



Nonparametric estimation of semivariogram functions  
by John Steven Cherry

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in  
Statistics

Montana State University

© Copyright by John Steven Cherry (1994)

Abstract:

Semivariogram functions are used by geostatisticians to describe the correlation structure of spatial random variables. Valid semivariogram functions must be conditionally negative definite. To ensure this restriction when estimating semivariogram functions, investigators typically fit a valid model to data in the form of a sample semivariogram. Two authors (Shapiro and Botha, 1991) suggested a nonparametric method of estimating semivariogram functions, but did not compare their method to the more traditional methods of fitting these functions. The method was evaluated using simulated data. The fits obtained using the nonparametric method were compared to nonlinear least squares fits obtained using four parametric models (exponential, Gaussian, rational quadratic, and power) The comparisons were made using the integrated squared errors of the resulting fits. In addition, a method of using the nonparametric method to estimate the variance (sill) of second order stationary random processes is described and evaluated. The nonparametric estimator always resulted in fits that were as good as those obtained using parametric models as measured by integrated squared error. The variance estimator as initially proposed resulted in biased and highly variable estimates, but a penalized fitting procedure resulted in estimates that were essentially unbiased and less variable than estimates obtained using the parametric models. This penalized estimation procedure did not affect the quality of the fitted functions. The nonparametric method is faster and easier to use, and more objective than parametric methods.

**NONPARAMETRIC ESTIMATION OF  
SEMIVARIOGRAM FUNCTIONS**

by

John Steven Cherry

A thesis submitted in partial fulfillment  
of the requirements for the degree

of

Doctor of Philosophy

in

Statistics

**MONTANA STATE UNIVERSITY**  
Bozeman, Montana

January 1994

NONPARAMETRIC ESTIMATION OF SEMIVARIOGRAM  
FUNCTIONS

John Steven Cherry

Advisor: William F. Quimby, Ph.D.

Montana State University  
1994

Abstract

Semivariogram functions are used by geostatisticians to describe the correlation structure of spatial random variables. Valid semivariogram functions must be conditionally negative definite. To ensure this restriction when estimating semivariogram functions, investigators typically fit a valid model to data in the form of a sample semivariogram. Two authors (Shapiro and Botha, 1991) suggested a nonparametric method of estimating semivariogram functions, but did not compare their method to the more traditional methods of fitting these functions. The method was evaluated using simulated data. The fits obtained using the nonparametric method were compared to nonlinear least squares fits obtained using four parametric models (exponential, Gaussian, rational quadratic, and power). The comparisons were made using

the integrated squared errors of the resulting fits. In addition, a method of using the nonparametric method to estimate the variance (sill) of second order stationary random processes is described and evaluated. The nonparametric estimator always resulted in fits that were as good as those obtained using parametric models as measured by integrated squared error. The variance estimator as initially proposed resulted in biased and highly variable estimates, but a penalized fitting procedure resulted in estimates that were essentially unbiased and less variable than estimates obtained using the parametric models. This penalized estimation procedure did not affect the quality of the fitted functions. The nonparametric method is faster and easier to use, and more objective than parametric methods.

© COPYRIGHT

by

John Steven Cherry

1994

All Rights Reserved

D378  
C4245

## APPROVAL

of a thesis submitted by

John Steven Cherry

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style and consistency and is ready for submission to the College of Graduate Studies.

1-26-94  
Date

William F. Quinby  
Chairperson, Graduate Committee

Approved for the Major Department

Jan 26, 1994  
Date

John Lund  
Head, Major Department

Approved for the College of Graduate Studies

2/4/94  
Date

Ed Brown  
Graduate Dean

**STATEMENT OF PERMISSION TO USE**

In presenting this thesis in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this thesis is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this thesis should be referred to University Microfilms International, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation for sale in and from microform or electronic format, along with the right to reproduce and distribute my abstract in any format in whole or in part."

Signature John Steven Chung

Date 1/26/94

## ACKNOWLEDGEMENTS

There were many individuals who contributed to the successful completion of this dissertation. In particular, I would like to acknowledge Dr. William Quimby, Dr. Jeff Banfield, Dr. Robert Boik, and Dr. Martin Hamilton. These four individuals were all members of my graduate committee, but also contributed much more than that to my education and training as a statistician. They were always supportive, and always available.

I also would like to thank Dr. Curt Vogel and Dr. John Lund of the Mathematical Sciences Department. Although they were not members of my committee, they gave freely of their time when I encountered problems in their areas of expertise. Dr. Vogel, in particular, helped me over a major obstacle.

Thanks are due to fellow graduate student Tim Carlson. His knowledge of the computing system and software used in the Department of Mathematical Sciences was invaluable. Tim, Don Daly, Bob Hoar, Sunil Tiwari, and Mike Sanford all were good friends who helped as only friends can help.

Finally, I wish to acknowledge my wife Marion. She has spent a good deal of time watching me go to school. I owe her a lot.

## TABLE OF CONTENTS

LIST OF TABLES .....	vii
LIST OF FIGURES .....	ix
ABSTRACT .....	xiv
1. INTRODUCTION .....	1
Spatial Stochastic Processes .....	4
Properties of Variogram and Covariance Functions .....	6
Estimation of Semivariogram Functions .....	8
Parametric Semivariogram Models .....	10
Some Isotropic Semivariogram Models .....	11
Fitting Semivariogram Models .....	15
Preview .....	18
2. THE SHAPIRO-BOTHA METHOD .....	19
Spectral Representation .....	20
The Fitting Procedure .....	23
3. FITTING NONPARAMETRIC SEMIVARIOGRAMS .....	26
Implementation of the SB Method .....	27
Node Selection .....	28
How Should the Nodes be Selected? .....	31
How Many Nodes are Enough? .....	32
Does the Saturation Lead to Overfitting? .....	34
Comparison of Fits .....	37
4. ESTIMATION OF THE SILL .....	51
The Sill .....	51
Stabilizing the Sill Estimate .....	64
Choosing the Penalty Term .....	65
5. ASYMPTOTICS .....	81

6. EXAMPLES .....	92
Example 1 .....	92
Example 2 .....	93
Example 3 .....	96
7. FUTURE RESEARCH NEEDS .....	99
Estimation of the Nugget Effect .....	99
Estimation of the Range .....	104
Sample Semivariogram as an Estimate .....	105
MINQU Estimation .....	109
$L_1$ Fitting .....	111
8. SUMMARY AND CONCLUSIONS .....	113
Summary .....	113
Implementation .....	114
REFERENCES CITED .....	116
APPENDICES .....	122
APPENDIX A SIMULATION OF DATA .....	123
APPENDIX B SOFTWARE USED .....	125

## LIST OF TABLES

Table	Page
1. The minimum, 25th percentile, median, 75th percentile, and maximum of the residual norms from 100 fits using four different selections of nodes. The simulated data are from an exponential model with sill and range of 10. ....	34
2. The minimum, 25th percentile, median, 75th percentile, and maximum ISE values from nonparametric and parametric semivariogram models fit to 100 simulated data sets. In each case the true semivariogram model was exponential with no nugget effect and a sill of 10. The ranges varied as indicated. ....	46
3. The minimum, 25th percentile, median, 75th percentile, and maximum ISE values from nonparametric and parametric semivariogram models fit to 100 simulated data sets. The true semivariogram model was a mixture of rational-quadratic and hole effect models with no nugget effect and a sill of 14. ....	48
4. The minimum, 25th percentile, median, 75th percentile, and maximum for sill estimates from five different simulations. In each case the true semivariogram model was exponential with no nugget effect and a sill of 10. The ranges varied as indicated. ....	58
5. Estimated mean squared error of sill estimates from simulated data sets for each model-range combination. The true sill is 10. ....	59
6. Means of the sill estimates from simulated data sets for each model-range combination. The true sill is equal to 10. ....	59
7. The minimum, 25th percentile, median, 75th percentile, and maximum for the sill estimates from the rational quadratic-hole effect model. The true sill value is 14. ....	60

8. Estimated mean squared error and means for sill estimates from a rational quadratic-hole effect model with true sill value of 14. (*rq-rational quadratic*). . . . . 62
9. Mean sill estimates and their variances from unpenalized and penalized fits at the estimated point of maximum curvature of plot of mean residual norm versus  $\log_{10}(\lambda)$  values. . . . . 71

## LIST OF FIGURES

Figure	Page
1. Examples of six parametric semivariogram models. ....	14
2. Semivariogram cloud with classical method of moments sample semivariogram estimator. ....	29
3. Two different nonparametric semivariograms based on two different collections of nodes. The solid curve was produced by the set of functions in the middle panel. The dashed curve was produced by the set of functions in the bottom panel. ....	30
4. Three different fits to simulated data. The solid line is based on 100 nodes equispaced in $(0, 4]$ and 100 nodes equispaced in $[4.16, 20]$ . The dashed line is based on 200 nodes equispaced in $(16, 20]$ , and the dotted line is based on 200 nodes equispaced in $(0, 4]$ . ....	33
5. This figure shows increasing <i>smoothness</i> of $1 - \Omega_d(x)$ for $d = 1$ (top panel), $d = 2$ (second panel), $d = 3$ (third panel), and $d = \infty$ (bottom panel). ....	38
6. Boxplots of integrated squared errors from nonparametric (SB) and parametric models (exponential, Gaussian) fit to 100 simulated data sets based on exponential model with sill of 10 and range of 2. ....	41
7. Boxplots of integrated squared errors from nonparametric (SB) and parametric models (exponential, Gaussian, rational quadratic, and power) fit to 100 simulated data sets based on exponential model with sill of 10 and range of 6. <i>rq=rational quadratic model</i> . ....	42
8. Boxplots of integrated squared errors from nonparametric (SB) and parametric models (exponential, Gaussian, rational quadratic, and power) fit to 100 simulated data sets based on exponential model with sill of 10 and range of 10. <i>rq=rational quadratic model</i> . ....	43

9. Boxplots of integrated squared error from nonparametric (SB) and parametric models (exponential, Gaussian, rational quadratic, and power) fit to 100 simulated data sets based on exponential model with sill of 10 and range of 14. *rq=rational quadratic model*. . . . . 44
10. Boxplots of integrated squared error from nonparametric (SB) and parametric models (exponential, Gaussian, rational quadratic, and power) fit to 100 simulated data sets based on exponential model with sill of 10 and range of 18. *rq=rational quadratic model*. . . . . 45
11. Boxplots of integrated squared error from nonparametric (SB) and parametric models (exponential, Gaussian, rational quadratic, and power) fit to 100 simulated data sets based on a mixture of rational quadratic and hole effect models. *rq=rational quadratic model*. . . . . 49
12. Boxplots of sill estimates from nonparametric (SB) and parametric models (exponential, Gaussian, and rational quadratic) fit to simulated data sets based on exponential model with sill of 10 and range of 2. . . . . 53
13. Boxplots of sill estimates from nonparametric (SB) and parametric models (exponential, Gaussian, and rational quadratic) fit to simulated data sets based on an exponential model with sill of 10 and range of 6. . . . . 54
14. Boxplots of sill estimates from nonparametric (SB) and parametric models (exponential, Gaussian, and rational quadratic) fit to simulated data sets based on an exponential model with sill of 10 and range of 10. . . . . 55
15. Boxplots of sill estimates from nonparametric (SB) and parametric models (exponential, Gaussian, and rational quadratic) fit to simulated data sets based on an exponential model with sill of 10 and range of 14. . . . . 56

16. Boxplots of sill estimates from nonparametric (SB) and parametric models (exponential, Gaussian, and rational quadratic) fit to simulated data sets based on an exponential model with sill of 10 and range of 18. .... 57
17. Boxplots of sill estimates from nonparametric (SB) and parametric models (exponential, Gaussian, and rational quadratic) fit to 100 simulated data sets based on a rational quadratic-hole effect model. The true sill is 14. .... 61
18. Two different nonparametric semivariogram fits to a sample semivariogram based on two different node collections. .... 63
19. Results of penalized fitting procedure on a simulated data set. Upper panel shows a plot of sill estimates versus the residual norm for 45 values of  $\lambda$ . Lower panel shows a plot of sill estimates and residual norms versus 45 values of  $\log_{10}(\lambda)$ . .... 66
20. This plot shows the relationship between the average of the sill estimates and the average of the residual norms determined from fits to 1100 data sets. Each point in the plot corresponds to one of 45 values of  $\lambda$  from  $10^{-9}$  (upper left) to 100 (lower right). The true semivariogram model is exponential with sill and range of 10. .... 68
21. This plot shows the relationship between the average of the sill estimates and the average of the residual norms determined from fits to 1000 data sets. Each point in the plot corresponds to one of 45 values of  $\lambda$  from  $10^{-9}$  (upper left) to 100 (lower right). The true semivariogram model is a mixture of rational quadratic and hole effect models with sill of 14. .... 69
22. This plot shows the relationship between the average of the sill estimates and the average of the residual norms determined from fits to 1000 data sets. Each point in the plot corresponds to one of 45 values of  $\lambda$  from  $10^{-9}$  to 100. The true semivariogram model is a mixture of spherical models with sill of 4. .... 70

23. Plot of the means of the sill estimates and the means of the residual norms versus  $45 \log_{10}(\lambda)$  values. The means were determined from fits to 1100 data sets. The true semivariogram model is exponential with sill and range of 10. .... 72
24. Plot of the means of the sill estimates and the means of the residual norms versus  $45 \log_{10}(\lambda)$  values. The means were determined from fits to 1000 data sets. The true semivariogram model is a mixture of a rational quadratic and hole effect models with sill of 14. . 73
25. Plot of the means of the sill estimates and the means of the residual norms versus  $45 \log_{10}(\lambda)$  values. The means were determined from fits to 1000 data sets. The true semivariogram model is a mixture of spherical models with sill of 4. .... 74
26. (a)-Sample semivariogram and two nonparametric fits. Solid line is unpenalized fit with sill estimate of 7.42. Dashed line is penalized fit with sill estimate of 7.29 (b)-Plot of residual norms versus  $\log_{10}(\lambda)$ . 76
27. This plot shows the relationship between the average of the sill estimates and the average of the residual norms determined from fits to 1000 simulated data sets. Each point in the plot corresponds to one of 45 values of  $\lambda$  from  $10^{-9}$  (upper left) to 100 (lower right). The true semivariogram model is exponential with sill of 10 and range of 2. .... 77
28. This plot shows the relationship between the average of the sill estimates and the average of the residual norms determined from fits to 1000 simulated data sets. Each point in the plot corresponds to one of 45 values of  $\lambda$  from  $10^{-9}$  (upper left) to 100 (lower right). The true semivariogram model is exponential with sill of 10 and range of 18. .... 78
29. Plot of the means of the sill estimates and means of residual norms versus  $45 \log_{10}(\lambda)$  values. The means were determined from fits to 1000 data sets. The true semivariogram model is exponential with sill of 10 and range of 2. .... 79

30. Plot of the means of the sill estimates and means of residual norms versus  $45 \log_{10}(\lambda)$  values. The means were determined from fits to 1000 data sets. The true semivariogram model is exponential with sill of 10 and range of 18. .... 80
31. Results of fitting a nonparametric semivariogram to data from known valid semivariogram models. .... 85
32. This figure shows the increasing accuracy of fitted nonparametric semivariograms to a true exponential semivariogram (points) under increasing domain and infill asymptotics. The 2 upper lines are 95th percentiles, the 2 middle lines are means, and the 2 lower lines are 5th percentiles from 1000 fits. The true sill and range were 10. .... 89
33. Sample semivariogram (points) and two semivariogram fits to Clark's (1979) silver data. Solid line is spherical model with sill = 11. Dashed line is nonparametric model with sill = 11.33. .... 94
34. Sample semivariogram (points) and two semivariogram fits to Clark's (1979) nickel data. Solid line is spherical model with sill = 2.15. Dashed line is nonparametric model with sill = 2.15. .... 95
35. Sample semivariogram (points) and two semivariogram fits to Cressie's (1985) iron ore data. Solid line is spherical model with sill = 3.59. Dashed line is nonparametric model with sill = 3.02. .... 97
36. Sample semivariogram with a nugget effect of 5 and two nonparametric semivariogram estimators. The solid line is a fit to the data with the nugget included. The dotted line resulted from a fit to the data with the nugget removed. The nugget was then added back to the resulting fit to get the dashed line. .... 102

## ABSTRACT

Semivariogram functions are used by geostatisticians to describe the correlation structure of spatial random variables. Valid semivariogram functions must be conditionally negative definite. To ensure this restriction when estimating semivariogram functions, investigators typically fit a valid model to data in the form of a sample semivariogram. Two authors (Shapiro and Botha, 1991) suggested a nonparametric method of estimating semivariogram functions, but did not compare their method to the more traditional methods of fitting these functions. The method was evaluated using simulated data. The fits obtained using the nonparametric method were compared to nonlinear least squares fits obtained using four parametric models (exponential, Gaussian, rational quadratic, and power). The comparisons were made using the integrated squared errors of the resulting fits. In addition, a method of using the nonparametric method to estimate the variance (sill) of second order stationary random processes is described and evaluated. The nonparametric estimator always resulted in fits that were as good as those obtained using parametric models as measured by integrated squared error. The variance estimator as initially proposed resulted in biased and highly variable estimates, but a penalized fitting procedure resulted in estimates that were essentially unbiased and less variable than estimates obtained using the parametric models. This penalized estimation procedure did not affect the quality of the fitted functions. The nonparametric method is faster and easier to use, and more objective than parametric methods.

## CHAPTER 1

### INTRODUCTION

Cressie (1991) defines geostatistics as the modeling of data as a partial realization of a stochastic process  $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$  where  $\mathbf{s}$  varies *continuously* over some  $d$ -dimensional spatial region of interest. Geostatistics was developed by geologists and mining engineers working outside the mainstream of classical statistics. The mining emphasis is evident in such texts as Journel and Huijbregts (1978) and David (1977). These investigators were primarily concerned with determining the economic feasibility of extracting gold, diamonds, copper, coal, oil, and so on from some space. The data they collected was correlated and they recognized that knowledge of the correlation structure could be put to good use in predicting what they would find at unsampled locations.

Geostatistical techniques have a wider range of applicability than just mining geology. Spatial problems arise in a wide variety of scientific disciplines, including geography, sociology, ecology, forestry, wildlife and range science, epidemiology, astronomy, and climatology. The decade of the 1980's saw an explosion of work in the area of geostatistics. Several good texts summarizing this and other work in spatial statistics include Cressie (1991), Haining (1990), Upton and Fingleton

(1985, 1989), and Ripley (1981). Issacs and Srivastava (1989) is a more accessible introduction to what could be termed classical geostatistics. Christensen (1991) discusses spatial data analysis within the context of linear models.

The main tool geostatisticians use to model correlation structure is the variogram function (Cressie, 1991), which is related to the more familiar covariance and correlation functions with which statisticians work. Variogram functions are useful not only as a description of spatial dependency, but also play a key role in the spatial prediction technique known as kriging.

Valid covariance functions must be positive definite. Similarly, valid variogram functions must satisfy a mathematical property known as conditional negative definiteness. The methods typically used to estimate variogram functions do not yield estimators that are guaranteed conditionally negative definite. Thus, geostatisticians have typically chosen to estimate variogram functions by fitting conditionally negative definite models to *data* in the form of a sample semivariogram. Such choices are usually made on a subjective basis, and imply assumptions about the underlying spatial process. This subjectivity can lead to significant differences in conclusions about the correlation structure of the process (Englund, 1990). It has been recognized that a nonparametric method of estimating variogram functions would be desirable (Cressie, 1991), but satisfying the property of conditional negative definiteness has been an obstacle.

Recently, several papers have addressed this issue. Shapiro and Botha (1991) and Sampson and Guttorp (1992) both viewed the problem of estimating variogram functions as a mixture (a linear combination) of a more general class of known valid functions. Shapiro and Botha (1991) presented their method as a nonparametric variogram estimator, but after describing it they illustrated its use on only a single data set. Sampson and Guttorp (1992) used essentially the same approach, but fit the nonparametric function differently. Again, they only used it on a single case. Hall, Fisher, and Hoffman (1993) used a modified kernel regression technique that they claim has nice theoretical properties, but did not use it to fit any data. Lele (1994) presented a method of constructing conditionally negative definite matrices for use in kriging. His method does not require the specification of a parametric family of variogram functions and it performs well in practice (Lele, personal communication). It does not actually produce an explicit variogram function.

In what follows, the Shapiro-Botha method of fitting valid variogram functions will be examined further. This nonparametric method will be compared with more traditional parametric methods. In addition, although Shapiro and Botha (1991) did not address this issue, their method can be used to derive a new nonparametric estimate of the variance of a second order stationary random process. Some properties of this estimator will be investigated. Throughout the remainder of the text the Shapiro-Botha method will be referred to as the SB method.

Spatial Stochastic Processes

Following Cressie (1991) let

$$\{Z(\mathbf{s}) : \mathbf{s} \in D\}$$

where  $D \subset \mathbf{R}^d$ , be a spatial stochastic process. Then  $Z(\mathbf{s})$  is a random variable for each location  $\mathbf{s}$  where  $\mathbf{s}$  is assumed to vary continuously over  $D$ . Generally, the dimension  $d$  will be 1, 2, or 3. The analysis of spatial data requires the modeling of  $Z(\mathbf{s})$ . Assuming the existence of first and second order moments  $Z(\mathbf{s})$  can be modeled as

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + \epsilon(\mathbf{s})$$

where  $\mu(\mathbf{s})$  is a constant mean function and  $\epsilon(\mathbf{s})$  is a zero mean random error process. Data collected from such a process represents only a partial sample of a single realization. Inference is impossible without additional assumptions. Accordingly, let

$$E[Z(\mathbf{s})] = \mu \quad \forall \mathbf{s} \in D. \quad (1)$$

and assume that the covariance function satisfies

$$\text{Cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_j)) = C(\mathbf{s}_i - \mathbf{s}_j) \quad \forall \mathbf{s}_i, \mathbf{s}_j \in D. \quad (2)$$

Thus, the covariance function depends only on the distance between two points,  $\mathbf{s}_i$  and  $\mathbf{s}_j$ .

**Definition 1** *Spatial processes satisfying (1) and (2) are said to be second-order stationary. If  $C(\mathbf{s}_i - \mathbf{s}_j)$  is a function of Euclidean distance (or lag)  $h_{ij} = (\|\mathbf{s}_i - \mathbf{s}_j\|)$  only, then  $C$  is said to be isotropic. Processes that are not isotropic are said to be anisotropic.*

The variogram function is defined to be

$$2\gamma(\mathbf{s}_i, \mathbf{s}_j) \equiv \text{Var}(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)) \quad \forall \mathbf{s}_i, \mathbf{s}_j \in D.$$

As with the covariance function, the variogram is a function of the increments  $\mathbf{s}_i - \mathbf{s}_j$  only, and is written

$$2\gamma(\mathbf{s}_i, \mathbf{s}_j) = 2\gamma(\mathbf{s}_i - \mathbf{s}_j). \quad (3)$$

The function  $\gamma(\mathbf{s}_i - \mathbf{s}_j)$  is called the *semivariogram*.

**Definition 2** *Spatial processes satisfying (1) and (3) are said to be intrinsically stationary.*

It is easily seen that second-order stationarity implies intrinsic stationarity, but the converse is not necessarily true. The variogram function always exists even though the covariance function may not exist. Existence of the covariance function and second-order stationarity imply,

$$2\gamma(\mathbf{s}_i - \mathbf{s}_j) = 2(C(\mathbf{0}) - C(\mathbf{s}_i - \mathbf{s}_j)). \quad (4)$$

where  $C(\mathbf{0}) = \text{Var}(Z(\mathbf{s}))$ . Cressie (1991), Christensen (1991), and Journel and Huijbregts (1978) give good discussions of stationarity and its implications in spatial stochastic processes.

In the subsequent text isotropy will be assumed. Also, let  $\mathbf{h}_{ij} = \mathbf{s}_i - \mathbf{s}_j$  and  $h_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$ . The subscripts on  $\mathbf{h}$  and  $h$  will frequently be suppressed.

### Properties of Variogram and Covariance Functions

Let  $Z(\mathbf{s})$  be a spatial stochastic process, with mean  $\mu$ , covariance function  $C(\mathbf{h})$ , and variogram function  $2\gamma(\mathbf{h})$ . Finite linear combinations of the type,

$$W = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i)$$

for any weights  $\lambda_i \in \mathbf{R}$  are of interest for prediction. Such a linear combination must have a nonnegative variance,

$$\text{Var}(W) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(\mathbf{s}_i - \mathbf{s}_j) \geq 0 \quad (5)$$

for all  $\lambda_i$  and all  $\mathbf{s}_i, \mathbf{s}_j \in D$ .

Under second-order stationarity (4) holds and

$$\text{Var}(W) = C(\mathbf{0}) \sum_{i=1}^n \lambda_i \sum_{j=1}^n \lambda_j - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j) \geq 0$$

for all  $\lambda_i$  and all  $\mathbf{s}_i, \mathbf{s}_j \in D$ , including all  $\lambda_i$  such that  $\sum_{i=1}^n \lambda_i = 0$ . Thus,

$$- \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j) \geq 0 \quad (6)$$

for all  $\mathbf{s}_i, \mathbf{s}_j \in D$ , and all  $\lambda_i$  such that  $\sum_{i=1}^n \lambda_i = 0$ .

**Definition 3** A function  $C(\mathbf{h}); (\mathbf{h} \in \mathbf{R}^d)$  satisfying (5) is said to be positive-definite in  $\mathbf{R}^d$ . A function  $\gamma(\mathbf{h}); (\mathbf{h} \in \mathbf{R}^d)$  satisfying (6) is said to be conditionally negative definite in  $\mathbf{R}^d$ .

Thus, valid covariance functions are required to be positive definite, and valid semivariogram (and variogram) functions are required to be conditionally negative definite.

The above properties of  $C(\mathbf{h})$  and  $2\gamma(\mathbf{h})$  imply

- $C(\mathbf{0}) = \text{Var}(Z(\mathbf{s})) \geq 0$
- $C(\mathbf{h}) = C(-\mathbf{h})$
- $|C(\mathbf{h})| \leq C(\mathbf{0})$  (Cauchy-Schwarz Inequality)
- $2\gamma(\mathbf{0}) = 0$
- $2\gamma(\mathbf{h}) = 2\gamma(-\mathbf{h}) \geq 0$

As the distance  $h$  increases it is often reasonable to assume that the correlation disappears; that is,

$$\lim_{h \rightarrow \infty} C(\mathbf{h}) = 0.$$

This implies that

$$\lim_{h \rightarrow \infty} \gamma(\mathbf{h}) = \text{Var}(Z(\mathbf{s})) = C(\mathbf{0}).$$

Conditional negative definiteness is not sufficient for a function to be a valid variogram. A further condition that a variogram function must satisfy for validity (Cressie, 1991) is

$$\lim_{h \rightarrow \infty} \frac{2\gamma(\mathbf{h})}{h^2} = 0.$$

Thus, the conditionally negative definite function  $f(h) = h^2$  is not a valid variogram function.

Note that if  $C_1(\mathbf{h})$  and  $C_2(\mathbf{h})$  are positive definite in  $\mathbf{R}^d$ , then so is  $\lambda_1 C_1(\mathbf{h}) + \lambda_2 C_2(\mathbf{h})$  for any nonnegative  $\lambda_i$ . Care must be taken with such combinations, however; because although positive definiteness in  $\mathbf{R}^d$  implies positive definiteness in  $\mathbf{R}^p$  where  $p \leq d$  it is not true that positive definiteness in  $\mathbf{R}^d$  implies positive definiteness in  $\mathbf{R}^p$  where  $p > d$ . Similar results hold for conditionally negative definite functions. It should also be noted that if  $C(\mathbf{h})$  is positive definite then  $k - C(\mathbf{h})$ , where  $k$  is any constant, is conditionally negative definite.

### Estimation of Semivariogram Functions

Up to this point the discussion has centered on variogram functions. Through an abuse of language, many articles that refer to fitting variograms are actually fitting semivariograms. In what follows the variogram function will not be mentioned again. The discussion will always pertain to estimation of the semivariogram function.

The starting point for estimation has been the semivariogram cloud (Cressie, 1991). This is a plot of the  $n(n-1)/2$  distinct values of

$$Y_{ij} = \frac{[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]^2}{2}$$

against  $h_{ij}$ .

If  $Z(\mathbf{s})$  is assumed to be Gaussian then

$$Y_{ij} \sim \gamma(h_{ij})\chi_1^2$$

implying the values of  $Y_{ij}$  are highly skewed.

Cressie and Hawkins (1980) and Hawkins and Cressie (1984) considered more robust approaches. Using Box-Cox power transformations they recommended plotting the transformed differences

$$\frac{[|Z(\mathbf{s}_i) - Z(\mathbf{s}_j)|]^{1/2}}{2}$$

against  $h_{ij}$ . Cressie (1991) refers to such a plot as a *square-root-differences-cloud* and argues that it is more useful as an exploratory tool because of its symmetry.

Sample semivariograms are determined by smoothing the semivariogram cloud in some fashion. The classical method of moments estimator for the sample semivariogram is,

$$\hat{\gamma}(\mathbf{h}) \equiv \frac{1}{2 |N(\mathbf{h})|} \sum_{|N(\mathbf{h})|} (Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2 \quad (7)$$

where  $\mathbf{h} \in \mathbf{R}^d$ ,  $N(\mathbf{h}) \equiv \{(\mathbf{s}_i, \mathbf{s}_j) : \mathbf{h} = \mathbf{s}_i - \mathbf{s}_j; i, j = 1, \dots, n\}$ , and  $|N(\mathbf{h})|$  is the number of pairs in  $N(\mathbf{h})$ . If the data are irregularly spaced, then  $N(\mathbf{h})$  is generally modified to be the set of points  $\{(\mathbf{s}_i, \mathbf{s}_j) : \mathbf{s}_i - \mathbf{s}_j \in T(\mathbf{h})\}$ , where  $T$  is a  $d$ -dimensional tolerance region surrounding  $\mathbf{h}$ . This estimator is generally attributed to Matheron and is unbiased, but is not resistant to outliers (Cressie, 1991).

Cressie (1991) recommends instead that one construct sample semivariograms using the robust estimator

$$\bar{\gamma}(\mathbf{h}) \equiv \left(\frac{1}{2}\right) \left( \frac{\left(\frac{1}{|N(\mathbf{h})|} \sum_{|N(\mathbf{h})|} (|Z(\mathbf{s}_i) - Z(\mathbf{s}_j)|^{1/2})\right)^4}{(0.457 + 0.494/|N(\mathbf{h})|)} \right). \quad (8)$$

He argues that the summands in (8) are less correlated than the summands in (7) implying that they are more efficient in estimating the semivariogram.

There are other methods of determining sample semivariograms. Cressie (1991) gives a good discussion. Dowd (1984) also addresses this issue. The most commonly used method is the method of moments estimator described above, and it is the one which will be used in this dissertation.

### Parametric Semivariogram Models

Generally, estimation of a semivariogram function involves fitting a conditionally negative definite model to data in the form of one of the sample semivariograms discussed above. Examples of commonly used models can be found in Cressie (1991)

and Journel and Huijbregts (1978), among others.

### Some Isotropic Semivariogram Models

Following are six examples of valid isotropic semivariogram models. These models and mixtures of them will be used extensively in subsequent sections and chapters, especially Chapters 3 and 4.

*Spherical Model:* The spherical model may be the most commonly used model. It is valid in  $\mathbf{R}^1$ ,  $\mathbf{R}^2$ , and  $\mathbf{R}^3$ , and has the form,

$$\gamma(\mathbf{h}; \theta) = \begin{cases} 0 & \mathbf{h} = \mathbf{0} \\ c_0 + s((3/2)(h/r) - (1/2)(h/r)^3) & 0 < h \leq r \\ c_0 + s, & h \geq r \end{cases}$$

$\theta = (c_0, s, r)$ , where  $c_0 \geq 0, s \geq 0, r \geq 0$ .

*Exponential Model:* Another commonly used model, the exponential model, is valid in all dimensions. It has the form,

$$\gamma(\mathbf{h}; \theta) = \begin{cases} 0 & \mathbf{h} = \mathbf{0} \\ c_0 + s(1 - \exp(-3h/r)) & h > 0 \end{cases}$$

$\theta = (c_0, s, r)$ , where  $c_0 \geq 0, s \geq 0, r \geq 0$ .

*Gaussian Model:* The Gaussian model is valid in all dimensions, and has the form;

$$\gamma(\mathbf{h}; \theta) = \begin{cases} 0 & \mathbf{h} = \mathbf{0} \\ c_0 + s(1 - \exp(-3(h/r)^2)) & h > 0 \end{cases}$$

$\theta = (c_0, s, r)$ , where  $c_0 \geq 0, s \geq 0, r \geq 0$ .

*Rational Quadratic Model:* The rational quadratic model is valid in all dimensions,

and has the form,

$$\gamma(\mathbf{h}; \theta) = \begin{cases} 0 & \mathbf{h} = \mathbf{0} \\ c_0 + sh^2/(1 + (h^2/r)) & h > 0 \end{cases}$$

$\theta = (c_0, s, r)$ , where  $c_0 \geq 0, s \geq 0, r \geq 0$ .

*Power Model:* The power model is valid in all dimensions. It has the form,

$$\gamma(\mathbf{h}; \theta) = \begin{cases} 0 & \mathbf{h} = \mathbf{0} \\ c_0 + b_p h^\lambda & h > 0 \end{cases}$$

$\theta = (c_0, b_p, \lambda)$ , where  $c_0 \geq 0, b_p \geq 0$ , and  $0 \leq \lambda < 2$ .

*Hole Effect Model:* There are several hole effect models. This version is valid in  $\mathbf{R}^1, \mathbf{R}^2$ , and  $\mathbf{R}^3$ , and has the form,

$$\gamma(\mathbf{h}; \theta) = \begin{cases} 0 & \mathbf{h} = \mathbf{0} \\ c_0 + s(1 - r \sin(h/r)/h) & h > 0 \end{cases}$$

where  $\theta = (c_0, s, r)$  with  $c_0, s$ , and,  $r \geq 0$ .

In all of the above models, the parameter  $c_0$  is referred to as the nugget effect. Cressie (1991) discusses the nugget effect in some detail. It is included in the model when it is believed that small scale variation is causing a discontinuity at the origin. Thus,  $\gamma(\mathbf{h}) \rightarrow c_0 > 0$ , as  $\mathbf{h} \rightarrow \mathbf{0}$ . It is a troublesome parameter because typically data are not available at lags small enough to estimate it. If continuity at the origin is expected then the only reason for  $c_0 > 0$  is measurement error. But, a nugget effect may arise as a natural part of a spatial stochastic process. It is typically modeled as a white-noise microscale process, and modeling it accurately can be

important. The sample semivariogram may give an indication of a nugget effect if it has values significantly greater than 0 at small lags.

The parameters  $s$  and  $r$  appear in all but the power model. Note that

$$\lim_{h \rightarrow \infty} \gamma(\mathbf{h}) = c_0 + s.$$

This asymptote is called the *sill* and is equal to  $C(\mathbf{0})$ . In the spherical model the semivariogram actually attains the sill, whereas in the others it approaches the sill asymptotically.

The parameter  $r$  in the above models is related to the lag at which the random variables  $Z(\mathbf{s})$  become uncorrelated, a value called the *range*. In the spherical model, the range is the lag at which  $\gamma(\mathbf{h}; \theta) = c_0 + s$ . In the other models the range is typically defined as the lag at which the semivariogram equals 95% of the sill.

The power model does not have a sill or range. It is an example of a semivariogram model for a spatial process for which the covariance function does not exist. Such processes with infinite variance may actually exist; however, power models are more commonly used to model semivariograms in cases where the distances between the observations  $Z(\mathbf{s}_i)$  are all less than the range.

Figure 1 shows examples of these six semivariogram models. The nugget effect in each of these was set to 0, and the sill and range were set to 3 (except for the power model).

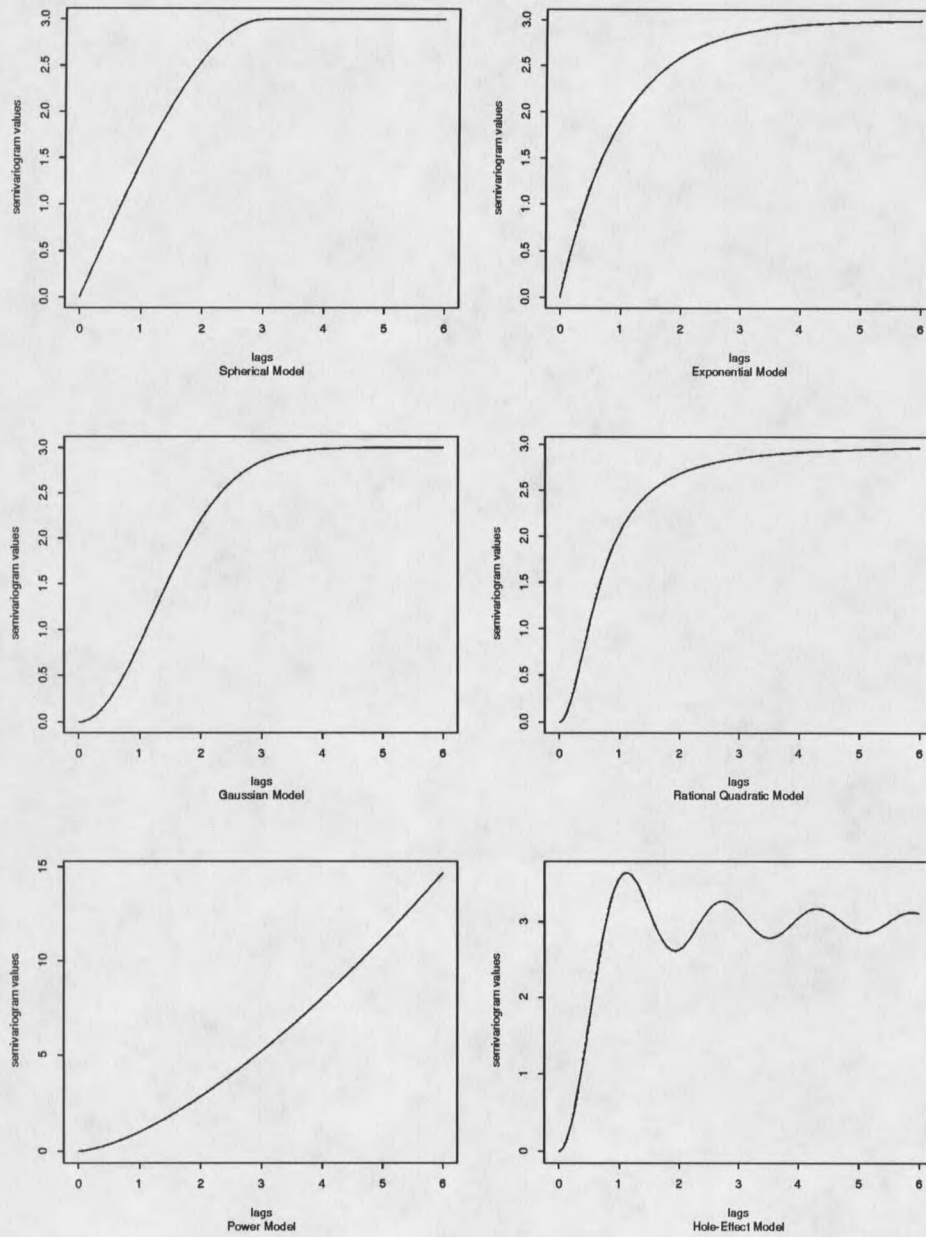


Figure 1: Examples of six parametric semivariogram models.

### Fitting Semivariogram Models

The most common methods of fitting semivariogram models are,

- By Inspection
- Maximum Likelihood
- Restricted Maximum Likelihood
- Minimum Norm Quadratic Estimation
- Least Squares

The first item above was not included facetiously. This is a popular and widespread method of fitting. Clark's (1979) introductory text and, more recently, Issacs and Srivastava (1989) deal only with this approach to fitting semivariograms. The GSLIB software (Deutsch and Journel, 1992) has no provisions for fitting semivariograms by any other method. Typically, a choice of a parametric family is made and a model is fit to the data by a trial and error process. Different parameter values are tried until a model that *looks good* is found. Frequently, it is discovered that one model alone will not suffice, and different mixtures of valid models will be examined. Statisticians, of course, do not think much of this approach. But if it is to be supplanted by other, supposedly superior techniques, then those techniques will need to be easy to implement and give good results.

Cressie (1991) gives an excellent summary of the other approaches. What they all have in common is the requirement of an *a priori* choice of a parametric family of semivariogram models.

Both maximum likelihood and restricted maximum likelihood have not been widely used (but see Jones and Vecchia, 1993). Maximum likelihood estimators tend to be biased, often badly so (Cressie, 1991). Other problems include possible multimodality of the likelihood function and global maxima that yield unrealistic parameter estimates (Warnes and Ripley, 1987; Ripley, 1988, Watkins, 1992). Restricted maximum likelihood estimators tend to not be as severely biased as maximum likelihood estimates. The reader who wishes more information on these two methods of estimating semivariogram functions is referred to Cressie (1991), and Kitinidis (1987). Minimum norm quadratic (MINQU) estimation will be discussed further in Chapter 7.

By far the most common approach to semivariogram modeling among statisticians is nonlinear least squares. Given a sample semivariogram (usually Matheron's method of moments estimator) find  $\theta$  that minimizes

$$\sum_{i=1}^k \{\hat{\gamma}(h_i) - \gamma(h_i; \theta)\}^2$$

where  $\hat{\gamma}(h_i)$  is the sample semivariogram value at lag  $h_i$  and  $\gamma(h_i; \theta)$  is a known valid semivariogram model that is nonlinear in its parameters.

This is a mathematically reasonable thing to do, but it does not take into account

the distributional properties of  $\hat{\gamma}$ . This can be accomplished by using generalized least squares. Let  $\mathbf{V}$  be the covariance matrix of

$$\hat{\gamma} = (\hat{\gamma}(h_1), \dots, \hat{\gamma}(h_k))'.$$

Then the problem becomes one of choosing  $\theta$  that minimizes

$$(\hat{\gamma} - \gamma(\theta))' \mathbf{V}^{-1} (\hat{\gamma} - \gamma(\theta)).$$

The calculation of  $\mathbf{V}$  can be difficult, and a reasonable compromise is to use weighted least squares where  $\mathbf{V}$  is set equal to a diagonal matrix with  $\text{var}(\hat{\gamma}(h_i))$  on the diagonal. Cressie (1985) discussed least squares fitting of semivariogram functions in detail, and recommended minimizing

$$\sum_{i=1}^k |N(h_i)| \left( \frac{\hat{\gamma}(h_i)}{\gamma(h_i; \theta)} - 1 \right)^2$$

as a good approximation to weighted least squares. Note that the more pairs of observations  $|N(h_i)|$  at lag  $h_i$  the more weight the residual at that lag gets. Also, smaller values of the theoretical semivariogram give more weight to the residual. Cressie (1985) also recommended using the robust estimator (8) in place of the method of moments estimator.

Zimmerman and Zimmerman (1991) compared the methods just described (with the exception of fitting by inspection) and recommended that least squares and weighted least squares be used to fit semivariogram functions.

Preview

Chapter 2 discusses the problem of nonparametric semivariogram estimation in general terms. The SB method will be described. The method depends on the spectral theory of positive definite and conditionally negative definite functions and this will also be discussed in Chapter 2. Chapter 3 describes the implementation of the SB method and its performance as a fitting procedure using simulated data. The nonparametric fits will be compared to least squares fits obtained using parametric models. The problems of node selection and overfitting will be discussed. Chapter 4 describes how the SB method can be used to estimate the sill. Comparisons of sill estimates obtained using parametric fitting procedures will be made. A method of stabilizing the sill estimates will be described. Chapter 5 deals with asymptotics, and Chapter 6 presents examples of the use of the SB nonparametric fitting method on real data. Chapter 7 discusses future research needs, including nonparametric estimation of the nugget effect and the range, MINQU estimation and  $L_1$  fits. Chapter 8 contains a brief summary and some conclusions about the best way to implement the SB method.

## CHAPTER 2

### THE SHAPIRO-BOTHA METHOD

All of the fitting criteria discussed in the previous chapter rely on estimating the parameters of an *a priori* choice of a class of semivariogram models. The sample semivariograms (7 and 8) are nonparametric, but are not likely to yield valid semivariograms. They are used to identify a reasonable form for a parametric semivariogram, and to find reasonable starting values for the parameter estimates in the nonlinear least squares fitting algorithms.

To the eye there really is not much difference between an exponential, a rational quadratic, and a Gaussian semivariogram model. Different choices of a semivariogram model can result in different estimates of parameters, even among such seemingly similar models. Frequently, the choice of a model is based more on familiarity and what seems to be popular in the literature. Spherical models are popular, despite the fact that nonlinear fits to such a model are difficult to do correctly. The difficulty arises because the spherical model is not differentiable in the parameter  $r$ . Of course, if one is fitting models by inspection then the differentiability of objective functions is not a concern.

The SB nonparametric estimation procedure relies on the spectral representation

of covariance and semivariogram functions which will be discussed first, followed by a description of the SB nonparametric procedure.

### Spectral Representation

Much of the work on the spectral representation of positive definite functions arose in the study of the theory of random functions. It is well known that a function is positive definite if and only if it is the covariance function for some second-order stationary random process, and that valid covariance functions are required to have certain spectral representations (*e.g.* see Bochner (1955), Doob (1953), Yaglom (1987*a*, 1987*b*), and Matern (1986)).

Schoenberg (1938) and von Neumann and Schoenberg (1941) established the following theorems about the spectral representation of isotropic positive definite and conditionally negative definite functions from a different perspective. Given a space  $\mathbf{X}$  with a metric (or pseudometric)  $d$ , does there exist a mapping  $\Phi : \mathbf{X} \rightarrow \mathbf{H}$  (where  $\mathbf{H}$  is a Hilbert space) such that

$$\| \Phi(x) - \Phi(y) \| = d(x, y)$$

for all  $x, y \in \mathbf{X}$ , (i.e. when is  $\mathbf{X}$  isometrically embeddable in  $\mathbf{H}$ )? If  $d^2$  is conditionally negative definite, then such a  $\Phi$  exists.

**Theorem 1**  $C(h)$  is positive definite in  $\mathbf{R}^d$  if and only if it has a spectral representation

$$C(h) = \int_0^\infty \Omega_d(ht) d\mu(t)$$

where

$$\Omega_d(r) = (2/r)^{(d-2)/2} \Gamma(d/2) J_{(d-2)/2}(t).$$

$J_\nu$  is the Bessel function of the first kind of order  $\nu$ , and  $\mu$  is a nondecreasing bounded function on  $t \geq 0$ .

**Theorem 2**  $C(h)$  is positive definite in all dimensions if and only if it has the form

$$C(h) = \int_0^\infty \exp\{-(ht)^2\} d\mu(t)$$

where  $\mu(t)$  is nondecreasing and bounded for  $t \geq 0$ .

Although the term conditionally negative definite never appears in their papers, the following are immediate corollaries of Theorem 7 in von Neumann and Schoenberg (1941; p.244) and Theorem 6 in Schoenberg (1938; p.828).

**Theorem 3**  $G(h)$  is conditionally negative definite in  $\mathbf{R}^d$  if and only if it is of the form

$$G(h) = \int_0^\infty \frac{(1 - \Omega_d(ht))}{t^2} d\mu(t)$$

where  $\Omega_d$  has the form given above,  $\mu(t)$  is nondecreasing for  $t \geq 0$ ,  $\mu(0) = 0$ , and

$$\int_1^\infty \frac{d\mu(t)}{t^2} < \infty$$

**Theorem 4**  $G(h)$  is conditionally negative definite in all dimensions if and only if it is of the form

$$G(h) = \int_0^\infty \frac{1 - \exp(-h^2 t)}{t} d\mu(t)$$

where  $\mu(t)$  is nondecreasing for  $t \geq 0$ , and

$$\int_1^\infty \frac{d\mu(t)}{t} < \infty.$$

Fortunately, the rather imposing  $\Omega_d(r)$ , reduces to

- $\Omega_1(r) = \cos(r)$  for one dimension.
- $\Omega_2(r) = J_0(r)$  for two dimensions.
- $\Omega_3(r) = \frac{\sin(r)}{r}$  for three dimensions.
- $\Omega_4(r) = \frac{2J_1(r)}{r}$  for four dimensions.

The Schoenberg-von Neumann results, derived within the context of the geometry of Hilbert spaces, appear to have escaped attention of those involved in geostatistical work. For example, much attention has been paid to developing techniques to check potential semivariogram functions for conditional negative definiteness. In particular, Christakos (1984), and Armstrong and Diamond (1984), both addressed this problem, but neither referenced the work of Schoenberg, who established many of the results presented in those two papers. The first acknowledgement of the significance of the work of Schoenberg and von Neumann in geostatistics is found

in Cressie (1991) who gives an accessible account of the spectral properties of covariance and semivariogram functions. A rigorous treatment of spectral and other properties of positive definite and related functions can be found in Berg, Christensen, and Ressel (1984).

### The Fitting Procedure

Shapiro and Botha (1991) formulated the problem as follows. Find a  $m \times 1$  vector  $\mathbf{p} = (p_1, \dots, p_m)'$  that minimizes

$$\mathbf{Q}(\mathbf{p}) = \sum_{i=1}^n w_i \left( \hat{\gamma}(h_i) - \sum_{j=1}^m (1 - \Omega_d(h_i t_j)) p_j \right)^2 \quad (9)$$

subject to the constraint that  $p_j \geq 0; j = 1, \dots, m$ . Let  $\hat{\gamma}$  be a  $n \times 1$  vector,  $\mathbf{A}$  be an  $n \times m$  matrix with elements  $\{a_{ij}\} = \{1 - \Omega_d(h_i t_j)\}$ ,  $\mathbf{p}$  be a  $m \times 1$  solution vector, and  $\mathbf{W}$  be an  $n \times n$  weight matrix. Then (9) can be written,

$$\mathbf{Q}(\mathbf{p}) = (\hat{\gamma} - \mathbf{A}\mathbf{p})' \mathbf{W} (\hat{\gamma} - \mathbf{A}\mathbf{p}). \quad (10)$$

For convenience they chose  $\mathbf{W}$  to be an identity matrix. Recalling Theorem 1, it can be seen that  $\sum_{j=1}^m (1 - \Omega_d(h_i t_j)) p_j$  is conditionally negative definite with  $\mu(t)$  taken to be a step function with nonnegative jumps  $p_j$  at nodes  $t_j$ . Finding a solution to (9),  $\tilde{\mathbf{p}} = (\tilde{p}_1, \dots, \tilde{p}_m)'$ , is a problem in quadratic programming. The resulting nonparametric estimate of the semivariogram has an explicit representation as

$$\tilde{\gamma}(h) = \sum_{j=1}^m (1 - \Omega_d(h t_j)) \tilde{p}_j \quad (11)$$

The class of all conditionally negative definite functions is larger than the class being fit by the SB procedure, as can be seen by comparing the results of Theorems 3 and 4 with the described fitting procedure. The following proposition shows that the SB approach does yield semivariograms valid for isotropic second-order stationary random processes, however.

**Proposition 1**  $\gamma(h)$  is a valid semivariogram for an isotropic second-order stationary random process if and only if it has the form

$$\gamma(h) = \int_0^{\infty} (1 - \Omega_d(ht)) dM(t)$$

where  $M(t)$  is a nonnegative bounded nondecreasing function on  $t \geq 0$ .

**Proof:** From Theorem 1 it is known that

$$C(h) = \int_0^{\infty} \Omega_d(ht) dM(t)$$

is positive definite and is thus a covariance function. Further,

$$C(0) = \int_0^{\infty} dM(t) < \infty.$$

Let  $\gamma(h)$  be a valid semivariogram for an isotropic second order stationary random process. Then,

$$\begin{aligned} \gamma(h) &= C(0) - C(h) \\ &= \int_0^{\infty} dM(t) - \int_0^{\infty} \Omega_d(ht) dM(t) \\ &= \int_0^{\infty} (1 - \Omega_d(ht)) dM(t). \end{aligned}$$

The converse is proved by reversing the steps. To see that  $M(t)$  is nonnegative note that by Theorem 3 there exists a nonnegative nondecreasing function  $\mu(t)$  such that  $dM(t) = (1/t^2)d\mu(t) \geq 0$  which implies  $M(t) \geq 0$ .  $\square$

Shapiro and Botha (1991) only demonstrated that the solution to (10) is conditionally negative definite and gave an example of its use on one data set. They did not compare it with more traditional parametric fitting methods. They did discuss some approaches designed to yield smoother estimates, as the solution to (10) may tend to track the data too closely. They also give examples of fitting valid semivariograms in one, two, and three dimensions. They do not discuss how to choose the nodes; choosing them in their examples in what they describe as an *ad hoc* manner.

SB make no reference to nonparametric estimation of the nugget effect, the sill, and the range. It appears at first glance that a nugget effect could be incorporated by simply adding an intercept term to the model. However, as will be discussed later, it is more difficult than that, at least for the SB method. Estimation of the range will also be seen to be a difficult problem. Estimating the sill will be easier. Note that as  $h \rightarrow \infty$ , (11) will converge to  $s_p = \sum \tilde{p}_j$ . Thus,  $s_p$  is an estimate of the sill.

## CHAPTER 3

### FITTING NONPARAMETRIC SEMIVARIOGRAMS

This chapter has 2 sections. The first section describes how the SB method was implemented. In particular, the problems of node selection and overfitting are discussed. Also, the selection of the appropriate dimensional version of  $\Omega_d(ht)$  is considered. In the second section, fits obtained using the SB method are compared with least squares fits obtained with four different parametric models.

The comparisons required a large amount of simulated data. The technique used to simulate spatial random fields is described in Appendix A. Random fields were simulated for 50 locations spaced one unit apart on a one dimensional transect. This yielded a total of 1225 distinct pairs of

$$Y_{ij} = \frac{[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]^2}{2}$$

over 49 lags. It is typical to only consider lags with an adequate number of observations, usually taken to be 30 (Cressie, 1991). Thus, the number of lags considered for the comparisons was taken to be 20, and the resulting semivariogram is a scatter plot of 790 points of  $Y_{ij}$  versus 20 lags with  $50 - h$  observations at lag  $h$ . The nugget effect was set to 0 for each of these models. The sample semivariogram was

computed using the method of moments estimator.

Figure 2 shows a simulated semivariogram cloud based on simulated observations of a spatial random process. The true semivariogram model is exponential with sill of 10 and range of 10. Also shown is the sample semivariogram based on the classical method of moments estimator.

The three dimensional version of the SB estimator was chosen for all nonparametric fits. That is,

$$\Omega_3(ht) = \frac{\sin(ht)}{ht}.$$

This version will yield nonparametric estimates guaranteed to be conditionally negative definite for spatial data from one, two, or three dimensions.

### Implementation of the SB Method

SB used a Fortran subroutine (QPROG) in IMSL to accomplish the constrained least squares fitting of their method. The program *NNLS* described in Lawson and Hansen (1974) was used for the fitting procedure described here. *NNLS* produces a solution that satisfies the Kuhn-Tucker optimality conditions and it always converges. Kennedy and Gentle (1980), discuss the technique and give a simple example. Lawson and Hansen (1974) provide Fortran code for *NNLS*. This program was implemented and interfaced with the programming language Splus.

### Node Selection

The solution to (10) requires that the nodes (*i.e.* the  $t_j$ 's) be chosen before minimization. The selection of a collection of nodes is actually a selection of a set of functions of the form  $(1 - \Omega_d(ht_j))$  that will be used to construct the estimated semivariogram. Clearly, the set of nodes selected will influence the fit.

Figure 3 shows the results of two different fits using the SB method with two different collections of nodes. The method of moments estimator is from the semivariogram cloud shown in Figure 2. The solid curve in 3 resulted from a choice of 20 nodes equispaced on  $[.1, 2]$  and the dashed curve resulted from a choice of 20 nodes equispaced on  $[.5, 10]$ . Also shown in Figure 3 are the collections of functions that these node selections produced. The first selection produced a set of functions that was better able to capture the behavior of the sample semivariogram at larger lags (see the middle panel of Figure 3), and produced a better fit to the sample semivariogram.

However, it is inefficient to try to customize the selection of nodes to a given fitting problem. Such a method of choosing nodes is also open to the criticism that one can achieve any fit desired. One of the strengths of the nonparametric method is that it has the potential of being less subjective than the parametric fitting procedures currently used. It seems clear that the choice of nodes should be done in some systematic, objective way that produces a collection of functions rich

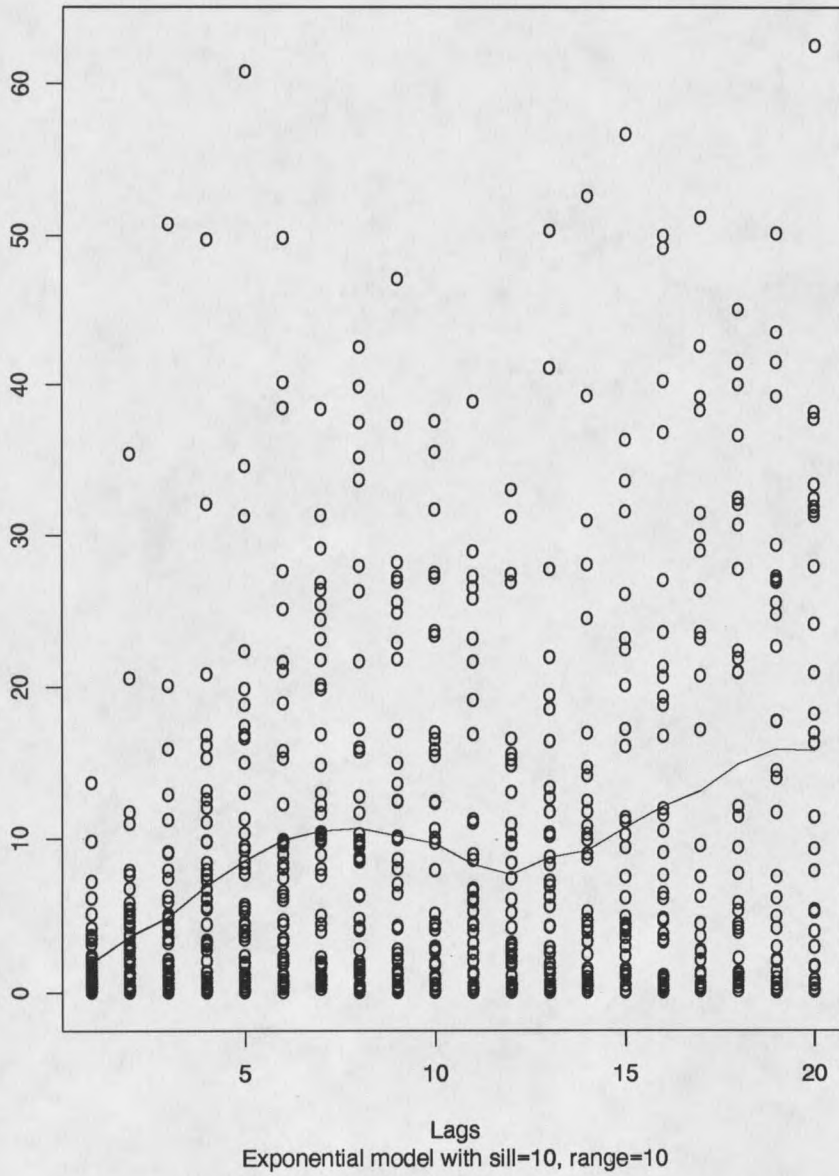


Figure 2: Semivariogram cloud with classical method of moments sample semivariogram estimator.

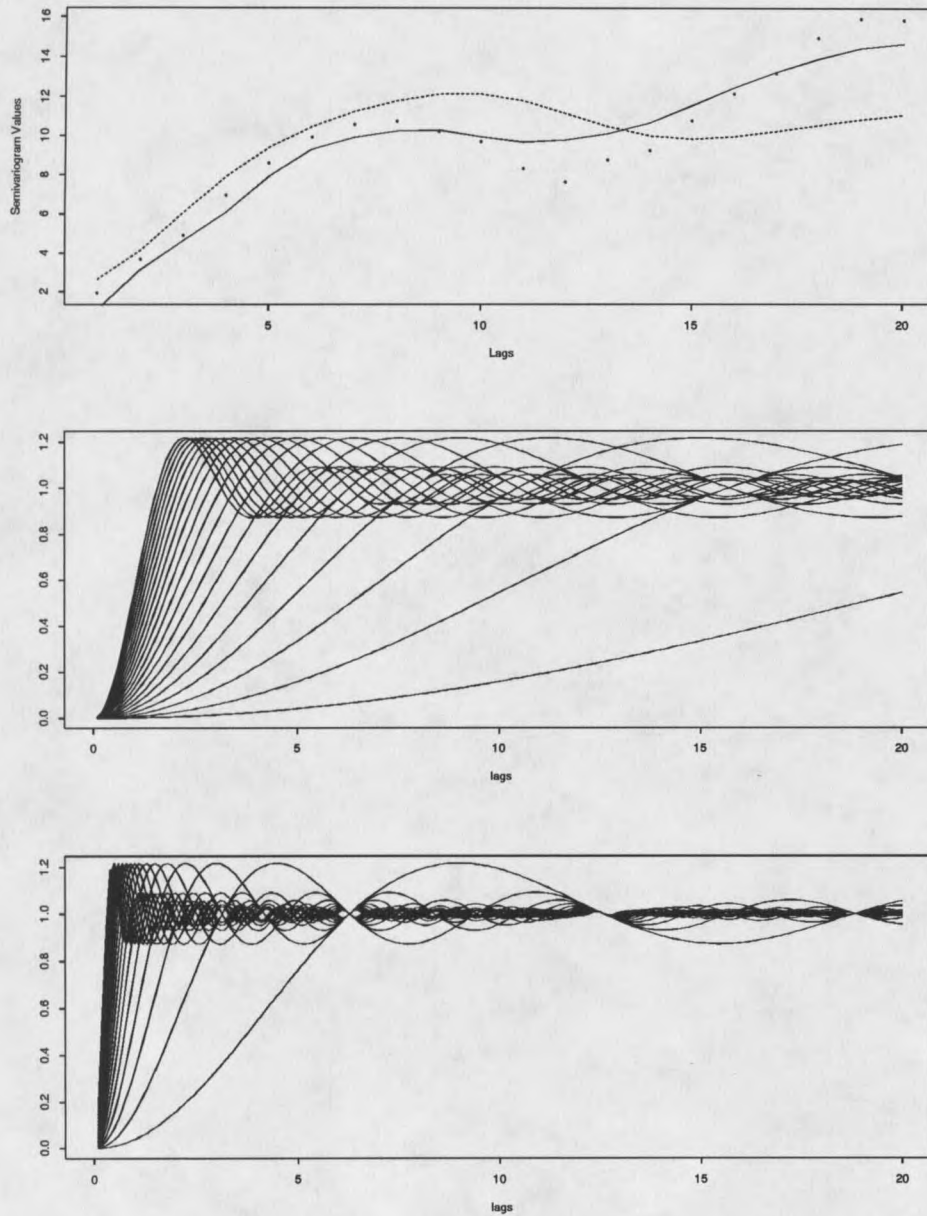


Figure 3: Two different nonparametric semivariograms based on two different collections of nodes. The solid curve was produced by the set of functions in the middle panel. The dashed curve was produced by the set of functions in the bottom panel.

enough (in some sense) to capture the behavior of the data.

There is no mathematical reason to restrict the number of nodes to be less than the number of observations. One could simply *saturate* the node space by choosing a large number of nodes and let the *NNLS* procedure, along with the data, pick the ones that are important. The rest will be assigned jumps of zero.

There are three questions that need to be resolved with the saturation approach.

1. How should the nodes be selected?
2. How many nodes are enough?
3. Does the saturation approach lead to overfitting?

#### How Should the Nodes be Selected?

Initially, fits based on collections of several hundred nodes were evaluated. The nodes were all restricted to the interval  $[0, 20]$ . Figure 4 shows the results of fitting valid three-dimensional semivariogram functions to simulated data using different collections of 200 nodes. The solid line is based on a fit using 100 nodes equispaced in  $(0, 4]$  and 100 nodes equispaced in  $[4.16, 20]$ . The dotted line is the result of a fit using 200 nodes equispaced in  $(0, 4]$  and the dashed line is the result of a fit using 200 nodes equispaced in  $(16, 20]$ . There is not much difference between the first two fits, but the third is clearly not adequate. Achieving a good fit requires a suitable

number of nodes near 0. These are associated with lower frequency curves that allow the behavior of the data at larger lags to be captured. However, the function

$$1 - \frac{\sin(ht)}{ht}$$

rises to an asymptote with increasing  $t$  so quickly that relatively few nodes larger than 4 are necessary to capture the behavior of the data at smaller lags. In fact, in many cases, putting all the nodes in the interval  $(0, 4]$  was all that was needed. Even if nodes greater than 4 were included they were frequently assigned weights of 0 by the *NNLS* algorithm. However, in looking at thousands of fits to simulated data it was noted that occasionally nodes greater than 4 were needed. It was extremely rare for nodes above 8 to receive positive weights.

#### How Many Nodes are Enough?

Theoretically, there is no upper limit to the number of nodes. Practically, the number chosen should be kept small enough to make the fitting computationally feasible.

Table 1 shows the minimum, median, maximum and the interquartile range of the residual norms from 100 fits to simulated data from a random process with an exponential semivariogram model with sill and range of 10. There were four different node selections. The selections were

- o 200 nodes with 100 nodes equispaced in  $(0, 4]$  and 100 equispaced in  $[4.16, 20]$ .

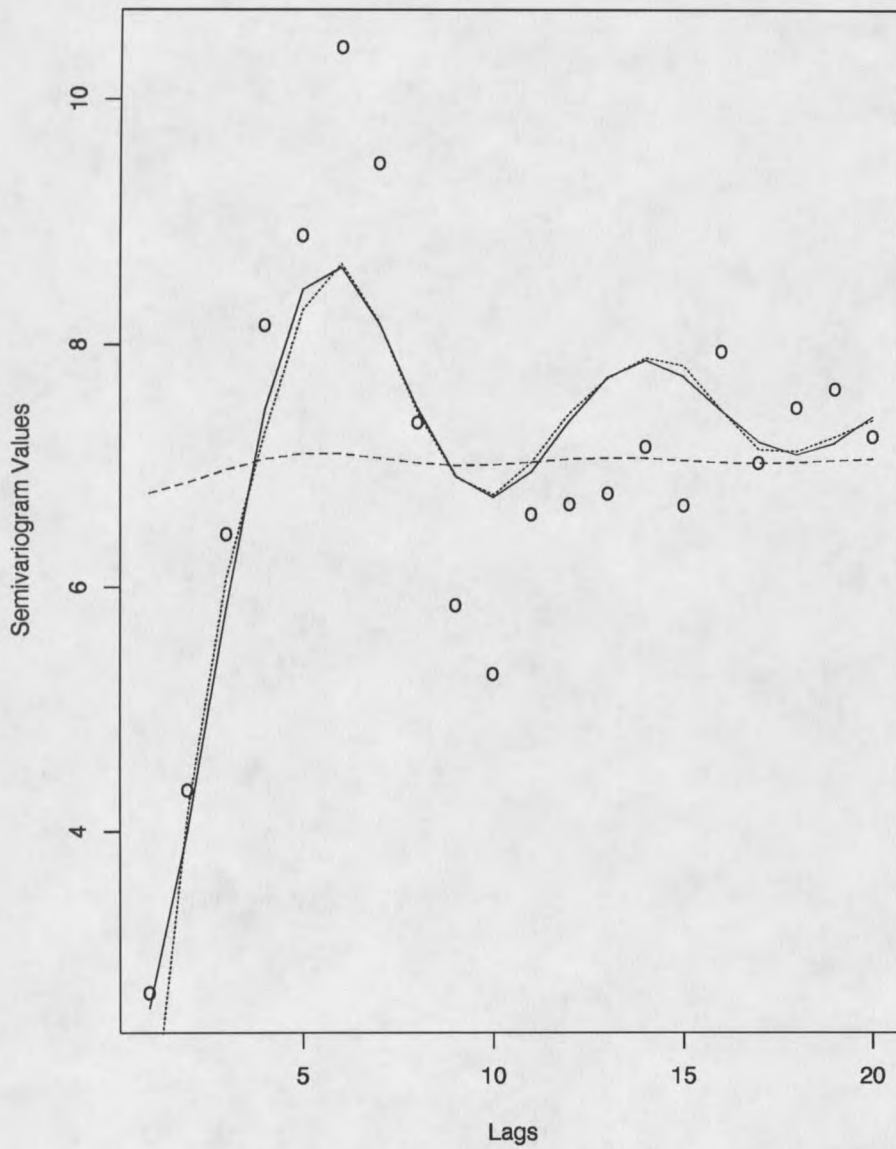


Figure 4: Three different fits to simulated data. The solid line is based on 100 nodes equispaced in  $(0, 4]$  and 100 nodes equispaced in  $[4, 16, 20]$ . The dashed line is based on 200 nodes equispaced in  $(16, 20]$ , and the dotted line is based on 200 nodes equispaced in  $(0, 4]$ .

Table 1: The minimum, 25th percentile, median, 75th percentile, and maximum of the residual norms from 100 fits using four different selections of nodes. The simulated data are from an exponential model with sill and range of 10.

	200 nodes	500 nodes (1st set)	1000 nodes	500 nodes (second set)
minimum	0.6026742	0.5959836	0.595294	0.602674
25th percentile	1.7908759	1.7792969	1.777326	1.789360
median	2.5501313	2.5240973	2.520496	2.546565
75th percentile	4.1687350	4.1480268	4.143043	4.144547
maximum	12.3698721	12.3212223	12.314255	12.328954

- 500 nodes with 250 nodes equispaced in  $(0, 4]$  and 250 equispaced in  $[4.064, 20]$ .
- 1000 nodes with 500 nodes equispaced in  $(0, 4]$  and 500 equispaced in  $[4.008, 20]$ .
- 500 nodes equispaced in  $(0, 20]$ .

Increasing the number of nodes did not increase the accuracy of the fit, but did increase computational time significantly. It appears that the first selection of 200 nodes works well.

### Does the Saturation Lead to Overfitting?

Table 1 provides evidence that increasing the number of nodes does not result in overfitting. There are several reasons for this.

One reason is that the *NNLS* algorithm, which is based on the Kuhn-Tucker optimality conditions, is not likely to produce solutions that just interpolate the data, even with large numbers of nodes.

The following statement of these conditions is from Lawson and Hansen (1974).

**Kuhn-Tucker Conditions:** Let  $\mathbf{A}$  be a  $m \times p$  matrix and  $\mathbf{b}$  and  $\mathbf{x}$  be  $m \times 1$  and  $p \times 1$  vectors respectively. A  $p \times 1$  vector  $\hat{\mathbf{x}}$  is a solution to the problem

$$\min_{\mathbf{x} \geq 0} \|\mathbf{Ax} - \mathbf{b}\|^2 \quad (12)$$

if and only if there exists a  $p \times 1$  vector  $\hat{\mathbf{y}}$  and a partitioning of the integers 1 through  $p$  into subsets  $\mathcal{E}$  and  $\mathcal{S}$  such that

$$\hat{\mathbf{y}} = \mathbf{A}'(\mathbf{A}\hat{\mathbf{x}} - \mathbf{b})$$

$$\hat{x}_i = 0 \text{ for } i \in \mathcal{E}, \quad \hat{x}_i > 0 \text{ for } i \in \mathcal{S}$$

$$\hat{y}_i \geq 0 \text{ for } i \in \mathcal{E}, \quad \hat{y}_i = 0 \text{ for } i \in \mathcal{S}.$$

Proofs of this statement can be found in many standard texts in the theory of constrained optimization and nonlinear programming (*e.g.* Boot, 1964).

The *NNLS* algorithm of Lawson and Hansen (1974) finds a solution that satisfies the Kuhn-Tucker conditions. The algorithm starts by initializing a  $p \times 1$  solution vector  $\hat{\mathbf{x}}$  and a  $m \times p$  matrix  $\mathbf{A}_S$  to a  $p \times 1$  vector of 0's and a  $m \times p$  matrix of 0's, respectively. *NNLS* adds positive elements to  $\hat{\mathbf{x}}$  sequentially. Every time a positive element  $x_j$  is added to  $\hat{\mathbf{x}}$  the corresponding column  $a_j$  of  $\mathbf{A}$  is added to  $\mathbf{A}_S$ . On termination,  $\hat{\mathbf{x}}$  is the unconstrained solution to the problem of minimizing

$$\|\mathbf{A}_S \mathbf{x} - \mathbf{b}\|.$$

Lawson and Hansen (1974) point out that only linearly independent columns will be added to  $\mathbf{A}_S$ . Thus, the number of positive elements in  $\hat{\mathbf{x}}$  will always be less than or equal to the rank of  $\mathbf{A}$ . In practice, the number of positive elements in  $\hat{\mathbf{x}}$  was usually on the order of 4 or 5.

Thus, associated with problem (12) is a well-posed problem in the constrained parameter space that is unlikely to have a solution that just interpolates the points.

In effect, constrained least squares fitting using *NNLS* is a form of regularization that imposes smoothness constraints. This is analogous to penalized least squares routines that are frequently used to deal with problems of overfitting and solving ill-posed inverse problems (*e.g.* Tikhonov regularization). The only time the SB method resulted in interpolation of points was when data was sampled without error from a valid semivariogram function.

Another reason for the lack of overfitting is that the three dimensional version of the SB estimator was chosen. This also automatically imposes smoothness constraints. Figure 5 shows a comparison of plots of  $1 - \Omega_d(x)$  with  $d = 1, 2, 3$ , and  $\infty$  against  $x$ . It is obvious that valid nonparametric semivariograms constructed from the higher dimensional versions will be smoother than those constructed from the lower dimensional versions. This heuristic argument is made rigorous in Schoenberg (1938, p. 822) where he proves that the class of functions positive definite in  $\mathbf{R}^d$  have  $[(d - 1)/2]$  derivatives (where  $[(d - 1)/2]$  denotes the largest integer

not exceeding  $(d - 1)/2$ ). This indicates the smoothing effect of the requirement of positive definiteness (and, hence, conditional negative definiteness) in higher dimensional Euclidean space. For  $d > 2$ , positive definite functions are everywhere differentiable, and the class of positive definite functions valid in any dimension are infinitely differentiable (Yaglom, 1987a, p. 357).

Even if interpolation is not a problem some investigators may still feel that the estimated functions are not smooth enough. There are several options available to deal with this problem. Shapiro and Botha (1991) show how it is possible to impose monotonicity and convexity constraints on the fits by imposing appropriate constraints on first and second derivatives. Also, choosing  $d = \infty$  will result in a monotonic increasing fit. It is also possible to get sample semivariograms that are smoother than the method of moments estimator used here by smoothing the semivariogram cloud in some other fashion, *e.g.* splines, kernel regression, locally weighted regression, and so on. The SB method could then be applied to these smoother sample semivariograms to yield valid fits. None of those methods of imposing smoothness constraints were used here.

### Comparison of Fits

This section will present a comparison of results from fitting a nonparametric model and various parametric models to simulated data sets. Five different models

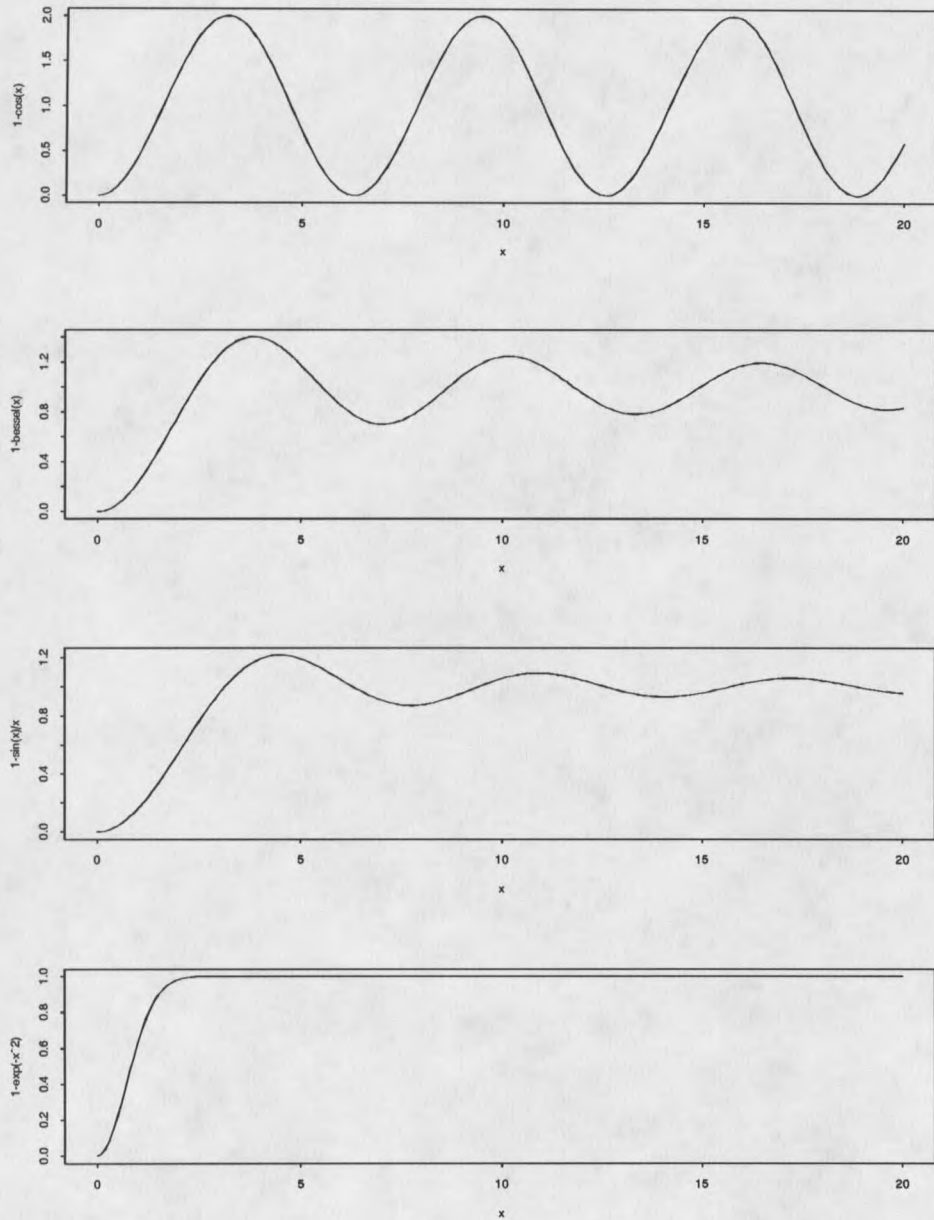


Figure 5: This figure shows increasing *smoothness* of  $1 - \Omega_d(x)$  for  $d = 1$  (top panel),  $d = 2$  (second panel),  $d = 3$  (third panel), and  $d = \infty$  (bottom panel).

were chosen for the simulations:

- exponential model with sill = 10 and range = 2
- exponential model with sill = 10 and range = 6
- exponential model with sill = 10 and range = 10
- exponential model with sill = 10 and range = 14
- exponential model with sill = 10 and range = 18

The exponential model was chosen because it is easy to simulate and is a popular choice for parametric fitting. A total of 100 random fields was simulated for 50 locations spaced one unit apart on a one dimensional transect. The nugget effect was set to 0 for each of these models.

Four different parametric models (exponential, Gaussian, rational quadratic, and power) were fit to the sample semivariograms using a nonlinear least squares program (UMSOLVE) in Matlab.

The UMSOLVE software uses Newton's method with some modifications to find a solution. The iteration limit for UMSOLVE was set to 500. If the algorithm failed to converge it indicated that by a termination code. This happened occasionally, and no special effort was made to correct the problem when it occurred.

The power model was included even though it cannot be the semivariogram for a second-order stationary process because it does not have a sill. Such models are

generally considered appropriate only if one believes that the random process has an infinite variance. However, the model is used in practice and can give good fits to data from a second-order stationary process in which all the  $Y_{ij}$ 's in the semivariogram cloud occur at lags less than the range.

Initial starting values for the parameters in the parametric fits were the true values of the underlying exponential model, except for the power model where starting values consistent with a straight line with a slope of 1 were chosen.

The fits were compared by computing the integrated squared error (*ISE*) over the interval  $[0, 20]$  where

$$ISE = \int_0^{20} (\tilde{\gamma}(h) - \gamma(h))^2 dh$$

was evaluated using the trapezoid rule ( $\tilde{\gamma}(h)$  is the estimated semivariogram and  $\gamma(h)$  is the true semivariogram function).

Figures 6 through 10 show boxplots of the resulting *ISE*'s from the fits. Figure 6 only gives results for the Shapiro-Botha fits and the exponential and Gaussian models because the rational quadratic and power models fit so poorly, their *ISE*'s swamped the results from the other fits if they were included in the figure.

Table 2 gives the minimum, maximum, and the 25th, 50th, and 75th percentiles for the *ISE* values calculated from the simulations.

The *NNLS* procedure always converged. The nonlinear least squares program occasionally failed to converge. This was a problem for the rational quadratic model

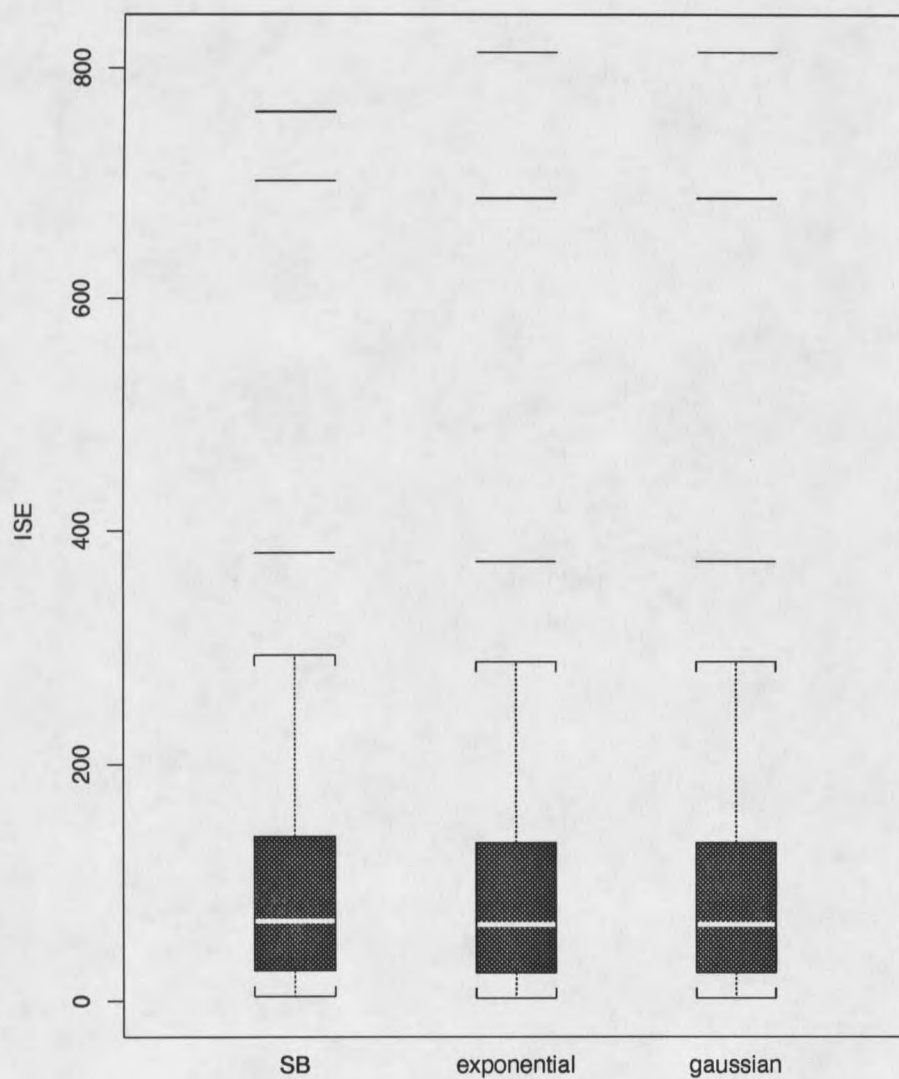


Figure 6: Boxplots of integrated squared errors from nonparametric (SB) and parametric models (exponential, Gaussian) fit to 100 simulated data sets based on exponential model with sill of 10 and range of 2.

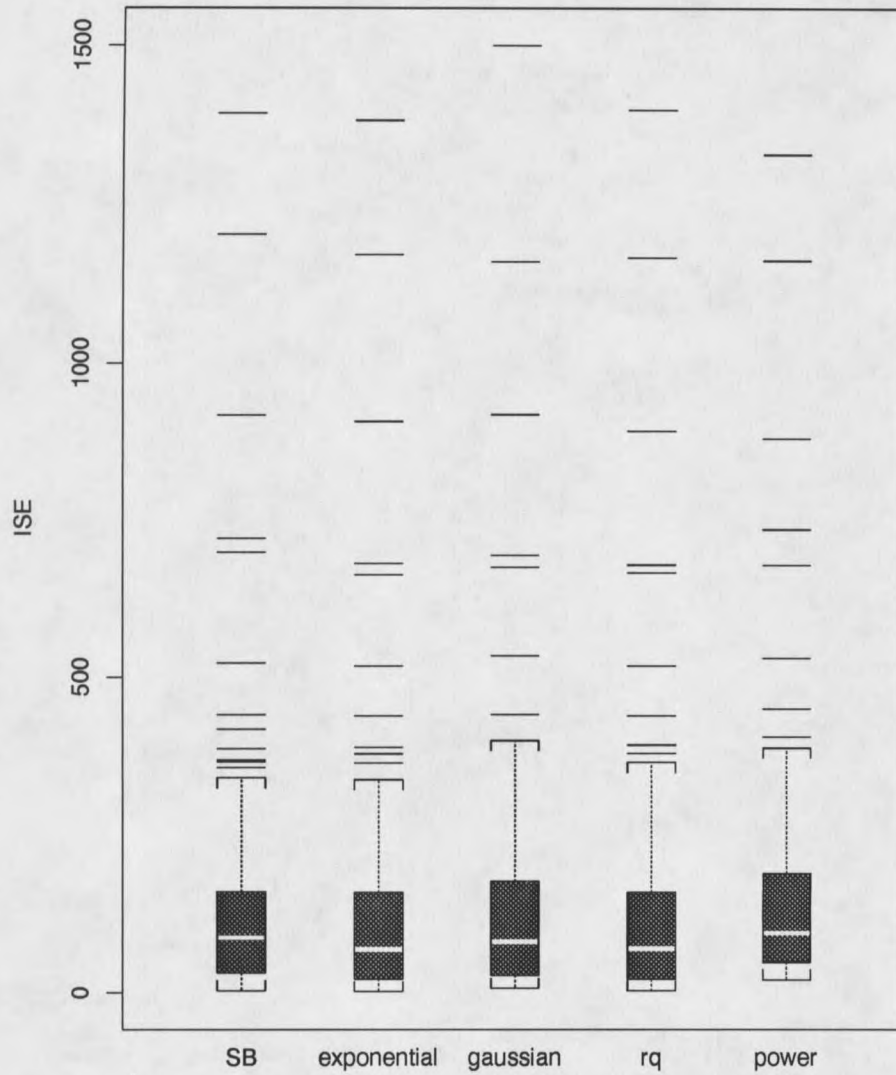


Figure 7: Boxplots of integrated squared errors from nonparametric (SB) and parametric models (exponential, Gaussian, rational quadratic, and power) fit to 100 simulated data sets based on exponential model with sill of 10 and range of 6. *rq=rational quadratic model*

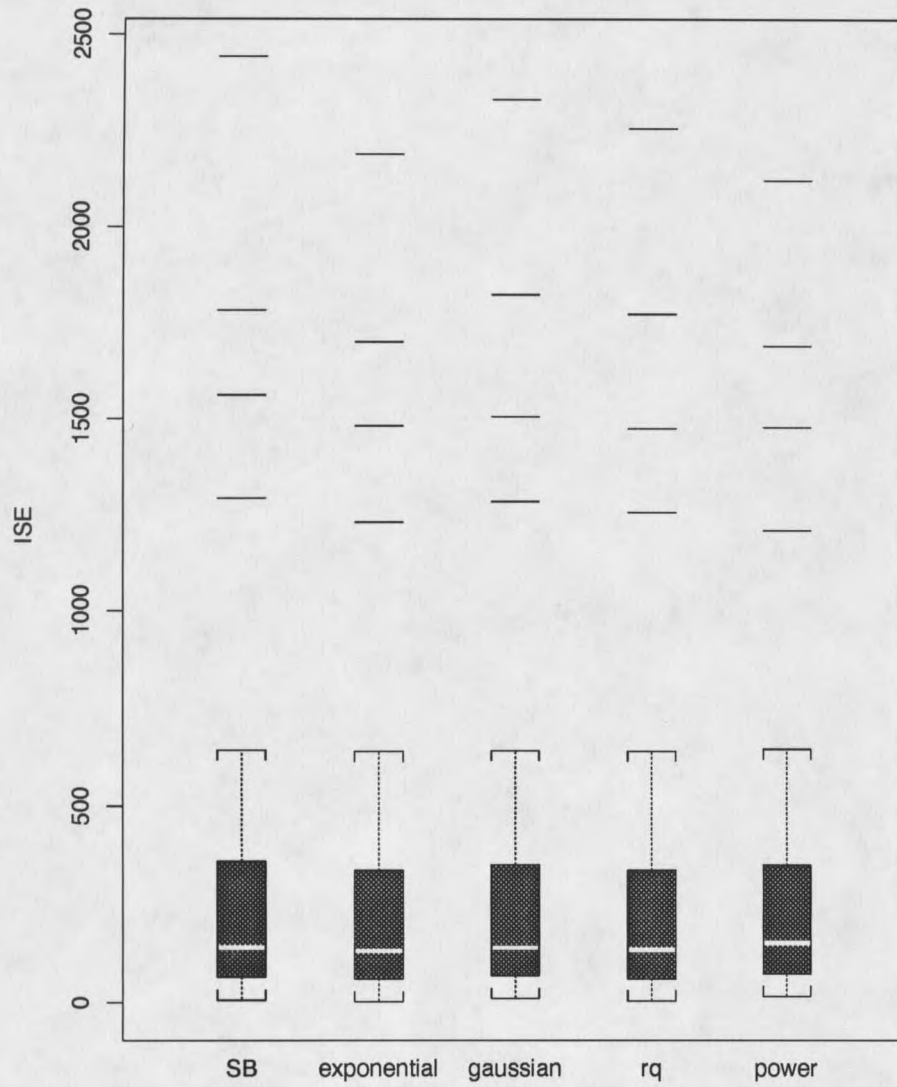


Figure 8: Boxplots of integrated squared errors from nonparametric (SB) and parametric models (exponential, Gaussian, rational quadratic, and power) fit to 100 simulated data sets based on exponential model with sill of 10 and range of 10. *rq=rational quadratic model*

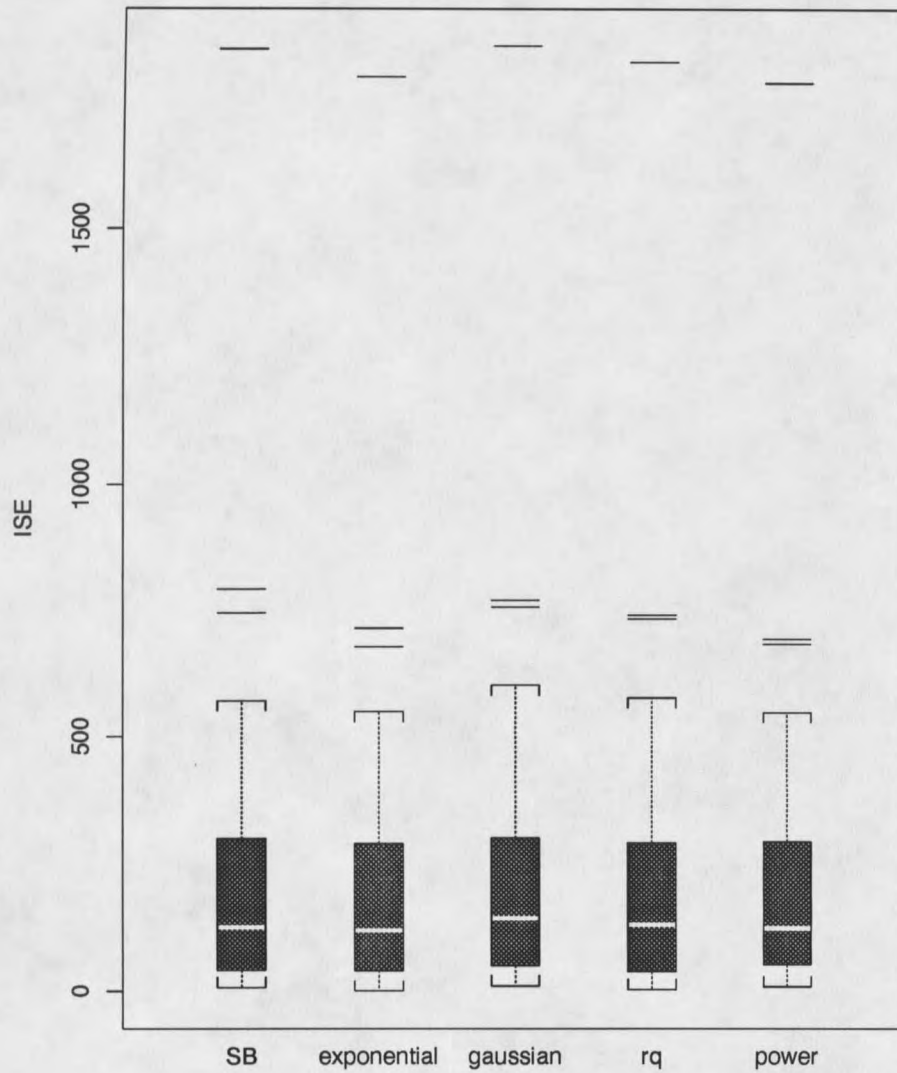


Figure 9: Boxplots of integrated squared error from nonparametric (SB) and parametric models (exponential, Gaussian, rational quadratic, and power) fit to 100 simulated data sets based on exponential model with sill of 10 and range of 14. *rq=rational quadratic model*

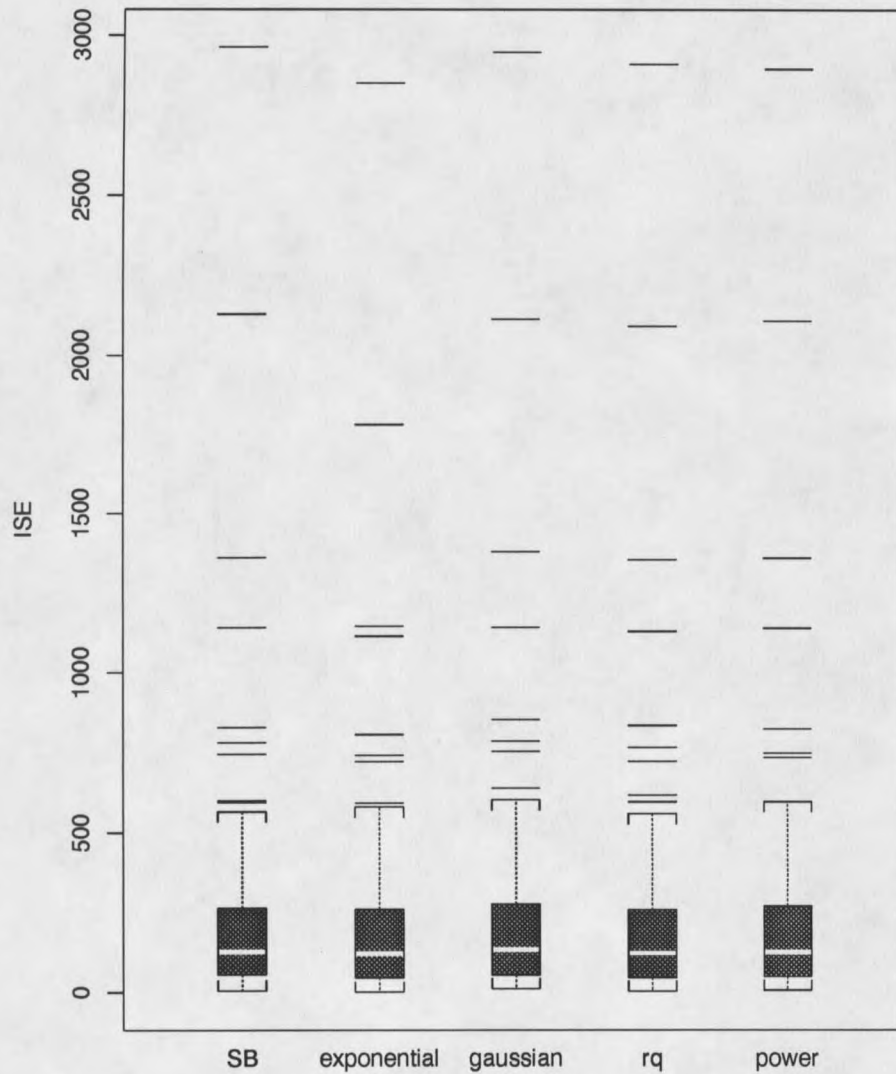


Figure 10: Boxplots of integrated squared error from nonparametric (SB) and parametric models (exponential, Gaussian, rational quadratic, and power) fit to 100 simulated data sets based on exponential model with sill of 10 and range of 18. *rq=rational quadratic model*

Table 2: The minimum, 25th percentile, median, 75th percentile, and maximum ISE values from nonparametric and parametric semivariogram models fit to 100 simulated data sets. In each case the true semivariogram model was exponential with no nugget effect and a sill of 10. The ranges varied as indicated.

Range	Percentile	SB model	exponential model	Gaussian model	rational quadratic	power model
2	minimum	3.967	2.294	2.294	9.709	79797.97
	25th	25.839	23.660	23.661	534.160	226248.16
	median	68.134	64.769	64.769	1857.099	356117.07
	75th	138.244	133.727	133.727	5477.218	536841.67
	maximum	762.897	813.441	813.441	66672.128	2639251.15
6	minimum	3.222	2.032	7.120	2.609	20.095
	25th	32.085	22.101	27.355	21.134	48.010
	median	86.630	68.382	80.821	69.362	93.699
	75th	159.088	157.198	174.979	157.651	184.748
	maximum	1392.476	1380.905	1500.226	1397.236	1327.763
10	minimum	5.944	1.786	9.919	3.525	15.001
	25th	65.502	59.870	68.678	60.039	72.241
	median	140.759	131.855	139.931	134.361	152.91545
	75th	357.362	327.911	345.526	330.643	342.240
	maximum	2441.958	2190.026	2331.541	2257.026	2121.190
14	minimum	7.551	0.940	10.857	3.557	9.206
	25th	42.272	40.685	50.050	39.226	53.733
	median	125.518	119.475	143.126	130.514	123.438
	75th	296.722	286.539	299.189	289.0342	287.391
	maximum	1859.669	1804.287	1866.105	1833.577	1791.904
18	minimum	4.939	0.379	10.825	3.480	5.514
	25th	55.466	44.949	53.186	46.889	48.732
	median	125.296	119.089	131.374	121.087	123.228
	75th	253.437	252.528	272.817	251.467	267.964
	maximum	2962.943	2850.684	2947.621	2910.382	2895.873

when the simulated data came from the exponential model with a range of 2 (10 of the 100 attempted fits failed to converge), and the exponential model when the simulated data came from the exponential model with a range of 18 (6 of the 100 attempted fits failed to converge).

The failure of the rational quadratic to adequately fit the data from the simulation with a range of 2 is puzzling. It has the same general shape as the true underlying exponential model. The failure of the power model is not surprising at all. The exponential model with a sill of 10 and a range of 2 rises quickly to its asymptote. The power model simply cannot track this behavior. It is not really fair to consider the power model in this case. This illustrates a problem with evaluating fits with simulated spatial data. The large number of simulations require an automated approach. But fitting semivariograms is, even with the nonparametric approach, subjective. No competent practitioner, looking at the semivariogram clouds from the 100 simulations based on the exponential model with sill of 10 and range of 2 would willingly choose to fit a power model. Similarly, many practitioners might well choose to fit a power model to the data from the simulations based on an exponential model with sill of 10 and range of 18. The semivariogram clouds determined from these simulations yielded sample semivariograms that look as if they could be fit very well with a power model, as indeed they were (Figure 10). The range of 18 is close to the maximum lag, and experimental evidence for the

Table 3: The minimum, 25th percentile, median, 75th percentile, and maximum *ISE* values from nonparametric and parametric semivariogram models fit to 100 simulated data sets. The true semivariogram model was a mixture of rational quadratic and hole effect models with no nugget effect and a sill of 14.

Percentile	SB model	exponential model	Gaussian model	rational quadratic	power model
minimum	3.694	39.055	17.633	30.417	105.201
25th	89.708	107.158	81.392	93.489	169.360
median	273.344	235.613	201.055	204.490	305.200
75th	535.758	490.064	479.795	480.109	547.420
maximum	2513.583	2729.247	2751.936	2740.215	2753.768

existence of the sill and range is missing from many of the sample semivariograms.

Data were also generated for a random process for which the true semivariogram function was a sum of a rational quadratic model and the hole effect model valid for three dimensional spatial data.

The nugget effect was set to zero and the simulated process had a true semivariogram given by

$$\gamma(h) = 10(1 - \sin(h)/h) + 2h^2/(1 + (h^2/2)).$$

The sill for this semivariogram is 14. A boxplot of the *ISE* values from the nonparametric fit and the four parametric fits is shown in Figure 11, and the minimum, 25th percentile, median, 75th percentile, and maximum *ISE* values for the five comparisons is shown in Table 3. The nonparametric fits again compare favorably with the parametric fits.

The results of this chapter show that the nonparametric method proposed by SB

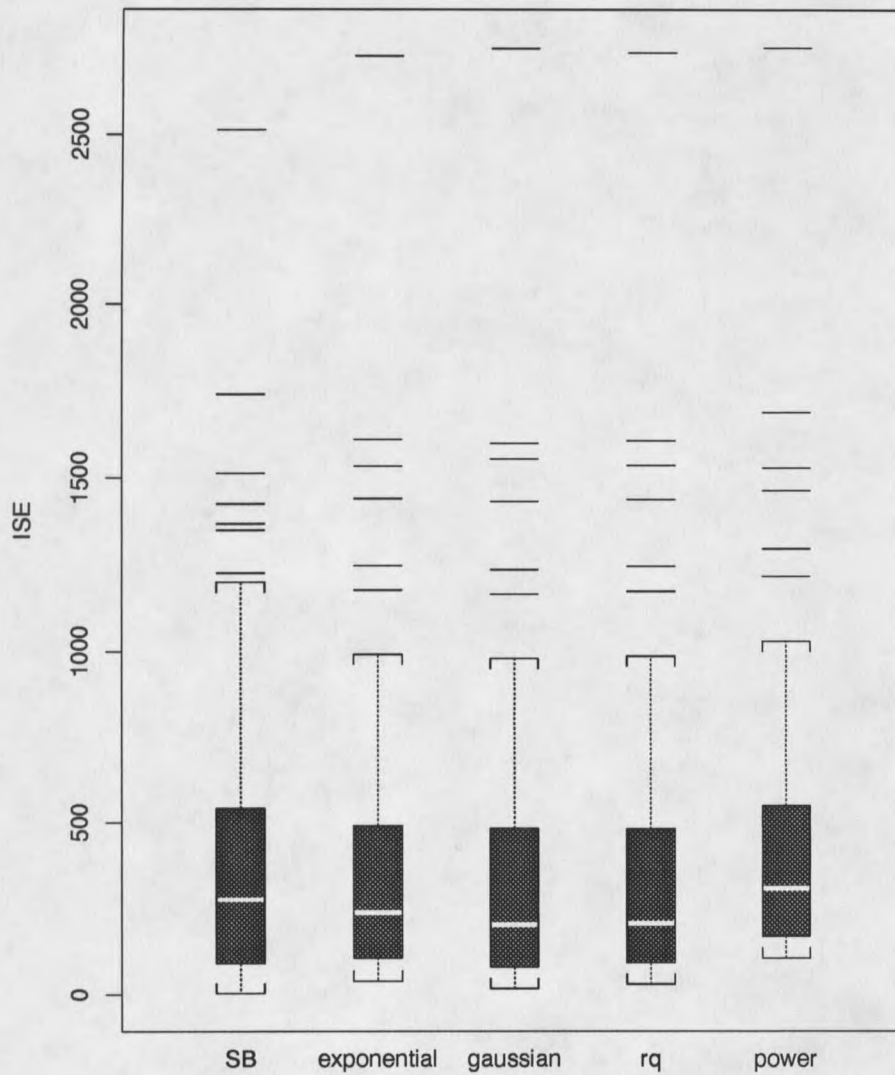


Figure 11: Boxplots of integrated squared error from nonparametric (SB) and parametric models (exponential, Gaussian, rational quadratic, and power) fit to 100 simulated data sets based on a mixture of rational quadratic and hole effect models. *rq=rational quadratic model.*

fits as well as the parametric models currently used (based on the comparison of *ISE*'s). It is more objective than parametric fitting and it is easier to implement, even if the parametric models are being fit by inspection. The *NNLS* algorithm is fast and always converges, and there are many similar algorithms available in the quadratic programming literature that would undoubtedly work as well. The method is fairly robust to the selection of nodes.

The advantages of the SB method become even more apparent when one realizes that fitting semivariograms using parametric models is rarely as straightforward as it was in this chapter. In general, isotropic data is rare and the first step in fitting semivariogram models is to identify axes of anisotropy and fit isotropic models along these axes. Also, it is common for geostatisticians to have to fit mixtures of parametric models to get good fits. This is one reason why fitting by inspection is so common. Fitting such semivariogram models using nonlinear least squares algorithms would be difficult at best.

## CHAPTER 4

### ESTIMATION OF THE SILL

There are three intrinsic parameters of interest in semivariogram estimation; the nugget effect, the sill, and the range.

The first section of this chapter will describe how the SB method can be used to yield an estimate of the sill. The sill estimates obtained tend to be biased and highly variable. The second section discusses how to stabilize the sill estimates.

Nonparametric estimation of the nugget effect and the range are more difficult problems and will be discussed further in Chapter 7.

#### The Sill

Recall that the sill ( $s_p$ ) is defined as

$$s_p = \lim_{h \rightarrow \infty} \gamma(h) = C(\mathbf{0}).$$

But,

$$C(\mathbf{0}) = \int_0^\infty dM(t)$$

and so the SB estimator of the semivariogram function yields an estimate of the sill ( $\hat{s}_p$ ) given by

$$\hat{s}_p = \sum_{i=1}^n \tilde{p}_i.$$

Accurate estimation of the sill is important for assessing the uncertainty of predictions at unsampled locations obtained using kriging. Prediction intervals in particular require good sill estimates. Even if prediction is not a goal, the estimation of the variance of  $Z(\mathbf{s})$  is required for any attempt to adequately characterize the properties of a spatial random process. Issacs and Srivastava (1989) and Cressie (1991) both discuss the importance of obtaining good estimates of the sill.

Figures 12 through 16 show boxplots of sill estimates from the SB nonparametric and parametric models fit to the simulated data sets based on the exponential models described in Chapter 3. It is obvious that the estimates provided by the SB method are biased. Further, these estimates are more variable than the estimates provided by fitting the parametric models.

Table 4 presents the minimum, 25th percentile, median, 75th percentile, and maximum of those sill estimates. The true sill value is 10. Tables 5 and 6 show the estimated mean squared error and means for the sill estimates. None of the tables present results from the power model fits, because there is no sill in the power model.

These estimates are based on 100 simulations for each range except for the

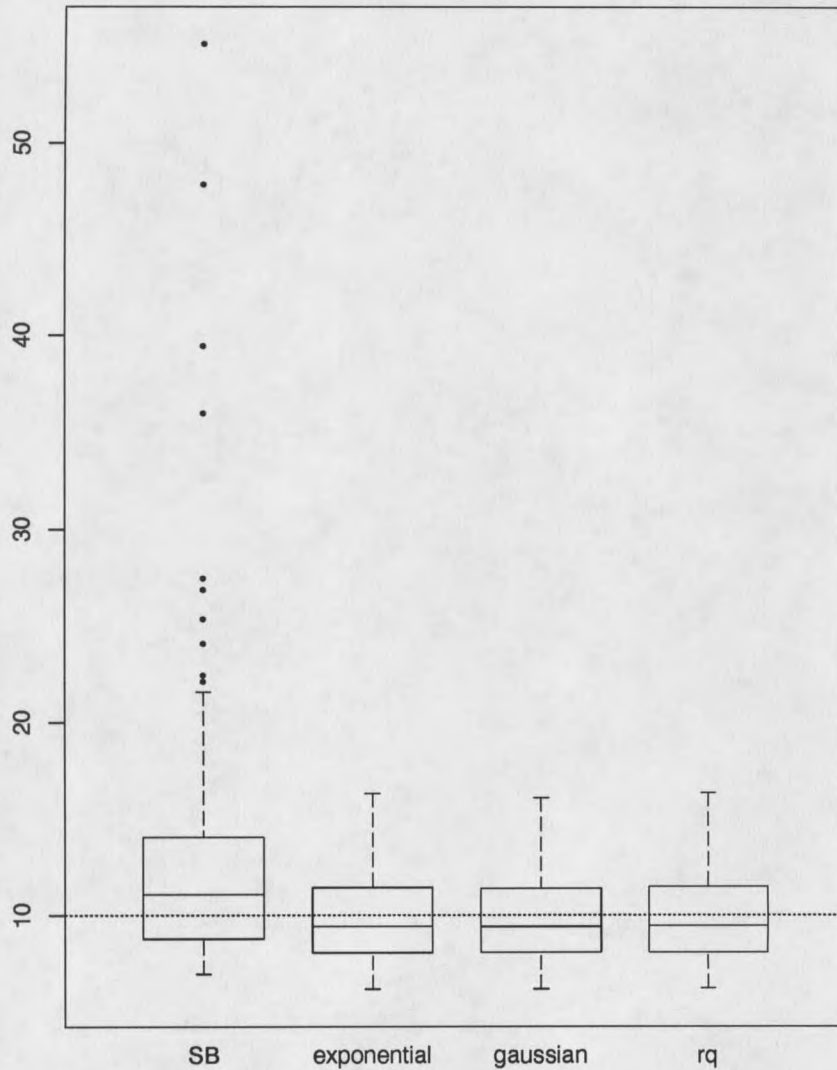


Figure 12: Boxplots of sill estimates from nonparametric (SB) and parametric models (exponential, Gaussian, and rational quadratic) fit to simulated data sets based on exponential model with sill of 10 and range of 2.

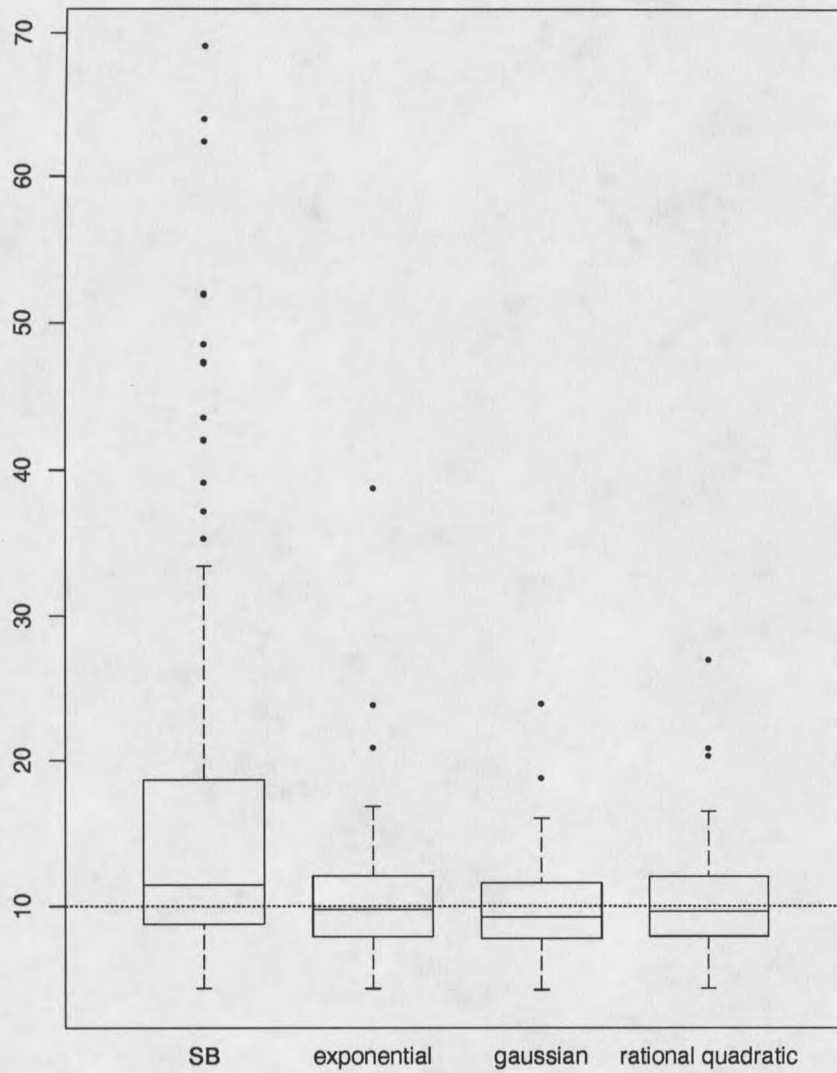


Figure 13: Boxplots of sill estimates from nonparametric (SB) and parametric models (exponential, Gaussian, and rational quadratic) fit to simulated data sets based on an exponential model with sill of 10 and range of 6.

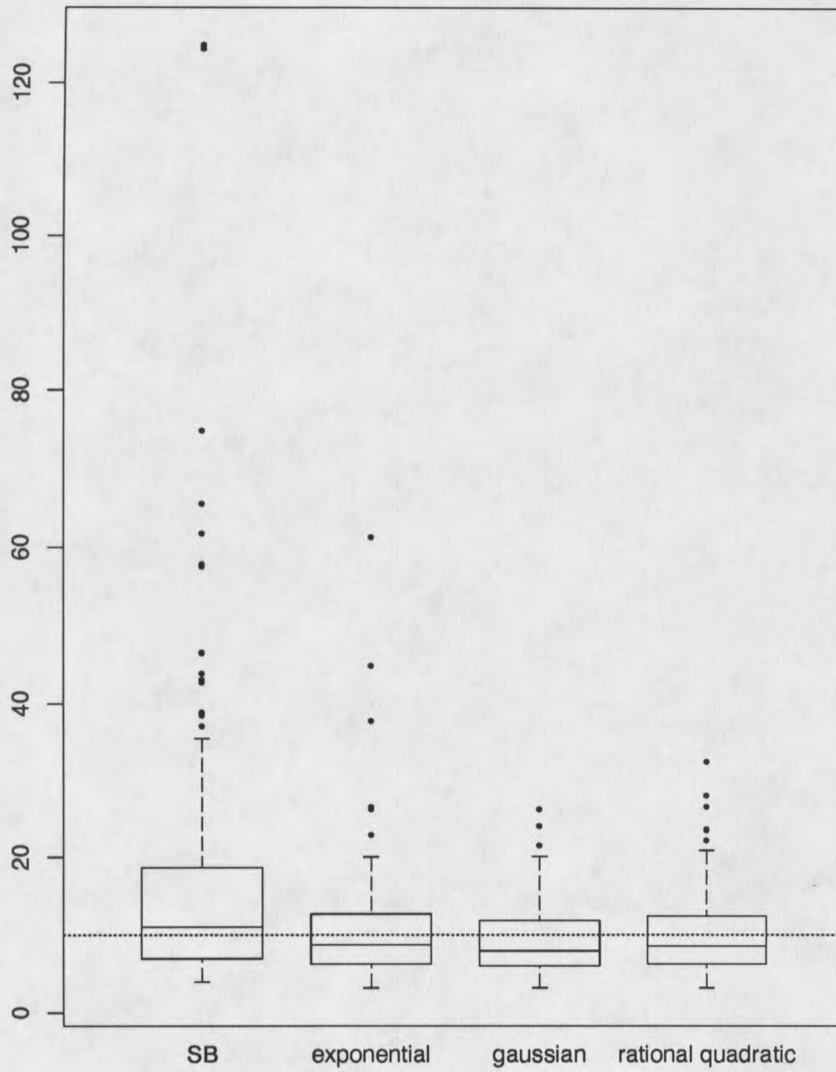


Figure 14: Boxplots of sill estimates from nonparametric (SB) and parametric models (exponential, Gaussian, and rational quadratic) fit to simulated data sets based on an exponential model with sill of 10 and range of 10.



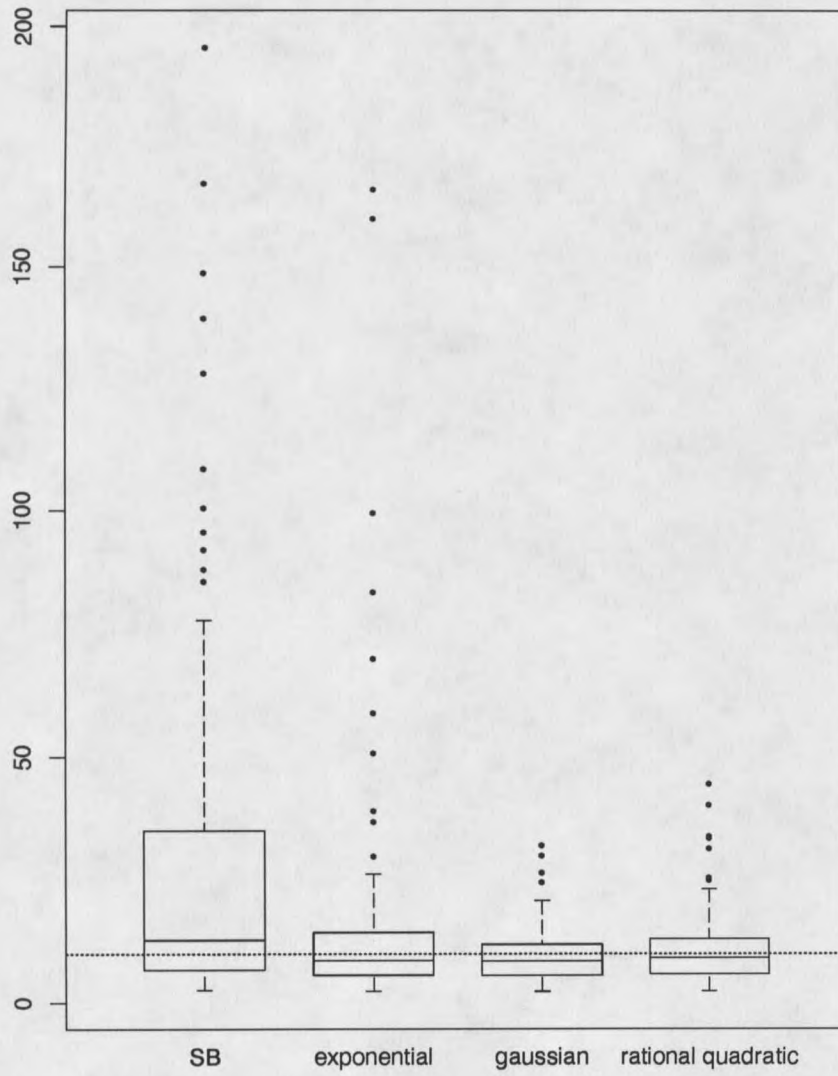


Figure 16: Boxplots of sill estimates from nonparametric (SB) and parametric models (exponential, Gaussian, and rational quadratic) fit to simulated data sets based on an exponential model with sill of 10 and range of 18.

Table 4: The minimum, 25th percentile, median, 75th percentile, and maximum for sill estimates from five different simulations. In each case the true semivariogram model was exponential with no nugget effect and a sill of 10. The ranges varied as indicated.

Range	Percentile	SB model	exponential model	Gaussian model	rational quadratic
2	minimum	6.930	6.146	6.141	6.188
	25th	8.807	8.072	8.077	8.086
	median	11.050	9.425	9.400	9.463
	75th	13.960	11.376	11.357	11.431
	maximum	55.167	16.291	16.064	16.325
6	minimum	4.331	4.317	4.228	4.338
	25th	8.793	7.920	7.766	7.931
	median	11.460	9.754	9.237	9.650
	75th	18.654	12.062	11.566	12.020
	maximum	69.070	38.736	23.899	26.944
10	minimum	3.984	3.186	3.177	3.192
	25th	6.923	6.317	6.027	6.272
	median	10.988	8.699	7.902	8.555
	75th	18.612	12.557	11.745	12.319
	maximum	124.882	61.194	26.211	32.403
14	minimum	3.203	3.103	3.065	3.132
	25th	7.261	5.933	5.611	5.884
	median	11.025	8.163	7.715	8.104
	75th	24.315	12.457	10.787	11.757
	maximum	143.519	185.851	24.892	29.911
18	minimum	2.612	2.372	2.264	2.346
	25th	6.772	5.655	5.570	5.847
	median	12.856	8.901	8.585	9.215
	75th	34.248	14.266	11.921	13.098
	maximum	195.521	165.919	31.962	44.530

Table 5: Estimated mean squared error of sill estimates from simulated data sets for each model-range combination. The true sill is equal to 10.

Range	SB model	exponential model	Gaussian model	rational quadratic
2	74.903	5.093	4.816	4.998
6	252.422	18.815	10.276	12.346
10	531.663	68.969	21.344	30.269
14	826.264	470.940	19.674	27.886
18	1854.353	771.269	33.652	63.040

Table 6: Means of the sill estimates from simulated data sets for each model-range combination. The true sill is equal to 10.

Range	SB model	exponential model	Gaussian model	rational quadratic
2	13.407	9.967	9.900	9.990
6	17.126	10.394	9.856	10.214
10	18.738	10.871	9.369	10.064
14	21.975	13.784	8.769	9.622
18	29.698	16.751	9.639	11.098

exponential model fit to the simulated data from the models with ranges of 10 and 18. There was one extreme sill estimate of 54,258.18 in the former case, and there were a total of 7 sill estimates that were over 29,000 in the latter setting. The nonlinear fitting procedure failed to converge for 6 of those 7. These outliers are not included in the tables or figures.

Even without those fits, the sill estimates from fitting an exponential model to the data were not as good as those from using the Gaussian or rational-quadratic models. This is surprising because the true underlying model is exponential. However, parameter estimation in nonlinear models of this general type is a surprisingly

Table 7: The minimum, 25th percentile, median, 75th percentile, and maximum for the sill estimates from the rational quadratic - hole effect model. The true sill value is 14.

Percentile	SB model	exponential model	Gaussian model	rational quadratic
minimum	4.425	4.338	4.233	4.375
25th	10.264	9.432	9.469	9.600
median	14.334	12.669	12.467	12.676
75th	19.639	15.641	15.412	15.781
maximum	139.897	32.779	26.122	27.199

difficult problem. Seber and Wild (1989) devote several pages to a discussion of identifiability and ill-conditioning problems encountered in fitting what is essentially an exponential semivariogram model.

The results from the data set with a range of 18 are suspect for all of the models considered. The sill is not reached until lag 18 which is almost outside the range of the data. Given the variability in the data, it is not surprising that the estimates from all the fits, both parametric and nonparametric, were the worst for this example.

Figure 17 is a boxplot of the 100 sill estimates from nonparametric and parametric fits to simulated data sets where the true semivariogram model is the rational quadratic hole effect model. Table 7 is a table of the minimum, 25th percentile, median, 75th percentile, and maximum of those 100 sill estimates. Table 8 presents the estimated mean squared error and means for the sill estimates.

One approach to getting better sill estimates is to choose nodes to custom fit a

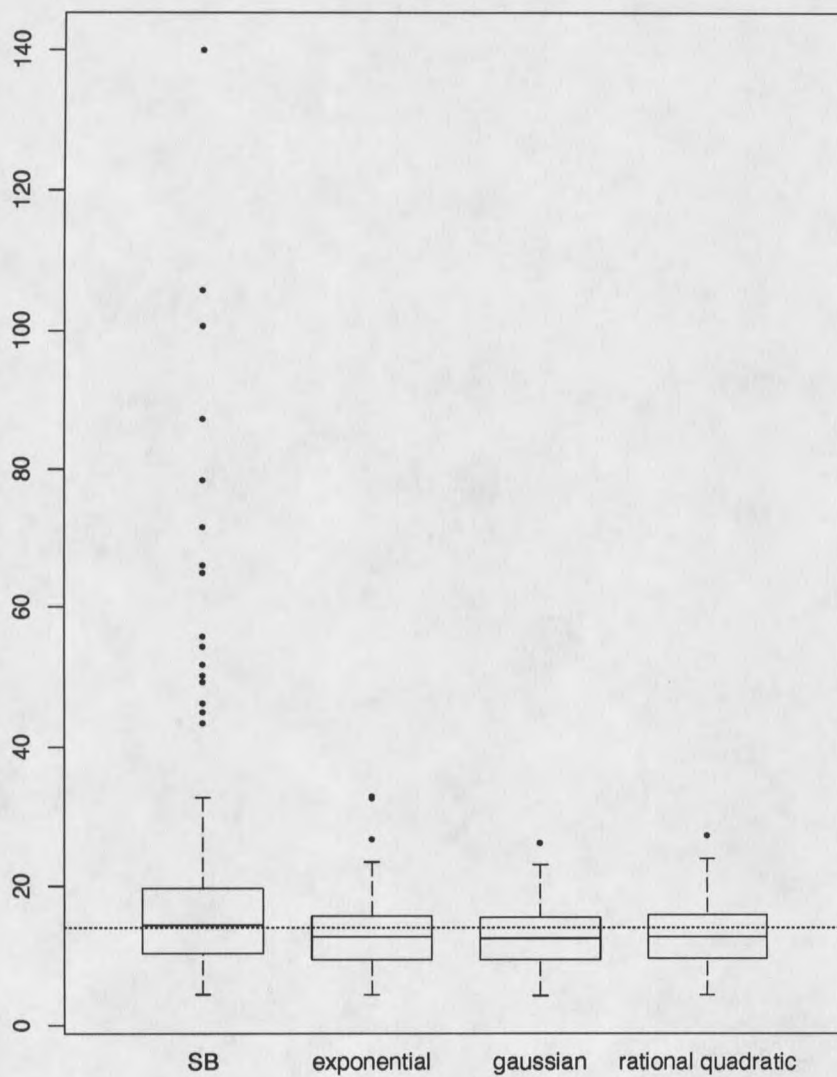


Figure 17: Boxplots of sill estimates from nonparametric (SB) and parametric models (exponential, Gaussian, and rational quadratic) fit to 100 simulated data sets based on a rational quadratic-hole effect model. The true sill is 14.

Table 8: Estimated mean squared error and means for sill estimates from a rational quadratic-hole effect model with true sill value of 14. (*rq-rational quadratic*)

Model	Estimated Mean Squared Error	Mean
SB	627.890	22.597
exponential	28.197	13.302
Gaussian	20.836	12.761
rq	22.030	13.113

given sample semivariogram. For example, Figure 18 shows a sample semivariogram from a simulation where the true semivariogram was exponential with sill and range of 10. The solid line is the estimated semivariogram based on the nodes used above. The fit is good, but the sill estimate is 33.32, which is clearly too high. The dashed line is a fit using a collection of 90 nodes equispaced in  $(0, 18]$ . The sill estimate based on this fit is 8.08 which is much better.

This does not seem to be a desirable way to proceed, however. It would be preferable to not have to worry about custom fitting node selection to the data. As noted above, one of the strengths of the nonparametric approach is that it has the capability of removing some of the subjectivity of current methods of fitting semivariogram functions. Given a suitable collection of nodes, can the sill estimates be constrained or stabilized in some statistically defensible way to provide better estimates without affecting the fit? This question is addressed in the following section.

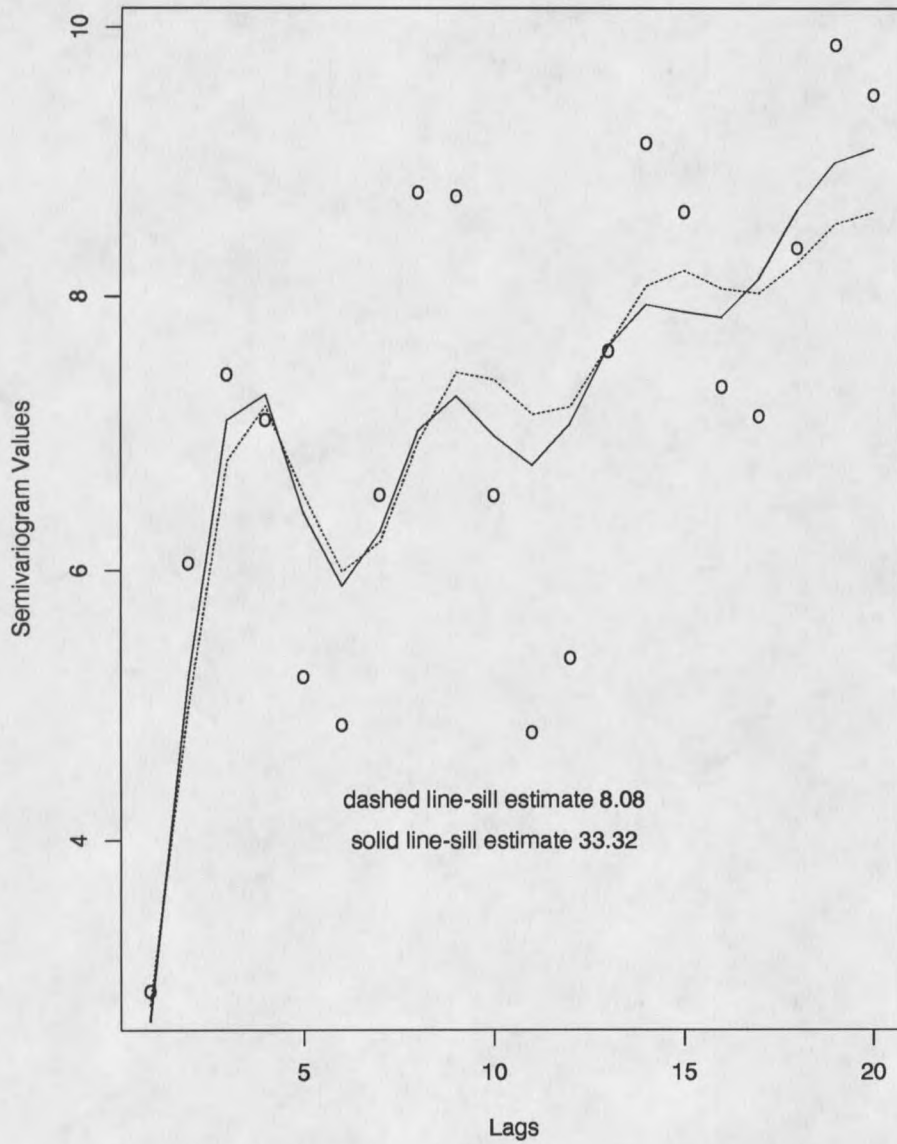


Figure 18: Two different nonparametric semivariogram fits to a sample semivariogram based on two different node collections.

Stabilizing the Sill Estimate

Recall that the problem under consideration is to minimize

$$\| \mathbf{A}\mathbf{p} - \hat{\gamma} \|$$

subject to the constraint that  $\mathbf{p} \geq 0$ . An additional constraint on the sill estimate can be incorporated by minimizing

$$\left\| \begin{pmatrix} \mathbf{A} \\ \sqrt{\lambda} \mathbf{1}'_p \end{pmatrix} \mathbf{p} - \begin{pmatrix} \hat{\gamma} \\ 0 \end{pmatrix} \right\|$$

subject to the constraint that  $\mathbf{p} \geq 0$ .  $\mathbf{1}_p$  is a  $p \times 1$  vector of ones, and  $\lambda$  is a nonnegative scalar penalty term. The goal is to choose a value of  $\lambda$  that gives a penalized sill estimate with little effect on fit. The sill estimates should thus be stabilized.

This approach is similar to ridge regression. Ridge regression could in fact be used to stabilize the parameter (jump) estimates. However, if the collection of nodes is large, then ridge regression results in a much larger system of equations that can become computationally awkward to handle. Also, the goal of ridge regression is to stabilize the estimates of the individual parameters (jumps) in the model. Thus, it only indirectly addresses the problem of instability in the sill estimate. The above approach results in a system of equations that is not much more computationally difficult to handle and directly addresses the question of stabilizing the sill estimate.

Note also that this penalized fitting procedure should produce sill estimates that are both less biased and less variable. In fact, it should theoretically be possible to find the value of  $\lambda$  that gives unbiased sill estimates. This is a difficult analytical problem, however. Practical ways of choosing  $\lambda$  are discussed below.

### Choosing the Penalty Term

Figure 19 shows two ways of looking at how increasing values of  $\lambda$  simultaneously affect sill estimates and fit. The upper panel shows a plot of sill estimates versus the residual norm for 45 values of  $\lambda$ , from  $\lambda = 10^{-9}$  to  $\lambda = 100$ . These values of  $\lambda$  were chosen so that log to base 10 ( $\log_{10}$ ) values of  $\lambda$  were equispaced on  $[-9, 2]$  in increments of 0.25. The sample semivariogram being fit is the same as that in Figure 18. The lower panel shows a plot of the sill estimates and residual norms plotted against the 45  $\log_{10}(\lambda)$  values. With small values of  $\lambda$  the sill estimates are high and then, as  $\lambda$  increases the sill estimates begin to decrease. This initial decrease is not accompanied by an increase in the residual norm. Thus, the fit is not appreciably affected by smaller values of  $\lambda$ .

A reasonable approach, similar to that taken in choosing the penalty term in ridge regression and other penalized fitting methods, is to choose a value of  $\lambda$  where the residual norm starts to increase and use the sill estimate from that fit. To get an idea of how such an approach might work, a large number of random

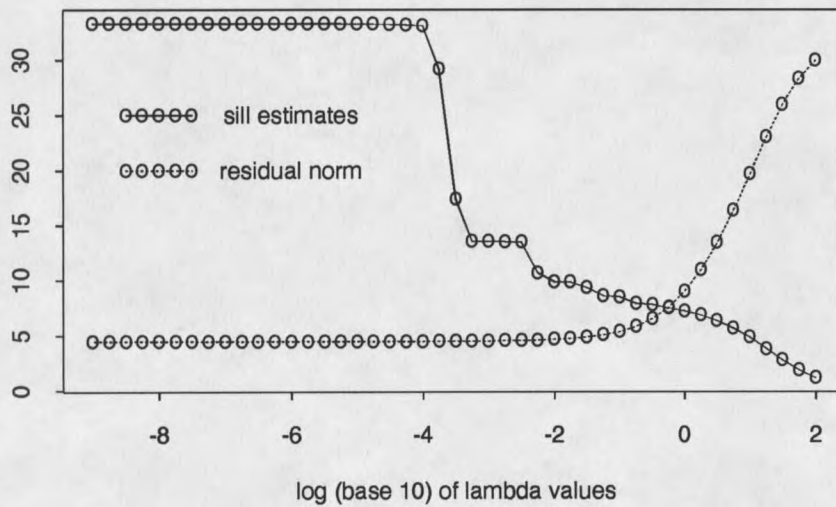
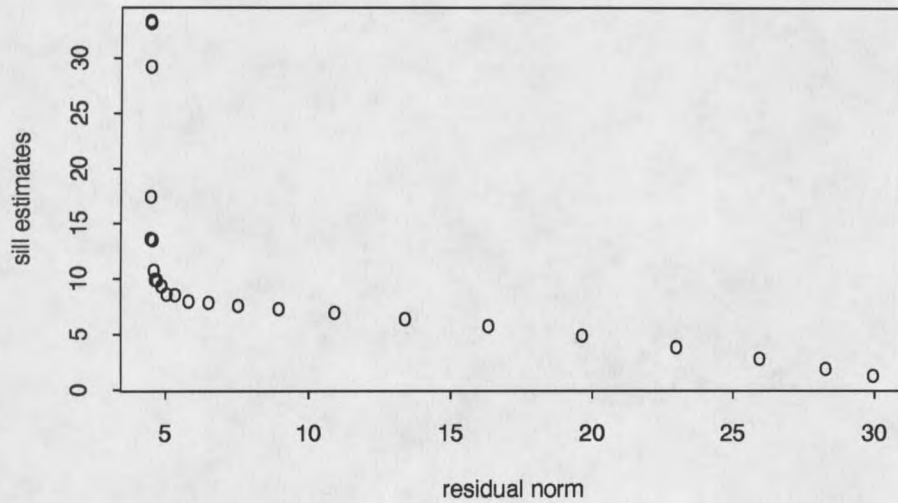


Figure 19: Results of penalized fitting procedure on a simulated data set. Upper panel shows a plot of sill estimates versus the residual norm for 45 values of  $\lambda$ . Lower panel shows a plot of sill estimates and residual norms versus 45 values of  $\log_{10}(\lambda)$ .

fields were simulated based on the following three models,

- an exponential model with sill and range of 10
- a mixture of a rational quadratic and a hole effect model with a sill of 14
- a mixture of two spherical models with a sill of 4

There were 1100 simulations for the exponential model and 1000 for the other two models.

As described above, the data was assumed to fall on a one dimensional transect with 50 observations one unit apart. The semivariogram clouds and method of moments estimators were determined for these random fields. The method of moments estimator was only applied to the first 20 lags. The penalized fitting procedure was carried out and the average of the sill estimates and the average of the residual norms were calculated from the fits for each model.

Figures 20, 21, and 22 show plots of the average sill values versus the average norms of the residuals for the three different models. There are 45 points in each of the graphs with the left most point corresponding to  $\lambda = 10^{-9}$  and the right most point corresponding to  $\lambda = 100$ . Thus, the plots show the relationship between average sill estimates and average fit under increasing values of  $\lambda$ . In each case the fit starts to deteriorate at a sill value close to the true sill of the underlying models.

To get a better idea of the relationship, the average sill and residual norm were

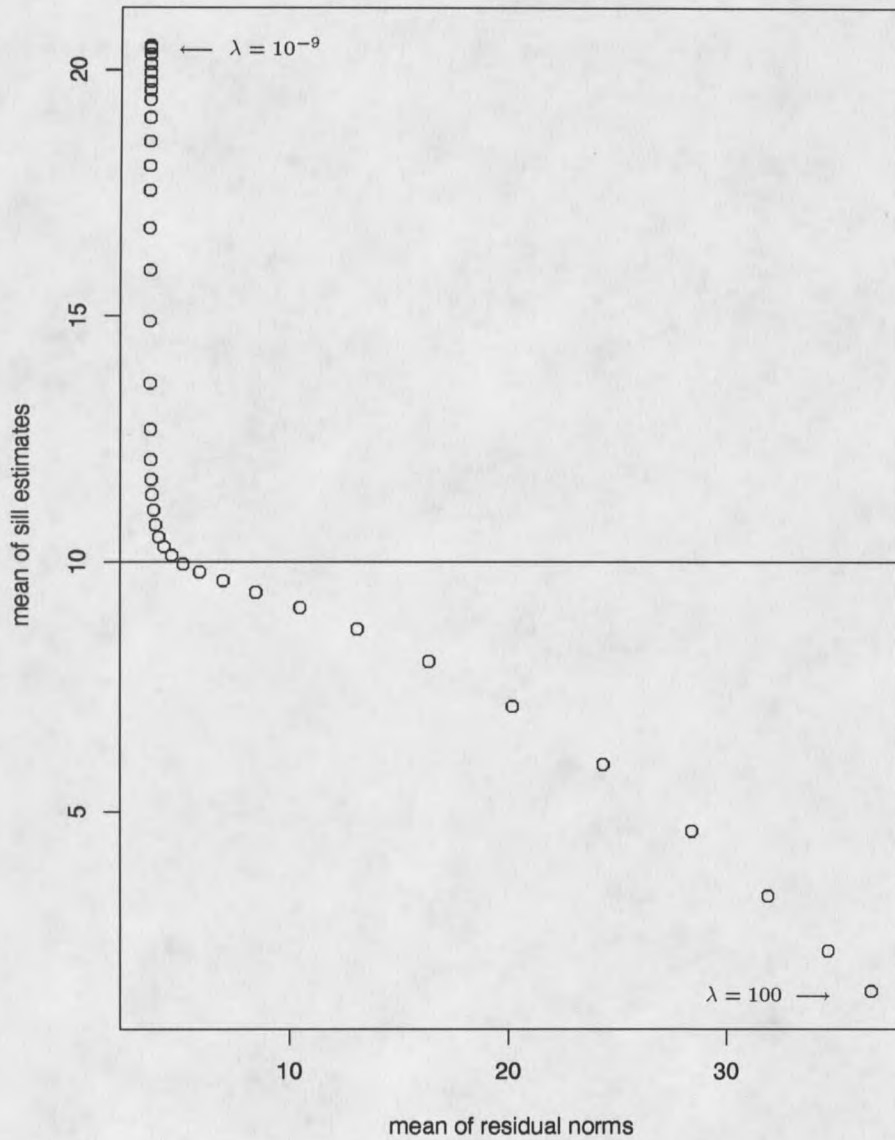


Figure 20: This plot shows the relationship between the average of the sill estimates and the average of the residual norms determined from fits to 1100 data sets. Each point in the plot corresponds to one of 45 values of  $\lambda$  from  $10^{-9}$  (upper left) to 100 (lower right). The true semivariogram model is exponential with sill and range of 10.

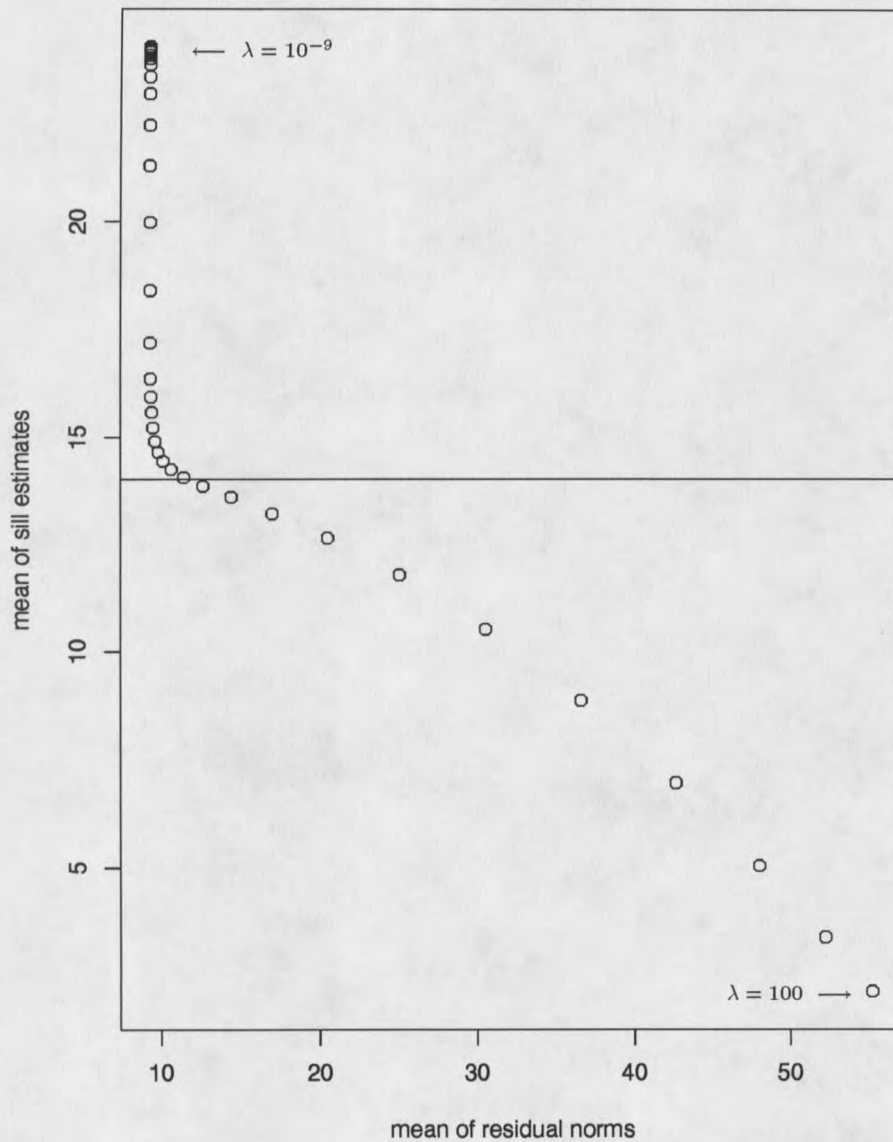


Figure 21: This plot shows the relationship between the average of the sill estimates and the average of the residual norms determined from fits to 1000 data sets. Each point in the plot corresponds to one of 45 values of  $\lambda$  from  $10^{-9}$  (upper left) to 100 (lower right). The true semivariogram model is a mixture of rational quadratic and hole effect models with sill of 14.

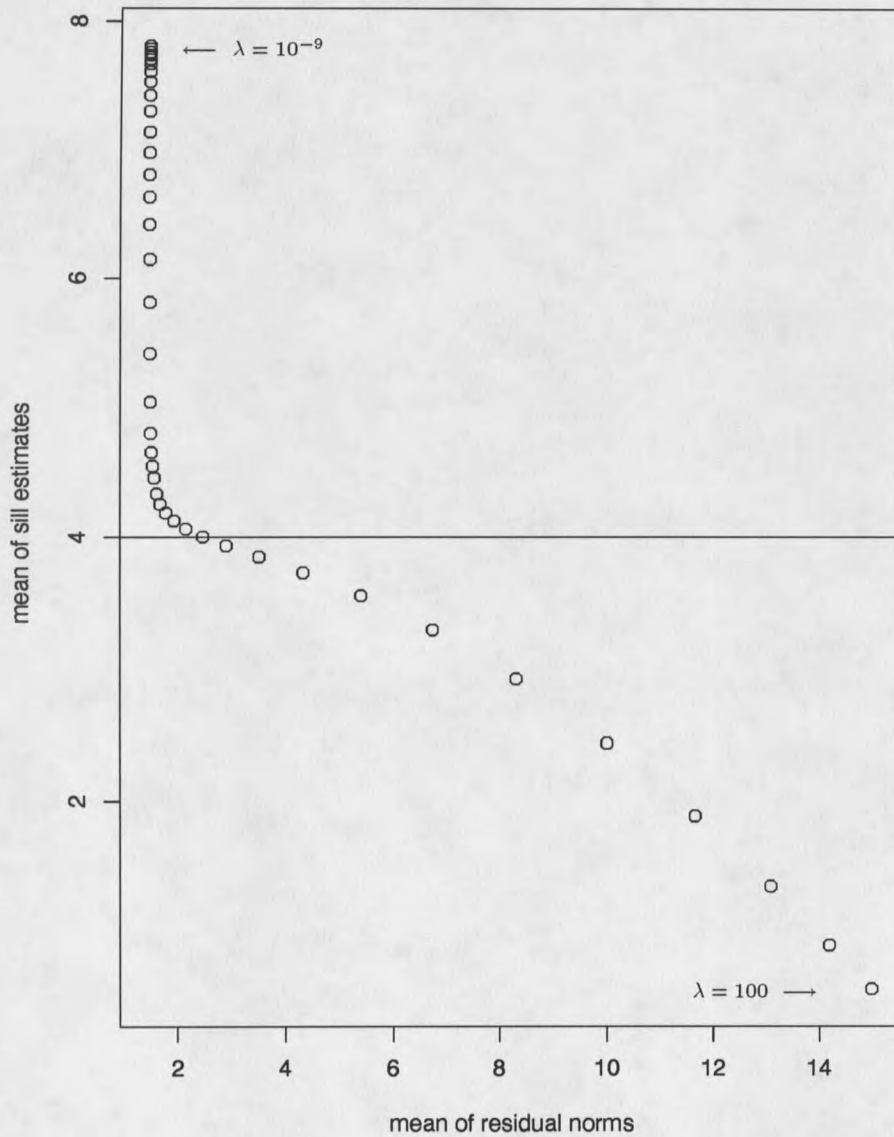


Figure 22: This plot shows the relationship between the average of the sill estimates and the average of the residual norms determined from fits to 1000 data sets. Each point in the plot corresponds to one of 45 values of  $\lambda$  from  $10^{-9}$  to 100. The true semivariogram model is a mixture of spherical models with sill of 4.

Table 9: Mean sill estimates and their variances from unpenalized and penalized fits at the estimated point of maximum curvature of plot of mean residual norm versus  $\log_{10}(\lambda)$  values.

Model	Exponential (Sill= 10)	
	Mean	Variance
Unpenalized	20.49	563.80
Penalized	9.96	17.52
Model	Rational Quadratic-Hole Effect (Sill= 14)	
	Mean	Variance
Unpenalized	24.01	578.09
Penalized	14.24	24.67
Model	Sphericals (Sill= 4)	
	Mean	Variance
Unpenalized	7.81	92.00
Penalized	4.06	3.82

plotted versus the 45  $\log_{10}(\lambda)$  values (Figures 23, 24, and 25). In each case, it appears that the value of  $\lambda$  that gives an unbiased estimate of the sill is close to the point where the curve of the residual norm has maximum curvature. This point was approximated numerically in each case and the results are shown in the Table 9. In each case the approximate point of maximum curvature of the residual norm curve occurs at  $\lambda$  values between 0.1 and 0.178.

If the sill estimate from a fit with  $\lambda = 0$  is reasonable, the penalized estimation procedure does not yield a penalized sill estimate that is unreasonable. This is illustrated in Figure 26. The top panel shows the sample semivariogram from one of the simulated random fields based on the exponential model with sill and range of 10. The solid line is a nonparametric fit of a semivariogram function with

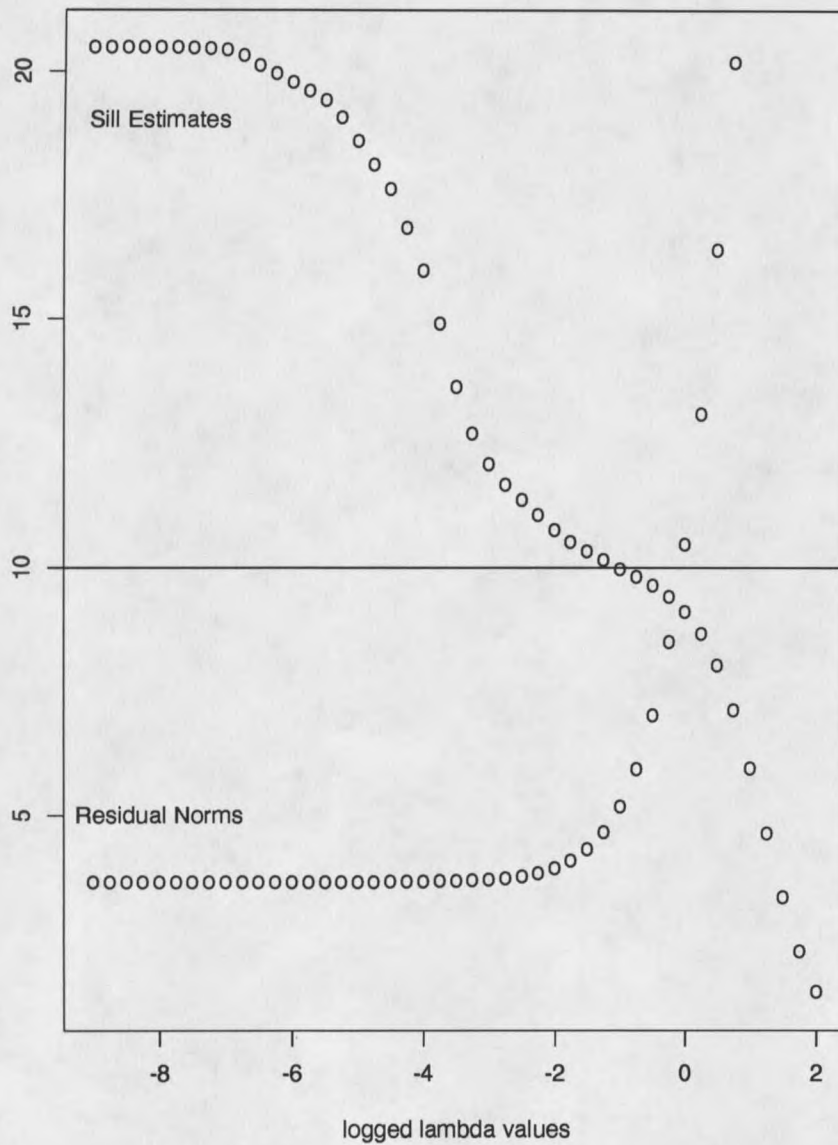


Figure 23: Plot of the means of the sill estimates and the means of the residual norms versus 45  $\log_{10}(\lambda)$  values. The means were determined from fits to 1100 data sets. The true semivariogram model is exponential with sill and range of 10.

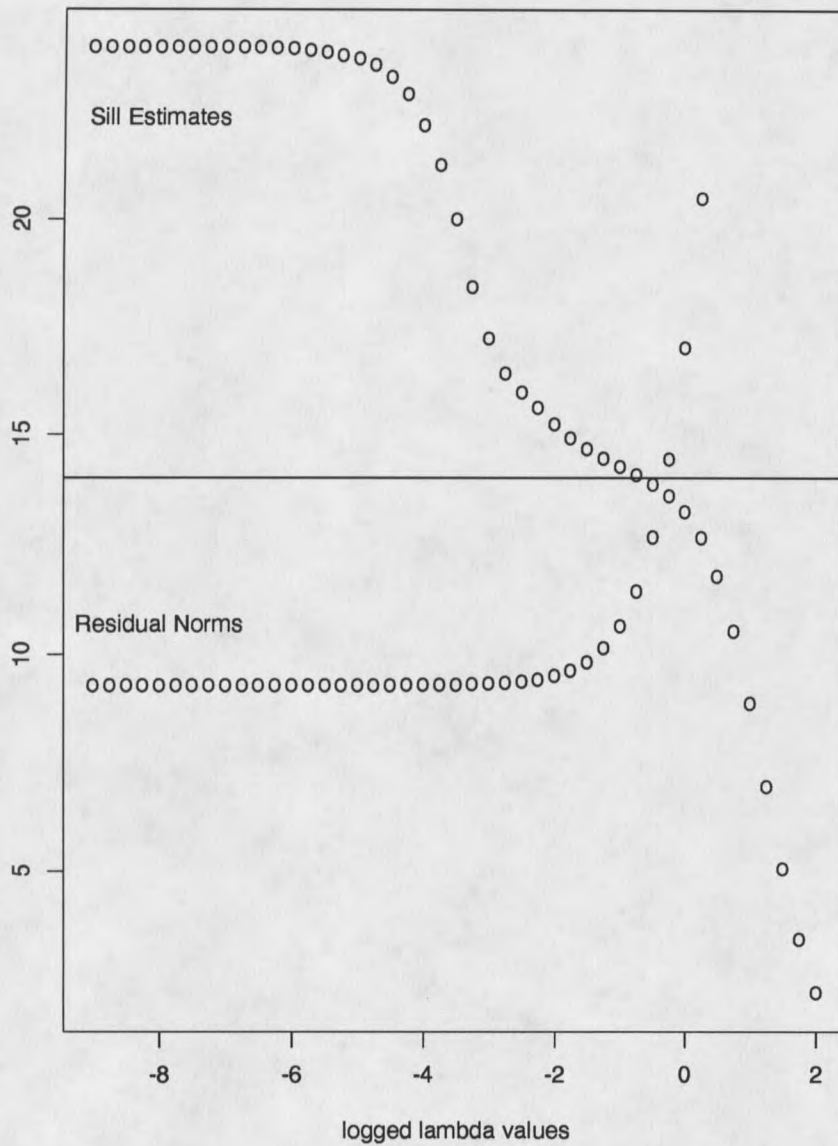


Figure 24: Plot of the means of the sill estimates and the means of the residual norms versus 45  $\log_{10}(\lambda)$  values. The means were determined from fits to 1000 data sets. The true semivariogram model is a mixture of a rational quadratic and hole effect models with sill of 14.

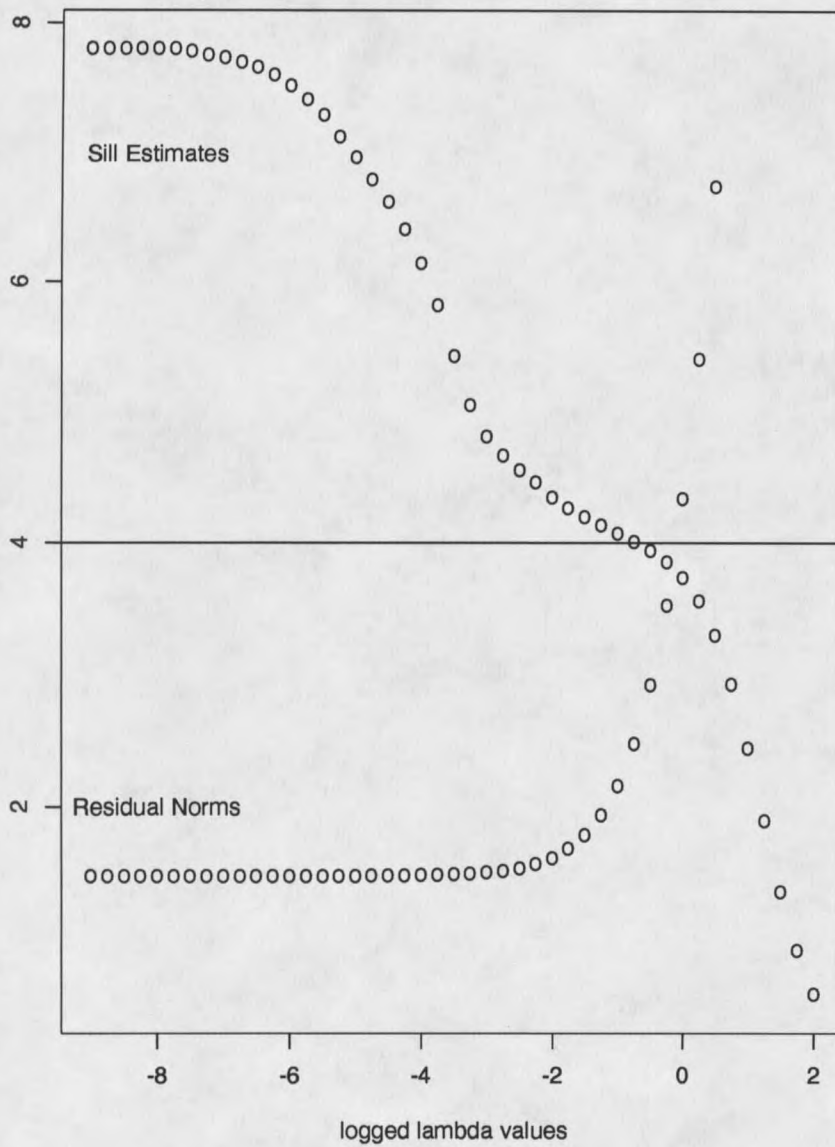


Figure 25: Plot of the means of the sill estimates and the means of the residual norms versus 45  $\log_{10}(\lambda)$  values. The means were determined from fits to 1000 data sets. The true semivariogram model is a mixture of spherical models with sill of 4.

$\lambda = 0$ . The sill estimate is 7.42, which is not unreasonable. The dashed line is a nonparametric semivariogram based on a fit with  $\lambda = .316$ . The sill estimate is 7.29. The bottom panel shows the relationship between the residual norm and the logged  $\lambda$  values. Maximum curvature occurs at about the point where  $\lambda = .316$ .

Figures 27, 28, 29, and 30 show plots similar to Figures 20 through 25 for 1000 simulations based on true exponential semivariogram models with sills of 10 and ranges of 2 and 18, respectively. The average sill estimate corresponding to the approximate point of maximum curvature on the residual norm versus  $\log_{10}(\lambda)$  values is 10.35 and 9.21 respectively, with variances of 6.20 and 23.98. The corresponding means for the unpenalized fits are 13.49 and 22.98 with variances of 69.62 and 872.71, respectively.

Thus, the penalized fitting procedure results in less biased and less variable sill estimates with little effect on fit. This general result holds over a wide variety of range, sill, and model combinations. The worst results from choosing  $\lambda$  on the basis of the approximate point of maximum curvature of the residual norm versus  $\log_{10}(\lambda)$  was in the case of the exponential model with sill of 10 and range of 18. But even here there is no questioning the fact that the sill estimates are not as badly biased and are much less variable than the unpenalized fits. Furthermore, in each case, the penalized nonparametric sill estimates compare favorably with the parametric sill estimates.

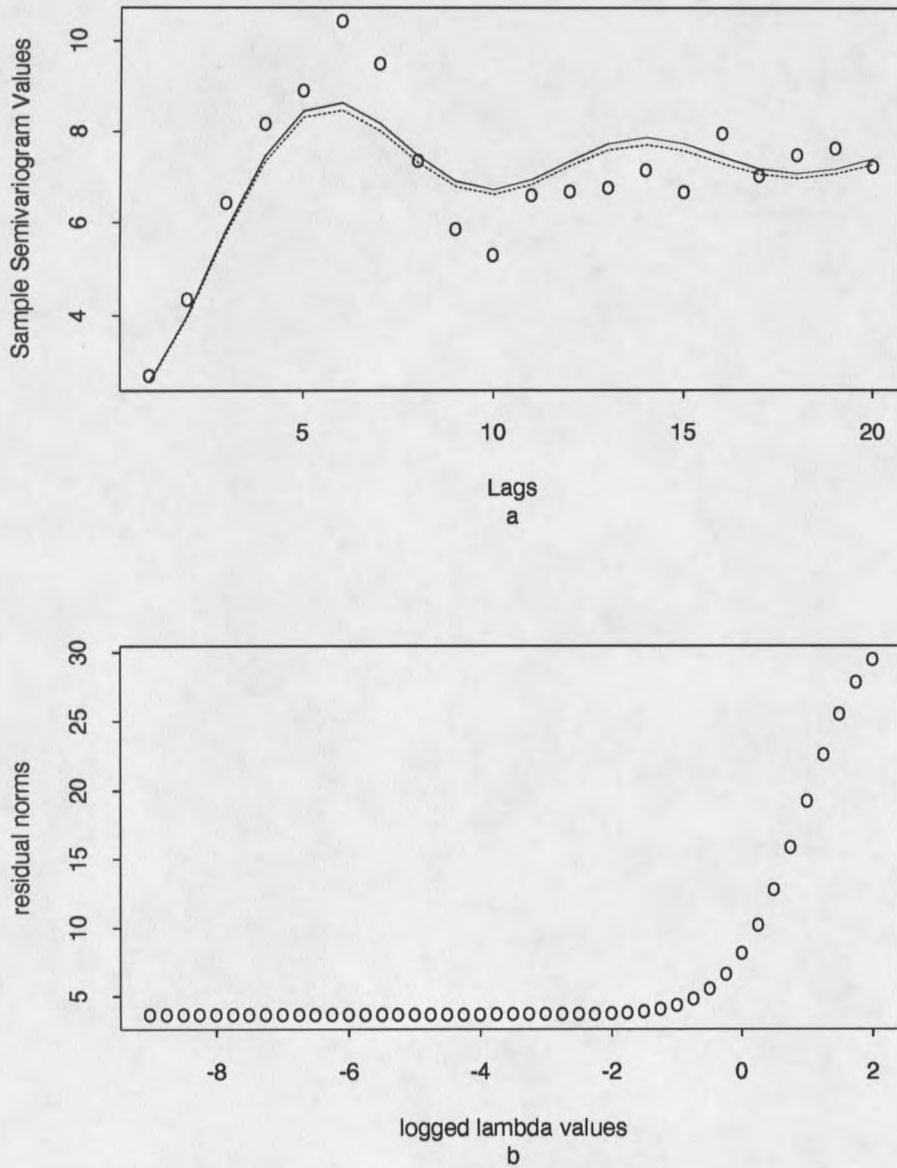


Figure 26: (a)-Sample semivariogram and two nonparametric fits. Solid line is unpenalized fit with sill estimate of 7.42. Dashed line is penalized fit with sill estimate of 7.29 (b)-Plot of residual norms versus  $\log_{10}(\lambda)$ .

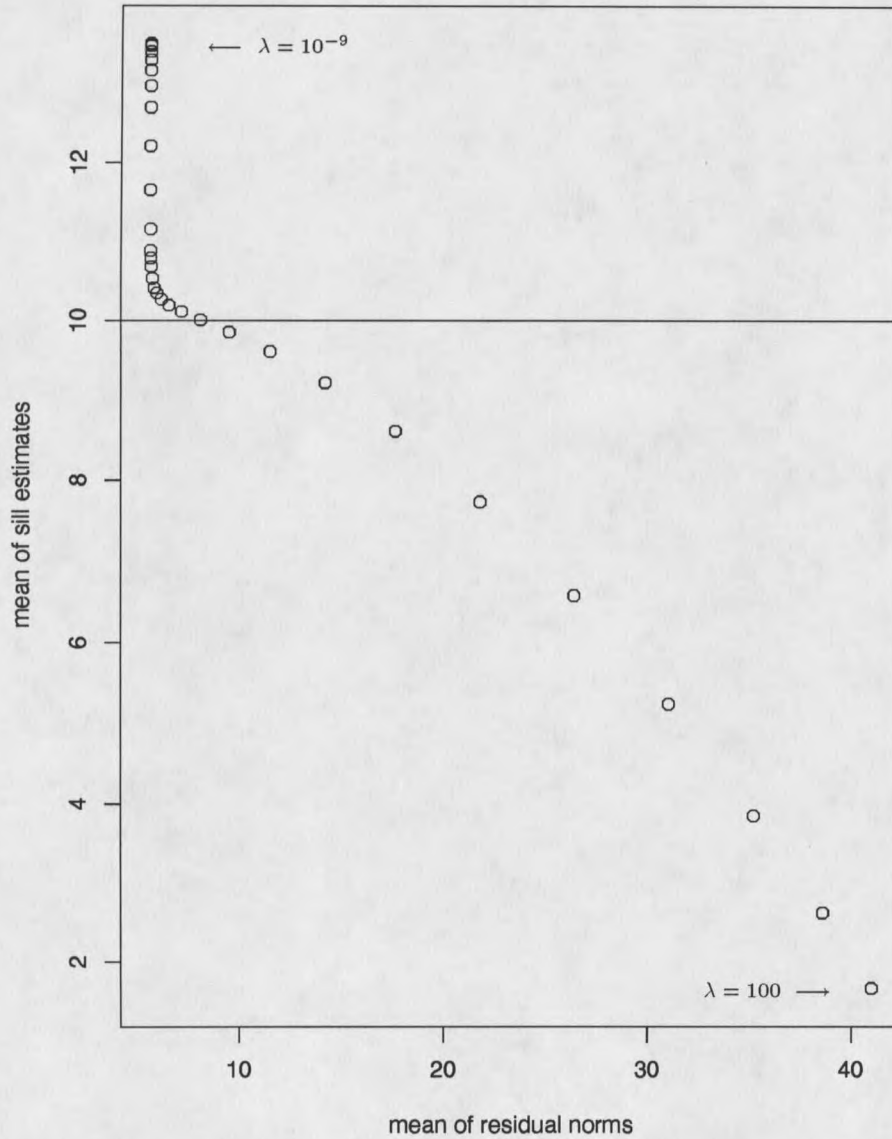


Figure 27: This plot shows the relationship between the average of the sill estimates and the average of the residual norms determined from fits to 1000 simulated data sets. Each point in the plot corresponds to one of 45 values of  $\lambda$  from  $10^{-9}$  (upper left) to 100 (lower right). The true semivariogram model is exponential with sill of 10 and range of 2.

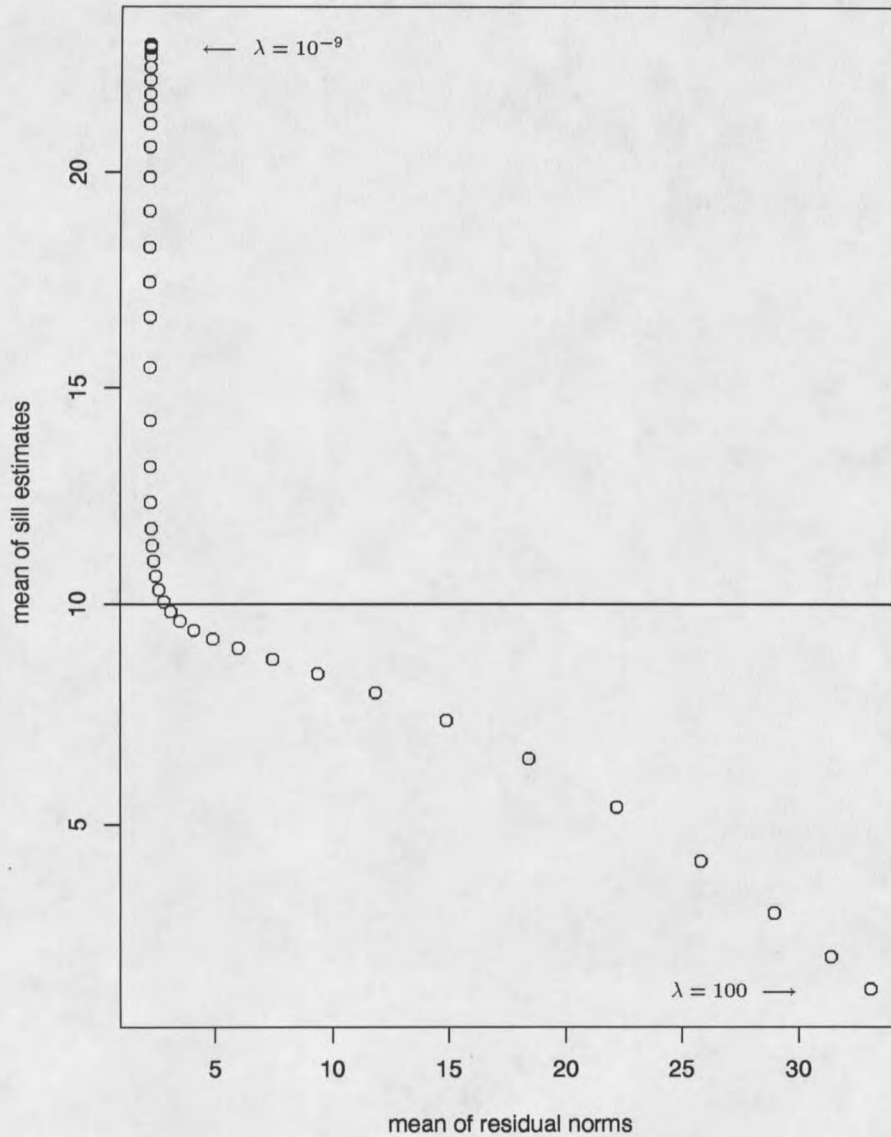


Figure 28: This plot shows the relationship between the average of the sill estimates and the average of the residual norms determined from fits to 1000 simulated data sets. Each point in the plot corresponds to one of 45 values of  $\lambda$  from  $10^{-9}$  (upper left) to 100 (lower right). The true semivariogram model is exponential with sill of 10 and range of 18.

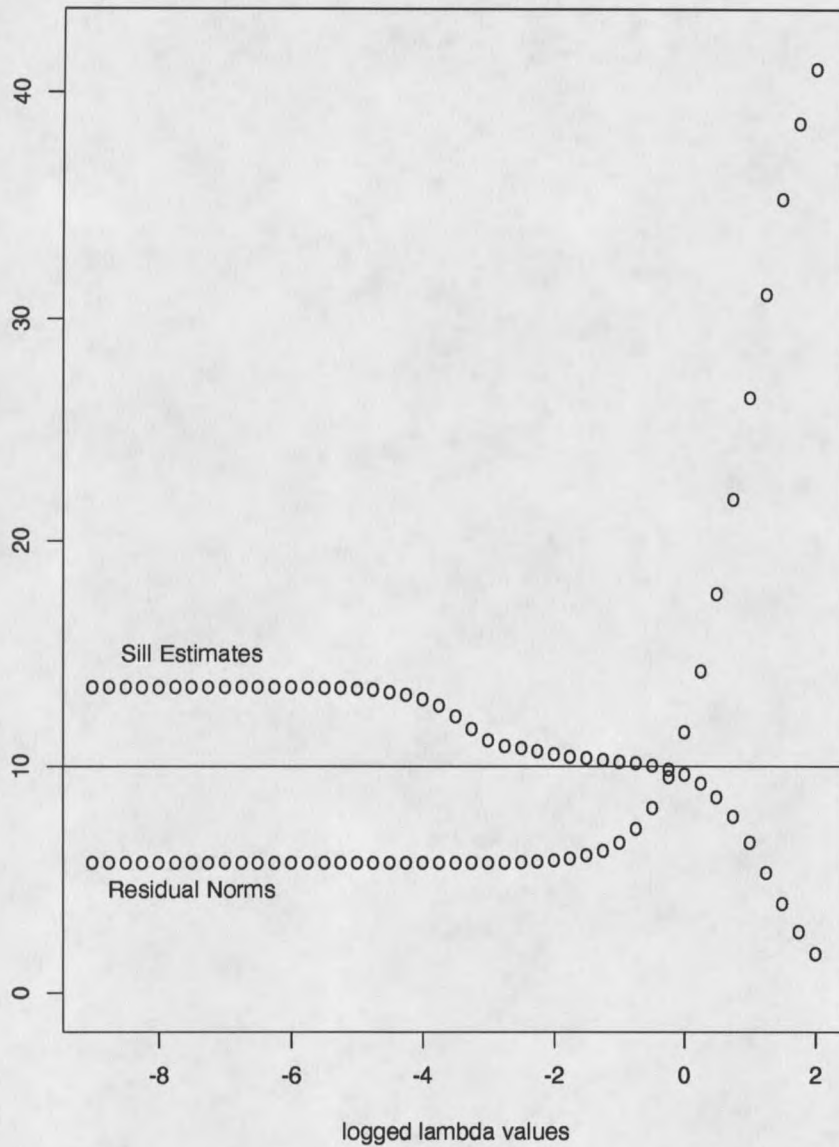


Figure 29: Plot of the means of the sill estimates and means of residual norms versus 45  $\log_{10}(\lambda)$  values. The means were determined from fits to 1000 data sets. The true semivariogram model is exponential with sill of 10 and range of 2.

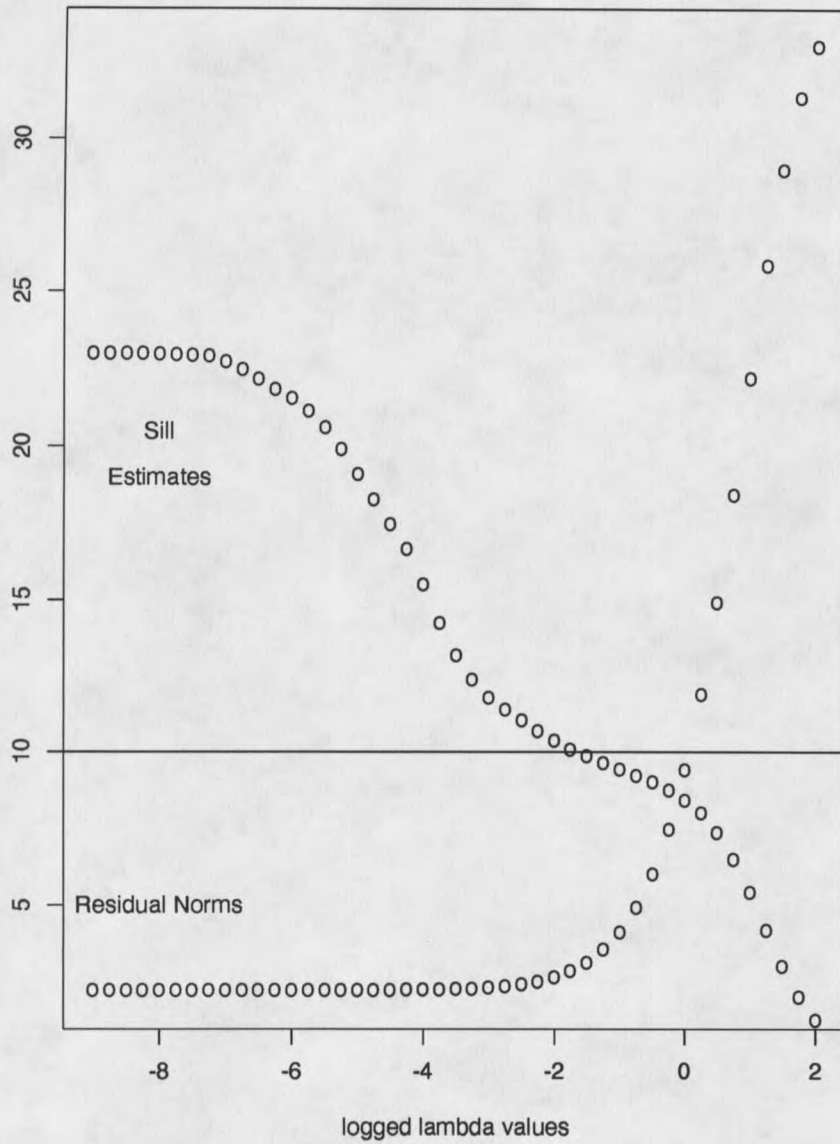


Figure 30: Plot of the means of the sill estimates and means of residual norms versus 45  $\log_{10}(\lambda)$  values. The means were determined from fits to 1000 data sets. The true semivariogram model is exponential with sill of 10 and range of 18.

## CHAPTER 5

### ASYMPTOTICS

The idea of saturating the node space is an appealing one. Even if the resulting problem has an ill-conditioned operator matrix and more parameters to estimate than data, the *NNLS* algorithm will tend to reduce the problem to a well-conditioned lower dimensional problem. The results of the simulations show that nonparametric semivariogram estimation yields good estimates of the true function and of an important intrinsic parameter of the second order stationary spatial process without the problem of having to choose a particular parametric form. Further, the nonparametric procedure using *NNLS* always converges (see Lawson and Hanson 1974 for a discussion of the convergence properties of *NNLS*), is quite fast, and is fairly robust to the selection of nodes.

However, it is generally accepted that a *good* estimator must possess some optimal large sample properties. In particular, as the amount of information becomes infinite the estimate should approach the truth. There are some who claim that this is not of much importance in geostatistics, arguing that dealing with the statistical properties of semivariogram estimators is an exercise in large sample theory of little

practical importance in linear geostatistics (Journel and Huijbregts, 1978, Journel, 1985).

It appears that consistency of the SB estimator  $\tilde{\gamma}(h)$  depends on the consistency properties of the sample semivariogram  $\hat{\gamma}(h)$ . If  $\hat{\gamma}(h)$  is a consistent estimator of the true semivariogram function  $\gamma(h)$  then the SB estimator seems to preserve those consistency properties.

To see this, consider the following argument. Let

$$\tilde{\gamma}_n(h) = \sum_{j=1}^n (1 - \Omega_d(ht_j)) \tilde{p}_j$$

be an SB estimate of a true semivariogram function  $\gamma_M(h)$ , based on minimizing

$$\sum_{i=1}^n \left( \sum_{j=1}^n (1 - \Omega_d(h_i t_j)) p_j - \hat{\gamma}_n(h_i) \right)^2 \quad (13)$$

subject to  $p_j \geq 0$  for  $j = 1, 2, \dots, n$ . Let  $\hat{\gamma}_n(h)$  be a consistent estimator of  $\gamma_M(h)$ .

Note that  $(1 - \Omega_d(ht))$  is uniformly continuous in both  $h$  and  $t$  for all  $d = 1, 2, \dots$ . This is perhaps most easily seen by noting that  $1 - \cos(x)$ ,  $1 - J_0(x)$ , and  $1 - \sin(x)/(x)$  are all uniformly continuous. Also  $1 - \exp(-x^2)$  is uniformly continuous and these represent the dimensional extremes of  $(1 - \Omega_d(ht))$ . Also, recall that  $M(t)$  is a nonnegative, nondecreasing, bounded function.

Let  $\gamma_{M_n}(h)$  be a semivariogram function where  $M_n(t)$  is a step function with  $n$  jumps. Then,  $\gamma_{M_n}(h)$  is a linear combination of uniformly continuous functions that have the same form as the kernel in  $\gamma_M(h)$ , and there will exist an  $n$  and a

collection of nodes  $t_j, j = 1, \dots, n$ , such that  $\gamma_{M_n}(h)$  can be made arbitrarily close to  $\gamma_M(h)$ . That is, given  $\epsilon > 0$  there exists a semivariogram function  $\gamma_{M_n}(h)$  with  $n$  step mixing function  $M_n(t)$  such that

$$\| \gamma_{M_n}(h) - \gamma_M(h) \| \leq \frac{\epsilon}{2}.$$

Further, the hypothesized consistency properties of  $\hat{\gamma}_n(h)$  imply the existence of an  $n$  large enough such that

$$\| \hat{\gamma}_n(h) - \gamma_M(h) \| \leq \frac{\epsilon}{2}.$$

So, if  $\tilde{\gamma}_n(h)$  is a solution based on the collection of nodes in  $\gamma_{M_n}$  then it is easy to see that,

$$\begin{aligned} \| \tilde{\gamma}_n(h) - \hat{\gamma}_n(h) \| &\leq \| \gamma_{M_n}(h) - \hat{\gamma}_n(h) \| \\ &\leq \| \gamma_{M_n}(h) - \gamma_M(h) \| + \| \gamma_M(h) - \hat{\gamma}_n(h) \| \\ &\leq \epsilon \end{aligned}$$

Some further comments are in order. First, note that any consistency results will only hold when estimating a semivariogram function valid in  $d$  or fewer dimensions with a nonparametric estimator valid in  $d$  or fewer dimensions. For example, if one is trying to estimate a semivariogram function valid in two dimensions with the three dimensional version of the nonparametric estimator, one is not necessarily going to achieve consistency, because the class of valid two dimensional semivariograms is larger than the class of three dimensional semivariograms.

Second, this argument says nothing about how the nodes should be allowed to go to infinity. It simply says that as the number chosen gets arbitrarily large, the nonparametric estimator will do a better and better job of fitting data sampled without error from  $\gamma_M$ . However, the rate of convergence of  $\tilde{\gamma}$  to  $\gamma_M$  may depend more on the convergence of  $\hat{\gamma}(h)$  to  $\gamma_M(h)$  than to the ability of  $\tilde{\gamma}_n(h)$  to fit the data. Figure 31 shows the results of fitting the nonparametric estimator to three sample semivariograms where the data are actually values from a known semivariogram model (*i.e.* the sampling was done without error). These data are the values of the true model at lags of 0.5 to 20 in units of 0.5. Thus, there are 40 values for each model. The first model is the exponential model with sill and range of 10. The second model is the rational quadratic-hole effect model with sill of 14, and the third is the mixture of sphericals with sill of 4. A collection of 200 nodes was used in the fitting. The first 100 were equispaced in the interval  $[.04, 4]$  and the second 100 were equispaced in the interval  $[4.16, 20]$ . The sill estimates from the resulting fits were 9.98, 13.9, and 4.03, respectively. In each case, the residual norm was 0. It is important to note that the resulting fit is a valid semivariogram function that interpolates exactly the points from the true underlying model.

It is obvious that any consistency properties the sill estimator discussed in the previous chapter possesses depends on the consistency of the SB estimator for the true semivariogram. If the SB estimator is consistent then, the unconstrained

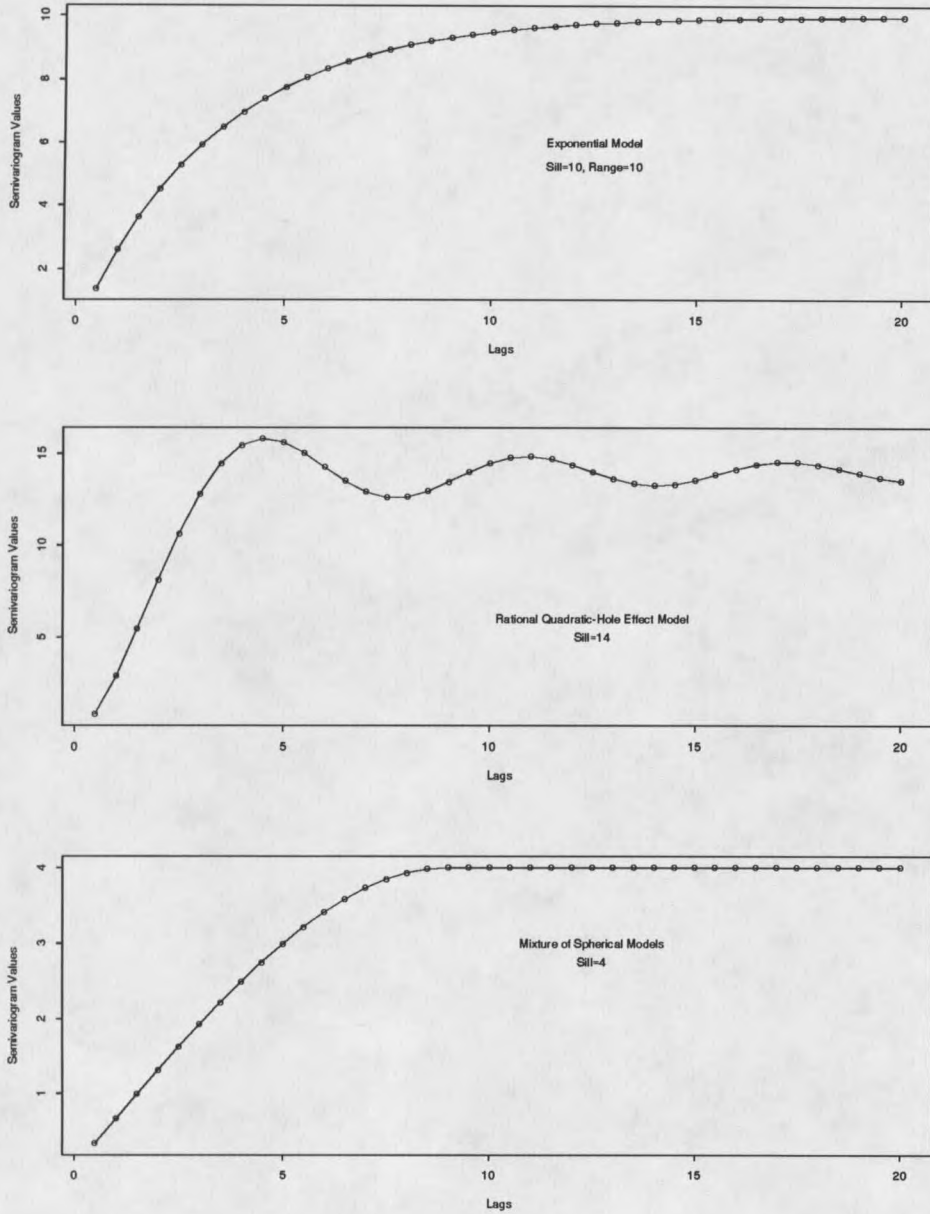


Figure 31: Results of fitting a nonparametric semivariogram to data from known valid semivariogram models.

estimator will be consistent. This follows from the following Lemma found in von Neumann and Schoenberg (1941).

**Lemma 1** *Let*

$$\gamma_n(h) = \int_0^\infty (1 - \Omega_d(ht)) dM_n(t); \quad d = 1, 2, \dots \quad (14)$$

*be a sequence of functions with  $M_n(t)$  a nonnegative nondecreasing function bounded above on  $t \geq 0$ . Let  $\gamma_n(h)$  converge as  $n \rightarrow \infty$ , uniformly in any finite interval, to a function  $\gamma(h)$ . Then  $\gamma(h)$  has the form*

$$\gamma(h) = \int_0^\infty (1 - \Omega_d(ht)) dM(t); \quad d = 1, 2, \dots \quad (15)$$

*with  $M(t)$  a nonnegative nondecreasing function bounded above on  $t \geq 0$ . Further,*

$$M_n(t) \rightarrow M(t) \quad (16)$$

*for all values of  $t$  which are continuity points of  $M(t)$  and*

$$\int_a^\infty dM_n(t) \rightarrow \int_a^\infty dM(t) \quad (17)$$

*in all continuity points  $t = a > 0$  of  $M(t)$ .*

*Conversely, if (16) and (17) hold and if  $\gamma_n(h)$  and  $\gamma(h)$  are defined by (14) and (15) then  $\gamma_n(h)$  converges as  $n \rightarrow \infty$ , uniformly in any finite interval, to  $\gamma(h)$ .*

The proof of this lemma is quite long and the interested reader is referred to von Neumann and Schoenberg (1941). Equation (17) is the relevant result for consistency of the sill estimator discussed in Chapter 4.

As noted above, the large sample properties of sample semivariogram estimators has not been a topic of much interest in geostatistics. Davis and Borgman (1979 and 1982) presented results on exact and asymptotic sampling distributions of the method of moments sample semivariogram. Cressie and Hawkins (1980), Cressie (1985), Hawkins and Cressie (1984), and Dowd (1984) also discuss some statistical (including large sample) properties of sample semivariograms. In general, consistency and asymptotic normality depend on fairly restrictive assumptions.

Much more attention has been paid to the problem of developing robust sample semivariograms. In fact, the work of Cressie and Hawkins (1980), Cressie (1985), and Hawkins and Cressie (1984) was motivated by this problem. Recall that the starting point for the estimation of semivariogram functions is the variogram or the semivariogram cloud. For the semivariogram cloud the points have the form,

$$Y_{ij} = \frac{[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]^2}{2}$$

and these tend to be skewed. It is important to have sample semivariograms that are robust to the presence of outliers.

Cressie (1985) presented some results on the statistical properties of the method of moments estimator ( $\hat{\gamma}(h)$ ). He derived an explicit expression for the variance of  $\hat{\gamma}(h)$  and showed that it is of order  $1/N(h)$  where  $N(h)$  is the number of observations of

$$\frac{[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]^2}{2}$$

at lag  $h$ . His derivation was based on several assumptions including normality of the observations  $Z(s)$ . Baczkowski and Mardia (1987, pp. 572 – 573), using a somewhat different approach established a similar result.

For the simulated data considered here, the method of moments estimator and the resulting SB nonparametric estimator exhibited some desirable large sample properties. This is illustrated in Figure 32. The results indicated in the figure were generated in the following manner.

A total of 1000 random fields was generated with 50 observations located one unit apart on a one dimensional transect. The true semivariogram function for the random process was an exponential model with sill and range of 10. A sample semivariogram (method of moments) was calculated for each of the simulated data sets over 20 lags. There were 20 points in each sample semivariogram calculated by smoothing a semivariogram cloud of 790 points. A nonparametric semivariogram function was fit to the sample semivariogram for each of the data sets. The 5th percentile, mean, and 95th percentile of the 1000 nonparametric estimators was determined for each of the 20 lags.

A second set of 1000 random field was generated with 200 observations located one-half unit apart on a one dimensional transect. Once again the true semivariogram function for the random process was an exponential model with sill and range of 10. Sample semivariograms were also calculated for each of the data sets

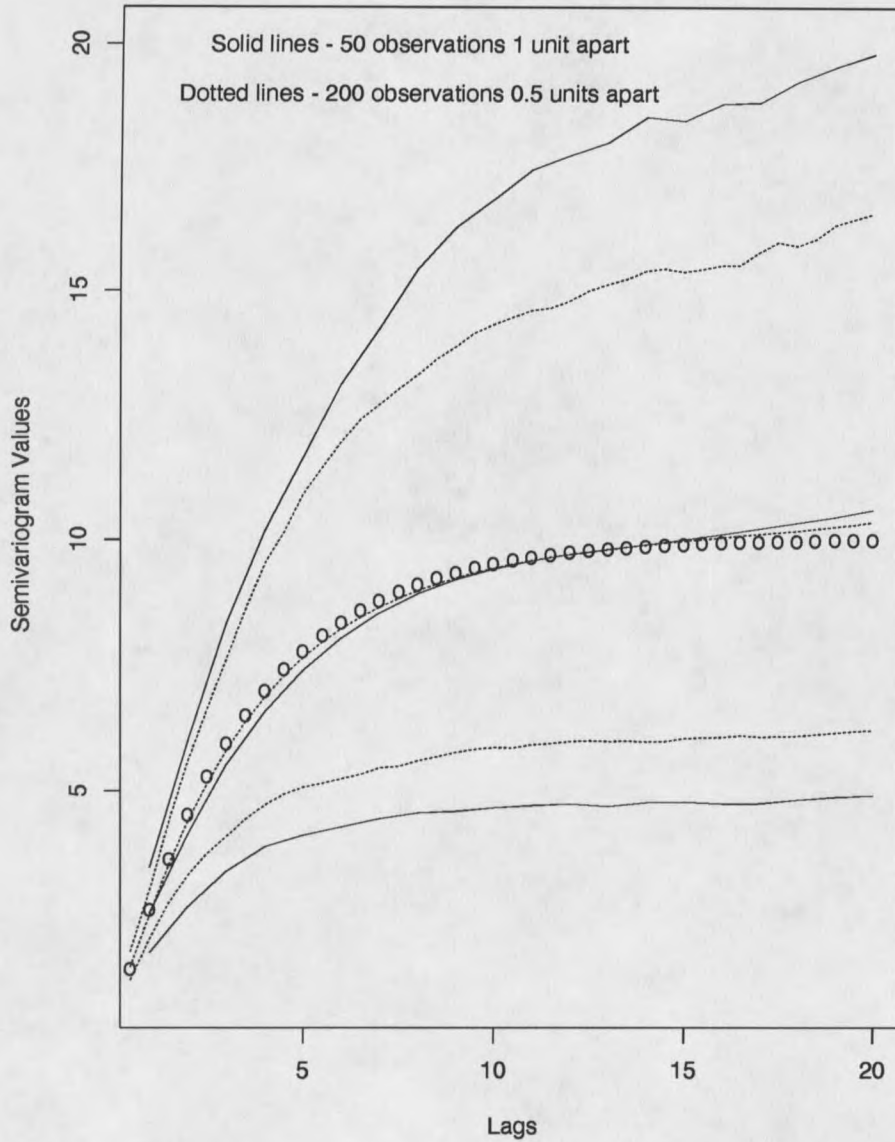


Figure 32: This figure shows the increasing accuracy of fitted nonparametric semi-variograms to a true exponential semivariogram (points) under increasing domain and infill asymptotics. The 2 upper lines are 95th percentiles, the 2 middle lines are means, and the 2 lower lines are 5th percentiles from 1000 fits. The true sill and range were 10.

for 40 one-half unit lags in the interval  $[0, 20]$ . Thus, these sample semivariograms had 40 points, based on smoothing a cloud of 7180 points over 40 lags equispaced in  $(0, 20]$ . Nonparametric semivariogram functions were again fit to the 1000 sample semivariograms, and the 5th percentile, mean, and 95th percentile of the 1000 nonparametric estimators was determined for each of the 40 lags.

The points in the figure lie along the true semivariogram function. The solid lines indicate the 5th (lower) and 95th (upper) percentiles and the mean (middle) of the 1000 functions fit to the first set of simulations. Thus, these are crude 90 percent confidence intervals for values of the fitted functions over the 20 lags. The dotted lines correspond to the mean and the 5th and 95th percentiles for the 1000 semivariogram functions fit to the second set of simulated data.

This situation thus corresponds to examining the large sample behavior of the nonparametric semivariogram estimator under both infill and increasing domain asymptotics. It is clear that in this setting more information results in a less variable and less biased estimator, providing empirical evidence of consistency.

It is surprising that more attention has not been paid to estimating semivariograms using smoothing techniques such as kernel regression or splines. There has been some work done on kernel regression with correlated errors. Hart and Wehrly (1986) used kernel smoothing methods on repeated measurements (growth curve) data. Altman (1990) also investigated smoothing with correlated errors. The con-

clusion seems to be that kernel regression performed well as long as the correlations died off over *short* distances. Diggle and Hutchinson (1988) have investigated spline smoothing with correlated errors.

The results of this chapter show that the SB nonparametric technique can be thought of as a filter that takes a sample semivariogram and converts it into an estimator that is itself a valid semivariogram. The SB estimator also seems to preserve the consistency properties (if any) of sample semivariogram estimators. It would seem to be worthwhile to further investigate the performance of kernel regression and splines on data with correlated errors.

## CHAPTER 6

### EXAMPLES

In this chapter, three examples of the use of the SB method will be presented. These are examples of nonparametric estimation of semivariogram functions and nonparametric estimation of the sill for three actual data sets taken from the literature. The nonparametric and parametric estimates are compared.

#### Example 1

The first data set comes from Clark (1979). The data is in the form of silver concentrations sampled from an ore body. Only data from the first 75 sample locations is used. This is the data set that SB used to illustrate their method. Clark (1979) fit a spherical semivariogram to this data by eye. Her estimated semivariogram took the form;

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ 11((3/2)(h/50) - (1/2)(h/50)^3) & 0 < h \leq 50 \\ 11, & h \geq 50 \end{cases}$$

The nonparametric fit was done using the same collection of nodes as was used above. The nonparametric semivariogram valid in 3 or fewer dimensions was used.

Figure 33 shows the actual data, Clark's fit (the solid line) and the nonparametric fit. Clark's sill estimate is 11. The sill estimate from the nonparametric fit is 11.3. The nonparametric model was fit with no constraints. There seems to be no reason to do a constrained fit in this case.

### Example 2

The second data set is also from Clark (1979). The data is logged nickel concentrations from an ore body. The experimental semivariogram is shown in Figure 34. Clark estimated a nugget effect of 0.40 and the data shown has been corrected to correspond to a nugget effect of 0. This has been done for convenience only and does not affect the comparison of the nonparametric fit with the parametric fit.

Clark fit a complicated mixture of spherical models to this data. Her model (absent the nugget effect) has the form,

$$\gamma(\mathbf{h}) = \begin{cases} 0 & \mathbf{h} = \mathbf{0} \\ \left[ \begin{array}{l} 1.15((3/2)(h/12) - (1/2)(h/12)^3) + \\ ((3/2)(h/60) - (1/2)(h/60)^3) \end{array} \right] & 0 < h \leq 12 \\ 1.15 + ((3/2)(h/60) - (1/2)(h/60)^3) & 12 < h \leq 60 \\ 2.15, & h \geq 60 \end{cases}$$

Figure 34 shows the actual data, Clark's fit (the solid line) and a penalized nonparametric fit (dashed line). Clark's sill estimate was 2.15. The sill estimate from an unpenalized ( $\lambda = 0$ ) fit was 3.74. This appears to be too high. The penalized fit-

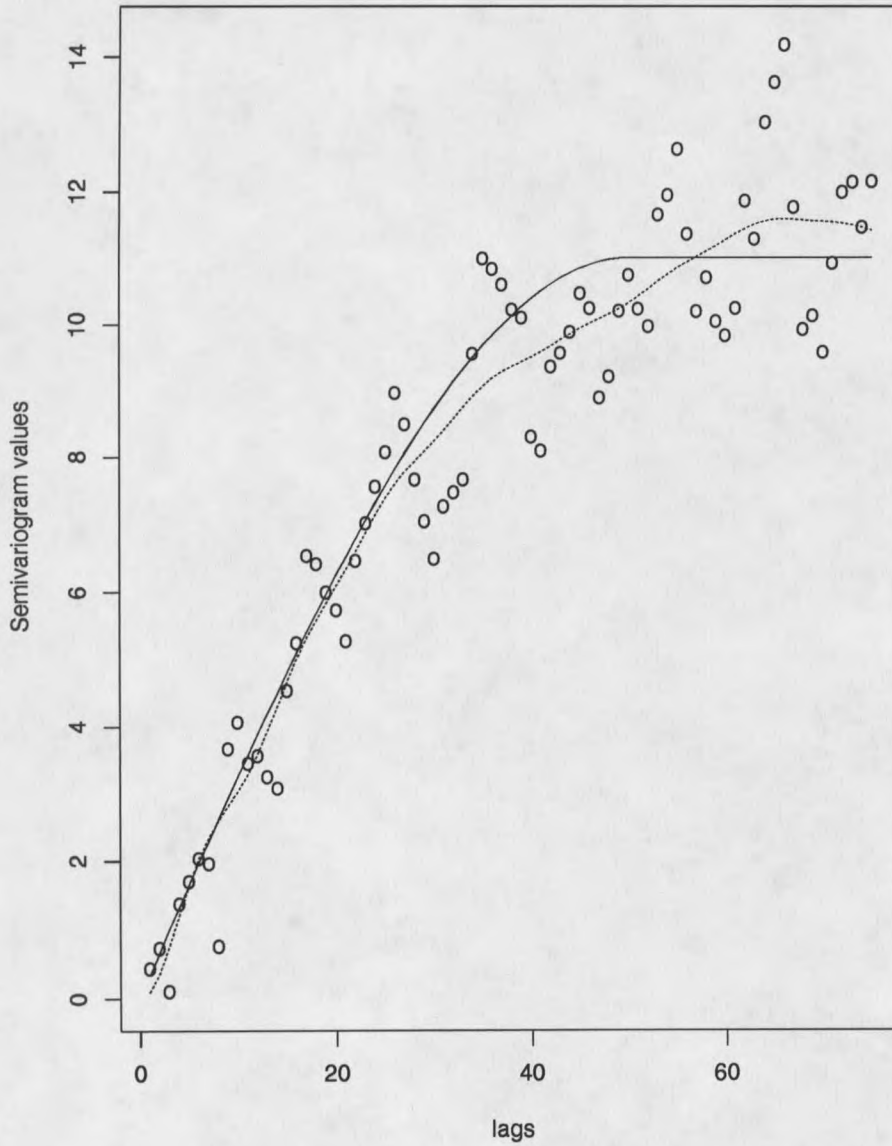


Figure 33: Sample semivariogram (points) and two semivariogram fits to Clark's (1979) silver data. Solid line is spherical model with sill = 11. Dashed line is nonparametric model with sill = 11.33.

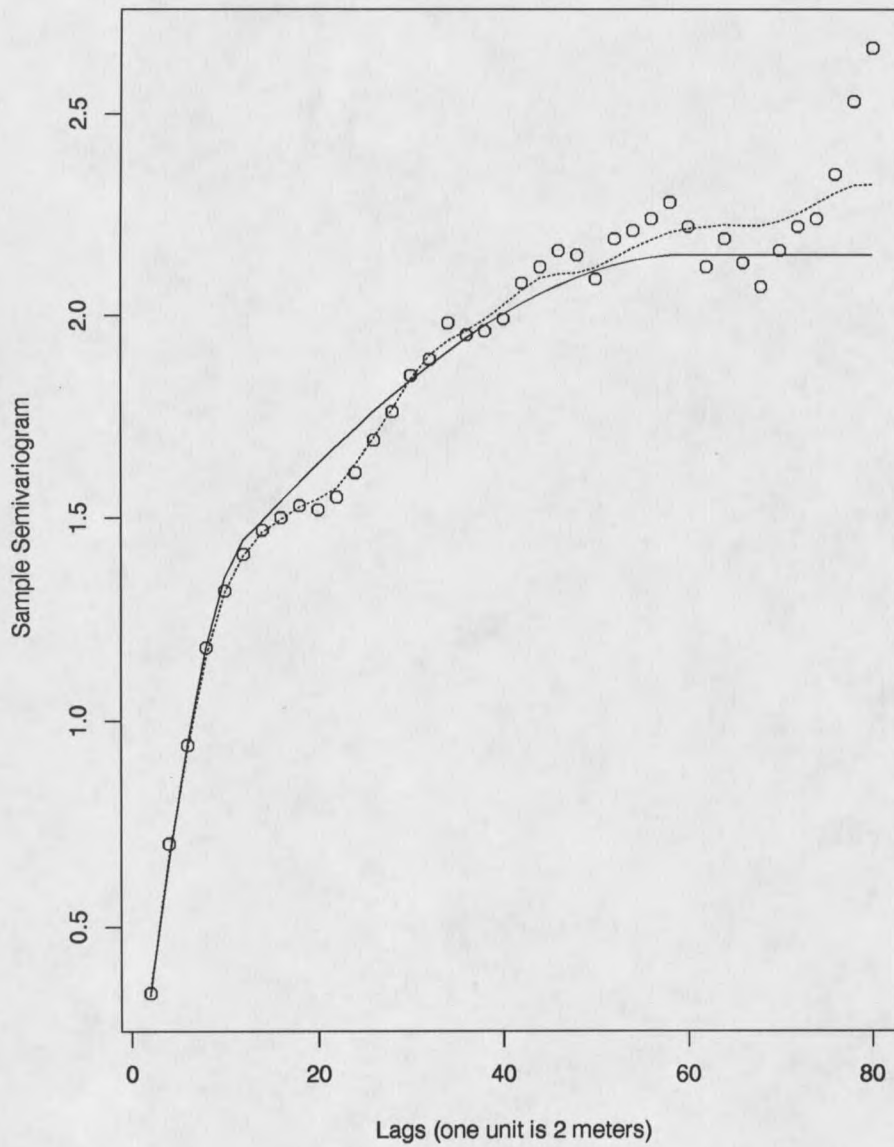


Figure 34: Sample semivariogram (points) and two semivariogram fits to Clark's (1979) nickel data. Solid line is spherical model with sill = 2.15. Dashed line is penalized nonparametric fit with sill = 2.15.

ting procedure described in Chapter 4 was used and the resulting sill estimate from the penalized nonparametric fit was also 2.15. Once again the parametric model was fit by eye. This rather complicated model should make it obvious why fitting by eye is so common. No single semivariogram model is going to fit well enough by itself, and trying to fit such a mixture of models by nonlinear least squares is difficult at best. The mixture of sphericals is not even differentiable in all its parameters (Cressie, 1985). This example also shows the strength of the nonparametric approach. It is much easier and faster than fitting a parametric model by nonlinear least squares, and, given the software, it is faster and easier than fitting by eye, even when the penalized fitting procedure is carried out.

### Example 3

The third example is from Cressie (1985). The data is actually the residuals from a median polish of the original data, which consisted of iron ore concentrations located in the east-west direction of an ore body. The data shown are not the classical method of moments estimators, but Cressie's robust estimator. The results are shown in Figure 35. Cressie estimated a nugget effect of 4.83 in the original and the data shown has been corrected for this. He fit a spherical model of the form,

$$\gamma(\mathbf{h}) = \begin{cases} 0 & \mathbf{h} = \mathbf{0} \\ 3.59((3/2)(h/50) - (1/2)(h/50)^3) & 0 < h \leq 8.73 \\ 3.59, & h \geq 8.73 \end{cases}$$

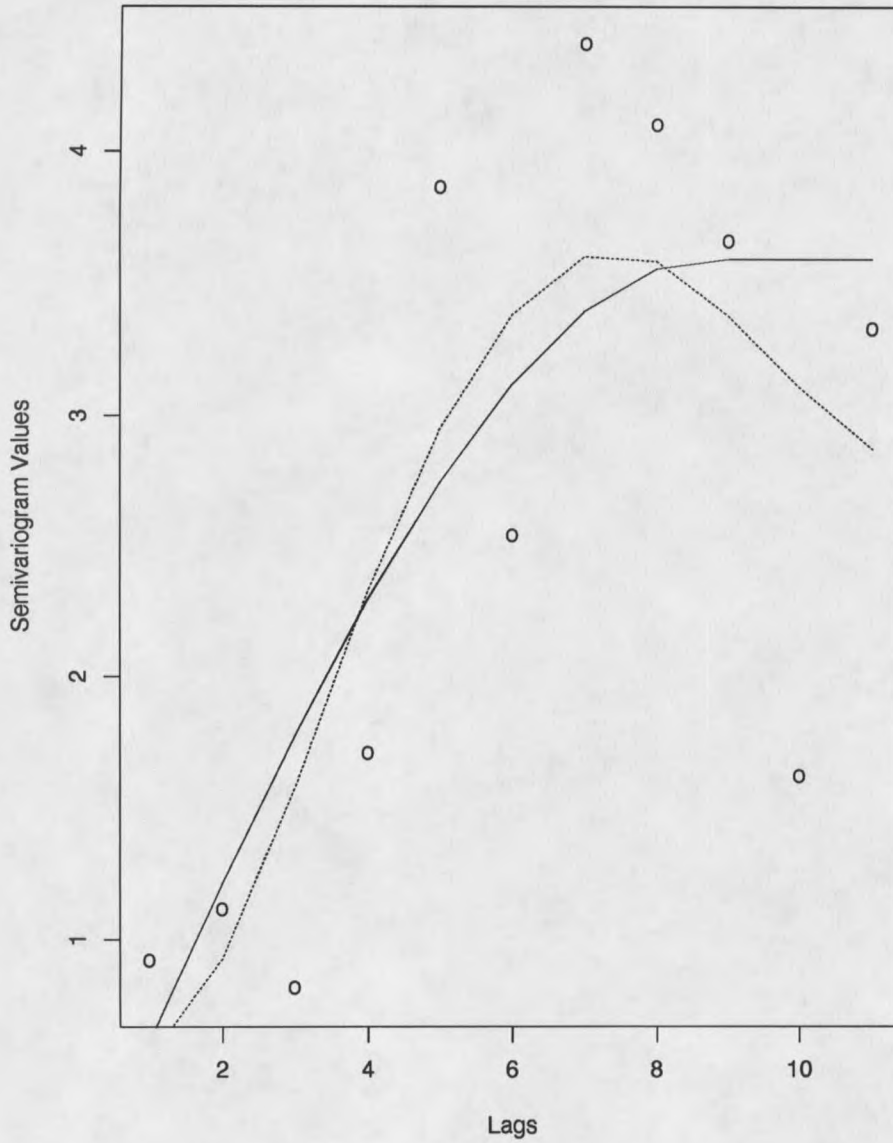


Figure 35: Sample semivariogram (points) and two semivariogram fits to Cressie's (1985) iron ore data. Solid line is spherical model with sill = 3.59. Dashed line is nonparametric model with sill = 3.02.

Cressie was able to fit this using nonlinear least squares. He discusses a procedure for circumventing the nondifferentiability problem for this semivariogram model. The solid line is the resulting fit. The dashed line is an unpenalized nonparametric fit. Cressie's sill estimate is 3.59 and the sill estimate from the nonparametric fit is 3.02.

## CHAPTER 7

### FUTURE RESEARCH NEEDS

This chapter will describe some future research needs in the nonparametric estimation of semivariogram functions and their parameters. The nonparametric estimation of the nugget effect and the range will be the topic of the first section. Subsequent sections will deal with the possible conditional negativeness of sample semivariograms, MINQU estimation of semivariograms, and  $L_1$  fitting.

#### Estimation of the Nugget Effect

In Chapters 3 and 4 all of the simulations were based on semivariogram models with no nugget effect. If a nugget effect is present then the SB method as formulated here would seem to have a fundamental problem. By necessity the SB estimator,  $\tilde{\gamma}(h)$ , must go through the origin. It is not possible to explicitly estimate a nugget effect using the estimator in this form.

One approach is to model the process as a nugget effect plus a semivariogram

that goes through the origin so that the sample semivariogram,  $\hat{\gamma}(h)$  is such that,

$$E[\hat{\gamma}(h)] = c_0 + \gamma_0(h). \quad (18)$$

The nugget effect  $c_0$  could be estimated first and then subtracted from the data yielding a new data set that is used for the determination of  $\gamma_0(h)$  using the SB approach. This is essentially what was done with the nickel data of Clark (1979) and the iron ore data of Cressie (1985).

The problem is to find a suitable estimator of the nugget effect. This will be difficult because the problem requires the estimation of a parameter that lies outside the range of the data.

This may not be much of a concern in many practical settings. Consider Figure 36. The points are from a sample semivariogram with a true underlying semivariogram model that is exponential with sill and range of 10 and no nugget effect. A value of 5 has been added to every one of the values in the original sample semivariogram, thus simulating a nugget effect of 5. The solid line is a nonparametric semivariogram fit to the data with the nugget effect included. This semivariogram goes through the origin, but it still gives a good fit to the data for lags greater than 1 and the sill estimate is 12.39. The sill estimate for the original data with no nugget effect was 7.42. The nugget effect is thus implicit in the nonparametric semivariogram for larger lags even though that particular semivariogram goes through the origin and has no nugget effect.

The dotted line was generated by taking the original data (with nugget effect of 0), fitting it with a semivariogram that properly goes through the origin, and then adding in the nugget effect (which was known in this case). This is a more appropriate way to proceed, but the fit is not much better for lags greater than 1.

If one is only interested in using this function to krig at lags greater than 1, then explicit estimation of the nugget would not be required because its effect is accounted for in the fit. A problem could arise if one wants to predict a value for the random field at lags outside the range of the data, but this is always true.

In looking at several other examples it seems that when nugget effects are present, the nonparametric estimator gives good estimates over the range of the data, even when the data is located near the origin. The only time there are problems is when data is *very* near the origin at which point one has a pretty good estimate of the nugget effect anyway. However, this topic needs further study, and there is a need for a good nonparametric estimator of the nugget effect.

Simply adding an intercept to the SB model will not work, because if there is no data at the origin the *NNLS* procedure will simply return an estimate of that parameter that is equal to 0.

There are several other possibilities worth exploring, however. One is to fit a linear model or nonlinear model to the first few lags. The fit could just be a straight line fit or something more complicated depending on the sample semivariogram.

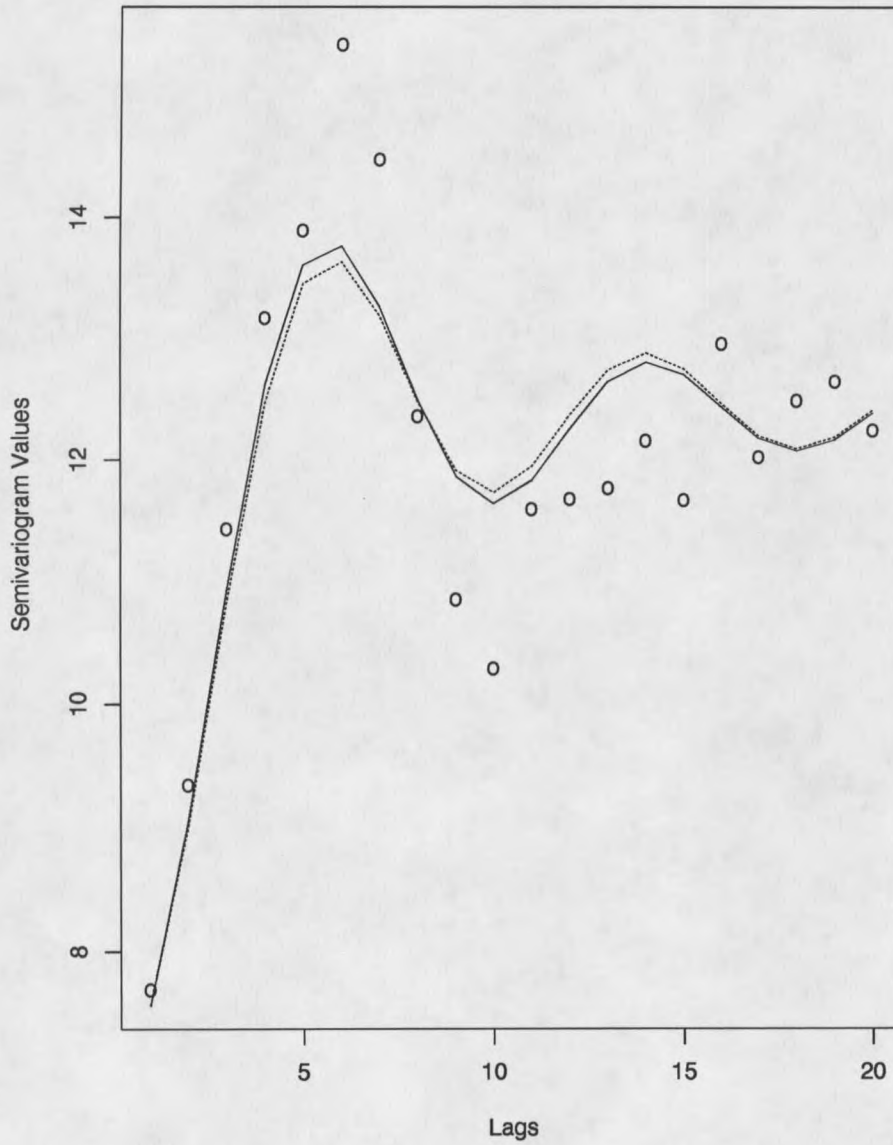


Figure 36: Sample semivariogram with a nugget effect of 5 and two nonparametric semivariogram estimators. The solid line is a fit to the data with the nugget included. The dotted line resulted from a fit to the data with the nugget removed. The nugget was then added back to the resulting fit to get the dashed line.

One nonlinear possibility is to fit a power model (using nonlinear least squares) to the first few lags.

Still another possibility is to fit (18) with no nonnegativity constraints. The nugget effect  $c_0$  can be incorporated as an explicit parameter with  $\gamma_0(h)$  having the form of the SB nonparametric estimator. This could be done with standard least squares software. The estimate of the nugget effect from this initial fit could then be subtracted from the sample semivariogram to yield a sample semivariogram with an estimated nugget effect of 0. The *NNLS* routine could then be used to fit a nonparametric estimator using this new sample semivariogram. Unfortunately, this has not produced good results in the few cases in which it has been tried.

Other possibilities include using various smoothers that are robust to boundary effects to estimate the nugget effect. These methods include locally weighted regression (Hastie and Loader, 1993) or kernel regression that has been modified to be robust to boundary effects (e.g. Hall and Wehrly, 1991).

The idea of estimating the nugget effect first and then using this estimate to produce a new sample semivariogram with estimated nugget effect of 0 has merit. This is analogous to a procedure common in time series analysis. Many time series software packages assume that the time series being analyzed has mean 0, and a prelude to analysis of data using these packages is to determine the sample mean of the time series in question and use this to generate a new time series with

approximate mean 0. Of course estimating the mean in these cases is considerably easier than estimating the nugget effect in the spatial setting.

### Estimation of the Range

The range is the lag at which the random variables of the spatial process become uncorrelated. This parameter is important in parametric fitting because the range is an explicit parameter in many parametric models (e.g. the spherical model) or is a function of some such parameter (e.g. the rational-quadratic model). The range is not an explicit parameter in the SB model. If the estimated semivariogram is to be used for kriging, then this is not a major concern because the semivariogram itself is the description of the correlation structure that is used in kriging.

If an estimate of the range is desired; however, then the SB model will not yield one. Nonparametric estimation of the range will undoubtedly be somewhat complicated. One approach is to examine the residuals from the fitted model and try to determine the lag at which the sample variance of the residuals stabilizes. Clearly, one needs to have more than the residuals of the fit to a sample semivariogram. What is needed is the residuals of the fit to the semivariogram cloud.

Sample Semivariogram as an Estimate

The method of moments estimator is not generally used as an estimate of the semivariogram in kriging because of the conditional negative definiteness property of semivariograms. This property is important in kriging. Without it such embarrassments as negative variances of the kriging estimator can result, and the predicted values can be poorly estimated. Cressie (1991) comments that the method of moments (and other smoothing techniques) are almost certain to fail this crucial test. But, there is little in the literature to substantiate this claim. It seems to be accepted with little overt evidence to back it up. The question of whether or not the method of moments estimator would yield a valid kriging matrix; that is, whether or not the appropriate submatrix ( $\Gamma$ , say) in the kriging matrix would be conditionally negative definite using the sample semivariogram was investigated. The submatrix,  $\Gamma$ , has elements

$$\{\gamma(\mathbf{h}_{ij})\} = \{\gamma(\mathbf{s}_i - \mathbf{s}_j)\}.$$

The following two propositions establish some results about the eigenstructure of conditional negative definite matrices that will be useful.

Let  $\mathbf{A}$  be an  $n \times n$  symmetric matrix with  $a_{ii} = 0$  and  $a_{ij} > 0$ . Recall that  $\mathbf{A}$  is conditionally negative definite (*cnd*) if  $\mathbf{y}'\mathbf{A}\mathbf{y} < 0$  for all vectors  $\mathbf{y}$  such that

$\mathbf{y}'\mathbf{1}_n = 0$ . Consider the maximum of

$$\frac{\mathbf{y}'\mathbf{A}\mathbf{y}}{\mathbf{y}'\mathbf{y}}$$

subject to the constraint  $\mathbf{y}'\mathbf{1}_n = 0$ .

Let  $\mathbf{J}_n$  be an  $n \times n$  matrix of ones. Let  $\mathbf{U}$  be the  $n \times (n - 1)$  matrix whose columns are the orthonormal eigenvectors corresponding to the  $(n - 1)$  eigenvalues of  $\mathbf{J}_n$  that are equal to 0. Then  $\mathbf{P} = \frac{1}{n}\mathbf{J}_n$  is the perpendicular projection operator onto the space spanned by the column vector  $\mathbf{1}_n$ . Let  $\mathbf{I}_n$  be the  $n \times n$  identity matrix. Then  $(\mathbf{I}_n - \mathbf{P})\mathbf{y} = \mathbf{y}$  if and only if  $\mathbf{y}'\mathbf{1}_n = 0$ . Further,  $(\mathbf{I}_n - \mathbf{P}) = \mathbf{U}\mathbf{U}'$ . Thus,  $(\mathbf{I}_n - \mathbf{P})\mathbf{y} = \mathbf{y}$  implies  $\mathbf{U}\mathbf{U}'\mathbf{y} = \mathbf{y}$ . Now, rewriting  $\mathbf{U}\mathbf{U}'\mathbf{y}$  as  $\mathbf{U}\mathbf{w}$ , it can be seen that the problem of finding the maximum of

$$\frac{\mathbf{y}'\mathbf{A}\mathbf{y}}{\mathbf{y}'\mathbf{y}}$$

subject to the constraint  $\mathbf{y}'\mathbf{1}_n = 0$  is the same as finding the maximum of

$$\frac{\mathbf{w}'\mathbf{U}'\mathbf{A}\mathbf{U}\mathbf{w}}{\mathbf{w}'\mathbf{w}} \tag{2}$$

with no constraints on  $\mathbf{w}$ . Note that  $\mathbf{U}'\mathbf{U} = \mathbf{I}_{n-1}$ . The maximum of the constrained problem is thus the largest eigenvalue of  $\mathbf{U}'\mathbf{A}\mathbf{U}$ . These results can be used to establish the following proposition.

**Proposition 2** *A is (cnd) if and only if  $\mathbf{U}'\mathbf{A}\mathbf{U}$  is negative definite.*

**Proof:** Suppose  $\mathbf{A}$  is *(cnd)*. Then

$$\frac{\mathbf{y}'\mathbf{A}\mathbf{y}}{\mathbf{y}'\mathbf{y}}$$

is less than 0 for all  $\mathbf{y}$  such that  $\mathbf{y}'\mathbf{1}_n = 0$ . But this implies that the largest eigenvalue of  $\mathbf{U}'\mathbf{A}\mathbf{U}$  is negative and thus  $\mathbf{U}'\mathbf{A}\mathbf{U}$  is negative definite.

Suppose  $\mathbf{U}'\mathbf{A}\mathbf{U}$  is negative definite. Then its largest eigenvalue is negative. But this is the maximum of

$$\frac{\mathbf{y}'\mathbf{A}\mathbf{y}}{\mathbf{y}'\mathbf{y}}$$

under the constraint that  $\mathbf{y}'\mathbf{1}_n = 0$ , and so  $\mathbf{A}$  must be *(cnd)*.

Another consequence of this result follows.

**Proposition 3** *If  $\mathbf{A}$  is (cnd) then  $\mathbf{A}$  has one positive eigenvalue and  $(n - 1)$  negative eigenvalues.*

**Proof:** Let

$$\lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{A}) \geq \dots \geq \lambda_n(\mathbf{A})$$

and

$$\lambda_1(\mathbf{U}'\mathbf{A}\mathbf{U}) \geq \lambda_2(\mathbf{U}'\mathbf{A}\mathbf{U}) \geq \dots \geq \lambda_{n-1}(\mathbf{U}'\mathbf{A}\mathbf{U})$$

be the ordered eigenvalues of  $\mathbf{A}$  and  $\mathbf{U}'\mathbf{A}\mathbf{U}$  respectively. Then, by Corollary 1 on page 293 of Lancaster and Tismenetsky, (1985)

$$\lambda_1(\mathbf{A}) \geq \lambda_1(\mathbf{U}'\mathbf{A}\mathbf{U}) \geq \lambda_2(\mathbf{A}) \geq \dots \geq \lambda_{n-1}(\mathbf{U}'\mathbf{A}\mathbf{U}) \geq \lambda_n(\mathbf{A}).$$

But,  $\mathbf{A}$  (*cond*) implies that  $\lambda_1(\mathbf{U}'\mathbf{A}\mathbf{U}) < 0$  and so  $\lambda_i(\mathbf{A}) < 0$  for  $i = 2, \dots, n$ . Further,  $Tr(\mathbf{A}) = 0$  implies that  $\sum_{i=1}^n \lambda_i(\mathbf{A}) = 0$ . Thus,  $\lambda_1(\mathbf{A}) = -\sum_{i=2}^n \lambda_i(\mathbf{A}) > 0$ .

The converse of Proposition 3 is not true. Proposition 2 can be used to generate a simple test for conditional negativeness of a matrix.

The results from determining method of moments estimators from 700 simulated data sets from the five exponential models, the rational quadratic-hole effect model, and the mixture of spherical models described in Chapter 3 were used. There were 100 simulated data sets for each of the 7 models. These method of moments estimators were used to generate the submatrices that would be used in the kriging matrix, and those submatrices were tested for conditional negative definiteness. The results are shown below.

- exponential model (sill=10, range=2): 92 conditionally negative definite.
- exponential model (sill=10, range=6): 90 conditionally negative definite.
- exponential model (sill=10, range=10): 83 conditionally negative definite.
- exponential model (sill=10, range=14): 72 conditionally negative definite.
- exponential model (sill=10, range=18): 84 conditionally negative definite.
- rational quadratic-hole effect model (sill=14): 30 conditionally negative definite

- mixture of sphericals model (sill=4): 71 conditionally negative definite.

There is a surprisingly large percentage of valid matrices. The conditional negativity of kernel smooths of semivariogram clouds was also investigated. For the simulated data used here, these fits essentially yielded the method of moments estimators, with the same results as far as conditionally negative definite estimators was concerned. But, it was always possible to get valid matrices for those few data sets for which the original method of moments smooths failed to yield a valid matrix by resmoothing using different bandwidths. If there is some relationship between bandwidth selection and conditional negative definiteness then this opens up the possibility of using kernel regression or splines to estimate semivariogram functions and get valid estimators. This topic is clearly worthy of further study.

A recent paper by Rousseeuw and Molenberghs (1993) discusses several methods of transforming covariance and correlation matrices to achieve the necessary positive definite condition. These methods could be applied to semivariogram matrices also. In theory then, one could estimate the semivariogram (or covariance) function using any reasonable method and then apply these methods or modifications of them to achieve valid kriging matrices.

#### MINQU Estimation

Minimum norm quadratic (MINQU) estimation was developed by Rao (1970,

1971a, 1971b, 1979) for estimation of variance-components. Kitinidis (1985, 1987), Marshall and Mardia (1985), Stein (1988) and others have discussed its use in spatial statistics. It can be used in the spatial setting if the variogram is linear in its parameters,

$$2\gamma(\mathbf{h}; \theta) = 2 \sum_{i=1}^k \theta_i \gamma(\mathbf{h}).$$

This implies that the covariance function is also linear in its parameters. Assume that the data  $\mathbf{Z}$  are distributed with a mean of  $\mathbf{X}\beta$  and covariance matrix  $\Sigma(\theta)$ .

Further, assume that

$$\Sigma(\theta) = \sum_{i=1}^k \theta_i \Sigma_i.$$

The MINQU estimator is the  $\hat{\theta}_i$  that minimizes  $E(\hat{\theta}_i - \theta_i)^2$ , and that can be written as  $\hat{\theta}_i = \mathbf{W}' H_j \mathbf{W}$ , where  $\mathbf{W} = A' \mathbf{Z}$  is an  $(n - q) \times 1$  vector of linearly independent error contrasts. Then the  $n - q$  columns of  $A$  are linearly independent and  $A' \mathbf{X} = \mathbf{0}$ . It is common to place restrictions on  $H_j$  such that the MINQU estimators are both unbiased and invariant to changes in  $\mathbf{X}$ , often referred to as MINQU(U,I) estimators.

Cressie (1991) does not think MINQU estimation is appropriate, in general, because of the restricted nature of allowable semivariogram functions. Certainly, most of the parametric semivariogram models used in fitting are nonlinear in their parameters. However, the nonparametric semivariograms that have been the subject of study here are linear in their parameters, and thus, in theory at least, capable

of being fit by MINQUE methodology.

### $L_1$ Fitting

It is surprising that no one has tried to fit semivariogram models using  $L_1$  fits. For the nonparametric method discussed here the problem would become; find a  $m \times 1$  vector  $\mathbf{p} = (p_1, \dots, p_m)$  that minimizes

$$\mathbf{Q}(\mathbf{p}) = \sum_{i=1}^n w_i \left| \hat{\gamma}(h_i) - \sum_{j=1}^m (1 - \Omega_d(h_i t_j)) p_j \right| \quad (19)$$

subject to the constraint that  $p_j \geq 0; j = 1, \dots, m$ .

For fitting parametric models the problem is; find  $\theta$  that minimizes

$$\sum_{i=1}^k | \{ \hat{\gamma}(h_i) - \gamma(h_i; \theta) \} |$$

where  $\hat{\gamma}(h_i)$  is the sample variogram value at lag  $h_i$  and  $\gamma(h_i; \theta)$  is some appropriately chosen semivariogram model.

The method of moments estimator is known to be badly skewed. Cressie (1985 and 1991) advocated the use of his robust estimator to try to circumvent this problem in some sense. There is no mention anywhere in the geostatistical literature of fitting semivariogram models by minimizing the  $L_1$  norm.

This approach has much to recommend it. There are many algorithms available, and a great deal of work has been done in recent years. Gonin and Money (1989)

give a good summary of both linear and nonlinear  $L_p$  norm estimation. This topic is also worthy of further study.

## CHAPTER 8

### SUMMARY AND CONCLUSIONS

This chapter presents a brief summary of the major results and some suggestions on implementing the SB nonparametric semivariogram estimator.

#### Summary

The results presented in Chapter 3 demonstrated that the SB method of estimating nonparametric semivariogram functions yields estimated functions that fit as well as the more traditional parametric methods currently used. The method is robust to the selection of nodes as long as the collection is varied enough to capture the behavior of the data, particularly at larger lags. The method is much faster than parametric fitting, is more objective, and is easier to implement. This is true regardless of how the parametric models are fit. The method does not, in general, overfit; that is, data interpolation is not a problem. If it is thought necessary, additional smoothness constraints could be imposed using a variety of methods discussed briefly in Chapter 3. As pointed out in Chapter 7, the estimated

nonparametric semivariograms all have nugget effects that are equal to 0, but the nugget effect is incorporated into the estimated function away from the origin. This is a topic of further study, but the SB method appears to be robust to its failure to estimate the nugget effect.

The results of Chapter 4 show that the SB nonparametric estimator can be used to estimate the sill. The sill estimates are biased and highly variable, but a simple penalized least squares approach results in sill estimates that are not nearly as biased or variable without affecting fit. The sill estimates are comparable to those resulting from fitting with parametric models.

The arguments in Chapter 5 show that the SB estimator works as a filter, preserving the consistency properties (if any) of the sample semivariogram estimator used while ensuring the conditional negative definiteness property that valid semivariograms must have. This opens up the possibility of using other smoothing techniques for which consistency may be easier to establish.

### Implementation

For adequate fitting, the choice of nodes used here should suffice for many data sets. Recall that this was a selection of 200 nodes with 100 equispaced in  $[0.04, 4]$  and 100 equispaced in  $[4.16, 20]$ . The three dimensional kernel was used throughout this dissertation and appears to be adequate for the majority of problems. It tends

to produce smoother results than the one or two dimensional versions. If one really wants smoothness then the version that produces semivariograms valid in any dimension will work well.

For practical purposes, the penalized estimation method described in Chapter 4 need not necessarily be used unless the sill estimate is clearly unreasonable. If it is thought necessary to use this approach, one should choose the value of the penalty term that corresponds to the point of maximum curvature on the plot of the residual norm versus the  $\log_{10} \lambda$  values as illustrated in Chapter 4.

Appendix B contains listings of the computer codes used to implement the method. The Fortran listing of the program *NNLS* can be found in Lawson and Hansen (1974).

**REFERENCES CITED**

- Altman, N. S. (1990). Kernel smoothing of data with correlated errors. *Journal of the American Statistical Association*. 85:749-759.
- Armstrong, M., and Diamond, P. (1984). Testing variograms for positive-definiteness. *Journal of the International Association of Mathematical Geology*. 16:407-421.
- Baczkowski, A. J., and Mardia, K. V. (1987). Approximate lognormality of the sample semi-variogram under a Gaussian process. *Communications in Statistics. Simulation and Computation*. 16:571-585.
- Berg, C., Christensen, J. P. R, and Ressel, P. (1984). *Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions*. New York: Springer-Verlag.
- Bochner, S. (1955). *Harmonic Analysis and the Theory of Probability*. Berkeley: Univ. of California Press.
- Boot, J. C. G. (1964). *Quadratic Programming: Algorithms- Anomalies-Applications*. Chicago: Rand McNally.
- Christakos, G. (1984). On the problem of permissible covariance and variogram models. *Water Resources Research*. 20:251-265.
- Christensen, R. (1991). *Linear Models for Multivariate, Time Series, and Spatial Data*. New York: Springer-Verlag.
- Clark, I. (1979). *Practical Geostatistics*. Essex, England: Applied Science Publishers.
- Cressie, N. A. (1985). Fitting variogram models by weighted least squares. *Journal of the International Association for Mathematical Geology*. 17:693-702.
- Cressie, N. A. (1991) *Statistics for Spatial Data*. New York: John Wiley.
- Cressie, N. A., and Hawkins, D. M. (1980). Robust estimation of the variogram, I. *Journal of the International Association for Mathematical Geology*. 12:115-125.
- David, M. (1977). *Geostatistical Ore Reserve Estimation*. Amsterdam: Elsevier.
- Davis, B. M., and L. E. Borgman. (1979). Some exact sampling distributions for variogram estimators. *Mathematical Geology*. 11:643-653.

- Davis, B. M., and L. E. Borgman. (1982). A note on the asymptotic distribution of the sample variogram. *Mathematical Geology*. 14:189-193.
- Deutsch, C. V., and Journel, A. G. (1992). *GSLIB: Geostatistical Software Library and User's Guide*. New York: Oxford University Press.
- Diggle, P. J., and M. F. Hutchinson, M. F. (1989). On spline smoothing with autocorrelated errors. *Australian Journal of Statistics*. 31:166-182.
- Doob, J. L. (1953). *Stochastic Processes*. New York: John Wiley.
- Dowd, P. A. (1984). The variogram and kriging: Robust and resistant estimators. In *Geostatistics for Natural Resources Characterization, Part I*, G. Verly, M. David, A. G. Journel, and A. Marechal (eds). 91-106. Dordrecht: D. Reidel.
- Englund, E. J. (1990). A variance of geostatisticians. *Mathematical Geology*. 22:417-455.
- Gonin, R., and Money, A. H. (1989). *Nonlinear  $L_p$  Norm Estimation*. New York: Marcel Dekker.
- Haining, R. P. (1990). *Spatial Data Analysis for the Social and Environmental Sciences*. London: Cambridge Univ. Press.
- Hall, P., Fisher, N. I., and Hoffman, B. (1993). On the nonparametric estimation of covariance functions. Unpublished manuscript.
- Hall, P. and Wehrly, T. E. (1991). A geometrical method for removing edge effects from kernel-type nonparametric regression estimates. *Journal of the American Statistical Association*. 86:665-672.
- Hart, D., and Wehrly, T. E. (1986). Kernel regression estimation using repeated measurements data. *Journal of the American Statistical Association*. 81:1080-1088.
- Hastie, T, and Loader, C. (1993). Local regression: Automatic kernel carpentry. *Statistical Science*. 8:120-143.
- Hawkins, D. M., and Cressie, N. A. (1984). Robust kriging- a proposal. *Journal of the International Association for Mathematical Geology*. 16:3-18.
- Isaaks, E. H., and R. M. Srivastava. (1989). *An Introduction to Applied Geostatistics*. New York: Oxford University Press.

- Jones, R. H., and Vecchia, A. V. (1993). Fitting continuous ARMA models to unequally spaced data. *Journal of the American Statistical Association*. 88:947-954.
- Journel, A. G. (1985). The deterministic side of geostatistics. *Journal of the International Association for Mathematical Geology*. 17:1-14.
- Journel, A. G., and Huijbregts, C. J. (1978). *Mining Geostatistics*. London: Academic Press.
- Kennedy, W. J., Jr., and Gentle, J. E. (1980). *Statistical Computing*. New York: Marcel Dekker.
- Kitinidis, P. K. (1985). Minimum variance unbiased quadratic estimation of covariances of regionalized variables. *Journal of the International Association for Mathematical Geology*. 17:195-208.
- Kitinidis, P. K. (1987). Parametric estimation of covariances of regionalized variables. *Water Resources Bulletin*. 23:557-567.
- Lancaster, P., and Tismenetsky, M. (1985). *The Theory of Matrices. 2nd Edition*. New York: Academic Press.
- Lawson, C. L., and Hanson, R. J. (1974). *Solving Least Squares Problems*. Englewood Cliffs, NJ: Prentice-Hall.
- Lele, S. (1993). Inner product matrices, kriging, and nonparametric estimation of the variogram. Submitted for publication to *Mathematical Geology*.
- Marshall, R. J. and Mardia, K. V. (1985). Minimum norm quadratic estimation of components of spatial covariance. *Journal of the International Association for Mathematical Geology*. 17:517-525.
- Matern, B. (1986). *Spatial Variation, 2nd Edition*. Lecture Notes in Statistics, No. 36. New York: Springer-Verlag.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C*. Cambridge: Cambridge University Press.
- Quimby, W. F. (1986). Selected topics in spatial statistical analysis: Nonstationary vector kriging, large scale conditional simulation of three dimensional Gaussian random fields, and hypothesis testing in a correlated random field. Unpublished PhD Dissertation. Univ. of Wyoming, Laramie.

- Rao, C. R. (1970). Estimation of heteroscedastic variances in linear models. *Journal of the American Statistical Association*. 67:161-172.
- Rao, C. R. (1971a). Estimation of variance and covariance components-MINQUE Theory. *Journal of Multivariate Analysis*. 1:257-275.
- Rao, C. R. (1971b). Minimum variance quadratic unbiased estimation of variance components. *Journal of Multivariate Analysis* 1:445-456.
- Rao, C. R. (1979). MINQE theory and its relation to ML and MML estimation of variance components. *Sankhya(B)*. 41:138-153.
- Ripley, B. D. (1981). *Spatial Statistics*. New York: John Wiley.
- Ripley, B. D. (1988). *Statistical Inference for Spatial Processes*. Cambridge: Cambridge University Press.
- Rousseeuw, P. J., and Molenberghs, G. (1993). Transformation of non-positive semidefinite correlation matrices. *Communications in Statistics. Theory and Methods*. 22:965-984.
- Sampson, P. D., and Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*. 87:108-119.
- Seber, G. A. F., and Wild, C. J. (1989). *Nonlinear Regression*. New York: John Wiley.
- Schoenberg, I. J. (1938). Metric spaces and completely monotone functions. *Annals of Mathematics*. 39:811-841.
- Shapiro, A., and Botha, J. D. (1991). Variogram fitting with a general class of conditionally nonnegative definite functions. *Computational Statistics and Data Analysis*. 11:87-96.
- Stein, M. L. (1987). Minimum norm quadratic estimation of spatial variograms. *Journal of the American Statistical Association*. 82:765-772.
- Upton, G. J. G., and Fingleton, B. (1985). *Spatial Data Analysis by Example, Volume 1: Point Pattern and Quantitative Data*. Chichester: John Wiley.
- Upton, G. J. G., and Fingleton, B. (1989). *Spatial Data Analysis by Example, Volume 2: Categorical and Directional Data*. Chichester: John Wiley.

- von Neumann, J., and Schoenberg, I. J. (1941). Fourier integrals and metric geometry. *Transactions of the American Mathematical Society*. 50:226-251.
- Warnes, J. J., and Ripley, B. D. (1987). Problems with likelihood estimation of covariance functions of spatial Gaussian processes. *Biometrika*. 74:640-642.
- Watkins, A. J. (1992). On models of spatial covariance. *Computational Statistics and Data Analysis*. 13:473-481.
- Yaglom, A. M. (1987a). *Correlation Theory of Stationary and Related Random Functions I: Basic Results*. New York: Springer-Verlag.
- Yaglom, A. M. (1987b). *Correlation Theory of Stationary and Related Random Functions II: Supplementary Notes and References*. New York: Springer-Verlag.
- Zimmerman, D. L., and Zimmerman, M. B. (1991). A Monte Carlo comparison of spatial semivariogram estimators and corresponding kriging predictors. *Technometrics*. 33:77-91.

**APPENDICES**

## APPENDIX A

### SIMULATION OF DATA

Data was simulated as follows. It is assumed that the data came from  $n$  locations one unit apart in  $\mathbf{R}^1$ . The nugget effect was set equal to 0. A valid covariance function is applied to an interpoint distance matrix generated from this scenario. For example, suppose observations are located at points 1, 2, 3, 4, and 5. A matrix of interpoint distances is formed. For example;

$$\mathbf{D} = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & 0 & 1 & 2 & 3 \\ 2 & 1 & 0 & 1 & 2 \\ 3 & 2 & 1 & 0 & 1 \\ 4 & 3 & 2 & 1 & 0 \end{pmatrix}.$$

Valid semivariogram models for a second-order stationary process have a corresponding covariance function. Thus, the exponential covariance function, is

$$C(\mathbf{h}; \theta) = s\{\exp(-3h/r)\}$$

where  $\theta = (s, r)'$ . Then  $C(\mathbf{D})$  is a valid covariance matrix for specified values of  $\theta$ . Take a Cholesky Decomposition of  $C(\mathbf{D})$  (this is possible because  $C(\mathbf{D})$  is positive definite). This results in  $C(\mathbf{D}) = \mathbf{R}\mathbf{R}'$  where  $\mathbf{R}$  is an upper triangular matrix. Now a random sample  $\mathbf{y}$  of size  $n$  is taken from a normal distribution with mean 0

and variance 1. Then  $\mathbf{z} = \mathbf{R}\mathbf{y}$  is multivariate normal with mean  $\mathbf{0}$  and covariance matrix  $\Sigma(\theta) = \mathbf{R}\mathbf{R}'$ . The  $n \times 1$  vector  $\mathbf{z}$  is a simulated spatial stochastic process with a known covariance structure. The generalization of this to higher dimensions is straightforward.

The simulation of spatial random fields is an active area of research in its own right. Cressie (1991) gives a good summary of current techniques. The one described above is easy and gives good results. It was used by Quimby (1986) and Zimmerman and Zimmerman (1991).

## APPENDIX B

### SOFTWARE USED

There are three main Splus functions that do the computations reported on in the dissertation. They are `vario.cloud`, `npvar`, and `meandist`. The function `vario.cloud` will compute a semivariogram cloud. The function `meandist` takes the output of `vario.cloud` and computes the method of moments sample semivariogram. The function `npvar` produces the estimated semivariogram function(s) using the SB method.

It is assumed that the user has  $n$  realizations that are equally spaced on a one dimensional transect. If the data is irregularly spaced and/or comes from two or three dimensions then `vario.cloud` and `meandist` will not work and the user will need to determine the semivariogram cloud and the sample semivariogram using some other method. For example, Deutsch and Journel (1992) provide Fortran code for determining sample semivariograms under these conditions. Also, Brian Ripley has produced a set of Splus functions for the analysis of spatial data. These functions are available on Statlib. To get them, send the e-mail message,  
  
`send spatial from S`

to `statlib@temper.stat.cmu.edu`. This entry is a shar archive, and if the user does not know how to deal with a shar archive then the message

```
send shar from general
```

should be sent. It will result in instructions on how to handle shar archives.

The function, `vario.cloud` returns a matrix that contains the information necessary to plot a semivariogram cloud. Generally, it is recommended that the user work only with lags for which there are at least 30 observations in the semivariogram cloud.

Suppose `raw.data` is a vector in Splus containing 50 realizations of  $Z(\mathbf{s}_i)$  located one unit apart on a one dimensional transect. There are 49 distinct lags ( $h$ ), but only lags 1 to 20 have at least 30 observations of the form

$$Y_{ij} = \frac{[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]^2}{2}.$$

There are 49 pairs at  $h = 1$ , 48 pairs at  $h = 2$ , and so on up to 30 pairs at  $h = 20$ .

Then

```
cloud_vario.cloud(raw.data)
```

results in a 790 by 2 matrix. The first column consists of the lags, and the second column has the unique observed  $y_{ij}$ 's at each  $h$ . If desired, it is possible to get the entire semivariogram cloud by the command,

```
cloud_vario.cloud(raw.data,thirty=F).
```

It is then possible to generate the sample semivariogram by

```
samp.semvar_meandist(cloud).
```

The result `samp.semvar` is a  $20 \times 2$  matrix whose first column contains the unique lags (numbered 1 to 20) and whose second column contains the sample semivariogram values at those lags. If plotted, (`plot(samp.semvar)`) the output produces a plot similar to that shown in Fig 2 (page 29).

Once the sample semivariogram has been computed `npvar` is used to generate the nonparametric semivariogram. It is recommended that the first fit be done with no penalties imposed. This is especially true if one is only interested in getting a nonparametric function and has no interest in the sill estimate. If the sill estimate is of interest, and the unpenalized estimate appears unreasonable, then `npvar` can be used to get a penalized fit. The penalized fitting procedure may also be used to impose additional smoothness constraints if that is desired.

Figure 32 on page 89 was generated using this function. The sample semivariogram is from Clark (1979) and was stored in the vector `clark.silver`. The commands used were,

```
fig32.result_npvar(clark.silver).
```

The function returns

- the solution vector (`fig32.result$xhat`)

- the sill estimate (`fig32.result$sill`)
- the norm of the residual vector (`fig32.result$resnorm`)

If `show=T` is included in the function call, a plot of the fit will be produced.

In this case there was really no need to do a penalized fit. However, if it is felt necessary to do a penalized fit to stabilize the sill estimate then a vector `lambda` of penalty terms needs to be specified. The default value of  $\lambda$  is a  $45 \times 1$  vector with values such that  $\log_{10}(\lambda) = (-9, -8.75, -8.5, \dots, 1.5, 1.75, 2)'$ . Then,

```
result_npvar(clark.silver,penalty=1)
```

returns the following,

- a  $45 \times 1$  vector of sill estimates (`result$sill.vec`)
- a  $45 \times 1$  vector of residual norms (`result$resid.vec`)
- a matrix of solution vectors (`result$xhat`)

It is possible for a user to specify other value(s) for the penalty term. Denoting, this selection by `my.lambda`, these fits are obtained by

```
result_npvar(clark.silver,penalty=1,lambda=my.lambda).
```

The function also can produce plots like those in Figures 23 through 25 on pages 72 to 74 by specifying `show=T` in the function call. This can assist the user in

choosing the optimal value of  $\lambda$ . In general the optimal value will be close to the point of maximum curvature in the plot of the residual norm versus the  $\log_{10}(\lambda)$  values.

These functions and supporting documentation are available on Statlib. To get them, send the e-mail message,  
send npvar from S  
to [statlib@temper.stat.cmu.edu](mailto:statlib@temper.stat.cmu.edu).

