Abstract

MA, LIYUN. Spectral Methods for Likelihood Approximation of Spatial Processes. (Under the direction of Dr. MONTSERRAT FUENTES.)

Likelihood methods for large spatial data sets are often very difficult, due to computational burden. Even for Gaussian random fields, exact calculation of the likelihood for a spatial process at n locations requires $O(n^3)$ operations and hence is impractical for large data sets. We present a version of Whittle's approximation to the Gaussian log likelihood for spatial regular lattices which only requires $O(nlog_2n)$ operations and does not involve calculating determinants. Due to edge effect the estimated covariance parameters using this approximated likelihood method are efficient only in one dimension. To remove this edge effect, we introduce a new form of data taper, a rounded data taper that gives more tapering for corner observations. Therefore, with less overall tapering, we get the amount of smoothing that we need without losing so much information and obtain better estimates of the spectral density. We present simulation and theoretical results to show the benefits and the performance of the new data taper and the spectral likelihood approximation for large spatial data sets. Applications include the spatial analysis of sea surface temperature (SST) data provided by National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center (NCDC).

Spectral Methods for Likelihood Approximation of Spatial Processes

by

Liyun Ma

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

STATISTICS

Raleigh

2006

APPROVED BY:

Dr. Montserrat Fuentes Chair of Advisory Committee	Dr. Richard Reynolds
Dr. Dennis D. Boos	Dr. Wenbin Lu

Biography

Liyun Ma was born on January 20th, 1979 to Yongcheng Ma and Xiangzheng Yin in Rudong, China. She graduated from Rudong High School in Rudong in 1997. She received a Bachelor of Management Science degree from University of Science and Technology of China in Hefei, China in 2001. She then joined the graduate program in the Department of Statistics at North Carolina State University in August 2001. She hopes to receive her Doctor of Philosophy degree in Statistics in August 2006.

Acknowledgements

First, I would like to express my deepest appreciation to my wonderful advisor Montserrat Fuentes for her guidance, training and support during this research. I feel so fortunate to have worked with her and learned to become a more independent researcher under her direction. Her efforts and help have meant a tremendous amount to me.

I also want to thank the support from NOAA National Climatic Data Center (NCDC). I am grateful to the valuable suggestions and comments of Dr Richard Reynolds. He gave me a lot help in the study of Sea Surface Temperature field.

My thanks also go to Dr. Dennis Boos, Dr. Marcia Gumpertz, and Dr. Wenbin Lu for their warm help with various problems in my research.

Finally I want to thank my husband Shuchun Wang, my parents and my sister for their selfless love and support.

Research funded by the U.S. National Oceanic and Atmospheric Administration. Grant Award Number NAO3NES4400015 through a Cooperative Agreement (Climate Weather Impacts on Society and the Environment) via the Charleston Coastal Services Center and the National Climatic Data Center.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Overview

1.1 Motivation

Estimating spatial covariance structures is fundamental in spatial statistics. However in the analysis of huge data sets, using a likelihood approximation to estimate the covariance parameters is very difficult since inversion of a large covariance matrix to compute a likelihood function may not be possible or may demand much effort and time in computation. Moreover, data sets that encompass thousands of location sites become more prevalent. One example of huge geostatistical data is global sea surface temperature (SST) data. SST fields are useful for weather forecasting, climate monitoring and climate prediction. The SST analysis is usually carried on using in situ (ship and buoy) and satellite data sets. The resolution of the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) SST data set is 0.25 degree longitude by 0.25 degree latitude, i.e. the global domain of the SST data contains more than hundred thousand locations.

The size of this spatial problem is not unusual. However, if the observations are on a regular complete lattice, we can use spectral methods to compute the likelihood function. Spectral methods are a powerful tool for studying the spatial structure of random processes and generally offer significant computational benefits. A common approach is Whittle's approximation to the Gaussian log likelihood for spatial regular lattice. However, due to edge effects, the Whittle's likelihood approximation is not consistent.

Data tapering is an important tool in two and higher dimensional problems where there are a large number of boundary observations. In spatial statistics, data tapers are often the tensor product of two one-dimensional tapers. However we generally need more tapering for the corner observations. Thus we introduce a new data taper, a rounded data taper, that gives more tapering to the observations at four corners. Therefore, with less overall tapering, we get the amount of smoothing that we need without losing so much information.

We describe spectral methods for spatial likelihood approximation to estimate covariance parameters. We use the new data taper to filter the raw data before computing the periodogram, then use the tapered periodogram to approximate the spectral density in Whittle's likelihood function. We present theoretical results to prove that under some regularity conditions the tapered Whittle estimate is asymptotically consistent. We investigate the performance of our spectral likelihood in fixed domain through a simulation study. We present an application of the proposed tapered spectral approximation to the likelihood to a large spatial data set: Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) TA monthly average sea surface Temperature data. This application illustrates the performance of the covariance estimators when dealing with a real large data set, presenting the impact of the method on practical issues.

This chapter is organized as follows. In section 2 we briefly review the approaches proposed to estimate covariance structures in spatial statistics and discuss their advantages and disadvantages. In section 3 we give a outline of the dissertation.

1.2 Literature Review

Mardia and Marshall (1984) have considered maximum likelihood approach to estimate the components of spatial covariance. But computing the likelihood function for a Gaussian process at n locations requires $O(n^3)$ operations. So if there are more than several hundred observations, exact likelihood calculations can be very difficult, if not infeasible.

Vecchia (1988) proposed a simple likelihood approximation in a spatial setting. In Vecchia's paper, a sequence of approximate likelihood function L_m is defined where m indicates that m observations out of n observations are included in the likelihood function. The smaller the value of m, the more efficient the computation but worse approximation. L_m approaches the conventional likelihood function as m goes to n. An iterative estimation procedure is developed such that estimates based on L_1

are used as initial values for estimates based on L_2, and so on. Statistics computed at each step of the procedure are used to determine when the iterative parameter estimates have converged. This method has been implemented in a computer program AMLE3D written by Pardo-Igúzquiza and Dowd (1997) for three dimensional spherical, exponential and Gaussian models of spatial covariance.

A related method for approximating the likelihood is to divide the observation region into numbers of subregions, calculate the likelihood for each subregion separately and then multiply these likelihoods together. Similar to Vecchia's method, smaller subregions lead to easier computation but worse approximation of the likelihood. Stein (1986) proposed such a procedure for minimum norm quadratic estimators, which also requires computing quadratic forms of the inverse covariance matrices.

Stein et al. (2004) adapted Vecchia's method to approximate the restricted likelihood of a Gaussian random process, and discussed the computational challenges to obtain the derivatives of the approximated likelihood and restricted likelihood functions.

Caragea (2003) proposed several approximations to the likelihood based on grouping the observations with clusters and building an estimating function by accounting the variability both between and within groups. In this way, the estimation becomes practical for considerably large data sets. However, for his approximation methods, calculating the variance of the second derivative of pseudo negative likelihood function is a much more difficult procedure and it is still part of Caragea's future research.

If the observations are on a regular complete lattice, it is possible to use the spectral likelihood approximation proposed by Whittle (1954) to estimate the covariance parameters. This method only requires $O(nlog_2n)$ operations and does not involve calculating determinants and the inverse of the covariance matrix. In this setting, the approximate likelihood can be calculated very efficiently by making use of the fast Fourier transform (FFT).

The spatial periodogram is a powerful tool to estimate the correlation structure of a stationary process. It is a nonparametric estimate of the spectral density, which is the Fourier transform of the covariance function. Guyon (1982) proved that when the periodogram is used to approximate the spectral density in Whittle's likelihood function, the periodogram bias is non-negligible for 2-dimensional processes, and for 3-dimensions the bias dominates the mean squared error (MSE) of the parameter estimates. Moreover, Stein (1995) showed that in the fixed domain asymptotic where the number of observations in a fixed study area increases, using the periodogram can yield highly misleading results. Thus, the covariance parameter estimates based on the Whittle's likelihood function are only efficient in one dimension, but not in two and higher dimensions. Guyon used another version of the periodogram, an "unbiased periodogram" to solve this problem. The unbiased periodogram is the discrete Fourier transform of an unbiased version of the sample covariance c_n^*. However c_n^* has some other unpleasant properties. The first disadvantage is that it is not always positive definite so that the spectral estimates based on c_n^* may be negative. Under some circumstances the minimum of Whittle's function may not even exist. Secondly, c_n^* has large variance, particularly for large lags. In practical situations the use of

c_n^* instead of c_n may make a big difference, as in the last example of Mardia and Marshall (1984). Thirdly, c_n^* has a severe bias if some nonstationarities are present at the boundary, for instance the boundary plots in a field trial behave differently.

Instead of smoothing biased periodogram estimates, direct filtering of the raw data with a data taper can also provide a consistent estimate of the covariance parameters. Dahlhaus and künsch (1987) used data tapers to avoid the bad properties of c_n^* and get asymptotically negligible bias. They proved a central limit theorem for the tapered Whittle estimate $\hat{\theta}_n$ and showed that under some regularity conditions it is asymptotically efficient. The result indicates that in two and three dimensions the tapered estimates is an great compromise between the classical biased and the unbiased versions avoiding their disadvantages and retaining many of the good properties.

1.3 Outline

In chapter 2 we introduce the study in global sea surface temperature (SST) fields. We propose a geostatistical approach including variogram estimation and ordinary kriging to estimate the biases of the satellite SST. We also get the parameter estimation of the noise to signal ratio, zonal spatial scale and meridional scale for the biases.

In chapter 3 we offer a general introduction of spectral domain. We introduce the notation of Fourier transform and its properties. We describe the spectral representation of a stationary spatial process, the notation of periodogram, spectral density and the Whittle's approximation to the Gaussian likelihood.

In chapter 4, we propose a new data taper function for spatial data, which we called rounded data taper. The benefit of the rounded data taper is that it gives more tapering to the corner observations. Therefore, with less overall tapering, we get the amount of smoothing that we need without losing so much information.

We present a method to choose the appropriate smoothing parameters for data taper by taking into account the trade-off between bias and variance of the tapered periodogram. Since the variance of the periodogram increases when the bias is reduced, we minimize the relative mean squared error (MSE) of the periodogram to get estimates of the taper parameters so that we can get better estimates of the spectral density.

By examining the impact of data taper on the spectral window of the periodogram, we found that both rounded taper and muitiplicative taper reduce the side lobes of the periodogram estimates. More importantly, the periodogram after applying the rounded taper has much more reduced side lobes than the periodogram after applying a multiplicative taper.

In chapter 5, we present a spectral method to approximate the likelihood for Gaussian lattice process. The key idea is to use the new data taper to filter the raw data before computing the periodogram then use the tapered periodogram to approximate the spectral density in the Whittle's likelihood function. We use similar proof as Dahlhaus and künsch (1987) to show that Whittle's likelihood estimates for covariance parameters after applying rounded taper hold the same asymptotic properties as Whittle's likelihood estimates after applying multiplicative taper. That's to say they are asymptotically consistent. We also present results from a simulation study to show that for finite samples the covariance estimates obtained from our spectral likelihood method are more appealing than covariance estimates obtained from spectral likelihood method using muitiplicative taper and covariance estimates from exact likelihood method.

Chapter 6 consists an application of proposed spectral likelihood approximation method to sea surface temperature (SST) data provided by National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center (NCDC).

The last chapter contains a discussion and suggestions for directions of future research. Fuentes (2006) introduced spatial likelihood approximation methods using spectral tools for irregularly spaced spatial data sets. We suggest a promising approach of following the principles outlined in her approach and study the potential impact of tapered spectral methods on the likelihood approximation, prediction and inference made for incomplete and irregularly spaced spatial data.

Chapter 2

Sea Surface Temperature Field

Global sea surface temperature (SST) fields are useful for monitoring climate change, as an oceanic boundary condition for atmospheric models and as a diagnostic tool for comparison with the SSTs produced by ocean models. SST fields are also important for fishery since SST is closely related to the generation of clouds and a sea fog in coastal area and over the ocean.

SSTs have been measured from Advanced Very High Resolution Radiometer (AVHRR) aboard the National Oceanic and Atmospheric Administration (NOAA) satellite since 1981. The great advantage of these satellite SSTs is their global, repeated coverage, compared with in situ measurements by ship and buoy. However, the operational feasibility of deriving SSTs from AVHRR is severely limited by the inability of AVHRR to penetrate cloud cover.

The measurement of SSTs through clouds by satellite microwave radiometers has been an elusive goal for many years. The early radiometers in the 1980's were poorly calibrated, and the later radiometers lacked the low frequency channels needed by the retrieval algorithm. Finally, in December 1997, microwave SSTs become available from the Tropical Rainfall Measuring Mission (TRMM) satellite Microwave Imager (TMI). The important feature of microwave retrievals is that SST can be measured through clouds. This is a distinct advantage over the traditional infrared SST observations from AVHRR that require a cloud-free field of view.

The use of satellite data can significantly improve the in situ analysis, especially in regions of sparse in situ data. But research has shown that both AVHRR and TMI data have biases that change with space and time as physical properties of the atmosphere change and as the satellite instruments and the orbits of the satellite change. It is critical to monitor differences between satellite instruments and the other products to quickly diagnose any of these changes.

Reynolds et al. (1989) showed that the largest biases of AVHRR data in the record were due to stratospheric aerosols from the April 1982 volcanic eruptions of EL Chi-Chon and June 1991 volcanic eruptions of Mount Pinatubo. The difference between the in situ and the nighttime satellite data also suggests that the aerosol-corrected equation did not correct all the biases caused by the aerosols and that it may not be completely independent of the aerosol concentration which has spatial and temporal variations. The largest errors in the TMI SST retrievals occur near land and at high wind speeds. In addition, errors are caused by engineering problems within the main reflector and the onboard satellite attitude control system.

Because strong satellite SST biases are evident, a preliminary step using Poisson's equation was added before performing optimum interpolation (OI) analysis to correct any large scale satellite biases. Here we present a statistical approach to combine the in situ data and the satellite data to model the spatial structure of the TMI satellite SST biases. We estimate the biases by first reexamining the differences between satellite and in situ data. Then a new system will be developed using the data which will be integrated with the ordinary kriging (OK) processing.

This chapter is organized as follows. In section 2, we introduce a statistical method for spatial interpolation, which includes semivariogram estimation and ordinary kriging. In section 3, we apply the methodology presented in section 2 to sea surface temperature (SST) data and estimate the noise to signal ratio, meridional spatial scale and zonal scale for the satellite SST biases. Section 4 has a discussion.

2.1 Statistical Methodology

2.1.1 Statistical Model

Assume we observe a random spatial process Z at locations $\boldsymbol{s}_1, \boldsymbol{s}_2 \ldots, \boldsymbol{s}_n$, the observations can be modeled as

$$Z(\boldsymbol{s}) = m(\boldsymbol{s}) + \epsilon(\boldsymbol{s}) \tag{2.1}$$

Here, $m(\boldsymbol{s}) = E(Z(\boldsymbol{s}))$ and $\{\epsilon(\boldsymbol{s}), \boldsymbol{s} \in D\}$ is a zero mean random field, associated with a covariance function

$$C(\boldsymbol{s}, \boldsymbol{t}) = \mathrm{cov}(Z(\boldsymbol{s}), Z(\boldsymbol{t})) = \mathrm{cov}(\epsilon(\boldsymbol{s}), \epsilon(\boldsymbol{t})). \tag{2.2}$$

The process is called stationary if the mean $m(\boldsymbol{s})$ is constant and the $\text{var}(Z(\boldsymbol{s}) - Z(\boldsymbol{t}))$ depends only on the sites relative position $(\boldsymbol{s} - \boldsymbol{t})$, i.e. there is a function such that:

$$\text{var}(Z(\boldsymbol{s}) - Z(\boldsymbol{t})) = 2\gamma(\boldsymbol{s} - \boldsymbol{t}) \tag{2.3}$$

It is easy to see that

$$\gamma(\boldsymbol{h}) = C(\boldsymbol{0}) - C(\boldsymbol{h}) \tag{2.4}$$

$\gamma(\boldsymbol{h})$ is called semivariogram. It describes the spatial correlation between observations and tells us how different, on average, are measurements at various distances apart.

2.2 Ordinary Kriging

The geostatistical technique known as kriging was developed in the late 1950's, building on the pioneering work of a South African mining engineer named Danie Krige. Formally, kriging now refers to a family of least-squares linear regression algorithms that attempt to predict values of a random variable at locations where data for the variable is not available, based on the spatial pattern of the available data.

Ordinary kriging is the most widely used kriging method which accounts for local fluctuations of the mean by limiting the area where the mean for Z is fixed to some local neighborhood. This variant of kriging considers the mean for Z to be constant but unknown for all locations in the local neighborhood.

Let $\boldsymbol{s}_0$ denote an arbitrary location in D, usually this will be an unsampled location, the ordinary kriging predictor of the value of $Z(\boldsymbol{s}_0)$ at interested locations $\boldsymbol{s}_0$ is a linear combination of the data values, i.e.,

$$\hat{Z}(\boldsymbol{s}_0) = \sum_{i=1}^{n} \lambda_i Z(\boldsymbol{s}_i) \tag{2.5}$$

where the weight λ_i is the ith element of $\Gamma^{-1}\gamma$. γ is a $n \times 1$ vector and Γ is a $n \times n$ matrix. The ith element of γ is the semivariogram $\gamma(\boldsymbol{s}_i - \boldsymbol{s}_0)$ and the ith, jth element of Γ is $\gamma(\boldsymbol{s}_i - \boldsymbol{s}_j)$.

The ordinary kriging predictor is also known as best linear unbiased predictor (BLUP) since it is unbiased and minimizes $\text{var}(\hat{Z}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0))$ subject to the restriction $\sum_{i=1}^{n} \lambda_i = 1$. The minimized variance is called the kriging variance.

Ordinary kriging is derived under the assumption that the semivariogram is known. In practice, the semivariogram is unknown and must be estimated, and the $\hat{\lambda}$ replaces λ in the kriging equations and in the expressions for the kriging variance.

2.2.1 Semivariogram Estimation

The sample semivariogram (also called empirical semivariogram), a statistical model computed by using the available sample data, is the key part of the kriging procedure. It expresses how the data vary spatially across the area of interest. The classical or method of moments estimator for the sample semivariogram is

$$\check{\gamma}(\boldsymbol{h}) = \frac{1}{2N(\boldsymbol{h})} \sum_{N(\boldsymbol{h})} [Z(\boldsymbol{s}) - Z(\boldsymbol{s} + \boldsymbol{h})]^2 \tag{2.6}$$

Figure 2.1: Sample semivariogram

Figure 2.2: Exponential semivariogram

where $N(\boldsymbol{h})$ is the number of pairs for lag $\boldsymbol{h}$.

To calculate the sample semivariogram, we need to divide the range of distances between sampled points (from zero to the maximum distance) into a set of discrete intervals, then assign each pair of sampled points to one of the distance intervals according to the distance between the points and use the equation (2.6) to compute the average of the square of difference between sampled points in each distance interval. The resulting value is then the semivariogram estimation for the midpoint distance of each distance interval $\boldsymbol{h}$. Figure 2.1 is a plot of sample semivariogram. As can be seen in the figure, the shape of the semivariogram indicates that at small separation distances, the variance is small. In other words, points that are close together have similar values.

Once the experimental semivariogram is computed, the next step is to define a model semivariogram. A model semivariogram is a simple mathematical function that models the trend in the experimental semivariogram. There are various theoretical semivariogram models such as triangular model, spherical model, exponential model, Matern class of models, linear and power models, etc.

In a semivariogram model, the variance levels off after a certain distance beyond which observations appear independent is called sill and the certain distance is called range. Nugget effect is the vertical jump from the value of semivariogram at extremely small separation distances due to sampling error or short scale variability. An example of exponential semivariogram model is presented in figure 2.2, where the sill is 1.0, the range is 0.2 and the nugget is 0.2.

There are generally two methods to fit a theoretical semivariogram model to the sample semivariogram: weighted nonlinear least squares (WSL) and maximum likelihood (ML) or restricted maximum likelihood (REML).

Let $\gamma(\boldsymbol{h};\boldsymbol{\theta})$ denote the parametric model to be fit to the sample semivariogram and let Θ denote the parameter space for $\boldsymbol{\theta}$. A weighted nonlinear least squares estimator of $\gamma(\boldsymbol{h};\boldsymbol{\theta})$ is defined as a value of $\hat{\boldsymbol{\theta}} \in \Theta$ that minimizes the weighted residual sum of squares function:

$$\omega(\boldsymbol{\theta}) = \sum_{u \in U} \frac{N(\boldsymbol{h}_u)}{[\gamma(\boldsymbol{h}_u;\theta)]^2}[\hat{\gamma}(\boldsymbol{h}_u) - \gamma(\boldsymbol{h}_u;\boldsymbol{\theta})]^2. \tag{2.7}$$

where U is a specified subset of lag classes believed to yield reliable estimated of $\gamma(\boldsymbol{h})$.

If we let $V(\boldsymbol{\theta})$ denote the covariance function matrix of $\mathbf{Z} = (Z(\boldsymbol{s}_1), Z(\boldsymbol{s}_2), \ldots, Z(\boldsymbol{s}_n))$ and let $\mathbf{X}$ denote the model matrix for the model $\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \epsilon$, the log likelihood function is

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{Z}) = -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})^{'}\mathbf{V}^{-1}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}) \tag{2.8}$$

A MLE value is a value $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$ that maximizes $L(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{Z})$. Generally a MLE must be found by numerical optimization routines. Thus we have to be concerned with starting values and convergence criteria.

A restricted MLE (REML) is defined as a value $\hat{\boldsymbol{\theta}} \in \Theta$ that maximizes the log likelihood function associated with $n - \text{rank}(\mathbf{X})$ linearly independent error contrasts. It is less biased than MLE.

2.3 Application

In this application, we choose a SST data set to demonstrate our approach of estimating the biases of the TMI satellite sea surface temperature. The area of the data set covers longitude 149W to 131W and latitude 21N to 39 N. The data set is a monthly average of all data within March, 1998.

We interpolate the in situ SST at the locations of the satellite data and compare the estimated in situ SST with the TMI SST to get the estimation of the biases of the TMI SST. After we get the estimation of the TMI SST biases, we estimate the covariance function of the bias surface. We are interested in three parameters of the covariance function: the noise to signal ratio, the meridional spatial scale and the zonal scale. The noise to signal ratio is the ratio of nugget and partial sill. The zonal spatial scale is the range at west to east direction. The meridional spatial scale is range at north to south direction.

The satellite data we worked with are the Tropical Rainfall Measuring Mission

(TRMM) Microwave Imager (TMI) TA data. They are processed using a Remote Sensing System algorithm to obtain sea surface temperatures (SST). The resolution of the data set is 0.25 degree longitude by 0.25 degree latitude. Figure 2.3 is a map of the TMI satellite data.

The monthly in situ data we studied are gathered from the Global Telecommunications System (GTS) by NOAA's National Centers for Environmental Prediction (NCEP). The in situ SST data are determined from observations from ships and buoys. Most ship observations are made from insulated buckets, hull contact sensors, and engine intakes at depths of one to several meters. The observations from drifting and moored buoys are made by thermistor or hull contact sensor and usually relayed in real time by satellite. Figure 2.4 gives the locations of the in situ data.

Figure 2.3: Map of TMI SST data

Figure 2.4: Locations of in situ SST data

First, we calculate the sample semivariogram of the in situ data. Since the sample semivariogram levels out and is curvy all the way up, we consider both the exponential model and the Gaussian model. We use weighted nonlinear least square (WLS) method to fit both models to the sample semivariogram and estimate the covariance parameters. Since using pairs of points that are independent adds nothing to the sample semivariogram plot but noise, we usually only use 1/2 to 3/4 of the maximum lag as the maximum distance for the semivariogram.

In Figure 2.5, the open circle is the sample semivariogram and the black dot-dash line is the sample semivariogram with smaller lags. The red solid line indicates the fitted exponential model and the green dashed line indicates the fitted Gaussian model. From the plot, we can see, both models capture the shape of sample semivariogram very well. Since the Gaussian model gives smaller weighted least square value (5.9) than the exponential model (13.2), we choose to use the Gaussian semivariogram model.

Figure 2.5 shows that the choice of Gaussian semivariogram model is adequate. We take this form and use the parameters estimated from WLS as the starting value for the restricted maximum likelihood (REML) method and maximum likelihood (ML) method to estimate the covariance parameters. The plot of the Gaussian semivariogram models fitted by REML and ML are shown in Figure 2.6. So that the estimated semivariogram model is of the form

$$\gamma(\boldsymbol{h}) = C_0 + C_1(1 - \mathrm{e}^{-\boldsymbol{h}^2/R^2}) \tag{2.9}$$

where C_0 is the nugget, $C_0 + C_1$ is the sill, C_1 is the partial sill and R is the range. The parameter estimates got from REML method are $\hat{C}_0 = 0.3744, \hat{C}_1 = 1.209, \hat{R} = 800$ and the estimates from ML method are $\hat{C}_0 = 0.4207, \hat{C}_1 = 0.9505, \hat{R} = 800$.

Figure 2.5: Theoretical semivariogram curves added to the sample semivariogram of the in situ SST.

Figure 2.6: Gaussian semivariogram models fitted by REML and ML.

We pick the parameter estimations get from REML method and use the fitted Gaussian model to get the kriging estimates and the associated standard errors for in situ SST . The map of the estimates and the kriging standard errors are plotted in figure 2.7 and figure 2.8.

The estimates of the TMI SST biases are the difference between the TMI SST data and the in situ SST estimates. Since we are interested in the noise to signal ratio, meridional spatial scale and the zonal scale of the biases, we fit theoretical semivariogram models to the sample semivariogram at west to east direction and north to south direction separately. The figure 2.9 is the plot of sample semivariogram of the biases estimates, the red line is the fitted exponential semivariogram. The figure 2.10

Figure 2.7: Map of TMI SST estimates.

Figure 2.8: Kriging standard errors.

is the plot of sample semivariogram of the biases estimates at west to east direction, the red line is the fitted exponential semivariogram. The figure 2.11 is the plot of sample semivariogram of the biases estimates at north to south direction, the blue line is the fitted exponential semivariogram.

The results of the parameter estimates for the TMI SST biases are presented in Table 2.1. The estimate of nugget for TMI biases is 0.0214 and the estimate of partial sill is 0.1675. The noise to signal ratio is the ratio of nugget and partial sill so that the estimate of noise to signal ratio is 0.13. The estimate of zonal spatial scale is the estimate of range at west to east direction, which is 821.6 kilometer (km). The estimate of meridional spatial scale is the estimate of range at north to south direction, which is 562.5 kilometer (km).

Figure 2.9: Theoretical semivariogram curves added to the sample semivariogram of the biases estimates.

Figure 2.10: Semivariogram of the biases estimates at west to east direction.

Figure 2.11: Semivariogram of the biases estimates at north to south direction.

Table 2.1: Parameter estimations of TMI SST biases.

Parameters	Units	Estimates
Noise to signal ratio		0.13
Zonal spatial scale	(km)	821.6
Meridional spatial scale	(km)	562.5

2.4 Discussion

In this work, we introduced a geostatistical method to estimate the biases of the satellite sea surface temperature (SST) biases. The overall purpose is to improve the climate-scale SST analysis produced at NOAA as described by Reynolds and Smith (1994) and Reynolds et al. (2002). This effort is designed to support development of a climate ocean observing system. The analysis is done by using semivariogram estimation and ordinary kriging to estimate any large scale satellite biases relative to the in situ data.

The analysis uses satellite data from Tropical Rainfall Measuring Mission (TRMM) satellite Microwave Imager (TMI) and in situ data from ships and buoys gathered from the Global Telecommunications System (GTS) by NOAA's National Centers for Environmental Prediction (NCEP). The data in the application is only from a small area. We have used this method to the global SST data and found that the zonal scale and meridional scale of satellite SST biases have spatial variations. The spatial variations are strongly linked to the types of biases that occurred over the time period. Those biases may be useful as guidance for estimating the general scale s biases, but their shape and locations should not be assumed to be always fixed. It may be most reasonable to use globally-constant bias statistics to represent the most typical scales.

Since the size of global SST data set is very large, we have to divide the global region into several subregions to be able to use the method we introduced in this chapter. However since the satellite observations are on a regular complete lattice, we can use spectral method to approximate the likelihood function. We will introduce the spectral likelihood approximation method in later chapters and use spectral method on the same data set to get the estimation of covariance parameters.

Chapter 3

Spectral Analysis of Spatial Process

The analysis of stationary processes with their spectral representations in the frequency domain is referred to as spectral analysis. Spectral analysis is a powerful tool for studying the spatial structure of random fields and generally offer significant computational benefits. It is particularly advantageous in analyzing large data sets and in studying properties of multivariate processes.

In section 1 of this chapter we review the Fourier transform and its properties. In section 2, we introduce the spectral representation of a stationary process and describe some commonly used classes of spectral densities. In section 3, we introduce the notation of periodogram and study its properties. In section 4, we present Whittle's approximation to the Gaussian likelihood.

3.1 Fourier Analysis

3.1.1 Fourier Transform

A Fourier analysis of a spatial process, also called a harmonic analysis, is a decomposition of the process into sinusoidal components (sines and cosines waves). The coefficients of these sinusoidal components are the Fourier transform of the process.

Suppose that $g()$ is a real or complex valued function, define

$$G(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} g(\boldsymbol{s}) \exp\{i\boldsymbol{\omega}^t \boldsymbol{s}\} d\boldsymbol{s}. \tag{3.1}$$

The function G in (3.1) is said to be the Fourier transform of g. Then g has the representation

$$g(\boldsymbol{s}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} G(\boldsymbol{\omega}) \exp\{-i\boldsymbol{\omega}^t \boldsymbol{s}\} d\boldsymbol{\omega}. \tag{3.2}$$

So that $|G(\boldsymbol{\omega})|$ represents the amplitude associated with the complex exponential with frequency $\boldsymbol{\omega}$. The right hand side of Equation (3.2) is called the Fourier integral representation of g. The function g and G are said to be a Fourier transform pair.

It is often useful to think of functions and their transforms as occupying two domains. These domains are usually referred as time (or space) domain and frequency (or spectral) domain respectively. Operations performed in one domain have corresponding operations in the other. So we can move between domains so that operations can be performed where they are easiest or most advantageous.

3.1.2 Properties of the Fourier Transform

Scaling Property

If $\mathcal{F}\{f(x)\} = F(s)$ is a real, nonzero constant function, where $\mathcal{F}$ denotes the Fourier transform, then

$$\mathcal{F}\{f(ax)\} = \frac{1}{|a|}F(s/a). \tag{3.3}$$

The scaling property states that if the width of a function is decreased while its height is kept constant, then its Fourier transform becomes wider and shorter. If its width is increased, its transform becomes narrower and taller.

Shifting Property

If $\mathcal{F}\{f(x)\} = F(s)$ and x_0 is a real constant, then

$$\mathcal{F}\{f(x - x_0)\} = F(s)\exp(i2\pi x_0 s). \tag{3.4}$$

From the shifting property, it is evident that the Fourier transform of a shifted function is just the transform of the unshifted function multiplied by an exponential factor having a linear phase.

Convolution Theorem

Suppose that $g(x) = f(x) \bigotimes h(x)$, then given $\mathcal{F}\{g(x)\} = G(s)$, $\mathcal{F}\{f(x)\} = F(s)$, and $\mathcal{F}\{h(x)\} = H(s)$,

$$G(s) = F(s)H(s). \tag{3.5}$$

This extremely powerful result demonstrates that the Fourier transform of a convolution is simply given by the product of the individual transforms; that is

$$\mathcal{F}\{f(x) \bigotimes h(x)\} = F(s)H(s).$$

The reverse is also true, that is

$$\mathcal{F}\{f(x)h(x)\} = F(s) \bigotimes H(s). \tag{3.6}$$

Parseval's Theorem

Parseval's Theorem states that the power of a signal represented by a function $h(t)$ is the same whether computed in signal space or frequency space; that is

$$\int_{-\infty}^{\infty} h^2(t)dt = \int_{-\infty}^{\infty} |H(f)|^2 df. \tag{3.7}$$

3.2 Spectral Analysis of Spatial Processes

Let $Z(\boldsymbol{s})$ be a random spatial process in $\mathbb{R}^2$. If it has a constant mean and $\text{Cov}\{Z(\mathbf{s}+\mathbf{u}), Z(\boldsymbol{s})\} = C(\mathbf{u})$, indicating that the covariance among pairs $\mathbf{u}$ unites apart is the same everywhere in the region, where C is the autocovariance function, $Z(\boldsymbol{s})$ is called weakly stationary. Suppose the autocovariance function satisfies

$$\int_{\mathbb{R}^2} |C(\mathbf{x})| d\mathbf{x} < \infty$$

then the spectral density of the random series $Z(\boldsymbol{s})$ at frequency $\boldsymbol{\omega}$ is defined to be the fourier transform of the autocovariance function

$$f(\boldsymbol{\omega}) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} \exp(-i\mathbf{x}^T\boldsymbol{\omega}) C(\mathbf{x}) d\mathbf{x}. \tag{3.8}$$

$$C(\mathbf{x}) = \int_{\mathbb{R}^2} \exp(i\mathbf{x}^T\boldsymbol{\omega})f(\boldsymbol{\omega})d\mathbf{x}. \tag{3.9}$$

From the relation above, we can see the spectral density of a stationary process is the counterpart of a covariance function in spectral domain. Thus the spatial structure of Z could be analyzed with a spectral approach or equivalently by estimating the autocovariance function.

3.2.1 Spectral Density of a Continuous Process Observed on a Lattice

The spectral density f is closely related to the spectral decomposition of the process Z into a superposition of harmonic oscillations. However, such a decomposition can not be uniquely restored from observations of Z in $\triangle\mathbb{Z}^2$, where $\triangle$ is the distance between neighboring observations, and $\mathbb{Z}^2$ is the integer lattice. Due to aliasing effect, it is impossible to distinguish the harmonic components with frequencies differing by an integer multiple of $2\pi/\triangle$ by observations in the integer lattice $\mathbb{Z}^2$.

If $Z(\boldsymbol{s})$ is observed only at uniformly spaced spatial locations $\triangle$ units apart, the spectrum of observations of the sample sequence $Z(\triangle\mathbf{s})$ is concentrated within the finite frequency band $[-\pi/\triangle, \pi/\triangle]^2$. The spectral density $f_\triangle$ of the process on the lattice can be written in terms of the spectral density f of the continuous process Z as

$$f_\triangle(\boldsymbol{\omega}) = \sum_{Q\in\mathbb{Z}^2} f(\boldsymbol{\omega} + \frac{2\pi Q}{\triangle}). \tag{3.10}$$

for $\boldsymbol{\omega} \in [-\pi/\triangle, \pi/\triangle]^2$.

Figure 3.1: Spectral density of sample at lattice.

From Figure 3.1 we can see the difference between the spectral density of the continuous process and the spectral density of sample at lattice. The spectral density of sample at lattice is a periodic function concentrated within $[-\pi/\triangle, \pi/\triangle]^2$.

3.2.2 Classes of Spectral Densities

Gaussian Model

Gaussian model is a commonly used form for the covariance function of a smooth process.

$$C(h) = \sigma^2 \exp(-\alpha h^2) \tag{3.11}$$

the corresponding spectral density is

$$f(\omega) = \frac{1}{2}\sigma^2(\pi\alpha)^{-1/2} \exp\big(-\omega^2/(4\alpha)\big) \tag{3.12}$$

The parameter σ^2 is the variance of the process. The parameter α^{-1} is a parameter that explains how fast the correlation decays.

Note that C is infinitely differentiable and correspondingly all moments of the spectral density are finite. This type of behavior usually would be considered unrealistic for a physical process.

Spherical Model

One of the most commonly used model for isotropic covariance functions in geological and hydrological applications is the spherical model

$$C(h) = \begin{cases} \sigma^2(1 - \frac{3}{2\rho} + \frac{1}{2\rho^2}h^3) & \text{for} \quad h \le \rho \\ 0 & \text{for} \quad h \ge \rho \end{cases}$$

The parameter σ^2 is the variance of the process. The parameter ρ is called the range and is the distance at which correlations become exactly 0. This function is only once differentiable at $h = \rho$ and this can lead to problems when using likelihood methods for estimating the parameters of this model.

Matérn Class

A class of practical variograms and autocovariance functions for a process Z can be obtained from the Matérn class of spectral densities.

$$f(\boldsymbol{\omega}) = \frac{\sigma^2\Gamma(\nu + d/2)}{\Gamma(\nu)\pi^{d/2}\rho^{2\nu}(\rho^{-2} + |\boldsymbol{\omega}|^{\nu+d/2})} \tag{3.13}$$

where d is the dimension of the process. Here, the vector of covariance parameter is $\theta = (\sigma^2, \rho, \nu)$. The parameter σ^2 is the variance of the process. The parameter ρ measures how the correlation decays with distance, and it is generally called range. The parameter ν measures the degree of smoothness of the process, the higher the value of ν the smoother the process would be.

The corresponding covariance is

$$C(h) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(h/\rho)^{\nu} K_{\nu}(h/\rho) \tag{3.14}$$

where K_{ν} is a modified Bessel function.

3.3 Periodogram

The spatial periodogram is a nonparametric estimate of the spectral density, which is the fourier transform of the covariance function. The periodogram is a powerful tool to study the properties of stationary process observed on a d dimensional lattice.

If Z is a stationary process observed at N equally spaced locations in $D(n_1 \times n_2)$, where $N = n_1 n_2$, and the spacing between neighboring observation is $\triangle$, then the periodogram is defined as

$$I_N(\boldsymbol{\omega}) = \triangle^2 (2\pi)^{-2} (n_1 n_2)^{-1} \left| \sum_{s_1=1}^{n_1} \sum_{s_2=1}^{n_2} Z(\triangle \mathbf{s}) \exp\{-i \triangle \boldsymbol{s}^T \boldsymbol{\omega}\} \right|^2 \tag{3.15}$$

It is the modulus square of a finite Fourier transform for the observed region of the random field. In practice, the periodogram estimate for $\boldsymbol{\omega}$ is computed in the set of Fourier frequencies $2\pi\mathbf{f}/\triangle\mathbf{n}$, where $\mathbf{f}/\mathbf{n} = (\frac{f_1}{n_1}, \frac{f_2}{n_2})$, and $\mathbf{f} \in J_N$, for

$$J_N = \{-[(n_1-1)/2], \ldots, [n_1/2]\} \times \{-[(n_2-1)/2], \ldots, [n_2/2]\}$$

and $[x]$ denotes the integer part of x.

The expected value of the periodogram at $\boldsymbol{\omega}_0$ is given by

$$EI_N(\boldsymbol{\omega}_0) = \triangle^2(2\pi)^{-2}(n_1n_2)^{-1}\int_{\Pi^2_\triangle} f_\triangle(\boldsymbol{\omega})W(\boldsymbol{\omega}-\boldsymbol{\omega}_0)\mathrm{d}\boldsymbol{\omega} \qquad (3.16)$$

where $f_\triangle$ is the spectral density of the sampled sequence $Z(\triangle\mathbf{x})$, $\Pi^2_\triangle = (-\pi/\triangle, \pi/\triangle)^2$, and

$$W(\boldsymbol{\omega}) = \prod_{j=1}^{2}\frac{[\sin(\frac{n_j\omega_j}{2})/n_j]^2}{\sin^2(\frac{\omega_j}{2})}$$

for $\boldsymbol{\omega} = (\omega_1, \omega_2) = 2\pi\mathbf{f}/\mathbf{n}$. The side lobes (subsidiary peaks) of the function W can lead to substantial bias in $I_N(\boldsymbol{\omega}_0)$ as an estimate of $f_\triangle(\boldsymbol{\omega}_0)$ since they allow the value of $f_\triangle$ at frequencies far from $\boldsymbol{\omega}_0$ to contribute to the expected value. Figure 3.2 shows a graph of W in the horizonal direction ($\omega_2 = 0$). We can clearly see the main lobe around the origin and some smaller side lobes at other frequencies.

If the side lobes of W were substantially smaller, we could reduce this source of bias for the periodogram considerably. Tapering is a technique that effectively reduces the side lobes associated with the spectral window W. We form the product $h(\boldsymbol{x})Z(\boldsymbol{x})$ for each value of $\boldsymbol{x} = (x_1, x_2)$, where $h(\boldsymbol{x})$ is a sequence of real valued constants called data taper, and then we compute the periodogram for the tapered data. Tapering technique will be discussed in next chapter.

Asymptotic properties of the periodogram

Consider a Gaussian stationary process Z with mean 0 and spectral density $f(\boldsymbol{\omega})$ on

Figure 3.2: Spectral window along the horizontal axis. The X axis shows the frequencies, while the Y axis shows the spectral window for the periodogram at fourier frequencies $(\frac{2\pi f_1}{n_1}, 0)$.

a lattice D. We assume Z is observed at N equally spaced locations in $D(n_1 \times n_2)$, where $N = n_1 n_2$, and the spacing between observations is $\triangle$. We define the periodogram function, $I_N(\boldsymbol{\omega})$, as in (3.15).

Assume $n_1 \to \infty, n_2 \to \infty$ and $n_1/n_2 \to a$ for a constant $a > 0$. Then we get:

1. The expected value of the periodogram, $I_N(\boldsymbol{\omega})$, is asymptotically $f_{\triangle}(\boldsymbol{\omega})$.

2. The asymptotic variance of $I_N(\boldsymbol{\omega})$ is $f_{\triangle}^2(\boldsymbol{\omega})$.

3. The periodogram values $I_N(\boldsymbol{\omega})$, and $I_N(\boldsymbol{\omega}')$ for $\boldsymbol{\omega} \neq \boldsymbol{\omega}'$, are asymptotically independent.

Brillinger (1981) gives a detailed proof of the above properties of periodogram.

In the time-space domain, empirical variogram estimates are most commonly used to estimate the correlation structure of a process. When a parametric variogram model is fit to empirical variogram estimates, frequently used techniques such as non-linear least squares or restricted maximum likelihood (REML) approaches generally do not take into account any correlation between estimated variogram values. The same data points are used to estimate the variogram at different lags, and the resulting variogram estimates are more correlated than the observations of the underlying process. There is not much know about the correlation structure of the variogram estimates and ignoring such correlation can mislead data analysis. One the other hand , in the spectral domain, part three of the asymptotic properties of the periodogram allow us to easily fit a parametric model to the periodogram models.

Asymptotic distribution of the periodogram

If the process Z is Gaussian, then the periodogram has asymptotically a distribution that is multiple of a χ_2^2. More specifically, the periodogram $I_N(\boldsymbol{\omega}_j)$ where $\boldsymbol{\omega}_j$ is a Fourier frequency has asymptotically a $f(\boldsymbol{\omega}_j)\chi_2^2/2$ distribution.

3.4 Whittle's Likelihood Approximation

For large data sets, calculating the determinants that we have in the likelihood function can be often infeasible. Spectral methods based on Whittle's (1954) likelihood approximation can be used to obtain the maximum likelihood estimates (MLE) of the covariance parameters $\theta = (\theta_1, \ldots, \theta_r)$.

Whittle's likelihood approximation to the Gaussian negative log likelihood is

$$\frac{N}{(2\pi)^2} \int_{[-\pi,\pi]^2} (\log f_{\triangle}(\boldsymbol{\omega}) + I_N(\boldsymbol{\omega}) f_{\triangle}(\boldsymbol{\omega})^{-1}) d\boldsymbol{\omega} \tag{3.17}$$

where $I_N(\boldsymbol{\omega})$ is the periodogram and $f_{\triangle}(\boldsymbol{\omega})$ is the spectral density of the process at lattice. This approximated likelihood can be calculated efficiently by using the fast fourier transform. This approximation requires only $O(nlog_2 n)$ operations and does not involve calculate the determinants. Simulation studies seem to indicate that N needs to be at least 100 to get good estimated MLE parameters using Whittle's approximation.

In practice, the above integral can be replaced by

$$\frac{N}{(2\pi)^2} \sum (\log f_{\triangle}(2\pi \mathbf{f}/\triangle \mathbf{N}) + I_N(2\pi \mathbf{f}/\triangle \mathbf{n}) f_{\triangle}(2\pi \mathbf{f}/\triangle \mathbf{n})^{-1}) \tag{3.18}$$

where $\mathbf{f}/\mathbf{n} = (\frac{f_1}{n_1}, \frac{f_2}{n_2})$, and $\mathbf{f} \in J_N$, for

$$J_N = \{-[(n_1-1)/2], \ldots, [n_1/2]\} \times \{-[(n_2-1)/2], \ldots, [n_2/2]\}$$

and $[x]$ denotes the integer part of x.

The asymptotic covariance matrix of the MLE estimates of $\theta_1, \ldots, \theta_r$ is

$$\left\{ \frac{2}{N} \left[\frac{1}{4\pi^2} \int_{[-\pi,\pi]} \int_{[-\pi,\pi]} \frac{\partial \log f(\boldsymbol{\omega})}{\partial \theta_j} \frac{\partial \log f(\boldsymbol{\omega})}{\partial \theta_k} d\boldsymbol{\omega} \right]^{-1} \right\}_{jk} \tag{3.19}$$

this is much easier to compute than the inverse of the Fisher information matrix.

Chapter 4

New Data Taper

The periodogram, I_N defined in Chapter 3 is also the discrete Fourier transform of the biased estimate of the covariance, the sample covariance $c_N(\mathbf{k}) = N^{-1}\sum(Z_{\boldsymbol{s}} - \bar{Z})(Z_{\mathbf{s+k}} - \bar{Z})$, rather than the unbiased estimate of covariance, where N is the total number of observations, $Z_{\boldsymbol{s}}$ and $Z_{\mathbf{s+k}}$ are observations at locations $\boldsymbol{s}$ and $\mathbf{s}+\mathbf{k}$, and $\bar{Z}$ is the sample mean. No matter how large N becomes, the periodogram always involves the tail of the sample covariance, which is a poor estimate of the corresponding theoretical covariance. One of the advantages of tapering is the reduction of the bias due to the boundary effect on the sample covariance. And we generally need more tapering for the corner observations. Here we propose a new data taper, a rounded data taper, that gives more tapering to the observations at four corners. Therefore, with less overall tapering, we get the amount of smoothing that we need without losing so much information.

4.1 Multiplicative Data Taper

Let w be a continuous increasing function with $w(0) = 0$ and $w(1) = 1$, we define a one dimensional taper $g(u)$, $u \in [0, 1]$, with smoothness parameter ρ, by

$$g(u) = \begin{cases} w(2u/\rho) & \text{for} \quad 0 \leq u \leq \frac{1}{2}\rho, \\ 1 & \text{for} \quad \frac{1}{2}\rho \leq u \leq \frac{1}{2}, \\ g(1-u) & \text{for} \quad \frac{1}{2} < u \leq 1, \end{cases}$$

The typical shape of a tapering function, $g(u)$ involves a steady increase from 0 to a maximum of 1, followed by a steady decrease to 0.

There are different types of taper functions. A common taper function is the Tukey-Hanning taper with $w(u) = \frac{1}{2}\{1 - \cos(u\pi)\}$.

In spatial statistics, data tapers are often the tensor product of two one-dimensional tapers, $h_M(\mathbf{s}/N) = g_1(s_1/n_1)g_2(s_2/n_2)$ where $\mathbf{s}/N = (s_1/n_1, s_2/n_2)$ and $N = n_1n_2$. In two dimensions, the taper function takes on the form of a rectangle centered in the middle of the data grid, then decreases smoothly to 0 at the edges of the grid, so it is called the multiplicative data taper.

Figures 4.1 shows an example of multiplicative data taper. It gives weight 1 to the observations inside the middle rectangle region and weight $h_M(\mathbf{s}/N)$ that goes smoothly from 1 to 0 to the outside of the rectangular region.

Figure 4.1: Multiplicative taper on a 12×10 grid with smoothing parameter 4 and 2.

4.2 Rounded Data Taper

To define the rounded data taper, we first define new coordinates (r_1, r_2) in terms of $(s_1, s_2) \in [0, n_1] \times [0, n_1]$,

$$\begin{aligned}
r_1 &= s_1 - (n_1 - 1)/2 \\
r_2 &= s_2 - (n_2 - 1)/2 \\
d &= \sqrt{[r_1 - (n_1/2 - \epsilon)]^2 + [r_2 - (n_2/2 - \epsilon)]^2} \\
S &= \{(r_1, r_2), s.t. \quad |r_1| > (n_1/2 - \epsilon) \quad \text{and} \quad |r_2| > (n_2/2 - \epsilon)\}
\end{aligned}$$

We now give a partition of the grid in terms of the new coordinates (r_1, r_2),

$$
\begin{aligned}
A &= \{(r_1, r_2), \quad s.t. \quad |r_1| \le (n_1/2 - \epsilon) \quad \text{and} \quad |r_2| > (n_2/2 - \delta)\} \\
B &= \{(r_1, r_2), \quad s.t. \quad |r_2| \le (n_1/2 - \epsilon) \quad \text{and} \quad |r_1| > (n_2/2 - \epsilon)\} \\
C &= \{(r_1, r_2), \quad s.t. \quad |r_1| \le (n_1/2 - \delta) \quad \text{and} \quad |r_2| < (n_2/2 - \epsilon), \\
&\quad \text{or} \quad |r_1| \le (n_1/2 - \epsilon) \quad \text{and} \quad |r_2| < (n_2/2 - \delta)\} \\
D &= \{(r_1, r_2), \quad s.t. \quad (r_1, r_2) \in S \quad \text{and} \quad d \in [\epsilon - \delta, \epsilon]\} \\
E &= \{(r_1, r_2), \quad s.t. \quad (r_1, r_2) \in S \quad \text{and} \quad d \ge \epsilon\} \\
F &= \{(r_1, r_2), \quad s.t. \quad (r_1, r_2) \in S \quad \text{and} \quad d \le \epsilon - \delta\}
\end{aligned}
$$

Finally we present the weight function, $h_R()$, that defines the rounded data taper,

$$
h_R(r_1, r_2) = \begin{cases}
\frac{1}{2}\{1 - \cos(\frac{\pi(n_2/2 - r_2)}{\delta})\} & \text{for} \quad (r_1, r_2) \in A, \\
\frac{1}{2}\{1 - \cos(\frac{\pi(n_1/2 - r_1)}{\delta})\} & \text{for} \quad (r_1, r_2) \in B, \\
1 & \text{for} \quad (r_1, r_2) \in C \quad \text{or} \quad F, \\
\frac{1}{2}\{1 - \cos(\frac{\pi(\epsilon - d)}{\delta})\} & \text{for} \quad (r_1, r_2) \in D, \\
0 & \text{for} \quad (r_1, r_2) \in E,
\end{cases}
$$

Figures 4.2 shows an example of rounded data taper. The weight in the region F and C is 1, the weight in the region E is 0, and the weight in the region A, B, and D is h_R that goes smoothly from 1 to 0. The parameter ϵ is the radius of the circle that defines the region D, and $\epsilon - \delta$ is the radius of the circle that defines the region F.

Figure 4.2: Rounded taper on a 12×10 grid with smoothing parameter $\epsilon = 4$ and $\delta = 2$.

4.3 The Choice of Taper Parameters

Consider a Gaussian stationary process Z with spectral density $f(\boldsymbol{\omega})$ on a lattice D. We assume Z is observed at N equally spaced locations in $D(n_1 \times n_2)$, where $N = n_1 n_2$ and the spacing between neighboring observation is $\triangle$. $E(Z(\boldsymbol{s})) = c_z, \mathrm{Cov}(Z(\mathbf{s}+\mathbf{u}), Z(\boldsymbol{s})) = c_{zz}(\mathbf{u})$. Suppose $\sum_{\mathbf{u}} c_{zz}(\mathbf{u}) < \infty$, after the data are filtered with data taper, the spatial periodogram for tapered data is given as:

$$I^{(T)}(\boldsymbol{\omega}) = \triangle^2(2\pi)^{-2}\Big(\sum_{\boldsymbol{s}} h(\triangle \boldsymbol{s}/N)^2\Big)^{-1} \left| \sum_{\boldsymbol{s}} Z(\triangle \boldsymbol{s}) h(\triangle \boldsymbol{s}/N) \exp\{-i\triangle \boldsymbol{s}^T \boldsymbol{\omega}\} \right|^2 \tag{4.1}$$

We refer to this modified periodogram obtained from the tapered data as the tapered spatial periodogram. Here we provide the expectation and the covariance (variance) of the spatial tapered periodogram to study the effect of data taper so that we can estimate the spectral density.

The expectation of the tapered spatial variogram $I^{(T)}(\boldsymbol{\omega})$ is

$$\begin{aligned} &EI^{(T)}(\boldsymbol{\omega}) \\ &= \triangle^2(2\pi)^{-2}(\sum_{\boldsymbol{s}} h(\triangle\boldsymbol{s}/N)^2)^{-1} \int_{\Pi_\triangle^2} |\sum_{\boldsymbol{s}} h(\triangle\boldsymbol{s}/N)\exp\{-i\triangle\boldsymbol{s}^T\boldsymbol{\omega}\}|^2 f_\triangle(\boldsymbol{\omega}-\mathbf{a})\mathrm{d}\mathbf{a} + \\ &\quad \triangle^2(2\pi)^{-2}(\sum_{\boldsymbol{s}} h(\triangle\boldsymbol{s}/N)^2)^{-1}|\sum_{\boldsymbol{s}} h(\triangle\boldsymbol{s}/N)\exp\{-i\triangle\boldsymbol{s}^T\boldsymbol{\omega}\}|^2 c_z^2 \end{aligned} \tag{4.2}$$

In the case the mean $c_z = 0$ and $\triangle = 1$,

$$EI^{(T)}(\boldsymbol{\omega}) = (2\pi)^{-2}(\sum_{\boldsymbol{s}} h(\boldsymbol{s}/N)^2)^{-1} \int_{\Pi^2} |\sum_{\boldsymbol{s}} h(\boldsymbol{s}/N)\exp\{-i\boldsymbol{s}^T\boldsymbol{\omega}\}|^2 f_\triangle(\boldsymbol{\omega}-\mathbf{a})\mathrm{d}\mathbf{a} \tag{4.3}$$

Now we can see the spectral window of periodogram after data taper is

$$W^T(\boldsymbol{\omega}) = \frac{|\sum_{\boldsymbol{s}} h(\boldsymbol{s}/N)\exp\{-i\boldsymbol{s}^T\boldsymbol{\omega}\}|^2}{\sum_{\boldsymbol{s}} h(\boldsymbol{s}/N)^2} \tag{4.4}$$

The expectation (4.3) is a weighted average of the spectral density with weight concentrated in the neighborhood of $\boldsymbol{\omega}$ and relative weight determined by the taper function. If $f_\triangle(\boldsymbol{\omega})$ has a substantial peak for α in the neighborhood of $\boldsymbol{\omega}$, then the expected value of periodogram, given by (4.3) can differ quite substantially from $f_\triangle(\boldsymbol{\omega})$. The advantage of employing a taper is now apparent. It can be taken to have a shape to reduce the effect of neighboring peak.

$$\text{Cov}(I^{(T)}(\omega_1), I^{(T)}(\omega_2)) =$$
$$(\sum_{\boldsymbol{s}} h(\boldsymbol{s}/N)^2)^{-2} \left\{ |H_2(\omega_1 + \omega_2)|^2 + |H_2(\omega_1 - \omega_2)|^2 \right\} f_{\triangle}^2(\boldsymbol{\omega}) + O(N^{-1})$$

where $H_2(\boldsymbol{\omega}) = \sum_{\boldsymbol{s}} h(\boldsymbol{s}/N)^2 \exp\{-i\boldsymbol{s}^T\boldsymbol{\omega}\}$

The mean square error (MSE) of the tapered periodogram at a particular frequency $\boldsymbol{\omega}$ is

$$\begin{aligned} \text{MSE}(I^{(T)}(\boldsymbol{\omega})) &= \text{Var}(I^{(T)}(\boldsymbol{\omega})) + \text{bias}^2(I^{(T)}(\boldsymbol{\omega})) \\ &= (\sum_{\boldsymbol{s}} h(\boldsymbol{s}/N)^2)^{-2} \Big\{ 1 + \Big| \sum_{\boldsymbol{s}} h(\boldsymbol{s}/N)^2 \exp\{-2i\boldsymbol{s}^T\boldsymbol{\omega}\} \Big|^2 \Big\} f_{\triangle}^2(\boldsymbol{\omega}) \\ &\quad + \Big(EI^{(T)}(\boldsymbol{\omega}) - f_{\triangle}(\boldsymbol{\omega}) \Big)^2 \end{aligned} \tag{4.5}$$

The relative MSE of the periodogram is

$$\text{MSE.rela}(I^{(T)}(\boldsymbol{\omega})) = \text{MSE}(I^{(T)}(\boldsymbol{\omega}))/(f_{\triangle}(\boldsymbol{\omega})) \tag{4.6}$$

Notice that MSE of the tapered periodogram depends on the spectral density observed on lattice $f_{\triangle}(\boldsymbol{\omega})$ which we don't have here, we propose using a plug-in method for the spectral density. The plug-in method provides an easy computation which relies directly on the data to find the effect of tapering in estimating the spectral density. Since the spectral density of the sample sequence $Z(\triangle\boldsymbol{s}), \boldsymbol{s} \in \mathbb{Z}^2$ can be expressed as $f_{\triangle}(\boldsymbol{\omega}) = \frac{\triangle^2}{(2\pi)^2} \sum_{\boldsymbol{s}\in\mathbb{Z}^2} \exp(-i\triangle\boldsymbol{s}^T\boldsymbol{\omega}) C_z(\triangle\boldsymbol{s})$, the spectral density of the sample sequence $Z(\triangle\boldsymbol{s})$ can be estimated by periodogram, the discrete Fourier transform of the sample covariance of the data.

Choosing appropriate value for the smoothing parameters (ϵ, δ) will enable us to summarize the spatial structure of the spatial process. The estimates of the parameters are chosen to minimize the summation of relative MSE of the periodogram over all Fourier frequencies.

4.4 The Impact of Data Taper on Periodogram

In this section, we study the impact of rounded taper on periodogram through a example. We simulate a regular complete lattice (40×40). The simulated process of interest Z is a stationary Gaussian process with an Matérn Covariance function C.

$$C(h) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(h/\rho)^{\nu} K_{\nu}(h/\rho)$$

the sill parameter σ^2 is 1, the range ρ is 3 and the smoothing parameter ν is 2.

We use rounded taper and multiplicative taper to filter the raw data and calculate the tapered periodogram separately. Our strategy is to construct 100 pairs of taper parameters $(1,1),(1,2),\ldots,(1,10),(2,1),(2,2),\ldots,(2,10),\ldots,(10,1),(10,2),\ldots,(10,10)$ and calculate the summation of relative MSE of the periodogram over all Fourier frequencies for each pair of taper parameters. We also calculate the taper percentage, the percent of the data being altered by the taper function. The higher the taper percentage, the more information we lose. So within the taper parameters giving the smallest range of relative MSE , we choose a pair that has the smallest taper percentage and set $(m_1, m_2) = (5,5)$ and $(\epsilon, \delta) = (5,3)$.

Figure 4.3 shows the spectral windows on a decibel scale along the vertical axis ($\omega_1 = 0$). Figure 4.4 shows the spectral windows on a decibel scale along the horizonal axis ($\omega_2 = 0$). Note that the spectral windows corresponding to the two data tapers have significantly smaller side lobes than W without taper. The spectral window for the rounded taper, at almost all frequencies along the horizontal axis, is much smaller than the spectral window for the multiplicative taper.

Figure 4.5 is a plot of the periodogram after applying the rounded data taper. Figure 4.6 is a plot of the periodogram after applying multiplicative taper. The open circles indicate the periodogram and the solid line indicates the spectral density of sample at lattice.

From these two plots we can easily see that at higher frequencies, the periodogram after applying rounded taper is a better estimate for the spectral density than the periodogram after applying multiplicative taper. This is consistent with the results shown in the plots of the spectral windows.

We also calculated the mean of the absolute distance (MSD) between the periodogram and the spectral density. The MSD of using rounded taper is 1.2 and the MSD of using multiplicative taper is 4.8. This also suggests that the rounded taper seems to do better in reducing the bias of the periodogram than the multiplicative taper.

In summary, tapering is an operation that replaces the spectral window of the periodogram with one having smaller side lobes, which is a useful method for reducing the bias due to the leakage phenomenon of spectral estimators. In practice, the

Figure 4.3: Spectral windows along the vertical axis.

Figure 4.4: Spectral windows along the horizontal axis.

Figure 4.5: Periodogram after rounded taper at frequencies $(0, \omega)$.

rounded data taper, especially in the border of the region, seems to be a reasonable choice, given an adequate amount of smoothing, without losing so much information.

Figure 4.6: Periodogram after multiplicative taper at frequencies $(0, \omega)$.

Chapter 5

Spectral Likelihood Approximation

5.1 Whittle's likelihood approximation

Whittle's likelihood approximation to the Gaussian negative log likelihood is

$$\frac{N}{(2\pi)^2}\sum(logf_{\triangle}(\boldsymbol{\omega}) + I_N(\boldsymbol{\omega})f_{\triangle}(\boldsymbol{\omega})^{-1}) \tag{5.1}$$

where the sum is evaluated at the Fourier frequencies, $I_N(\boldsymbol{\omega})$ is the periodogram and $f_{\triangle}(\boldsymbol{\omega})$ is the spectral density of the process at lattice. This approximated likelihood can be calculated efficiently by using the fast fourier transform (FFT). This approximation requires only $O(nlog_2 n)$ operations, where N needs to be at least 100 to get good estimated MLE parameters using Whittle's approximation.

The asymptotic covariance matrix of the MLE estimates of $\theta_1, \ldots, \theta_r$ is

$$\Big\{\frac{2}{N}\Big[\frac{1}{4\pi^2}\int_{[-\pi,\pi]}\int_{[-\pi,\pi]}\frac{\partial \log f(\boldsymbol{\omega})}{\partial\theta_j}\frac{\partial \log f(\boldsymbol{\omega})}{\partial\theta_k}d\boldsymbol{\omega}\Big]^{-1}\Big\}_{jk} \tag{5.2}$$

where $\boldsymbol{\omega} = (\omega_1, \omega_2)$. This is much easier to compute than the inverse of the Fisher information matrix.

Guyon (1982) proved that when the periodogram is used to approximate the spectral density in the Whittle's likelihood function, the periodogram bias is non-negligible for 2-dimensional processes, and for 3-dimensions the bias dominates the mean squared error (MSE) of the parameter estimates. Moreover, Stein (1995) showed that using fixed domain asymptotics, in which the observations get increasingly dense in some fixed region as their number increase, the periodogram can yield highly misleading results. Thus, the covariance parameter estimates based on the Whittle likelihood function are only consistent in one dimension, but not in two and higher dimensions.

Direct filtering of the raw data with a data taper can provide a consistent estimate of the covariance parameters. We use the new data taper, rounded taper, to filter the raw data before computing the periodogram then we use the tapered periodogram to approximate the spectral density using the Whittle's likelihood function. We present theoretical results to prove that under some regularity conditions, the tapered Whittle estimate is asymptotically consistent using increasing domain asymptotics. We also present results from simulation studies to show that for a finite sample the covariance estimates obtained from our spectral likelihood approximation method are more appealing than covariance estimates obtained from other spectral likelihood approximation methods.

5.2 Asymptotic Properties of the Estimated Covariance Parameters

If $X = (X_{\boldsymbol{s}}, \boldsymbol{s} \in Z^2)$ is a zero mean, second order stationary spatial process, observed on a rectangle $P_N = \{1, \dots, n_1\} \times \{1, \dots, n_2\}$ of sample size $N = n_1 n_2$. The spectral measure of the process is defined by the spectral density $f(\boldsymbol{\theta}, \boldsymbol{\omega})$. It is well know that $X_{\boldsymbol{s}}$ has the representation

$$X_{\boldsymbol{s}} = \sum_{\boldsymbol{u} \geq 0} a_{\boldsymbol{u}} \varepsilon_{\boldsymbol{s}-\boldsymbol{u}}, \quad a_0 = 1, \tag{5.3}$$

and $E(\varepsilon_{\boldsymbol{s}}) = 0$, $E(\varepsilon_{\boldsymbol{s}} \varepsilon_{\boldsymbol{t}}) = 0 \, (\boldsymbol{s} \neq \boldsymbol{t})$, $E(\varepsilon_{\boldsymbol{s}}^2) = \sigma^2$ where

$$\sigma^2 = (2\pi)^2 \exp\Big\{(2\pi)^{-2} \int_{T^2} \log f(\boldsymbol{\theta}, \boldsymbol{\omega}) d\boldsymbol{\omega}\Big\}. \tag{5.4}$$

In this section, we study the asymptotic properties of the tapered MLE $(\hat{\boldsymbol{\theta}}_N, \hat{\sigma}_N^2)$. We use two theorems to prove that under regularity conditions, if $(\boldsymbol{\theta}_0, \sigma_0^2)$ is the true value of $(\boldsymbol{\theta}, \sigma^2)$, $\hat{\boldsymbol{\theta}}_N$ is a consistent estimate of $\boldsymbol{\theta}_0$ and $\hat{\sigma}_N^2$ is a consistent estimate of σ_0^2.

A modification of whittle's approximation to the log likelihood is

$$L_N(\boldsymbol{\theta}) = -\frac{1}{2} \frac{N}{(2\pi)^2} \int_{[-\pi,\pi]^2} (\log f_{\boldsymbol{\theta}}(\boldsymbol{\omega}) + I_N^T(\boldsymbol{\omega}) f(\boldsymbol{\theta}, \boldsymbol{\omega})^{-1}) d\boldsymbol{\omega} \tag{5.5}$$

where

$$I_N^T(\boldsymbol{\omega}) = (2\pi)^{-2} \{\sum_{\boldsymbol{s}} h_R(\boldsymbol{s}/N)^2\}^{-1} \Big|\sum_{\boldsymbol{s}} X(\boldsymbol{s}) h(\boldsymbol{s}/N) \exp\{-i\boldsymbol{s}^T \boldsymbol{\omega}\}\Big|^2 \tag{5.6}$$

Since $\sigma^2 = (2\pi)^2 \exp\Big\{(2\pi)^{-2} \int_{T^2} \log f(\boldsymbol{\theta}, \boldsymbol{\omega}) d\boldsymbol{\omega}\Big\}$, $L_N(\boldsymbol{\theta})$ can be written as

$$L_N(\boldsymbol{\theta}) = -\frac{N}{2} \log \frac{\sigma^2}{(2\pi)^2} - D_N(\boldsymbol{x}, \boldsymbol{\theta})/(2\sigma^2) \tag{5.7}$$

with

$$\begin{aligned} D_N(\boldsymbol{x}, \boldsymbol{\theta}) &= \frac{N}{(2\pi)^2} \int_{[-\pi,\pi]^2} I_N^T(\boldsymbol{\omega}) g(\boldsymbol{\theta}, \boldsymbol{\omega})^{-1} d\boldsymbol{\omega} \\ &= N \sum_{\boldsymbol{t} \in P_1(\boldsymbol{t})} \alpha_{\boldsymbol{t}}(\boldsymbol{\theta}) C_t \end{aligned} \tag{5.8}$$

where

$$g(\boldsymbol{\theta}, \boldsymbol{\omega}) = ((2\pi)^2/\sigma^2) f(\boldsymbol{\theta}, \boldsymbol{\omega}), \tag{5.9}$$

$$P_1(\boldsymbol{t}) = \{\boldsymbol{t} : |\boldsymbol{t}| \leq (n_1 - 1, n_2 - 1)\} \tag{5.10}$$

$$C_{\boldsymbol{t}} = (\sum_{\boldsymbol{s}} h(\boldsymbol{s}/N)^2)^{-1} \sum_{\boldsymbol{t},\boldsymbol{s}} Z(\boldsymbol{s}) Z(\boldsymbol{s}+\boldsymbol{t}) h(\boldsymbol{s}/N) h((\boldsymbol{s}+\boldsymbol{t})/N) \tag{5.11}$$

is the sample covariance of the tapered data and

$$\alpha_{\boldsymbol{t}}(\boldsymbol{\theta}) = \frac{1}{(2\pi)^2} \int_{[-\pi,\pi]^2} \exp\{i\boldsymbol{t}^T \boldsymbol{\omega}\} g(\boldsymbol{\theta}, \boldsymbol{\omega})^{-1} d\boldsymbol{\omega}. \tag{5.12}$$

The following assumptions (H) are made throughout.

(H1)$f(\boldsymbol{\theta}, \boldsymbol{\omega})$ is continuous and nonzero; also $X_{\boldsymbol{s}}$ is a linear process and we suppose that ε has fourth cumulant $\kappa_4(\boldsymbol{\theta})$;

(H2)The parameter $\boldsymbol{\theta}, \sigma^2$ lie in a region defined as $0 < \sigma^2 < \infty, \boldsymbol{\theta} \in \Theta$, where Θ is a bounded closed set contained in a set S in 2-dimensional Euclidean space;

(H3) If $\boldsymbol{\theta} \neq \boldsymbol{\theta}'$, $g(\boldsymbol{\theta}, \boldsymbol{\omega}) \neq g(\boldsymbol{\theta}', \boldsymbol{\omega})$;

(H4) If $\boldsymbol{\theta} \in S$, $g(\boldsymbol{\theta}, \boldsymbol{\omega})$ and $\{g(\boldsymbol{\theta}, \boldsymbol{\omega})\}^{-1}$ are continuous function for $\boldsymbol{\omega} \in [-\pi, \pi]^2$. The partial derivative of $\{g(\boldsymbol{\theta}, \boldsymbol{\omega})\}^{-1}$ with respect to the components of $\boldsymbol{\theta}$ are also continuous function for $\boldsymbol{\omega} \in [-\pi, \pi]^2$.

All classical models(e.g Matern covariance model) comply with these assumptions.

5.2.1 Consistency of $\hat{\boldsymbol{\theta}}_N$

Lemma 1: *Let $\boldsymbol{\theta}_0$ be the true value of $\boldsymbol{\theta}$ and let $\boldsymbol{\theta}^*$ be any other point in Θ. Then there is a positive constant $K(\boldsymbol{\theta}_0, \boldsymbol{\theta}^*)$ such that*

$$\lim_{N\to\infty} p\{N^{-1}[D_N(\boldsymbol{\theta}_0) - D_N(\boldsymbol{\theta}^*)] < -K(\boldsymbol{\theta}_0, \boldsymbol{\theta}^*)\} = 1 \tag{5.13}$$

For simplicity, we write $D_N(\boldsymbol{\theta})$ for $D_N(\boldsymbol{x}, \boldsymbol{\theta})$. The proof of Lemma 1 is in Appendix A.

From Lemma 1 it is easily seen that $\hat{\boldsymbol{\theta}}_N$ is consistent when Θ is a finite set. To obtain consistency when Θ is an arbitrary bounded closed set, we require $D_N(\boldsymbol{\theta})$ to satisfy a suitable continuity condition.

Lemma 2: *let $|N^{-1}[D_N(\boldsymbol{\theta}_2) - D_N(\boldsymbol{\theta}_1)]| < H_{\delta,N}(\boldsymbol{x}, \boldsymbol{\theta}_1)$ for all $\boldsymbol{\theta}_1 \in \Theta, \boldsymbol{\theta}_2 \in S$ such that $|\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1| < \boldsymbol{\delta}$ where*

$$\lim_{\boldsymbol{\delta}\to 0} E(H_{\boldsymbol{\delta},N}) = 0 \qquad \textit{uniformly in } N, \tag{5.14}$$

and

$$\lim_{N\to\infty} var(H_{\boldsymbol{\delta},N}) = 0 \qquad \textit{for each } \boldsymbol{\delta}. \tag{5.15}$$

Then $P\lim_{N\to\infty} \hat{\boldsymbol{\theta}}_N = \boldsymbol{\theta}_0$ in P_0 probability.

The proof of Lemma 2 is in Appendix B.

A theorem of Walker (1964) indicates that assumption (H4) implies the above continuity condition and so is sufficient for $\hat{\boldsymbol{\theta}}_N$ to be consistent.

Theorem 1: *If all the assumptions in (H) are satisfied, $\hat{\theta}_N$ is a consistent estimator; that is $\lim \hat{\theta}_N = \theta_0$ in P_0 probability.*

Theorem 1 is a straightforward result from Lemma 1 and Lemma 2.

5.2.2 Consistency of $\hat{\sigma}_N^2$

Theorem 2: *If all the assumptions in (H) are satisfied, $\hat{\sigma}_N^2$ is a consistent estimator; that is* $\lim_{N\to\infty} \hat{\sigma}_N^2 = \sigma_0^2$ *in P_0 probability.*

Proof: Since the $(\hat{\boldsymbol{\theta}}_N, \hat{\sigma}_N^2)$ are MLE estimates of tapered whittle's likelihood function,

$$\frac{\partial}{\partial \sigma^2} L_N(\hat{\boldsymbol{\theta}}_N, \hat{\sigma}_N^2) = -\frac{N}{2\hat{\sigma}_N^2} + \frac{D_N(\hat{\boldsymbol{\theta}}_N)}{2\hat{\sigma}_N^4} = 0$$

That is

$$\hat{\sigma}_N^2 = N^{-1} D_N(\hat{\boldsymbol{\theta}}_N)$$

Follow lemma 1 and Lemma 2, we have

$$\lim_{N\to\infty} N^{-1}[D_N(\hat{\boldsymbol{\theta}}_N) - D_N(\boldsymbol{\theta}_0)] = 0 \quad \text{in } P_0 \text{ probability.}$$

That is

$$\lim_{N\to\infty} \hat{\sigma}_N^2 = \sigma_0^2 \quad \text{in } P_0 \text{ probability.}$$

5.3 Simulation Study

The goal of the simulation study is to investigate the impact of the proposed rounded taper on the spectral likelihood method for finite samples, and to understand the relationship between the performance of spectral likelihood approximation method and the sample size.

5.3.1 The Design of the Simulation Study

The simulation study is a two steps procedure in which we estimate the taper parameters and the covariance parameters separately. In the first step, we obtain the estimates of the taper parameters by minimizing the summation of the relative MSE of the periodogram over all Fourier frequencies. In the second step, we use the estimated taper parameters in the tapered Whittle's spectral likelihood function to estimate the covariance parameters.

We simulate n observations from a Gaussian spatial process on a regular complete lattice. This process has a stationary Matérn Covariance function C

$$C(\boldsymbol{h}) = \tau^2 I(\boldsymbol{h}) + \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\boldsymbol{h}/\rho)^{\nu} K_{\nu}(\boldsymbol{h}/\rho) \tag{5.16}$$

where $I(h)$ is an indicator function, it takes the value 1 when $\boldsymbol{h} = (0, 0)$ and 0 everywhere else. Nugget $\tau^2 = 0.25$, partial sill $\sigma^2 = 1.0$, range $\rho = 2$ and the smoothing parameter $\nu = 3.0$.

The spectral density of the continuous process is

$$f(\boldsymbol{\omega}) = \frac{\pi^{-1}\sigma\nu\rho^{-2\nu}}{(\rho^{-2} + |\boldsymbol{\omega}|^2)^{\nu+1}} \tag{5.17}$$

and the spectral density of the process on the lattice is

$$f_{\triangle}(\boldsymbol{\omega}) = \sum_{Q\in\mathbb{Z}^2} f(\boldsymbol{\omega} + \frac{2\pi Q}{\triangle}). \tag{5.18}$$

Here, the sum is truncated after $2n$ terms.

To investigate the relationship between the performance of our spectral likelihood approximation method and the sample size, we simulate observations from above spatial process on different sample grids with different sample sizes. For this exercise, we consider a 10×10 grid with sample size $n = 100$, a 20×20 grid with sample size $n = 400$, and a 30×30 grid with sample size $n = 900$. The grids are equally spaced with the distance between neighboring locations equal to 1 unit. Under each condition, we generate 50 independent data sets and estimate the parameters for each data set and compute the Monte carlo (MC) mean and MC standard deviation(SD) of the parameters estimates obtained from the 50 simulations of the underlying process. We also calculate the Mean relative absolute error (MRAE) of the parameters to compare different likelihood approximation methods. MRAE is defined as

$$MRAE = \sum_{i=1}^{m} \left| \frac{\theta_i^{(0)} - \hat{\theta}_i}{\theta_i^{(0)}} \right| \tag{5.19}$$

where m is the number of covariance parameters, $\theta_i^{(0)}$ is the true value of parameter θ_i and $\hat{\theta}_i$ is the maximum likelihood estimate of θ_i.

5.3.2 Simulation Results

As stated earlier, the estimation process is a two step procedure. First we use rounded taper and multiplicative taper to filter the raw data and calculate the tapered periodogram. For each of the 50 simulated data sets on grid 10×10, we estimate the taper parameters by minimizing the summation of the relative MSE of the periodogram function (4.6) over all Fourier frequencies and the taper percentage. The estimated taper parameters on this grid is the mean of the 50 estimated taper parameters. We use similar method to obtain the estimated taper parameters for grid

20×20 and 30×30. The results are presented in Table (5.1).

Table 5.1: Estimates of the taper parameters: $(\hat{m}_1, \hat{m}_2)$ is the mean of the estimates of multiplicative taper parameters of 50 data sets and $(\hat{\epsilon}, \hat{\delta})$ is the mean of the estimates of rounded taper parameters of 50 data sets .

Taper parameters	10×10 lattice	20×20 lattice	30×30 lattice
$(\hat{m}_1, \hat{m}_2)$	(1,1)	(2,2)	(3,3)
$(\hat{\epsilon}, \hat{\delta})$	(2,1)	(4,2)	(5,3)

After we obtain the estimation of the taper parameters on three grids, we use these estimates in the tapered Whittle's spectral likelihood function to estimate the covariance parameters. The estimated parameters are the Monte Carlo mean of the parameter estimation from 50 simulations. The number in the bracket is the Monte Carlo standard deviation (SD) from 50 simulations.

Table (5.2) contains MLE results from the simulation on a grid 10×10. Table (5.3) and table (5.4) contain MLE results from the simulations on sample grids 20×20 and 30×30. The results show that both rounded taper and multiplicative taper can improve the performance of the Whittle's spectral likelihood approximation method significantly. The classical likelihood estimates and the tapered whittle's likelihood estimates for τ^2, σ^2 and ρ are close to the true value. But it's not the case for smoothing parameter ν. Classical likelihood method always underestimate ν while the spectral likelihood method always overestimate ν. This is because the smoothness parameter ν is always very difficult to estimate.

Table 5.2: Estimated covariance parameters (mean from the 50 simulations) on sample grid 10×10. The values in parenthesis are the Monte Carlo (MC) standard deviations (SD) from the 50 simulations.

Covariance parameters	τ^2	σ^2	ρ	ν	MRAE
Truth	0.25	1.0	1.0	3.0	
Classical MLE	0.21 (0.21)	0.92 (0.22)	1.12(0.26)	1.03 (2.25)	0.25
Spectral MLE without taper	0.54 (0.34)	2.32 (0.46)	3.63 (0.38)	9.25 (3.94)	1.80
Spectral MLE with Rounded taper	0.18 (0.13)	0.75 (0.11)	1.51 (0.15)	5.15 (3.13)	0.44
Spectral MLE with Multi. taper	0.15 (0.16)	0.73 (0.15)	1.64 (0.18)	6.02 (3.21)	0.60

Table 5.5 contains the covariance estimation of the tapered spectral approximation to the likelihood function on three different grids. We present the mean of the estimated parameters from the 50 simulations, the Monte Carlo standard deviations obtained from the 50 simulations and the mean of the estimated standard deviations using asymptotic results. The result indicates that $n = 100$ is enough to get good estimated MLE parameters using tapered whittle's approximation. However, as the sample size increase, the tapered spectral likelihood approximation method gives better parameter estimations, especially for the smoothness parameter ν. The fact that the Monte Carlo standard deviation and the estimated standard deviation are similar for τ^2, σ^2 and ρ suggests that our estimation of the uncertainty associated to the parameters are also reliable.

Table 5.3: Estimated covariance parameters (mean from the 50 simulations) on sample grid 20×20. The values in parenthesis are the Monte Carlo (MC) standard deviations (SD) from the 50 simulations.

Covariance parameters	τ^2	σ^2	ρ	ν	MRAE
Truth	0.25	1.0	1.0	3.0	
Classical MLE	0.23 (0.20)	0.94 (0.21)	1.12 (0.26)	1.26 (2.67)	0.21
Spectral MLE without taper	0.45 (0.29)	2.25 (0.23)	2.56 (0.35)	8.41 (4.23)	1.35
Spectral MLE with Rounded taper	0.18 (0.12)	0.83 (0.13)	1.43 (0.13)	5.06 (2.92)	0.39
Spectral MLE with Multi. taper	0.17 (0.14)	0.79 (0.16)	1.60 (0.11)	5.43 (2.81)	0.49

Table 5.4: Estimated covariance parameters (mean from the 50 simulations) on sample grid 30×30. The values in parenthesis are the Monte Carlo (MC) standard deviations (SD) from the 50 simulations.

Covariance parameters	τ^2	σ^2	ρ	ν	MRAE
Truth	0.25	1.0	1.0	3.0	
Classical MLE	0.23 (0.16)	0.95 (0.18)	1.11(0.23)	1.45 (2.32)	0.19
Spectral MLE without taper	0.42 (0.33)	2.17 (0.21)	2.39 (0.32)	8.33 (4.06)	1.25
Spectral MLE with Rounded taper	0.21 (0.14)	0.88 (0.15)	1.37 (0.11)	4.63 (3.15)	0.30
Spectral MLE with Multi. taper	0.19 (0.17)	0.82 (0.17)	1.61 (0.16)	4.97 (3.43)	0.42

Table 5.5: Estimated covariance parameters using 50 simulations. We present the mean of the estimated parameters from the 50 simulations, the Monte Carlo standard deviations obtained from the 50 simulations and the mean of the estimated standard deviations using asymptotic results.

parameters	10×10 grid			20×20 grid			30×30 grid		
	MC mean	MC SD	$\hat{SD}$	MC mean	MC SD	$\hat{SD}$	MC mean	MC SD	$\hat{SD}$
τ^2	0.18	0.13	0.11	0.18	0.12	0.10	0.21	0.14	0.15
σ^2	0.75	0.11	0.10	0.83	0.13	0.11	0.88	0.15	0.14
ρ	1.51	0.15	0.11	1.43	0.13	0.12	1.37	0.11	0.12
ν	5.15	3.13	0.21	5.06	2.92	0.20	4.63	3.15	0.18

Chapter 6

Application: Sea Surface Temperature Analysis

In this application, our goal is to estimate the large scale spatial structure of the monthly average satellite SST data. The estimation of spatial parameters is performed through the classical maximum likelihood technique, as well as through the tapered Whittle's likelihood approximation method.

6.1 Data

The satellite SST data set we use to illustrate the impact of tapered Whittle's likelihood approximation method is the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) TA monthly average data. The area of interest covers longitude 149W to 131W and latitude 21N to 39 N. The data set is a monthly average of all data within March, 1998. The resolution of the data set is 0.25 degree longitude by 0.25 degree latitude. Because the Earth is approximately spherical, we need to calculate the great-circle distance matrix for any two sets of longitude/latitude locations. Figure 6.1 is a map of the TMI satellite data plotted at the original longitude and latitude. Figure 6.2 is a map of the TMI satellite data plotted at transformed coordinates(in kilometer).

Figure 6.1: Map of TMI SST data at original coordinates.

Figure 6.2: Map of TMI SST data at transformed coordinates.

6.2 Spatial Analysis

To estimate the covariance parameters using classical maximum likelihood technique, we first calculate the sample variogram of satellite SST data. Since the sample variogram levels out and is curvy all the way up, we consider both the exponential model and the Gaussian model. We use weighted nonlinear least square (WLS) method to fit both models to the sample variogram and estimate the covariance parameters. Since using pairs of points that are independent adds nothing to the sample variogram plot but noise, we usually only use 1/2 to 3/4 of the maximum lag as the maximum distance for the variogram.

Figure 6.3: Theoretical Variogram Curves Fitted by WLS Added to the Sample Variogram of Satellite SST Data.

In Figure 6.3, the open circle is the sample variogram and the black dotdash line is the sample variogram with smaller lags. The red solid line indicates the fitted exponential model and the green dashed line indicates the fitted gaussian model.

From the plot, we can see, both models capture the shape of sample variogram very well. Since the Gaussian model gives much smaller weighted least square value (25740) than the exponential model (71528), we choose to use the Gaussian variogram model. So that the estimated variogram model is of the form

$$\gamma(\boldsymbol{h}) = C_0 + C_1(1 - \mathrm{e}^{-\boldsymbol{h}^2/R^2}) \tag{6.1}$$

where C_0 is the nugget, $C_0 + C_1$ is the sill, C_1 is the partial sill and R is the range.

Figure 6.3 shows that the choice of Gaussian variogram model is adequate. We take this form and use the parameters estimated from WLS as the starting value for the maximum likelihood (ML) method to estimate the covariance parameters. The parameter estimates got from the ML method are $\hat{C}_0 = 0.0590, \hat{C}_1 = 0.1175, \hat{R} = 334$.

To estimate the covariance parameters using rounded tapered Whittle's likelihood method, we need to estimate the taper parameters by minimizing the approximated mean square error(MSE) of the tapered periodogram. Notice that the MSE of the tapered periodogram depends on the spectral density observed on lattice $f_{\triangle}(\boldsymbol{\omega})$ which we don't have here, we propose using a plug-in method for the spectral density. The plug-in method provides an easy computation which relies directly on the data to find the effect of tapering in estimating the spectral density. Since the spectral density of the sample sequence $Z(\triangle \boldsymbol{s}), \boldsymbol{s} \in \mathbb{Z}^2$ can be expressed as $f_{\triangle}(\boldsymbol{\omega}) = \frac{\triangle^2}{(2\pi)^2} \sum_{\boldsymbol{s} \in \mathbb{Z}^2} \exp(-i\triangle \boldsymbol{s}^T \boldsymbol{\omega}) C_z(\triangle \boldsymbol{s})$, the spectral density of the sample sequence $Z(\triangle \boldsymbol{s})$ can be estimated by the discrete Fourier transform of the sample covariance of the data. By using this plug-in method to compute the summation of the relative MSE over all fourier frequencies, we get $\hat{\epsilon} = 400$ and $\hat{\delta} = 200$, so the partition of the grid in terms of the transformed coordinates (r_1, r_2) is,

$$
\begin{aligned}
A &= \{(r_1, r_2), \quad s.t. \quad |r_1| \leq 600 \quad \text{and} \quad |r_2| > 800\} \\
B &= \{(r_1, r_2), \quad s.t. \quad |r_2| \leq 600 \quad \text{and} \quad |r_1| > 600\} \\
C &= \{(r_1, r_2), \quad s.t. \quad |r_1| \leq 800 \quad \text{and} \quad |r_2| < 600, \\
&\quad \text{or} \quad |r_1| \leq 600 \quad \text{and} \quad |r_2| < 800\} \\
D &= \{(r_1, r_2), \quad s.t. \quad (r_1, r_2) \in S \quad \text{and} \quad d \in [200, 400]\} \\
E &= \{(r_1, r_2), \quad s.t. \quad (r_1, r_2) \in S \quad \text{and} \quad d \geq 400\} \\
F &= \{(r_1, r_2), \quad s.t. \quad (r_1, r_2) \in S \quad \text{and} \quad d \leq 200\}
\end{aligned}
$$

where $d = \sqrt{(r_1 - 600)^2 + (r_2 - 600)^2}$ and $S = \{(r_1, r_2), s.t. \quad |r_1| > 600 \text{ and } |r_2| > 600\}$.

The weight function, $h_R()$, that defines the rounded data taper is,

$$
h_R(r_1, r_2) = \begin{cases}
\frac{1}{2}\{1 - \cos(\frac{\pi(1000-r_2)}{200})\} & \text{for} \quad (r_1, r_2) \in A, \\
\frac{1}{2}\{1 - \cos(\frac{\pi(1000-r_1)}{200})\} & \text{for} \quad (r_1, r_2) \in B, \\
1 & \text{for} \quad (r_1, r_2) \in C \quad \text{or} \quad F, \\
\frac{1}{2}\{1 - \cos(\frac{\pi(\epsilon-d)}{200})\} & \text{for} \quad (r_1, r_2) \in D, \\
0 & \text{for} \quad (r_1, r_2) \in E,
\end{cases}
$$

The plot of the rounded taper function are presented in figure 6.4.

After we get the parameter estimation of the taper function, we use the rounded taper to filter the raw SST data and compute the tapered Whittle's likelihood approx-

Figure 6.4: Rounded taper for the TMI SST data.

imation and get the MLE for the covariance parameters. The estimation for nugget, partial sill and range are $\hat{C}_0 = 0.02, \hat{C}_1 = 0.1502, \hat{R} = 210$. The Gaussian variogram model fitted by classical maximum likelihood method and tapered Whittle's likelihood approximation method are plotted in figure 6.5. The estimation of covariance parameters are presented in table 6.1.

It is clear from Figure 6.5 and Table 6.1 that the tapered Whittle's likelihood methods leads to very close estimation of nugget and partial sill to the classical maximum likelihood estimators. Even though those two approximation methods give different range estimation, since range parameter is not very critical in kriging prediction (Stein, 1999), we still conclude that the tapered method does lead to results comparable to the classical case. Since the tapered spectral method is computationally more efficient, it is therefore more advantageous to employ them for large data sets.

Figure 6.5: Gaussian variogram models fitted by classical maximum likelihood method and tapered Whittle's likelihood approximation method.

Table 6.1: Estimated covariance parameters by classical Maximum likelihood methods and tapered Whittle's likelihood approximation method. The value in parenthesis are the estimated standard deviations.

Covariance parameters	Nugget τ^2	Partial sill σ^2	Range ρ
Classical MLE	0.0590 (0.23)	0.1175 (0.56)	334 (0.31)
Tapered Whittle's MLE	0.02 (0.34)	0.1502(0.27)	210 (0.85)

Chapter 7

Conclusion

7.1 Summary

In this thesis, we proposed and studied a new spectral method to approximate the likelihood for spatial processes. This spectral method is based on the Whittle's likelihood approximation method to the Gaussian negative log likelihood.

The key feature of our new spectral likelihood method is the use of new data taper to filter the raw data. Tapering is a highly effective technique to remove edge effects in high dimensional problems. Data tapers are generally used to reduce the bias and variance of the variogram. This new data taper, also called rounded taper, gives more tapering to the corner observations. Therefore, with less overall tapering, we get the amount of smoothing that we need without losing so much information. We prove theoretical results to show that under some regularity conditions the MLE obtained from our spectral method for the covariance parameters are asymptotically consistent.

We studied the impact of data taper by comparing the spectral windows of the periodogram corresponding to two kinds of data tapers, the multiplicative taper and the rounded taper. The side lobes of the spectral window can lead to substantial bias in periodogram as we discussed earlier. The study shows that overall the spectral window for the rounded taper has smaller side lobes than the spectral window for the multiplicative taper. We also illustrated that if we equalize the amount of tapering in the corners, we generally lose more information with a multiplicative taper. So in practice, the rounded data taper, especially in the border of the region, seems to be a reasonable choice, given an adequate amount of smoothing, without losing so much information.

We conducted a simulation study to show the performance of the tapered spectral likelihood approximation for a lattice spatial process. The results show that the MLE from our spectral method is appealing compared the to the MLE obtained using multiplicative tapering and the exact likelihood method.

We used the new spectral approach in the application of sea surface temperature (SST) data to analyze the spatial structure of the monthly averaged TMI satellite SST data. This application showed that the tapered Whittle's likelihood method leads to spatial parameter estimation that is very close to the classical maximum likelihood method. In practice, the tapered spectral likelihood approach is very attractive because of its simplicity and because it is very computationally fast compared to other known likelihood approximation methods for spatial data that give consistent estimates. It is therefore more advantageous to use it with large data sets, where the maximum likelihood method is slow or infeasible to use.

7.2 Future Research

One of the main disadvantages of the use of the tapered Whittle's likelihood approximation method is that the data have to be observed on a complete regular lattice. Parzen(1963), Bloomfield (2000), Neave (1970) have studied the spectral methods for irregular time series in the context of estimating the periodogram of a time series. Fuentes (2006) has introduced spectral methods to approximate the likelihood for large spatial data sets with missing values and irregularly spaced data sets. For incomplete lattice, the first step in her approach is to fill in zero values to the process at the locations in this grid where we have no data then efficiently calculate the periodogram using FFT. For irregular spaced data, assume Z is a continuous Gaussian spatial process of interest observed at M irregularly spaced location with spectral density f_Z. A process Y at location $\boldsymbol{x}$ is defined as the integral of Z in a block of area $\triangle^2$ centered at $\boldsymbol{x}$,

$$Y(\boldsymbol{x}) = \triangle^{-2} \int h(\boldsymbol{x} - \boldsymbol{s}) Z(\boldsymbol{s}) d\boldsymbol{s} \tag{7.1}$$

where for $\boldsymbol{u} = (u_1, u_2)$

$$h(\boldsymbol{u}) = \begin{cases} 1 & \text{if} \quad u_1 < \triangle/2, u_2 < \triangle/2, \\ 0 & \text{otherwise} \end{cases}$$

Then, Y is also a stationary process with spectral density f_Y given by

$$f_Y(\boldsymbol{\omega}) = \triangle^{-2}|\Gamma(\boldsymbol{\omega})|^2 f_Z(\boldsymbol{\omega}) \tag{7.2}$$

where $\boldsymbol{\omega} = (\omega_1, \omega_2)$ and $\Gamma(\boldsymbol{\omega}) = \int h(\boldsymbol{u}) \exp(-i\boldsymbol{\omega}\boldsymbol{u}) = [2\sin(\triangle\omega_1/2)/\omega_1][2\sin(\triangle\omega_2/2)/\omega_2]$

By equation (7.1), $Y(\boldsymbol{x})$ can be treated as a continuous spatial process defined for all $\boldsymbol{x} \in D$. But if we only consider the process Y at a lattice $(n_1 \times n_2)$ with sample size $N = n_1 n_2$, i.e. the values of (x) are the centroids of the N grid cells in the lattice, having spacing $\triangle$ betweeing neighboring sites. Then, we have that the spectral density of the lattice process Y is

$$f_{\triangle,Y}(\boldsymbol{\omega}) = \sum_{Q \in \mathbb{Z}^2} |\Gamma(\boldsymbol{\omega} + 2\pi Q/\triangle)|^2 f_Z(\boldsymbol{\omega} + 2\pi Q/\triangle) \tag{7.3}$$

If we apply Whittle's likelihood to $f_{\triangle,Y}$, written in terms of f_Y, we can obtain MLE for the covariance parameters of Z by writing the likelihood of the process Y.

We can follow the principles outlined in this approach by Fuentes (2006) and study the potential impact of tapered spectral methods on the likelihood approximation, prediction and inference made for spatial data.

We have used the tapered Whittle's likelihood approximation method with satellite sea surface temperature data. Large spatially correlated data sets are emerging from EPA and NOAA resources where there is need for estimating tools that can handle huge data sets. I plan to apply the method proposed in this thesis to those data sets.

Bibliography

[1] Besag, J.E. (1974). Soatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B,* 36, 76-86.

[2] Besag, J.E. and Moran P.A. (1975). On the estimation and testing of spatial interaction in Gaussian lattice processes. *Biometrica* 62, 555-562.

[3] Bloomfield, P. (2000). *Fourier Analysis of Time Series.* Wiley, New York.

[4] Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory.* Expanded edition. Holden-Day, Inc, San Francisco.

[5] Caragea, P. (2003) Approximate likelihoods for spatial processes. Ph.D. Dissertation at UNC.

[6] Constantine, A. G. and Hall, P. (1994) Characterizing surface smoothness via estimation of effective fractal dimension. *Journal of the Royal Statistical Society, Series B,* 56, 96-113.

[7] Cramér, H. and Leadbetter, M. R. (1967). *Stationary and related stochastic processes: Sample function properties and their applications.* Wiley, New York.

[8] Cressie, N. (1993). *Statistics for Spatial Data.* Wiley, New York.

[9] Dalhaus,R. and künsch, H. (1987). Edge effects and efficient parameter estimation for stationay random fields. *Biometrika,* 74, 877-882.

[10] Fuentes, M. (2006). Spectral methods to approximate the likelihood for irregularly spaced spatial data. *Journal of the American Statistical Association,* In press.

[11] Guyon, X. (1982). Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika,* 69, 95-105.

[12] kitanidis. P.K. (1997), *Introduction to Geostatistics: Applications in hydrogeology*, Cambridge University Press, New York.

[13] Mardia, K. V. and Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71, 135-146.

[14] Marcotte, D. (1996). Fast variogram computation with FFT. *Computers and Geosciences,* 22, 1175-1186.

[15] 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.

[16] Neave, H. R. (1970). Spectral analysis of a stationary time series using initially scarce data. *Biometrika*, 57, 111-122.

[17] Pardo-Igúzquiza and Dowd. (1997). AMLE3D: a computer program for the inference of spatial covariance parameters by approximate maximum likelihood estimation. *Computational Geosciences*, 23, 793-805.

[18] Parzen, E. (1963). On spectral analysis with missing observations and amplitude modulation. *Sankhya,Series. A*,25, 383-392.

[19] Priestley, M. B. (1981). *Spectral Analysis and Time Series.* Academic Press, London.

[20] Ripley, B. D. (1981). *Spatial Statistics.* Wiley, New York.

[21] Reynolds, R.W., (1988), A real-time global sea surface temperature analysis. *Journal of Climate*, 1, 75-86.

[22] Reynolds, R.W., (1993), Impact of Mount Pinatubo aerosols on satellite-derived sea surface temperatures.*Journal of Climate*, 6,768-774.

[23] Reynolds, R. W. and D. C. Marsico, (1993), An improved real-time global sea surface temperature analysis. *Journal of Climate*, 6, 114-119.

[24] Reynolds, R. W. and T. M. Smith, (1994) Improved global sea surface temperature analyses. *Journal of Climate*, 7, 929-948.

[25] Reynolds, R. W. and T. M. Smith, (1995) A high resolution global sea surface temperature climatology. *Journal of Climate*, 8, 1571-1583.

[26] Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes and W. Wang, (2002) An improved in situ and satellite SST analysis for climate. *Journal of Climate*, 15, 1609-1625.

[27] Smith, T. M. and R. W. Reynolds, (1998) A high resolution global sea surface temperature climatology for the 1961-90 base period. *Journal of Climate*, 11, 3320-3323.

[28] Stein, M. L. (1987). Minimum norm quadratic estimation of components of spatial variogram. *Journal of the American Statistical Association*, 82, 765-772.

[29] Stein,M. L. (1995). Fixed domain asymptotics for spatial periodograms. *Journal of the American Statistical Association*, 90, 1277-1288.

[30] Stein, M. L. (1999). *Interpolation of Spatial Data: some theory for kriging.* Springer-Verlag, New York.

[31] Stein, M. L., Chi, Z. and Welty, L. J. (2004). Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society, Series B,* 66, 297-312.

[32] Vecchia, A. V. (1988). Estimation and model identification for continuous spatial process. *Journal of the Royal Statistical Society, Series B,* 50, 297-312.

[33] Whittle, P. (1954). On stationary processes in the plane. *Biometrika,* 41, 434-449.

[34] Wu,X., Menzel, P.W., and Wade,G. S., (1999), Estimation of Sea Surface Temperatures Using GOES-8/9 Radiance Measurements. *Bulletin of the American Meteorological Society*, 80, 11271138.

[35] Yaglom, A. M. (1987). *Correlation theory of stationary anf related random functions.* Springer-Verlag, New York.

Appendices

Appendix A

Proof for Lemma 1

Lemma 1:*Let* $\boldsymbol{\theta}_0$ *be the true value of* $\boldsymbol{\theta}$ *and let* $\boldsymbol{\theta}^*$ *be any other point in* Θ. *Then there is a positive constant* $K(\boldsymbol{\theta}_0, \boldsymbol{\theta}^*)$ *such that*

$$\lim_{N\to\infty} p\{N^{-1}[D_N(\boldsymbol{\theta}_0) - D_N(\boldsymbol{\theta}^*)] < -K(\boldsymbol{\theta}_0, \boldsymbol{\theta}^*)\} = 1 \tag{A.1}$$

PROOF:

$$\begin{aligned} D_N(\boldsymbol{x}, \boldsymbol{\theta}) &= \frac{N}{(2\pi)^2} \int_{[-\pi,\pi]^2} I_N^T(\boldsymbol{\omega}) g(\boldsymbol{\theta}, \boldsymbol{\omega})^{-1} d\boldsymbol{\omega} \\ &= N \sum_{\boldsymbol{t} \in P_1(\boldsymbol{t})} \alpha_{\boldsymbol{t}}(\boldsymbol{\theta}) C_t \end{aligned} \tag{A.2}$$

where

$$g(\boldsymbol{\theta}, \boldsymbol{\omega}) = ((2\pi)^2/\sigma^2) f(\boldsymbol{\theta}, \boldsymbol{\omega}), \tag{A.3}$$

$$P_1(\boldsymbol{t}) = \{\boldsymbol{t} : |\boldsymbol{t} = (t_1, t_2)| \leq (n_1 - 1, n_2 - 1)\} \tag{A.4}$$

$$C_{\boldsymbol{t}} = (\sum_{\boldsymbol{s}} h(\boldsymbol{s}/N)^2)^{-1} \sum_{\boldsymbol{t},\boldsymbol{s}} Z(\boldsymbol{s}) Z(\boldsymbol{s} + \boldsymbol{t}) h(\boldsymbol{s}/N) h((\boldsymbol{s} + \boldsymbol{t})/N) \tag{A.5}$$

is the sample covariance of the tapered data and

$$\alpha_{\boldsymbol{t}}(\boldsymbol{\theta}) = \frac{1}{(2\pi)^2}\int_{[-\pi,\pi]^2} \exp\{i\boldsymbol{t}^T\boldsymbol{\omega}\}g(\boldsymbol{\theta},\boldsymbol{\omega})^{-1}d\boldsymbol{\omega}. \tag{A.6}$$

Write

$$Y_N = N^{-1}[D_N(\boldsymbol{\theta}_0) - D_N(\boldsymbol{\theta}^*)] = \sum_{\boldsymbol{t}\in P_1(\boldsymbol{t})} C_t[\alpha_{\boldsymbol{t}}(\boldsymbol{\theta}_0) - \alpha_{\boldsymbol{t}}(\boldsymbol{\theta}^*)] \tag{A.7}$$

Writing f_0, g_0, g^* respectively for $f(\boldsymbol{\theta},\boldsymbol{\omega}_0), g(\boldsymbol{\theta},\boldsymbol{\omega}_0), g(\boldsymbol{\theta},\boldsymbol{\omega}^*)$, then

$$\mathrm{E}(Y_N) = \sum_{\boldsymbol{t}\in P_1(\boldsymbol{t})} \gamma_{\boldsymbol{t}}^{(0)}(\frac{(n_1-t_1)(n_2-t_2)}{n_1n_2})[\alpha_{\boldsymbol{t}}(\boldsymbol{\theta}_0) - \alpha_{\boldsymbol{t}}(\boldsymbol{\theta}^*)] \tag{A.8}$$

with $\gamma_{\boldsymbol{t}}^{(0)} = \mathrm{E}[(\sum_{\boldsymbol{s}} h(\boldsymbol{s}/N)^2)^{-1}\sum_{\boldsymbol{t},\boldsymbol{s}} Z(\boldsymbol{s})Z(\boldsymbol{s}+\boldsymbol{t})h(\boldsymbol{s}/N)h((\boldsymbol{s}+\boldsymbol{t})/N)]$.

By Parseval's theorem we have

$$\begin{aligned}\sum_{\boldsymbol{t}=-\infty}^{\infty} \gamma_{\boldsymbol{t}}^{(0)}\alpha_{\boldsymbol{t}}(\boldsymbol{\theta}_0) &= \frac{(\sum_{\boldsymbol{s}} h(\boldsymbol{s}/N)^2)^{-1}}{(2\pi)^2}\int_{[-\pi,\pi]^2}\frac{(2\pi)^2 f_0}{g_0}d\boldsymbol{\omega} \\ &= (\sum_{\boldsymbol{s}} h(\boldsymbol{s}/N)^2)^{-1}\sigma_0^2\end{aligned} \tag{A.9}$$

and

$$\begin{aligned}\sum_{\boldsymbol{t}=-\infty}^{\infty} \gamma_{\boldsymbol{t}}^{(0)}\alpha_{\boldsymbol{t}}(\boldsymbol{\theta}^*) &= \frac{(\sum_{\boldsymbol{s}} h(\boldsymbol{s}/N)^2)^{-1}}{(2\pi)^2}\int_{[-\pi,\pi]^2}\frac{(2\pi)^2 f_0}{g^*}d\boldsymbol{\omega} \\ &= \frac{(\sum_{\boldsymbol{s}} h(\boldsymbol{s}/N)^2)^{-1}}{(2\pi)^2}\sigma_0^2\int_{[-\pi,\pi]^2}\frac{(2\pi)^2 g_0}{g^*}d\boldsymbol{\omega}\end{aligned} \tag{A.10}$$

Hence

$$\begin{aligned}\lim_{N\to\infty}\mathrm{E}(Y_N) &= \sum_{\boldsymbol{t}=-\infty}^{\infty} \gamma_{\boldsymbol{t}}^{(0)}[\alpha_{\boldsymbol{t}}(\boldsymbol{\theta}_0) - \alpha_{\boldsymbol{t}}(\boldsymbol{\theta}^*)] \\ &= (\sum_{\boldsymbol{s}} h(\boldsymbol{s}/N)^2)^{-1}\sigma_0^2\{1 - \frac{1}{(2\pi)^2}\int_{[-\pi,\pi]^2}(g_0/g^*)d\boldsymbol{\omega}\}\end{aligned} \tag{A.11}$$

Now equation (5.4) may be written

$$\log\frac{\sigma^2}{(2\pi)^2} = \frac{1}{(2\pi)^2}\int_{[-\pi,\pi]^2}\log\Big\{\frac{\sigma^2 g(\boldsymbol{\omega})}{(2\pi)^2}\Big\}d\boldsymbol{\omega} \tag{A.12}$$

i.e.

$$\int_{[-\pi,\pi]^2} \log g(\boldsymbol{\omega}) d\boldsymbol{\omega} = 0.$$

Since the logarithmic function is concave and by assumption (H 3), g_0 and g^* are not equal almost everywhere, we have

$$0 = \frac{1}{(2\pi)^2} \int_{[-\pi,\pi]^2} \log(g_0/g^*) d\boldsymbol{\omega} < \log \Big\{ \frac{1}{(2\pi)^2} \int_{[-\pi,\pi]^2} (g_0/g^*) d\boldsymbol{\omega} \Big\} \tag{A.13}$$

So that $\frac{1}{(2\pi)^2} \int_{[-\pi,\pi]^2} (g_0/g^*) > 1$, From (A.6) and (A.8)

$$\lim_{N\to\infty} E(Y_N) = -\mu(\boldsymbol{\theta}_0, \boldsymbol{\theta}^*), \tag{A.14}$$

where $\mu(\boldsymbol{\theta}_0, \boldsymbol{\theta}^*) > 0$.

Also using the another formula for $D_N(\boldsymbol{\theta})$ given in (5.8)

$$\text{var}Y_N = \text{var}\Big\{ \frac{1}{(2\pi)^2} \int_{[-\pi,\pi]^2} I_N^T(\boldsymbol{\omega})(g_0^{-1} - (g^*)^{-1}) d\boldsymbol{\omega} \Big\} \tag{A.15}$$

varY_N is of order N^{-1} by a theorem due to Grenander and Rosenblatt (1953,1957), then

$$\lim_{N\to\infty} \text{var}Y_N = 0. \tag{A.16}$$

By a simple application of Chebyshev's inequality,

$$\lim_{N\to\infty} p\{N^{-1}[D_N(\boldsymbol{\theta}_0) - D_N(\boldsymbol{\theta}^*)] < -\mu(\boldsymbol{\theta}_0, \boldsymbol{\theta}^*)\} = 1 \tag{A.17}$$

For any constant $K(\boldsymbol{\theta}_0, \boldsymbol{\theta}^*)$ that less than $\mu(\boldsymbol{\theta}_0, \boldsymbol{\theta}^*)$,

$$\lim_{N\to\infty} p\{N^{-1}[D_N(\boldsymbol{\theta}_0) - D_N(\boldsymbol{\theta}^*)] < -K(\boldsymbol{\theta}_0, \boldsymbol{\theta}^*)\} = 1 \tag{A.18}$$

So that Lemma 1 holds.

Appendix B

Proof for Lemma 2

Lemma 2: *Let* $|N^{-1}[D_N(\boldsymbol{\theta}_2) - D_N(\boldsymbol{\theta}_1)]| < H_{\delta,N}(\boldsymbol{x}, \boldsymbol{\theta}_1)$ *for all* $\boldsymbol{\theta}_1 \in \Theta, \boldsymbol{\theta}_2 \in S$ *such that* $|\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1| < \boldsymbol{\delta}$ *where*

$$\lim_{\boldsymbol{\delta}\to 0} E(H_{\boldsymbol{\delta},N}) = 0 \qquad \textit{uniformly in } N, \tag{B.1}$$

and

$$\lim_{N\to\infty} var(H_{\boldsymbol{\delta},N}) = 0 \qquad \textit{for each } \boldsymbol{\delta}. \tag{B.2}$$

Then $P\lim_{N\to\infty} \hat{\boldsymbol{\theta}}_N = \boldsymbol{\theta}_0$ *in* P_0 *probability.*

Suppose $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_0$, by Lemma 1,

$$\lim_{n\to\infty} P\{N^{-1}[U_N(\boldsymbol{\theta}_0) - U_N(\boldsymbol{\theta}_1)]| < -K(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1)\} = 1 \tag{B.3}$$

$(B.1)$ and $(B.2)$ implies that

$$\lim_{n\to\infty} P\{H_{\boldsymbol{\delta},N}(\boldsymbol{x}, \boldsymbol{\theta}_1) < K(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1)\} = 1 \tag{B.4}$$

Since simultaneous occurrence of the events in $(B.3)$ and $(B.4)$ implies

$$N^{-1}[D_N(\boldsymbol{\theta}_0) - D_N(\boldsymbol{\theta}_2)] < 0 \tag{B.5}$$

for $|\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1| < \boldsymbol{\delta}$. We have

$$\lim_{n\to\infty} P\{\sup_{\boldsymbol{\theta}_2 \in N(\boldsymbol{\theta}_1)}[D_N(\boldsymbol{\theta}_0) - D_N(\boldsymbol{\theta}_1)] < 0\} = 1 \tag{B.6}$$

Now the collection of sets $\{N(\boldsymbol{\theta}_1) : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_0\}$ obtained by letting $\boldsymbol{\theta}_1$ run through all points in Θ except $\boldsymbol{\theta}_0$ and the set $\{N(\boldsymbol{\theta}_0) : |\boldsymbol{\theta} - \boldsymbol{\theta}_0| < \boldsymbol{\delta}_0\}$ where $\boldsymbol{\delta}_0$ is arbitrary, together constitute an open covering of Θ, this contains a finite open covering $N(\boldsymbol{\theta}_j) : 0 \leq j \leq m$, then gives

$$\lim_{n\to\infty} P\{\sup_{\boldsymbol{\theta}_2 \in \cup_{j=1}^{m} N(\boldsymbol{\theta}_j)}[U_N(\boldsymbol{\theta}_0) - U_N(\boldsymbol{\theta}_2)] < 0\} = 1 \tag{B.7}$$

or

$$\lim_{n\to\infty} P\{\inf_{\boldsymbol{\theta}\in\Theta} U_N(\boldsymbol{\theta}) = \inf_{\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)} U_N(\boldsymbol{\theta}) < 0\} = 1 \tag{B.8}$$

so that

$$\lim_{n\to\infty} P\{|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|\} = 1 \tag{B.9}$$

Thus since δ_0 can be arbitrarily small

$$p(\lim_{n\to\infty} \hat{\theta}_N) = \theta_0 \tag{B.10}$$