CRWR Online Report 03-01

Spatial and Temporal Trends of the
Western Arctic Ocean Benthic Community

by

Jonathan L. Goodall, M.S.E.

Graduate Research Assistant

and

David R. Maidment, Ph.D.

Kenneth H. Dunton, Ph.D.

Principle Investigators

May 2003

Spatial and Temporal Trends of the
Western Arctic Ocean Benthic Community

By

Jonathan L. Goodall, B.S. Civil Engineering

Report

Presented to the Faculty of the

Department of Environmental and Water Resources Engineering of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Engineering

University of Texas at Austin

May 2003

To my parents

Acknowledgments

I want to thank those that made this report possible: my advisors Dr. Maidment and Dr. Dunton for their guidance, mentorship, and financial support, Dr. Barrett for helping me in the early days of the project, and Dr. Maidment's research group members for their friendship and teaching. I have learned a great deal from all of you.

Abstract

The Space Time Distribution of Benthos in the Western Arctic Ocean

Jonathan L. Goodall, M.S.E.

The University of Texas at Austin, 2003

Supervisor: David R. Maidment

The spatial and temporal distribution of the benthic community of the Western Arctic Ocean was analyzed for trends in biomass through geostatistical interpolation of a retrospective (1970 – 1995) database of 1,093 point samples. A Geographic Information System (GIS) was used to manage, analyze, and display the spatially referenced point samples, as well as the interpolated continuous surface of benthic biomass. The geostatistical interpolation produced both mean predictions and prediction standard errors on a continuous scale within the study region. Natural variability of benthic biomass was evident in the standard errors, which were of the same magnitude as the prediction mean values. The final result of the spatial analysis revealed very high benthic biomass in the Bering and Chukchi Seas and lowest organism abundance on the Alaskan Beaufort Sea shelf. Areas of relatively low localized biomass were also noted at the outlets of the two major river basins within the study region: The Yukon and Mackenzie River systems. The temporal analysis of samples spanning three decades measured within a 40,000 km^2 area south of St. Lawrence Island in the Bering Sea showed an increase in benthic biomass from 1970 to 1995, but the ability to detect decadal temporal trends throughout the study area was hindered by an insufficient spatial overlap of data sampled through time.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 BACKGROUND

The purpose of this study was to identify the spatial and temporal trends of

benthic biomass in the Western Arctic Ocean (Figure 1) by applying geostatistical

techniques to a retrospective dataset collected from 1970 to 1995. For the

purpose of this study, I define benthic biomass as the abundance, in gm^{-2} wet

weight, of all macrofauna larger than 1 mm in size that live on or in the bottom

sediment of the ocean floor.
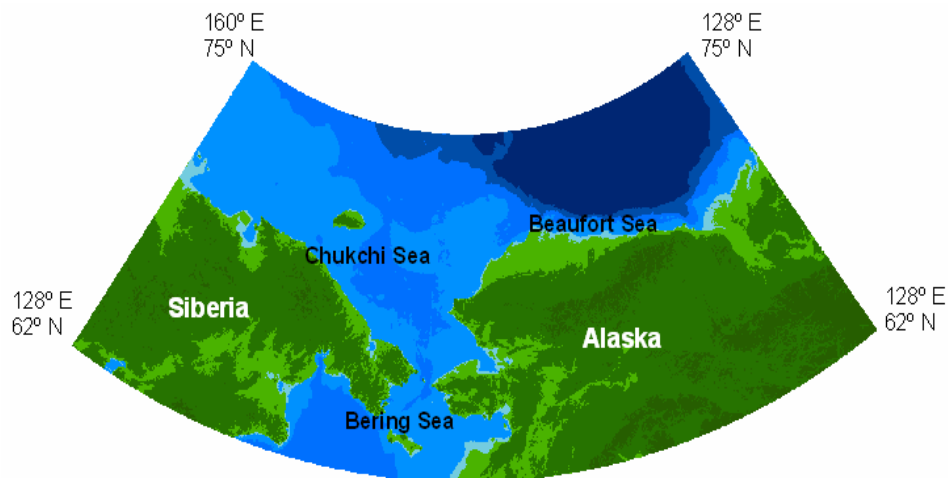
Figure 1 – The Western Arctic Ocean, the study region for this report.

The spatial and temporal patterns in benthic faunal biomass can provide

important information about overlying water column productivity and the

coupling of carbon between pelagic and benthic communities. An understanding

of the linkages between arctic circulation processes and spatial patterns in benthic

faunal biomass is critical to our ability to predict the consequences of global climate change on arctic marine ecosystems with respect to biological productivity on arctic selves.

This project was funded by a National Science Foundation (NSF) grant through the Arctic System Science (ARCSS) research program. The goal of the ARCSS program is to develop a better understanding of the ecosystem-wide impacts of climate change on the arctic system. The Shelf Basin Interactions (SBI) project is one of several ARCSS initiatives that address the coupling of carbon between arctic shelves and basins. The SBI project consists of three phases that are scheduled to be completed over the next six years.

Phase I was designed to undertake retrospective studies of datasets in the North Bering, Chukchi, and Beaufort Seas that would help focus field research efforts in phase II on the cycling and transformations of carbon on the western arctic shelves. Large and potentially invaluable historical databases existed in unpublished records from the Western Arctic region that had never been synthesized, including thousands of records on the biomass, density, and composition of benthic organisms from the North Bering, Chukchi, and Beaufort. Work completed by a previous graduate student within the Department of Environmental and Water Resources at the University of Texas at Austin, Jóna Finndís Jónsdóttir, resulted in a first draft of these data into a single, spatially referenced relational database (Jónsdóttir 2000). Work detailed in this paper describes revisions made to the relational database including the addition of

sample stations, temporal and spatial synthesis, and integration of water column

chlorophyll *a* data. This spatially referenced database represents a substantial and

extremely valuable collection of baseline data of the benthic community over the

past thirty years.

Phase II of the SBI project (2001-2006) will constitute the experimental

phase of the research effort in the North Bering, Chukchi, and Beaufort Seas. The

SBI program will conclude with a regional modeling effort in Phase III (2007-

2009) that will address the effects of global change on the ecosystems of the

western arctic shelves and basin (Grebmeier et. al 2001). All three phases have

the overarching goal of utilizing the benthos as a spatial and temporal integrator

of oceanographic processes that could provide a valuable indicator of potential

global change impacts.

1.2 STUDY OBJECTIVES

This project's primary objective is to describe the spatial and temporal

pattern of benthic biomass within the Western Arctic Ocean. Two main tasks

were employed to meet this objective (Figure 2). The first task was to retrieve

additional biological and chemical data relevant to the study area, including

additional benthic biological and integrated chlorophyll *a* data from the National

Oceanographic Data Center (NODC) and other published and unpublished

Western and Russian sources. These datasets were then be added to the current

arctic database to create a more complete account of the benthic community. The

second task employed Geographic Information Systems (GIS) software in conjunction with geostatistical techniques to examine and graphically display the spatial and temporal trends of the benthic community within the Western Arctic Ocean.

The first study task was further divided into four steps. First, additional benthic biomass datasets from unpublished sources were gathered to close data gaps in the East Siberian Sea off the Russian coast. Second, chlorophyll *a* measurements collected in the region were consolidated from all known sources, both published and unpublished. Third, other datasets were located to test hypotheses that benthic biomass was correlated with temperature or depth. Finally, these additional datasets were incorporated into the arctic database for analysis using GIS and geostatistical software.

The second study task, to determine the spatial and temporal patterns of the benthic community using geostatistical models, was accomplished in four steps. First, an exploratory data analysis was performed on the raw samples to understand their basic statistical distributions in time and space. Second, a geostatistical model was selected to interpolate benthic biomass at unmeasured locations. Third, a spatially continuous surface of benthic biomass in the Western Arctic Ocean was derived from the geostatistical model and a temporal analysis was conducted to determine whether any long-term trends were evident in the dataset. Finally, the spatial and temporal analyze were subjected to an uncertainty analysis resulting in prediction errors for the interpolated surfaces.

Primary Goal: *Determine the spatial and temporal trends of benthic biomass in the Western Arctic Ocean*

Task 1: *Complete geodatabase of chemical and biological data*

Sub-tasks
  - add benthic biomass data
  - add Chlorophyll data
  - add depth & temperature data
  - compile into geodatabase

Task 2: *Use geostatistics to identify spatial and temporal trends for benthic biomass*

Sub-tasks
  - exploratory data analysis
  - geostatistical model development
  - prediction surface generation
  - uncertainty analysis

Figure 2 – Summary of project objectives and tasks.

1.3 REPORT OUTLINE

Following this introduction, Chapter 2 will review the relevant literature for the use of geostatistical methods using GIS. Chapter 3 will introduce the retrospective datasets, discuss how they were gathered, and show the processing steps necessary for viewing the spatial data in a GIS. Chapter 4 will present a complete exploratory data analysis of the benthos in both space and time. The results of the temporal and spatial trend analysis of benthos in the Western Arctic Ocean will be presented in Chapter 5. Finally, Chapter 6 will provide concluding

thoughts on the trend of benthos over space and time and offer recommendations

for future research and data collection within the study area.

Chapter 2: Background

2.1 INTRODUCTION TO ArcGIS AND Geodatabase Design

GIS proved critical in the analysis of the spatially referenced station measurements of benthic biomass and related attributes. The specific GIS software used in this study was ArcGIS 8.2. ArcGIS is a commercial GIS software system for the visualization, creation, presentation, and map creation of spatial data (Minami *et al.* 1999). It is the latest development from Environmental Systems Research Institute (ESRI) meant to combine the company's previous two GIS systems, ARC/INFO and ArcView, with a common interface, but with two licensing levels: ArcView 8 and ArcInfo 8. ArcView 8 provides only a portion of the capabilities of ArcInfo 8, but costs significantly less. ArcGIS is a collective term for both ArcView 8 and ArcInfo 8.

One important advance first introduced with ArcGIS is the geodatabase data model. For the geodatabase data model, feature classes are data layers within the personal geodatabase. Feature classes can be grouped into a feature dataset where all the feature classes have the same coordinate system (McDonald 1999). Feature datasets provide a logical grouping of datasets. For example, biological data can be separated from chemical data or physical data. Table 1 describes feature datasets and feature classes and how they related to previous GIS data structures (i.e. coverages and shapefiles).

	Geodatabase	Coverage	Shapefile
Collections of datasets	A geodatabase is a collection of feature datasets	An ArcInfo workspace is a collection of coverages	A shapefile folder is a collection of shapefiles
Datasets	A feature dataset is a collection of feature classes	A coverage is a collection of coverage feature classes	A shapefile has one shapefile feature class
Collections of features	A feature class is a collection of features of the same type	A coverage feature class is a collection of coverage features	A shapefile feature class is a collection of shapefile features
Features	Point, multipoint, polylines, polygon, annotation, and network	Primary cover feature classes; point or label point, arc, and node. Secondary feature classes: polygon, tic, link, section, and annotation. Compound feature classes; region and route.	Point, multipoint, line, and polygon.

Table 1 – Comparing the structure of vector datasets (Minami 1999)

An advantage of the geodatabase data model is that it supports object-oriented relationships between data layers (McDonald 1999).   For example, the recently developed Arc Hydro Data Model is a water resources geodatabase with relationships between watersheds and a point placed at the watershed's outlet (Maidment 2002). The relationships between the watershed and outlet point feature classes allows for spatially distributed data within the watershed to be summarized at a single related point at the outlet. The same idea proved very useful for creating moving window statistics for exploratory data analysis, as will be shown in Chapter 4.

2.2 INTRODUCTION TO GEOSTATISTICS

The spatial and temporal analysis of the benthic community structure relies heavily on geostatistical modeling of benthic biomass to provide an estimate (with uncertainty) of biomass densities at unmeasured locations. Geostatistical techniques, while originally developed for mining engineering, have become increasingly important to the earth and natural sciences. More recent applications of geostatistics include estimating marine biomass for primarily economic motivations (i.e. where fishermen might find the most lobsters). Prior to discussing these recent applications, however, it is important to introduce the basics of geostatistical theory to familiarize the reader with the essential mathematics and terminology.

Most classical statistical methods do not account for spatial information within data, so regression techniques using classical statistical theory provide less accurate prediction results for data correlated over space (Isaaks and Srivastava 1989). Even if one attempts to predict an attribute value over space by using nonlinear regression, the attribute is often too irregular, again producing inaccurate results (Burrough 1998). The extension of classical statistics using geostatistic techniques is of great value in the assessment of spatial correlation in data when predicting values at unmeasured locations (Isaaks and Srivastava 1989). Geostatistics accomplishes this task by calculating the attribute value at an unmeasured location, $Z(x)$, as the sum of three factors: (1) the deterministic trend,

m(x), (2) the autocorrelated variation, $\varepsilon'(x)$, and (3) a constant correction factor

for statistical noise, $\varepsilon''(x)$ (Burrough 1998). The general equation for kriging

interpolation is:

$$Z(m) = m(x) + \varepsilon'(x) + \varepsilon'' \qquad (1)$$

Ordinary kriging, the geostatistical technique used in this study, assumes

that m(x) is constant over space and equal to the mean of the sample space

(Johnston *et al.* 1999). Thus, the attribute value at an unmeasured location x,

Z(x), is simply a function of the autocorrelation factor, $\varepsilon'(x)$, and a constant

correlation factor for statistical noise, $\varepsilon''(x)$. Universal kriging relaxes the

assumption of a constant mean, but often does not produced more accurate results

because additional parameters must be introduced to describe the changing mean

(Johnston *et al.* 1999).

Autocorrelation is the tendency for samples separated by less distance to

be more similar than samples separated by a greater distance. The spatial

correlation present in many earth science datasets is commonly called

autocorrelation to reflect the fact that the dataset is correlated with itself (Isaaks

and Srivastava 1989). The autocorrelation factor used in kriging is obtained from

a model fit to a plot of separation distance (lag, h) vs. semivariance ($\gamma(h)$) for

pairs of points separated by a distance h±Δh (Figure 3). For all pairs of samples

separated by a distance h±Δh, the semivariance, $\hat{\gamma}(h)$, can be calculated using

equation 2.

$$\varepsilon'(x) = \hat{\gamma}(h) = \frac{1}{2n}\sum_{i=1}^{n}\left(Z(X_i) - Z(X_i + h)\right)^2 \qquad (2)$$

Once a semivariance vs. lag plot has been generated (also known as an

experimental semivariogram), a function is fit to the data to model the

relationship between separation distance and the semivariance, $\gamma(h)$ (Figure 3).

Common fitting functions include the spherical model, the exponential model, the

linear model, and the Gaussian model. For this study, the spherical model

(Equation 3) was selected because it is a commonly used model and because it

modeled our experimental data well.
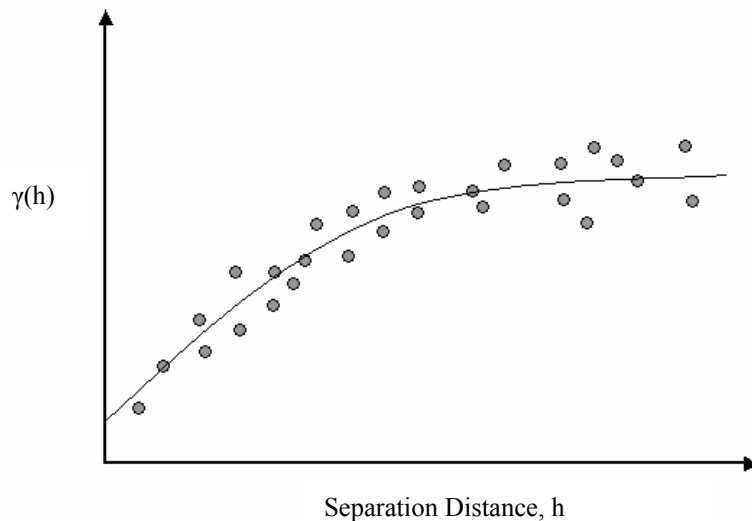
Figure 3 – An example empirical semivariogram with a fitting model.

$$\gamma(h) = c_o + c_1 \left\{ \frac{3h}{2a} - \frac{h^2}{2a^3} \right\} \qquad \text{for } 0 < h < a \qquad (3)$$

$$= c_o + c_1 \qquad\qquad \text{for } h \geq a$$

There are a few commonly used terms to refer to significant characteristics of the semivariogram (Figure 3). First, the example model predicts a nonzero variance for points separated by no distance. This discontinuity is referred to as a nugget or a nugget effect. Nugget effects can be the result of either intrinsic micro-scale variability within the data or sampling errors, and their magnitude provides insight into the smoothness of the data (variability over small separation distances). Another important characteristic of the semivariogram is the range. The range is the separation distance for which points separated by more distance are not correlated. The range can be thought of as the radius of influence for each sample location. Each measurement will have some influence on the prediction at unmeasured locations within a radius equal to the range around the measurement. Finally, the sill is a term given to the semivariance for separation distances greater than or equal to the range (Isaaks and Srivastava 1989).

Once a model for the semivariance has been computed and optimized, it can be used to find the weighting factors assigned to all measured points neighboring an unmeasured point, λ_i, by solving a series of equations. The value at the unmeasured point is then predicted by summing the product of all neighboring locations and their associated weight (Equation 4). To assure the

prediction is unbiased, the sum of the weighting factors must equal one (Isaaks and Srivastava 1989).

$$\hat{Z}(s_o) = \sum_{i=1}^{N} \lambda_i Z(s_i) \qquad (4)$$

Because geostatistics predicts measurements at unmeasured locations using statistical theory (i.e. correlation), the prediction values are accompanied by prediction standard errors or measurements of the uncertainty in the predicted values. This feature separates geostatistical techniques from other deterministic techniques such as inverse distance weighting. The output of the geostatistical model can be used to generate a probability density function (PDF) for benthic biomass at each unmeasured location.

2.3 GEOSTATISTICS AND GIS

One unique quality of this project is its use of geostatistical techniques to describe the spatial and temporal distribution of benthos from within a GIS environment. The geostatistical modeling is fully integrated within the GIS program as an extension developed by ESRI. The Geostatistical Analyst extension is new to ArcGIS 8 and it includes a number of deterministic and stochastic tools for interpolating surfaces from point samples such as inverse distance weighting, polynomial fitting, and kriging.

Although the Geostatisical Analyst is a relatively new addition to the ESRI product line, researchers have long been investigating the ways to link

geostatistical techniques with GIS for a number of years. P.A. Burrough of

Utrecht Centre for Environmental and Landscape Dynamics at Utrecht University

in the Netherlands provided a comprehensive explanation of the mutual benefit of

the joining of geostatistics with Geographic Information Systems (Burrough

2001). Burrough focused on the advantage of merging GIS and geostatistical

software from both the prospective of adding geostatistics to GIS (a GIS expert's

perspective) and of adding GIS to geostatistics (a geostatistical expert's

perspective). Because of its relevance to this report, the two paragraphs below

provide a summary of Burrough's thoughts.

The advantage of having GIS functionality coupled with geostatistical

techniques, Burrough (2001) states, is that the GIS provides a spatial context for

interpolation and conditional simulation, and tools for the visualization and

presentation of the geostatistically derived surfaces. GIS can also serve as a

storage system for various data layers, providing a context for the layers being

analyzed with geostatistics. For example, if one is interested in the distribution of

precipitation over a basin, the GIS can have the basin outline and the river

network to provide a context for the precipitation gages and resulting precipitation

prediction surface. Zonal statistics can easily be calculated for the precipitation

over the basin using in-house GIS functionality.

The advantage of having geostatistical functionality within a GIS is that

the user can benefit from statistically derived methods for interpolating point

samples over an area, and in the process, quantify the confidence in the

interpolated surface. Thus, if one has geo-referenced data already in a GIS environment, one does not need to bring that data into a different software system for geostatistical analysis. The partnership simply removes the extra step of transferring data between GIS and the modeling system. As a result, geostatistical techniques are brought to many who would otherwise not make the effort.

2.4 USING GEOSTATISTICS FOR BIOMASS ESTIMATION

Geostatistics have been used in previous research to assess the temporal and spatial components of marine life structure. One notable example is the work done by Maynou et al. (1998) on the spatial structure and biomass evaluation of lobster populations in the northwestern Mediterranean Sea. The researchers used geostatistical techniques to analyze lobster catches by commercial fishing trawlers during two periods separated by six months. Lobster is the most economically important species for Europe, thus the motivation of their research was to predict the spatial distribution patterns of lobster at the highest resolution possible. They succeeded in doing so by use of a geostatistical model.

Another application of geostatistics for biomass estimation in the marine science community is Defeo and Rueda (2002). The objective of their research was to quantify the population structure and abundance of sandy beach macroinfauna. In doing so, the researchers used two different approaches to test the reliability of each. The first method was stratified random sampling (design-based) and the second was geostatistics (model-based). The researchers

concluded that geostatistics is a powerful tool for providing key information on the spatial organization and abundance estimation of sandy beach macroinfauna.

A final application of geostatistical techniques for marine population estimation noted in this paper was done by Romaine *et al.* (2002). Their research compared population size estimates using simple block averages vs. geostatistical block averages. Geostatistical block averages, or block kriging, can be used to estimate an attribute value over an area instead at one point. The authors concluded that, when data have a degree of autocorrelation (that is correlations over space), block averages may not be appropriate and, instead, geostatistical averages using block kriging should be used. Block kriging results, the authors concluded, provide detailed distribution plots and reduce variance error estimates.

Chapter 3: Data

3.1 BENTHIC BIOMASS

Work on collecting all known measurements of benthic biomass within the Western Arctic Ocean began four years ago through a joint effort between the Center for Research in Water Resources (CRWR) and the Marine Science Institute (MSI) of the University of Texas at Austin. Investigators at MSI built a database of benthic biomass samples in the Western Arctic Ocean which provided a starting point for this research (Jónsdóttir 2000). The database included samples collected by five researchers where each record was minimally defined by the latitude, longitude, date, and biomass weight (g/m^2) at each collection site. Additional information including integrated chlorophyll a, temperature, and depth were included for some collection sites.

A portion of the benthic biomass data came from the National Oceanographic Data Center (NODC). The majority of the NODC data for the Western Arctic Ocean was compiled by Broad (1981). The data was delivered to the Marine Science Institute in text file format. Computer programming scripts were needed to transform the text files into database format. This work was carried out by researchers at the Marine Science Institute. Additional information about the NODC benthic biomass, is provided by Jónsdóttir (2000).

Additional benthic biomass data came from sources outside of the NODC including Grebmeier (1987, 1985), Carey (1984), Wacasey (1974), and Feder

(1982). Grebmeier submitted four spreadsheets that included latitude, longitude, date, and benthic biomass at each collection site. Carey and Wacasey's work focused on data collection in the Western Beaufort Sea on the northern coast of Alaska, while Feder collected samples in the Bering Sea.

Over the past two years, additional benthic biomass data have been collected and added to database (Figure 4). The additional data, collected by Stoker from 1970 to 1974 and from Grebmeier from 1993 to 1995, was intended to close obvious gaps noted in the previous datasets, particularly in the East Siberian Sea. The additional data were processed and added to the Arctic geo-referenced database feature class following the procedure outlined in Section 3.4.
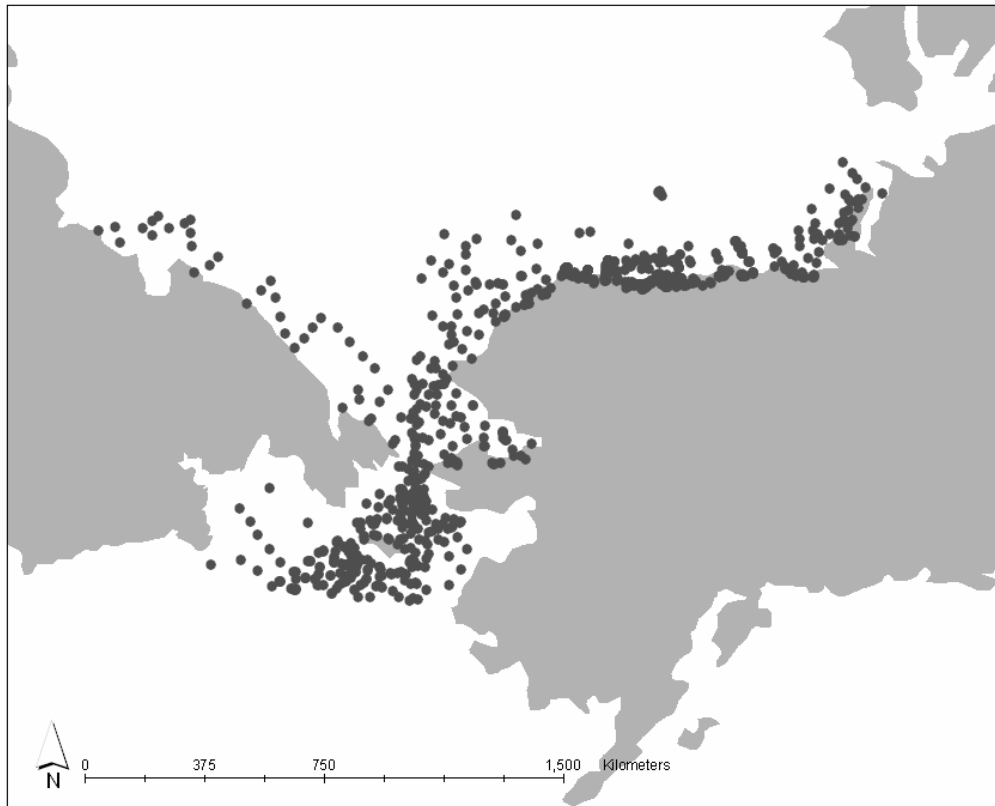
Figure 4 – The 1,093 benthic biomass sampling stations

3.2 INTEGRATED CHLOROPHYLL *a*

In addition to the supplemental benthic biomass samples, I analyzed integrated chlorophyll *a* data from the study area. The chlorophyll data provides an opportunity to examine the linkages between water column productivity and benthic faunal biomass. Chlorophyll *a* reflects the abundance of water column phytoplankton, which is the major carbon source for bottom-dwelling benthos. I hypothesized that high levels of chlorophyll *a* in the water column would be

reflected in higher levels of benthic biomass, since ungrazed phytoplankton sinks directly to the seabed.

Some gathered data were not integrated over the water column, but represented discreet water column measurements of chlorophyll *a* at various depths. To create depth integrated results, a Visual Basic for Applications macro was written in EXCEL to calculate the integrated chlorophyll *a* for a sampling location based on the measurements of chlorophyll at various depths (see Appendix A for code). The macro was written to be independent of the number of depth samples for each location.

Each station was also analyzed to assure an adequate representation of chlorophyll over the total depth. Single measurements of chlorophyll *a* were excluded from the database on the basis that they did not accurately represent the integrated chlorophyll at the measurement location. The final dataset included 2,514 chlorophyll samples (Figure 5).
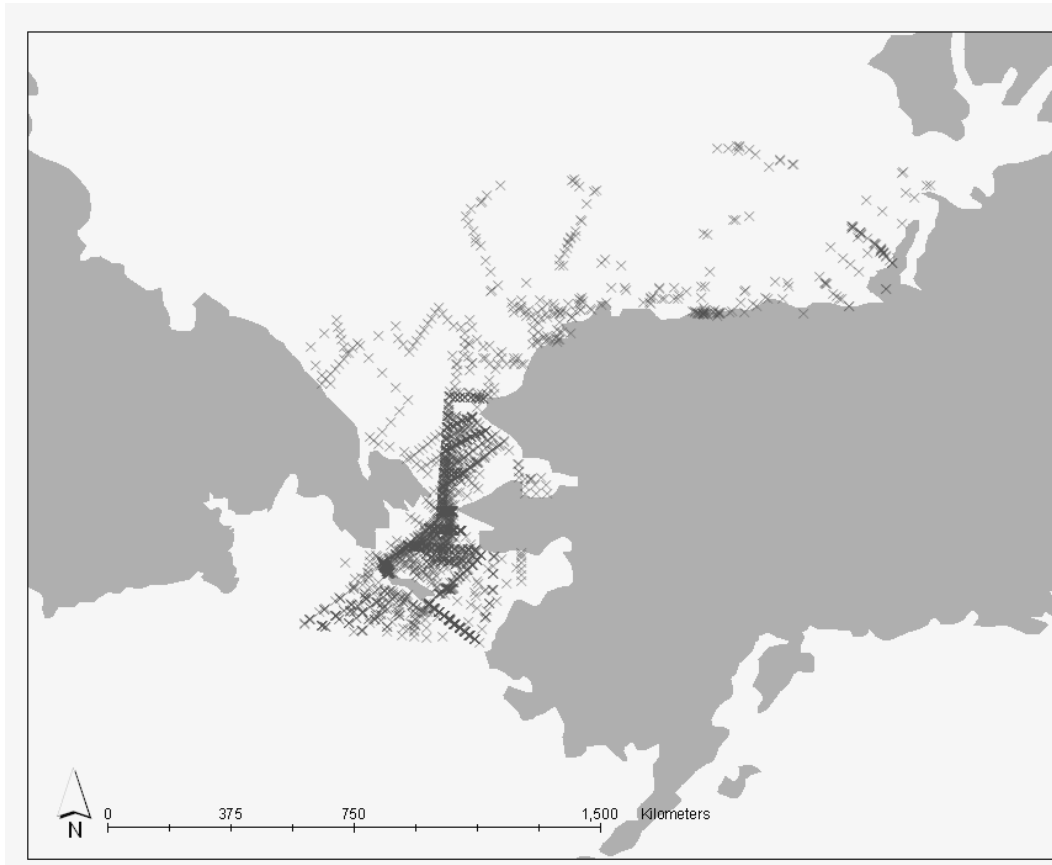
Figure 5 – Map of the chlorophyll a measurements in the Western Arctic Ocean.

3.3 CORRELATED MEASUREMENTS

It seemed logical that temperature and depth might be correlated to higher biomass density. If so, this additional information could be included into a geostatistical model to test these hypotheses. In addition to chlorophyll *a,* I hypothesized that benthic biomass would be inversely correlated to depth, but positively correlated with temperature. The result of correlations will be presented in the exploratory data analysis section of Chapter 4.

3.4 INCORPORATION INTO GIS

Once the biomass, chlorophyll, depth, and temperature data were gathered and preprocessed, the next step was to incorporate the data into a personal geodatabase as feature classes (Figure 6). Physical, Biological, Chemical, and Descriptive are the four feature datasets, providing the highest level of hierarchal data storage. Within the *Biological* dataset are two feature classes, *biomass* and *chlorophyll*. The *Descriptive* dataset houses a number of feature classes with the purpose of providing the background context for the sampling feature classes, i.e. land surface and moving window statistics grids. At present, the Chemical and Physical datasets are empty, but could be utilized if future efforts incorporate these data into programs that address the linkages between biological and physicochemical processes.
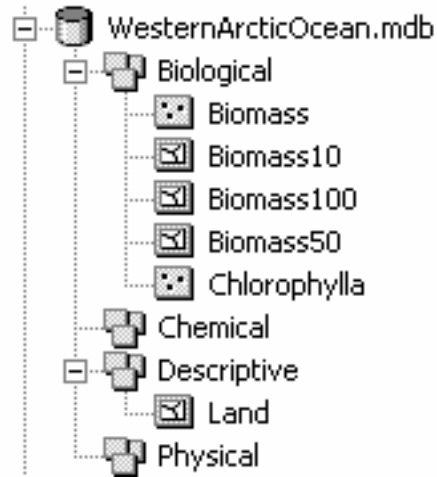
Figure 6 – The structure of the geodatabase used to integrate all previously collected benthos measurements into one GIS compatible format.

For the geostatistical analysis, it was necessary to transform the data from geographic coordinates (latitude and longitude) to a reference system (x and y coordinates). Numerous mathematical transformations, commonly referred to as map projections, can be used to convert geographic locations in three dimensions to reference locations in two dimensions. As a result of transforming a 3D image to a 2D image, each map projection distorts one or more of the four properties of the geographical system: shape, area, distance or direction (Minami 1999). For this project, the Lambert Azimuthal Equidistance projection was selected because kriging relies heavily on the separation distance between points, thus maintaining accurate distance measurements is critical. Appendix B contains the specific parameters of the Lambert Azimuthal Equidistance projection used in this research.

Area preservation is also important because one objective of this research is to quantify temporal changes in biomass within a fixed area in the Arctic Ocean, thus the area of the projected map must be the same as the real-world area of a 3D globe. It is not possible, unfortunately, to project the geographic coordinates into a reference system that preserves both distance and area (Minami 1999). Therefore, there were inaccuracies in the computation of temporal changes in biomass over fixed regions due to distortions of the area size due to using a distance preserving projection. These inaccuracies should not be significant, since they will be identical for each temporal period and will not bias the analysis towards one time period over another.

Chapter 4: Methodology

The previous section addressed the collection of benthic biomass, chlorophyll *a*, depth, and temperature data from both published and unpublished sources and how that data was input into a geodatabase. In the subsequent sections I describe the spatial and temporal analysis portion of the research. First, prior to the application of any geostatistical models, there must be a thorough exploratory data analysis (EDA) to better understand the statistical properties of the dataset. This knowledge is critical for creating a statistical model to quantify spatial and temporal patterns present in the benthic community. The spatial and temporal exploratory data analyses will be handled independently, starting first with the spatial analysis.

4.1 EXPLORATORY SPATIAL DATA ANALYSIS (ESDA)

The objective of exploratory spatial data analysis is to familiarize the user with the data to aid in the selection of a geostatistical model for biomass estimation at unmeasured locations (Kitanidis 1993). The Exploratory Spatial Data Analysis must identify the following features of the dataset: the presence of outliers and large-scale trends, its distribution, and the small or micro-scale variability (Johnston *et al.* 1999).

4.1.1 Outliers

As with any probability model, the presence of outliers can greatly influence the model prediction accuracies (Isaaks and Srivastava 1999). Thus,

prior to the selection and development of a geostatistical model, it is beneficial to identify any potential outliers. Outliers are only removed with adequate justification. It is often difficult to classify extreme measurements as outliers, especially when the researchers conducting the statistical analysis were not involved in the sample and data collection (as is the case here). When suspicious points are identified, the values and calculations used to derive the values are double checked. Unfortunately, this is not possible here.   For this study, the criterion for removing potential outlier removal will be based on the benefit derived from lowering the estimation error. This is, admittedly, a subjective decision, but the best that can be done given the circumstances.

One tool within the Geostatistical Analyst Extension for ArcGIS that proved helpful in identifying outliers and their effect on modeling the spatial correlation among data is the semivariogram cloud. A semivariogram cloud is a plot of the semivariance between a pair of measurements with respect to the distance separating the pair (the lag). This produces a spatially summarized benthic biomass dataset (Figure 7).

There exists a group of measurements with semivariance values far exceeding all others. The semivariogram cloud also shows that all of the high semivariance values are caused by a combination of one specific cell with many other cells. Thus, the one common cell is causing semivariance values that are distinctly different than those resulting from all other cell pairings. If the cell is kept in the dataset, its influence will significantly increase the variability of the

prediction results. I considered this cell (and its one underlying point) as an outlier and ignored it in the geostatistical estimation process in order to improve prediction confidence.
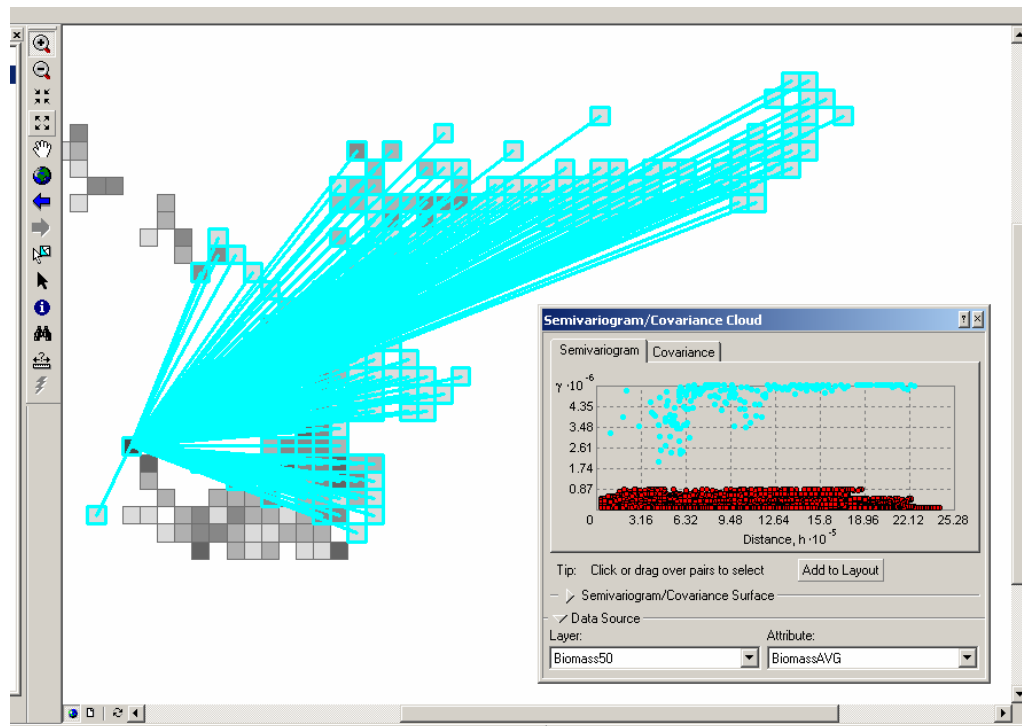
Figure 7 – Identification of outliers using the semivariance cloud tool

4.1.2 Distribution Analysis

Normally distributed random variables are beneficial for geostatistical interpolation for two reasons: 1) errors from the generation process are also normally distributed (Johnston *et al.* 1999), and 2) to account for the correlation between local variability and local mean. Non-normal distributions are often subject to this correlation of local variability and local mean, termed the

27

Proportional Effect (Cressie 1991) and can bias prediction results. For these two reasons, the biomass dataset was mathematically transformed to a normal distribution prior to estimation and then mathematically transformed back to its original distribution following interpolation.

A histogram of the biomass samples shows a positive skew nearly resembling a lognormal distribution (Figure 8). A log transformation of the dataset brought the distribution closer to normal, but could be improved by using a Box-Cox transformation with $\lambda = 0.15$. The Box-Cox transformation (Equation 5) is a method for transforming the frequency distribution of a dataset $Z(s)$ to Normal $Y(s)$ by adjusting the fitting parameter, λ, between zero and one (Johnston et al. 1999).

$$Y(s) = \frac{\left(Z(s)^{\lambda} - 1\right)}{\lambda} \tag{5}$$
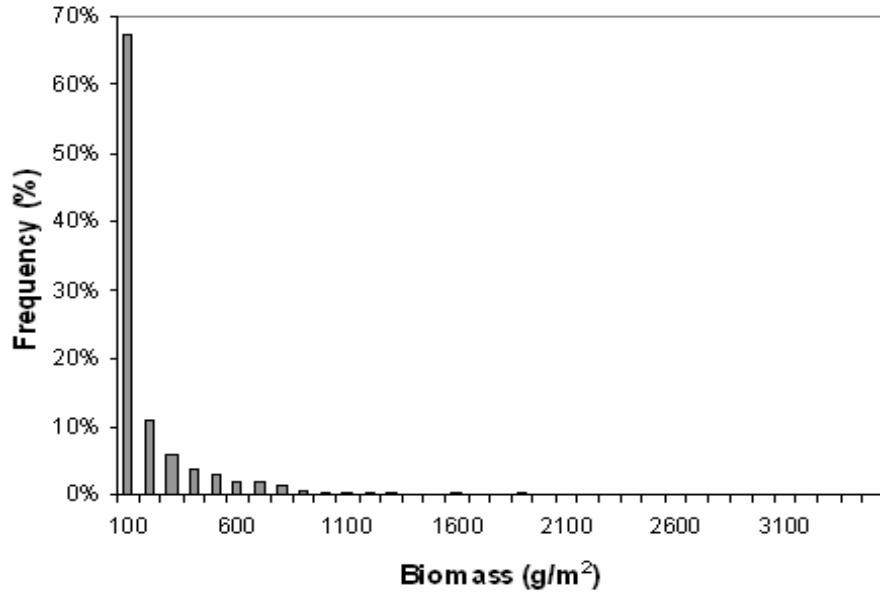
Figure 8 – Histogram of the benthic biomass dataset

For the benthic biomass dataset, a fitting parameter of 0.15 was used to produce a normally distributed dataset from the original dataset (Figure 9). This transformed dataset was input into the geostatistical model to estimate benthos at unmeasured locations. Once the estimations were calculated from the transformed dataset, the model applies an inverse Box-Cox transformation to calculate the estimations with respect to the original dataset.

Box-Cox (0.15) Transformation
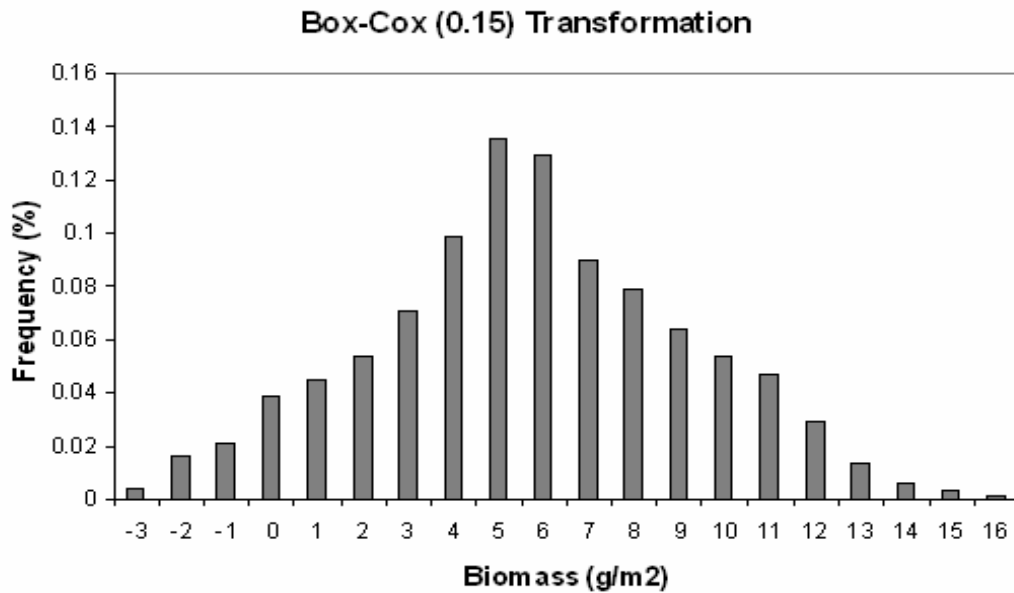
Figure 9 – Histogram of the transformed benthic biomass dataset

4.1.3 Large-Scale Trends

The presence of large-scale trends within a dataset is important when selecting a geostatistical model because different models deal with trends in different ways. As mentioned in the section, ordinary kriging interpolates a variable at an unmeasured location based on the assumption that the mean of the variable is constant. In other words, ordinary kriging assumes no large-scale trends exist in the dataset. Universal kriging is an interpolation method which allows for large-scale trends within datasets, but at the expense of introducing additional parameters which must be estimated to describe the nature of the trend. That said, it is sometimes the case that, even with large-scale trends present in data, ordinary kriging produces results as accurate as universal kriging (Johnston

et al. 1999) because fewer fitting parameters need to be estimated for an ordinary kriging model compared with a universal kriging model.

A three dimensional representation of the benthic biomass in the Bering Sea show high values that greatly exceed those measured in the Beaufort Sea off Alaska's northern coast (Figure 10). Because of this discrepancy in magnitude between Bering Sea and Beaufort Sea, a trend exists in the North to South direction. To account for these trends, the points were detrended by a cubic function prior to the geostatistical modeling using ordinary kriging. This is another way of accounting for trends without using the more sophisticated universal kriging model.
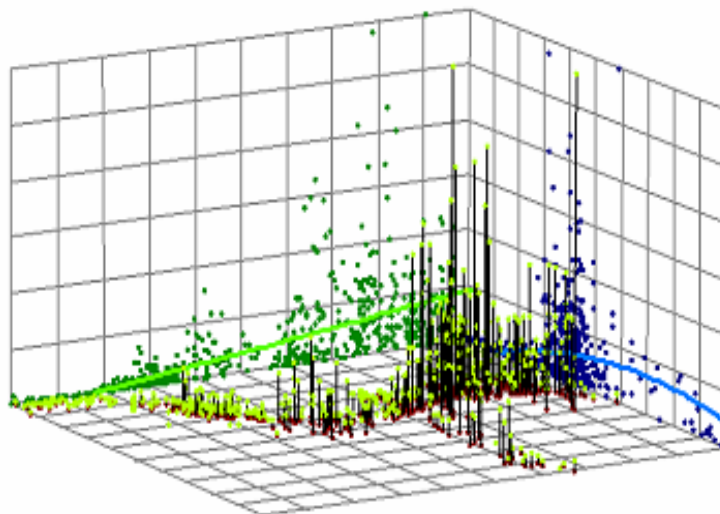

Figure 10 – 3D trend analysis of benthic biomass. Perspective is from Northern Siberia looking southeast.

4.1.4 Small-Scale Variability

The ability to accurately predict the overall surface of benthic biomass on a continuous scale relies heavily on the presence of small-scale variability of the

dataset. The fundamental assumption of the geostatistical interpolation method is that spatial continuity exists within the data, that is to say that two samples close to one another are more likely to be similar compared with two samples taken far apart (a property termed autocorrelation and introduced in Section 2.2, *Introduction to Geostatistics*). It is this correlation between separation distance and attribute value difference for all pairs of points that is captured in the semivariogram and hence used in the estimation of the attribute at unmeasured locations (Isaaks and Srivastava 1999).

If a dataset does not exhibit spatial continuity, it is virtually impossible to predict unmeasured locations on the basis of geostatistical models, no matter what model is used (Isaaks and Srivastava 1999). One measure of spatial discontinuity is the nugget effect (Johnston *et al.* 1999). If the semivariance does not approach zero as the lag (separation distance) approaches zero, a nugget exists within the data. Sources of a nugget effect can be either measurement errors or spatial sources of variation at distances smaller than the sampling interval (Johnston *et al.* 1999). While it is nearly impossible to eliminate all sources of measurement error, it is possible to reduce the spatial sources of variation at distances smaller than the sampling interval. This can be done by carefully considering the appropriate scales of spatial variation that are most important to the research at hand.

For the benthic biomass dataset, a preliminary look at the spatial continuity suggested that small-scale variability might significantly hinder the

development of a geostatistical model for point prediction. In the Bering Sea,

measurements varying by three orders of magnitude were recorded in nearly the

same geospatial location (Figure 11). This observation raises the question of the

spatial scale at which biomass can (and should) be predicted. Obviously, it will

be virtually impossible to make accurate predictions of the benthic biomass

spatial distribution at the sampling scale of $1m^2$. Kriging methods can fit a

surface to the raw sampling points, yet the prediction standard error (or the

standard deviation) for the predictions will be quite significant. The samples are

simply too random at such a fine resolution to predict with certainty. In addition,

the variability will introduce small-scale noise in the prediction. Thus, to reduce

the variability of the predictions, the samples may be binned and up-scaled to

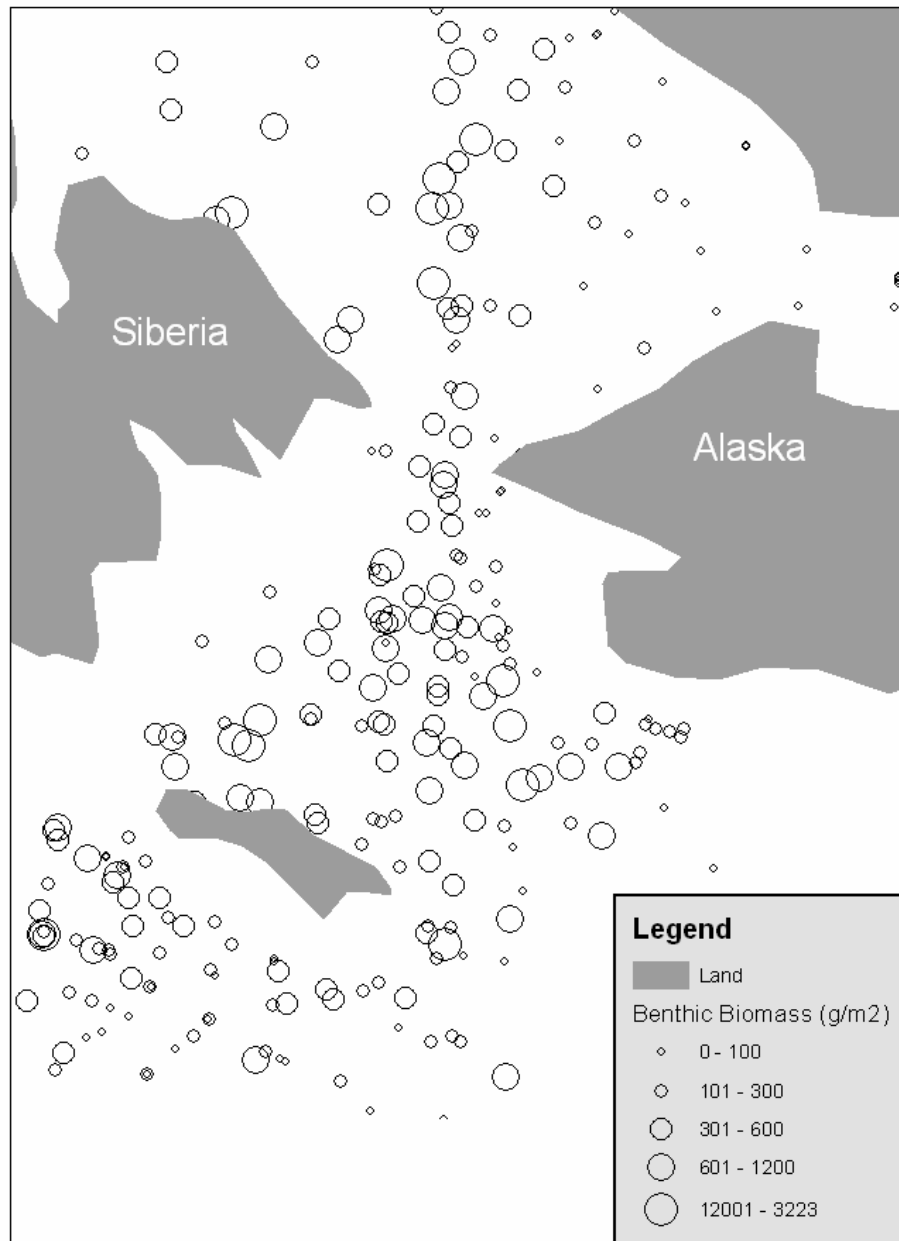smooth the small-scale variability prior to geostatistical interpolation.

Figure 11 – A preliminary look at the small-scale variability of the benthic biomass dataset focusing on the Bering Sea

The individual point samples of biomass were up-scaled using square bins of 1, 50, and 100 km in length to investigate the most appropriate scale for accurate yet sufficiently detailed interpolation. The aggregating of individual

samples into spatial bins was accomplished by first creating three grid feature classes with 1, 50, and 100 km length squares. Next, a *one to many* relationship was created to link each grid cell to the points that fall within its boundaries. The relationship was built using Visual Basic for Applications macros written in ArcGIS (Appendix C). All points were given a CellID corresponding to the cell's ObjectID and summary statistics were calculated for each grid cell based on the points within that grid cell. The summary statistics calculated were mean, standard deviation, and count.

The grid cells provide moving window statistics for the study area. If the grid cells are used as the input dataset for interpolation, the resulting prediction map will have less detail, but also less prediction error (variance) compared to a prediction surface generated using the exhaustive dataset. The goal is to balance these two qualities of the prediction map: local detail and overall appearance and accuracy (Isaaks and Srivastava 1999), a subjective task that depends on the final use for the prediction maps. In this case, the maps will serve as informational tools for the marine science community where large-scale trends are important. Loss of detail in the prediction results is tolerable in return for a smoother prediction surface that clearly shows large-scale trends.

Another advantage of the grid cells is to test the correlation between local mean and local variance (Isaaks and Srivastava 1999). If the local mean and variance are correlated, a phenomenon commonly referred to as a proportional effect, the information can be used to obtain unbiased estimates of standard

deviation values for predicted benthic biomass mean values. This is often a better

method for predicting standard deviation when compared to the prediction

standard error surface generated by the Geostatistical Analysis extension of

ArcGIS (Rossi 2002).

Plots of local mean with respect to local variance of the benthic biomass

grid datasets show that there is a positive relationship between biomass and

variability. Furthermore, the local variability is nearly as high as the local mean

for the study region. A linear relationship for the two variables using 100km grid

cells as bins with at least six samples in each bin showed very good correlation

between the two variables (correlation coefficient $\rho = 0.96$ for the thirty-nine

cells). The resulting equation can be used to calculate local variance ($\hat{\sigma}$) given

local mean (m).

$$\hat{\sigma} = 0.846m + 6.69 \tag{6}$$

Proportional Effect
n > 5
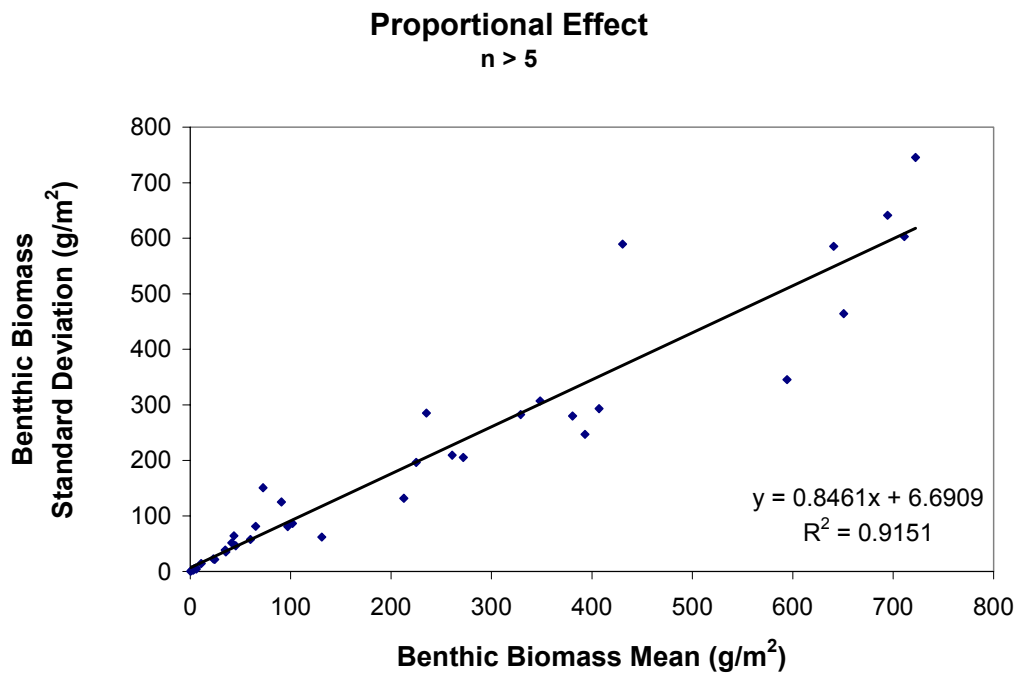
Figure 12 – Linear regression of local benthic biomass mean versus local benthic biomass standard deviation.

A linear regression that relates local mean with local variation for benthic biomass indicates that the variability of benthic biomass is nearly equal to the magnitude of the benthic biomass sample. This is evident in the Bering Sea where benthic biomass measurements varying by three orders of magnitude were sampled at nearly identical locations. Because the local variability of the dataset is significant, confidence intervals for predictions at benthic biomass point locations will be wide when the predicted mean biomass is high, no matter what model is used (Isaaks and Srivastava 1989).

4.1.5 Correlations with Other Variables

In addition to the benthic biomass samples, additional datasets were collected that might be related to benthic biomass densities to provide additional information that will improve interpolation results. The additional datasets are chlorophyll *a* integrated over the column depth, water temperature at sampling locations, and depth at sampling locations. Scatterplots of all combinations of benthic biomass with the other three variables were made for each scale (sample, 1, 50, and 100 km). The best results of each are presented in Figure 12.
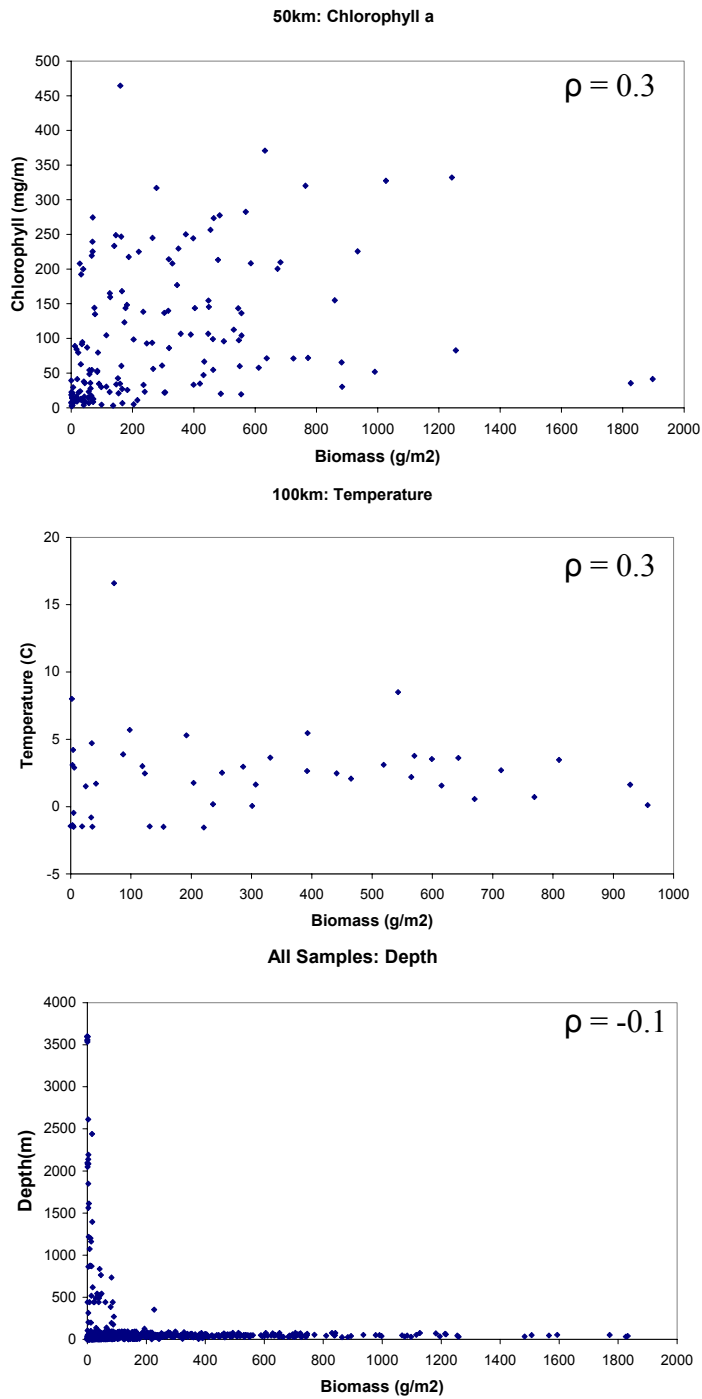
50km: Chlorophyll a

ρ = 0.3

100km: Temperature

ρ = 0.3

All Samples: Depth

ρ = -0.1

Figure 13 – Cross correlation between benthic biomass and (a) chlorophyll, (b) water temperature, and (c) depth.

Unfortunately, the scatterplots show little to no correlation between benthic biomass and integrated chlorophyll a, depth, or temperature. The correlation coefficients for all three were less than 0.4. The correlation coefficient only accounts for linear correlations, however, so it is possible that the data might be nonlinearly related. In addition, other factors not accounted for may be masking cross-correlations between biomass and the other variables. One suspicious combination is between biomass vs. depth. The scatterplot of biomass with depth shows that, if deeper than 200 to 300 m, the benthic biomass is not likely to exceed approximately 100 g/m^2. Because of this relationship and due to the fact that the cross-correlations will only help and never hurt the geostatistical model, all three cross correlations shown above were included in the co-kriging model (co-kriging to indication correlations with other variables are included within the model (Isaaks and Srivastava 1989)).

The potential to include correlated data for spatial interpolation can be very powerful if cheap data can be shown to be related to expensive data. While minimal correlation exists in the datasets used here, this approach should not be discounted. In fact, if additional resources were available, this might be a worthwhile place to start to refine the geostatistical model developed through this research.

To summarize, the findings of the Exploratory Spatial Data Analysis that will be used in the development of a geostatistical model are as follows.

1.) A 100 x 100 km grid of spatially averaged biomass values is used to reduce small-scale noise in the raw dataset.

2.) One cell of the gridded dataset is removed prior to geostatistical modeling. This cell has an average biomass value that is not representative of the spatially averaged benthic biomass at its location.

3.) The data are transformed using a Box-Cox function to produce a normal distribution.

4.) A cubic function is fit to the data prior to Ordinary Kriging to remove trends in the data. (Note: The model used for the geostatistical analysis, Geostatistical Analysts for ArcGIS, does not require the specification of exact coefficients for a detrending function, only the order of the function to be used. The fitting of the function and generation of coefficients is all done internally and hidden from the user.)

5.) Point samples of depth, water temperature, and chlorophyll a are included in the model because of possible cross-correlation with benthic biomass (although the correlation between biomass and the other variables appears to be minimal.)

4.2 EXPLORATORY TEMPORAL DATA ANALYSIS

The preceding discussion of the spatial distribution of benthos assumed the measurements taken were independent of the time they were measured. The

appropriateness of this assumption, and the general structure of the temporal

distribution of measurements, will be investigated in this section. The goals of

the exploratory temporal data analysis are to test assumptions made by the spatial

analysis (that the data are independent of collection date) and to ascertain

information necessary for the temporal statistical modeling.

4.2.1 Temporal Distributions

The yearly distribution shows that the majority of samples were collected

in the 1970s and subsequent voyages in the 1980s and 1990s have added to the

complete benthic dataset (Figures 14a). The monthly distribution shows that

nearly all of the samples were taken during the summer months (Figure 14b).

This summer-weighted monthly histogram is an important characteristic of the

dataset because it limits any seasonality effects that might exist.

4.2.2 Long-Term Temporal Trends

The complete dataset contains benthic biomass measurements from three

decades. If benthic biomass densities are to be considered independent of

collection date for the spatial analysis, there must not be any significant long term

trends in the overall dataset. On the other hand, if a significant long term trend

does exist in the dataset, while it may complicate the spatial interpolation, it may

also provide valuable information about how climate change is affecting the

benthic community. It is difficult to interpret the time series because the samples

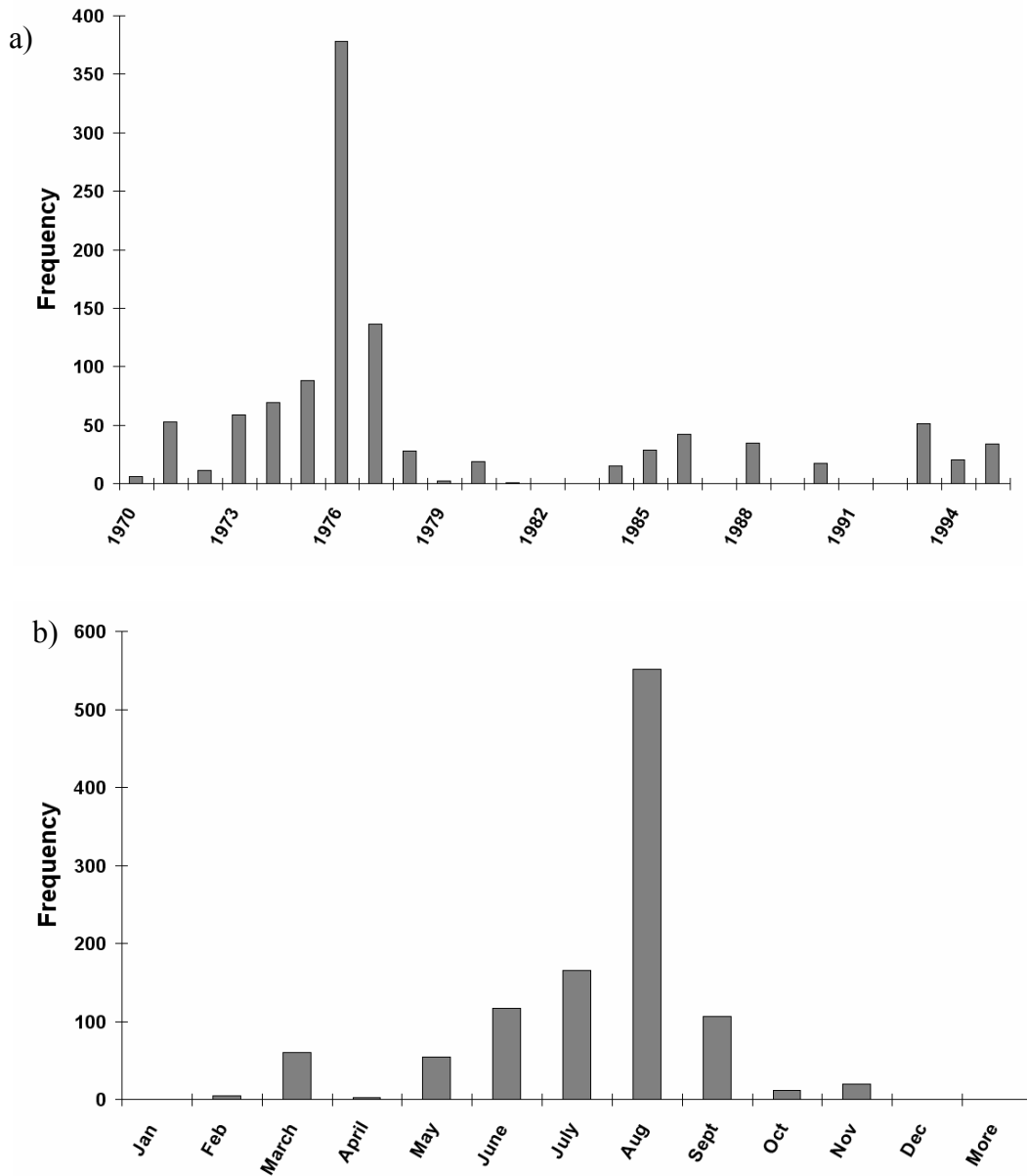collected for each year are not uniformly distributed over the sampling

Figure 14 – (a) The yearly distribution of biomass collection dates. (b) The monthly distribution of biomass collection dates

region (Figure 15). This means that, while average benthic biomass is lower in 1979, it is because the samples collected during this time were mainly in the Beaufort Sea instead of the Bering Sea.
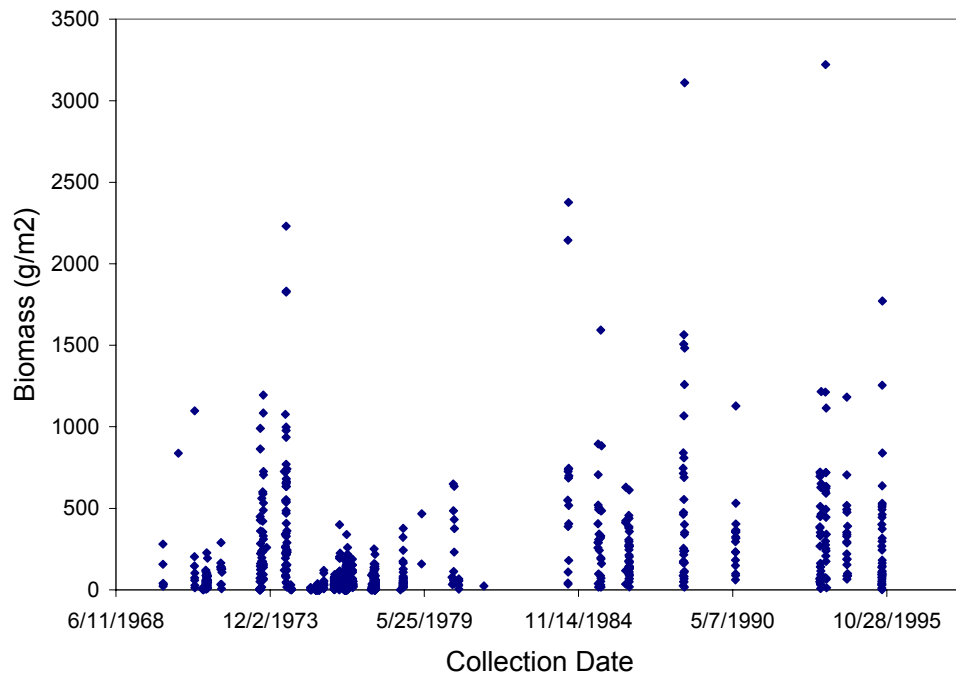
Figure 15 – Time series of benthic biomass measurements

In order to account for the difference in benthic biomass mean within the different seas when accounting for temporal trends, it is only appropriate to compare points within generally the same geographical location, because the main benthic biomass is not constant over the entire study area. Two locations were found where sufficient measurements from two or more decades exist so that temporal changes might be deduced. One of the areas is just north of St. Lawrence Island (southwest corner: 170°54'15"W 63°32'59"N, northeast corner 167°8'16"W 65°30'59"N) and the other is just south of the same island (southwest corner: 170°20'49"W 61°36'42"N, northeast corner 168°56'41"W 63°37'60N); both are 200km by 200km square regions (Figure 16). These two

44

areas will be used to detect if the benthic biomass density shows significant yearly change.
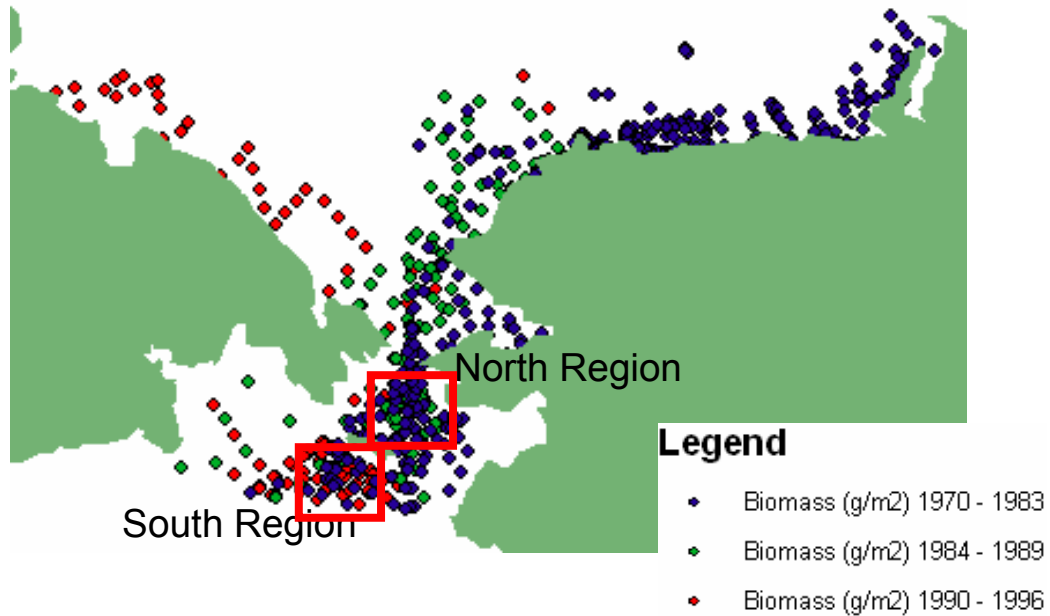
Figure 16 – The Two Comparison Areas

The northern comparison region contains twenty-eight samples from 1970-1984 along with thirty-nine samples from 1984-1990 and shows similar log-normal distributions as seen in the overall biomass dataset (Figure 17). A similar analysis was conducted for the southern comparisons region. The region contains twenty-one samples from 1970-1984 and thirty-three samples from 1984-1990 (Figure 18). Initial inspection suggested that the two temporal datasets are not as similar as those in the northern region, a possibility that will be discussed in the results section.

Figure 17 – The Northern Temporal Comparison Region

Figure 18 – The Southern Temporal Comparison Region

To summarize the findings of the exploratory temporal data analysis, two regions with overlapping benthic biomass samples from different decades were identified. Initial visual inspection of the two regions suggested that the northern region showed no change in benthic biomass from 1984 to 1996, but the southern region shows evidence of an increase in benthic biomass from 1970 to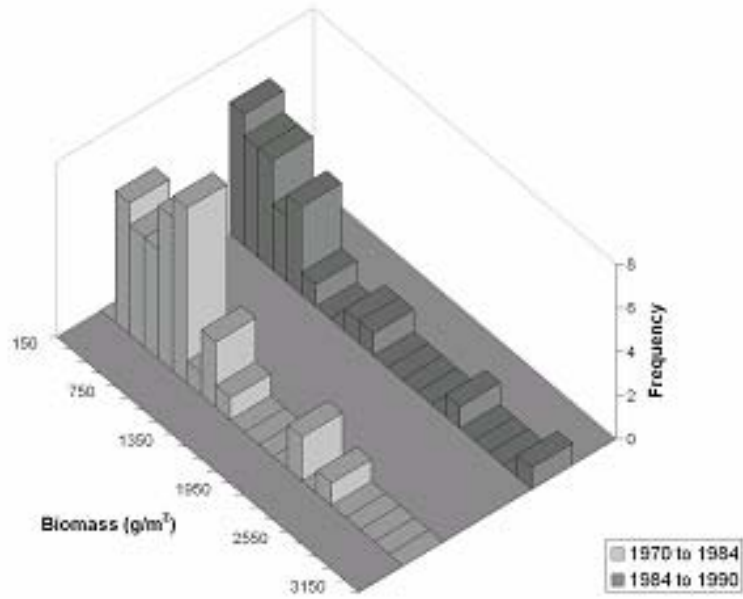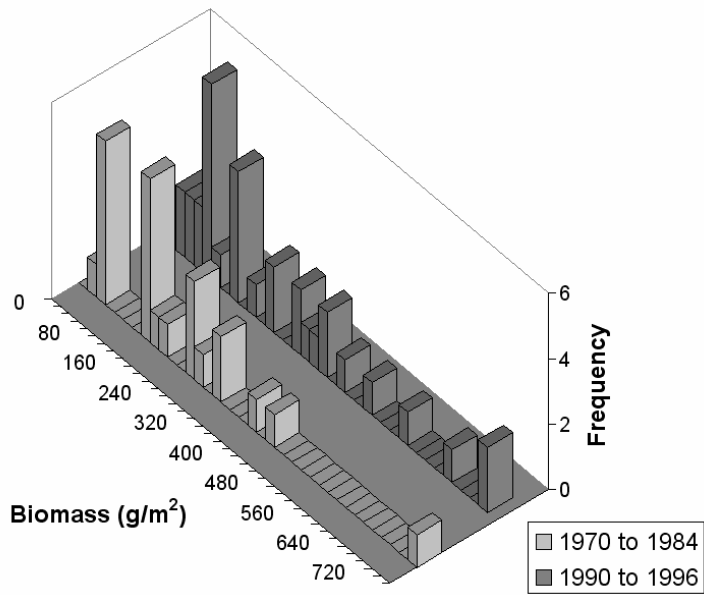 1996. I hypothesize that the population means of the two temporal datasets are different. This hypothesis will be tested following the generation of an interpolated surface (Section 5.2). The interpolated surface will better account for biases due to the spatial distribution of the sampling points.

4.3 MODELING THE AUTOCORRELATION: THE SEMIVARIOGRAMS

This section contains the critical part of geostatistical analysis: to construct a semivariogram that accurately represents the autocorrelation within the benthic biomass dataset. The usual first step in constructing a semivariogram is to make broad assumptions about the dataset to produce a first draft semivariogram. From there, different assumptions may be relaxed to observe their effects on the model. The broad assumptions generally used are 1) that the dataset has spatial continuity without significant noise so that a surface can be fit to the model, 2) that autocorrelation is dependent upon the distance separating points only and not the direction between two points (isotropy) and, 3) that the entire benthic biomass dataset is similar enough to be grouped into one semivariogram.

The semivariogram developed from these three assumptions has a fairly significant nugget of approximately 6.5 g^2/m^4 (Figure 19). This infers that if two samples were taken at precisely the same location, the difference between the samples would likely be ± 2.5 g/m^2. The autocorrelation of the dataset can be seen by the increasing semivariance with increasing separation distance. The range is the separation distance where the semivariance first becomes constant. In the benthic biomass dataset, the range is approximately 160 km. This means that each measurement has a radius of influence of 160 km – an important characteristic of the overall dataset when planning future data collection voyages. The semivariance at the separation distance equal to the range is referred to as the sill. It is an indication of the semivariance between two uncorrelated points and is approximately 15 g^2/m^4 for the benthic biomass dataset. The ratio of the sill to the nugget gives a measure of the autocorrelation present in the dataset.
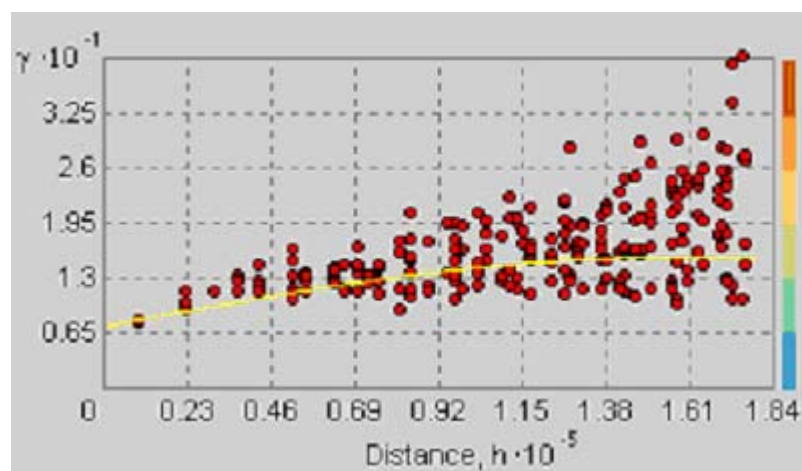
Figure 19 – First draft of semivariogram

Next I investigated the spatial continuity of the dataset at various spatial scales. The semivariogram of the samples at observation scale ($0.1m^2$) is the starting point for this analysis (Figure 19). As discovered through the exploratory spatial data analysis section, benthic biomass at this scale is highly variable, thus making the prediction standard error of the model relatively high. It might be advantageous, therefore, to "zoom out" from the raw dataset to get a better idea of the benthic biomass hotspots. Thus, three grids were generated, one with 1 km grid cells, one with 50 km grid cells, and the final with 100 km grid cells. Then the raw data samples were averaged for each grid cell. Semivariograms for all of the three up-scaled spatially averaged grids were computed (Figure 20) and compared to the observational scale semivariogram (Figure 19).
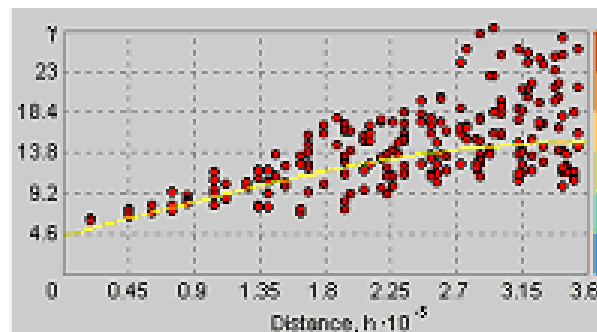
10 km x 10 km
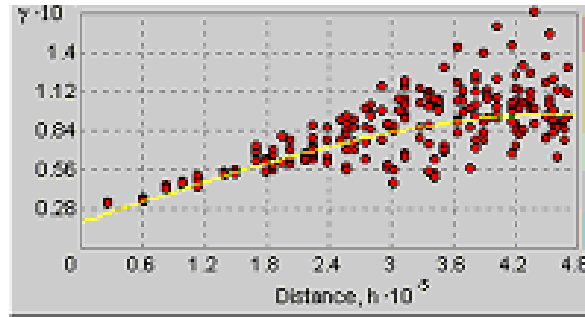
50 km x 50 km
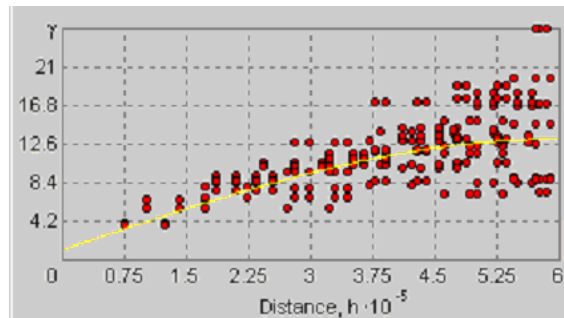
100 km x 100 km

Figure 20 – The up-scaled semivariograms

A summary of the important characteristics for each semivarigoram shows that the greater the spatially averaged prior to producing the semivariogram, the greater the datasets show an autocorrelation between separation distance and semivariance, as evident by in the increasing sill to nugget ratio (Table 2). The range of the data, or the radius of influence for each sample, also increases, partly because the samples on average are further apart and partly because the samples have been smoothed to remove small-scale variability.

	all benthos samples	10km x 10km grid dataset	50km x 50km grid dataset	100km x 100km grid dataset
Range km	153	360	470	590
Sill $(g/m^2)^2$	14.6	10.6	9.3	12.7
Nugget $(g/m^2)^2$	7.3	4.3	1.8	1.1
Sill/Nugget	2.0	2.5	5.2	11.5
RMS	243	248	185	147

Table 2 – Summary of up-scaled semivariograms

Choosing the best semivariogram from the four presented here is a subjective task. The spatially averaged grid data produce predictions with lower variance, but at the expense of a loss in information. One the other hand, the raw dataset surface contains the most information, but the variance of the predictions is often as high as the means of the predictions. The most appropriate dataset to use, therefore, depends on how the dataset will be used and how important accuracy is compared to local detail for that particular use. For this study, it was decided to use the 100 x 100 km grid cells because they best represent the large scale spatial trends and the location of large scale abnormalities in benthic biomass density. Although the 100 x 100 km scale dataset is used here, it would be possible to follow the same procedures with any of the other datasets, if one

thought the increase in local detail of the predictions to be of more value than the large scale trends.

The second assumption that was investigated is that the semivariance depends only on separation distance and not on separation direction. This idea that the correlation between points could be based on direction as well as distance is called anisotropy. One example of assuming anisotropic conditions is for modeling the dispersion of a pollutant in a groundwater aquifer. The concentration of the pollutant is likely to be more correlated in the direction of groundwater flow than in directions perpendicular to flow.

Based on this idea, it makes sense that benthic biomass might be anisotropic with a more significant correlation in the direction of general ocean currents. However, this research found that semivariograms with different search directions did not show any improvements on the omni-directional semivariogram. The exploratory spatial data analysis supported this finding by showing no significant preference for benthic biomass in a specific direction. Thus, the omni-directional semivariogram was kept. It should be noted that while there is not specific preference for directional correlation looking at the overall dataset, within particular seas where the direction of water flow is predominately in one direction, benthic biomass might be anisotropic, an idea that will be addressed later.

The final assumption implicit within the first draft semivariogram is that all benthic biomass data are of similar statistical distribution and relationship and, thus, it is appropriate to group all samples into one semivariogram. The exploratory spatial data analysis, however, suggested that the benthic biomass distribution in the Beaufort Sea off the northern coast of Alaska is different than the benthic biomass distribution in the Bering Sea. To investigate this assumption more fully, the entire benthic biomass dataset was divided by their respective Seas (Table 3).

	Beaufort Sea	Bering Sea	East Siberian Sea	Chukchi Sea
Count	416	340	27	73
Min	0.01	0.01	1.17	0.01
Max	377.4	3222.2	637.75	838
Mean	32.921	369.57	225.16	166.95
Std. Dev.	46.401	446.27	192.27	174.98
Skewness	3.1151	2.7849	0.601	1.4394
1st - Quartile	2.685	77.2	69.032	15.61
Median	18.07	231.07	138.78	134.06
3rd - Quartile	43.675	509.1	394.03	253.68

Table 3 – Summary statistics from the benthic biomass samples binned by Sea

Clearly, the benthic biomass population in the Beaufort Sea is distinct from that of the other seas. The mean benthic biomass for the Beaufort Sea was 33 g/m^2, which is significantly lower than the means of 370, 225, and 167 g/m^2 for the Bering, East Siberian, and Chukchi Seas, respectively. Following the development of a geostatistical model based on the exhaustive dataset, the

Beaufort Sea dataset is isolated from the overall dataset and a geostatistical model is developed for just this sea to better predict the spatial distribution of benthos.

4.4 SELECTING A GEOSTATISTICAL MODEL

The purpose of the exploratory data analysis conducted in the previous section was to aid in the selection of a geostatistical model for the prediction of benthic biomass at unmeasured locations. When selecting a model, there are three important guidelines to follow (Kitanidis 1993).

1) Choose the simplest model consistent with the data.

2) For practical estimation purposes, the right model is the one which represents available information about the structure of a spatial variable.

3) Make use of available information.

Based on these guidelines, the model selected at the outset of this research was an isotropic (direction independent) ordinary kriging model. Through the modeling process, simplifying assumptions were made and then relaxed to judge their relative impact on the prediction accuracy (see the exploratory data analysis section). It is important to understand that the modeling process is an iterative one where one starts with an exploratory data analysis and simple geostatistical model and then adjusts assumptions and parameters to minimize the model error (Figure 21).
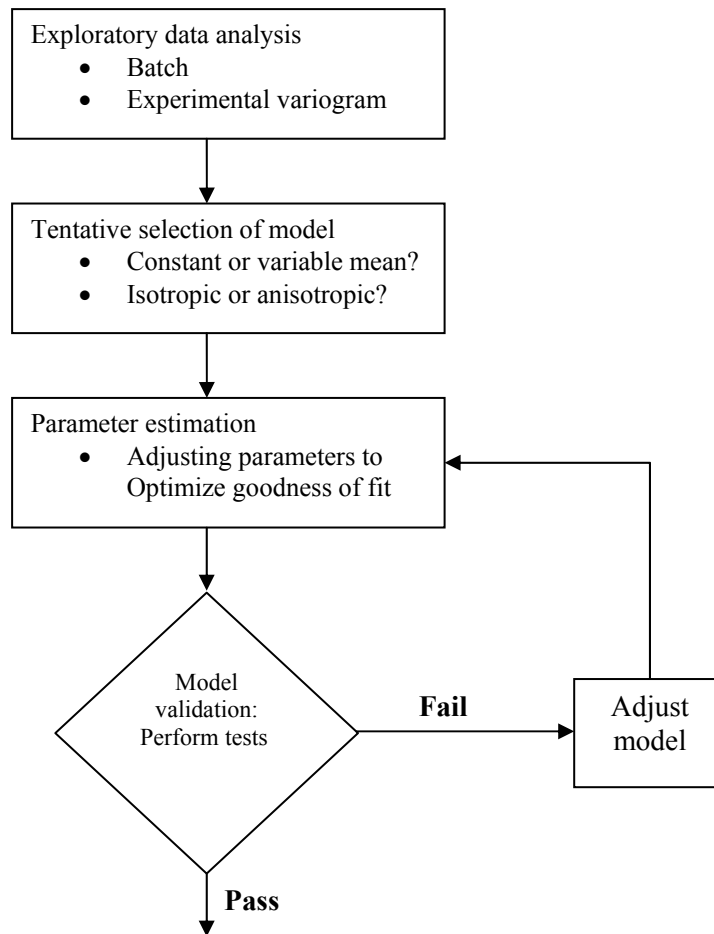
Figure 21 – Flowchart of geostatistical model development (Kitanidis 1993)

The criterion for model validation is typically the root mean square error (RMS) between the estimated and observed measurements. The Geostatistical Analyst extension of ArcGIS estimated measured locations by ignoring the measurement and using the model to predict what the value would be at that location. It is, therefore, not necessary to separate the original dataset into two sets: one for modeling and one for error analysis. The model assumptions and

parameters are adjusted to decrease the RMS to an acceptable level (or the lowest level possible). Once an acceptable RMS has been reached, the model is accepted.

4.5 PREDICTION AND STANDARD PREDICTION ERROR (PSE) SURFACES

Following the semivariance model development, it was possible to use the model to predict benthic biomass at all locations and, because geostatistics is a stochastic tool built from statistical theory, quantify the accuracy of all point predictions. These two quantities, estimated benthic biomass and prediction standard error, were generated as raster datasets by the geostatistical analyst and will be presented in the Results Chapter 5 of this report.

4.6 VISUALIZATIONS

One advantage of using GIS is that there exists a wealth of in-house tools for the visualization of spatial data in one, two, three, or even four dimensions. With these visualization abilities, it is possible to show the space and time distribution of benthos within the Western Arctic in an easily understood manner. Within the ArcGIS suite of programs is ArcScene, designed for viewing multiple data layers in three-dimensional space. Using ArcScene it is possible to raise the biomass values in the z-direction to show the magnitude of biomass measured at each location. Likewise, the continuous prediction surfaces can also be raised according to the biomass values for each cell to show the spatial distribution of predicted benthic biomass density.

To visualize the temporal trends in the benthic community in 4D, it is possible to extend ArcGIS to show how an attribute will change over time. One example is the Time Series Viewer which is part of the Arc Hydro Toolset. The Time Series Viewer allows ArcGIS to show changes in attribute value over time by repopulating an attribute value with time sequential values for each time step. For example, to show the movement of a storm over a river basin, the Time Series Viewer will, for each time step, change the precipitation value for each catchment. The result is an illusion of how the storm is moving over the basin. This Time Series Viewer could be used to show how benthic biomass is changing over time in a similar manner, if sufficient temporal data existed in a relatively uniform distribution of the study area.

Chapter 5: Results

This section will present the findings obtained from the geostatistical analysis of the benthic biomass measurements, with the spatial and temporal results presented separately. The main objectives of the spatial trend analysis were to 1) predict the continuous spatial distribution of benthic biomass and, 2) identify hot and cold spots where biomass is abnormally high or low. The main objective of the temporal trend analysis was an attempt to isolate any evidence of long term variations in benthic biomass density. Such evidence could provide significant insight into the impacts of observed climate change on benthic community structure.

5.1 SPATIAL TRENDS OF BENTHIC BIOMASS

The raw dataset of benthic biomass point measurements (Figure 22) provides some evidence of the spatial patterns of the benthic community. From this plot, it is clear that benthic biomass is lower in the Beaufort Sea and higher in the Bering Sea, but there is a significant amount of scatter (or noise) in this raw dataset due to a high variability of measurements in the Bering Sea, making the identification of trends difficult. Geostatistical techniques predict the mean and standard deviation of benthic biomass on a continuous scale for discrete point samples. If kriging, a geostatistical technique, is used to predict the benthic biomass distribution from this raw data, the variability (or prediction error) of the

kriging estimates will be quite significant and might dampen otherwise significant large scale trends. Thus, reducing the variability of the raw dataset by "zooming up" will lessen small scale variability in order to amplify large-scale trends.

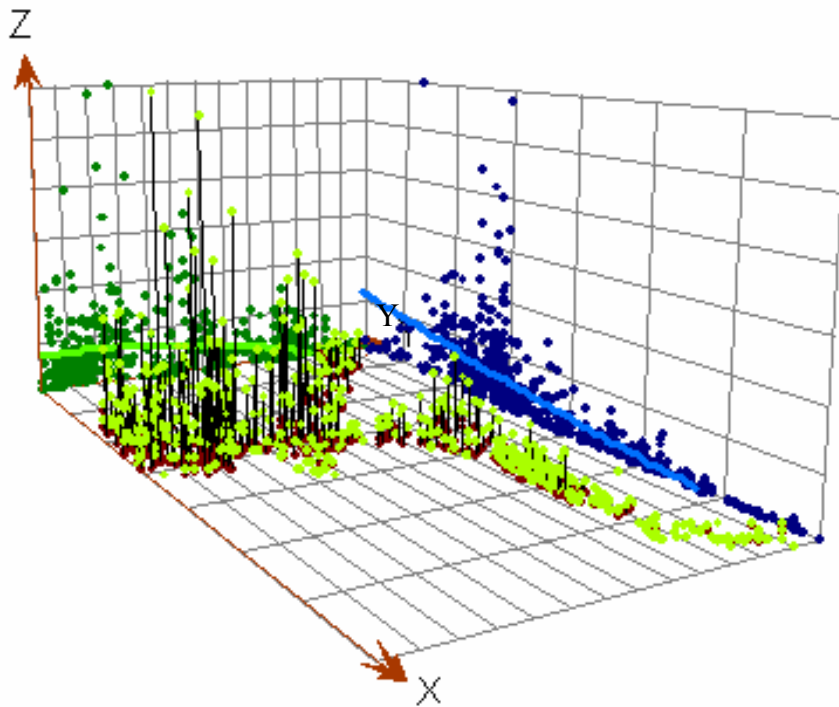**Figure 22 – 3D representation of benthic biomass point measurements. X is to the East, Y is to the North.**

In order to reduce the variability of the biomass sample, bins were used to spatially average local samples and thus provide an average value over a given area. The most obvious result of the spatial averaging over 100 km grid cells (Figure 23) is the lessening of variability within measurements at a single

location. The smoothing has traded smoothness for detail. It is also important to

note that one sample is significantly different than all other samples (as discussed

in section 4.1.1). The prediction surface was generated using this dataset,

ignoring the one abnormally high sample point.
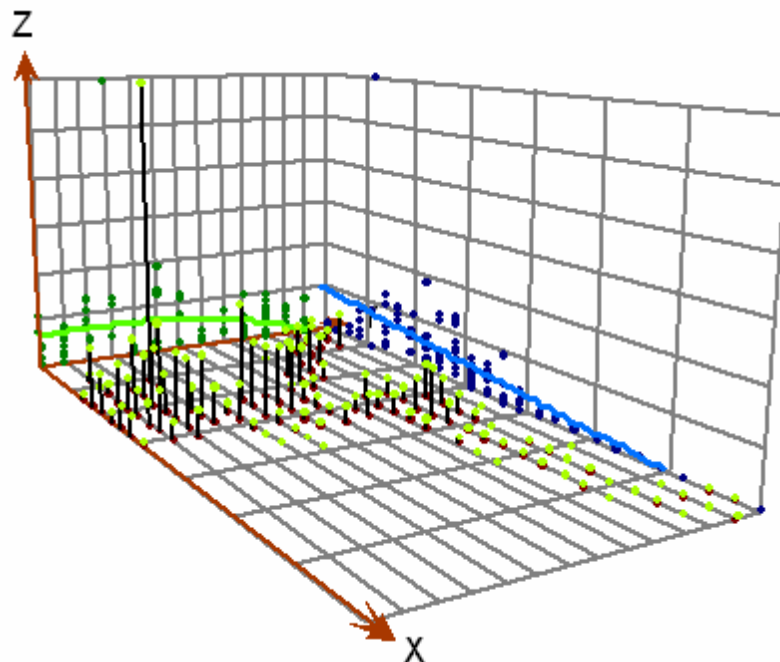
Figure 23 – 3D representation of benthic biomass 100km by 100km averaged data

The continuous surface of benthic biomass interpolated using ordinary

kriging and the spatially averaged data indicate that regions of highest biomass

occur in the Bering Sea. A semivariogram, which captures the autocorrelation

critical for generating the continuous surface, shows a radius of influence for each

sample of approximately 350 kilometers (Figure 25). This means that each point

is predicted as a weighted average of samples within 350 kilometers of that point.

Measurements closer to the prediction point are weighted more heavily than

distance points according to the function determined by the semivariogram.

Figure 24 – The contours of benthic biomass (g/m^2) in the Western Arctic Ocean based on geostatistical interpolation of the spatially averaged (100km by 100km) dataset.
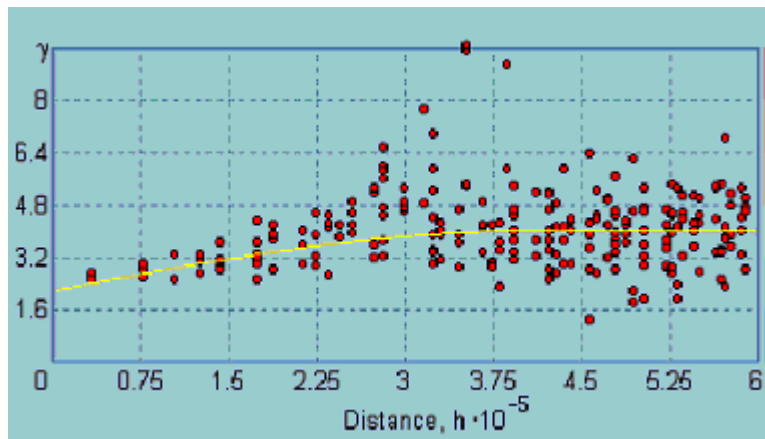
Figure 25 – The semivariogram model using to generate the continuous surface.
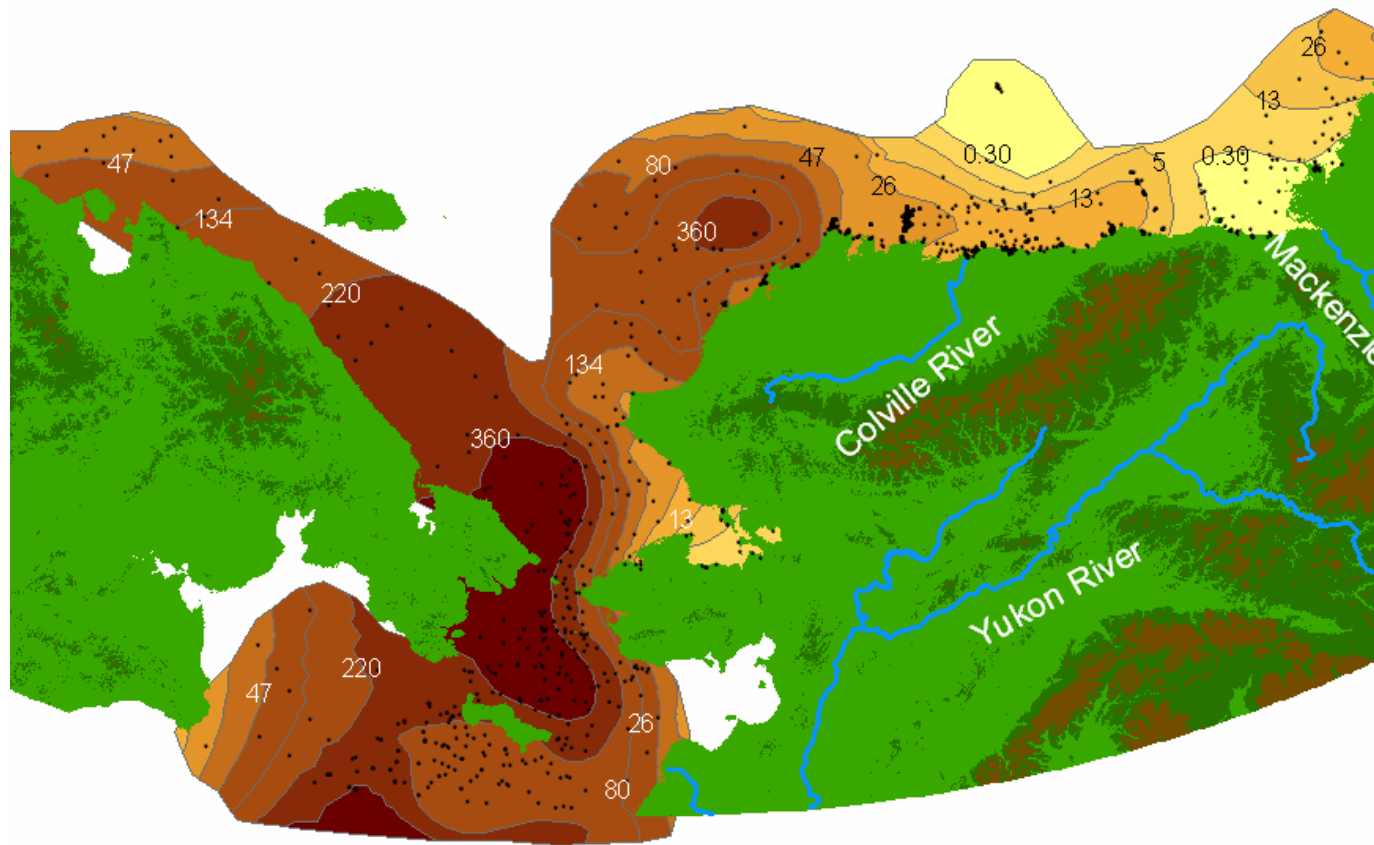
One potentially meaningful realization from these contours is the reduction in mean benthic biomass seemingly related to the outlets of major river basins. Both the Yukon and the MacKenzie Rivers have an apparent negative impact on benthos. This is due to the fresh water discharging from the rivers into the sea water and lowering the sea water's salinity.

Also significant is the high benthic biomass mean predictions in the Barrow Canyon, off the northwest coast of Alaska with predicted mean benthic biomass of above 360 g/m^2. This region utilized frequently by bowhead whales which are grazing on zooplankton; much of the phytoplankton produced here also falls directly to the seabed and provides food for the benthic community. These features are well defined by the geostatistical techniques.

The geostatistical model also produces predictions of the uncertainty associated with each estimate (Figure 24). These two parameters fully describe

the benthic biomass distribution at point samples modeled as a random variable.

Thus, the prediction mean and standard error can be used to generate a probability

density function (PDF) of benthic biomass at any point within the study area.

The geostatistical program used in this research does output prediction standard

errors, but these are biased unless the sample set used is multivariate normally

distributed. A sometimes better (less prone to biases) measure of standard error is

to use the proportional effect (relationship between mean and standard deviation)

documented in the raw dataset (Rossi 2002). Doing so allows one to calculate

standard errors based on the predicted mean values.

Using this approach, standard deviation and mean values were calculated

for six locations within the study region. Figure 26 shows the probability density

function for these six locations.   The distribution assumes a Box-Cox distribution

with λ=0.15 (as found for the raw benthic biomass dataset).   The probability

density functions can be used to calculate confidence intervals for benthic

biomass at any location. The plot shows what is expected: high variability in the

Bering Sea where biomass means are high and low variability in the Beaufort Sea

where means are low.
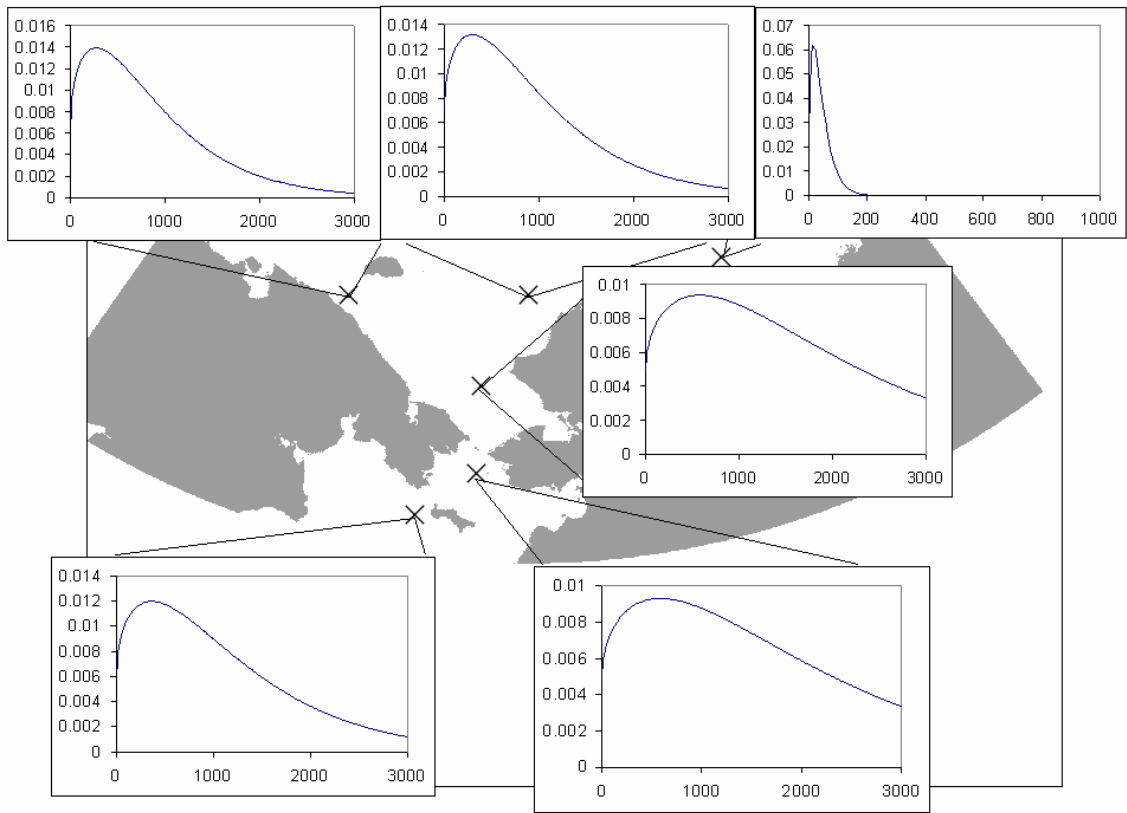
***Figure 26 – The Probability Density Function for benthic biomass predictions
at six locations***

As mentioned previous, the disadvantage of the 100 by 100 km

smoothing is the loss of detail in the spatial variability of benthos (Figure

27). Of particular importance is the Beaufort Sea where recorded highs in

benthic biomass have been lost due to the spatial averaging. This

information, however, is very important to the marine community. Taking

this observation into account suggests that, while smoothing over 100 km

may be appropriate for the Bering Sea where the samples have very high

variability, it may not be an appropriate scale for smoothing in the Beaufort

Sea. For this reason, and the reason stated in the prior section (that the

statistical distribution of biomass in the Beaufort is different than that in the

other seas), the Beaufort dataset was considered separately and a continuous

surface of benthic biomass samples was generated just for the Beaufort

biomass samples
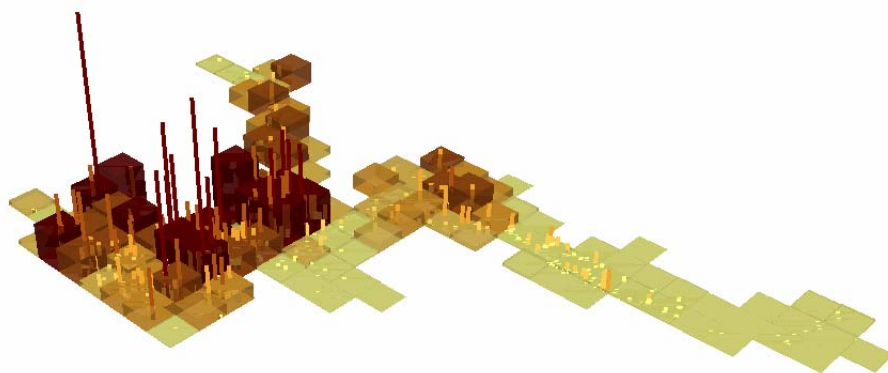
Figure 27 – Raw Samples vs. 100 x 100 km grid averages showing the loss of small scale detail

The biomass contours in the Beaufort Sea show local high levels of benthic biomass near the mouth of the Colville River (Figure 29). These hotspots of benthic biomass are products of the spikes in benthic biomass measured in the Sea. The interpolation procedure predicts a high density of benthic biomass in the northeastern corner of the map. This is due to the presence of a few samples with medium to high measurements of benthic biomass. There are insufficient samples in this region to have much confidence that the predictions in this region are accurate.

Figure 28 –Beaufort Sea benthic biomass spatial distribution

67

5.2 TEMPORAL TRENDS OF BENTHIC BIOMASS

To isolate temporal trends, it is first necessary to find regions with an adequate overlap of temporally varying measurements. The original biomass dataset was divided into three time periods to locate areas with overlapping temporal data. The time divisions (1970-1984, 1984-1990, and 1990-1996) were chosen because they represent the logical breaks in the collection dates. By using the 100 km x 1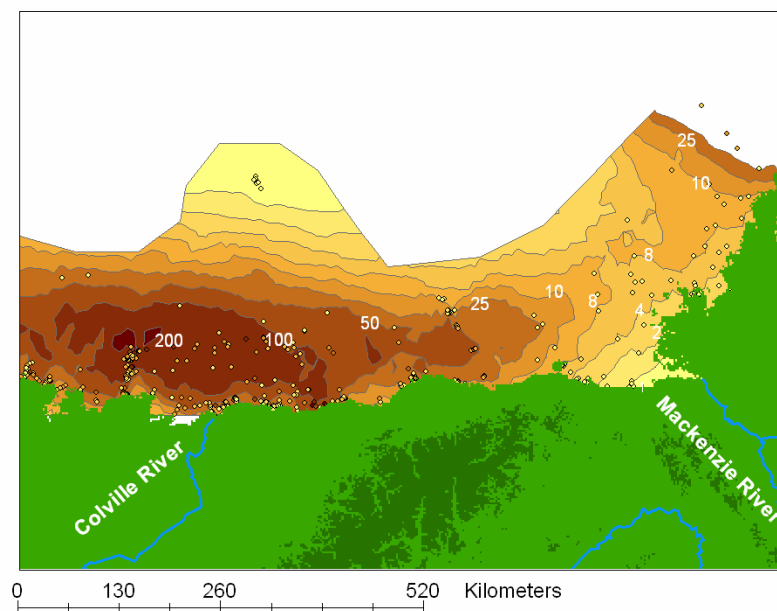00 km grid generated previously, cells for which more than three samples of any two of the time divisions fell were identified. This resulted in two comparison areas of 200 km x 200 km. The first was just south of St. Lawrence Island (Figure 29a) with twenty-eight samples from 1970-1984 along with thirty-nine samples from 1984-1990. The second was located just North of St. Lawrence Island (Figure 29b) and contained twenty-one samples from 1970-1984 and thirty-three samples from 1984-1990.

A geostatistical technique called block kriging was used to evaluate the change in biomass over time for the two regions. Block kriging is useful when the spatial distribution of samples is not as important as the overall measure of one sample within a fixed region (Isaaks and Srivastava 1989). Block kriging was used to estimate the total amount of benthic biomass within the square region for each time period. This procedure can be thought of as a weighted average over the area, but in actuality the average is found by interpolating a surface of

estimated benthic biomass predictions over the area, and then summing all predictions. This method accounts for the spatial distribution of the samples so that the clustering of the samples is less of a factor.

A semivariogram and corresponding model were generated for each time division within each comparison area using the methodology described in the spatial distribution section. The summary data about the raster file of predictions can provide an understanding of the benthic community over that time period within that region.

(a)

1970-1984
Mean = 743.33
S.D. = 414.17

1980-1989
Mean = 656.14
S.D. = 492.13

(b)

1970-1984
Mean = 187.36
S.D. = 37.90

1990-1996
Mean = 232.40
S.D. = 67.67
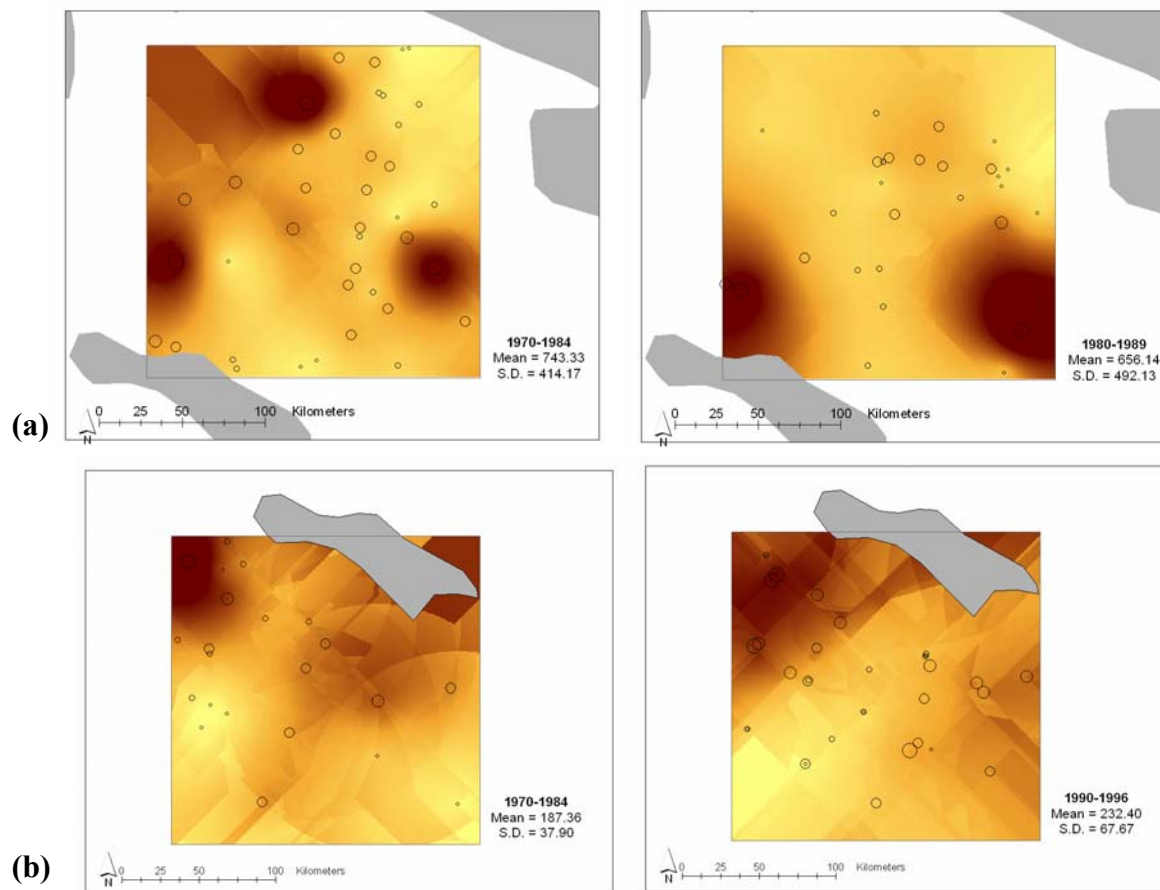
Figure 29 –The two comparison areas (a) just north of St. Lawrence Island in the Bering Sea and (b) just south of St. Lawrence Island. Darker brown indicates higher predicted benthic biomass (g/m2)

5.2.1 Northern Comparison Region

In the northern comparison region, there exists an overlapping of benthic biomass samples measured in the 1970s and in the 1980s. The samples from the two decades were divided and a prediction surface was calculated for each. These two surfaces represent and estimate the continuous distribution of benthos within the region for that decade. Based on the histogram of these interpolated benthic biomass predictions, there does not appear to be a significant difference between the benthic biomass distribution in the 1970s compared with the 1980s (Figure 30). A two-tailed, unequal variance t Test of all the raw sample points within the study region supported this observation. I concluded that mean benthic biomass did not change within the region from 1970 to 1980 (P-value = 0.825).

North Region Interpolated Benthic Biomass
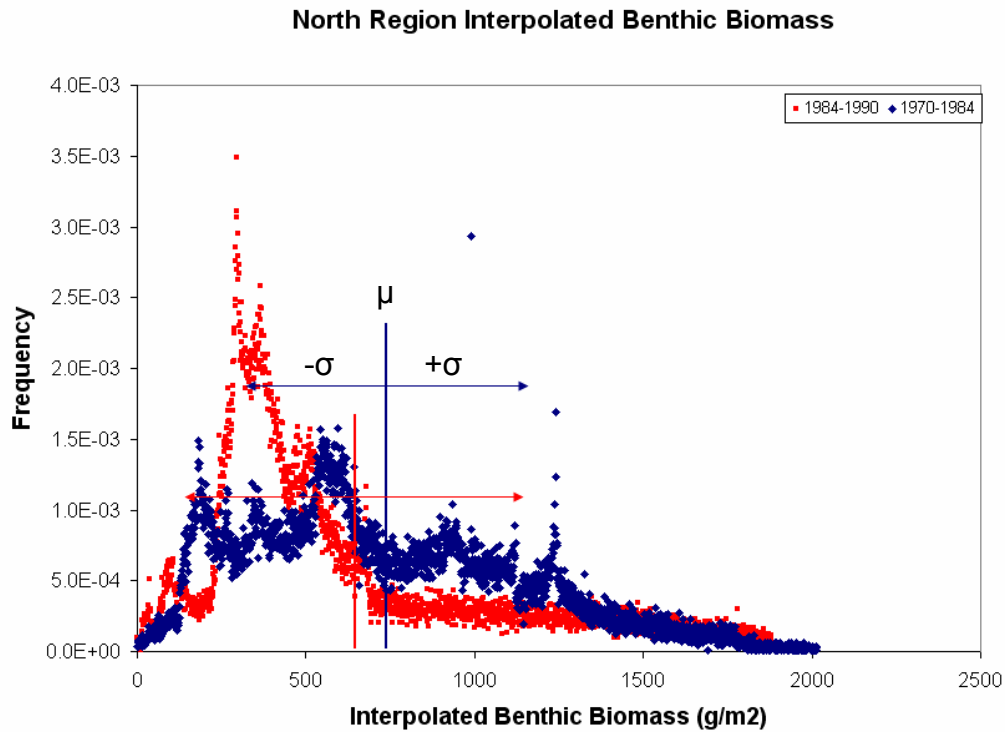
Figure 30 – Benthic community in the Northern Region in the 1970s and 1980s.

5.2.2 Southern Comparison Region

A similar procedure was conducted within the southern region with temporally overlapping data. Within this region, data from the 1970s were compared with data collected in the 1990s. The histogram of interpolated benthic biomass values shows that, unlike the northern study region, there appears to be a significant difference between the distributions of biomass in the 1970 when compared to the 1990 (Figure 31). A large percentage of the area (just under one-half) within the region has interpolated benthic biomass values greater than 250

g/m^2 for the 1990s. A two-tailed, unequal variance t-test of the raw datasets for each decade supported this observation. I concluded that there existed significant (P-value = 0.025) difference between the mean benthic biomass in 1970 and 1990, with the 1990 biomass being greater.
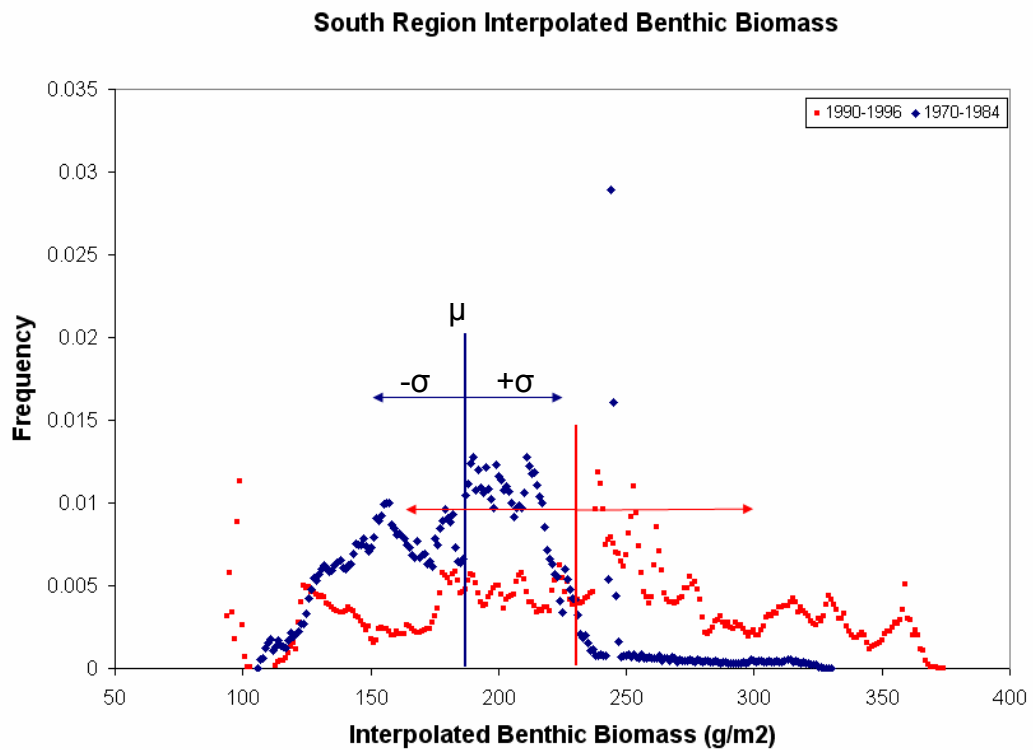

South Region Interpolated Benthic Biomass

Figure 31 – Benthic community in the Southern Region in the 1970s and 1990s.

Chapter 6: Conclusions

An extensive effort was put forth to gather all known benthic biomass measurements taken within the Western Arctic Ocean. The result is a retrospective (1970-1995) dataset of 1,093 benthic biomass point measurements. Data possibly correlated with benthic biomass were also collected to provide additional information regarding the spatial distribution of benthic biomass. These data include integrated chlorophyll *a*, depth, and water temperature. There were a total of 2,514 chlorophyll *a* samples integrated over the water column depth, 222 temperature readings at the time of sampling, and depth readings at all sample points (based on bathymetry data).

Geostatistical techniques were applied to these datasets to interpolate the continuous spatial distribution of benthos within the Western Arctic Ocean. The best model of benthic biomass over space showed the benthic population distribution as depicted in Figure 24. This figure was developed using up-scaled datasets of benthic biomass aggregated on100km square grid cells. The purpose of the up-scaled grid was to soften small scale variability in order to amplify large scale trends. One-hundred kilometers was chosen over 50km, 10km, and 1km grid sizes because it provided the appropriate balance of detail with smoothness necessary for the purposes of this study.

One of the benefits of the geostatistical model was the identification of abnormal benthic biomass densities at specific locations. For example, a benthic hotspot, that is where the density of biomass is unusually high in comparison to neighboring locations, was identified in Barrow Canyon off the Northwestern coast of Alaska. Benthic cold spots were identified at the outlets of the Mackenzie and Yukon Rivers, most likely due to low salinity. Providing reasons for these locations with abnormal benthic biomass, however, was not the within the scope of this study.

The interpolation of benthos was calculated using the correlation between separation distance and attribute value difference (autocorrelation) and the correlation between benthic biomass and integrated chlorophyll a, depth, and water temperature on the collection date. Scatterplots of benthic biomass and these other environmental datasets, however, showed little evidence of correlation. Therefore the primary information used to predict benthic biomass at unmeasured locations was the correlation between separation distances and attribute value differences as modeled by the semivariogram.

To investigate temporal trends within the data, two 200 km by 200 km locations were identified in which a significant number ($n \geq 20$) of benthic measurements existed from different decades. Both areas were in the Bering Sea, one was just north of St. Lawrence Island and the other just south. For each

location, the raw samples were divided into their respective decades. The southern region contained measurements from the 1970s and the 1990s and the northern region contained measurements from the 1970s and 1980s.

Using ordinary kriging, benthic biomass was interpolated for each time period within each region. Area averaging of the interpolated biomass surface (block kriging) for each decade at each of the two locations gave evidence of long term (decadal) upward trend for the southern region and no evidence of long term trend for the northern region. The ability to detect temporal trends in the benthic data, however, was weakened by insufficient time series of benthic biomass at specific locations.

One suggestion for future research is that there should be a more systematic and coordinated approach for the collection of biological, chemical, and physical data within the Western Arctic Ocean. Establishing specific locations where benthos can be measured repeated over time should provide more useful information in regards to the temporal trends of benthos. This research shows that benthic biomass is roughly correlated over 150 km (The range of the semivariogram for the raw sample points). Knowing this, it is recommended that future explorations collect samples on the range of 50-100 km. Anything denser will only provide unnecessary information; anything less dense will miss spatial correlations within benthic biomass

References

Broad, A.C. 1975-1980. Unpublished data available in : Interidal Organisms and Habitats (File 030). National Oceanographic Data Center. Washington, DC.

Burrough, P.A.. 1998. <u>Principles of Geographical Information Systems</u>. Oxford University Press, New York.

Burrough, P.A. 2001. *GIS and geostatistics: essential partners for spatial analysis.* Environmental and Ecological Statistics, **8**, 361-377.

Carey, A.G. 1971. Websec-71. Unpublished data.

Carey, A.G. 1976. OCSEAP. Unpublished data.

Carey, A.G., Boudrias, M.A., Kern, J.C., and Ruff, R.E., 1984, *Selected ecological studies on continental shelf benthos and sea ice fauna in the southwestern Beaufort Sea*. Final Report. Outer Continental Shelf Environmental Assessment Program. Fairbanks, AK, 23:1-164.

Cressie, N.A.C., 1991, <u>Statistics for Spatial Data</u>. John Wiley & Sons, Inc, New York.

Feder, H.M. 1986. OCSEAP. Unpublished data.

Feder, H.M., Day, R.H., Jewett, S.C., McCumby, K., McGee, S., and Schonberg, S.V., 1982, Infauna of the northeastern Bering and southeastern Chukchi Seas. Final Report. Outer Continental Shelf Envrionment Assessment Program. Fairbanks, AK, 120p.

Grebmeier, J.M., T.E. Whitledge, K. Aagaard, M. Bergmann, E.C. Carmack, L.A. Codispoti, D. Darby, K.H. Dunton, I.A. Melnikov, S. Moore, T. Takizawa, J.J. Walsh, P. Wassman, and P. Wheeler (eds). 2001. Arctic System Science Ocean-Atmosphere-Ice Interactions Western Arctic Shelf-Basin Interactions (SBI) Phase II Field Implementation Plan, SBI Project Office, The University of Tennessee, Knoxville, Tennessee, USA, 30 pp.

Isaaks, E.H., and R.M. Srivastava. 1989. <u>An Introduction to Applied Geostatistics</u>. Oxford University Press, New York.

Johnston, K., J.M. Ver Hoef, K. Krivoruchko, N. Lucas. 1999. <u>Using ArcGIS Geostatistical Analyst</u>. ESRI Press, Redlands, CA.

Jonsodottir, J.F., D.R. Maidment, and K.H. Dunton. 2000. *A GIS based analysis of the benthic community in the Western Arctic Ocean*, CRWR Online Report 2000-5.

Kitanidis, P.K.. 1993. Geostatistics, Chapter 20 in <u>Handbook of Hydrology</u>, Edited by D.R. Maidment, pp. 20.1-20.39, MacGraw-Hill, New York.

Maidment, D.R. (ed.). <u>Arc Hydro: GIS for Water Resources</u>. ESRI Press, Redlands, CA.

McDonald, A. 1999. <u>Building a Geodatabase</u>. ESRI Press, Redlands, CA.

Minami, M., M. Sakala, and J. Wrightsell. 1999. <u>Using ArcMap</u>. ESRI Press, Redlands, CA.

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

Rossi, R.E., 2002, Phone c*onversation with ArcGIS Geostatistical Analyst Extension Developer*.

Zeiler, M. 1999. <u>Modeling Out World</u>. ESRI Press, Redlands, CA.

Appendix A

Below is the computer code used in this research. All code was written in Visual Basic for Applications for either ArcGIS or EXCEL. Following the code is a list of the specific parameters used for the projected coordinate system.

Chlorophyll Data Processing (EXCEL)

```
Function cla_m3(station, depth, clam2)
'calculates the integrated chlorophyl for each depth
If station = station.Offset(1, 0) Then
    If station = station.Offset(-1, 0) Then
    'middle cells
        cla_m3 = (clam2.Offset(1, 0) + clam2) * (depth.Offset(1,
0) - depth) / 2
    Else
    'first cell
        cla_m3 = (clam2.Offset(1, 0) + clam2) * (depth.Offset(1,
0) - 0) / 2
    End If
Else
    If station = station.Offset(-1, 0) Then
    'last cell
        cla_m3 = 0
    Else
    'if only one cell
        cla_m3 = "only one value"
    End If
End If
End Function


Function SumNumber(station)
'creates a count column to be used to sum integrated chl numbers
Dim count As Integer
count = 1
If station = station.Offset(1, 0) Then
    If station = station.Offset(-1, 0) Then
    'middle cells
        SumNumber = 0
    Else
    'first cell
        Do While station = station.Offset(1, 0)
            count = count + 1
            Set station = station.Offset(1, 0)
```

```
        Loop
        SumNumber = count
    End If
Else
    If station = station.Offset(-1, 0) Then
    'last cell
        SumNumber = 0
    Else
    'if only one cell
        SumNumber = "only one value"
    End If
End If
End Function


Function sumcla(station, cla, count)
'sums the integrated chl values for each station
Dim sum
Sum = 0
If station = station.Offset(1, 0) Then
    If station = station.Offset(-1, 0) Then
    'middle cells
       sumcla = 0
    Else
    'first cell
        Do While count > 0
            sum = sum + cla
            Set cla = cla.Offset(1, 0)
            count = count - 1
        Loop
        sumcla = sum
    End If
Else
    If station = station.Offset(-1, 0) Then
    'last cell
        sumcla = 0
    Else
    'if only one cell
        sumcla = "only one value"
    End If
End If
End Function
End Sub


Sub DeleteRows()
'deletes extra station depths
Set currentCell = Worksheets("Sheet1").Range("B2")
```

```
Do While Not IsEmpty(currentCell)
    Set nextCell = currentCell.Offset(1, 0)
    If nextCell = currentCell Then
        If currentCell = currentCell.Offset(-1, 0) Then
            currentCell.EntireRow.Delete
        End If
    Else
        currentCell.EntireRow.Delete
    End If
    Set currentCell = nextCell
Loop
End Sub


Sub Insertdepths()
'to insert sonic depth values
Dim currentCell As Range, nextCell As Range, currentdepth As
Range, nextdepth As Range
Set currentCell = Worksheets("Sheet1").Range("B2")
Set currentdepth = Worksheets("sheet2").Range("D2")
Do While Not IsEmpty(currentCell)
    Set nextCell = currentCell.Offset(1, 0)
    If nextCell.Value <> currentCell.Value Then
        nextCell.EntireRow.Insert
        Worksheets("Sheet1").Range("A" & (currentCell.Row + 1)) =
currentCell.Offset(0, -1).Value
        Worksheets("Sheet1").Range("B" & (currentCell.Row + 1)) =
currentCell.Value
        Worksheets("Sheet1").Range("C" & (currentCell.Row + 1)) =
currentdepth.Value
        Worksheets("Sheet1").Range("D" & (currentCell.Row + 1)) =
0
        Set currentdepth = currentdepth.Offset(1, 0)
    End If
    Set currentCell = nextCell
Loop
End Sub
```

Appendix B

Specific Parameters of the Projected Coordinate System

Projected Coordinate System:
Name: Sphere_ARC_INFO_Lambert_Azimuthal_Equal_Area
  Alias:
  Abbreviation:
  Remarks:
Projection: Lambert_Azimuthal_Equal_Area
Parameters:
  False_Easting: 2000000.000000
  False_Northing: 4000000.000000
  Central_Meridian: -165.000000
  Latitude_Of_Origin: 90.000000
Linear Unit: Meter (1.000000)
Geographic Coordinate System:
  Name: GCS_Sphere_ARC_INFO
  Alias:
  Abbreviation:
  Remarks:
  Angular Unit: Degree (0.017453292519943299)
  Prime Meridian: Greenwich (0.000000000000000000)
  Datum: D_Sphere_ARC_INFO
   Spheroid: Sphere_ARC_INFO
    Semimajor Axis: 6370997.000000000000000000
    Semiminor Axis: 6370997.000000000000000000
    Inverse Flattening: 0.000000000000000000

X/Y Domain:
  Min X: -2324963.836265
  Min Y: -2794114.084286
  Max X: 6471129.181847
  Max Y: 6001978.933826
  Scale: 244.140625

M Domain:
  Min: 0.000000
  Max: 21474.836450

Scale: 100000.000000

Z Domain:
 Min: 0.000000
 Max: 21474.836450
 Scale: 100000.000000

Appendix C

Moving Window Statistics (ArcGIS)

```
'consolidate like attributes
Public Sub AssignAverageToCell()
  Dim pDoc As IMxDocument
  Dim pMap As IMap
  Dim pPLayer As IFeatureLayer
  Dim pALayer As IFeatureLayer
  Dim i As Long, j As Long, k As Long
  Dim ppro As IStepProgressor
  Dim dSum As Double, dAvg As Double
  Dim lAreaValField As Long, lPointValField As Long
  Dim sAreaValField As String, sPointValField As String
  Dim lAreaKeyField As Long, lPointKeyField As Long
  Dim sAreaKeyField As String, sPointKeyField As String

  'Specify field names
  sAreaValField = "AvgDepth_m" 'Where you will store the average
  sPointValField = "Estimated_Depth_m" 'the value for each point
  sAreaKeyField = "OBJECTID" 'Where you will store the average
  sPointKeyField = "CELLID50" 'the value for each point


  Set pDoc = ThisDocument
  Set pMap = pDoc.FocusMap

  Set pPLayer = pMap.Layer(0) 'point layer
  Set pALayer = pMap.Layer(1) 'area layer

  Dim pFCursor As IFeatureCursor
  Dim pAFeat As IFeature
  Set pFCursor = pALayer.Search(Nothing, False)
  Set pAFeat = pFCursor.NextFeature

  'Get the fields
  lAreaValField =
pALayer.FeatureClass.Fields.FindField(sAreaValField)
  lPointValField =
pPLayer.FeatureClass.Fields.FindField(sPointValField)
  lAreaKeyField =
pALayer.FeatureClass.Fields.FindField(sAreaKeyField)
  lPointKeyField =
pPLayer.FeatureClass.Fields.FindField(sPointKeyField)
```

```
  If lAreaValField = -1 Or lPointValField = -1 Or lAreaKeyField =
-1 Or lPointKeyField = -1 Then
    MsgBox "Invalid Field Name"
    Exit Sub
  End If

  'Get the editor
  Dim pEditor As IEditor
  Dim pUID As New UID
  pUID = "esriCore.Editor"
  Set pEditor = Application.FindExtensionByCLSID(pUID)
  pEditor.StartOperation

  'Status bar
  Set ppro = Application.StatusBar.ProgressBar
  Dim pTrackCancel As ITrackCancel 'you can hit Esc to quit
  Set pTrackCancel = New CancelTracker
  With ppro
    .MinRange = 0
    .MaxRange = pALayer.FeatureClass.FeatureCount(Nothing)
    .StepValue = 1
    .Position = 0
    .Show
  End With

  Dim pQF As IQueryFilter
  Dim pPCursor As IFeatureCursor
  Dim pPFeat As IFeature
  Dim pRelatedFeats As esriCore.ISet
  Dim lNumFeats As Long
  Dim lID As Long

  Do Until pAFeat Is Nothing
    If Not pTrackCancel.Continue Then
      ppro.Message = "Cancelling"
      pEditor.StopOperation "bleh"
      ppro.Hide
      Exit Sub
    End If

    dSum = 0
    lNumFeats = 0
    'Get the ID of the current area
    If Not IsNull(pAFeat.Value(lAreaKeyField)) Then
      lID = pAFeat.Value(lAreaKeyField)
      'Create the queryfilter
```

```
        Set pQF = New QueryFilter
        pQF.WhereClause = sPointKeyField & " = " & lID
        'Get the related features and sum
        Set pPCursor = pPLayer.Search(pQF, False)
        Set pPFeat = pPCursor.NextFeature
        Do Until pPFeat Is Nothing
          If Not IsNull(pPFeat.Value(lPointValField)) And
pPFeat.OID <> 2514 And pPFeat.OID <> 2499 Then
            dSum = dSum + pPFeat.Value(lPointValField)
            lNumFeats = lNumFeats + 1
        pAFeat.Store
          End If
          Set pPFeat = pPCursor.NextFeature
          pAFeat.Store
        Loop
      End If

      If lNumFeats > 0 Then
        dAvg = dSum / lNumFeats
      Else
        dAvg = 0
      End If

      pAFeat.Value(lAreaValField) = dAvg 'lNumFeats


      Set pAFeat = pFCursor.NextFeature
      'Progress bar
      ppro.Step
      j = j + 1
      ppro.Message = j & " of " & ppro.MaxRange
    Loop

  ppro.Message = "Finishing Edits (please wait)"
  pEditor.StopOperation ("bleh")

  ppro.Hide
End Sub

'spatial overlap
Public Sub SpatialOverlap()
  Dim pDoc As IMxDocument
  Dim pMap As IMap
  Dim pPLayer As IFeatureLayer
  Dim pALayer As IFeatureLayer

  Set pDoc = ThisDocument
  Set pMap = pDoc.FocusMap
```

```
Set pPLayer = pMap.Layer(0)
Set pALayer = pMap.Layer(1)

Dim pFCursor As IFeatureCursor
Dim pAFeat As IFeature
Dim pSF As ISpatialFilter
' assign cell value to points
Set pFCursor = pALayer.Search(Nothing, False)
Set pAFeat = pFCursor.NextFeature

Dim pEditor As IEditor
Dim pUID As New UID
pUID = "esriCore.Editor"
Set pEditor = Application.FindExtensionByCLSID(pUID)
pEditor.StartOperation

Dim pPCursor As IFeatureCursor
Dim pPFeat As IFeature
Dim j As Long

'Status bar
Dim ppro As IStepProgressor
Set ppro = Application.StatusBar.ProgressBar
Dim pTrackCancel As ITrackCancel 'you can hit Esc to quit
Set pTrackCancel = New CancelTracker
With ppro
  .MinRange = 0
  .MaxRange = pALayer.FeatureClass.FeatureCount(Nothing)
  .StepValue = 1
  .Position = 0
  .Show
End With


Do Until pAFeat Is Nothing
  Set pSF = New SpatialFilter
  With pSF
  Set .Geometry = pAFeat.Shape
    .SpatialRel = esriSpatialRelIntersects
  End With

  Set pPCursor = pPLayer.Search(pSF, False)
  Set pPFeat = pPCursor.NextFeature
  Do Until pPFeat Is Nothing
    pPFeat.Value(12) = pAFeat.Value(0)
    pPFeat.Store
    Set pPFeat = pPCursor.NextFeature
  Loop
```

```
      Set pAFeat = pFCursor.NextFeature

      'Progress bar
      ppro.Step
      j = j + 1
      ppro.Message = j & " of " & ppro.MaxRange
   Loop

   ppro.Message = "Finishing Edits (please wait)"
   pEditor.StopOperation ("bleh")

   ppro.Hide

   pEditor.StopOperation ("bleh")
End Sub
```