and Kelly *et al.* (1993). The statistical effort focused on the following questions:

- What are the spatial scales of variability in key water column parameters measured in Massachusetts Bay?

- Are temporal and spatial scales for the Phase I baseline nearfield water column sampling design adequate?

- What is the change in the system that the present monitoring design can detect?

- Are these detectable changes sufficient to measure change before levels considered to be meaningful are approached?

- Are there alternative sampling designs that could be considered without loss of statistical confidence?

Finally, the statistical techniques were examined to determine the best approach for evaluating variability and, at the same time, providing an ability to evaluate design options. The method chosen for this effort is a geostatistical technique known as kriging. This technique and reverse power analysis were used to develop information on detectable change in the system.

1.3 Report Content

Section 2 provides an overview of the statistical techniques and data sets used in the statistical treatments. Section 3 presents the results and discussion of the semivariogram and kriging analyses, detectable changes as a function of spatial scales, and estimates of detectable change based on reverse power analysis. Section 4 briefly summarizes information related to each of the above questions and provides a set of recommendations for the Phase II water column monitoring. Note that the results presented in this report do not give final answers to the questions posed above. Rather, they provide information that can be used to support programmatic decisions and guide further statistical evaluations necessary to determine the optimal design of the water column monitoring.

2.0 METHODS

Several statistical techniques were evaluated to determine the most appropriate method for addressing these exploratory analyses. The methods chosen (semivariogram analysis, kriging, and reverse power analysis) are discussed below. Brief descriptions of the data sets and data manipulations used to conduct the evaluations are also provided.

The semivariogram analysis examined the temporal and spatial scales of variability of key water column parameters measured in the nearfield and farfield monitoring program. The kriging analysis used the results from the semivariogram analysis to estimate detectable change as a function of spatial scale and alternate water column monitoring network designs. The reverse power analysis also estimated detectable change, although in this case the estimates were for the nearfield and farfield monitoring network as a whole, rather than as a function of spatial scale and reduced monitoring networks.

2.1 Semivariogram and Kriging Analyses

The geostatistical data analysis methods known as semivariogram analysis and kriging (Journel and Huijbregts, 1981; Isaaks and Srivastava, 1989; Cressie, 1993) can be used to evaluate temporal and spatial variability as well as changes in the statistical confidence of water quality monitoring estimates as a function of different monitoring designs (*i.e.*, different numbers and locations of sampling stations or different frequency of sampling surveys or both). Because each different monitoring design can be associated with different field sampling and chemical analysis costs, kriging, when employed to its fullest capability, can be used as a quantitative basis for evaluating trade offs between statistical confidence and monitoring costs.

The main steps in implementing this geostatistical approach for a single type of pollutant measurement or other water quality measurement can be summarized as follows:

1. Evaluate temporal and spatial correlations in the available monitoring data using a semivariogram analysis.

2. Use the resulting semivariogram correlations and the available monitoring data to interpolate (*i.e.*, estimate) water quality parameter concentrations at unsampled times and locations, and calculate the statistical uncertainty and detectable change associated with each estimated value.

3. Develop a set of hypothetical new monitoring network designs and, using the semivariogram correlations and kriging, evaluate the marginal gains or losses in statistical uncertainty associated with each new design.

Evaluating Temporal and Spatial Correlations. The cornerstone of any kriging analysis is the evaluation of temporal and spatial correlation using the semivariogram function. The semivariogram summarizes temporal and spatial variances (square of the difference between pairs of measurements) as a function of separation distance or direction. Therefore, low semivariogram values (*i.e.*, small differences) correspond to high correlation and high semivariogram values correspond to low correlation. Depending on the goals of the statistical analysis and on the temporal and spatial data available, semivariogram analysis can evaluate both temporal correlation and spatial correlation at various scales for multiple water column parameters.

Interpolating Water Quality Parameter Levels With Kriging. In water quality monitoring, actual measurements are made at only a limited number of time periods and monitoring stations. Water quality parameter levels at all other times and locations must be estimated. That is, a temporal-spatial interpolation of pollutant or other concentrations at unsampled times and locations within a body of water must be made based on concentrations measured at other fixed times and locations in the area. Kriging is a linear geostatistical estimation method which uses the temporal-spatial semivariogram function to determine the optimal weighing of the measured pollutant concentrations to be used for the required estimates. The kriging can be also be used to calculate the statistical uncertainty associated with the estimates. Thus, kriging is different from classical interpolation algorithms in that it produces statistically optimal estimates and associated uncertainty measures. These uncertainty measures can then be used to construct a confidence interval for each estimate. The confidence interval defines the range of values within which the unknown true concentration can be expected to lie. The estimation uncertainty and resulting width of the confidence interval depend not only on the strength of the temporal and spatial correlations in the data, but also on the number of monitoring stations and time periods for which data are available, and the proximity of those available measurements to the location and time for which the unknown concentration is being estimated. In addition, the kriging uncertainty measurements can be used in a reverse power analysis to determine the change that can be detected at each unsampled location in the system being monitored.

Evaluate Monitoring Design Alternatives. In the final analysis, the number and placement of the sampling stations is critical to detecting change and to the overall costs of the monitoring program (including sample collection and analysis). That is, the incremental costs for each additional station can be significant yet the potential benefits derived from added monitoring data (spatial or temporal) can also be significant. When coupled with knowledge of sample collection and analysis costs, kriging can provide quantitative information on the benefits (costs and confidence) that can be gained from collection of additional or fewer data points. Furthermore, kriging can provide the estimation uncertainty measures and associated detectable changes based only on the semivariogram values, and the locations and times of the samples to be collected in a proposed new monitoring scheme. Thus, evaluating the monitoring design is simplified because new samples do not need be collected, only their locations and times need to be specified. Therefore, a series of potential new monitoring schemes can be defined and evaluated to select the best design from both a statistical and a cost perspective.

The utility of the kriging analysis to the MWRA monitoring program was evaluated through exploratory analysis of two types of data collected from the water column during the baseline period. One set includes spatial and temporal results from nearfield surface fluorescence (also referred to as chlorophyll) over a two-year period (1992-1993); the other set includes spatial data from both nearfield and farfield monitoring at two different time periods. These data sets are described below. The results of the various semivariogram and kriging analyses performed are described in Section 3. The nearfield spatial and temporal data from 1992 and 1993 were used to develop preliminary information on detectable change in the nearfield as a function of sampling frequency and number of stations sampled; the spatial data sets were used to develop information on spatial scales of variability within the system and change that can be detected as a function of the spatial scales of variability.

2.1.1 Nearfield Temporal Data Set

The exploratory analysis of the sampling design was based on a single parameter — *in-situ* fluorescence as calibrated to measured chlorophyll concentrations. This measure of chlorophyll was selected because it is considered a key indicator/endpoint parameter (Hunt and Steinhauer, 1994a), because it shows large spatial and seasonal variability, and because it is readily measured through *in-situ* and traditional laboratory techniques. For this analysis, the data set was restricted to near-surface chlorophyll data, collected in the nearfield between February 1992 and December 1993. This data set includes two annual cycles. A file containing spatial (position) information, sample collection dates, and the chlorophyll data from the 0-5 m depth interval at each nearfield station was created. The data set consisted of results from 21 stations sampled during 32 surveys. Of a possible 667 records, 672 were available. The analysis was conducted on the entire data set; seasonal groupings from November to March (winter) and April to October (summer) were also evaluated.

The semivariogram analysis also enables directional trends in the data (gradients) to be evaluated by examining variability along specified directions. For this type of analysis, the semivariogram program searches for all data pairs that fit defined directional and spatial separation criteria (*e.g.*, directional separation from north and spatial separation distances of 100, 200, 300 m, etc.). The program then calculates the square of the difference between all pairs of measurements matching the criteria. If directional trends are present in the data, the semivariogram plots for the various directions will not closely overlay. In this analysis, separations in the north, northeast, east, and southeast directions were considered and, in all cases, a directional tolerance of $\pm 22.5°$ was allowed.

2.1.2 Spatial Data Sets

Assessment of the spatial scales of variance in the water column was conducted on two types of data — near surface (0-5 m) and photic zone (defined by the 1% light level). Data from two sampling periods (August 1993 and February 1994) were selected for the analysis and used to compare spatial variability within each survey. Comparison of the spatial results obtained for each survey provides an indication of seasonal differences. The parameters included in this analysis were salinity, temperature, fluorescence as chlorophyll *a*, beam attenuation, dissolved oxygen (DO), and dissolved inorganic nitrogen (DIN). Positional data for each data set were also included and used to calculate the separation distances. To evaluate surface variability, data from four tow-yo transects (each side of the outer nearfield area only), nearfield profiles, and farfield profiles were used to develop semivariograms. The locations of the stations are shown in Figure 1. The data files for this evaluation included the following:

Data Type	August 1993	February 1994
Tow-yo	3000 files (records)	200 files (records)
Nearfield	21 stations	21 stations
Farfield	25 stations	25 stations

The large difference in the number of tow-yo records between the two sampling periods is due to changes in operating procedures that enable the towed system to sample closer to the ocean surface for periods of time.

To conduct the photic zone evaluation, the depth of the 1% light level in the water column was determined using the *in-situ* light profile for each station in the nearfield and farfield. The average value over the photic depth was calculated for each *in-situ* parameter measured (salinity, temperature, fluorescence as chlorophyll *a*, beam attenuation, and dissolved oxygen) and used with the station position information to develop spatial semivariograms. For chlorophyll, the average value in the photic zone was multiplied by the depth of the photic zone to convert the value from a volume basis (μg/L) to an area basis (mg/m^2).

Data Type	August 1993	February 1994
Nearfield	21 stations	21 stations
Farfield	25 stations	25 stations

By developing semivariograms that present variability along various directions, the presence of spatial trends (gradients) in the data (*e.g.*, systematic directional changes in the chlorophyll *a* values with distance offshore) can be evaluated. Directional trends in the spatial data sets were evaluated by examining semivariograms developed along the two main tow-yo directions (*i.e.*, N75°E and N165°E) as well as two intermediate directions (N30°E and N120°E). Note that east is 90°, south is 180°, and west is 270°. Under this coordinate system, the northern and southern tow-yo transects extend along a N75°E direction from the western end of the transects while the eastern and western transects extend N165°E from their northern terminus. The specific models used for the kriging analyses discussed in Section 3 are provided in Appendix A.

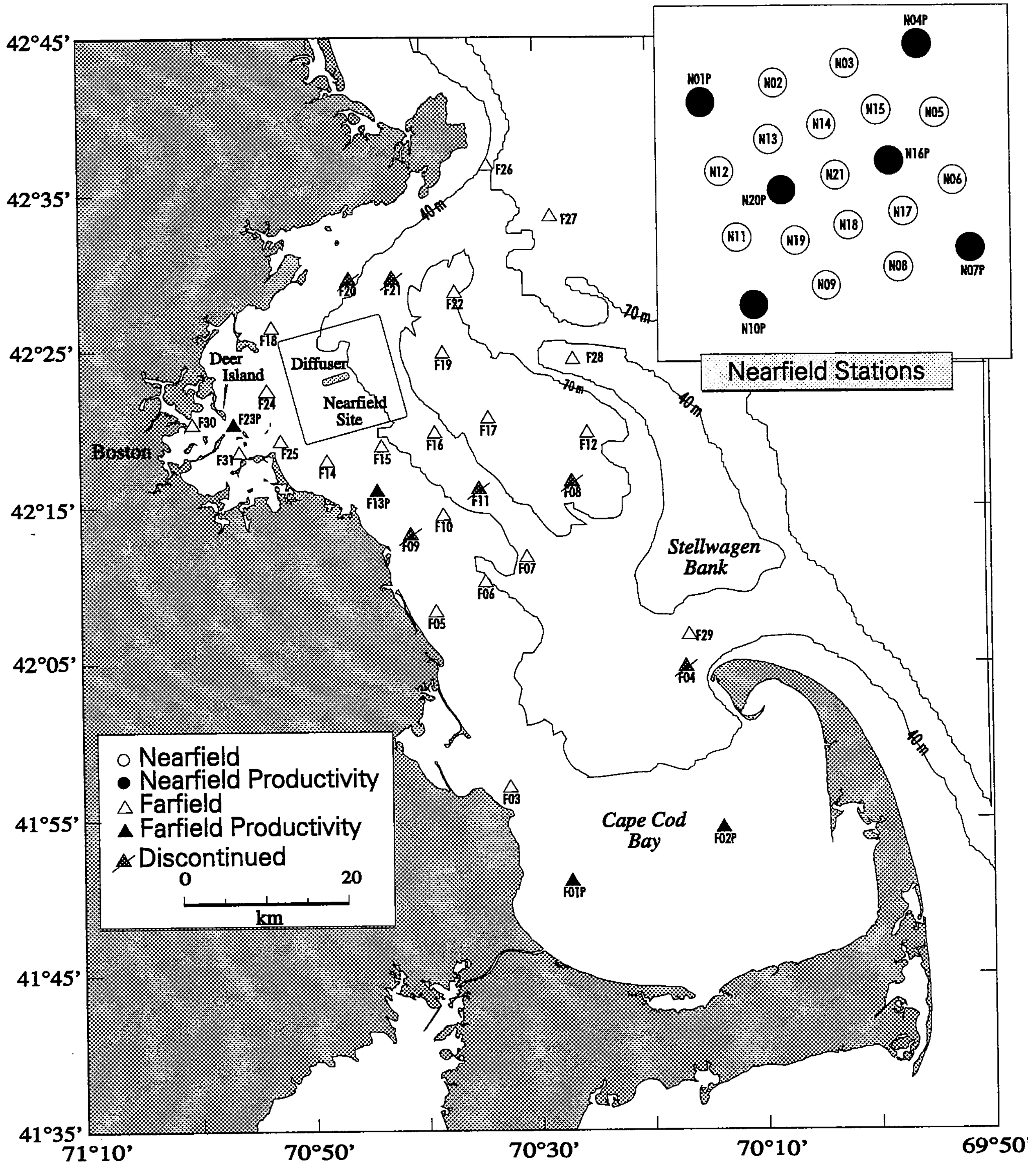


Figure 1. Location of nearfield and farfield water column stations sampled in 1992, 1993, and 1994. For 1994, several farfield stations were dropped and replaced by stations located farther towards the Gulf of Maine (F26, F27, F28, F29) or in Boston Harbor (F30 and F31). Nearfield stations remained the same over all years.

2.2 Power Analysis

Power analysis is used in this report to provide estimates of the change in key monitoring parameters that can be detected within the nearfield area. The results of this analysis are best viewed from the context of detecting change against a defined limit (standard), not as the ability to detect change in a system over time or space.

2.2.1 Approach

Application of statistical power analysis is gaining widespread use in environmental monitoring programs. Depending on the data that are available, the principles involved in power analysis can be used to evaluate a number of parameters, thereby increasing confidence in the statistical approach and design of a monitoring program. To conduct a power analysis the following variables are needed:

n	=	Number of samples or measurements used to formulate an estimate for decisionmaking
α	=	Probability of committing a Type I decision error (confidence level)
$1-\beta$	=	Power (β = probability of committing a Type II error)
σ	=	Standard error of the estimate (determined from real data)
es	=	Effect size (detectable change)

Generally, power analysis is used to develop estimates of the power of the measurements. However, knowledge of any four of these variables enables the fifth to be calculated. Peterman (1989) describes an application of power analysis, termed reverse power analysis, that can be used to develop estimates of the level of change (effect size) that can be detected for a given set of results. This approach was used previously in the MWRA monitoring program to examine the change that can be detected in concentrations of contaminants in the tissue of organisms (Hunt and Baptiste, 1993; Hillman *et al.*, 1994). This approach was used with the kriging analysis discussed above to determine the change in chlorophyll concentration that could be detected with various monitoring designs. This same concept was applied to nearfield water column results to develop estimates of the change in dissolved oxygen, chlorophyll *a*, DIN, and total suspended solids (TSS) that can be detected over the entire nearfield by the design employed for the nearfield monitoring during Phase I of the monitoring program.

2.2.2 Data Sets

To complete this analysis, each set of nearfield data (21 stations, 5 depths, 105 samples per survey) from February 1992 through May 1994 was treated as follows:

- For each parameter, the mean and standard deviation (nonbiased or "n-1" method) for each nearfield data set (all points) were calculated and graphed as a function of time.
- The percent detectable change for each survey using 80% power and 95% confidence level was calculated using reverse power analysis.
- The concentration of each variable that can be statistically detected for each survey was graphed against the mean concentration for each survey.

3.0 RESULTS AND DISCUSSION

3.1 Kriging Analysis

3.1.1 Nearfield Temporal Data Set

Individual surface chlorophyll *a* measurements for each station (indicated by different plotting symbols) show a relatively strong annual cycle (Figure 2a). Higher chlorophyll concentrations are observed in the summer and fall months, and lower concentrations are detected in the winter months. Over this two-year period, the minimum and maximum chlorophyll *a* ranged from 0 μg/L to 20.04 μg/L, respectively. The mean chlorophyll *a* concentration (all values) was 2.20 μg/L with a standard deviation of 2.59 μg/L. Because the probability distribution of the chlorophyll data exhibited positive skewness, the data were transformed by taking the natural logarithm before conducting the semivariogram analysis. These normalized results are shown in Figure 2b.

The results of the temporal semivariogram analysis for chlorophyll *a* are shown in Figure 3. The x-axis of this figure shows the separation in time (*i.e.*, days) between various pairs of monitoring surveys; the y-axis lists the semivariogram value, the average squared difference between the fluorescence values taken at the same sampling station during each of two surveys. Note that the semivariogram values increase with greater separation times up to 180 days and then decrease with still greater separations from 180 to 360 days. These temporal semivariogram changes relate directly to the cyclic annual trend observed for the chlorophyll values in Figure 2a.

Figures 4, 5, and 6 show semivariogram results that consider separation distances in space (*i.e.*, meters north and east) rather than time. Figure 4 shows spatial semivariogram results for data collected during the summer months (June-September). Figure 5 shows similar results for the winter months (November-March). Figure 6 shows spatial semivariogram results that include all months. These figures also consider separations in the northeast (NE) and southeast (SE) directions as well as semivariogram results averaged across all separation directions (omnidirectional). The semivariograms generally increase as distance between stations increases.

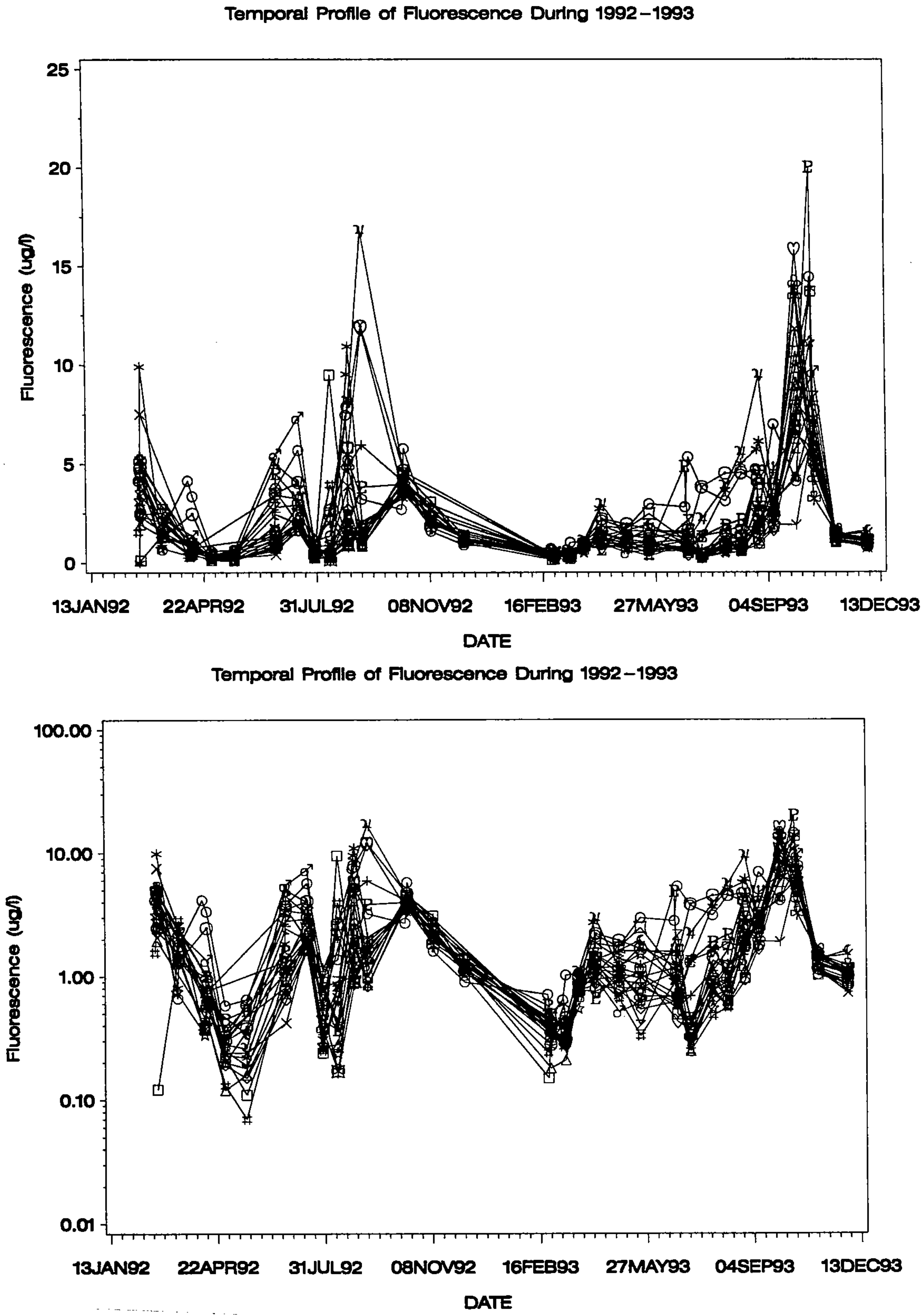


Figure 2. Nearfield surface chlorophyll (calibrated fluorescence) measurements from individual 1992 and 1993 stations plotted as a function of time. (a) linear concentration scale; (b) log-transformed concentration scale

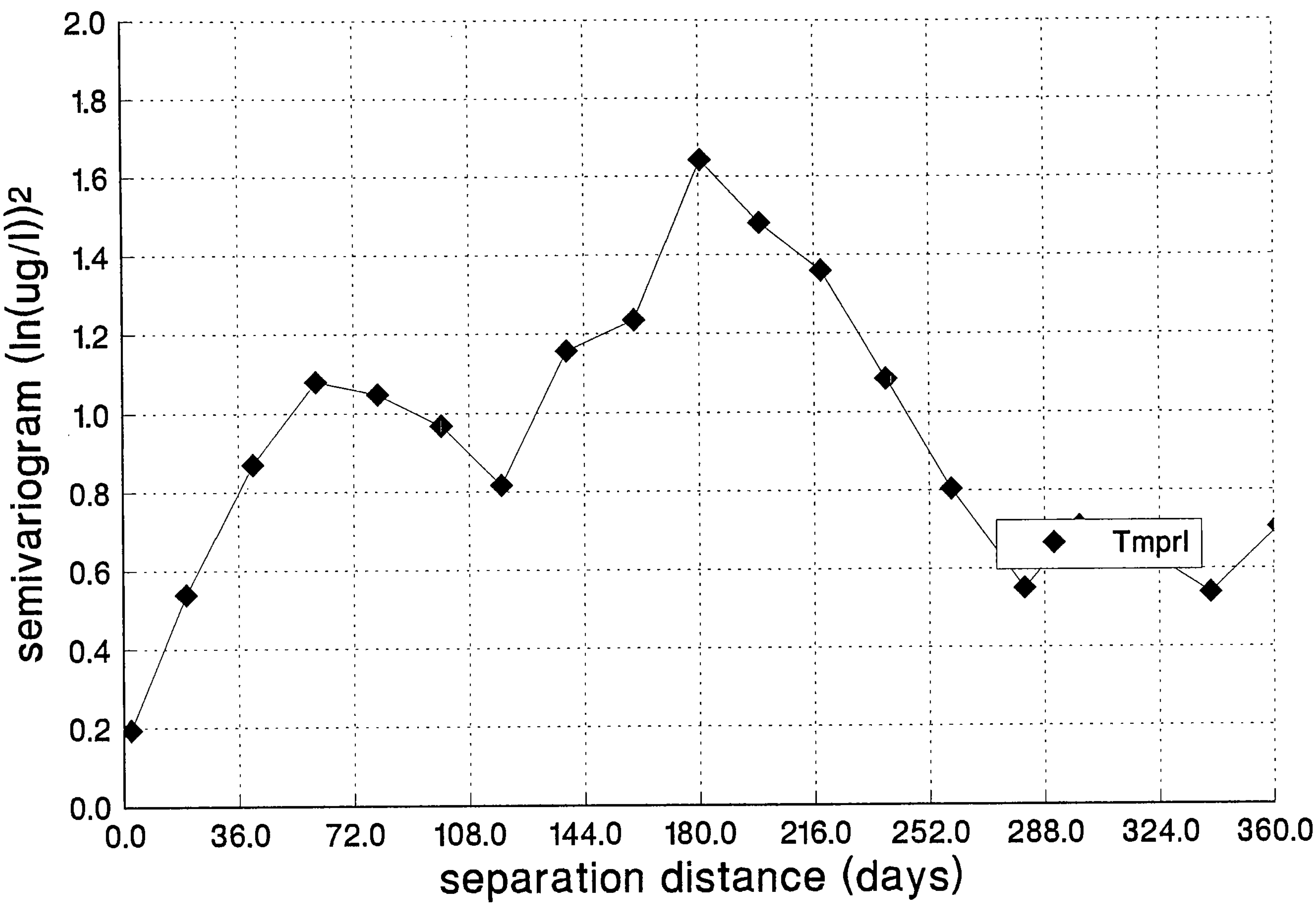


Figure 3. Experimental temporal semivariogram of log-transformed chlorophyll.

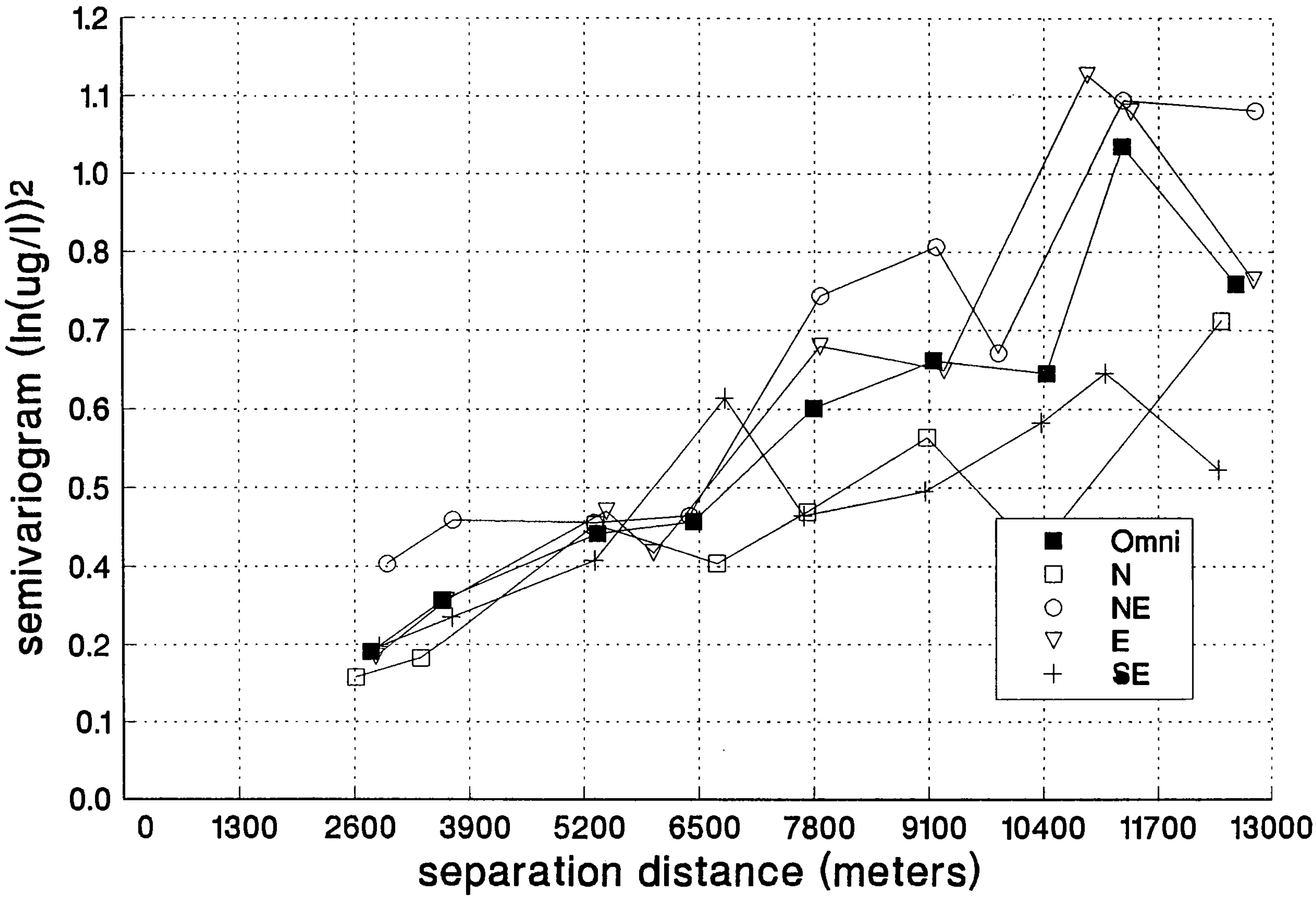


Figure 4. Experimental spatial semivariograms for log-transformed chlorophyll measured during the summer months (June - September).

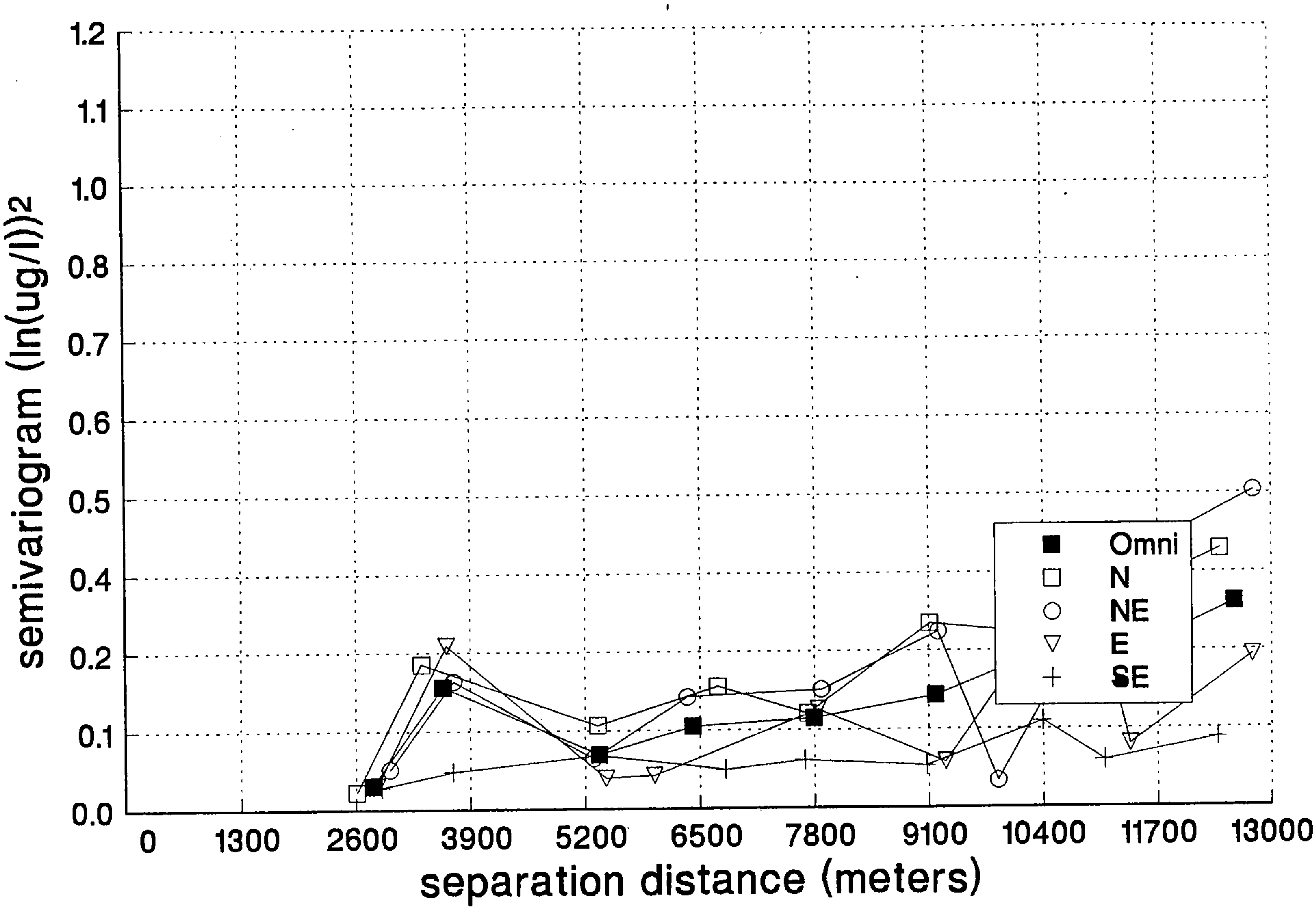


Figure 5. Experimental spatial semivariograms for log-transformed chlorophyll measured during the winter months (November - March).

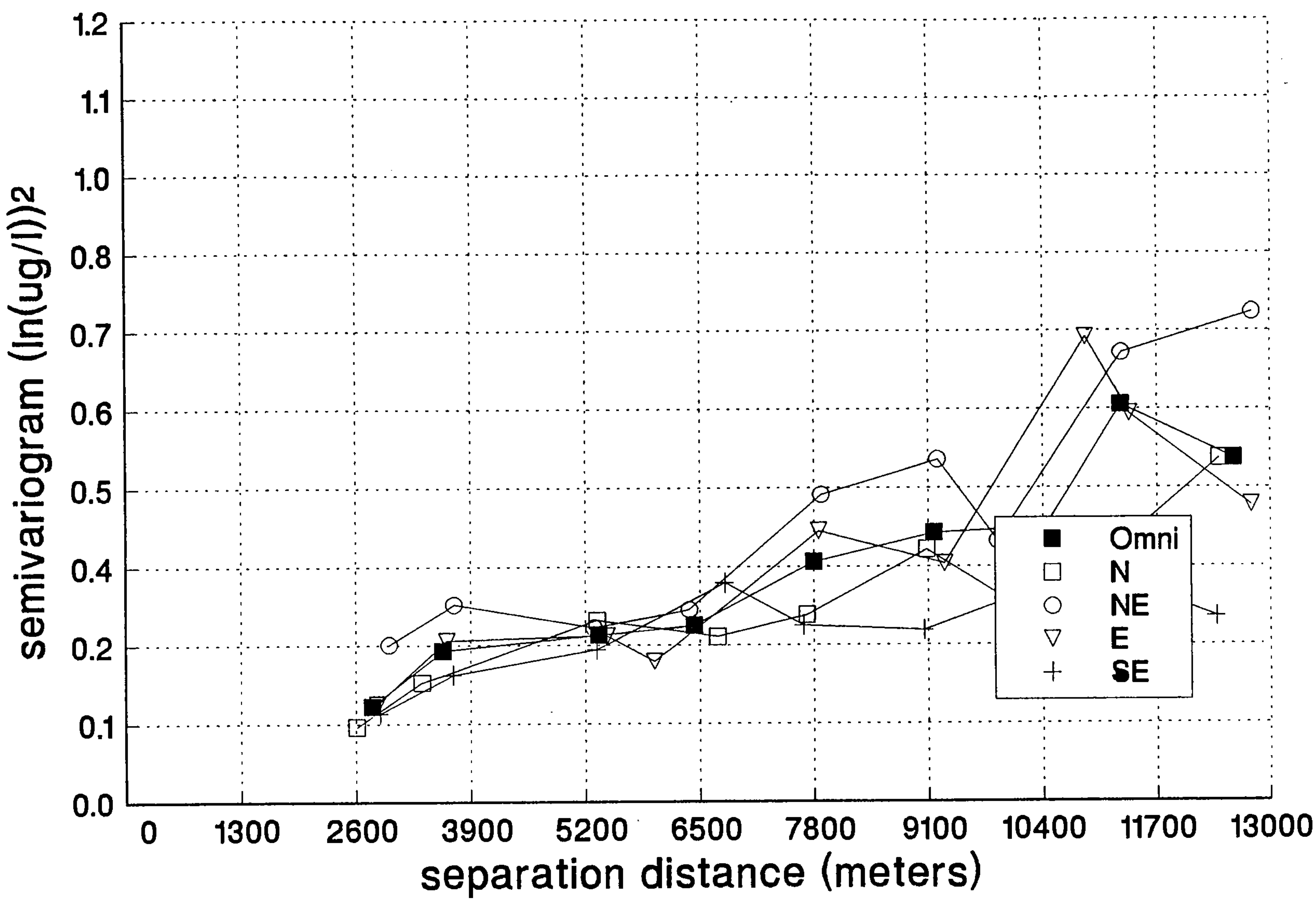


Figure 6. Experimental spatial semivariograms for log-transformed chlorophyll measured during all months.

A comparison of the semivariograms for summer (Figure 4) and winter (Figure 5) reveals distinct differences. First, the semivariogram values for the winter (0.01 to $\approx$0.5) are generally lower than values for the summer (0.2 to 1.1). Second, a clear trend of increasing semivariogram value occurs in the summer data but not in the winter data, indicating that the summer period is characterized by higher variability than the winter months. This observation can be tied to higher chlorophyll *a* values measured during the summer months relative to the winter months. Further, the results suggest that the surface chlorophyll variability between stations located further apart increases to a greater extent in the summer than in the winter, a result consistent with gradients observed in the actual data (Kelly, 1993). Incorporation of all data into one semivariogram (Figure 6) results in lower semivariogram values (0.1 to 0.7) relative to the summer semivariogram values. Increasing semivariogram values, indicative of increasing variability (gradients) with increasing separation distance, are evident for the entire data set.

Close examination of the summer spatial semivariogram (Figure 4) also reveals that semivariogram values in the eastward and northeastward directions at separation distances greater than about 6500 m within the nearfield are significantly higher than those towards the north and southeast. These data demonstrate that directional trends in chlorophyll are more apparent in the eastward and northeastward directions than in the other directions tested. This directional trend is not as evident in the winter semivariogram, again supporting more uniform distributions of surface chlorophyll within the nearfield in the winter period. These observations are also consistent with the timing of the development of the offshore gradient reported in Kelly (1993).

Finally, note that the overall magnitude of the spatial semivariogram values shown in Figures 4, 5, and 6 is generally less than the magnitude of values for the temporal semivariogram (Figure 3). Comparing the omnidirectional semivariogram values shown in Figure 6 to the temporal values shown in Figure 3 suggests that the spatial values are two to three times less than the temporal values. This result further indicates that the overall variability in the chlorophyll *a* data can be broken down into temporal differences between monitoring surveys which are over and above spatial differences observed among the sampling stations.

3.1.2 Monitoring Design

Because kriging produces statistically optimal estimates and associated uncertainty measures, the uncertainty measures can be used to construct confidence intervals, confidence bounds, or detectable changes associated with the estimates chlorophyll *a*. An example of the kind of uncertainty information that kriging can provide is represented in the form of detectable changes in Figure 7. In this case, a hypothetical situation is assumed where a single chlorophyll *a* measurement has been taken at one centrally located sampling station. Figure 7 shows the statistical uncertainty associated with using that single measurement as an estimator of the unknown chlorophyll *a* level at neighboring locations. These uncertainty values can be interpreted as the difference, with respect to a fixed standard (*e.g.*, regulatory standard, exposure limit, etc.), which could be detected (with 95% confidence and 80% power) at various locations around the central nearfield monitoring station. Note that close to the monitoring station the estimates are most certain and thus the smallest differences can be detected. For example, at a distance of about 1 km from the monitoring station, differences of about 8.0 μg/L could be detected, while differences of about 8.6 μg/L could be detected at a distance of 5 km.

Three other examples of the kind of information that can be developed by the kriging approach are now considered. These examples illustrate the benefits of sampling additional monitoring stations. Two examples consider the estimation of chlorophyll *a* concentrations at a given position **in a given month** using measurements taken in that month plus the measurements from either the previous 17 or 11 months. The third example considers estimation of chlorophyll concentrations at this location using measurements taken for only **the current month**. For these examples, the spatial point at which the chlorophyll concentrations is assumed to be estimated was approximately halfway between stations N14 and N21. When the number of stations was reduced, stations were selected to cover the nearfield region as well (uniformly) as possible. The last station retained was N16 which is close to the location being estimated. This station was selected because it is in an area of higher variability and provides a better illustrative example of the concepts discussed here than a station located exactly in the middle of the nearfield where variability tends to be slightly lower. Each of these cases is presented in Figure 8. The y-axis represents the gain in precision (*i.e.*, the decrease in the difference that could be detected) as additional monitoring stations are added to the network.

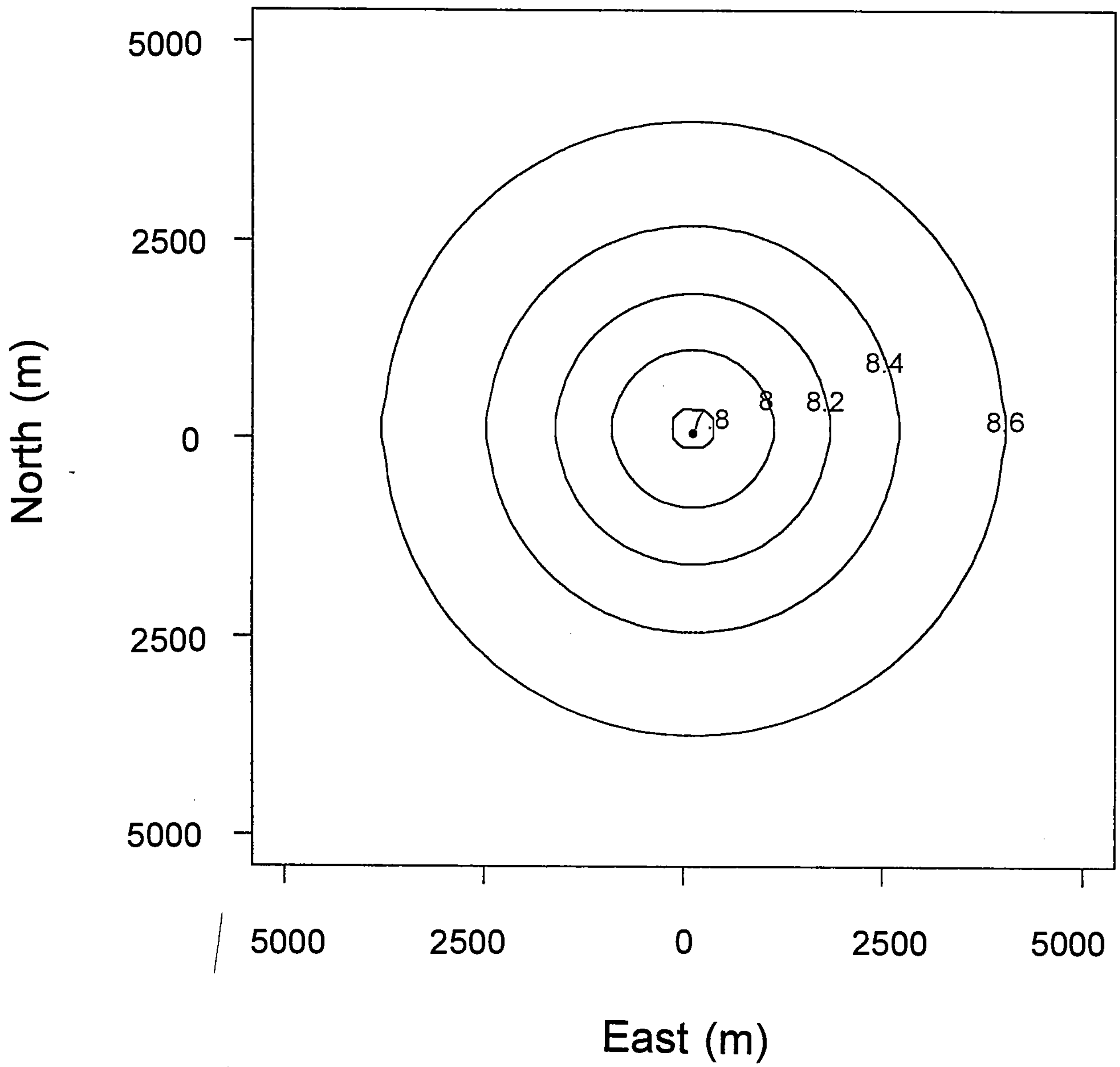


Figure 7. Statistical uncertainty (multiplicative factor) associated with using a single chlorophyll measurement to estimate chlorophyll concentrations at neighboring locations.

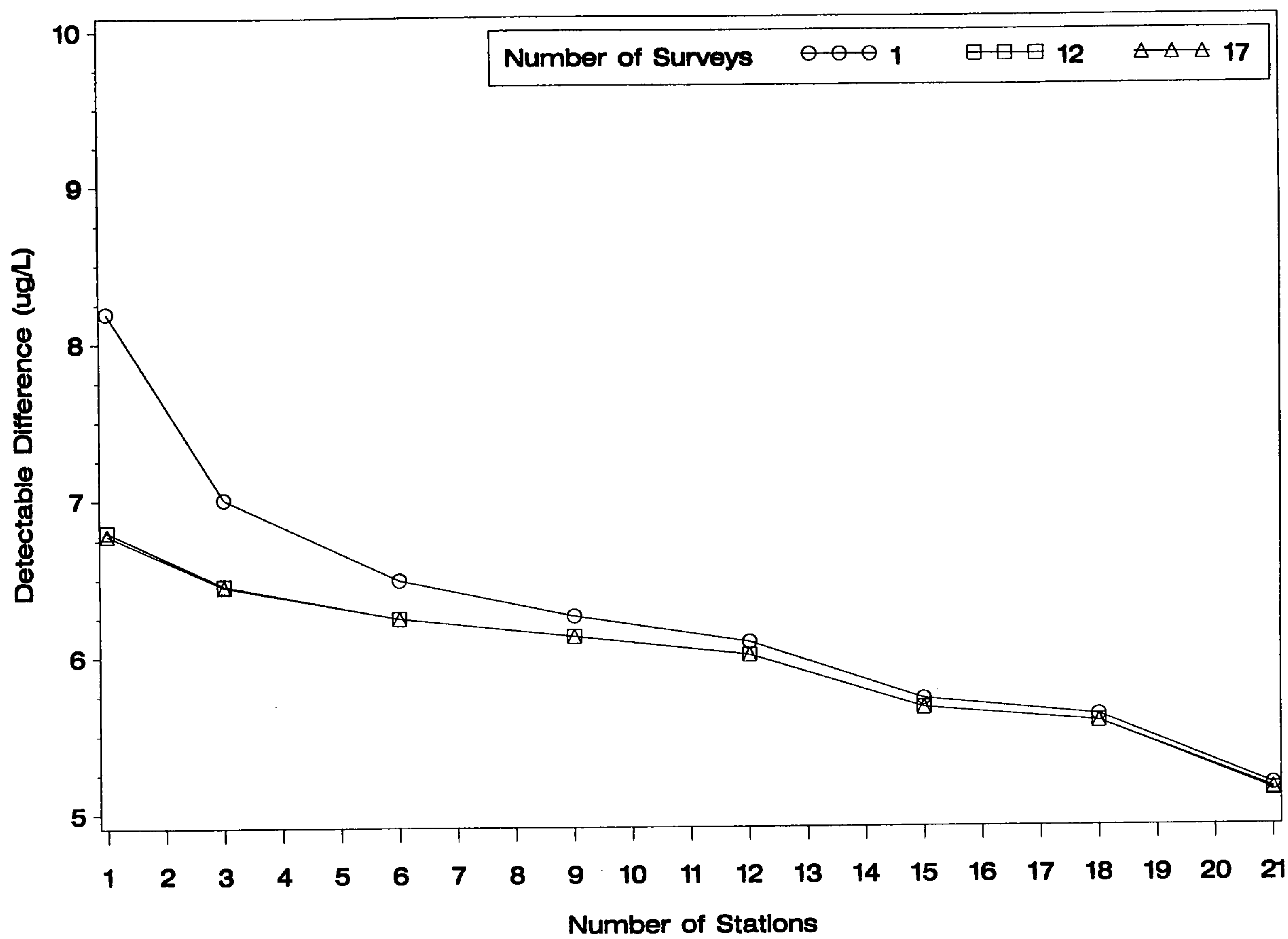


Figure 8. Gain in detectable change as a function of adding sampling stations. Curves assume either a single sampling (number of surveys = 1), sampling once per month for 12 months (number of surveys = 12), or 17 samplings per year (number of surveys = 17).

As an example, the difference that can be detected with just one survey improves from 8.2 μg/L for a single monitoring station to 5.2 μg/L for 21 monitoring stations, a 36% improvement in detectability. In contrast, the difference that can be detected with 17 surveys (*i.e.,* 1 survey every 21 days over the year) improves from 6.8 μg/L for a single monitoring station to 5.2 μg/L for 21 monitoring stations, a 24% improvement in detectability. It is interesting to note that in this example, monitoring with either 17 surveys or 12 surveys per year leads to the same level of statistical uncertainty. That is, the addition of five surveys does not help to reduce uncertainty in this case.

This result related to the fact that the uncertainty estimate is a function of both time and space considerations. However, this is not necessarily a general result for all combinations of surveys, stations, and locations to be estimated. For example, if the location to be estimated was near a given station, then repeated surveys at that nearby station may lead to more precise estimates than obtained from a single survey at other more distant stations. This points to the need to conduct a more thorough evaluation that is focused by specific questions and with a goal of deriving the optimum sampling design based on a defined confidence level for the various parameters measured under the water column component of the monitoring program.

3.1.3 Discussion

The semivariogram and kriging analysis show that temporal and spatial variability in the surface waters of the nearfield is low and seasonal in nature. Directional trends are also evident in the data as has been reported from the field program (Kelly, 1993). Suggestions of directional trends (east-west) in the variability are also evident. The fact that the semivariogram analysis shows temporal variance to be greater than the spatial variability suggests that reductions in the number of nearfield stations may be achieved without substantial loss in the precision obtained. The results shown in Figure 8 suggest that the estimates of precision for a single survey are similar to those obtained with 12 surveys over an annual cycle when approximately 12 stations are included in the sample design; differences of about 6.2 μg/L can be detected for both cases. Improvements in the estimation precision that are obtained by increasing the number of surveys from 12 to 17 are negligible.

Relative to the spatial variability, the higher observed temporal variability further indicates that greater emphasis should be placed on temporal sampling. Extension of this result to the entire

Massachusetts Bay system can not be made based on this preliminary evaluation and will require additional study.

Even though this exploratory analysis suggests design modifications could be implemented without loss of statistical confidence, further analysis to define the optimum number of stations or surveys must be conducted. Effective use of this analysis depends, however, on definition of the precision (and associated detectable change) that is most appropriate for the monitoring program. The preliminary results presented here suggest that more frequent water column monitoring at fewer stations may be a better design. Other designs, such as a program that focuses temporal monitoring into fewer surveys during less variable periods of the year and more frequent measurement during the more variable period, should also be considered. Thus, without conducting additional statistical analysis of the temporal data, including data collected during 1994, the optimal design can not be provided. Finally, these preliminary results can be used to develop and focus questions from which the optimal spatial and temporal nearfield monitoring design can be completed.

3.2 Spatial Data Sets

3.2.1 Semivariograms for Surface Waters

For this analysis, tow-yo data from the four outer transects (northern, southern, eastern, western) of the nearfield area (see Albro *et al.,* 1993 for pictorial representations of the sampling design) were combined with data from the nearfield and farfield stations and treated as a single data set. The intense sampling rate (every 2 seconds) of the tow-yo surveys required that a minimum separation distance be defined so that small-scale variability caused by positional overlaps (due to noise in the positioning data) could be removed. For the August data, separation distance was lagged (stepped) at 20-m intervals to a distance of 4000 m, providing a large number of data points over this spatial distance. For large scales, the data were lagged by 2000 m, which is approximately the same distance that separates the nearfield stations. Regardless of the lag used, all semivariograms were developed from the combined data sets defined in above in Section 2.1.2. The resulting semivariograms are shown on two different spatial scales for ease of presentation and discussion. The largest separation distance between the stations approached 50,000 m. However, because the semivariogram results are highly variable when calculated for lags greater than approximately one-half the largest separation distance, semivariograms beyond 25,000 m are not presented here. Also, in several cases the semivariograms display noisy behavior at separation distances > 10,000 m. No attempt was made to interpret this behavior because it is most likely the result of the dense and non-uniform spatial arrangement of the tow-yo data in contrast to the more uniform spatial coverage of the nearfield and farfield data.

Short-Scale Variability. Figures 9 and 10 show representative short-scale semivariogram salinity results for August 1993 and February 1994. Variability along the 75° and 165° directions is shown. The variability in salinity in August is low (0.008 to 0.018) and relatively constant over separation distances up to 4000 m. Only at scales from 0 to 400 m in the 165° direction is there an increase in the variability from 0.01 to about 0.014. The variability remains relatively constant from 400 m to a distance of 3000 m before decreasing. Along the 75° direction, the variability is very constant and consistently lower than in the 165° direction. The semivariogram values for salinity in February (0.01 to 0.06) are greater than the values for August.

Both semivariograms indicate relatively uniform variability across the nearfield region along the directions examined. The consistently higher semivariogram values along the 165° versus 75° direction indicate that there may be a weak directional trend in the salinity and that the difference is oriented more in the south-southeast (165°) direction than in the east-northeast (75°) direction. The results from February 1994 do not show this directional trend within the nearfield.

Similar observations can be made for the chlorophyll results at this scale of presentation, although the highest variability was observed in the August data (compare Figure 11 with Figure 12) and the separation between the two directions is the largest observed for all of the parameters evaluated (Appendix B). The chlorophyll results differ from the salinity results in that clearly increasing semivariogram values with increasing separation distance are observed in both February and August. As observed with salinity, the August chlorophyll results show a weak directional trend, while the February results only suggest directional trends at separation distances greater than 1600 m. The generally increasing variability in the surface chlorophyll as separation distance increases is consistent with observed trends in the data reported in Kelly *et al.* (1993c) and Kelly *et al.* (1994f), and reproduced in Figures 13 and 14.

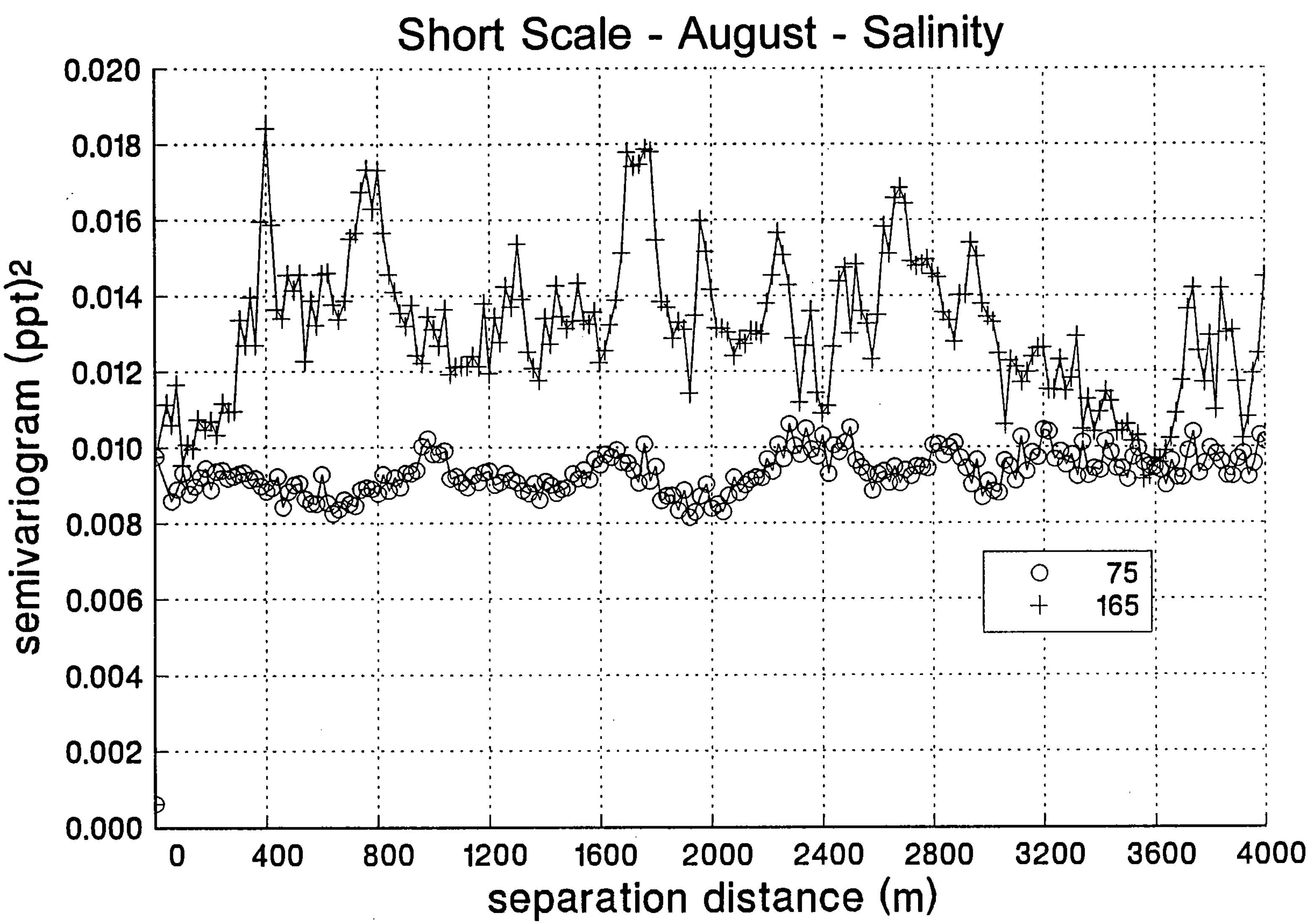


Figure 9. Short-scale spatial semivariogram showing variability in salinity along the direction of the outer boundaries (75° and 165°) of the nearfield box in August 1993.

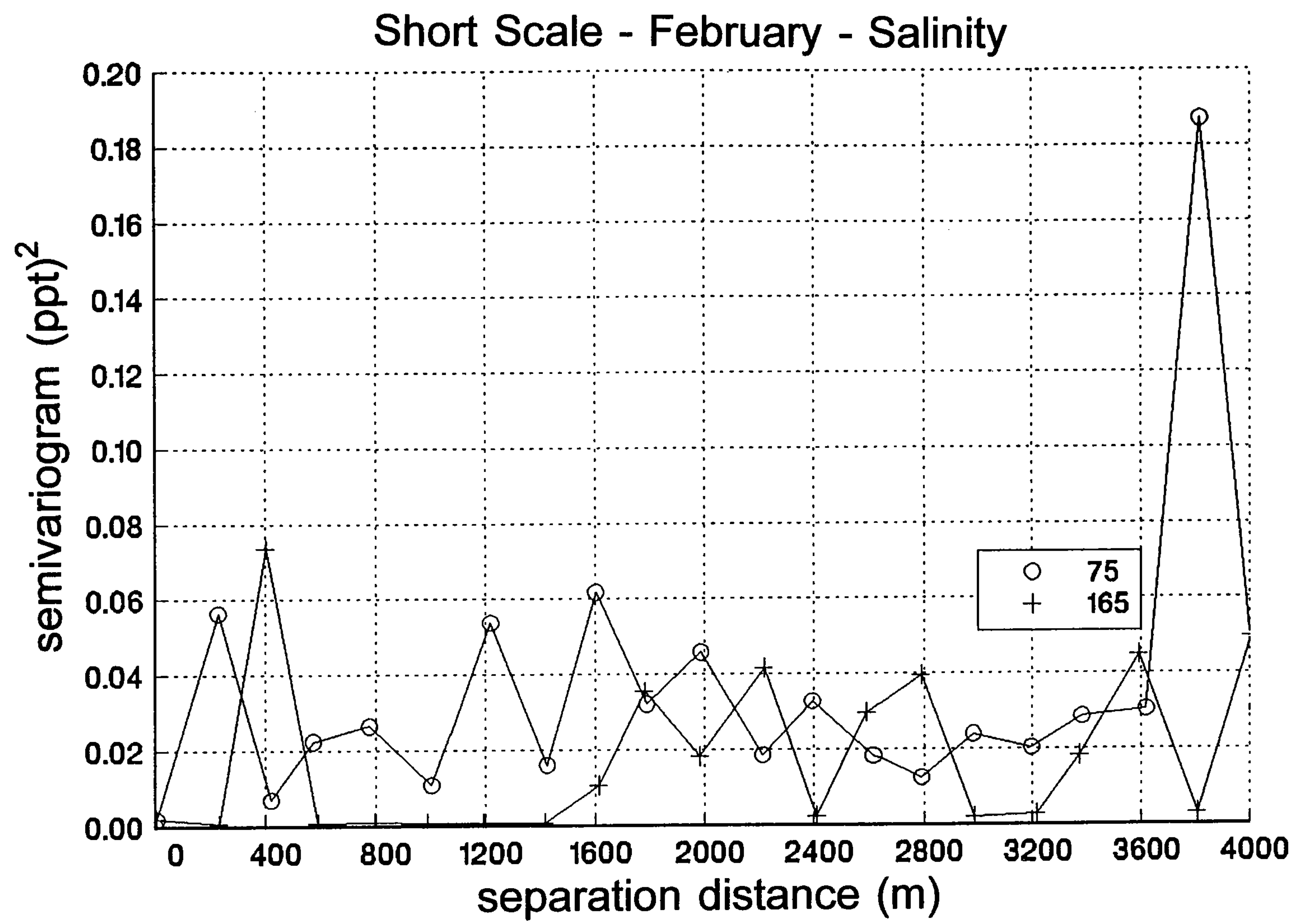


Figure 10. Short-scale spatial semivariogram showing variability in salinity along the direction of the outer boundaries (75° and 165°) of the nearfield box in February 1994.

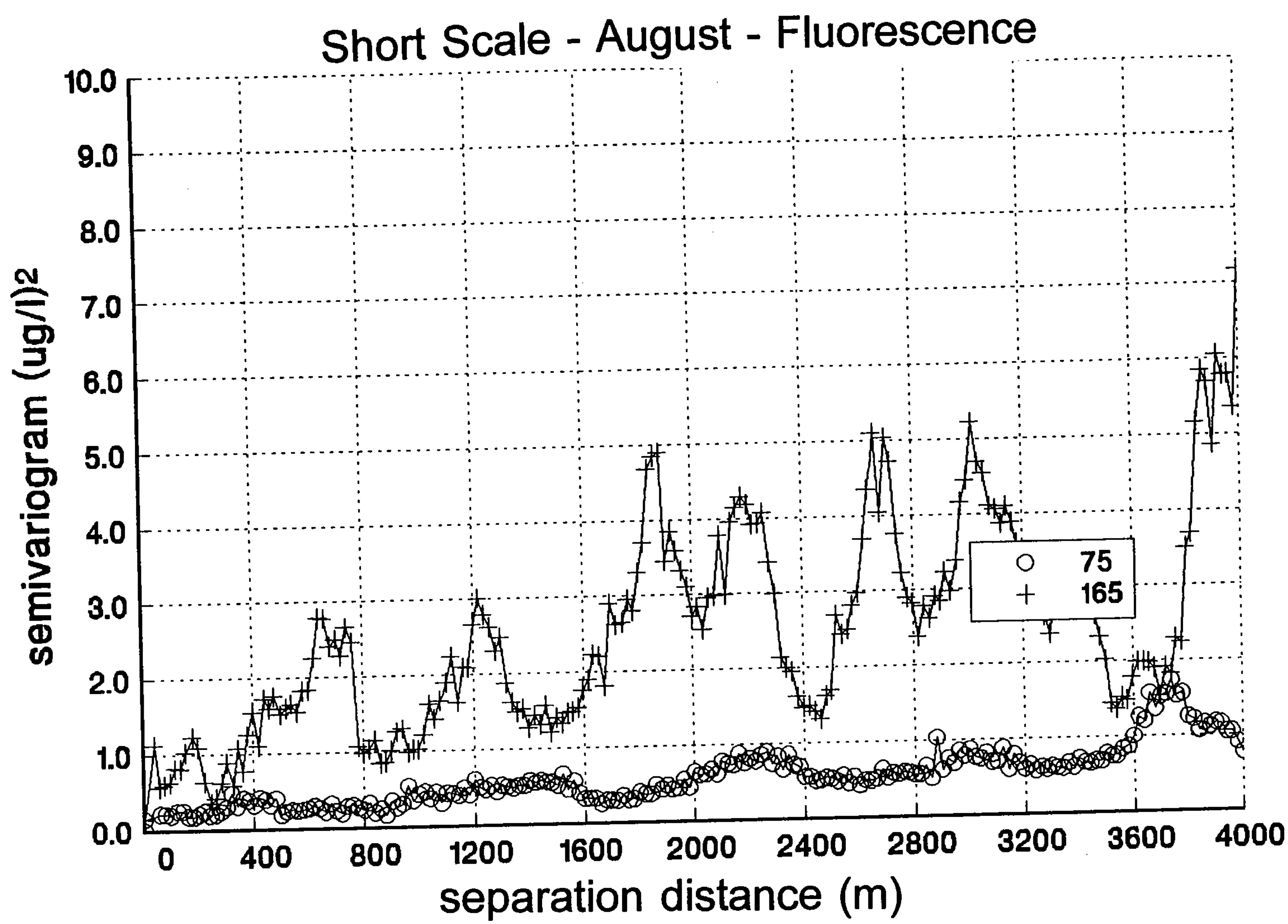


Figure 11. Short-scale spatial semivariogram showing variability in chlorophyll along the direction of the outer boundaries (75° and 165°) of the nearfield box in August 1993.

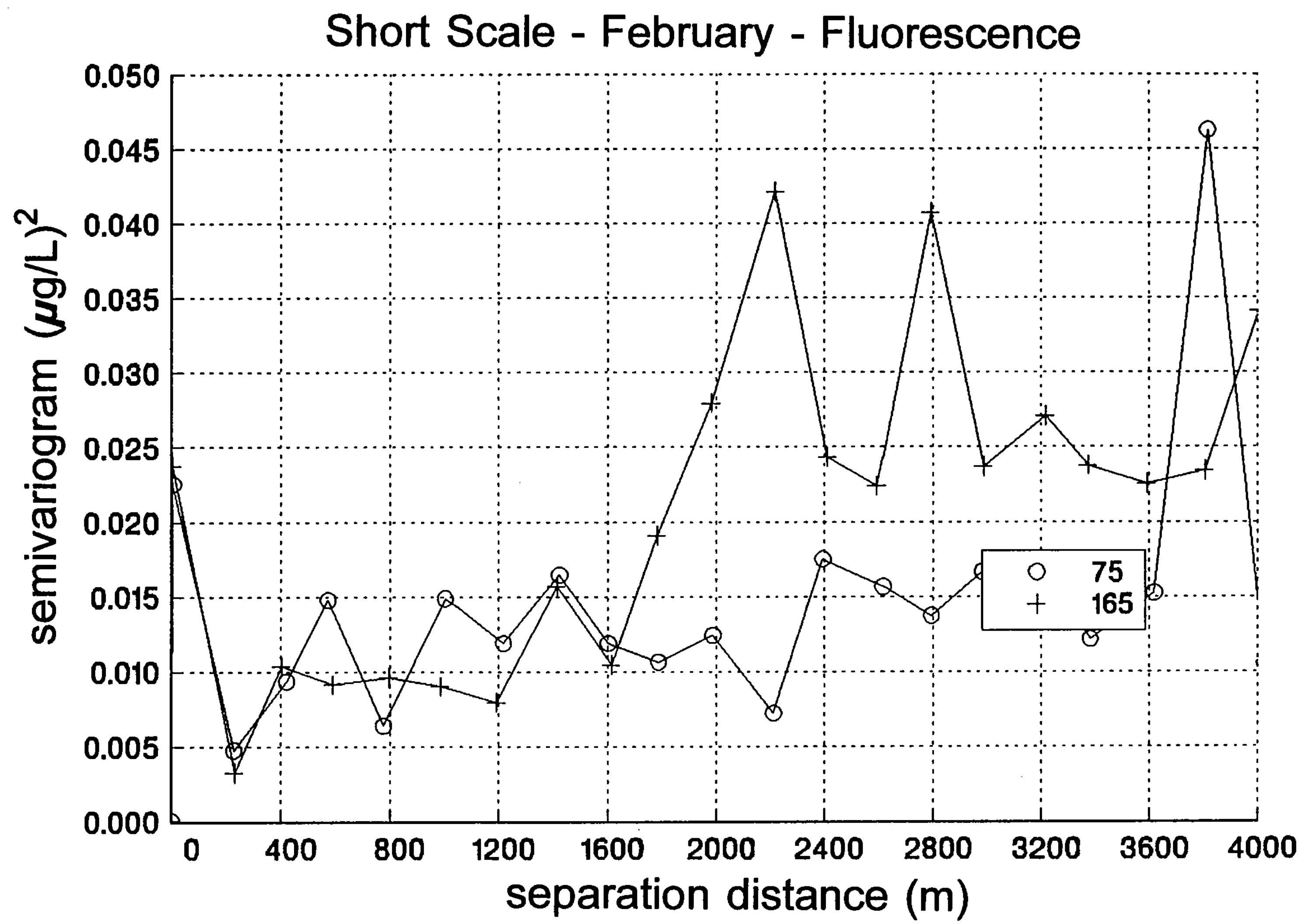


Figure 12. Short-scale spatial semivariogram showing variability in chlorophyll along the direction of the outer boundaries (75° and 165°) of the nearfield box in February 1994.

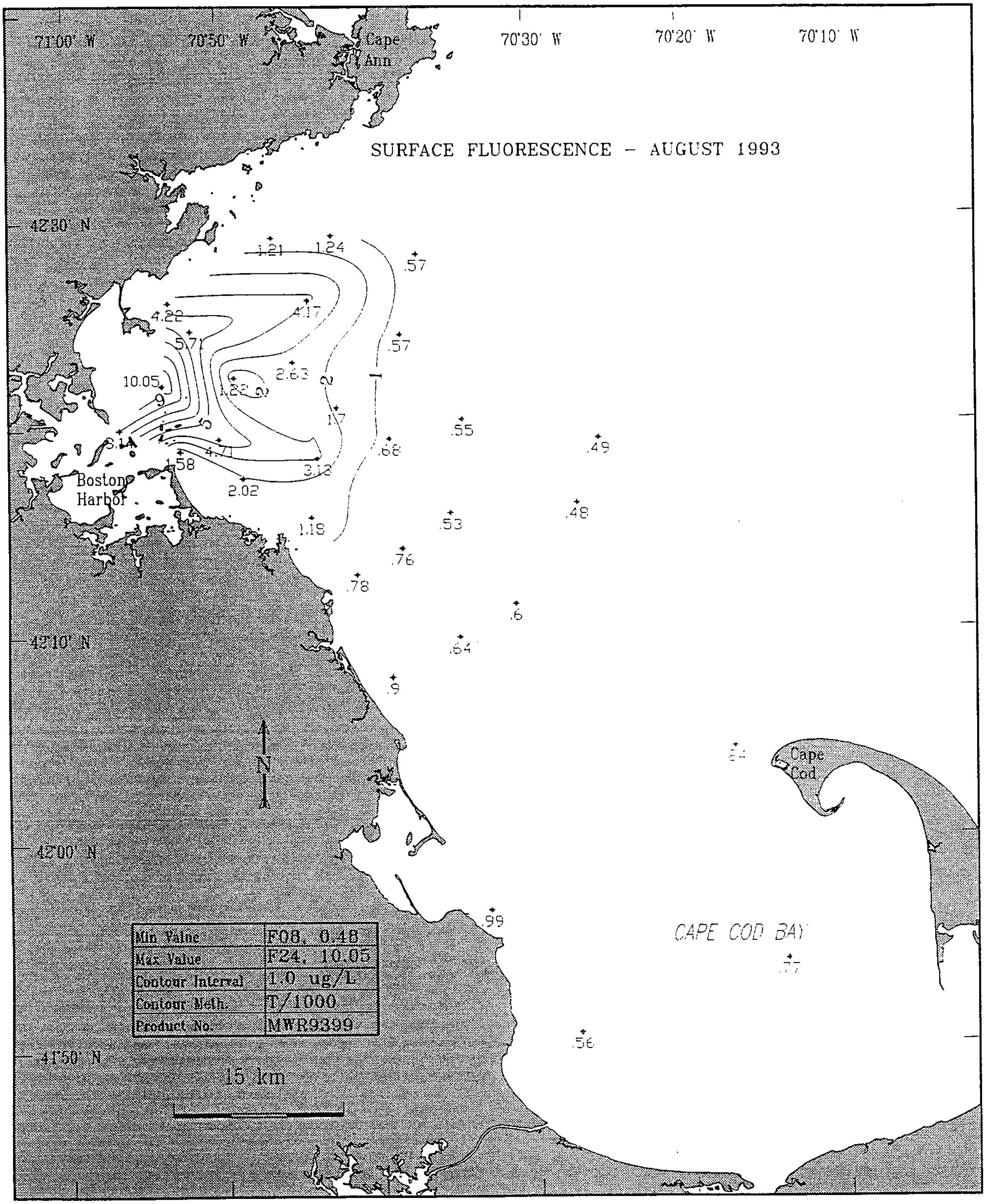


Figure 13. Surface *in-situ* chlorophyll *a* in the Massachusetts and Cape Cod Bay area in late August 1993 (from Kelly *et al.*, 1994d).

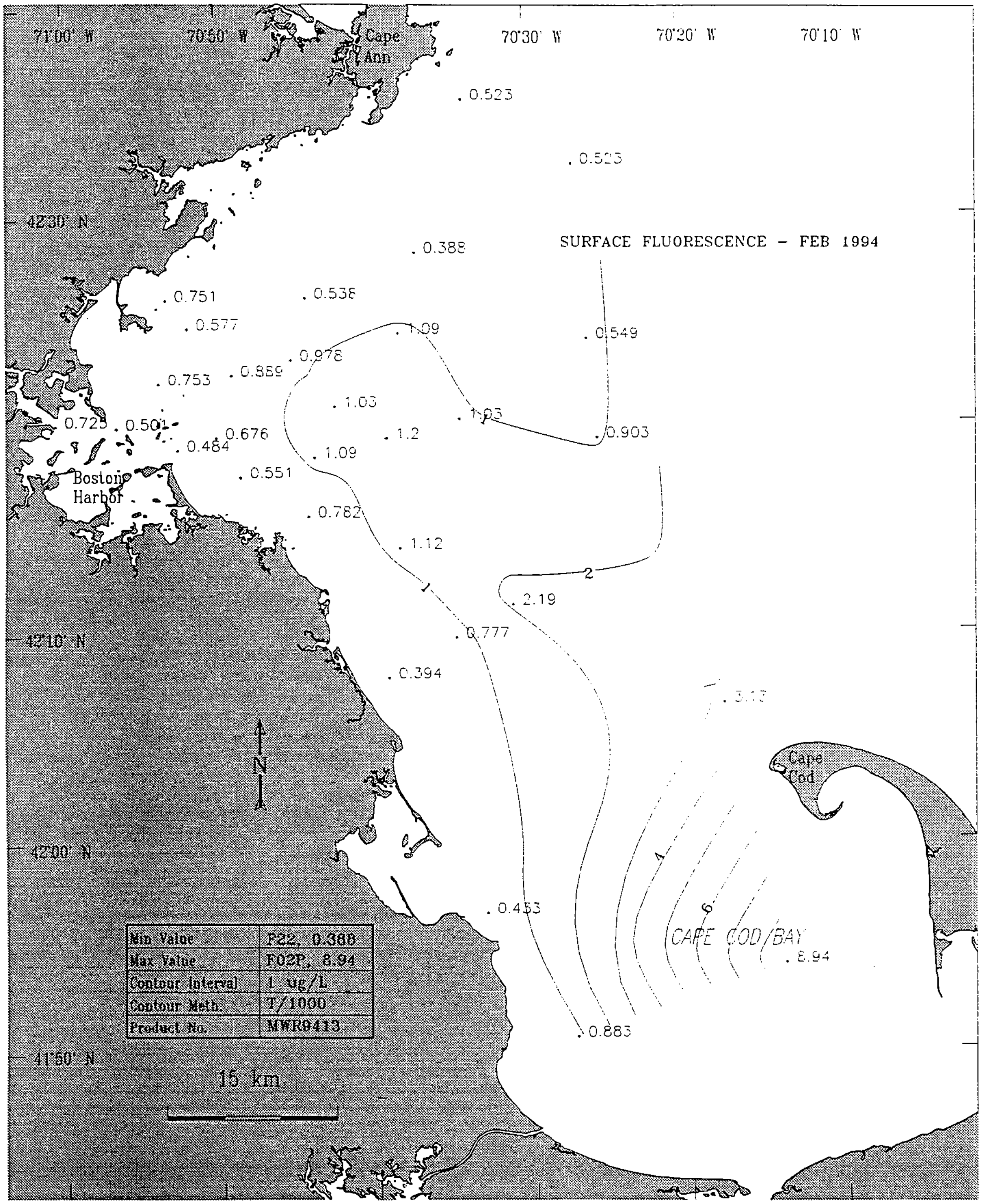


Figure 14. Surface *in situ* chlorophyll *a* in the Massachusetts and Cape Cod Bay area in February 1994 (from Kelly *et al.*, 1994f).

The characteristics of the semivariograms for the other parameters are summarized in Table 1. Results similar to those discussed above were apparent in the temperature and beam attenuation semivariograms which can be found in Appendix B. The results are similar to salinity and chlorophyll a, except that the directional trends appear to be weaker and that there appears to be substantial variability in the semivariogram results in the 165° direction.

The short-scale semivariograms point to relatively uniform variability across the nearfield as separation distance increases from 0 to 4000 m. The apparent trends in the direction of variability are also characteristic of the system and should be analyzed in greater depth than provided in this initial exploratory analysis. In addition, the fine-scale spatial resolution that is evident in the tow-yo data provides an indication of the small-scale variability over the nearfield. Further analysis and interpretation of this type of variability are not pursued in this report. However, such representations of the small-scale variability may provide an ability to statistically describe "patchiness" in the nearfield, and could prove useful for evaluating changes in the small-scale patterns and variability in the post-discharge period.

Long-Scale Variability. The long-scale semivariogram results expand the spatial scales of the observations described above. Comparison of the long-scale semivariograms demonstrates that the five parameters examined can be classified into three categories based on (1) the level of variability (semivariogram values) between August and February, (2) trends in increased variability as a function of separation distance, and (3) presence or absence of directional trends in the data. Representative semivariograms are discussed here. The full set of semivariograms can be found in Appendix B.

The classification of these parameters based on the semivariogram include: (1) salinity and dissolved oxygen, (2) temperature, chlorophyll a, and beam attenuation, and (3) DIN. The first category is characterized by similar semivariograms between August and February, and lack of directional trends in both the August and February surveys. Note that while salinity (Figures 15 and 16) shows a relatively flat semivariogram between 500 m and 12,000 m for both surveys, the dissolved oxygen semivariograms for each survey (Figures 17 and 18) consistently increase to a separation distance of about 22,000 m.

Table 1. Summary of short-scale semivariogram characteristics for the surface (0-5 m) water column throughout Massachusetts Bay.

Parameter	Semivariogram Value Range[1]	Variability August vs February	Variability increases as separation distance increases	Directional difference east-west and north-south transects
Salinity	A: 0.008 to 0.018 F: 0 to 0.1	February is **higher** than August	A: From 0 to 400 m F: No	A: Yes F: No
Dissolved Oxygen	A: 0.25 to 2.2 F: Not available	Not available	A: From 0 to 400 m along 165°, then cycles F: Not available	A: Yes, but is variable F: Not available
Temperature	A: 0.01 to 0.7 F: 0 to 0.16	February is **lower** than August	A: Along 165° from 0 to 400 m, then cycles F: No	A: Yes, but is variable F: No
Chlorophyll *a*	A: 0.02 to 7.0 F: 0.005 to 0.045	February is **lower** than August	A: Yes, but cycles F: No	A: Yes F: Along 165° greater than 1600 m
Beam Attenuation	A: 0.01 to 2.7 F: 0.001 to 0.32	February is **lower** than August	A: Slightly with cycles F: No	A: Small scale variability along 165° F: No

[1] A = August
F = February

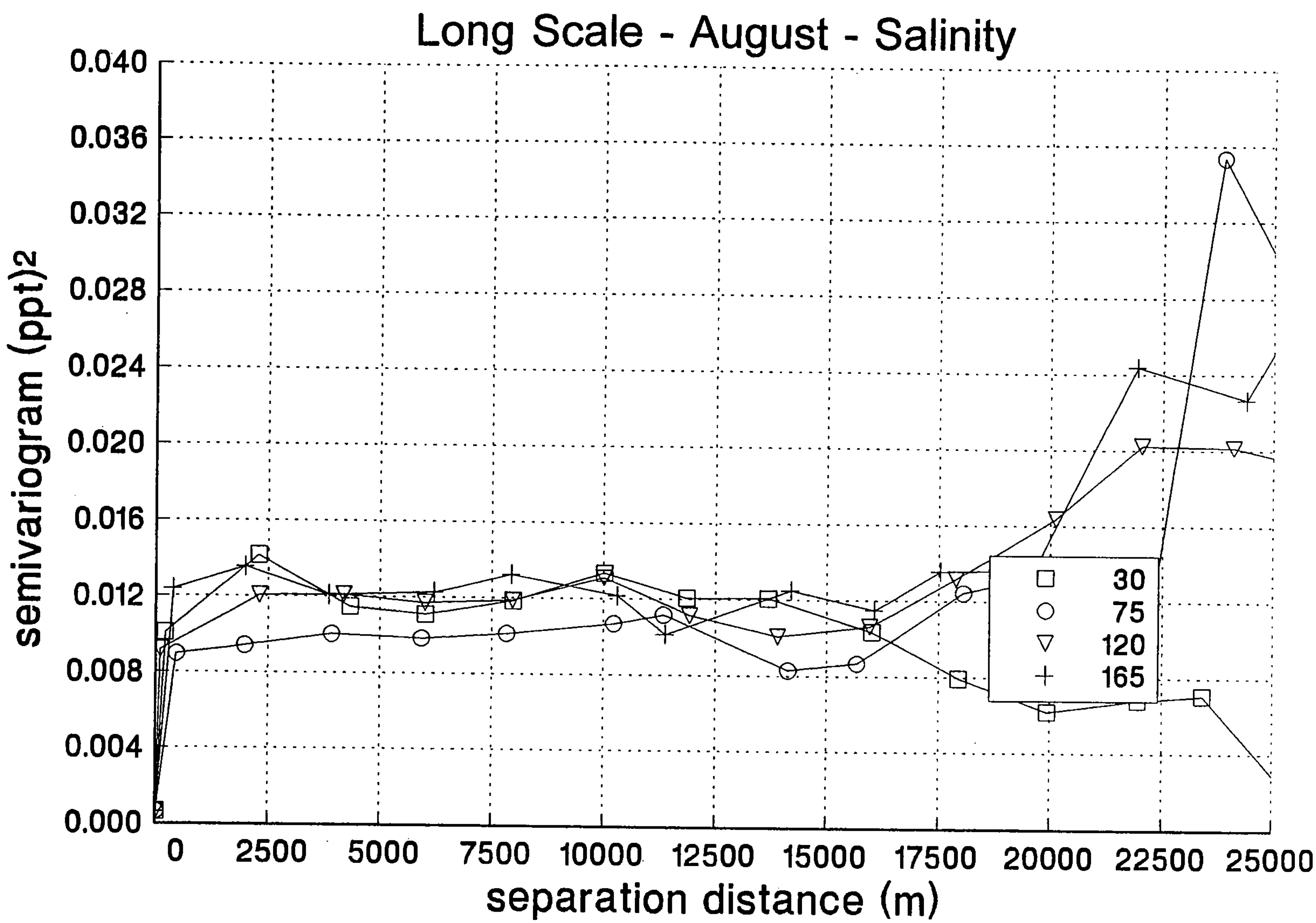


Figure 15. Long-scale spatial semivariogram showing variability in surface (0-5 m) salinity along the directions 30°, 75°, 120°, and 165° in August 1993.

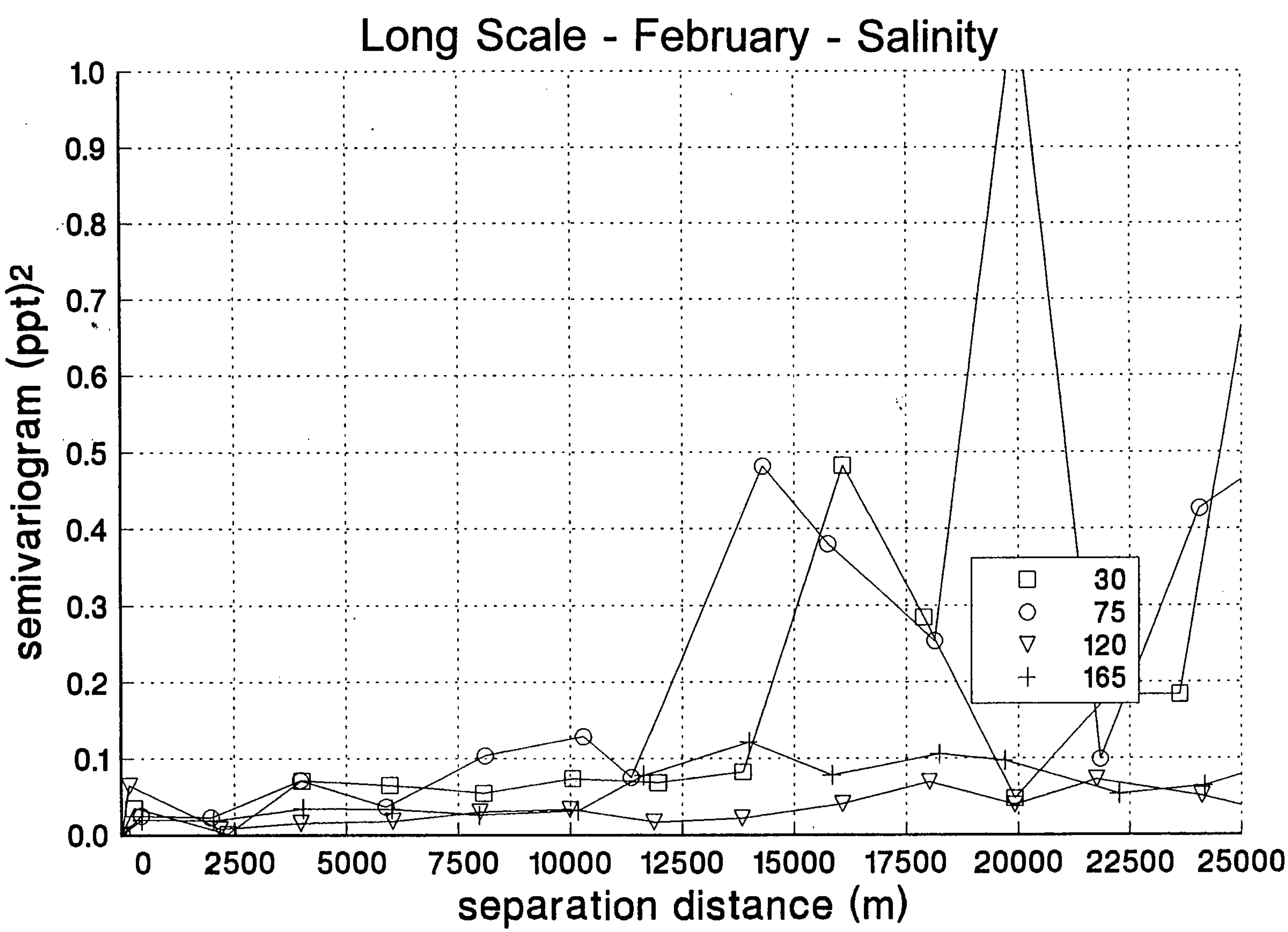


Figure 16. Long-scale spatial semivariogram showing variability in surface (0-5 m) salinity along the directions 30°, 75°, 120°, and 165° in February 1994.

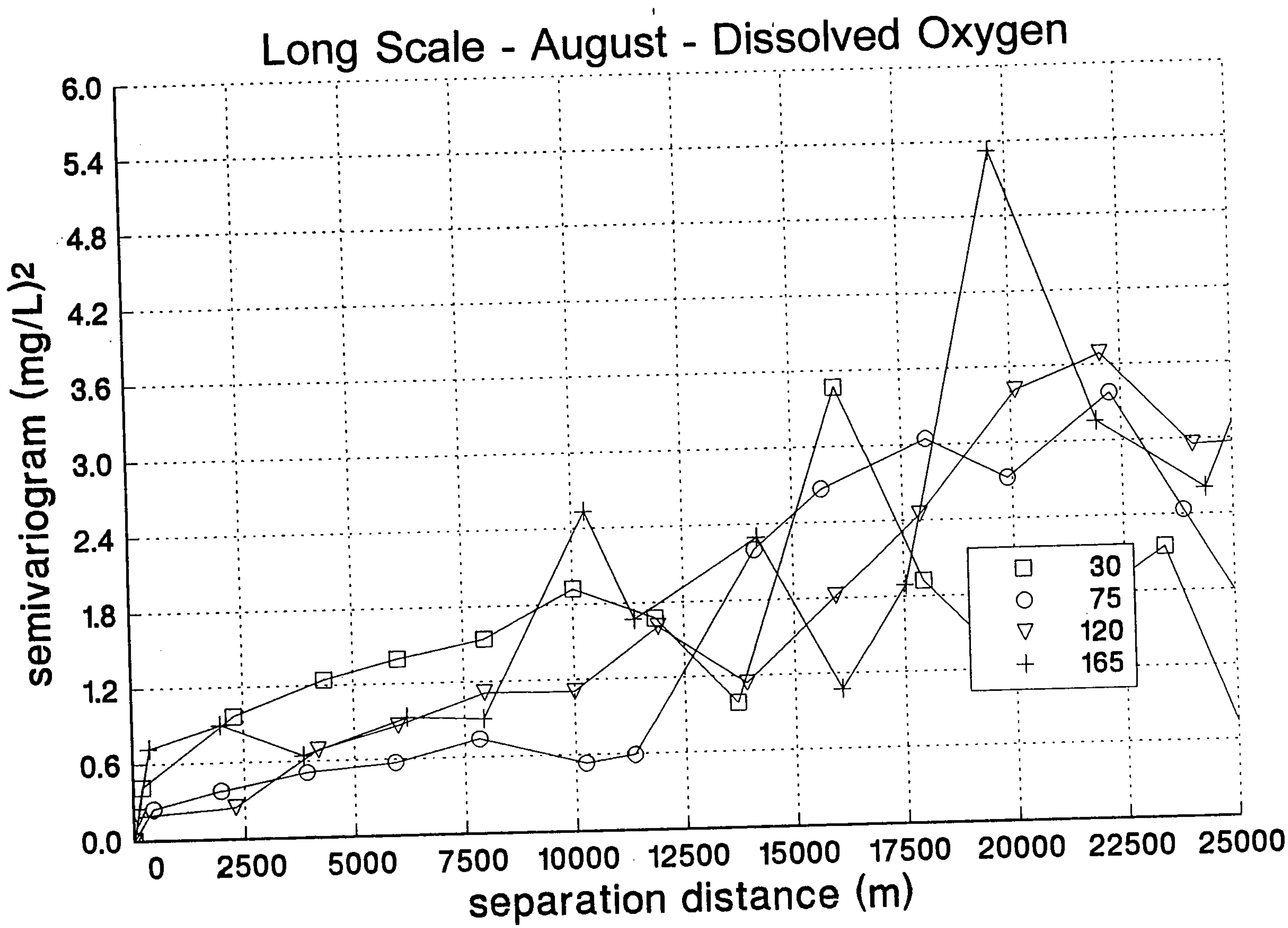


Figure 17. Long-scale spatial semivariogram showing variability in surface (0-5 m) dissolved oxygen along the directions 30°, 75°, 120°, and 165° in August 1993.

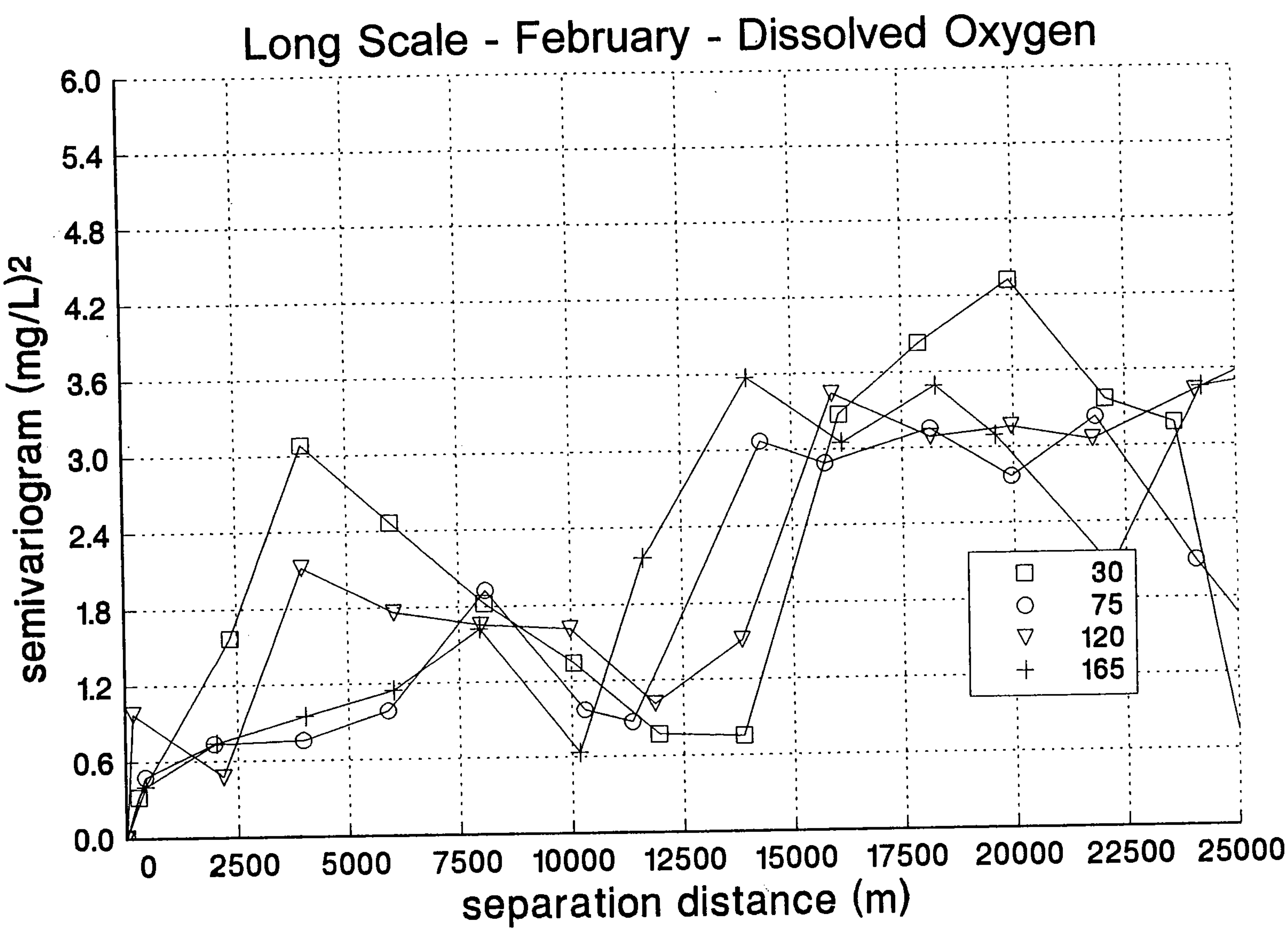


Figure 18. Long-scale spatial semivariogram showing variability in surface (0-5 m) dissolved oxygen along the directions 30°, 75°, 120°, and 165° in February 1994.

The second category is characterized by the February semivariogram values being much lower than those from August, increasing variability with separation distance in August but not February, and weak to moderate directional trends in August but not February. Figures 19 and 20 present the chlorophyll *a* semivariograms for August and February, respectively. Semivariograms for temperature and beam attenuations can be found in Appendix B. Comparison of the two chlorophyll *a* semivariograms clearly demonstrates the higher variability observed in August (0 to 40) relative to February (0.1 to 0.24) and the directional trends in the variability observed in the August data. The August chlorophyll *a* data show immediate separation of the variability as a function of separation distance in the 120° direction (east-southeast) relative to the other directions examined. Comparison of this directional trend to the contoured surface chlorophyll data (Figure 13) confirms the trend in the data. These results indicate the gradients in the data which should be considered when designing the Phase II water sampling plan.

The characteristics for the third category, which includes DIN, are similar to the second except that the variability in February is higher than in August, a reversal from the parameters included in the second category, and the weak directional trend in the 30° to 75° direction (north-northeast to east-northeast).

Table 2 summarizes the long-scale semivariogram results for all parameters. Note that generally well behaved semivariograms extend to about 10,000 m for all parameters except dissolved oxygen which behaves well to distances of about 22,000 m. These scales are 25 and 50% of the maximum distances among the stations used for the analysis, thus representing distances or block sizes that are 1-2 times the size of the nearfield. Further evaluations are required to determine if variability in larger block sizes can be confidently represented.

3.2.2 Semivariograms for the Photic Zone

Data from depths greater than 5 m were added to the analysis to evaluate the spatial scales of depth-integrated chlorophyll and related parameters. The characteristics of the photic depth semivariograms for salinity (Figures 21 and 22) were similar to those for the surface: a similar range in variability (0.005 to 0.6), slightly increasing semivariogram values as a function of separation distance, and larger variability in February versus August. They differed in that the overall variability in February was higher, primarily due to a weak directional trends (30° and 75°) that

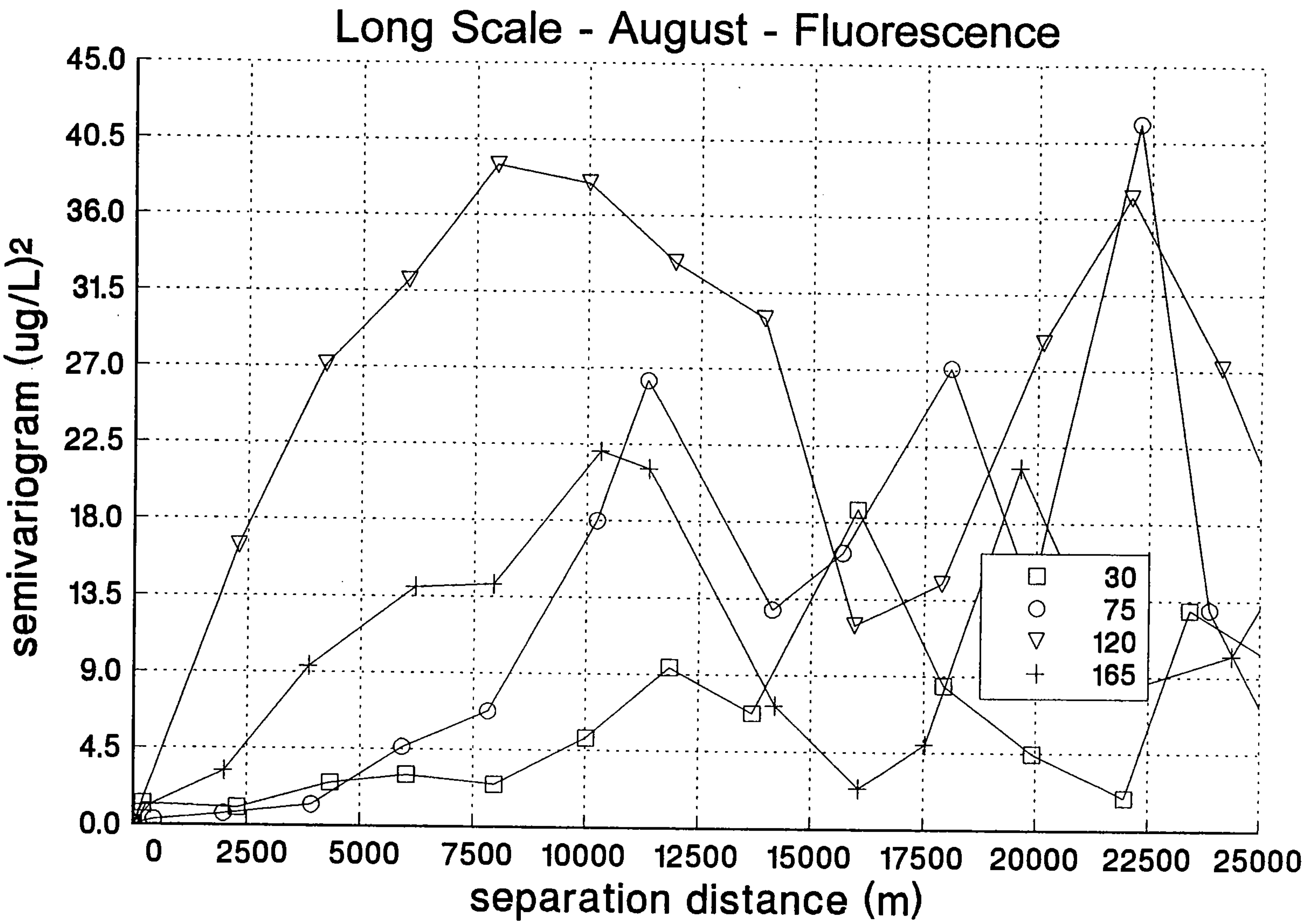


Figure 19. Long-scale spatial semivariogram showing variability in surface (0-5 m) chlorophyll a along the directions 30°, 75°, 120°, and 165° in August 1993.

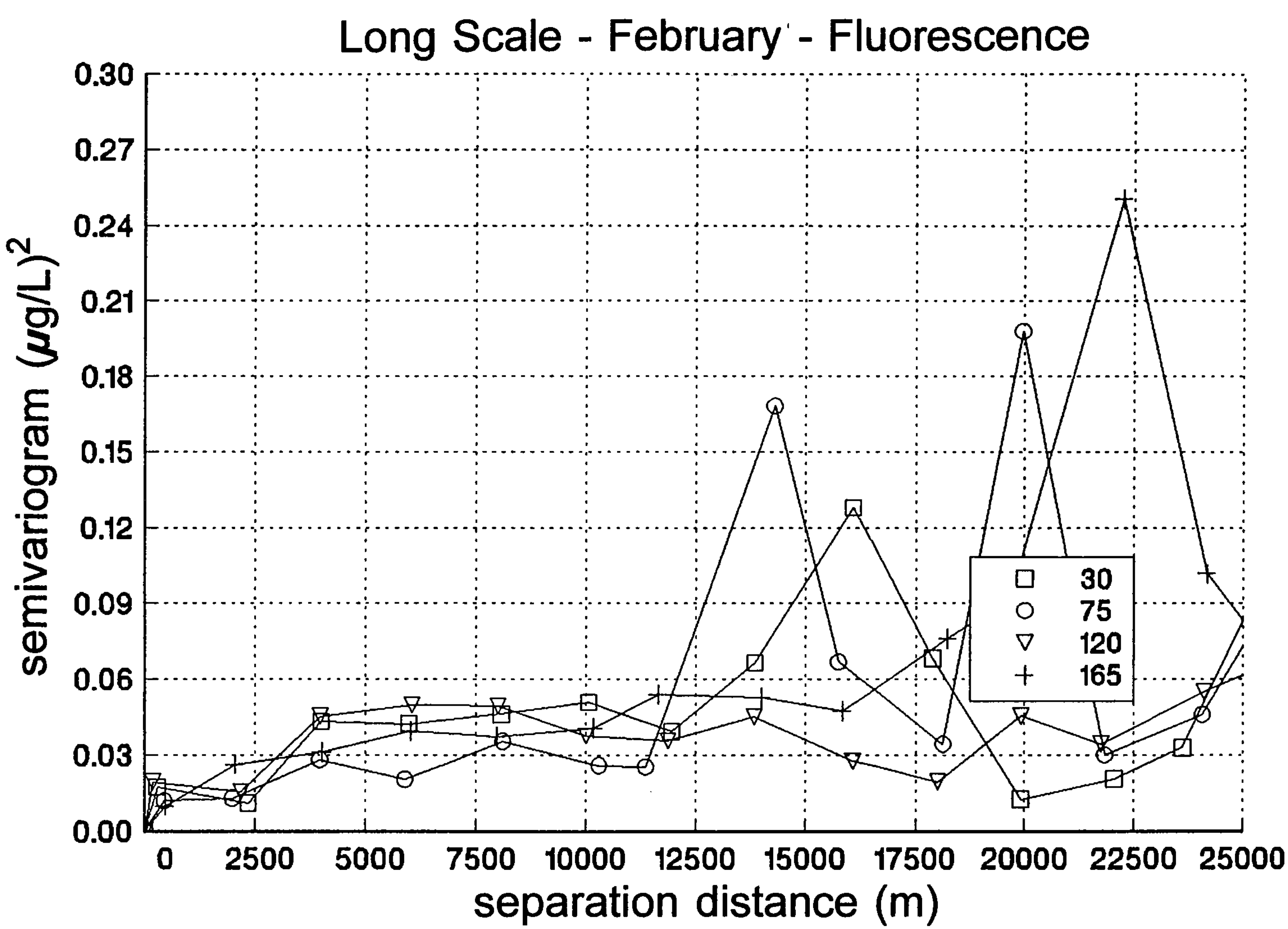


Figure 20. Long-scale spatial semivariogram showing variability in surface (0-5 m) chlorophyll *a* along the directions 30°, 75°, 120°, and 165° in February 1994.

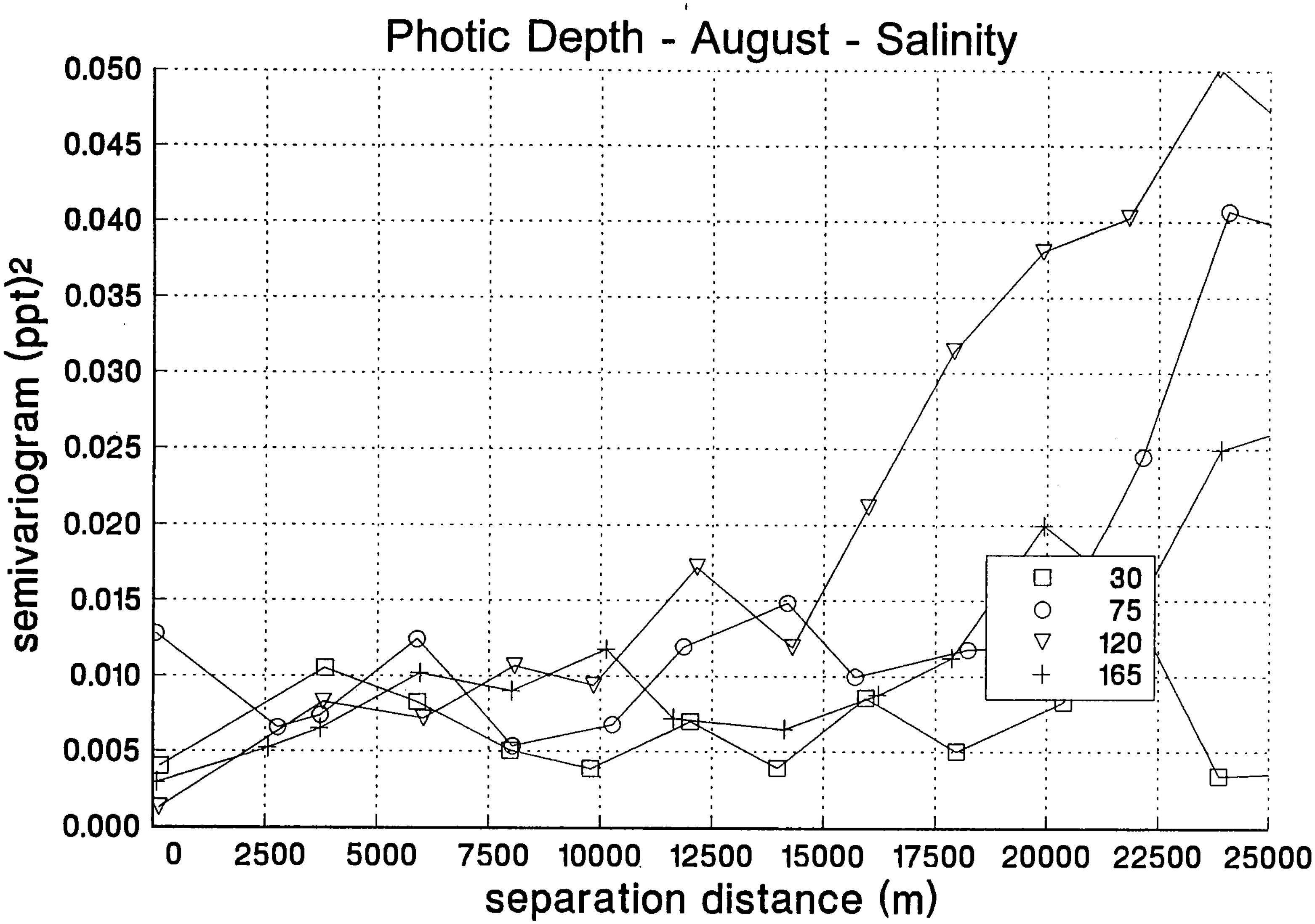


Figure 21. Long-scale spatial semivariogram showing variability in photic depth salinity along the directions 30°, 75°, 120°, and 165° in August 1993.

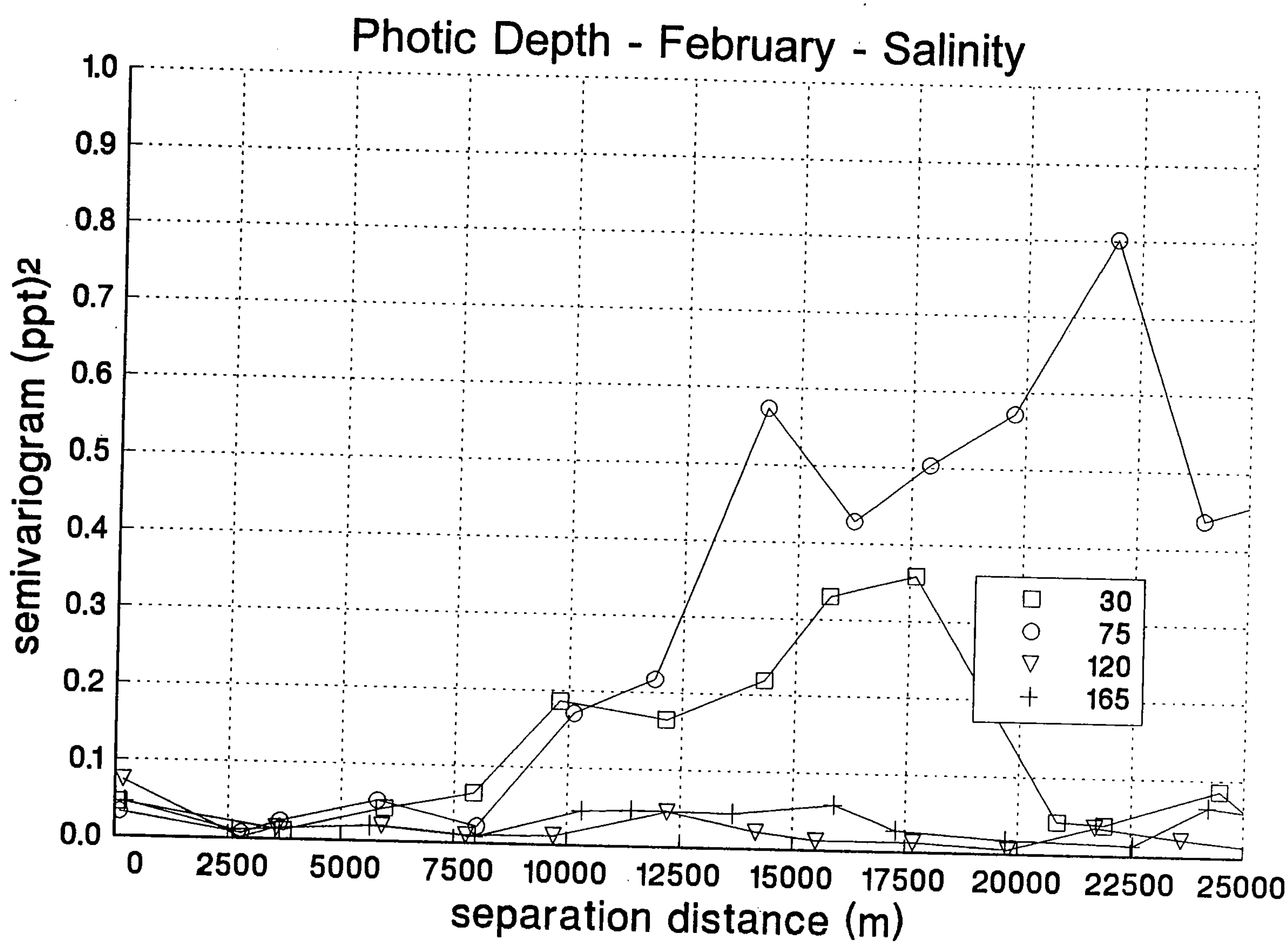


Figure 22. Long-scale spatial semivariogram showing variability in photic depth salinity along the directions 30°, 75°, 120°, and 165° in February 1994.

Table 2. Summary of long-scale semivariogram characteristics for the surface (0-5 m) water column throughout Massachusetts Bay.

Parameter	Semivariogram Value Range[1]	Variability August vs February	Variability increases with increasing separation distance	Directional Trends
Salinity	A: 0.008 to 0.012 F: 0 to 0.1	No differences	A: No F: No	A: None F: None
Dissolved Oxygen	A: 0.25 to 3.5 F: 0.25 to 3.5	No differences	A: Yes F: Yes	A: None F: None
Temperature	A: 0.1 to 1.5 F: 0.1 to 0.3	February is **lower** than August	A: Yes F: No	A: Yes F: None
Chlorophyll *a*	A: 0.01 to 40 F: 0.01 to 0.2	February is **lower** than August	A: Yes F: No	A: Along 120° F: None
Beam Attenuation	A: 0.02 to 3 F: 0.01 to 1	February is **lower** than August	A: Yes F: Yes	A: Along 120° F: None
Dissolved inorganic nitrogen	A: 0.005 to 0.03 F: 2 to 32	February is **higher** than August	A: No F: Yes	A: None F: Weak along 30° and 75°

[1] A = August
F = February

developed at separation distances greater than 7500 m. A weak directional trend (120°) was evident at distances greater than 15,000 m in August.

Other differences in the parameters include loss of directional tendency in the August photic zone data. This may be caused by removal of tow data from the analysis or inclusion of data from greater depths. Large increases in the variability in chlorophyll (Figures 23 and 24), compared to the surface data, may be due to the inclusion of higher concentrations at greater depths, and suggest that inclusion of the entire photic zone may provide a better measure of variability and trends than the surface results alone. Likewise, the differences in the photic depth temperature and salinity, relative to the surface-only semivariograms, are consistent with inclusion of higher variability induced through the inclusion of the pycnocline. In contrast, the beam attenuation is less variable in the photic zone data set relative to the surface data. Variance has some tendency to increase with distance for fluorescence and dissolved oxygen.

3.2.3 Spatial Scales and Detectable Change

Similar to the approach described in Sections 2.1 and 3.1.2, the spatial semivariograms presented in Sections 3.2.1 and 3.2.2 can be used, along with kriging and power analysis, to evaluate water column monitoring designs. The approach illustrated in this section is for both surface and photic zone chlorophyll *a* measured during August 1993 and February 1994. In contrast to the analysis presented in Section 3.1.2, this section considered estimation at various spatial scales by evaluating chlorophyll at three fixed locations (denoted A, B, C):

- Location A represents short spatial scales in the nearfield area, approximately 2 km north-northwest of station N11
- Location B represents larger spatial scales in the farfield area, approximately 5 km north-northwest of station F10
- Location C represents the largest spatial scales in the extended farfield area, approximately 15 km north-northeast of station F03

In this analysis, spatial scale was measured as the approximate distance to the nearest available monitoring stations under each of three monitoring designs:

- The first design included all 46 of the current nearfield and farfield stations.
- The second design included only 23 monitoring stations by retaining every other nearfield and farfield station (N1, N3, N4, N7, N8, N10, N12, N13, N16, N17, N19, N20, F2, F3, F7, F9, F11, F12, F14, F16, F19, F20, and F24).

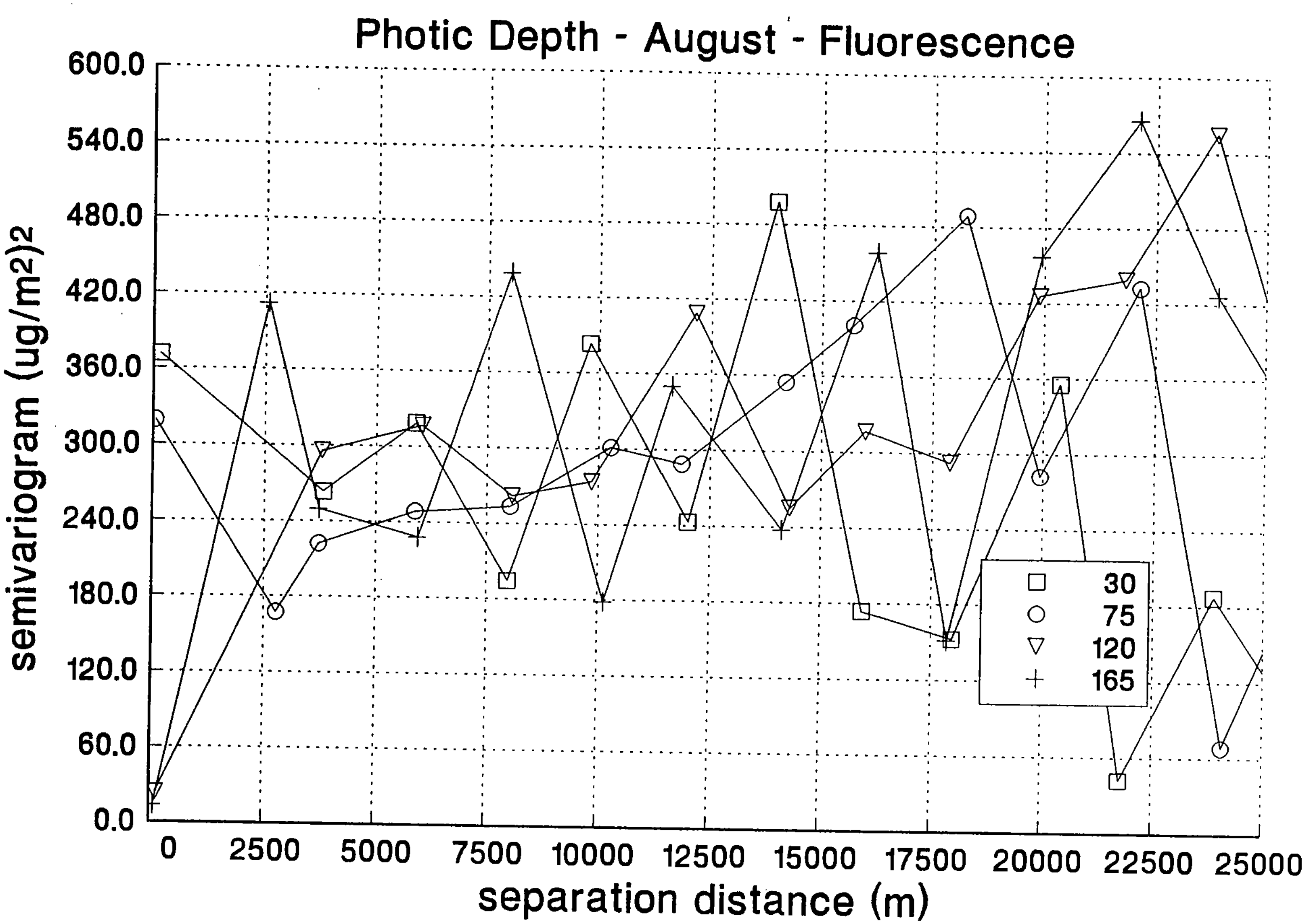


Figure 23. Long-scale spatial semivariogram showing variability in photic depth chlorophyll a along the directions 30°, 75°, 120°, and 165° in August 1993.

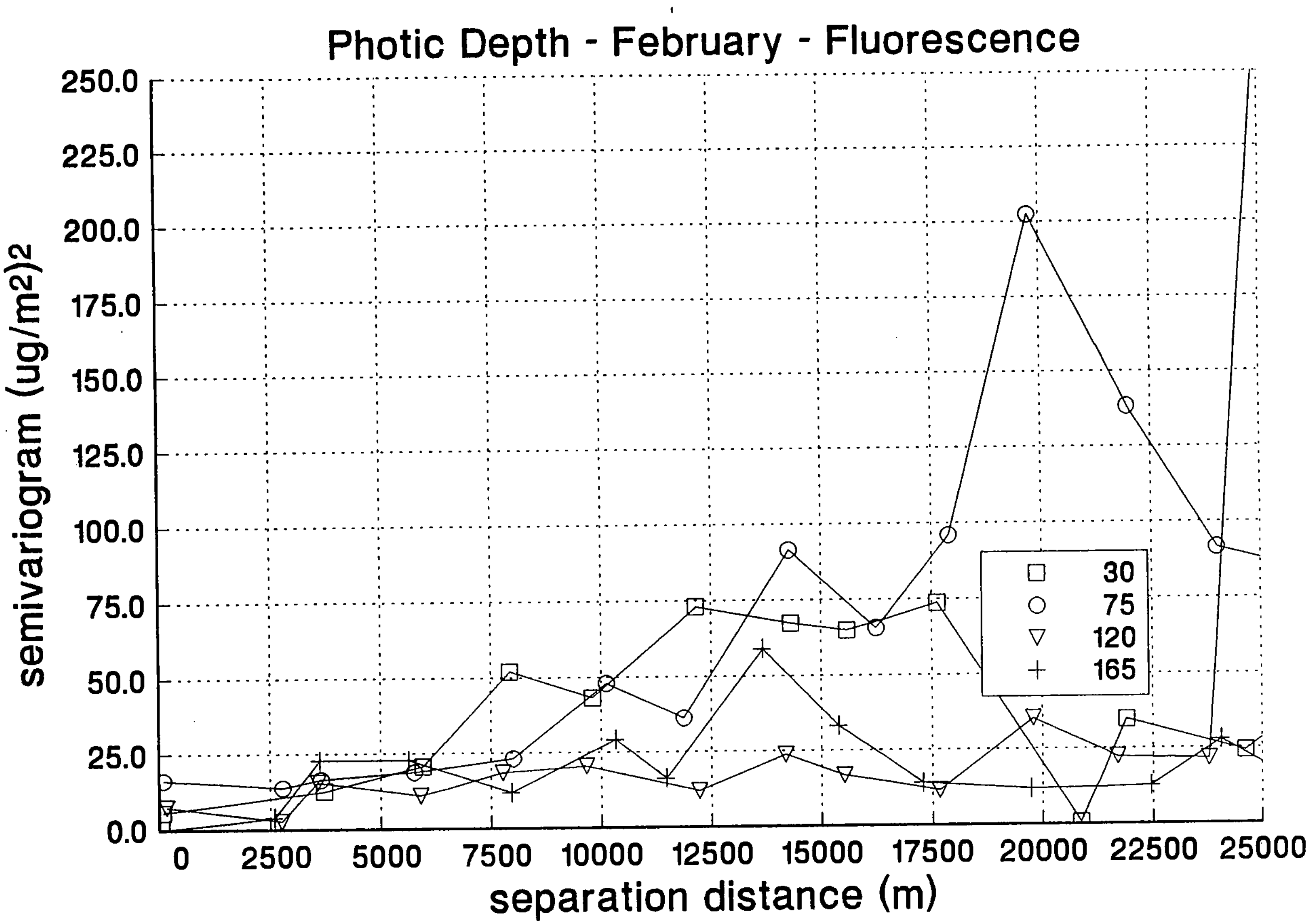


Figure 24. Long-scale spatial semivariogram showing variability in photic depth chlorophyll *a* along the directions 30°, 75°, 120°, and 165° in February 1994.

- The third design included only 11 monitoring stations by retaining every other station from the second design (N1, N4, N7, N10, N16, N20, F3, F9, F12, F14, and F20).

The statistical uncertainty associated with each monitoring design was evaluated by estimating the chlorophyll *a* concentration in both the surface waters and the photic zone at locations A, B, and C using a kriging analysis and the monitoring stations listed above for each of the three monitoring designs. Uncertainty is presented in Table 3 in terms of the change in chlorophyll *a* concentration that can be detected with each of the monitoring designs at each of the three fixed locations (A, B, and C). This analysis assumed a 95% confidence level and 80% power when calculating detectable change.

As an example, consider the result for the August chlorophyll measured in the surface waters. At the shortest spatial scale, location A is situated approximately 2 km from the nearest monitoring stations under the full station design, and a change of 3.7 μg/L could be detected in the chlorophyll concentration at that location. As the number of monitoring stations is reduced to 23 and to 11 stations, location A lies further from the nearest monitoring station (*i.e.*, 3 km and 4 km, respectively), and the statistical uncertainty and associated detectable difference increase to 3.8 and 7.8 μg/L, respectively. Also note that the trend of increasing uncertainty with increasing spatial scale continues for the second fixed location, B; however, uncertainty remains about the same when moving to the largest spatial scales represented by location C. This characteristic is directly linked to the fact that the corresponding semivariograms (Figure 19), which were used in the kriging analysis, were not found to increase beyond separation distances of about 7 km.

As an additional example, consider the results for August chlorophyll measured in the photic zone. At the shortest spatial scale under the full 46 station design, a change of 29.8 $\mu g/m^2$ could be detected in the chlorophyll concentration at location A. As noted in the surface chlorophyll, decreasing the number of monitoring stations to 23 and to 11 stations, increases the statistical uncertainty and associated detectable difference to 34.7 $\mu g/m^2$ and 40.3 $\mu g/m^2$, respectively.

Note that the difference in chlorophyll that can be detected at location A under the 46 and 11 station designs decreases by about 110% in the surface waters, compared to 35% over the photic zone. This suggests that inclusion of the entire photic zone data in the evaluation improves the difference that can

Table 3. Detectable Change as a Function of the Number of Monitoring Stations and Spatial Scale

No. of Stations	Fixed Location for Estimation[1]					
	A		B		C	
	Spatial Scale[2]	Detectable Difference	Spatial Scale[2]	Detectable Difference	Spatial Scale[2]	Detectable Difference
August Surface Chlorophyll						
	km	μg/L	km	μg/L	km	μg/L
46	2	3.7	5	7.8	15	6.9
23	3	3.8	7	7.9	19	6.9
11	4	7.8	8	11.1	23	6.9
February Surface Chlorophyll						
	km	μg/L	km	μg/L	km	μg/L
46	2	0.27	5	0.45	15	0.47
23	3	0.32	7	0.47	19	0.47
11	4	0.40	8	0.49	23	0.49
August Photic Zone Chlorophyll						
	km	$\mu g/m^2$	km	$\mu g/m^2$	km	$\mu g/m^2$
46	2	29.8	5	41.0	15	41.0
23	3	34.7	7	41.5	19	41.5
11	4	40.3	8	42.5	23	42.5
February Photic Zone Chlorophyll						
	km	$\mu g/m^2$	km	$\mu g/m^2$	km	$\mu g/m^2$
46	2	5.9	5	9.4	15	14.1
23	3	6.7	7	10.7	19	14.3
11	4	8.6	8	12.7	23	14.7

[1] Location A is approximately 2 km NNW of Station N11.
Location B is approximately 5 km NNW of Station F10.
Location C is approximately 15 km NNE of Station F03.

[2] Approximate distance to nearest monitoring stations.

be detected relative to the surface data alone. In contrast, the relative detectable differences between the 46 and 23 station designs are only 2.7 and 16% for the surface versus photic zone, respectively.

One general result that can be observed in Table 3 is that uncertainty, and thus detectable difference, is much lower in February chlorophyll than in the August chlorophyll. This fact is directly related to lower chlorophyll concentrations in February, along with lower spatial variability in those concentrations. In most cases, the results in Table 3 also indicate a greater increase in uncertainty when changing from the 23 station monitoring design to the 11 station design, than when changing from the full 46 station design to the 23 station design. This result is noteworthy given the fact that changing from the full design to the 23 station design required elimination of significantly more monitoring stations (23) than changing from the 23 station to the 11 station design (elimination of 12 stations).

3.3 Reverse Power Analysis

The mean chlorophyll *a*, dissolved oxygen, and DIN concentrations across stations for each nearfield survey completed between February 1992 and May 1994 are shown in Figures 25, 26, and 27, respectively. A repeating annual cycle for each of these parameters is evident. In addition to the mean concentration, the standard deviation for each survey is included on the graph and is represented by the horizontal line above and below the mean concentration connected by a vertical line. For chlorophyll and DIN, the heavier horizontal line that falls between the mean concentration and the upper error limit represents the increase in concentration (effect size) that can be detected for each survey. For oxygen, lower and upper detection limits are shown. Note that the change in concentration that can be detected varies as the magnitude of the mean concentration changes. Smaller changes are detectable in the winter months for chlorophyll and DIN; larger changes are detectable in the summer. This is due, in part, to the greater variability in these parameters with depth during the summer periods (winter conditions tend to be more homogeneous). Note that the change that can be detected in dissolved oxygen appears to be relatively constant.

To more easily visualize the change in detectable concentrations for each survey and parameter, the percent detectable change relative to the mean for each survey was calculated. This is simply the percent increase (or decrease) in concentration that can be observed based on the available data. It is

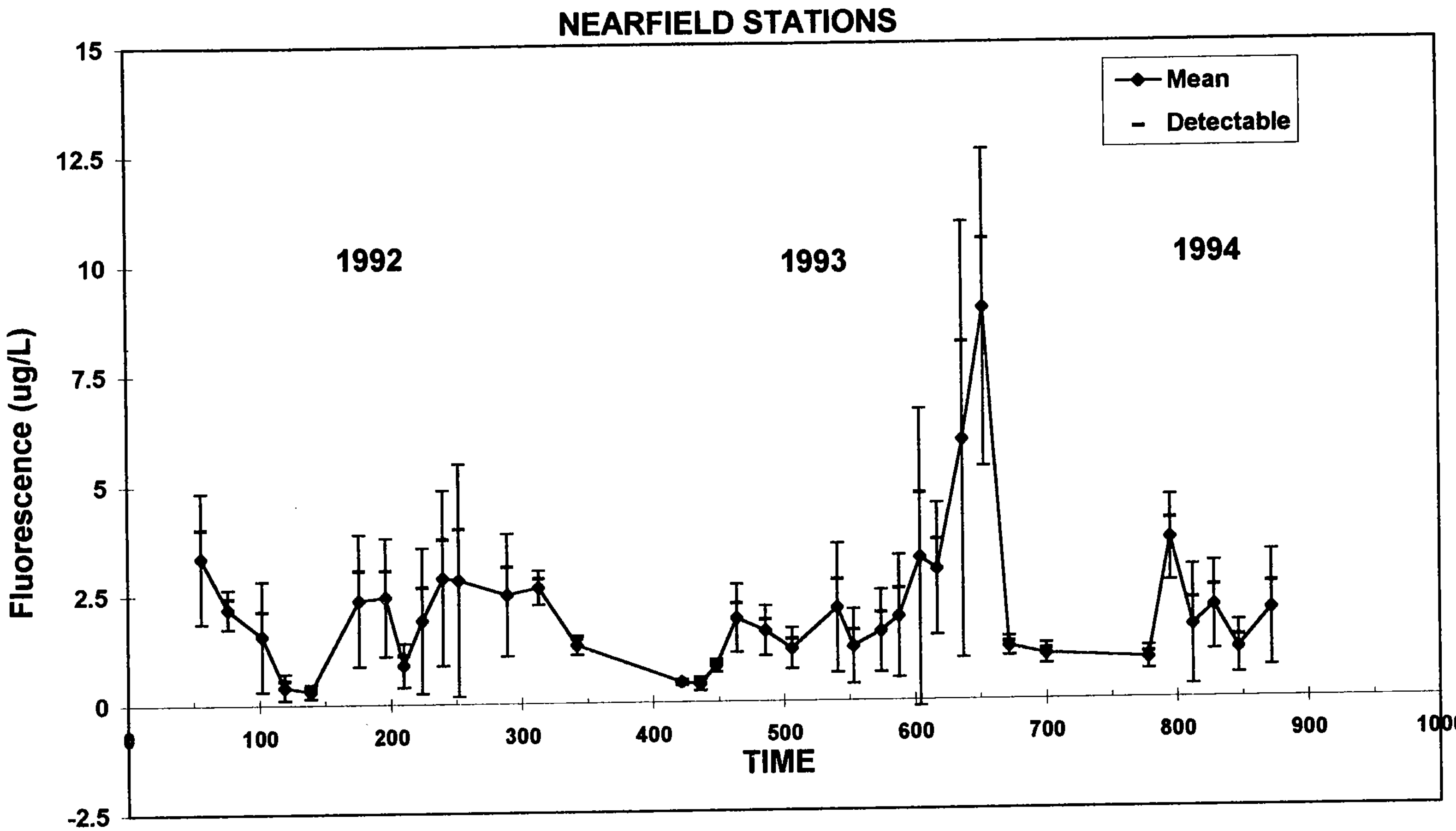


Figure 25. Chlorophyll *a* concentrations as a function of time from nearfield surveys completed between February 1992 and May 1994. The mean concentrations and standard deviations for each survey are shown. The single heavy bar between the mean and the standard deviation is the concentration that can be detected for the survey assuming $\alpha = 95\%$ and $1\text{-}\beta = 80\%$.

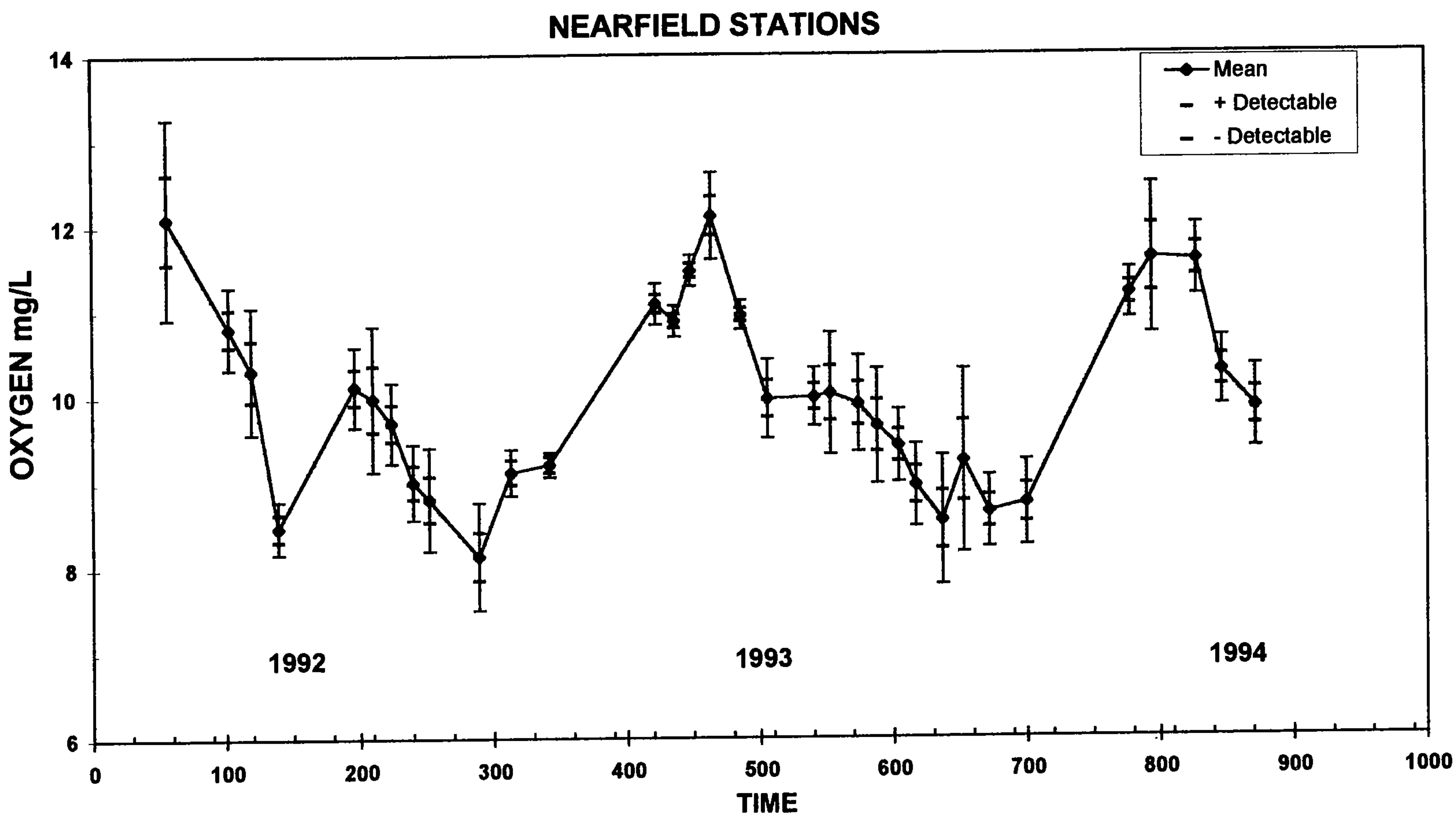


Figure 26. Dissolved oxygen concentrations as a function of time from nearfield surveys completed between February 1992 and May 1994. The mean concentrations and standard deviations for each survey are shown. The single heavy bar between the mean and the standard deviation is the concentration that can be detected for the survey assuming α = 95% and 1-β = 80%.

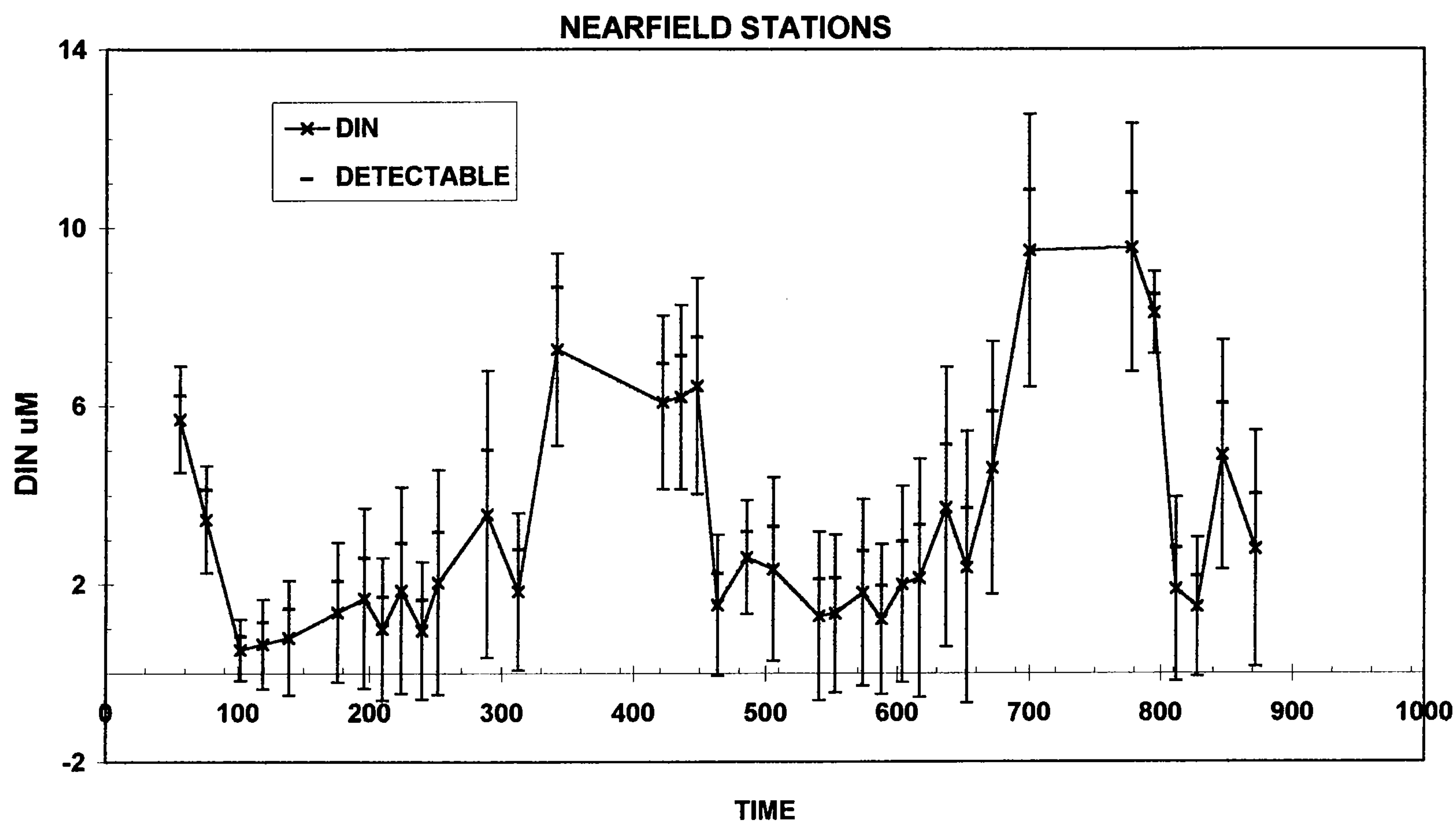


Figure 27. Dissolved inorganic nitrogen (DIN) concentrations as a function of time from nearfield surveys completed between February 1992 and May 1994. The mean concentrations and standard deviations for each survey are shown. The single heavy bar between the mean and the standard deviation is the concentration that can be detected for the survey assuming $\alpha = 95\%$ and $1\text{-}\beta = 80\%$.

calculated as the upper (lower) detectable concentration minus the mean divided by the mean (*e.g.*, if the mean = 10 and the detectable change is = 1, then (11-10)/10 = 0.1 or a 10% change). The results of this exercise for chlorophyll, DIN, and dissolved oxygen are shown in Figure 28. Note that the percentages are for each parameter and are presented independently and are not cumulative. The smallest changes that can be detected are for dissolved oxygen, generally <5% of the mean for any given survey. The changes in chlorophyll that can be detected range from 7 to 45%, while the detectable changes for DIN range from 12 to 85%. These data clearly show that the detectable change (effect size) varies seasonally. Also note that the data have been treated as though temporal correlations between surveys and seasons do not exist. This may result in an overestimate (or underestimate) of the true change that can be detected. Regardless, the present water column sampling design provides adequate ability to statistically detect small changes within the nearfield. The temporal variability in the detectable change also indicates, as concluded from the kriging analysis, that the final Phase II monitoring design must take into account both spatial and temporal responses. This further dictates that the monitoring plan (1) include clear indicator and endpoint values, (2) define the attributes (*i.e.*, averaging period, spatial scales) associated with these values, and (3) include decision criteria (*i.e.*, duration/frequency) for any hypothesis developed to test change.

Summary statistics were developed for these parameters to provide guidance on changes that can be detected under the present design (Table 4). For this summary, the mean (mean of the means) and associated standard deviation of survey averages were calculated for each parameter. Similarly, the mean detectable change (in percent and concentration units) was calculated. These summary statistics suggest changes of about 25, 2.3, 100, and 43% of the mean nearfield concentration can be used as guidance levels for chlorophyll, dissolved oxygen, TSS, and DIN, respectively.

The level of detectable change that should be measured for the various water column monitoring parameters has not been established. To provide guidance in this area, the minimum and maximum concentrations that might be detected under various levels of detectable change were calculated (Table 5). For example, if it is assumed that the mean nearfield dissolved oxygen concentration is 10 mg/L (no temporal period assumed) and the desired level of detectable change is ±2%, then the upper and lower concentrations that could be detected for the entire nearfield would be 9.8 and 10.2 mg/L, respectively. Similarly, setting the desirable level of detectable for chlorophyll, the

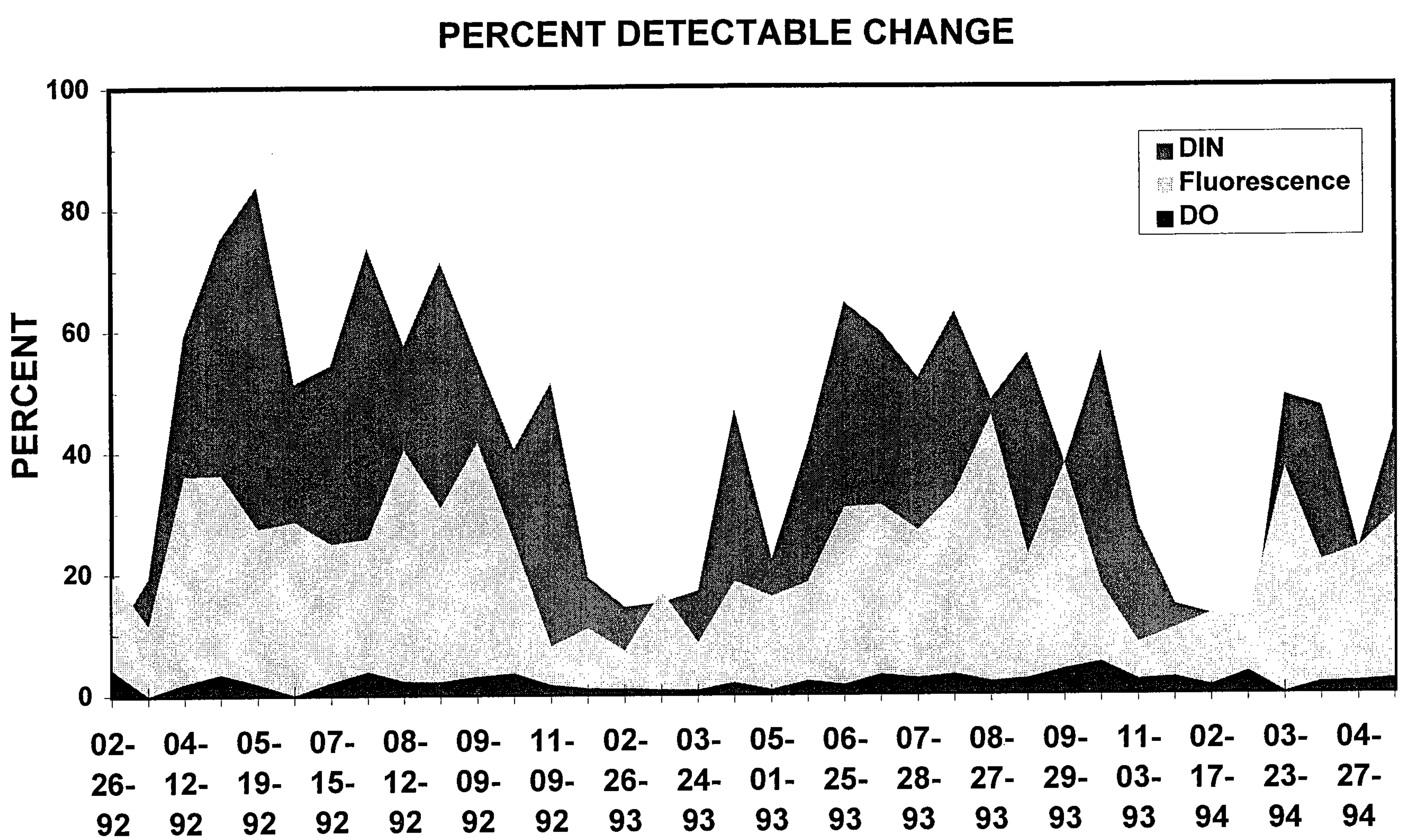


Figure 28. Percent detectable change for chlorophyll *a*, dissolved oxygen, and dissolved inorganic nitrogen as a function of time for all nearfield surveys completed between February 1992 and May 1994.

Table 4. Summary of detectable change in parameters measured on nearfield surveys between February 1992 and May 1994.

Parameter	Background Concentration	Detectable Change		
		%	Concentration	n (Surveys)
Chlorophyll *a* (μg/L)	2.1 ± 1.6	23.7 ± 10.7	0.5 ± 0.5	36
Dissolved Oxygen (mg/L)	10.0 ± 1.1	2.3 ± 1.1	0.23 ± 0.11	33.00
TSS (mg/L)	1.5 ± 0.7	99.1 ± 44.8	1.6 ± 1.6	24
DIN (μM)	3.2 ± 2.5	42.7 ± 21.2	0.93 ± 0.30	36

Table 5. Comparison of detectable concentrations based on desirable levels of detectable change.

Parameter	Mean Concentration February 1992 to May 1994	Detectable Change (mean %) of all surveys	Desirable Level of Detectable Change (%)					
			20[a]		100 (2x)		200 (3x)	
			Min	Max	Min	Max	Min	Max
Chlorophyll *a* (μg/L)	2	23.7	1.6	2.6	<1	4	<1	6
Dissolved Oxygen (mg/L)	10	2.3	8	12	-	-	-	-
TSS (mg/L)	1.5	99	-	-	-	3.0		4.5
DIN (μM)	3.2	42.7	-	-	7.2	-	-	9.6

[a]For dissolved oxygen, the 2% minimum and maximum are 9.8 and 10.2 mg/L, respectively; for 10%, the 2% minimum and maximum are 9 and 11 mg/L, respectively.

minimum effect level that could be detected would be 1.6 mg/L and the maximum would be 2.6 mg/L.

The actual effect size that the monitoring plan should target depends on a number of factors including the present concentrations in the water column, spatial and temporal scales of concern, and dissolved oxygen levels that are considered to be detrimental in the receiving waters. Given the limited analysis described above, it appears reasonable to state that the present nearfield design, without consideration of water depth or other ways to stratify the data, is more than sufficient to detect any levels of concern and to describe temporal trends in response to the outfall relocation. For example, if the level of meaningful change or the endpoint for dissolved oxygen in the system is considered to be 6 mg/L and the present levels of dissolved oxygen range from 7.5 to 11 mg/L, the detectable changes under the present design would be more than sufficient to provide warning of approach to the endpoint. If the limits for chlorophyll *a* are set at 18 or even 12 μg/L, then the 25% detectable change (against the average chlorophyll concentrations of $\approx$2.0 μg/L) provided by the present design would be more than capable of statistically detecting an approach towards the endpoint. Similar arguments can be made for the other parameters, as summarized in Table 5. Depending on the endpoints chosen, it appears that detection of twofold (100%) or threefold (200%) changes may be more than adequate for the monitoring program. This will, of course, depend greatly on the endpoint values chosen. Given the ability to detect small changes in the nearfield and expecting that endpoints that will be defined for the system are much larger (or smaller) than the concentrations presently observed in the system, these preliminary estimates suggest that the nearfield area may be over sampled relative to the detectable change required for the monitoring program. If so, the cost of the monitoring program could be lowered by implementing a simpler field program. The optimal spatial and temporal design awaits, however, definition of the confidence levels and change to be detected, and agreement on meaningful levels of change and endpoints for the key parameters, and clearly stated hypotheses with well-defined attributes (decision criteria) against which monitoring data can be used to test for unacceptable change. Once defined, more extensive statistical treatments of the data (such as that discussed in Section 3.1) can be applied.

4.0 SUMMARY

4.1 Conclusions

The exploratory statistical treatments of the water column data were designed to provide information that could contribute to resolving questions related to the MWRA post-discharge (Phase II) monitoring program. Preliminary information related to several of these questions is summarized below.

What are the spatial scales of variability in key water column parameters in Massachusetts Bay? Semivariogram analysis of key monitoring parameters in the Massachusetts Bay system indicates that the spatial variability depends greatly on the parameter of interest. Overall, the variability is low and for key biological parameters (*e.g.*, dissolved oxygen and chlorophyll *a*) generally increases as distance between sampling points increases. Fine-scale spatial variability (20 to 100s of meters) is evident from the high-resolution tow-yo data. Directional trends for several parameters were detected, although they are seasonally and possibly depth dependent. The statistically detected directional trends in the data are consistent with gradients observed in the field data. Overall, semivariogram results suggest that the system is well behaved at scales up to 10-12 km for most parameters and at scales up to 20 km for dissolved oxygen. Clear seasonality in the variability within the nearfield is evident. This temporal variability is greater than the spatial variability, suggesting that the monitoring program should focus more on the annual cycle than on spatial differences.

Are temporal and spatial scales for the Phase I baseline water column sampling adequate? Kriging analysis, based on two years of nearfield surface chlorophyll *a* data, suggests that temporal variability is greater than the spatial variability. Preliminary estimates of precision, gained by increasing the number of stations sampled, does not increase greatly once 12 stations are included in the nearfield spatial design. Increasing the number of annual surveys from 12 to 17 improves the ability to detect change in chlorophyll concentrations in the nearfield. Overall, the spatial and temporal water column sampling for the baseline period appears more than adequate to develop information from which statistically optimum and cost-effective sampling designs can be developed. However, this question can not be fully and definitively answered until the level of detectable change that is acceptable for the monitoring program is defined.

Are there alternative sampling designs that could be considered without loss of statistical confidence? The preliminary data suggest that alternate designs can be developed without loss of statistical confidence and possibly more cost effectively. Alternate designs, however, depend greatly on the specific questions being asked of the monitoring program and on the spatial/temporal attributes associated with those questions. Development of specific alternatives depends on definition of the issues identified above.

What is the change in the system that the present monitoring design can detect? The baseline water column monitoring design is capable of detecting changes well below any levels of concern. However, the level of detectable change is specific to each individual parameter. For the nearfield area, the detectable change ranged from 2% (dissolved oxygen) to 100% (total suspended solids) of the present baseline concentrations. Evaluation of the detectable change in the chlorophyll concentration as a function of spatial scale for representative summer and winter periods demonstrates that the change that can be detected is seasonally dependent. The smallest detectable differences were 0.27 μg chlorophyll *a*/L in February and 3.7 μg chlorophyll *a*/L in August at stations at spatial scales of about 2 km. Little difference in the detectable change at all scales (2-20 km) was observed in February while detectable differences substantially increased over this scale in August. Little difference in the detectable change was noted over scales of intermediate and longer scales (5-20 km). More important, the ability to detect change in chlorophyll is not greatly diminished when a 23 station design is employed relative to the present 46 station design. In contrast, a decrease to 11 sampling stations substantially decreases the change that can be detected relative to both the present 46 station design and a 23 station design. These results further point to the need to articulate the level of change that the monitoring program will be designed to detect.

Are these detectable changes sufficient to measure change before meaningful levels are approached? As indicated above, the preliminary assessments suggest that the baseline design is more than adequate to provide warning of unacceptable change or movement towards unacceptable levels. The assessment will benefit greatly when unacceptable levels of change are defined. Such information will enable more quantitative refinement of the monitoring program and will likely reduce the monitoring cost.

4.2 Recommendations

Recommendations for completing the statistical design of the outfall monitoring program include the following:

(1) Finalize detectable levels of change that will be applied to each key parameter so that the optimal monitoring program design can be developed. The exploratory studies discussed in this report suggest that, relative to present conditions, detectable change goals could be set at

10-20% for dissolved oxygen
100-200% (2x-3x) for chlorophyll
100% for total suspended solids
100-200% for dissolved inorganic nitrogen

(2) Finalize endpoints and warning levels that will be used to develop monitoring hypotheses.

(3) Further evaluate the directional trends in the data to establish whether farfield station locations are appropriately placed relative to final monitoring hypotheses and detection of changes outside of the nearfield.

(4) Further evaluate the apparent influence of water column depth on the estimates of variability to determine impact on detectable change levels and monitoring design.

(5) Optimize the spatial sampling scales and temporal sampling frequency through additional kriging analysis and based on information developed in response to recommendations (1) though (5).

5.0 REFERENCES

Albro, C.S., J.R. Kelly, J. Hennessy, P. Doering, and J. Turner. 1993. Combined work/quality assurance project plan for baseline water quality monitoring. Massachusetts Water Resources Authority, ENQUAD Technical Report Series , ms-14. 73 pp.

Cressie, N.A.C. 1993. *Statistics for Spatial Design*. John Wiley & Sons, New York. 900 pp.

Hillman, R.A., M.J. Moore, C.S. Peven, D.A. Lewis, L. Hanson, C.D. Hunt, and J. Stegeman. 1994. 1993 Annual fish and shellfish report. Massachusetts Water Resources Authority, ENQUAD Technical Report Series 94-9. 195 pp.

Hunt, C.D. and E. Baptiste. 1993. Evaluation of detectable changes in chemical contaminants in flounder and lobster tissue. Appendix B *In:* D. Shea, annual review of toxic contaminants discharged by MWRA: 1993, Massachusetts Water Resources Authority, ENQUAD Technical Report Series 93-18. 68 pp. plus appendices.

Hunt, C.D. and M. Steinhauer. 1994a. MWRA nutrient indicator workshop. Massachusetts Water Resources Authority, ENQUAD Technical Report Series, ms-20. 180 pp.

Hunt, C.D. and M. Steinhauer. 1994b. Massachusetts Water Resources Authority fish and shellfish science review meeting. Massachusetts Water Resources Authority, ENQUAD Technical Report Series, ms-21. 114 pp.

Hunt, C.D., M. Steinhauer, and K. Keay. 1994. Massachusetts Water Resources Authority benthic and sediment science meeting. Massachusetts Water Resources Authority, ENQUAD Technical Report Series ms-22. 119 pp.

Isaaks, E.H. and R.M. Srivastava. 1989. *An Introduction to Applied Geostatistics*. Oxford University Press, New York. 561 pp.

Journel, A.G. and Ch.J. Huijbregts. 1981. *Mining Geostatistics*. Academic Press, London. Reprinted with corrections. 600 pp.

Kelly, J.R. 1993. Nutrients and Massachusetts Bay: an update of eutrophication issues. Massachusetts Water Resources Authority, ENQUAD Technical Report Series 93-17. 119 pp.

Kelly, J.R., C.S. Albro, J.T. Hennessy, and D. Shea. 1992. Water quality monitoring in Massachusetts and Cape Cod Bays: February – March 1992. Massachusetts Water Resources Authority, ENQUAD Technical Report Series 92-08. 171 pp.

Kelly, J.R., C.S. Albro, J.T. Hennessy, and D. Shea. 1993a. Water quality monitoring in Massachusetts and Cape Cod Bays: April – August 1992. Massachusetts Water Resources Authority, ENQUAD Technical Report Series 93-1. 270 pp.

Kelly, J.R., C.S. Albro, and J.T. Hennessy. 1993b. Water quality monitoring in Massachusetts and Cape Cod Bays: August – November 1992. Massachusetts Water Resources Authority, ENQUAD Technical Report Series 93-15. 213 pp.

Kelly, J.R., C. Albro, K. Foster, J. Hennessy, P. Doering, L. Reed, E. Requintina, J. Turner, and D. Borkman. 1993c. Water quality monitoring in Massachusetts and Cape Cod Bays: annual report for 1992. Massachusetts Water Resources Authority, ENQUAD Technical Report Series 93-116. 129 pp.

Kelly, J.R., C.S. Albro, J.T. Hennessy, J. Turner, D. Borkman, and P. Doering. 1994a. Water quality monitoring in Massachusetts and Cape Cod Bays: December 1992, February 1993, and March 1993. Massachusetts Water Resources Authority, ENQUAD Technical Report Series 94-2. 197 pp.

Kelly, J.R., C.S. Albro, J.T. Hennessy, J. Turner, D. Borkman, and P. Doering. 1994b. Water quality monitoring in Massachusetts and Cape Cod Bays: April and May 1993. Massachusetts Water Resources Authority, ENQUAD Technical Report Series 94-3. 143 pp.

Kelly, J.R., C.S. Albro, J.T. Hennessy, J. Turner, D. Borkman, and P. Doering. 1994c. Water quality monitoring in Massachusetts and Cape Cod Bays: June – July 1993. Massachusetts Water Resources Authority, ENQUAD Technical Report Series 94-11. 152 pp.

Kelly, J.R., C.S. Albro, J.T. Hennessy, P.S. Libby, J. Turner, D. Borkman, and P. Doering. 1994d. Water quality monitoring in Massachusetts and Cape Cod Bays: August – September 1993. Massachusetts Water Resources Authority, ENQUAD Technical Report Series 94-12. 172 pp.

Kelly, J.R., C.S. Albro, J.T. Hennessy, J. Turner, D. Borkman, and P. Doering. 1994e. Water quality monitoring in Massachusetts and Cape Cod Bays: October – November 1993. Massachusetts Water Resources Authority, ENQUAD Technical Report Series 94-13. 143 pp.

Kelly, J.R., P.S. Libby, C.S. Albro, J.T. Hennessy, J. Turner, D. Borkman, and P. Doering. 1994f. Water quality monitoring in Massachusetts and Cape Cod Bays: February – March 1994. Massachusetts Water Resources Authority, ENQUAD Technical Report Series 94-14. 176 pp.

Kelly, J.R., P.S. Libby, C.S. Albro, J.T. Hennessy, J. Turner, D. Borkman, and P. Doering. 1994g. Water quality monitoring in Massachusetts and Cape Cod Bays: April – May 1994. Massachusetts Water Resources Authority, ENQUAD Technical Report Series 94-15. 136 pp.

MWRA. 1991. Massachusetts Water Resources Authority effluent outfall monitoring plan phase I: baseline studies. Massachusetts Water Resources Authority, Boston, MA. 95 pp.

Peterman, R.M. 1989. Application of statistical power analysis to the Oregon coho salmon (*Oncorhynchus kisutch*) problem. *Can. J. Fish. Aquat. Sci.* 46:1183-1187.

APPENDIX A

SPECIFICATIONS OF THE KRIGING MODELS USED IN THIS REPORT

Ordinary kriging was utilized in five cases to provide illustrative examples to MWRA of a general approach for quantifying statistical tradeoffs. These cases and the corresponding semivariogram models are summarized below with references to the applicable text section.

(Section 3.1.1)

Temporal and spatial variability in log-transformed surface chlorophyll measurements --The temporal (Figure 3) and spatial semivariograms for data across all months (Figure 6) were fitted (for subsequent kriging purposes) with a zonal anisotropic correlation model containing two components, first a geometric anisotropic spherical component (nugget variance = 0.01 $[\ln(\mu g/L)]^2$, total sill variance = 0.41 $[\ln(\mu g/L)]^2$) with an isotropic spatial range of 11,000 meters and a temporal range of 30 days, and second a zonal spherical component (total sill variance = 0.70 $[\ln(\mu g/L)]^2$) in the temporal direction with a range of 90 days.

(Section 3.2.1)

Spatial variability in August 1993 surface chlorophyll measurements -- The kriging analysis for August 1993 surface chlorophyll used a geometric anisotropic spherical semivariogram model (nugget variance = 1 $(\mu g/L)^2$, total sill variance = 32 $(\mu g/L)^2$) with a range in the N120°E direction of 8000 meters and a range in the N30°E direction of 240,000 meters.

Spatial variability in February 1994 surface chlorophyll measurements -- The kriging analysis for February 1994 surface chlorophyll used an isotropic spherical semivariogram model (nugget variance = 0.006 $(\mu g/L)^2$, total sill variance = 0.040 $(\mu g/L)^2$) with a range of 8000 meters.

(Section 3.2.2)

Spatial variability in August 1993 photic zone chlorophyll measurements -- The kriging analysis for August 1993 photic zone chlorophyll used an isotropic spherical semivariogram model (nugget variance = 30 $(mg/m^2)^2$, total sill variance = 300 $(mg/m^2)^2$) with a range of 5000 meters.

Spatial variability in February 1994 photic zone chlorophyll measurements kriging analysis for February 1994 photic zone chlorophyll used an isotropic spherical semivariogram model (nugget variance = 3 $(mg/m^2)^2$, total sill variance = 33 $(mg/m^2)^2$) with a range of 15,000 meters.

Response to Andrew Solow Comments

November 30, 1995

1. The report suffers considerably from the absence of a clear definition (or definitions) of what would constitute a change of interest. For example, the discussion in Sections 3.1.2 and 3.2.3 revolves around the problem of detecting a change through time in the measured value of a variable of interest at a single location. This case does not seem particularly relevant to monitoring water quality. Interest may more reasonably center, for example, on detecting a year-to-year change in mean chlorophyll-a concentration within a region. The failure to specify more clearly what sort of changes are of interest - while presumably not the fault of the authors of this draft report - is a serious problem.

   Response. We agree that several different statistical monitoring objectives could be formulated and that one possibility is monitoring for year-to-year changes in the mean chlorophyll concentration within a region. Clearly, both the size of the spatial zone of interest must be established (e.g., choices range from a single point in the Bay to spatial averages across various regions in the Bay), as well as the date and length of time of interest (e.g., choices range from a single day at a specified date during the year to temporal averages across specified days, weeks, or months). At several places in the report (e.g., last paragraph of Executive Summary; last paragraph of Section 3.1.2; third, fourth, fifth, and sixth paragraphs of Section 4.1; and points 1, 2, and 5 in Section 4.2) we indicate that our analyses are exploratory and are intended to demonstrate the kinds of statistical evaluations that could be performed, but that final design of the monitoring program requires clear articulation by MWRA and other interested parties of the monitoring objectives.

2. Throughout the report, the measure of change is the difference between the actual value and a standard. This is somewhat different from a change in the actual value from year to year. In this latter case, the value in neither year is known and the problem is somewhat more difficult.

   Response. We agree that another possible monitoring objective could be formulated to evaluate the change in various water quality parameters from year to year rather than monitoring for the difference from a fixed standard. Once again, this is a choice of monitoring objectives that needs to be made by MWRA before the appropriate statistical tradeoffs can be assessed and an adequate monitoring design determined. The objective illustrated in our report is certainly appropriate for cases where a fixed environmental standard (as used in the MWRA draft Phase II monitoring plan and the Outfall Contingency Plan) already exists or can be established (e.g., minimum human or aquatic exposure level to prevent adverse health effects), and it does present a (marginally) simpler illustrative example of the statistical approach for MWRA.

3. My interpretation of the semi-variograms included in the draft report is that a reasonable statistical model for many of the variables contains a seasonal cycle, a spatial trend

(possibly a function of distance from shore), and additional variability that may be nearly uncorrelated over temporal and spatial sampling scales.

Response. The alternative trends model which the reviewer discusses compared with the no-trends model which we utilized amounts to a choice between universal kriging and ordinary kriging approaches (Journel and Huijbregts, 1981, p. 313). Experience has often shown that if the objective is to estimate water quality itself (rather than the trend in water quality), which is MWRA's problem, at locations and times close to existing water quality measurements (i.e., no distant extrapolation is required), then universal and ordinary kriging give quite similar results. Use of a trends model can provide marginally higher statistical confidence, but only by adopting more stringent assumptions--namely, that the forms of the trends are known without error and that the residual semivariogram for the additional variability above and beyond the trends can be reliably inferred by detrending the data. Incidentally, this additional variability might, or might not, be correlated in time or space. Part of our objective was simply to evaluate whether temporal or spatial trends might exist in the water quality parameters. As can be seen by our discussion of numerous semivariograms, the strongest case for a trend can probably be made for temporal changes in chlorophyll levels. Spatial trends in this and other parameters appear to be localized and/or relatively weak, in which case a spatial trends model would be less appropriate. The intent of our correlation modeling was to provide an illustration to MWRA of the kinds of statistical tradeoffs that can be made, not to necessarily recommend a final temporal or spatial correlation model.

4. The draft report does not contain sufficient detail (e.g., about the semi-variogram models used in the kriging) for a thorough evaluation.

Response. Ordinary kriging was utilized in five cases to provide illustrative examples to MWRA of a general approach for quantifying statistical tradeoffs. These cases and the corresponding semivariogram models are summarized below.

(1) Temporal and spatial variability in log-transformed surface chlorophyll measurements (Section 3.1.1)--The temporal (Figure 3) and spatial semivariograms for data across all months (Figure 6) were fitted (for subsequent kriging purposes) with a zonal anisotropic correlation model containing two components, first a geometric anisotropic spherical component (nugget variance = 0.01 $[\ln(\mu g/L)]^2$, total sill variance = 0.41 $[\ln(\mu g/L)]^2$) with an isotropic spatial range of 11,000 meters and a temporal range of 30 days, and second a zonal spherical component (total sill variance = 0.70 $[\ln(\mu g/L)]^2$) in the temporal direction with a range of 90 days.

(2) Spatial variability in August 1993 surface chlorophyll measurements (Section 3.2.1) --The kriging analysis for August 1993 surface chlorophyll used a geometric anisotropic spherical semivariogram model (nugget variance = 1 $(\mu g/L)^2$, total sill variance = 32 $(\mu g/L)^2$) with a range in the N120°E direction of 8000 meters and a range in the N30°E direction of 240,000 meters.

(3) Spatial variability in February 1994 surface chlorophyll measurements (Section 3.2.1)--The kriging analysis for February 1994 surface chlorophyll used an isotropic

spherical semivariogram model (nugget variance = 0.006 $(\mu g/L)^2$, total sill variance = 0.040 $(\mu g/L)^2$) with a range of 8000 meters.

(4) Spatial variability in August 1993 photic zone chlorophyll measurements (Section 3.2.2)--The kriging analysis for August 1993 photic zone chlorophyll used an isotropic spherical semivariogram model (nugget variance = 30 $(mg/m^2)^2$, total sill variance = 300 $(mg/m^2)^2$) with a range of 5000 meters.

(5) Spatial variability in February 1994 photic zone chlorophyll measurements (Section 3.2.2)--The kriging analysis for February 1994 photic zone chlorophyll used an isotropic spherical semivariogram model (nugget variance = 3 $(mg/m^2)^2$, total sill variance = 33 $(mg/m^2)^2$) with a range of 15,000 meters.

5. There are occasional curiosities. For example, the overall levels of the directional semi-variograms in Figure 9 are different. This implies that the variances of the observations used to construct these semi-variograms were different. It is difficult to understand how this can be, unless different sets of observations were used to construct these semi-variograms, in which case the differences may not be directional at all.

   Response. The data set used for the semivariogram analysis in Section 3.2.1 is described in Section 2.1.2, and Section 3.2.1 clearly states that a single combined (i.e., tow-yo, nearfield, and farfield) data set was used in the analysis. However, in Figure 9 one must remember that 3000 of the 3046 data lie along the four transects, and because of the extreme density of data (i.e., at a 20 meter spacing) the N75°E semivariogram is mostly estimated from data on the northern and southern transects, while the N165°E semivariogram is mostly estimated from data on the western and eastern transects. Also, the different overall variance levels for these two directions are not unusual; in fact, different directional variances are commonly indicative of a zonal anisotropy (Journel and Huijbregts, 1981, p. 181) which can often be interpreted as a localized or weak trend in the zonal direction.

6. As another example, Figure 11 appears to contain a periodic component with period of around 600 m out to a range of about 2000 m. Could this be an artifact of the sampling?

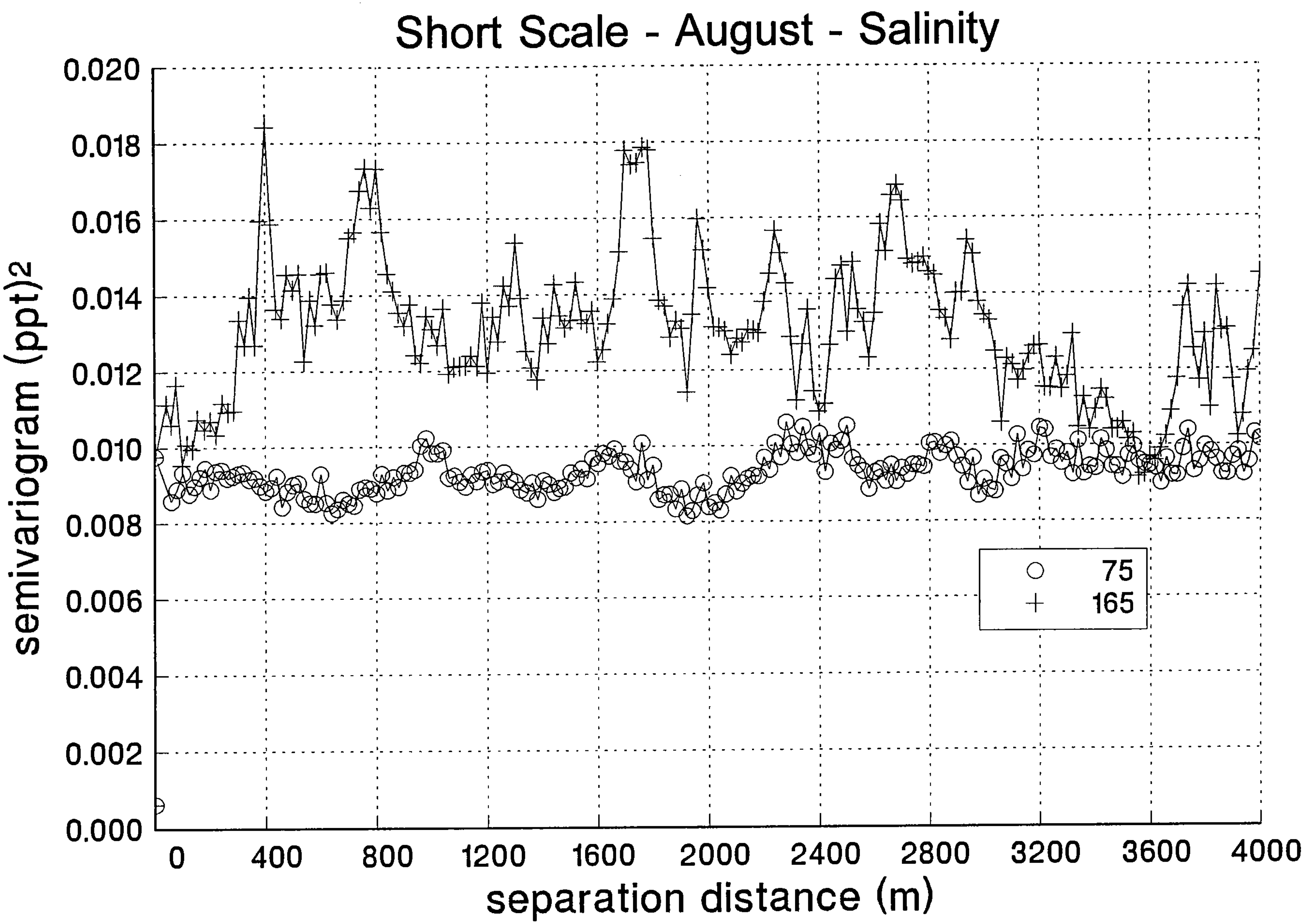
   Response. Yes, this apparent periodic component might be an artifact of sampling. The tow-yo data were taken in clusters of samples (spaced 20 meters apart) where the clusters were separated by anywhere from 200 meters to more than 1000 meters. We did not investigate the possible effects on semivariogram estimation of declustering the tow-yo data; however, it is interesting to note that this apparent periodicity is not seen in the N75°E semivariogram even though clustering was present in all four tow-yo transects.

7. A more careful specification of the requirements of the monitoring plan is needed before specific recommendations along these lines can be formulated.

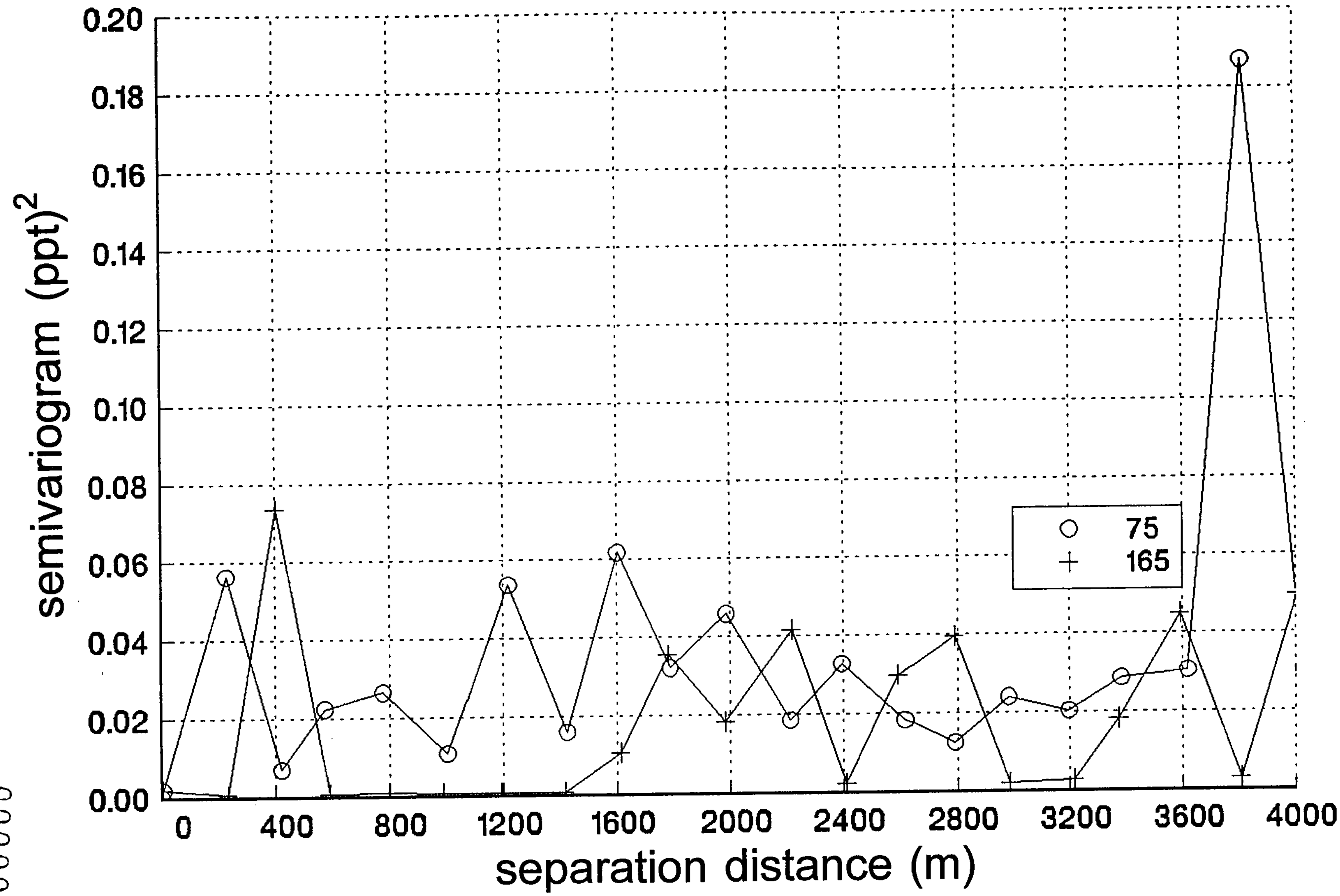
Response. As detailed in our response to Comment No. 1, the reviewer is simply reinforcing the same recommendation which we made to MWRA in several places in the report.

APPENDIX B

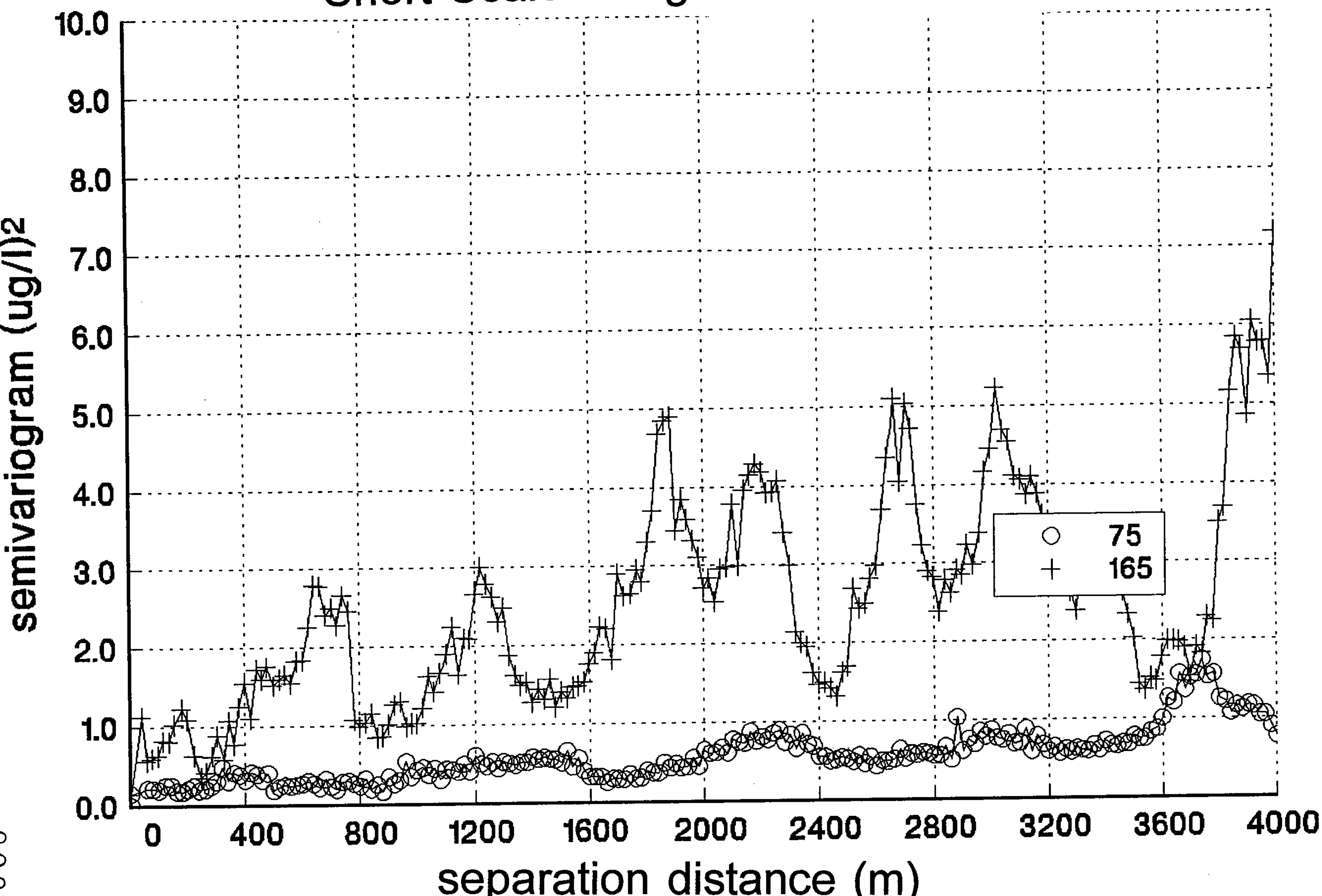
KRIGING RESULTS

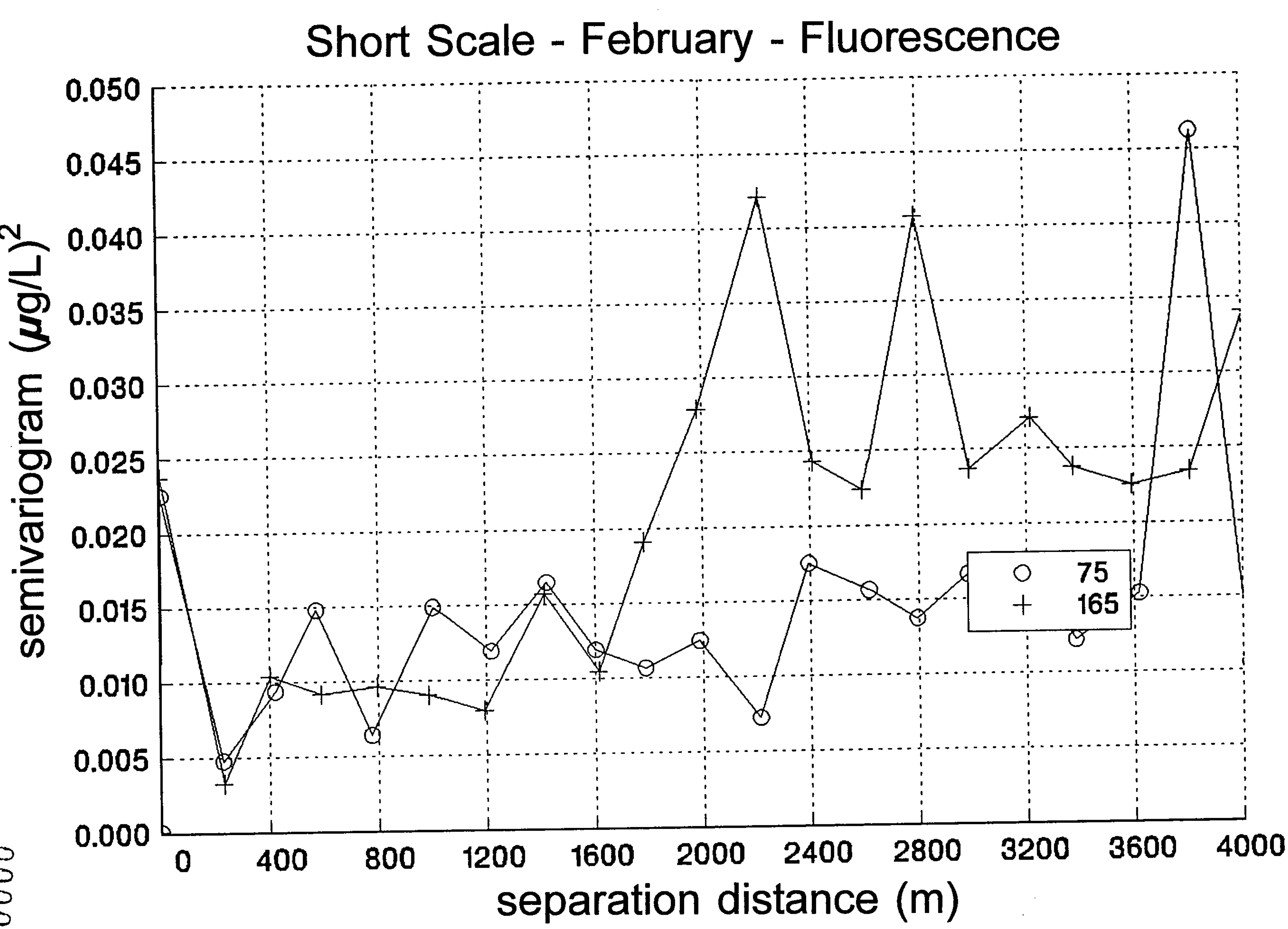

Short Scale - August - Salinity
semivariogram (ppt)2
0.020
0.018
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
400
800
1200
1600
2000
2400
2800
3200
3600
4000
separation distance (m)
75
165

Short Scale - February - Salinity



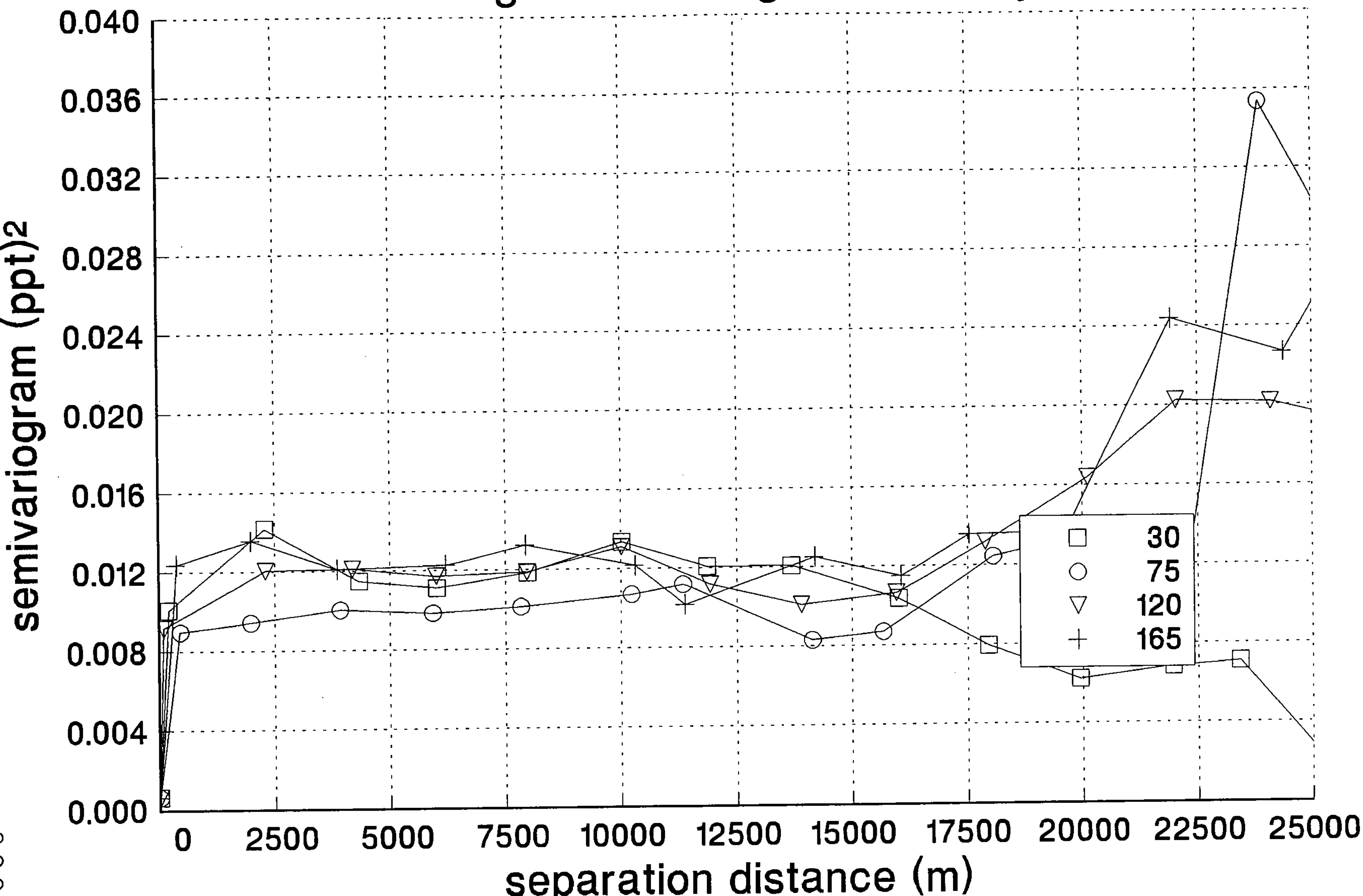
Short Scale - August - Fluorescence



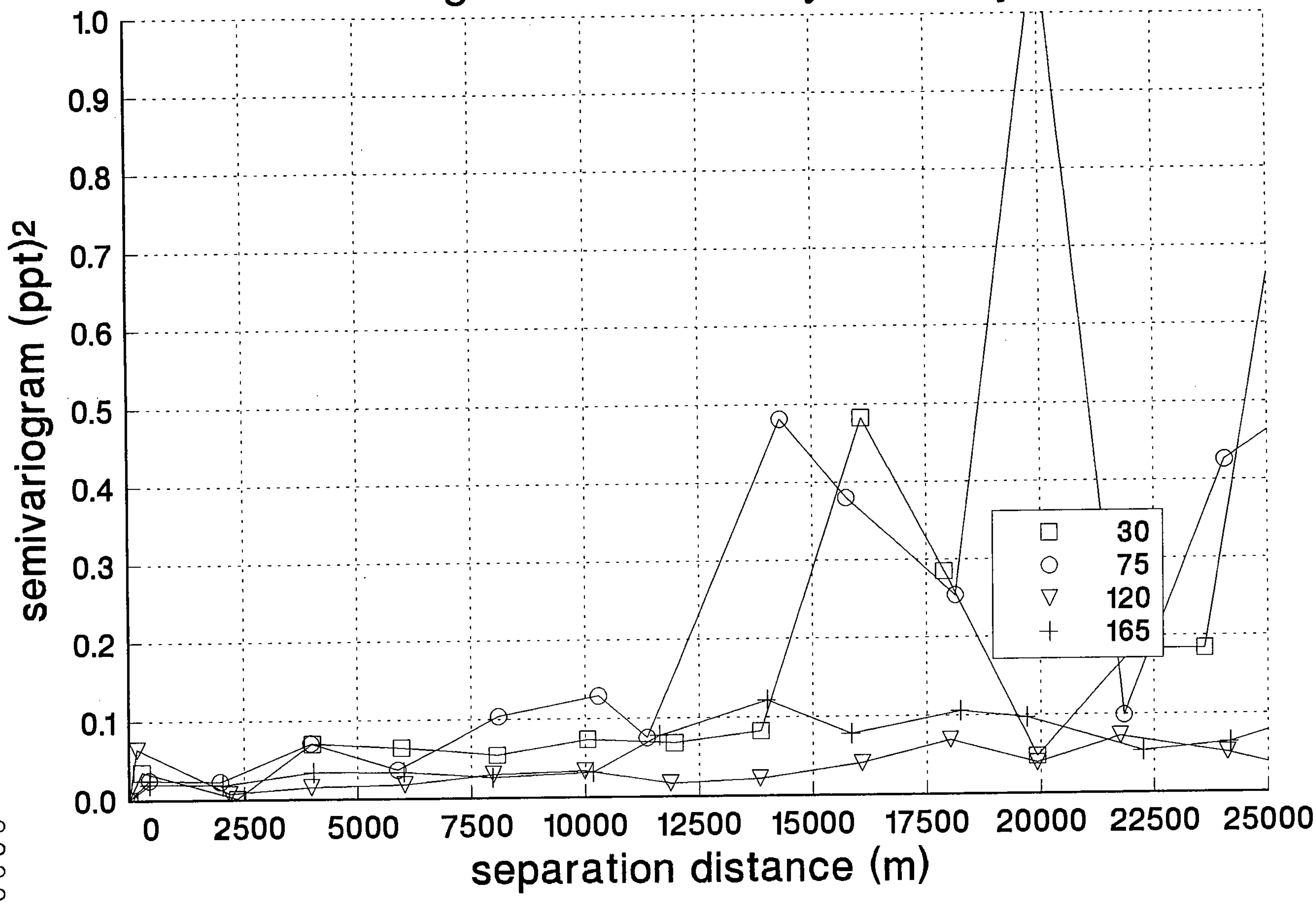

Short Scale - February - Fluorescence
semivariogram (μg/L)^2
0.050
0.045
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
0
400
800
1200
1600
2000
2400
2800
3200
3600
4000
separation distance (m)
75
165

000004

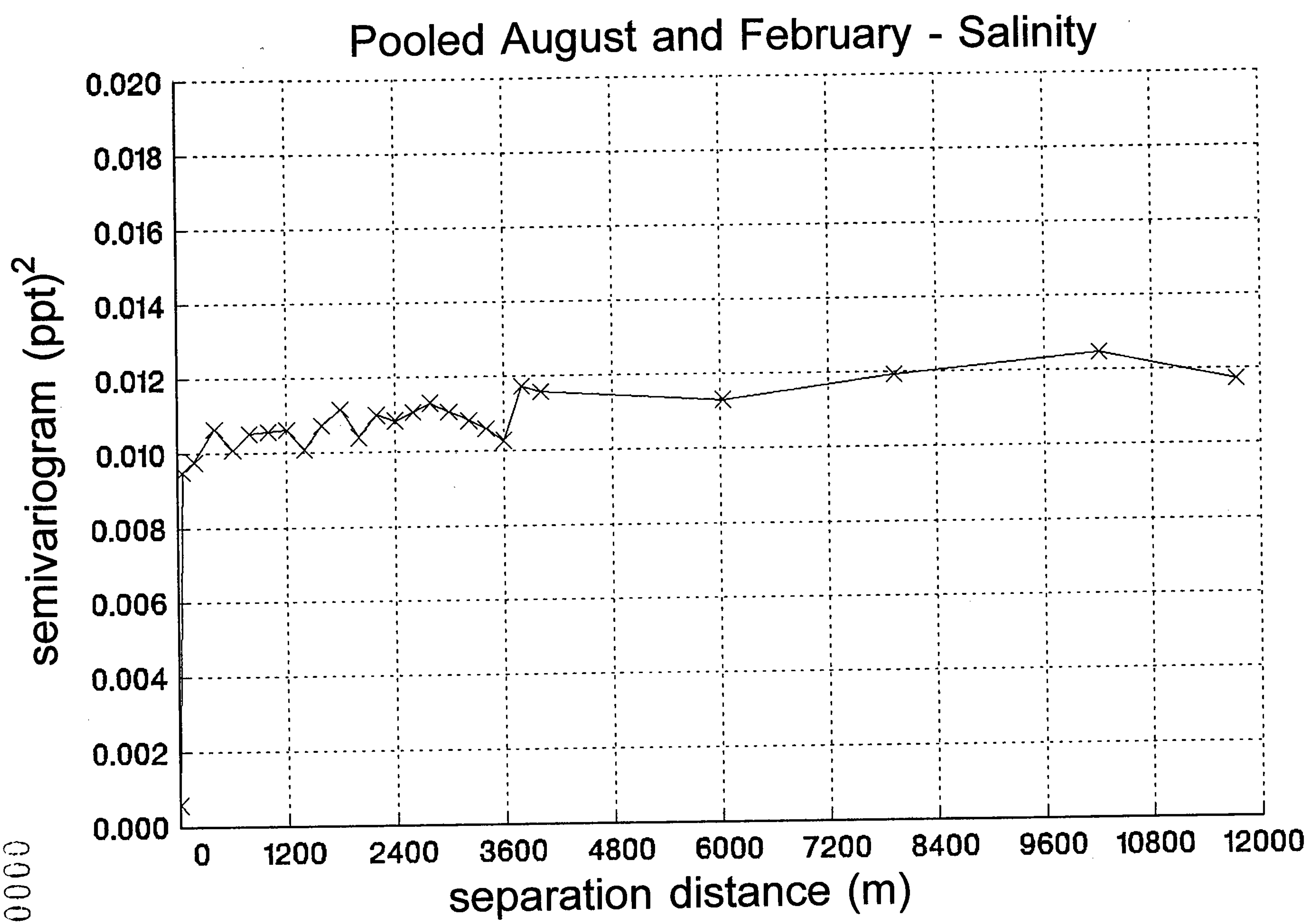
Long Scale - August - Salinity



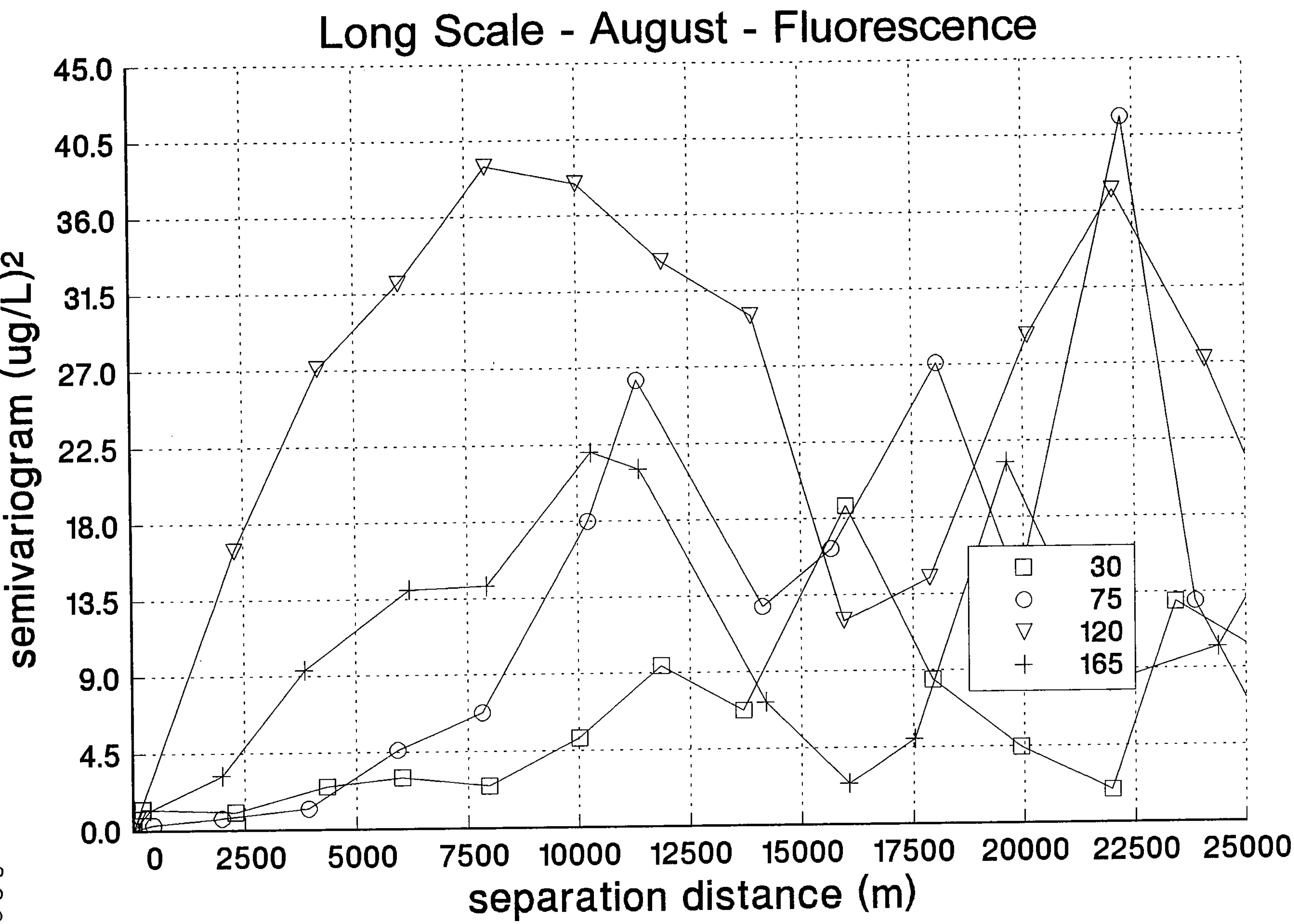
Long Scale - February - Salinity



000006

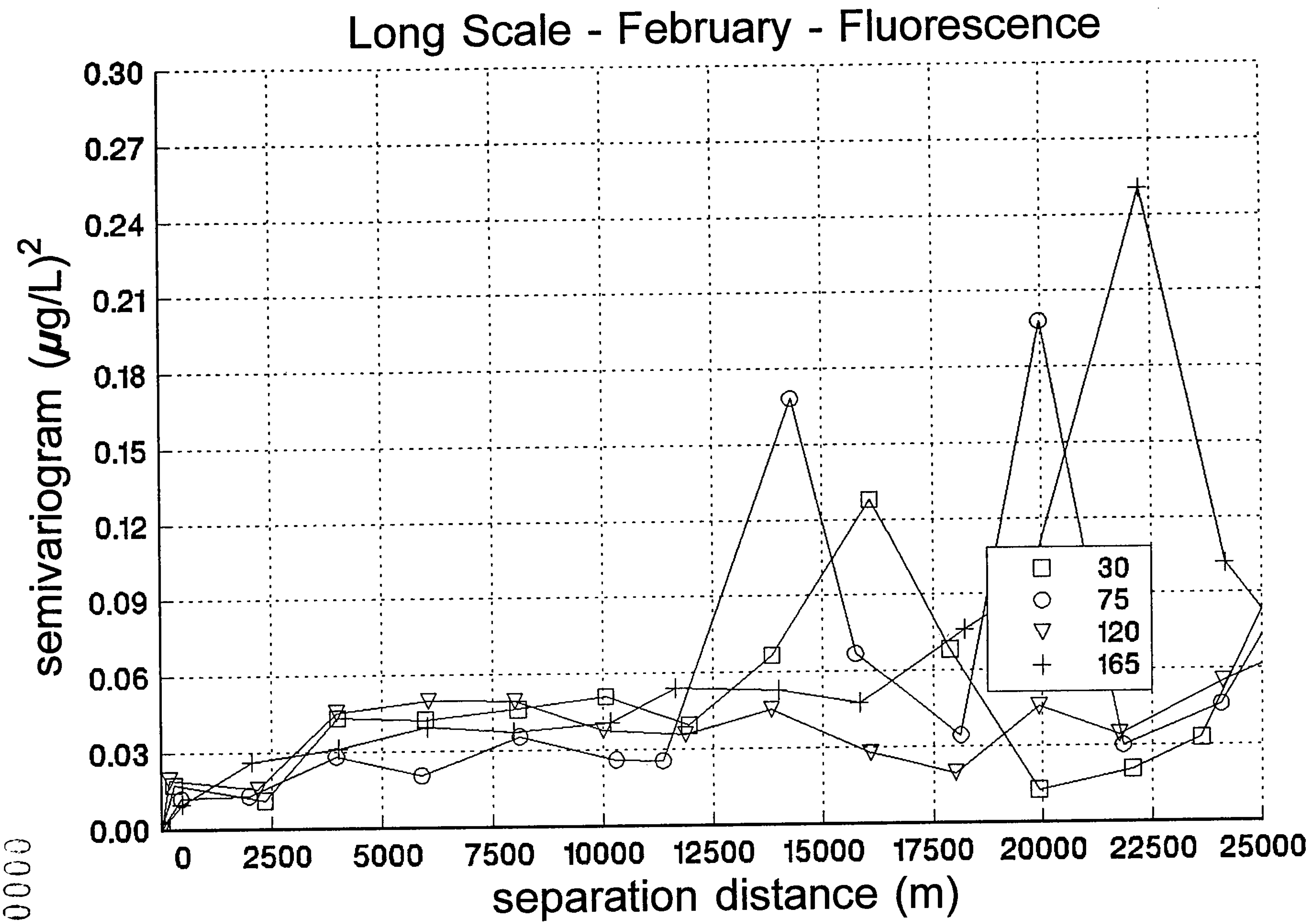

Pooled August and February - Salinity
semivariogram (ppt)²
0.020
0.018
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
1200
2400
3600
4800
6000
7200
8400
9600
10800
12000
separation distance (m)

000007



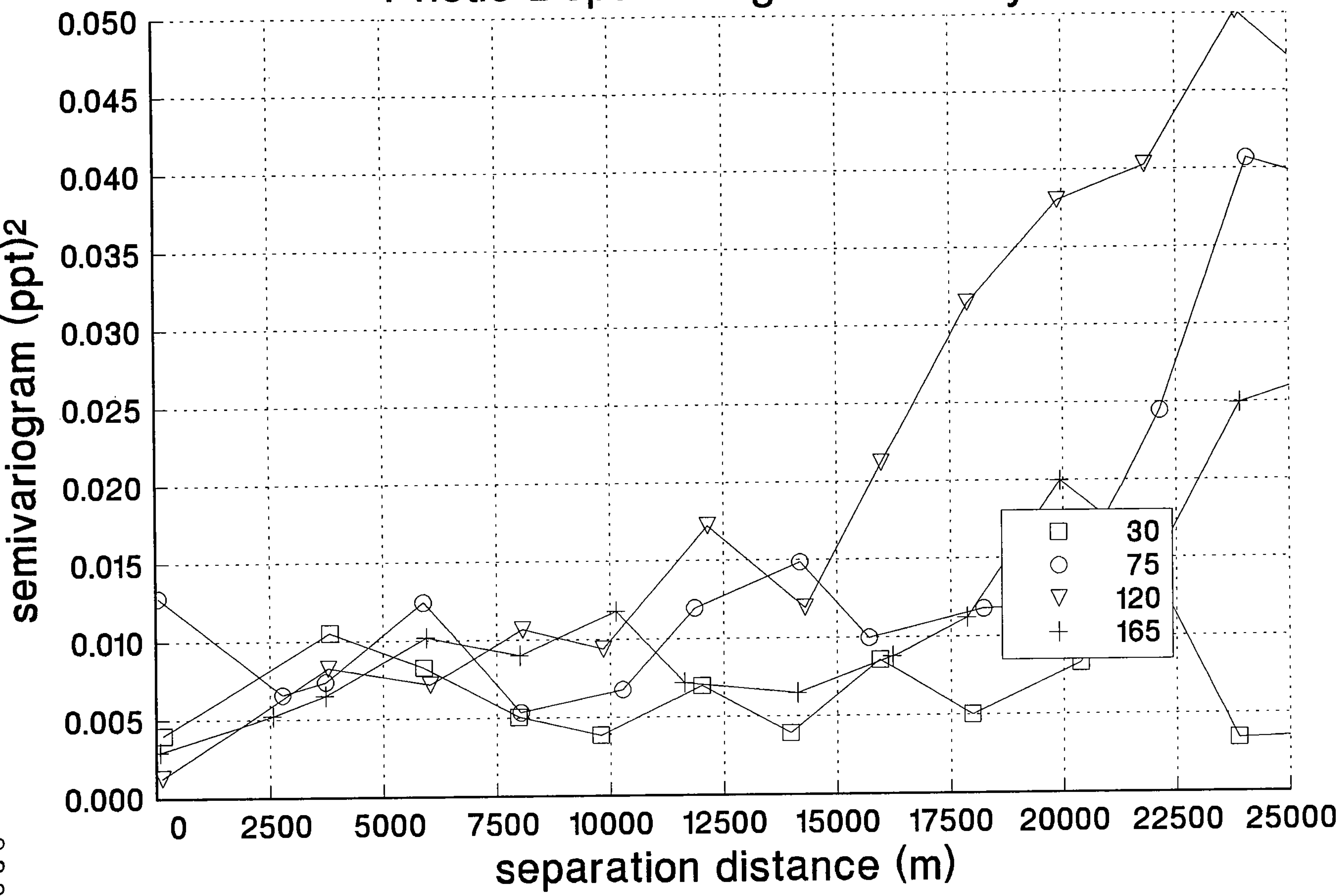
Long Scale - August - Fluorescence
semivariogram (ug/L)2
separation distance (m)
45.0
40.5
36.0
31.5
27.0
22.5
18.0
13.5
9.0
4.5
0.0
0
2500
5000
7500
10000
12500
15000
17500
20000
22500
25000
30
75
120
165

000008

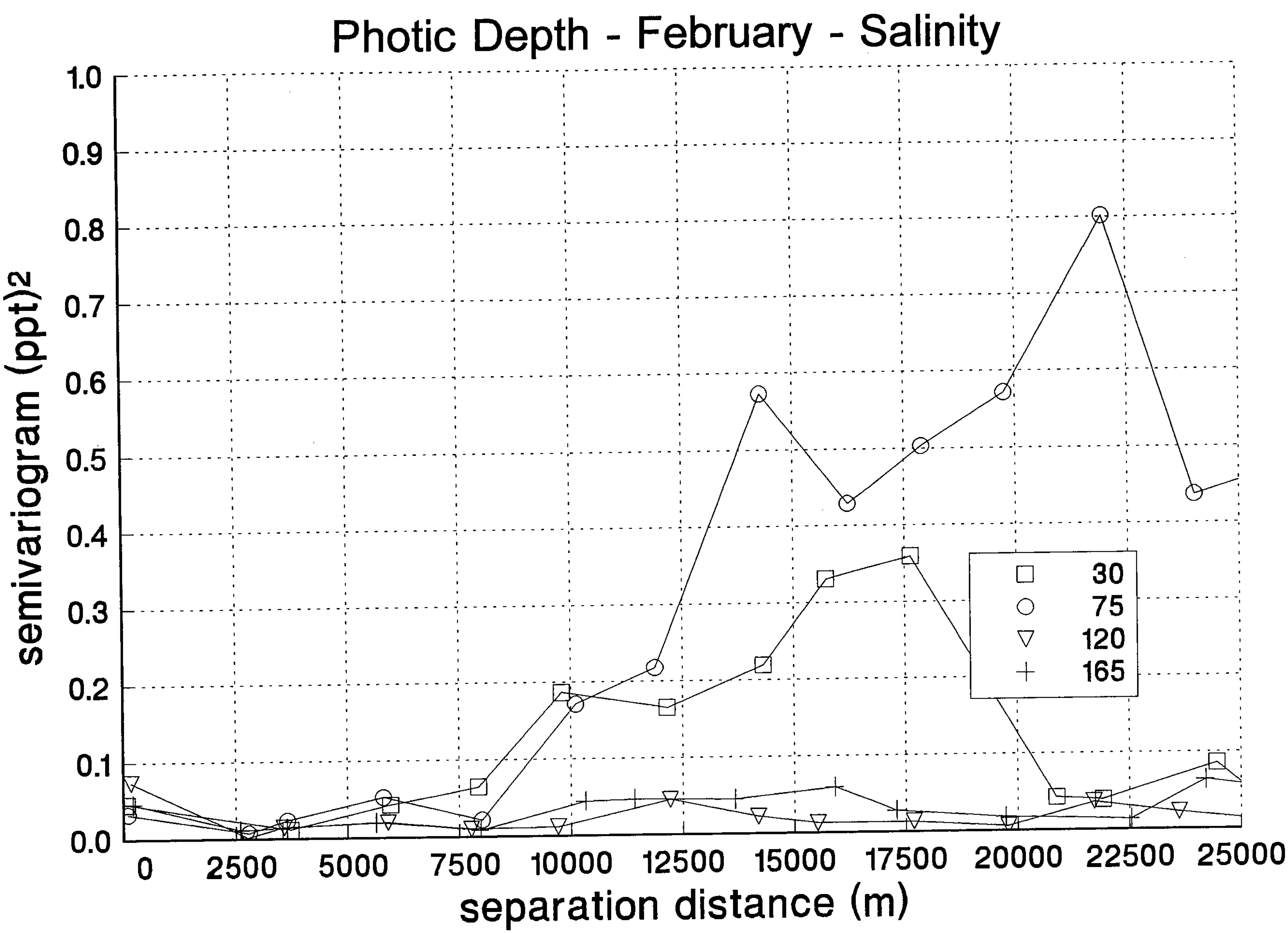

Long Scale - February - Fluorescence
semivariogram (μg/L)^2
0.30
0.27
0.24
0.21
0.18
0.15
0.12
0.09
0.06
0.03
0.00
0
2500
5000
7500
10000
12500
15000
17500
20000
22500
25000
separation distance (m)
30
75
120
165

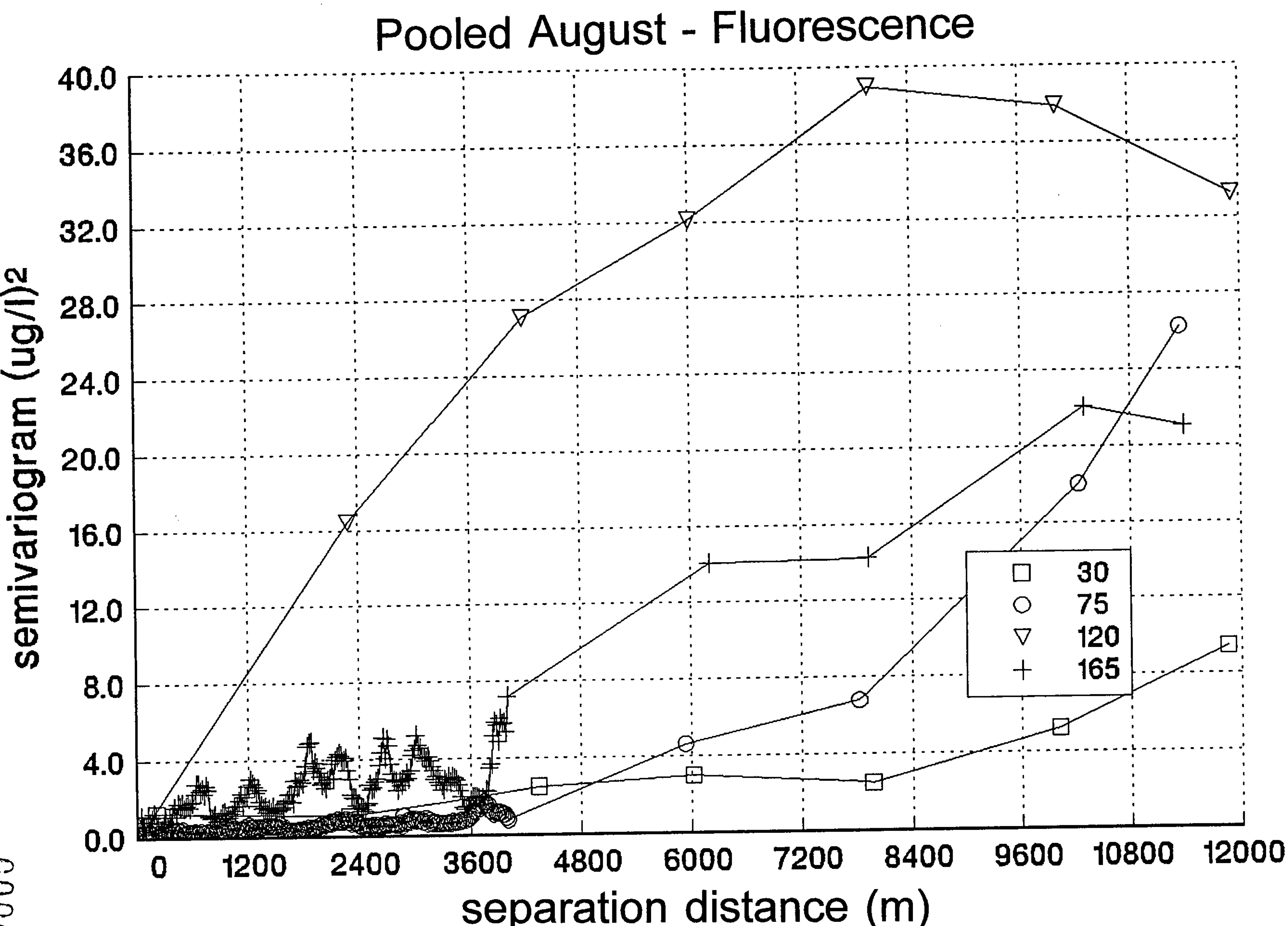
000009

Photic Depth - August - Salinity

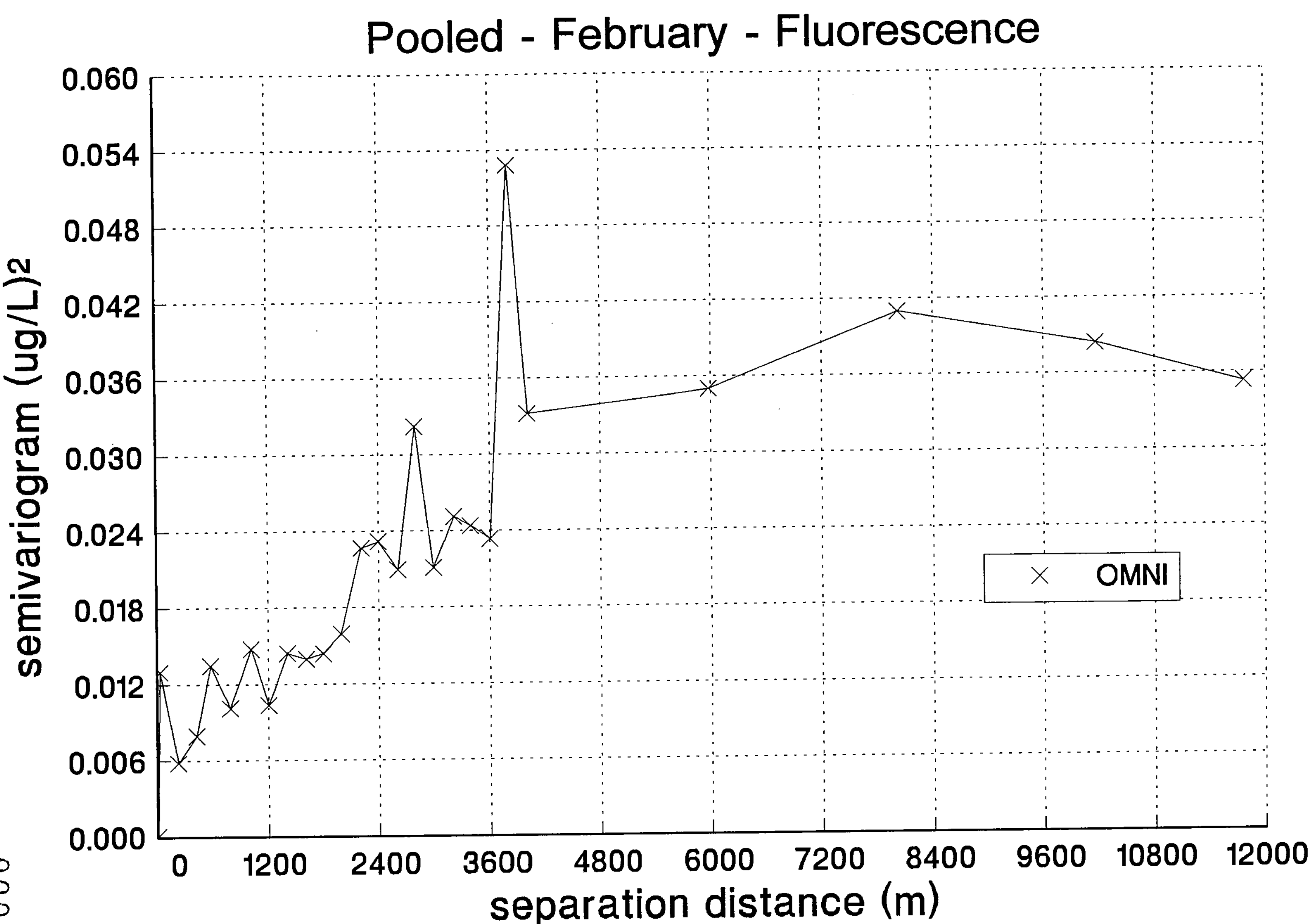


000010

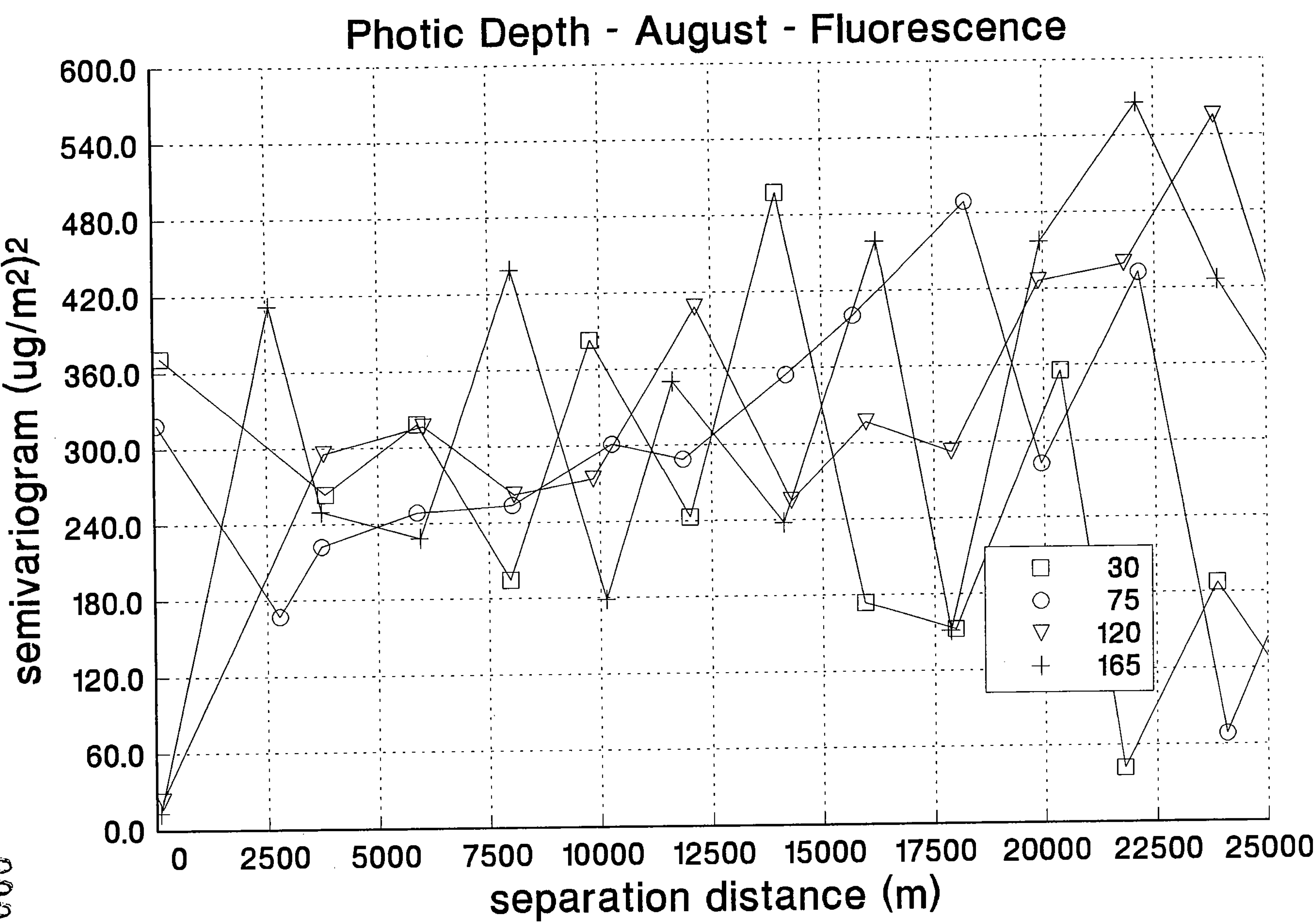

Photic Depth - February - Salinity
semivariogram (ppt)2
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
2500
5000
7500
10000
12500
15000
17500
20000
22500
25000
separation distance (m)
30
75
120
165

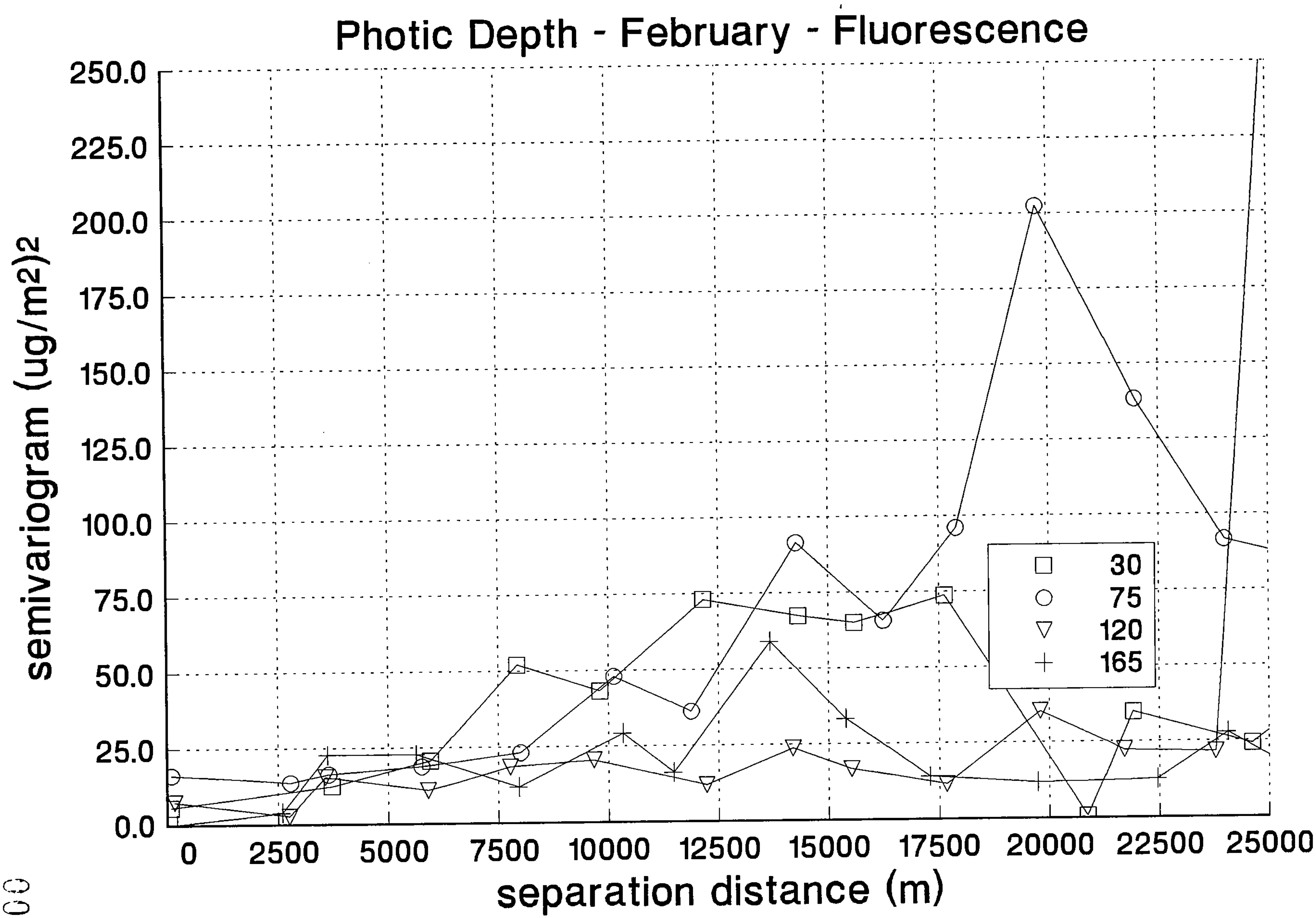

Pooled August - Fluorescence
semivariogram (ug/l)2
separation distance (m)
30
75
120
165
0.0
4.0
8.0
12.0
16.0
20.0
24.0
28.0
32.0
36.0
40.0
0
1200
2400
3600
4800
6000
7200
8400
9600
10800
12000

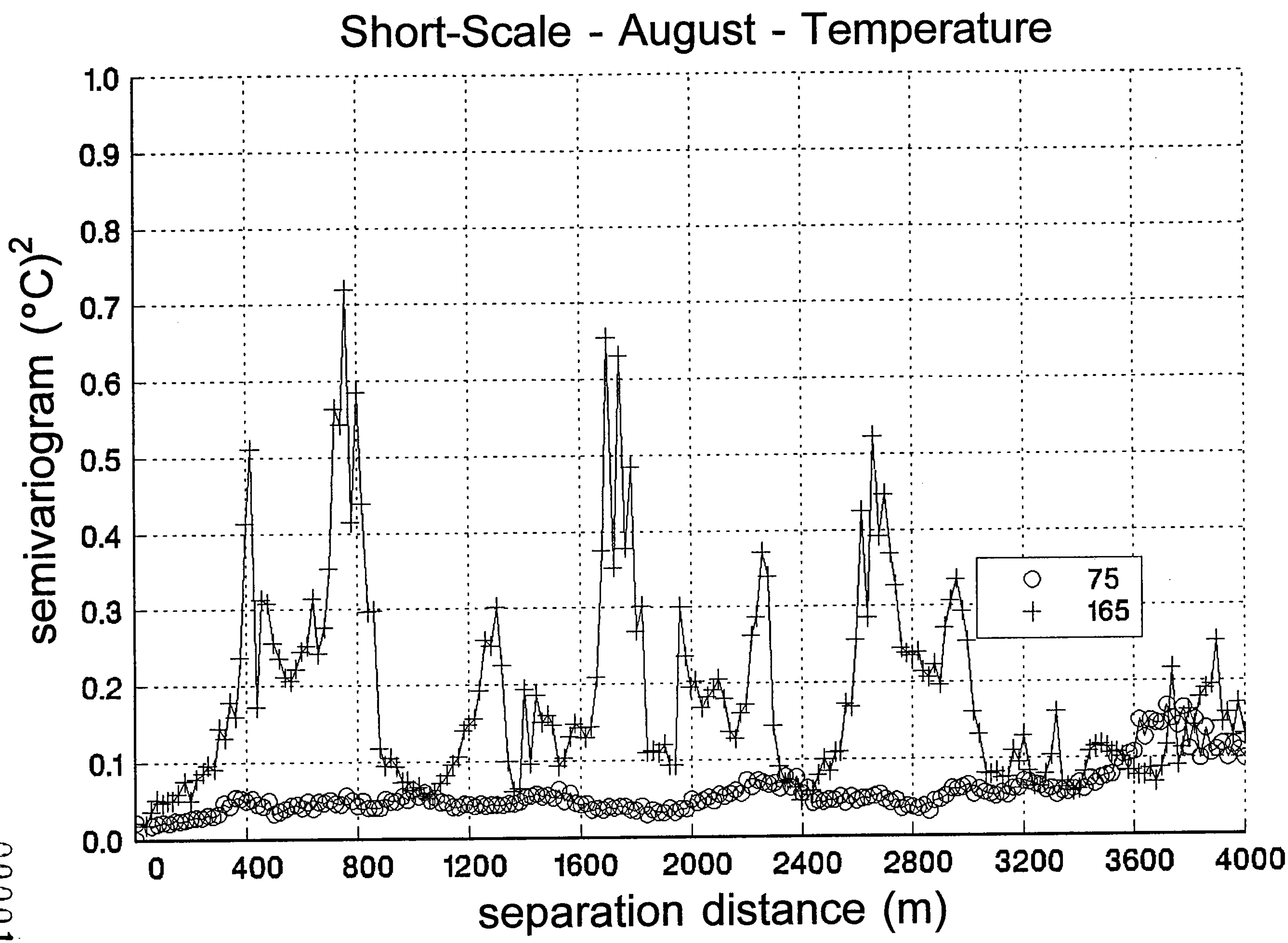
000012



Pooled - February - Fluorescence
semivariogram (ug/L)2
0.060
0.054
0.048
0.042
0.036
0.030
0.024
0.018
0.012
0.006
0.000
0
1200
2400
3600
4800
6000
7200
8400
9600
10800
12000
separation distance (m)
OMNI

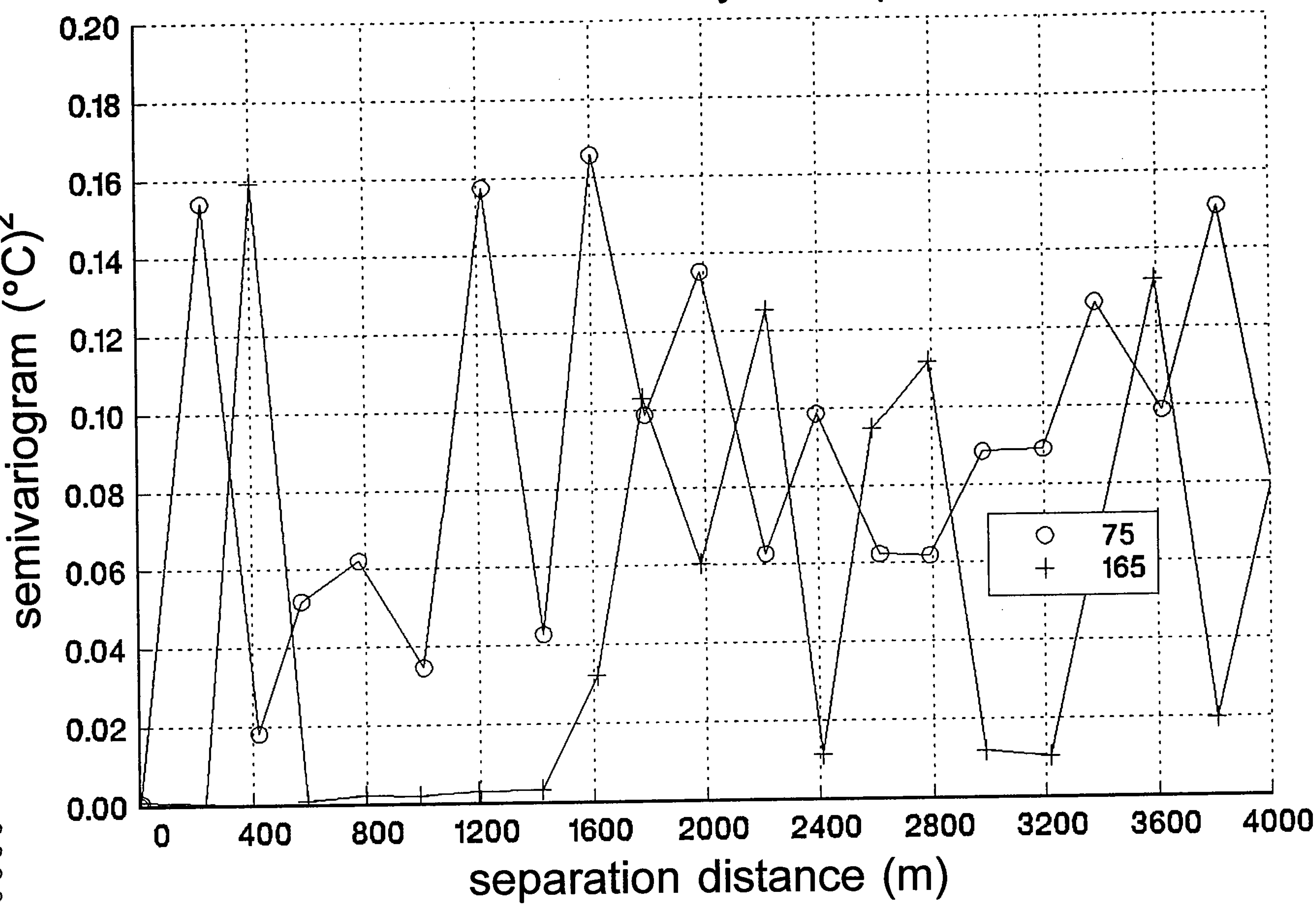

Photic Depth - August - Fluorescence
semivariogram (ug/m2)2
600.0
540.0
480.0
420.0
360.0
300.0
240.0
180.0
120.0
60.0
0.0
0
2500
5000
7500
10000
12500
15000
17500
20000
22500
25000
separation distance (m)
30
75
120
165

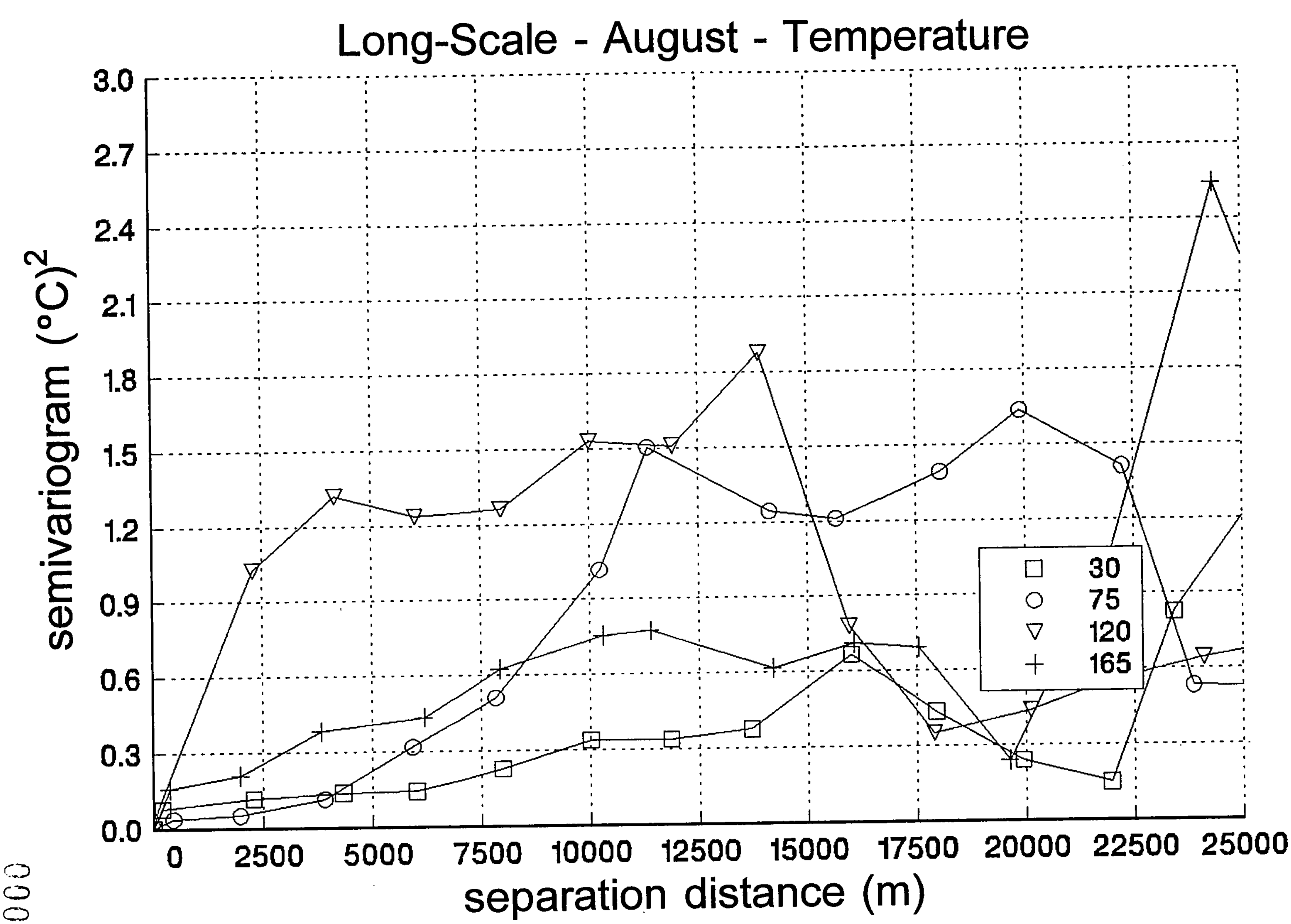

Photic Depth - February - Fluorescence
semivariogram (ug/m2)2
250.0
225.0
200.0
175.0
150.0
125.0
100.0
75.0
50.0
25.0
0.0
0
2500
5000
7500
10000
12500
15000
17500
20000
22500
25000
separation distance (m)
30
75
120
165


Short-Scale - August - Temperature
semivariogram (°C)²
separation distance (m)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
400
800
1200
1600
2000
2400
2800
3200
3600
4000
75
165

000016

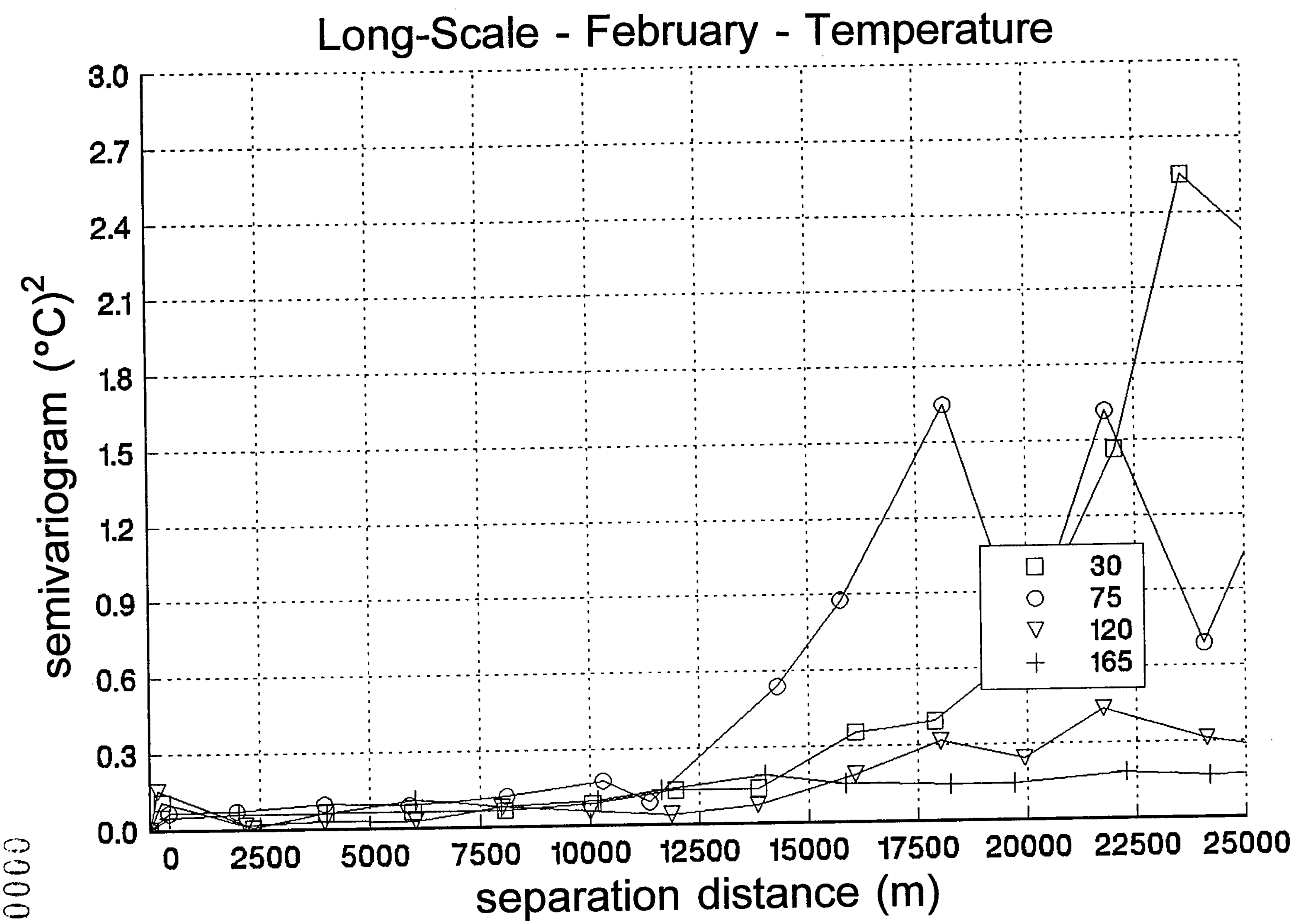
Short-Scale - February - Temperature





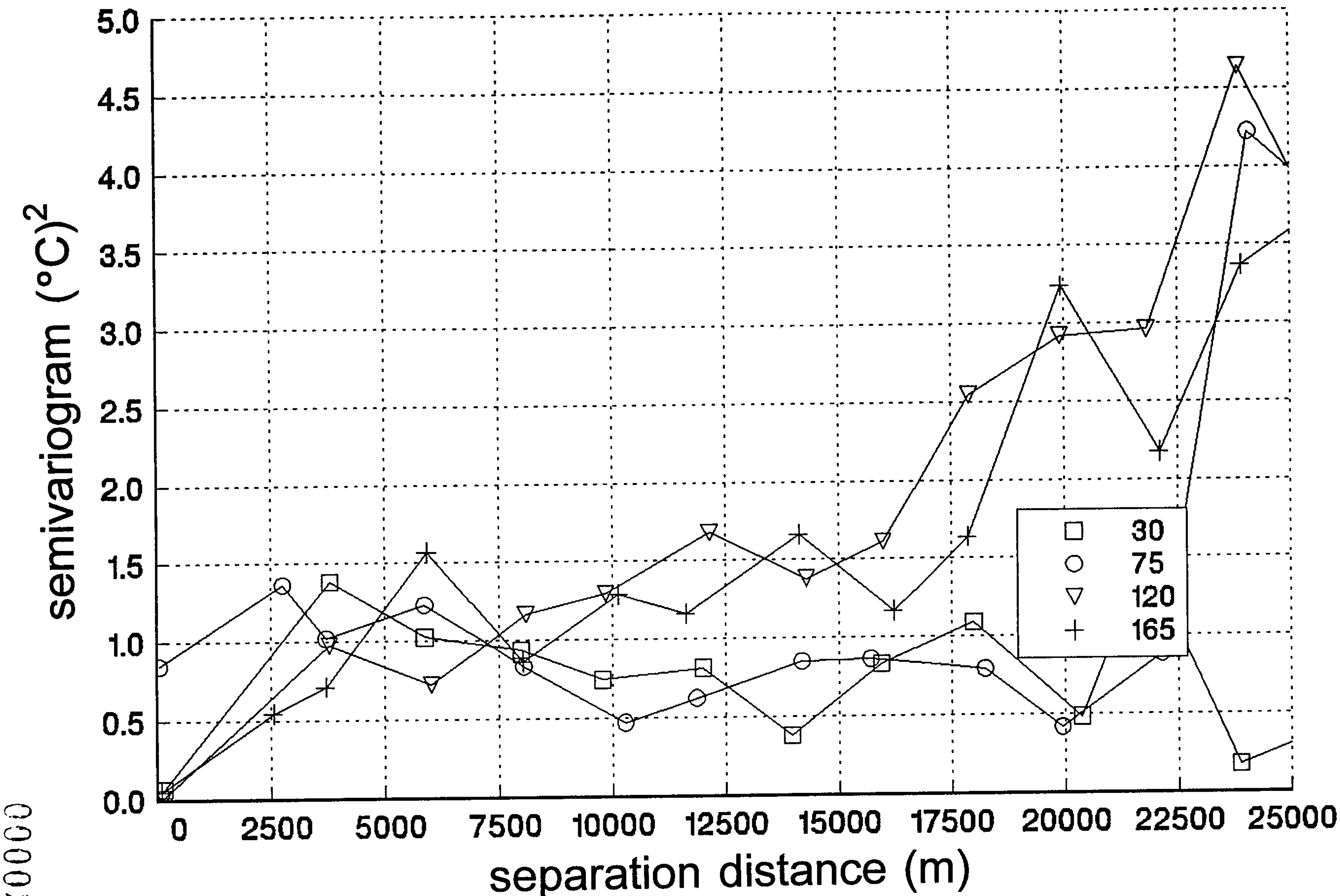
Long-Scale - August - Temperature
semivariogram (°C)^2
3.0
2.7
2.4
2.1
1.8
1.5
1.2
0.9
0.6
0.3
0.0
0
2500
5000
7500
10000
12500
15000
17500
20000
22500
25000
separation distance (m)
30
75
120
165

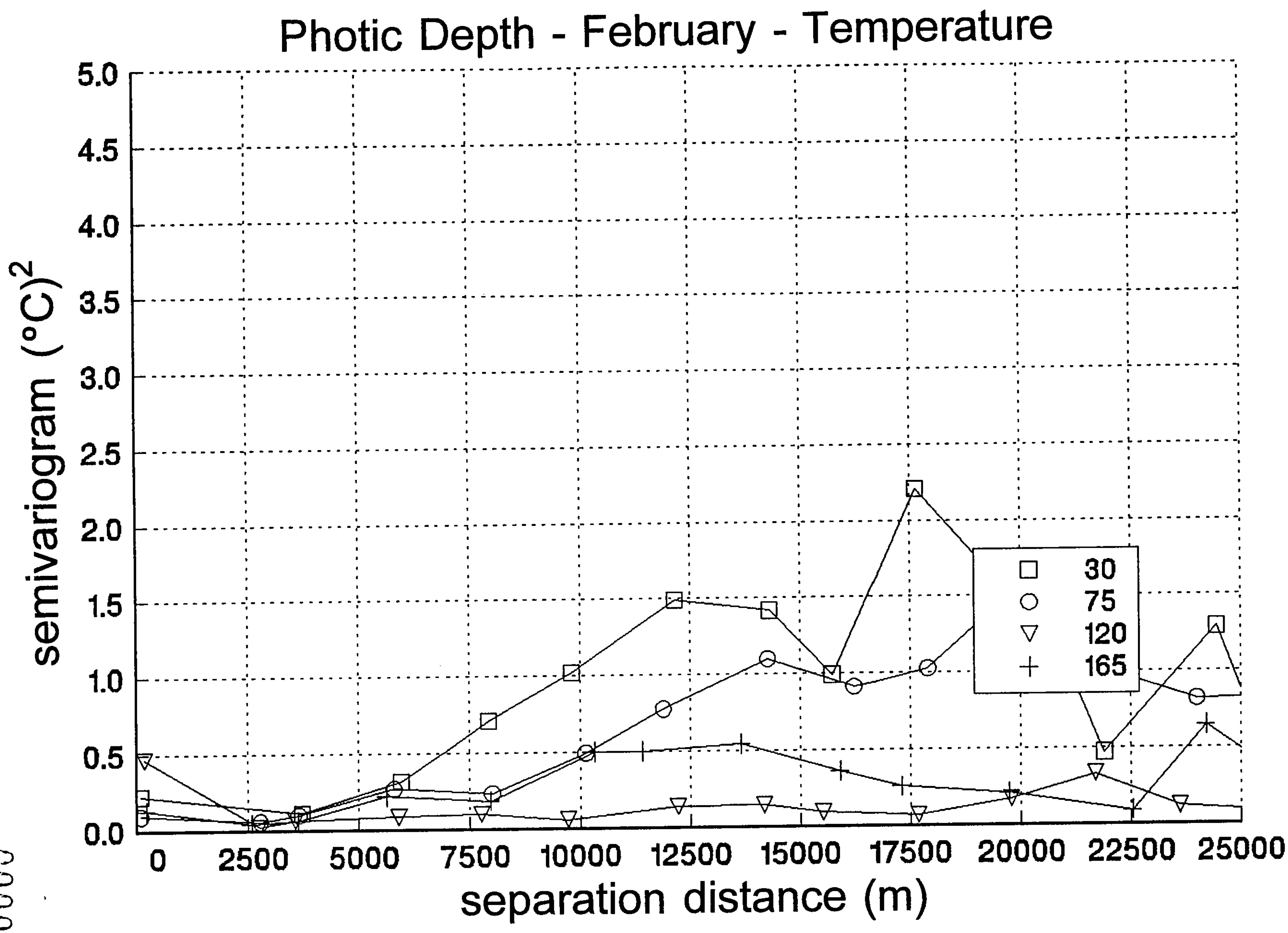
000018


Long-Scale - February - Temperature
semivariogram (°C)²
3.0
2.7
2.4
2.1
1.8
1.5
1.2
0.9
0.6
0.3
0.0
0
2500
5000
7500
10000
12500
15000
17500
20000
22500
25000
separation distance (m)
30
75
120
165

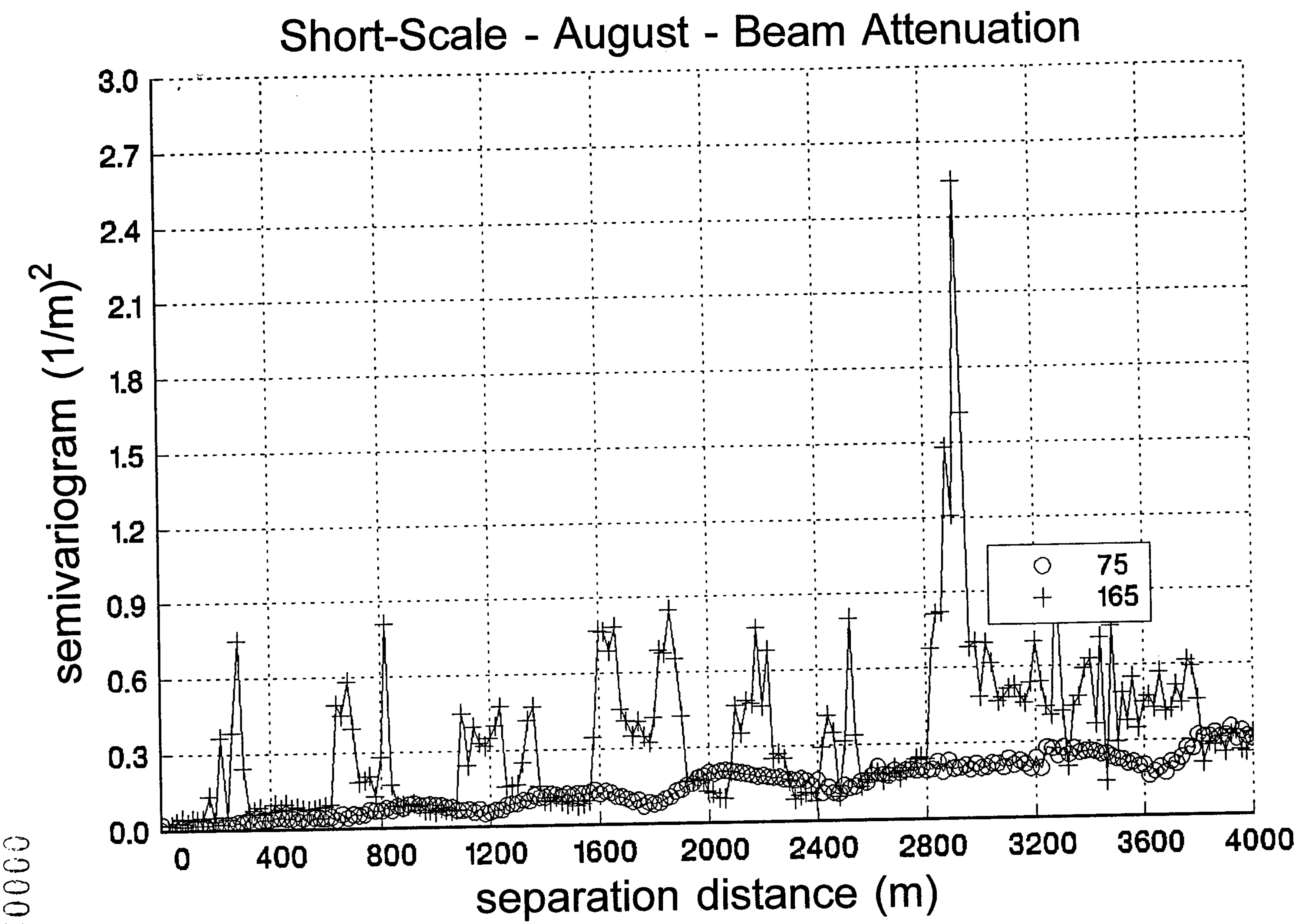
000019

Photic Depth - August - Temperature

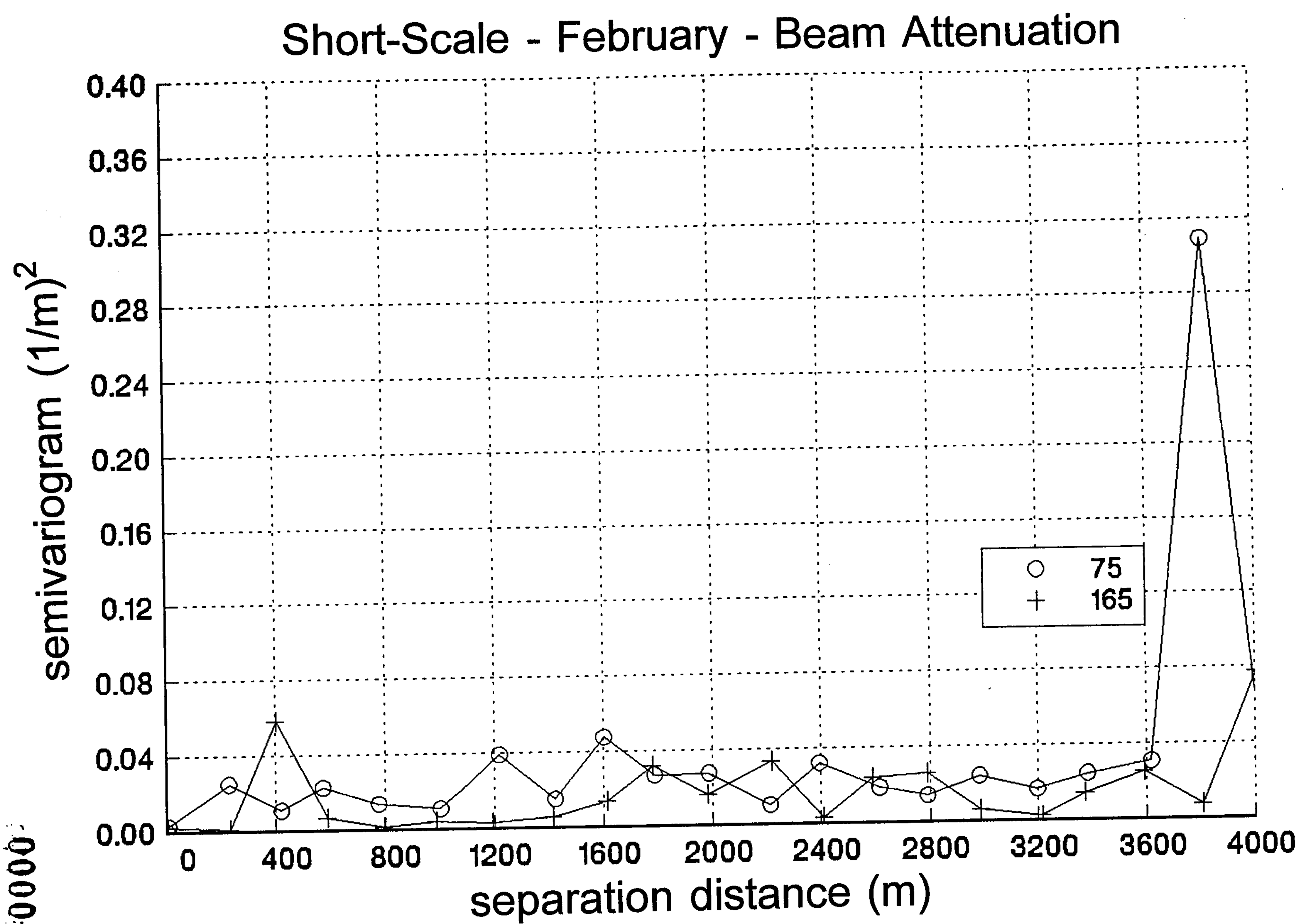



Photic Depth - February - Temperature
semivariogram $(°C)^2$
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
2500
5000
7500
10000
12500
15000
17500
20000
22500
25000
separation distance (m)
30
75
120
165

000021

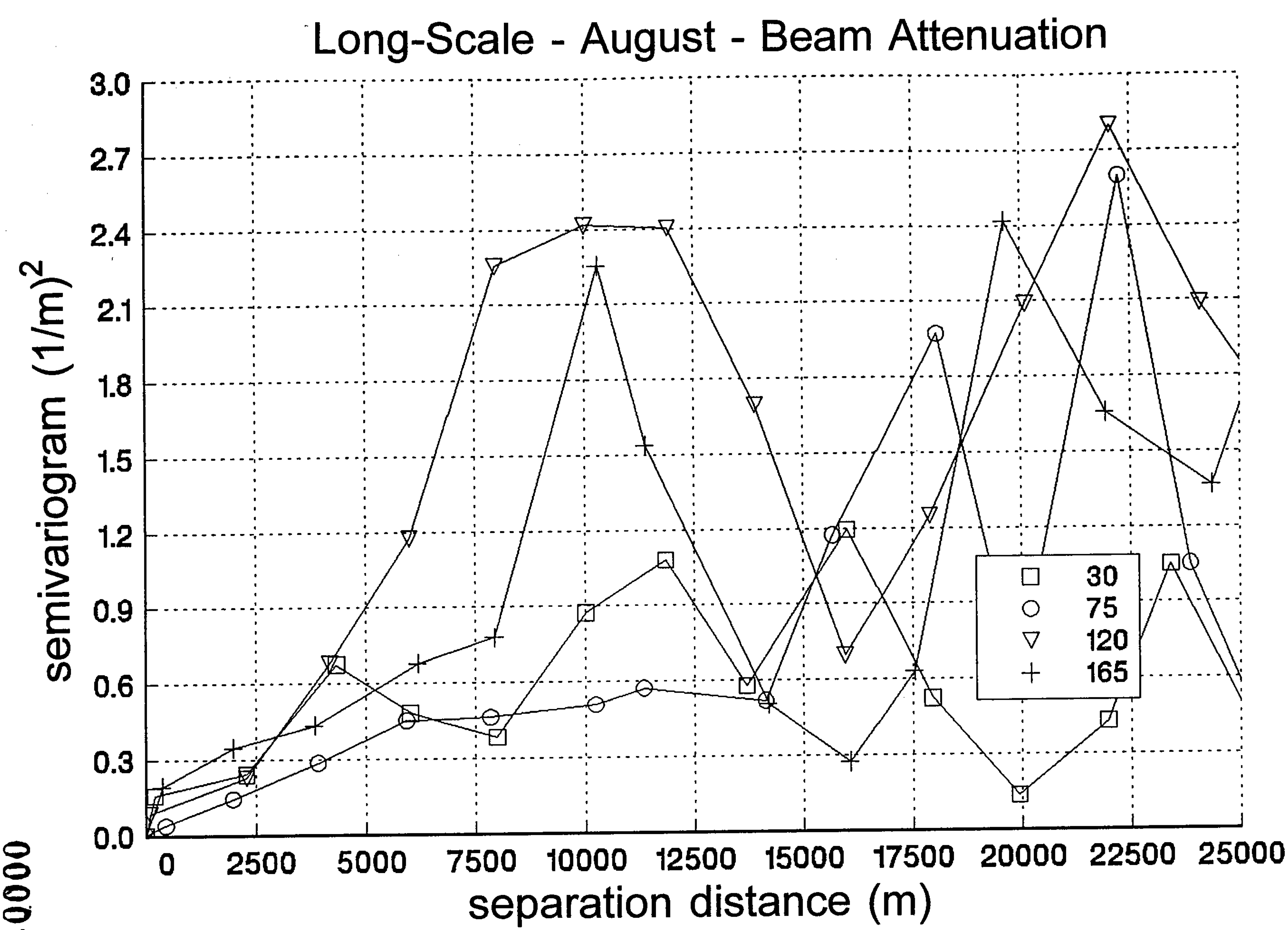

Short-Scale - August - Beam Attenuation
semivariogram $(1/m)^2$
3.0
2.7
2.4
2.1
1.8
1.5
1.2
0.9
0.6
0.3
0.0
0
400
800
1200
1600
2000
2400
2800
3200
3600
4000
separation distance (m)
75
165

000022



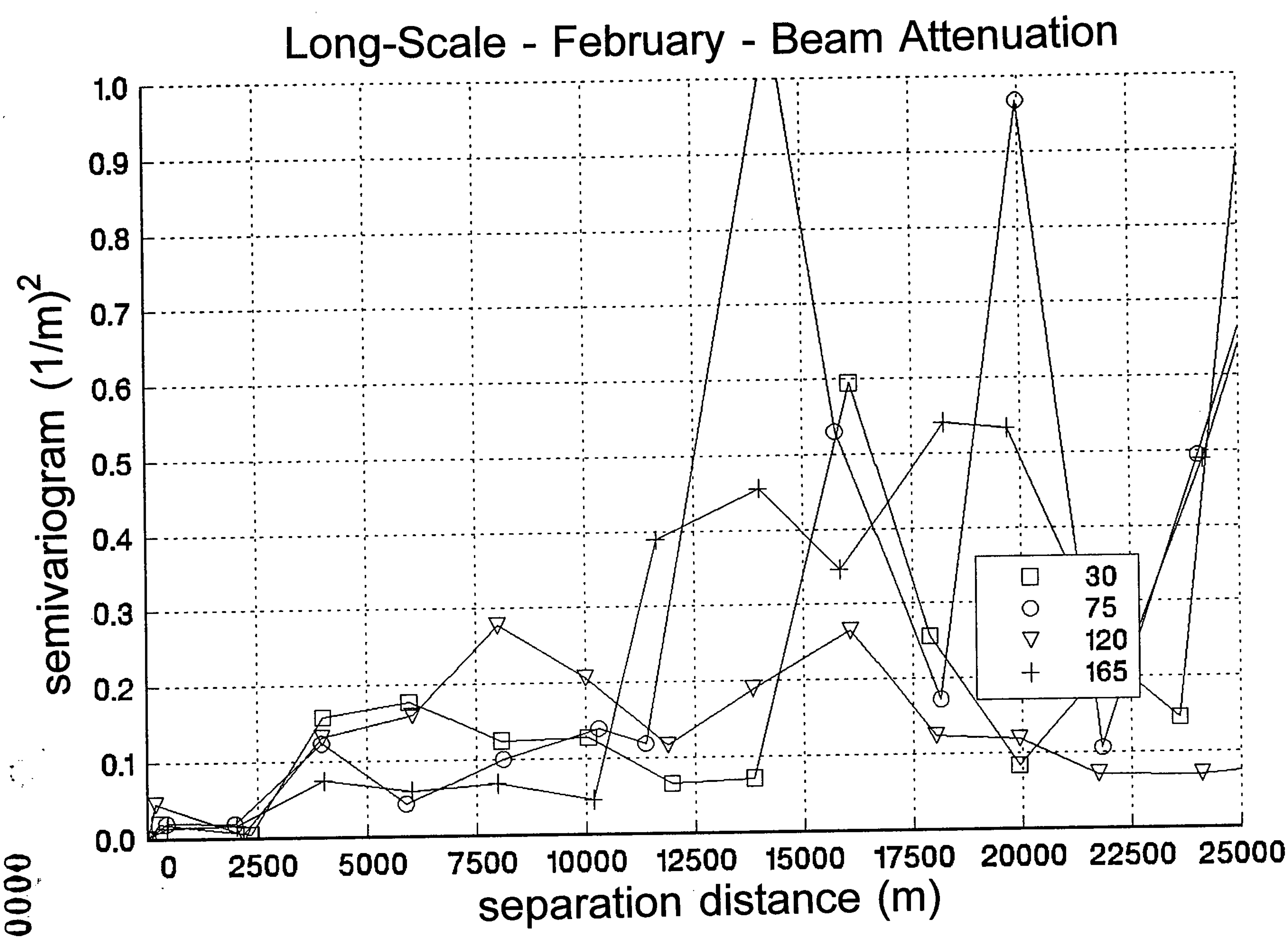
Short-Scale - February - Beam Attenuation
semivariogram (1/m)^2
0.40
0.36
0.32
0.28
0.24
0.20
0.16
0.12
0.08
0.04
0.00
0
400
800
1200
1600
2000
2400
2800
3200
3600
4000
separation distance (m)
75
165

000023



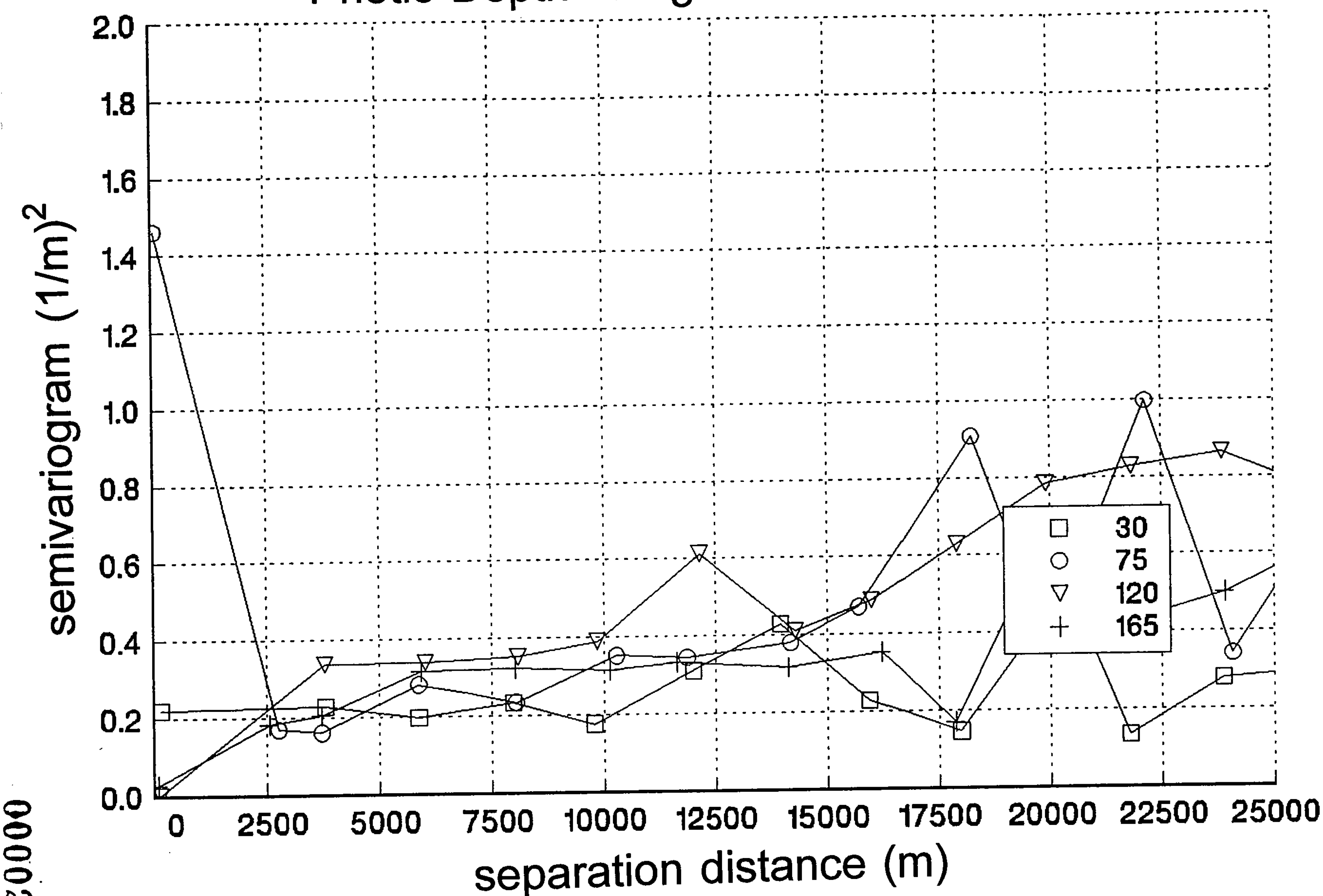
Long-Scale - August - Beam Attenuation
semivariogram $(1/m)^2$
separation distance (m)
3.0
2.7
2.4
2.1
1.8
1.5
1.2
0.9
0.6
0.3
0.0
0
2500
5000
7500
10000
12500
15000
17500
20000
22500
25000
30
75
120
165

000024

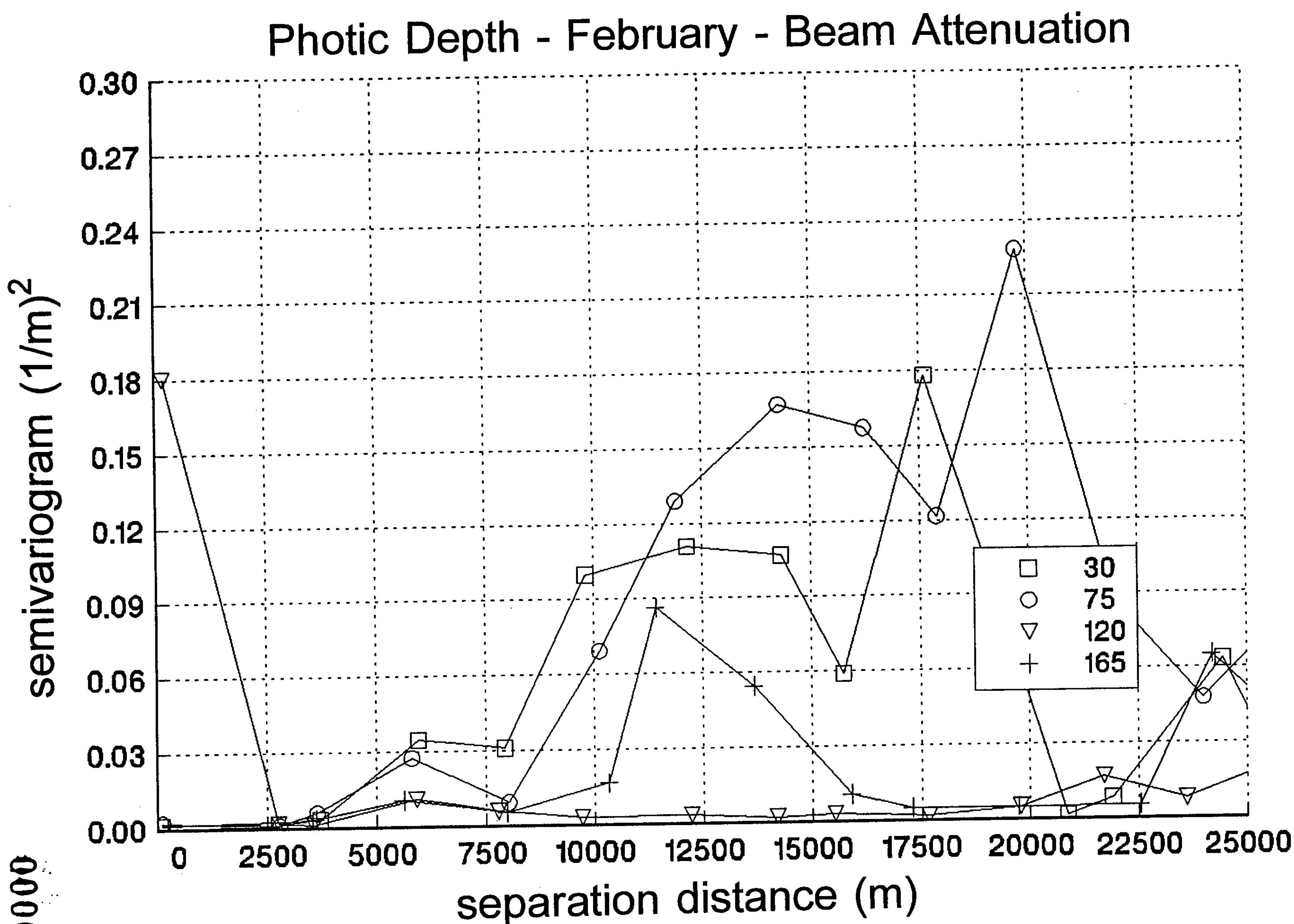


Long-Scale - February - Beam Attenuation
semivariogram (1/m)^2
separation distance (m)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
2500
5000
7500
10000
12500
15000
17500
20000
22500
25000
30
75
120
165

Photic Depth - August - Beam Attenuation

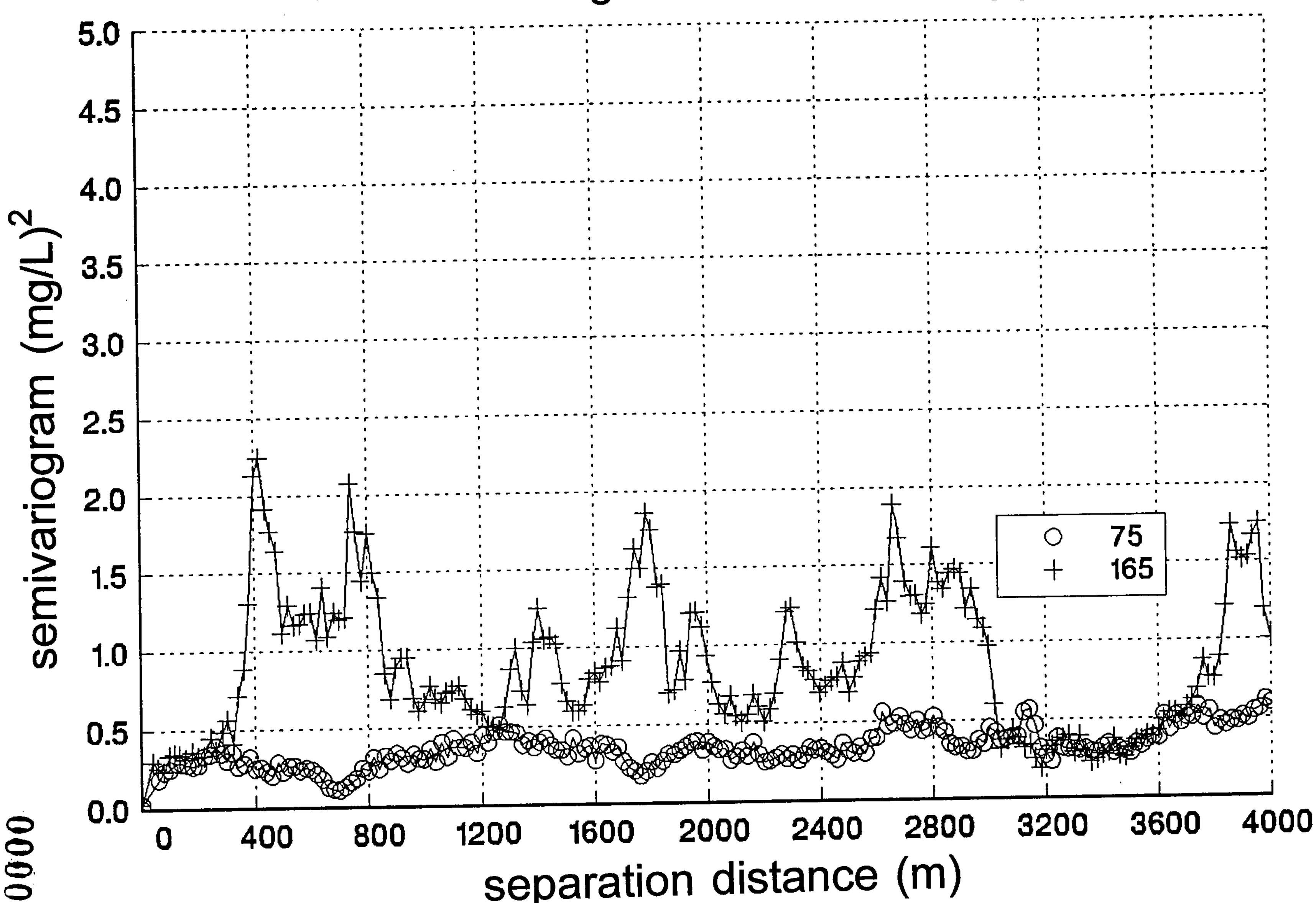


000026

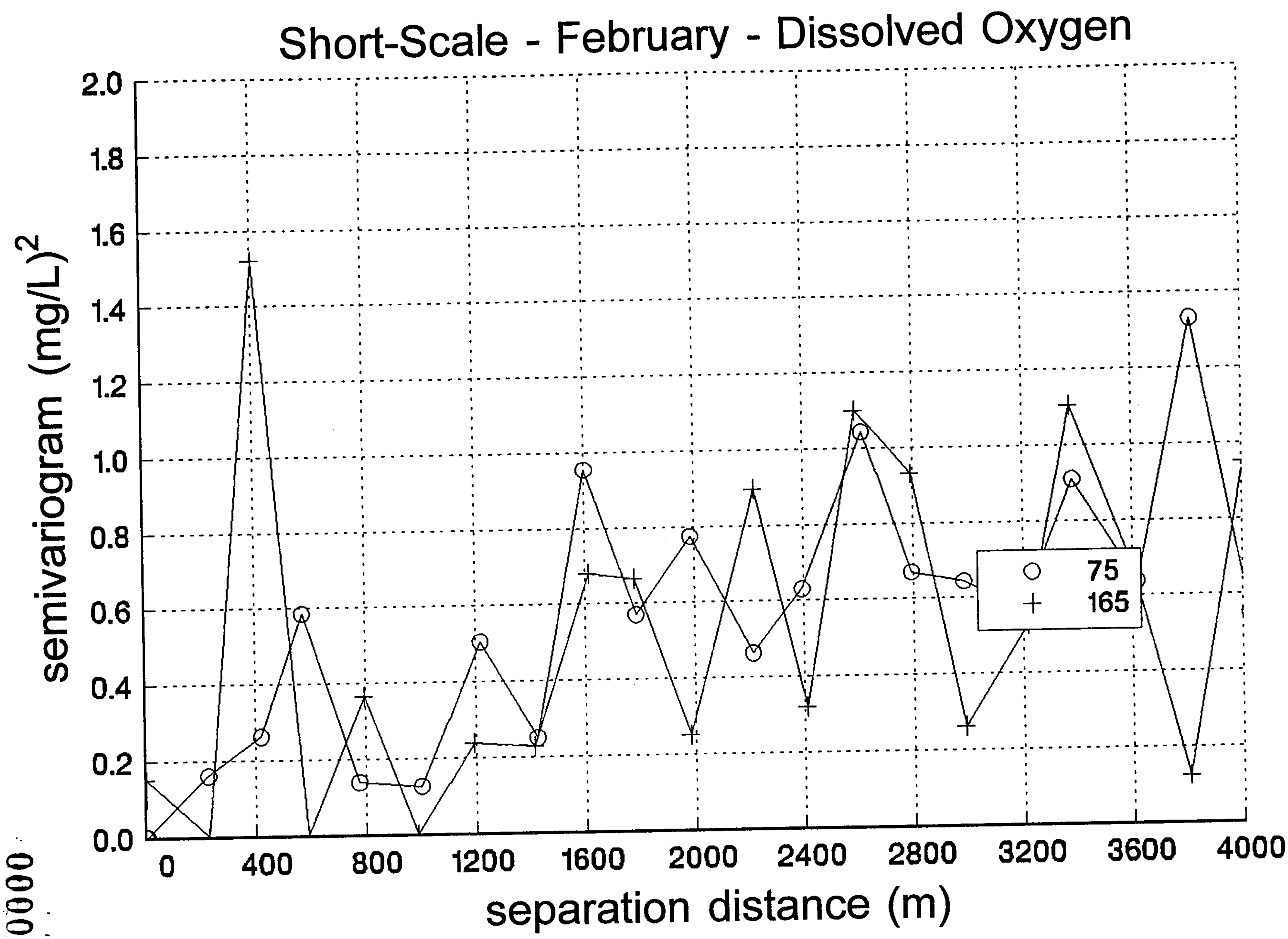

Photic Depth - February - Beam Attenuation
semivariogram (1/m)^2
0.30
0.27
0.24
0.21
0.18
0.15
0.12
0.09
0.06
0.03
0.00
0
2500
5000
7500
10000
12500
15000
17500
20000
22500
25000
separation distance (m)
30
75
120
165

000027

Short-Scale - August - Dissolved Oxygen

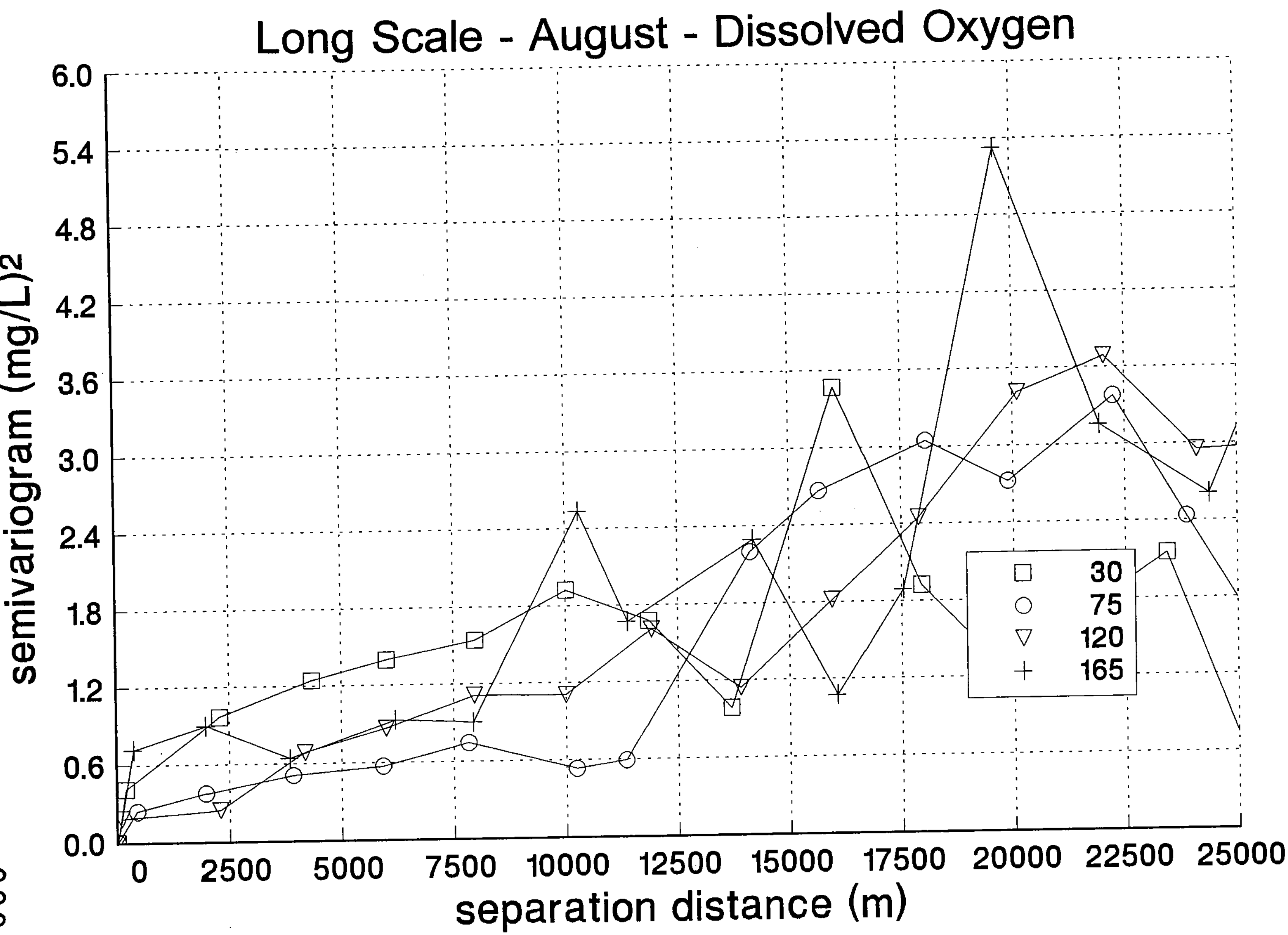


000028

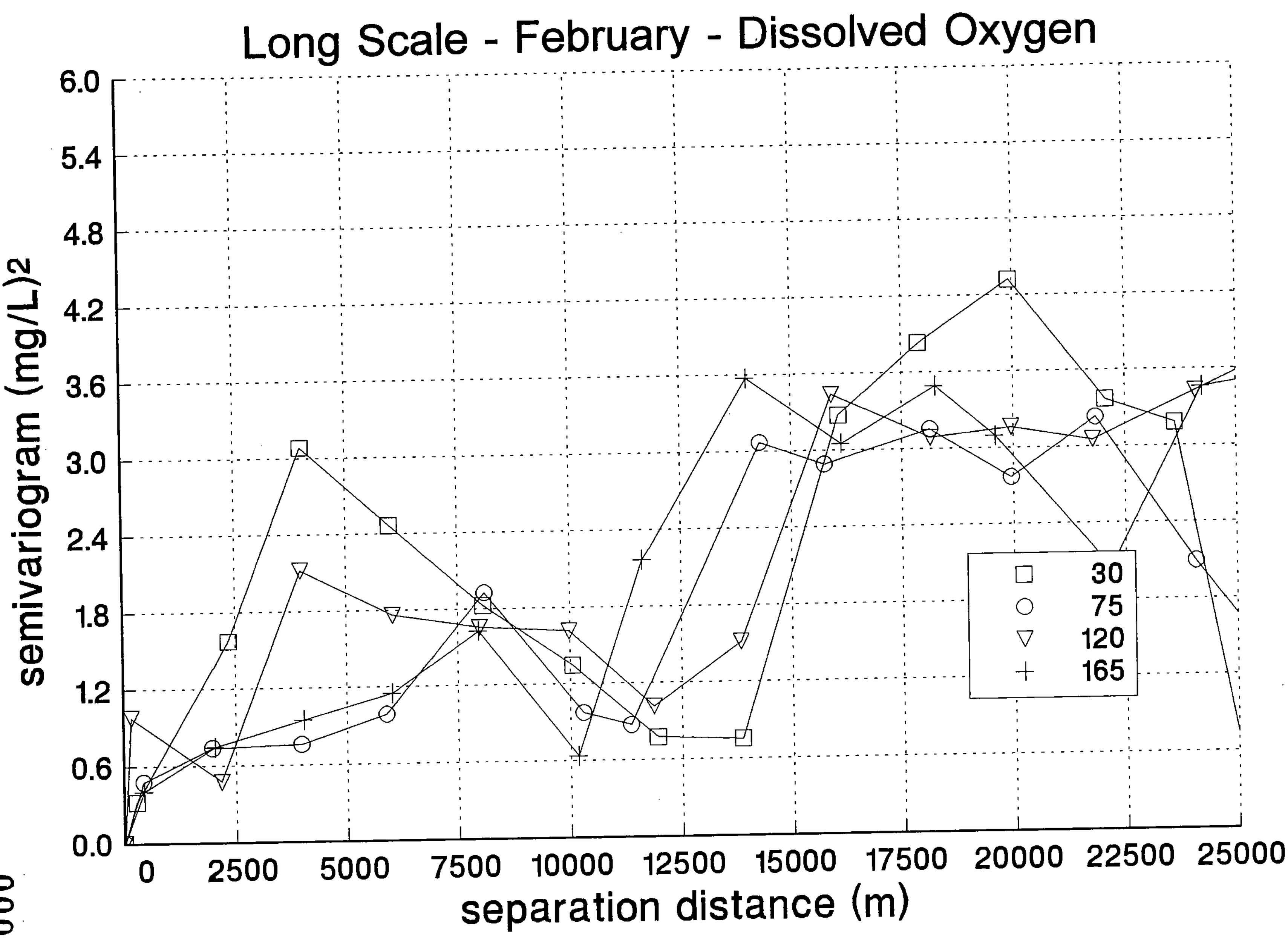


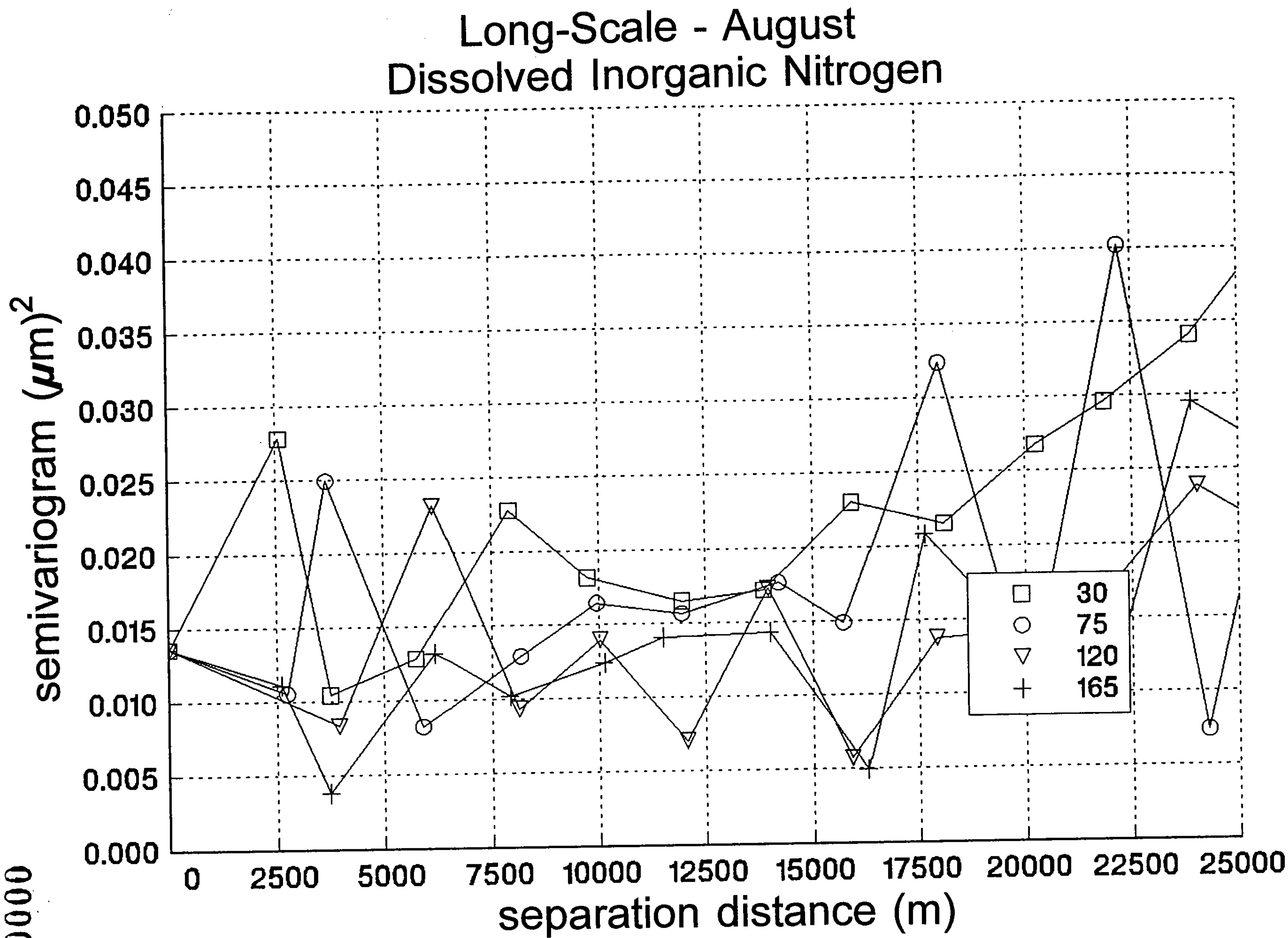
Short-Scale - February - Dissolved Oxygen
semivariogram (mg/L)^2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
400
800
1200
1600
2000
2400
2800
3200
3600
4000
separation distance (m)
75
165

000029



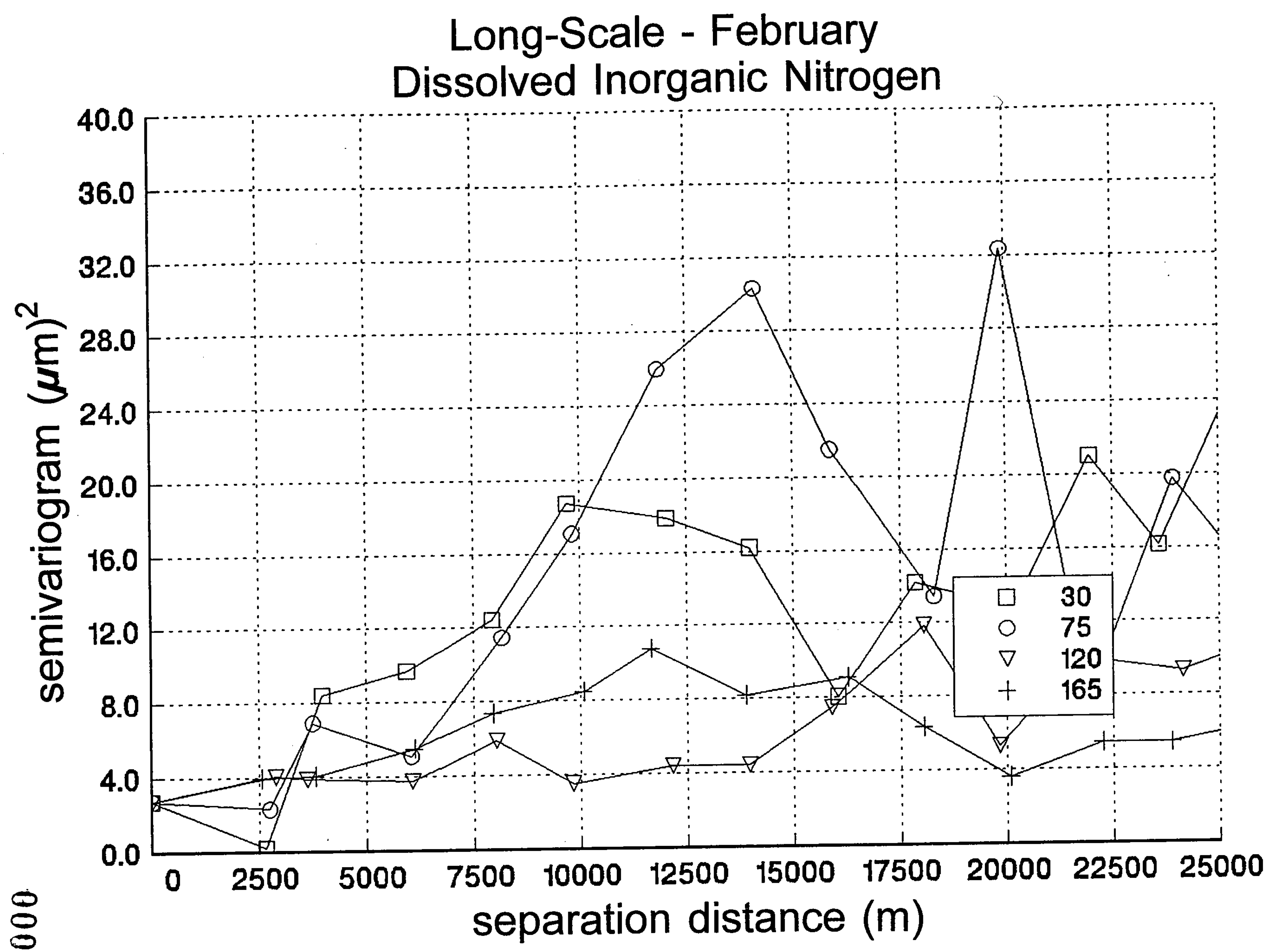
Long Scale - August - Dissolved Oxygen
semivariogram (mg/L)2
0.0
0.6
1.2
1.8
2.4
3.0
3.6
4.2
4.8
5.4
6.0
0
2500
5000
7500
10000
12500
15000
17500
20000
22500
25000
separation distance (m)
30
75
120
165

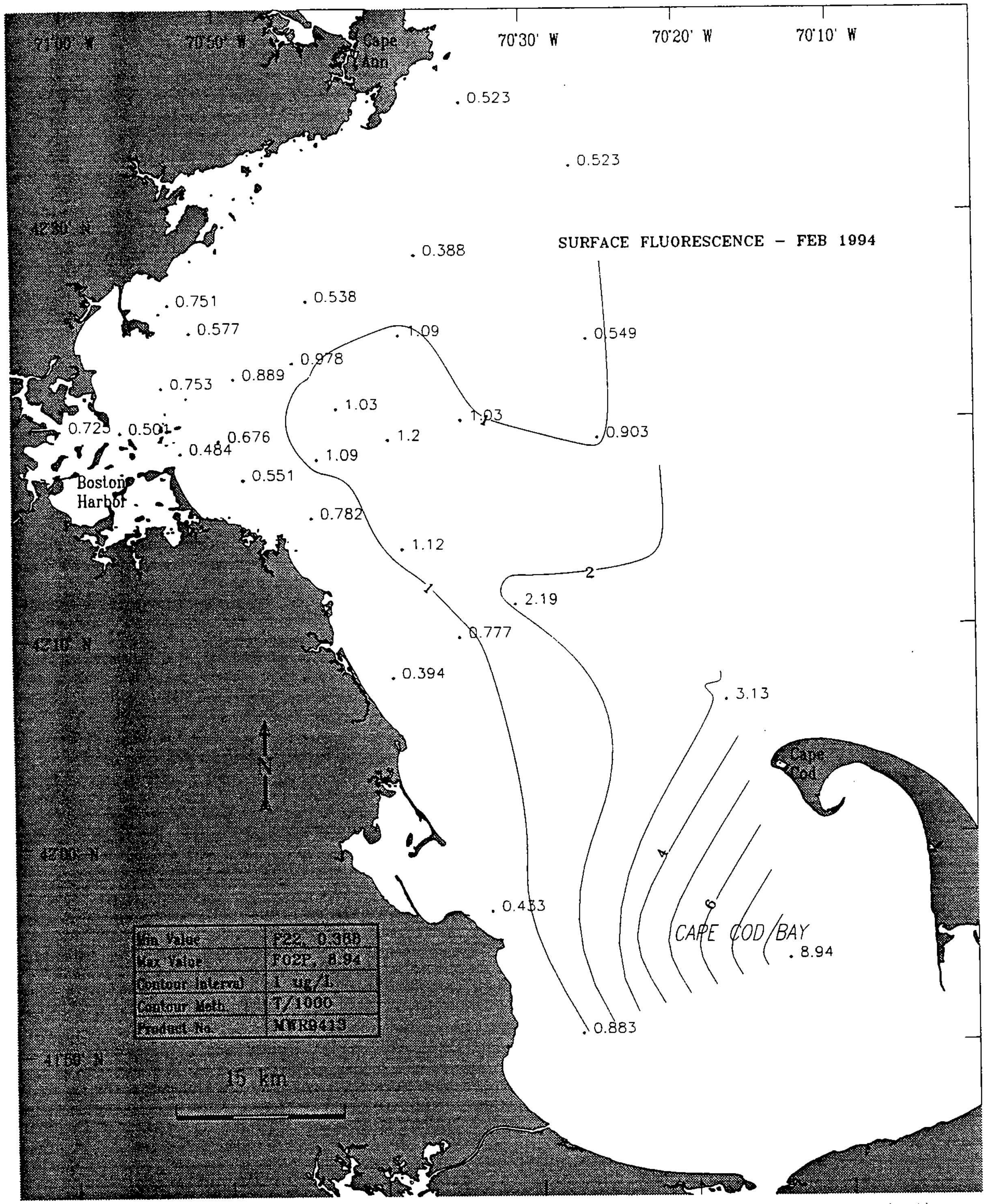

Long Scale - February - Dissolved Oxygen
semivariogram (mg/L)2
6.0
5.4
4.8
4.2
3.6
3.0
2.4
1.8
1.2
0.6
0.0
0
2500
5000
7500
10000
12500
15000
17500
20000
22500
25000
separation distance (m)
30
75
120
165

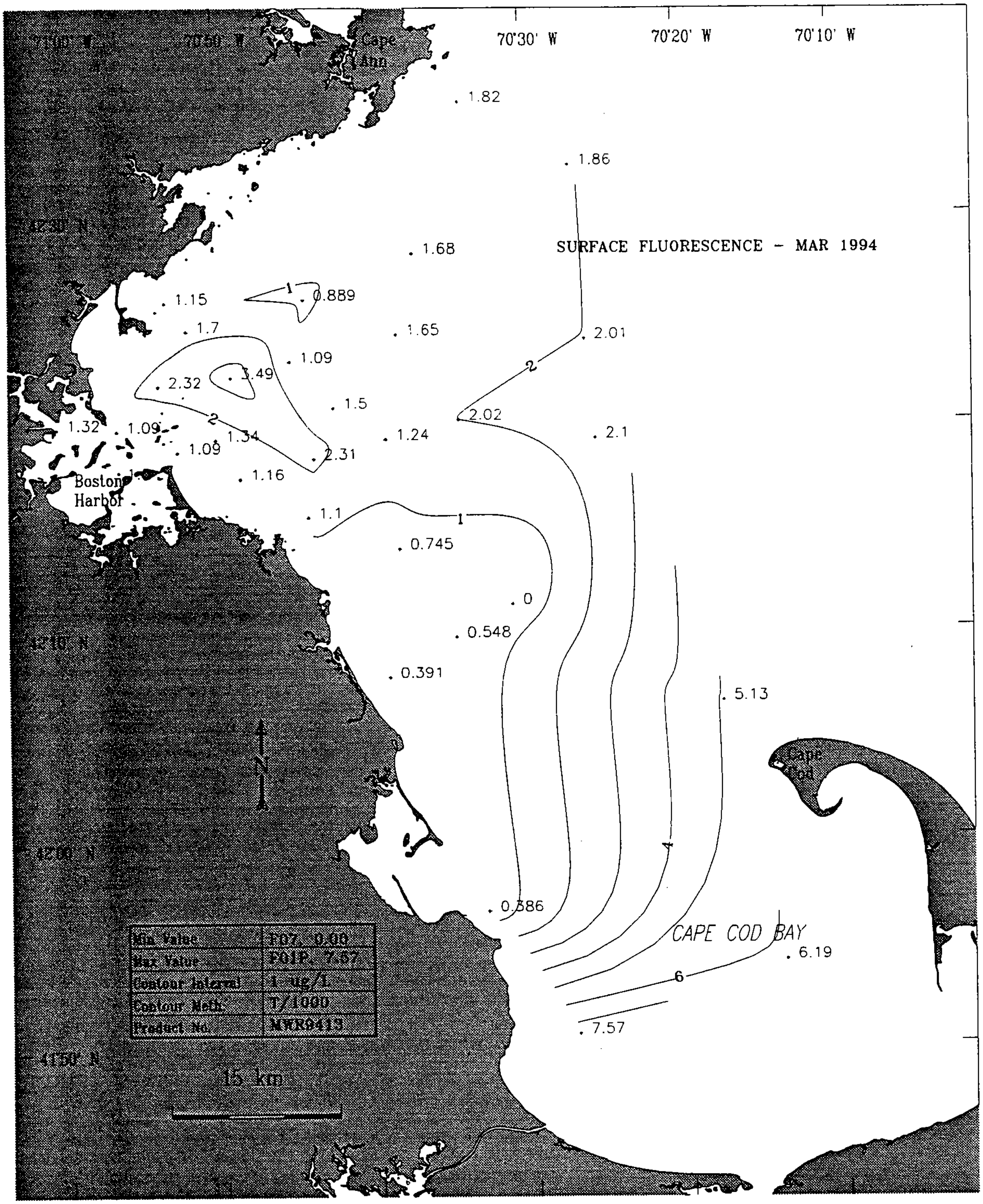


Long-Scale - August
Dissolved Inorganic Nitrogen
semivariogram (μm)^2
0.050
0.045
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
0
2500
5000
7500
10000
12500
15000
17500
20000
22500
25000
separation distance (m)
30
75
120
165

000032


Long-Scale - February
Dissolved Inorganic Nitrogen
semivariogram (μm)²
40.0
36.0
32.0
28.0
24.0
20.0
16.0
12.0
8.0
4.0
0.0
0
2500
5000
7500
10000
12500
15000
17500
20000
22500
25000
separation distance (m)
30
75
120
165


SURFACE FLUORESCENCE – FEB 1994
71°00' W
70°50' W
70°30' W
70°20' W
70°10' W
42°30' N
42°10' N
42°00' N
41°50' N
Cape Ann
Boston Harbor
Cape Cod
CAPE COD BAY
N
0.523
0.523
0.388
0.751
0.538
0.577
1.09
0.549
0.978
0.889
0.753
1.03
1.03
0.903
0.725
0.501
0.676
1.2
0.484
1.09
0.551
0.782
1.12
2
2.19
1
0.777
0.394
3.13
4
6
0.433
8.94
0.883
Min Value | P22, 0.388
Max Value | F02P, 8.94
Contour Interval | 1 ug/l
Contour Meth. | T/1000
Product No. | MWR9413
15 km


SURFACE FLUORESCENCE - MAR 1994
70'30' W
70'20' W
70'10' W
Cape Ann
Boston Harbor
CAPE COD BAY
Cape Cod
1.82
1.86
1.68
1.15
0.889
1.7
1.65
2.01
1.09
2.32
3.49
1.5
1.32
1.09
2.02
2.1
1.34
1.24
1.09
2.31
1.16
1.1
0.745
0
0.548
0.391
5.13
0.386
6.19
7.57
Min Value F07, 0.00
Max Value F01P, 7.57
Contour Interval 1 ug/l
Contour Meth T/1000
Product No MWR9413
15 km

The Massachusetts Water Resources Authority
Charlestown Navy Yard
100 First Avenue
Charlestown, MA 02129
(617) 242-6000

REVIEW OF THE DRAFT REPORT
"OPTIMIZATION OF WATER COLUMN MONITORING: STATISTICAL TREATMENTS"

Andrew R. Solow
Woods Hole Oceanographic Institution
Woods Hole, MA 02543

This report describes some exploratory statistical analysis of water column data from a baseline monitoring program in Massachusetts Bay developed by the Massachusetts Water Resources Authority. In broad terms, the analysis was aimed primarily at determining whether the current spatial and temporal sampling effort is adequate to detect change.

The report suffers considerably from the absence of a clear definition (or definitions) of what would constitute a change of interest. For example, the discussion in Sections 3.1.2 and 3.2.3 revolves around the problem of detecting a change through time in the measured value of a variable of interest at a single location. This case does not seem particularly relevant to monitoring water quality. Interest may more reasonably center, for example, on detecting a year-to-year change in mean chlorophyll-a concentration within a region. The failure to specify more clearly what sort of changes are of interest - while presumably not the fault of the authors of this draft report - is a serious problem.

Throughout the report, the measure of change is the difference between the actual value and a standard. This is somewhat different from a change in the actual value from year to year. In this latter case, the value in neither year is known and the problem is somewhat more difficult. Incidentally, under the assumptions of the report, higher power is achieved in assessing the significance of this kind of change by exploiting the fact that observations are spatially paired (i.e., taken at - or nearly at - the same locations in each year).

My interpretation of the semi-variograms included in the draft report is that a reasonable statistical model for many of the variables contains a seasonal cycle, a spatial trend (possibly a function of distance from shore), and additional variability that may be nearly uncorrelated over temporal and spatial sampling scales. The authors of the draft report chose to treat the seasonal cycle and spatial trends as autocorrelated stochastic processes. However, it is not just the case that values separated by short temporal or spatial lags are relatively close, but it is also the case that these values are predictably high at particular times of the year and in particular parts of Massachusetts Bay. This has clear implications for monitoring. If the problem is to detect exceedances of an absolute standard, then monitoring effort should be concentrated at times and locations where the background concentration is high. It seems to me that some minimal consideration of the physical and biological setting would be useful in designing a monitoring plan.

The geostatistical analysis presented in the draft report is rather mechanical. The draft report does not contain sufficient detail (e.g., about the semi-variogram models used in the kriging) for a thorough evaluation. There are occasional curiosities. For example, the overall levels of the directional semi-variograms in Figure 9 are different. This implies that the variances of the observations used to construct these semi-variograms were different. It is difficult to understand how this can be, unless different sets of observations were used to construct these semi-variograms, in which case the difference may not be directional at all. As another example, Figure 11 appears to contain a periodic component with period of around 600 m out to a range of about 2000 m. Could this be an artefact of the sampling?

In overall terms, my evaluation of this draft report is that, given the information at hand, it does an adequate job of addressing the questions that it raises. However, I would not recommend proceeding on the basis of the recommendations of this draft report. A more careful specification of the requirements of the monitoring plan is needed before specific recommendations along these lines can be formulated.

TO: Dave Aubrey

FROM: Andy Solow Andy

SUBJECT: Draft report

DATE: September 17, 1995

Dave,

I got your fax about my comments. I can answer the first question now and I will think about the second one over the next few days.

Let's focus on the problem of estimating the value of the variable of interest at an unsampled location from synoptic observations taken at a number of sample locations. The method used in the report - kriging - uses a linear combination of the observations as the estimator. The weights in this linear combination are chosen to minimize the error variance. Both the optimal weights and the corresponding minimized value of the error variance depend on spatial covariance.

A common approach - and the one followed in the report - is to use the observations to estimate spatial covariance at a number of lags. For example, if the observations lie on a grid with unit spacing, then estimates can be formed at lags 1, 2, 3, ...; at integer multiples of the square root of 2; etc. A continuous model is then fit - often by eye - to the estimates at discrete lags. A key parameter of this model is the magnitude of the discontinuity (if any) at the origin. In geostatistics, this is called the nugget effect.

Once the continuous model, including the nugget effect, has been fit, the analysis proceeds as if the fitted model were correct. This has always bothered the hell out of me, because it ignores the error in estimating the spatial covariance. The error in estimating the nugget effect may be especially large unless there some observations are taken at very short scales. Unfortunately, both the optimal weights and the associated error variance can be sensitive to the nugget effect, so that the error in its estimation can be serious.

Turning to the report, the nearfield sampling scale does not go below 2600 m, so the structure of the spatial covariance below that scale will be missed. It would also have been nice to see the number of sample pairs used to estimate the spatial covariance at the shorter lags. The model fitted to the semi-variograms are also not shown. It is to this kind of thing that I was referring in my report when I noted that insufficient detail is given to evaluate the analysis.

I hope this is helpful.