ABSTRACT

Chen, Li. Bayesian Hierarchical Spatial-Temporal Models for Wind Prediction. (Under the direction of Dr. Montserrat Fuentes and Dr. Jerry M. Davis).

Wind fields along coastlines are composed of many features that are spatially and temporally complex in nature, and they are well recognized as nonstationary spatial-temporal processes. The difficulties of spatial-temporal modeling can be overcome by using separable spatial-temporal processes. This subclass of spatial-temporal processes has several advantages, including rapid fitting and simple extensions of many techniques developed and used in time series analysis and classical geostatistics. However, these separable models are not always realistic. An empirical test for separability is proposed to aid in understanding spatial-temporal structure. Also a new class of nonseparable stationary covariance models is introduced. A special case in this new nonseparable covariance class is the Matérn covariance model with a scale parameter. The scale parameter is used to take into account the change of units between the spatial and temporal domain. For nonstationarity, we represent a nonstationary process as a mixture of local orthogonal stationary spatial-temporal processes. Compared to the moving-cylinder spatial-temporal kriging method by Haas, this new nonstationary model is simultaneously defined everywhere. The empirical test for separability can be used to determine the covariance structure of local stationary spatial-temporal processes. We apply the methodology to the wind fields over Chesapeake Bay. The goal of this application is to evaluate the ability of MM5, a mesoscale meteorological model, to forecast the wind fields over the Bay. Simple models are specified for observed wind data and MM5 output in terms of the underlying true wind process which is nonstationary and nonseparable, and estimate them in a Bayesian way. Then the numerical model evaluation consists of comparing the observed wind values with their predictive distributions given the MM5 output. Moreover, the improved wind fields can be simulated via the Bayesian posterior distribution of the underlying true winds.
To my dear family
Biography

Li Chen was born in Tianjin, China on the 26th of September, 1976. She got her Bachelor of Science degree in Information Science from Nankai University in June, 1997. She then continued her graduate studies at Nankai University, and developed an interest in statistics. After she got her Master of Science degree in Probability and Mathematical Statistics in June, 2000, she came to Raleigh and pursued a doctoral degree in statistics at North Carolina State University. She successfully defended her dissertation on May 3, 2004. She has been accepted by the Center of Integrating Statistical and Environmental Science at the University of Chicago as a research associate.
Acknowledgements

I would like to thank the members of my advisory committee for their advice and support during the preparation of my dissertation.

I am extremely grateful for having Dr. Montserrat Fuentes as my advisor. Her enthusiasm, knowledgeable guidance, and constant support, including financial support, have meant a tremendous amount to me. She is not only an excellent researcher who imposes the highest standards on her academic work, but also an outstanding advisor who provides continuing encouragement for her students. I thank her for her encouragement to become an independent researcher, and I also admire her tireless energy and her devotion to her professional career.

I am also extremely grateful to Dr. Jerry M. Davis, my co-advisor, for his endless patience, direction and time. I am very thankful to him for his careful reading of my drafts and for his insightful comments that have helped me to understand the material better. He has not only been helpful in answering my questions but also helped me with his kindness and encouragement in everyday life.

Thanks to Dr. Dave A. Dickey for his advice and encouragement during my work on my dissertation. Thanks to Dr. Sujit K. Ghosh for his valuable comments on the Bayesian methods. Thanks to Dr. Gary M. Lackmann for many enjoyable and helpful discussions on meteorology. Thanks to Dr. John F. Monahan for his helpful questions and suggestions that have helped me to present the material better. I also would like to thank Dr. Pantula, Dr. Swallow and all my professors, and Terry Byron for all their help and encouragement over these past 4 years.

Finally, I would like to thank my parents and my sister for their unceasing support and endless love.
# Contents

List of Figures vii  
List of Tables ix  

1 Introduction 1  
1.1 Motivation 1  
1.2 Description of data 4  
1.3 Objectives 7  

2 Spatial-Temporal Processes 11  
2.1 Introduction of spatial-temporal processes 11  
2.2 Recent approaches to spatial-temporal processes 14  
2.2.1 Maximum likelihood method 15  
2.2.2 Moving window method 16  
2.2.3 Bayesian method 20  
2.2.4 Some models for nonseparable stationary spatial-temporal covariance 23  
2.3 Summary 26  

3 Nonseparable and nonstationary spatial-temporal models 27  
3.1 An empirical test for separability 27  
3.2 A new class of nonseparable stationary covariance models 35  
3.3 New models for nonseparable and nonstationary processes 42  
3.3.1 Mixture of local spectrums (Spectral model) 43  
3.3.2 Mixture of local space-time models (Spatial-temporal domain) 46  
3.4 Spatial-temporal Trend 49  

4 Combining Spatial-Temporal Data 52  
4.1 Statistical models for disparate spatial-temporal data 52  
4.2 Change of support 55  
4.3 Statistical assessment of numerical model performance 56  
4.4 Bayesian spatial-temporal prediction 58
5 Application

5.1 Description of the data ......................... 64
5.2 Exploratory analysis ................................ 65
5.3 Spatial-temporal structure of wind speed ........... 68
  5.3.1 Nonstationarity for wind speed .................. 68
  5.3.2 Empirical test for separability ................. 70
  5.3.3 Spatial-temporal trend .......................... 73
  5.3.4 Spatial-temporal covariance structure .......... 74
5.4 Statistical assessment of MM5 performance .......... 77
5.5 Wind field mapping ................................ 78

6 Summary and Future Work .......................... 92

7 Appendix ............................................ 94

  7.1 The adjustment of wind fields from different height .... 94
  7.2 The function in (3.13) is a valid spectral density .... 95

Bibliography .......................................... 96
List of Figures

1.1  The location of the Chesapeake Bay. ................................. 2
1.2  Location maps. ...................................................... 5

2.1  MWRK. “x” indicates the predict location. ......................... 17
2.2  MCSTK. “x” indicates the predict point. .......................... 18

3.1  The contour plot for separable covariance S1. ...................... 33
3.2  The contour plot for nonseparable covariance N1. .................. 33
3.3  The contour plot for nonseparable covariance N2. .................. 34
3.4  The contour plot for nonseparable covariance N3. .................. 34
3.5  The contour plot for a separable spatial-temporal covariance. .... 38
3.6  The contour plot for a nonseparable spatial-temporal covariance. 40
3.7  Contour plots for some nonseparable spatial-temporal covariances. 41

4.1  The general modeling framework. .................................. 53

5.1  The location map of monitoring sites. ............................... 65
5.2  Grid for MM5 output. .............................................. 66
5.3  Wind field map from MM5 model forecast. ......................... 67
5.4  Subregions of stationarity. .......................................... 69
5.5  The time series of wind speed has similar structure over space within each subregion. ................................. 71
5.6  The spatial process of wind speed has same pattern over time within each subregion except for subregion 4. ................................. 72
5.7  Spatial-temporal trend for wind speed. ................................ 73
5.8  Empirical covariance. ............................................... 75
5.9  Posterior for sill. .................................................. 77
5.10 Posterior for spatial range. .......................................... 78
5.11 Posterior for conversion parameter. .................................. 79
5.12 Posterior for smoothness parameter. ................................ 80
5.13 Bayesian covariance. ............................................... 81
5.14 MM5 model evaluation at 12pm on July 21, 2002. .................. 82
5.15 MM5 model evaluation at 3pm on July 21, 2002. .................. 83
5.16 MM5 model evaluation at 6pm on July 21, 2002. ................. 84
5.17 Original MM5 output wind speed at 12pm on July 21, 2002. .......... 86
5.18 Improved wind speed by combing data at 12pm on July 21, 2002. .... 87
5.19 Original MM5 output wind speed at 3pm on July 21, 2002. ............ 88
5.20 Improved wind speed by combing data at 3pm on July 21, 2002 . . . 89
5.21 Original MM5 output wind speed at 6pm on July 21, 2002. ............ 90
5.22 Improved wind speed by combing data at 6pm on July 21, 2002 . . . 91
List of Tables

1.1 Information of monitoring locations ........................................... 6
1.2 48 variables from MM5 model output. ........................................ 8
1.3 48 variables from MM5 model output (continued). ........................ 9

2.1 Hierarchical model by Wikle, Berliner, and Cressie (1998). ............. 21

3.1 Separable covariance models. .................................................. 31
3.2 Nonseparable covariance models. .......................................... 31

4.1 Priors for the parameters in $\sigma^2(s)$ ...................................... 61

5.1 The posterior distribution for $\epsilon_i$. .................................. 76
5.2 The posterior means for covariance parameters. .......................... 81
5.3 The comparison of MSEs. ..................................................... 85
Chapter 1

Introduction

1.1 Motivation

The location of the Chesapeake Bay is shown by the bold box in Figure 1.1. To forecast the wind fields over the Chesapeake Bay is of long-standing interest, since the wind field in such a location is composed of many features that are spatially and temporally complex in nature, which makes for a very challenging forecasting problem. Despite the challenge, obtaining accurate wind field forecasts in the coastal zone is also critical to recreational interests, e.g., sailing, fishing and wind surfing; commercial interests, e.g., shipping and aviation; and environmental problems such as coastal erosion impacts. Increased emphasis on homeland security issues provides an additional incentive for developing accurate forecasts of the coastal wind fields; forecast of chemical or biological emissions is critically dependent upon accurate prediction of near-surface wind speed and direction.

The wind field in coastal regions is influenced by complex physical processes associated with small scale variations in terrain, and strong gradients of pressure, temperature, moisture and surface roughness. The fundamental diurnal changes in the structure of the atmospheric boundary layer differ greatly from the land surface to the water surface. In part this is due to the fact that radiant energy arriving at the water surface is distributed over a large volume of water, while over land this energy is concentrated at the earth’s surface. In
Figure 1.1: The location of the Chesapeake Bay.
addition, over the water surface persistent evaporation acts to lower the temperature of the air over the water surface in comparison to that over the land surface. Differences in the specific heats of water and soil also play a role, especially during the cooling cycle which begins in the evening hours. During the day, the land heats more quickly than the adjacent water, and the intensive heating of the air above produces a shallow thermal low. The air over the water remains cooler than the air over the land; hence, a shallow thermal high exists above the water. The overall effect of this pressure distribution is a sea breeze that blows from the sea toward the land within several kilometers of the coast. At night, the land cools more quickly than the water. The air above the land becomes cooler than the air over the water. With higher surface pressure now over the land, the wind reverses and becomes a land breeze which flows from the land toward the water (1). These effects lead to fundamental differences in heat flux between the two surfaces. As a result, the formation of a thermal internal boundary layer often takes place. Another result of these processes is that the stability regimes over water can differ greatly from those over land. Even the larger scale processes associated with these differences between the land boundary layer and marine boundary layer are difficult for a mesoscale model to capture.

Meteorological models have had a major impact on forecasting in this region. Synoptic-scale models have aided in providing wind speed and direction forecasts along with predictions of cloud cover and the occurrence of precipitation. Cloud cover may inhibit the development of a sea-breeze circulation while precipitation may inhibit the strength of pre-frontal winds. Unfortunately, due to the coarse resolution, synoptic-scale models are not capable of capturing the mesoscale dynamical flow features that often occur in the region in connection with complex land/sea meteorological processes. Mesoscale models have been found to be very useful in capturing these small scale features (32). The advantage of using a mesoscale model is that it has the capability to discern and resolve small-scale features that synoptic scale models can not handle. However, recent work by 32) has indicated that the current versions of the available mesoscale models are incapable of producing reliable high-resolution wind forecasts during periods of weak synoptic-scale forcing. Although sea-breeze processes have been adequately simulated in research studies (e.g., Rao and Fuelberg 27), operational prediction of small-scale processes such as the land-sea breeze circulation remains problematic for real-time numerical weather prediction (NWP) models available from National Center for Environmental Prediction (NCEP). This is true even at high reso-
olution (e.g., 4 km) grid spacing. The ability of NWP models to represent coastal wind fields is limited even when the models are run at very high resolution, owing to limitations in the ability of these models to parameterize sub-grid scale process (19). Moreover, the lack of a dense observing network over coastal waters precludes accurate high-resolution initial conditions for atmospheric models.

Currently, Jay Titlow of WeatherFlow Enterprises uses the output from NCEP models as one component in his efforts to provide detailed wind field forecasts over the Chesapeake Bay. Titlow does not use a formal model (either dynamical, statistical, or a combination of the two) to produce the forecast. He mentally integrates the information which comes from a variety of sources along with an intelligence-system that he has developed, based on his forecasting experience, to produce the wind field forecast. Any procedure that would make these meteorological model output wind fields more useful to him and others like him would be of great benefit to recreational interests, commercial interests and environmental interests.

Our main objective is to develop a statistical model which is capable of evaluating the mesoscale numerical model and providing reliable wind field maps over the Chesapeake Bay during the period of weak synoptic scale forcing.

1.2 Description of data

In the summer of 2001, National Aeronautic and Space Administration (NASA) and the Defense Threat Reduction Agency (DTRA) sponsored a field experiment (The Chesapeake Bay Numerical Weather Prediction Model Experiment) over the Chesapeake Bay (see Figure 1.1) to collect data pertinent to the numerical modeling community. Data were collected over the period July 16 to July 31. These days were chosen in the hopes that weakly forced flow conditions would prevail for a good portion of this period. Unfortunately, only the period 21-23 July fully met that criterion. It was hoped that the data collected could be used to improve the ability of mesoscale models to forecast the state of the atmosphere over complex terrain when small scale processes dominate the flow regime, either by improving initial conditions or in the development of a post-processing technique that
could be applied to meteorological model output. Limited data collection was continued into the summer of 2002. The criterion of weakly forced conditions was better met during the study period in 2002 than in 2001. The data from the 2002 period are used in analysis. During this period both observed data and MM5 (Penn State University/National Center for Atmospheric Research Mesoscale Model Version 5) output fields are available.

The observed data from the field experiment were measured hourly at 17 stations. These 17 stations are represented by the number in Figure 1.2 (a). The information of these 17 monitoring stations is given in the Table 1.1. At each station, the anemometer was at a different height above ground level. It varied from 9 meters to 18 meters. The observed data include not only the wind but also other weather information such as air temperature, pressure and humidity. The advantage of using the data from the experiment is that these special supplementary observed wind data were not integrated into the NCEP analysis used to initialize MM5, so that the MM5 model output and the observed wind data are independent.

The MM5 model output used here were generated using a 15-km Arakawa C grid. The points in Figure 1.2 (b) are the centers of each grid cell. This grid spacing was selected
<table>
<thead>
<tr>
<th>Location</th>
<th>ID</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Sensor Height</th>
<th>Land/Water</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ocean City</td>
<td>KX034</td>
<td>38.328</td>
<td>75.086</td>
<td>15</td>
<td>water</td>
</tr>
<tr>
<td>Point Lookout</td>
<td>KX054</td>
<td>38.04</td>
<td>76.322</td>
<td>15</td>
<td>land</td>
</tr>
<tr>
<td>Raccoon Point</td>
<td>KX095</td>
<td>38.141</td>
<td>75.785</td>
<td>9</td>
<td>water</td>
</tr>
<tr>
<td>Smith Island</td>
<td>KX096</td>
<td>37.974</td>
<td>76.041</td>
<td>10</td>
<td>land</td>
</tr>
<tr>
<td>Tangier Island</td>
<td>KX097</td>
<td>37.823</td>
<td>75.991</td>
<td>17</td>
<td>land</td>
</tr>
<tr>
<td>Onancock</td>
<td>KX098</td>
<td>37.813</td>
<td>75.89</td>
<td>12</td>
<td>land</td>
</tr>
<tr>
<td>Crisfield</td>
<td>KX099</td>
<td>37.97</td>
<td>75.87</td>
<td>12</td>
<td>water</td>
</tr>
<tr>
<td>Windmill Point</td>
<td>KX100</td>
<td>37.616</td>
<td>76.291</td>
<td>11</td>
<td>land</td>
</tr>
<tr>
<td>Cove Point</td>
<td>KX101</td>
<td>38.384</td>
<td>76.382</td>
<td>13</td>
<td>land</td>
</tr>
<tr>
<td>Lower Hooper Island</td>
<td>KX102</td>
<td>38.258</td>
<td>76.179</td>
<td>10</td>
<td>land</td>
</tr>
<tr>
<td>New Point Comfort</td>
<td>KX103</td>
<td>37.328</td>
<td>76.273</td>
<td>15</td>
<td>land</td>
</tr>
<tr>
<td>Great Wicomico Light</td>
<td>KX104</td>
<td>37.783</td>
<td>76.277</td>
<td>12</td>
<td>water</td>
</tr>
<tr>
<td>3rd Island CBBT</td>
<td>KX105</td>
<td>37.036</td>
<td>76.077</td>
<td>15</td>
<td>land</td>
</tr>
<tr>
<td>Bishop’s Head</td>
<td>KX106</td>
<td>38.217</td>
<td>76.033</td>
<td>18</td>
<td>land</td>
</tr>
<tr>
<td>Deltaville</td>
<td>KX107</td>
<td>37.562</td>
<td>76.301</td>
<td>9</td>
<td>water</td>
</tr>
<tr>
<td>Silver Beach</td>
<td>KX108</td>
<td>37.485</td>
<td>75.962</td>
<td>14</td>
<td>water</td>
</tr>
<tr>
<td>Coles Point</td>
<td>KX109</td>
<td>38.142</td>
<td>76.614</td>
<td>9</td>
<td>water</td>
</tr>
</tbody>
</table>

Table 1.1: Information of monitoring locations
because it is similar to the finest mesh produced by operational models run at the NCEP, and based on the discussion with Jay Titlow at WeatherFlow Inc. The MM5 model output has 48 variables (see Table 1.2 and Table 1.3) over 17 layers. The 10 meter wind fields were used for analysis.

1.3 Objectives

Our main objective is to develop a statistical model which is capable of statistically evaluating wind fields from MM5 and providing reliable wind field maps over the Chesapeake Bay.

The objective comparison between numerical model output and observed data provides a mean for assessing numerical model performance. However, it is not reasonable to compare them directly, since the numerical model output usually has a different spatial/temporal scale compared with observed data. The MM5 model output is grided over regular spatial cells, while data from the field experiment are point observations measured at the irregular spaced monitoring sites. These various data fields have different spatial scales. The hierarchical Bayesian approach is ideal for this application because it provides a mechanism for combining data from different resources with different spatial/temporal resolution, and it also provides a natural framework in which it is possible to include scientific knowledge in the model. The strategy is based on the formulation of three primary hierarchical stages:

- Stage 1. Data model: [data|underlying true process, parameters],
- Stage 2. Process model: [underlying true process|parameters],
- Stage 3. Prior: [parameters],

where the bracket notation denotes a probability distribution. The essence of this approach is to break complex problems into simpler pieces (3). In our case we break the problem into a series of conditional models. In practice, the distribution of the process of interest, which in our case is the wind field with the associated parameters, can be specified as a product
<table>
<thead>
<tr>
<th>Variable</th>
<th>Unit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBL</td>
<td>m</td>
<td>planetary boundary layer height</td>
</tr>
<tr>
<td>RN</td>
<td>cm</td>
<td>cumulative non-convective rainfall</td>
</tr>
<tr>
<td>RC</td>
<td>cm</td>
<td>cumulative convective rainfall</td>
</tr>
<tr>
<td>TGD</td>
<td>K</td>
<td>ground temperature</td>
</tr>
<tr>
<td>USTAR</td>
<td>m/s</td>
<td>friction velocity</td>
</tr>
<tr>
<td>CFRACH</td>
<td>none</td>
<td>MM5 high-cloud fractional coverage</td>
</tr>
<tr>
<td>CFRACM</td>
<td>none</td>
<td>MM5 medium-cloud fractional coverage</td>
</tr>
<tr>
<td>CFRAACL</td>
<td>none</td>
<td>MM5 low-cloud fractional coverage</td>
</tr>
<tr>
<td>MOL</td>
<td>1/m</td>
<td>inverse Monin-Obukhov length</td>
</tr>
<tr>
<td>RNET</td>
<td>W/m²</td>
<td>net radiation at ground</td>
</tr>
<tr>
<td>FSOIL</td>
<td>W/m²</td>
<td>heat flux in soil layers</td>
</tr>
<tr>
<td>WSTAR</td>
<td>m/s</td>
<td>convective velocity</td>
</tr>
<tr>
<td>TA</td>
<td>K</td>
<td>air temperature for layer 1</td>
</tr>
<tr>
<td>QV</td>
<td>gH₂O/gdryair</td>
<td>water vapor mixing ratio for layer 1</td>
</tr>
<tr>
<td>QR</td>
<td>gH₂O/gdryair</td>
<td>rain mixing ratio for layer 1</td>
</tr>
<tr>
<td>QNI</td>
<td>gH₂O/gdryair</td>
<td>snow mixing ratio for layer 1</td>
</tr>
<tr>
<td>PRES</td>
<td>Pa</td>
<td>pressure for layer 1</td>
</tr>
<tr>
<td>TV</td>
<td>K</td>
<td>virtual temperature for layer 1</td>
</tr>
<tr>
<td>RHLW</td>
<td>%</td>
<td>relative humidity for layer 1, one-phase water-only</td>
</tr>
<tr>
<td>UX</td>
<td>m/s</td>
<td>cross point U wind component for layer 1</td>
</tr>
<tr>
<td>VX</td>
<td>m/s</td>
<td>cross point V wind component for layer 1</td>
</tr>
<tr>
<td>WSPD</td>
<td>m/s</td>
<td>wind speed for layer 1</td>
</tr>
<tr>
<td>RSD</td>
<td>W/m²</td>
<td>downward shortwave radiation</td>
</tr>
<tr>
<td>ALB</td>
<td>none</td>
<td>albedo</td>
</tr>
<tr>
<td>EMISS</td>
<td>none</td>
<td>emissivity</td>
</tr>
<tr>
<td>GLW</td>
<td>W/m²</td>
<td>longwave radiation at ground</td>
</tr>
<tr>
<td>GSW</td>
<td>W/m²</td>
<td>shortwave radiation at ground</td>
</tr>
<tr>
<td>MAVAIL</td>
<td>none</td>
<td>moisture availability (0-1)</td>
</tr>
</tbody>
</table>

Table 1.2: 48 variables from MM5 model output.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Unit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>QFX</td>
<td>$W/m^2$</td>
<td>MM5 latent heat flux</td>
</tr>
<tr>
<td>HFX</td>
<td>$W/m^2$</td>
<td>MM5 sensible heat flux</td>
</tr>
<tr>
<td>TMN</td>
<td>$K$</td>
<td>constant temperature of infinite reservoir below slab</td>
</tr>
<tr>
<td>ZNT</td>
<td>$m$</td>
<td>roughness length</td>
</tr>
<tr>
<td>KFTOP</td>
<td>$m$</td>
<td>convective cloud top, or 0.0</td>
</tr>
<tr>
<td>KFEXP</td>
<td>$s$</td>
<td>time until expiration for KF cloud, or 0.0</td>
</tr>
<tr>
<td>QV2</td>
<td>$g_{H_2O}/g_{air}$</td>
<td>2 meter mixing ratio</td>
</tr>
<tr>
<td>TA2</td>
<td>$K$</td>
<td>2 meter air temperature</td>
</tr>
<tr>
<td>UX10</td>
<td>$m/s$</td>
<td>10 meter U wind component</td>
</tr>
<tr>
<td>VX10</td>
<td>$m/s$</td>
<td>10 meter V wind component</td>
</tr>
<tr>
<td>WSPD10</td>
<td>$m/s$</td>
<td>10 meter wind speed</td>
</tr>
<tr>
<td>TSEASFC</td>
<td>$K$</td>
<td>sea surface temperature</td>
</tr>
<tr>
<td>SNODPT</td>
<td>$mm$</td>
<td>water equivalent snow depth</td>
</tr>
<tr>
<td>SRNOFF</td>
<td>$mm$</td>
<td>surface runoff</td>
</tr>
<tr>
<td>SOILT010</td>
<td>$K$</td>
<td>soil temperature for soil layer 1</td>
</tr>
<tr>
<td>SOILM010</td>
<td>$m^3/m^3$</td>
<td>soil moisture for soil layer 1</td>
</tr>
<tr>
<td>SLPRS</td>
<td>$mb$</td>
<td>sea level pressure</td>
</tr>
<tr>
<td>DZ1000-850</td>
<td>$m$</td>
<td>thickness between 1000mb and 850mb surfaces</td>
</tr>
<tr>
<td>DZ1000-500</td>
<td>$m$</td>
<td>thickness between 1000mb and 500mb surfaces</td>
</tr>
<tr>
<td>DZ850-700</td>
<td>$m$</td>
<td>thickness between 850mb and 700mb surfaces</td>
</tr>
</tbody>
</table>

Table 1.3: 48 variables from MM5 model output (continued).
Bayesian analysis relies on the posterior distribution of the process of interest and the parameters given the data: $[\text{underlying true process, parameters}|\text{data}]$. This distribution is proportional to the product of the component distributions:

$[\text{data}|\text{underlying true process, parameters}][\text{underlying true process}|\text{parameters}][\text{parameters}]$.

Therefore, the solutions to our problems are straightforward. We model both the observed wind field and MM5 model output in terms of the underlying true process. For numerical model evaluation it consists of comparing the observed wind values with their posterior distributions given MM5 model output. To obtain improved wind field maps, we simulate the underlying true wind process given all the available data. The key is to model the underlying true wind process, which is a nonstationary and nonseparable spatial-temporal process. In next chapter, we will introduce spatial-temporal processes and review some techniques for spatial-temporal modeling.
Chapter 2

Spatial-Temporal Processes

2.1 Introduction of spatial-temporal processes

Environmental data usually have both spatial and temporal components. For example, MM5 model output were generated at regular grids over regular time intervals, say every hour. Thus the data analysis has to take into account the spatial dependence among the grid cells, but also that the model output at each grid cell is not independent over time but forms a time series. In other words, one must take account of temporal correlations as well as spatial correlations. Therefore, it is essential to have statistical models to describe how the data vary across space and time.

Following 6), the data
\[ f(Z(x_1,i,t_i), ..., Z(x_{n_i},i,t_i)) : i = 1, \ldots, m \]  
are assumed to be a finite sample of the stochastic process
\[ \{Z(x,t) : x \in D(t); t \in T\}, \]  
where the domain \( D(t) \subset \mathbb{R}^d \) may vary with time and \( T \subset \mathbb{R} \). Usually \( n_i = n \), \( D(t) \equiv D \) and \( T = 1, 2, \ldots \).

Definition 2.1.1 (Weak Stationarity). If the process (2.2) satisfies:
(1) $E\{Z(x, t)\} = \mu$, where $\mu$ is a constant;

(2) and $\text{cov}\{Z(x + h, t + u), Z(x, t)\} = C(h, u) < \infty$, where $h \in \mathbb{R}^d$ and $u \in \mathbb{R}$,

then the process (2.2) is a weakly stationary spatial-temporal process. It is also called a second-order stationary spatial-temporal process.

Thus the covariance of $Z$, which is weakly stationary, depends only on the separation vector $(h, u)$. Define $s = (x, t) \in \mathbb{R}^{d+1}$ to simplify the notation and write the spatial-temporal process as $Z(s)$. If $Z(s)$ is weakly stationary, $\text{cov}\{Z(s_1), Z(s_2)\} = C(s_1 - s_2)$, where $C(\cdot)$ is positive definite and defined on $\mathbb{R}^{d+1}$. Positive definite means

$$\sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j C(s_i - s_j) \geq 0$$

for all finite $N$, all $s_1, \ldots, s_N \in D \times T \subset \mathbb{R}^{d+1}$ and all real $a_1, \ldots, a_N$. If $C(s_1 - s_2)$ depends on $(s_1 - s_2)$ only through the Euclidean distance $|s_1 - s_2|$, then $Z(s)$ is isotropic.

Weak stationarity describes the property of shift invariance. A stronger form of shift invariance is strong stationarity.

**Definition 2.1.2 (Strong Stationarity).** A process $Z(s)$ is strongly stationary if all of its finite dimensional distribution functions are shift invariant. That is, for any finite $N$ and domain $D \times T$, with $\{s, s_1, \ldots, s_N\} \subseteq D \times T$ and $z_1, \ldots, z_N \in \mathbb{R}$ then

$$P(Z(s_1 + s) \leq z_1, \ldots, Z(s_N + s) \leq z_N) = P(Z(s_1) \leq z_1, \ldots, Z(s_N) \leq z_N).$$

Gaussian with finite second moment means strong stationarity. In general, strong stationarity implies weak stationarity, but the reverse is not true. For what follows, we work with weak stationarity and the word *stationary* refers to weakly stationary.

**Definition 2.1.3 (Separability).** If the covariance of process (2.2) can be written as

$$\text{cov}\{Z(s_1), Z(s_2)\} = \text{cov}\{Z(x_1, t_1), Z(x_2, t_2)\} = C_S(x_1, x_2)C_T(t_1, t_2),$$

where $C_S(\cdot)$ is a spatial covariance and $C_T(\cdot)$ is a temporal covariance, then the covariance is separable. The process is then a separable spatial-temporal process.
If we represent the data from a separable spatial-temporal process in the following vector form
\[ Z = (Z(x_1, t_1), \cdots, Z(x_n, t_1), \cdots, Z(x_1, t_m), \cdots, Z(x_n, t_m))^T, \]
then the separable spatial-temporal covariance matrix, \( \Sigma \), can be written as the Kronecker product of a covariance matrix for time only with one for space only,
\[ \Sigma = \Sigma_T \otimes \Sigma_S, \]
where \( \Sigma_T \) is a \( m \times m \) temporal covariance matrix and \( \Sigma_S \) is a \( n \times n \) spatial covariance matrix. The Kronecker product of two matrices \( V = (v_{ij})_{m \times m} \) and \( U = (u_{ij})_{n \times n} \) is defined as
\[ V \otimes U = \begin{pmatrix} v_{11}U & \cdots & v_{1m}U \\ \vdots & \ddots & \vdots \\ v_{m1}U & \cdots & v_{mm}U \end{pmatrix}_{mn \times mn}. \]
An advantage of a separable covariance is the great benefit for computation, because the inverse of \( \Sigma \) can be computed as
\[ (\Sigma)^{-1} = (\Sigma_T \otimes \Sigma_S)^{-1} = (\Sigma_T)^{-1} \otimes (\Sigma_S)^{-1}, \]
and the determinant of \( \Sigma \) is given by
\[ |\Sigma| = |\Sigma_T \otimes \Sigma_S| = |\Sigma_T|^n |\Sigma_S|^m. \]
Moreover, many techniques that have been developed and successfully used in time series analysis and geostatistics are available to this subclass of separable spatial-temporal processes.

From the definition of separability, we know that a separable process may or may not to be stationary. If a spatial-temporal process is separable and stationary, then
\[ \text{cov}\{Z(x_1, t_1), Z(x_2, t_2)\} = C_S(x_1 - x_2)C_T(t_1 - t_2) = C_S(h)C_T(u), \]
where \( h = x_1 - x_2 \) and \( u = t_1 - t_2 \). This implies that any pair of time series separated by spatial distance \( h \) is stationary and has the same structure; also any pair of spatial processes separated by temporal distance \( u \) is stationary and has the same pattern.
The generic space-time problem is to use data (2.1) to predict \( Z(x_0, t_0) \) where \( x_0 \in D \) and \( t_0 \in T \). In the next section, we review some approaches to modeling spatial-temporal processes.

### 2.2 Recent approaches to spatial-temporal processes

The classical geostatistics approach is based on spatial processes which are stationary, but it is widely recognized that real environmental processes are rarely stationary. In recent years, many methods for modeling nonstationary spatial processes have been developed. 29) present a nonparametric estimation procedure for nonstationary spatial covariance structure. The maximum likelihood approach is developed by 22) and 30). 13) introduces a moving-window kriging technique. 17) propose an alternative model to account for a nonstationary spatial covariance function, which is based on a moving average specification of a Gaussian process. This spatial model allows the spatial dependence structure to vary as a function of location. 9) introduce a new class of nonstationary processes which is based on the convolution of local stationary processes. In contrast to a “moving-window” approach, this model has the advantage that it is simultaneously defined everywhere. Also 24) and 18) develop methods that extend the empirical orthogonal functions (EOF) approach which is popular among atmospheric scientists in the eighties and in the early nineties.

These are the methodologies mainly used to deal with nonstationarity for spatial processes. In fact, environmental data usually have both spatial and temporal components. We should not ignore the dependency between space and time. A spatial-temporal statistical analysis can give us more insight than a purely spatial analysis. Most applications in spatial statistics involve the modeling of complex spatial-temporal dependency structures, and many of the problems of space and time modeling can be overcome by using separable processes. This subclass of spatial-temporal processes has several advantages, including rapid fitting and simple extensions of many techniques developed and successfully used in time series and classical geostatistics. In particular, a major advantage of these processes is that the covariance matrix for a realization can be expressed as the Kronecker product of two smaller matrices that arise separately from the temporal and purely spatial processes,
and hence its determinant and inverse are easily determinable. However, in real applications spatial-temporal processes are rarely separable and stationary.

The following subsections introduce some techniques for space-time process modeling. They are the maximum likelihood method, the moving window method and the Bayesian method. Some spatial-temporal covariance models are briefly reviewed too.

2.2.1 Maximum likelihood method

22) present a maximum likelihood approach to model multivariate spatial-temporal data when the sites are sparse but the time series is long.

Consider models for an $m$-variable spatial-temporal process, $\mathbf{Z} = \mathbf{Z}(x, t), \mathbf{Z} \in \mathbb{R}^m$ where the location $x$ is in some domain of interest $D \subset \mathbb{R}^p$ and $t$ indexes time. Then $\mathbf{Z}$ may be written as

$$\mathbf{Z}(x, t) = \mathbf{\mu}(x, t) + \mathbf{\epsilon}(x, t) \quad (2.3)$$

where $E(\mathbf{Z}(x, t)) = \mathbf{\mu}(x, t)$, the set of $m$ trend surfaces, and $\mathbf{\epsilon}$ is a mean zero second-order stationary spatial-temporal Gaussian process with covariance $\Sigma$. Since the data in Mardia and Goodall’s application are equally spaced in time, they generalize the repeated measurements model, with $N$ vectors of observations at each site, to a fully-factored model

$$\Sigma = \Sigma_{TSV} = \Sigma_T \otimes \Sigma_S \otimes \Sigma_V$$

where the $N \times N$ covariance matrix $\Sigma_T$ is from a temporal process. $\Sigma$ is the notation for a generic covariance, and subscripts $T$, $S$, $V$ denote temporal, spatial and multi-variate components.

Mardia and Goodall propose the idea, which is to remove the effect of time from the data and view the data as repeated measurements in space, then apply the maximum likelihood approach to estimate the covariance $\Sigma$. After transforming the data and detrending (subtracting the smooth of each variable at each spatial location with respect to time), model (2.3) is obtained, where for each $x$, the $\mathbf{\epsilon}(x, t)$ are i.i.d. as $t$ varies. Then it is regarded as a multivariable spatial model $\mathbf{Z}(x) = \mathbf{\mu}(x) + \mathbf{\epsilon}(x)$. The model of trend is
simplified by assuming the same form for the trend surface, which means \( \mu(x) = \beta^T f \). \( f \) is a polynomial function of spatial locations with coefficients \( \beta \). For covariance, they impose the conditions of stationarity and spatial isotropy. The maximum likelihood estimation is carried out by an iterative algorithm. Then multivariable prediction is done by plugging in the maximum likelihood estimates. They also discuss the transformation to stationary isotropic spatial covariance using a complete deformation from the “geographic” space to the new “isotropic” space.

Mardia and Goodall extend the multivariable spatial model to the case where a temporal variation is included. By using the factored model, they basically assume a separable model for space-time covariance., which is not always the case in application. The approach should have the ability to take into account space-time dependency.

### 2.2.2 Moving window method

13) presents a method for prediction under spatial nonstationarity. He called this method “moving-window regression residual kriging” (MWRRK). He then extends the method to the nonstationary spatial-temporal case (14 14, 15). In the spatial case, the method simply restricts the estimation and prediction procedure to the collection of sample locations within a circular subregion, as a window, centered at each point of prediction, which is illustrated in Figure 2.1. For spatial-temporal case, the restriction is to sample points within a cylinder. The center of the cylinder is the prediction point \( s_0 = (x_0, t_0) \). The cylinder is constructed as follows. First, choose observations within a time window \( (t_0 - \frac{m_T}{2}, t_0 + \frac{m_T}{2}) \) where \( m_T \) is the parameter for the width of time window. Second, within that time window, select spatial points in order of nearness to \( x_0 \), until \( n_c = n f_c \) points have been selected. The parameter \( f_c \) is the sampling fraction. The width of time window (i.e. height of the cylinder) is chosen by the analyst and should be consistent with the assumption of local temporal stationarity. To be consistent with the assumption of local spatial stationarity, the sample fraction is chosen as small as possible for homogeneity. Haas calls this procedure “moving-cylinder spatial-temporal kriging” (MCSTK). Figure 2.2 shows how the cylinder is chosen.
Figure 2.1: MWRKK. “x” indicates the predict location.
Figure 2.2: MCSTK. “x” indicates the predict point.
MCSTK consists of a two stage nonlinear regression procedure similar to that described by Cressie (1993, p. 22-24) for a two stage approach to linear estimated generalized least squares (EGLS).

The regression model in each prediction cylinder is

\[ Y(s) = \mu(s, \beta) + \psi(\mu(s, \beta), s)R(s), \]

where \( s = (x, t) \) is the space-time coordinate, and \( \mu \) is a parametric function of \( \beta \). The second term on the right side of the equation consists of a stationary spatial-temporal residual process, \( R(s) \), and a model for non-homogeneous variance, \( \psi(\mu(s, \beta), s) \).

The two stages are as follows. First, ordinary least square (OLS) residuals from the estimated space-time mean model are used to calculate the sample space-time semivariogram using the typical method of moments estimator (Cressie 1993, p.69). A separable space-time semivariogram model is fit to the sample semivariogram using Cressie’s (1993, p. 98-99) weighted nonlinear least squares (WLS) criterion, where the weights include an approximation to the variance of the sample semivariogram ordinates but do not account for the correlation between ordinates at different lags. The fitted variogram model is used to estimate the covariance matrix in a second EGLS fitting of the mean model, these residuals being used to obtain a refined estimate of the covariance structure via WLS variogram modeling. In these two stages, the heteroscedastic residual variance function \( \psi(\cdot) \) is assumed to be one. The mean and covariance estimates from the second stage are used to obtain a prediction, which is the sum of the estimated mean and the multiplication of the estimated \( \psi \) and the ordinary kriging prediction of residual at the prediction point, \( s_0 \). The estimated \( \psi \) is obtained by using the sample variance of the second stage residuals found at the locations close to \( s_0 \).

While his model allows for variance non-homogeneity, estimating \( \psi(\mu(s, \beta), s) \) in the EGLS procedures does not guarantee positive definiteness of the resulting covariance matrix estimate of the residuals, which led to problems when kriging. Instead, he adjusts for non-homogeneity after estimating the kriging covariance and only at the point of prediction rather than at all points in the cylinder. This adjustment improves cross-validation prediction standard error bias, but Haas suggests further investigation to gain theoretical
understanding of this method for accommodating non-homogeneous variance.

Because the MCSTK procedure defines a process local to each prediction cylinder, a global model is not defined. Hence a global covariance model does not exist. 15) does, however, give a method for calculating a global covariance matrix which he uses to perform Monte-Carlo hypothesis tests for pollutant trends and meteorological transport models. He defines the pairwise covariances between space-time points using the covariance of the prediction cylinder to which the midpoint of the points is closest. In this way, he arrives at a matrix of pairwise covariance for all prediction points. However, the matrix is not guaranteed to be positive definite. Using a result for real symmetric matrices, he obtains a positive definite matrix that is closest to the pairwise covariance matrix in a matrix norm sense. The positive definite approximation of the covariance matrix is obtained by replacing the nonpositive eigenvalues of the original covariance by small positive values.

The moving window approach (14 14, 15) has the advantage of retaining the familiarity of existing stationary techniques (i.e, kriging based on stationary variogram models). It allows for nonstationarity by letting the prediction system change with local windows centered on prediction locations. Haas’ method allows for irregularly spaced locations in space and time.

As mentioned above, the covariance estimated by Haas’ approach does not lead to a positive definite covariance matrix over all sites, 15) proposes a way around this using numerical analysis algorithms to find the nearest positive definite matrix. This is not as good as finding an explicit stochastic model for the entire process.

2.2.3 Bayesian method

In recent years, Bayesian methods are becoming quite popular for space-time modeling. It is well known that space-time data in environmental sciences always contain many different scales of spatial and temporal variability. Such data are often nonstationary in space and time and may involve many observation/prediction locations. These factors can limit the effectiveness of traditional space-time statistical models and methods. Also the
### Table 2.1: Hierarchical model by Wikle, Berliner, and Cressie (1998).

<table>
<thead>
<tr>
<th>Stage</th>
<th>Variables</th>
<th>Model</th>
<th>Sub-model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Data</td>
<td>$Z</td>
<td>Y, \theta_1$</td>
</tr>
<tr>
<td>2</td>
<td>Process</td>
<td>$Y</td>
<td>\mu, \beta, X, \theta_2$</td>
</tr>
<tr>
<td>3</td>
<td>Large and small scales</td>
<td>$\mu, \beta, X</td>
<td>\theta_3 = (\theta_\mu, \theta_\beta, \theta_X)$</td>
</tr>
<tr>
<td></td>
<td>Spatial prior: means</td>
<td>$[\mu</td>
<td>\theta_\mu]$</td>
</tr>
<tr>
<td></td>
<td>Spatial prior: seasonalities</td>
<td>$[\beta</td>
<td>\theta_\beta]$</td>
</tr>
<tr>
<td></td>
<td>Space-time dynamics</td>
<td>$[X</td>
<td>\theta_X]$</td>
</tr>
<tr>
<td>4</td>
<td>Model parameters</td>
<td>$[\theta_1, \theta_2, \theta_3</td>
<td>\theta_4 = (\theta_4(1), \theta_4(2), \theta_4(3))$</td>
</tr>
<tr>
<td></td>
<td>Measurement variances</td>
<td>$[\theta_1</td>
<td>\theta_4(1)]$</td>
</tr>
<tr>
<td></td>
<td>Model variances</td>
<td>$[\theta_2</td>
<td>\theta_4(2)]$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$[\theta_\mu</td>
<td>\theta_4(\mu)]$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$[\theta_\beta</td>
<td>\theta_4(\beta)]$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$[\theta_X</td>
<td>\theta_4(X)]$</td>
</tr>
<tr>
<td>5</td>
<td>Hyperparameters</td>
<td>$[\theta_4] = [\theta_4(1)</td>
<td>[\theta_4(2)]</td>
</tr>
</tbody>
</table>

difficulty of using point estimates of the covariance structure as a surrogate for the “true” covariance structure is that the uncertainty in the estimate is not directly translated to the final inference. One approach to include the uncertainty in the estimate is to use the Bayesian framework and base inference on the posterior distributions of the quantities of interest. This approach takes into account the complete likelihood surface rather than plugging in the maximum likelihood estimate of the covariance structure.

There are several recent examples of hierarchical Bayesian modeling for spatial-temporal processes. For example, 33) employ a hierarchical model for mapping disease rates. A series of recent papers by 34), 4), 28) and 35) show the use of the hierarchical Bayesian space-time models in atmospheric sciences.

We introduce the model proposed by 34) in detail. Suppose $Y(s, t)$ denotes the value of the process of interest at location $s$ and time $t$, where $(s, t) \in \mathcal{M}$ and $\mathcal{M}$ is a lattice or grid in space-time. A casual summary of the five stages of the basic hierarchical space-time model is presented in Table 2.1, where $[\cdot]$ represent the probability density function, and $[\cdot|\cdot]$ represent the conditional density (10). The five stages are as follows.

1. **First stage: measurement process**
Let \( Z \) denote observational data. A statistical measurement error model is then specified as \([Z|Y, \theta_1]\), where \( \theta_1 \) represents a collection of parameters. A standard example is to assume that, conditional on \( Y \) and \( \theta_1 \), the \( Z(s, t) \) are independent and that in each case, \( Z(s, t) \sim \text{Gau}(Y(s, t), \sigma^2_{s,t}) \). In this case, \( \theta_1 = \{ \sigma^2_{s,t} \} \) is the set of measurement error variances.

2. Second stage: large- and small-scale features

The authors suggest that the following modeling strategies are particularly relevant for atmospheric and oceanographic processes, since such processes are expected to display both strong seasonal variations and regional structures. The model for \( Y \) is conditional on three processes, denoted by \( \mu, \beta \), and \( X = \{X(s, t) : (s, t) \in \mathcal{M}\} \), and a collection of parameters \( \theta_2 \). Assume that for each site and time point

\[
Y(s, t) = \mu(s) + M(t; \beta(s)) + X(s, t) + \gamma(s, t)
\]

where \( \mu(s) \) represents a site-specific mean. \( M(t; \beta(s)) \) is a large-scale temporal model for seasonal effect with site-specific parameters \( \beta(s) \). The \( X(s, t) \) represent a short time scale dynamic process. The \( \gamma(s, t) \) are zero-mean random variables which model noise. The role of the \( X \)-process is to account for both spatial and temporal dynamics beyond those accounted for in long-term means and seasonal behavior. Since the modeled features, such as \( X \), explain much of the space-time structure of the \( Y \) process, instead of modeling \( \gamma(s, t) \), one can assume that the \( Y(s, t) \) are all conditionally independent random variables. Therefore

\[
Y(s, t) \sim \text{Gau}(\mu(s) + M(t; \beta(s)) + X(s, t), \sigma^2_Y(s))
\]

where \( \theta_2 = \{ \sigma^2_Y(s) \} \).

3. Third stage: spatial structures and dynamics

Define \( \mu = \{ \mu(s) : \text{all location } s \} \) and \( \beta = \{ \beta(s) : \text{all location } s \} \). The authors assume that \( \mu, \beta \) and \( X \) are mutually independent, conditional on third-stage parameters \( \theta_3 = (\theta_\mu, \theta_\beta, \theta_X) \), which leads to

\[
[\mu, \beta, X|\theta_3] = [\mu|\theta_\mu][\beta|\theta_\beta][X|\theta_X].
\]

To capture the gross spatial structures, the authors envisage the use of spatial models for \([\mu|\theta_\mu]\) and \([\beta|\theta_\beta]\). Physically and/or experimentally based dynamical models are
possible for $X$. Within the class of “statistical” or “stochastic” models, the most common example of $[X|\theta_X]$ is a (conditional) vector autoregression model.

4. Fourth stage: priors on parameters

The fourth stage is the specification of the priors for all model parameters, which is the specification of $[\theta_1, \theta_2, \theta_3|\theta_4]$. $\theta_4$ is some collection of hyperparameters. It is convenient to assume $\theta_4 = (\theta_4(1), \theta_4(2), \theta_4(3))$ associated with each stage, and a conditional independence relation

$$[\theta_1, \theta_2, \theta_3|\theta_4] = [\theta_1|\theta_4(1)][\theta_2|\theta_4(2)][\theta_3|\theta_4(3)].$$

Further, $\theta_4(3)$ would typically be partitioned as $\theta_4(3) = (\theta_4(\mu), \theta_4(\beta), \theta_4(X))$, and coupled with a further conditional independence assumption

$$[\theta_3|\theta_4(3)] = [\theta_\mu|\theta_4(\mu)][\theta_\beta|\theta_4(\beta)][\theta_X|\theta_4(X)].$$

5. Fifth stage: hyperpriors

Finally, hyperparameter priors are specified. The standard assumption is that

$$[\theta_4] = [\theta_4(1)][\theta_4(2)][\theta_4(\mu)][\theta_4(\beta)][\theta_4(X)].$$

Often, the formulation is simplified by taking $\theta_4(1)$ and/or $\theta_4(2)$ to be either empty or known, so that the corresponding terms on the right hand side of above equation drop out.

This hierarchical approach offers a flexible approach to modeling a large class of environmental space-time processes. It provides not only a natural framework in which to include scientific knowledge in the model, but also posterior distributions on quantities of interest that can be used for scientific inference. Moreover, the Bayesian method also provides a mechanism for combining data from very different sources. However, from a statistical point of view, it is still hard to find the covariance structure for space-time dependency directly.

2.2.4 Some models for nonseparable stationary spatial-temporal covariance

In this subsection, we introduce some nonseparable stationary spatial-temporal covariances proposed by 5) and 11).
5) propose a generic approach to developing parametric models for spatial-temporal processes. The method relies heavily on spectral representations for the theoretical space-time covariance structure, and generalizes the results of 23) for pure spatial processes. In essence, Matérn constructs a number of parametric families for spatial processes by direct inversion of spectral densities. Cressie and Huang show that the same ideas can be used to construct families of spatial-temporal covariances.

First, Cressie and Huang represent the stationary spatial-temporal covariance $C(h, u)$ as

$$C(h, u) = \int \int e^{i(k^T \omega + u\tau)} g(\omega, \tau) d\omega d\tau$$

(2.4)

where $C(h, u)$ is a stationary spatial-temporal covariance function in which $h$ represents a $d$-dimensional spatial vector and $u$ is a scalar time component. The function $g(\omega, \tau)$, where $\omega$ is $d$-dimensional and $\tau$ is scalar, is the spectral density of the covariance function $C$. The function $g$ may be written as a scalar Fourier transform in $\tau$,

$$g(\omega, \tau) = \frac{1}{2\pi} \int e^{-iu\tau} h(\omega, u) du$$

with inverse

$$h(\omega, u) = \int e^{iu\tau} g(\omega, \tau) d\tau.$$  

(2.5)

Putting (2.4) and (2.5) together,

$$C(h, u) = \int e^{ih^T \omega} h(\omega, u) d\omega.$$  

(2.6)

The next step is to write

$$h(\omega, u) = k(\omega) \rho(\omega, u)$$

(2.7)

where $k(\omega)$ is the spectral density of a pure spatial process and $\rho(\omega, u)$ for each $\omega$ is a valid temporal autocorrelation function in $u$. Cressie and Huang remark that any smooth space-time covariance function can be written in the form (2.6) and (2.7), and they also impose the conditions:
(C1) For each $\omega$, $\rho(\omega, \cdot)$ is a continuous temporal autocorrelation function, $\int \rho(\omega, u) du < \infty$ and $k(\omega) > 0$;

(C2) $\int k(\omega) d\omega < \infty$.

Under those conditions, the generic formula for $C(h, u)$ becomes

$$ C(h, u) = \int e^{ihT\omega}k(\omega)\rho(\omega, u) d\omega. \tag{2.8} $$

When $\rho(\omega, u)$ is independent of $\omega$, (2.8) reduces again to a separable model. 5) developed seven special cases of (2.8). For example,

$$ \rho(\omega, u) = \exp\left(-\frac{\|\omega\|^2 u^2}{4}\right) \exp(-\delta u^2), \quad (\delta > 0) $$

$$ k(\omega) = \exp\left(-\frac{c_0\|\omega\|^2}{4}\right), \quad (c_0 > 0) $$

which lead to

$$ C(h, u) \propto \frac{1}{(u^2 + c_0)^{d/2}} \exp\left(-\frac{\|h\|^2}{u^2 + c_0}\right) \exp(-\delta u^2). \tag{2.9} $$

The condition $\delta > 0$ is needed to ensure the condition (C1) is satisfied at $\omega = 0$, but the limit of (2.9) as $\delta \to 0$ is also a valid spatial-temporal covariance function, leading to the three parameter family

$$ C(h, u) = \frac{\sigma^2}{(a^2u^2 + 1)^{d/2}} \exp\left(-\frac{b^2\|h\|^2}{a^2u^2 + 1}\right). $$

Cressie and Huang’s approach is novel and powerful but depends on Fourier transform pairs in $\mathbb{R}^d$. 11) takes the approach of 5) and provides a very general class of valid spatial-temporal covariance models. The key result of 11) can be formulated as follows.

Let $\psi(t), t \geq 0$, be a completely monotone function$^1$, and let $\phi(t), t \geq 0$, be a positive function with a completely monotone derivative. Then

$$ C(h, u) = \frac{\sigma^2}{\phi(|u|^2)^{d/2}} \psi\left(\frac{\|h\|^2}{\phi(|u|^2)}\right), \tag{2.10} $$

$^1$A continuous function $\psi(t)$, defined for $t \geq 0$, is said to be completely monotone if it possesses derivatives $\psi^{(n)}$ of all orders and $(-1)^n \psi^{(n)}(t) \geq 0$ for $t > 0$ and $n = 0, 1, 2, \cdots$. 

is a space-time covariance function, where $h \in \mathbb{R}^d$ represents a $d$-dimensional spatial vector and $u \in \mathbb{R}$ is a scalar time component.

For example, putting $\psi(t) = \exp(-ct^\gamma)$ and $\phi(t) = (at^\alpha + 1)^\beta$ in (2.10) leads to

$$C(h, u) = \frac{\sigma^2}{(a|u|^{2\alpha} + 1)^{\beta d/2}} \exp\left(-\frac{c\|h\|^{2\gamma}}{(a|u|^{2\alpha} + 1)^{\beta \gamma}}\right),$$

where $(h, u) \in \mathbb{R}^d \times \mathbb{R}$. The product with the purely temporal covariance function $(a|u|^{2\alpha} + 1)^{-\delta}$, $u \in \mathbb{R}$, then gives the class

$$C(h, u) = \frac{\sigma^2}{(a|u|^{2\alpha} + 1)^{\delta + \beta d/2}} \exp\left(-\frac{c\|h\|^{2\gamma}}{(a|u|^{2\alpha} + 1)^{\beta \gamma}}\right),$$

where $(h, u) \in \mathbb{R}^d \times \mathbb{R}$. $a$ and $c$ are nonnegative scaling parameters of time and space, respectively; the smoothness parameters $\alpha$ and $\gamma$ take values in $(0, 1]$; $\beta \in [0, 1]$, $\delta \geq 0$ and $\sigma^2 > 0$. A separable covariance function is obtained when $\beta = 0$.

### 2.3 Summary

In the previous section, we have reviewed several existing approaches to modeling spatial-temporal processes. We have identified some advantages and disadvantages of these methods. But it is still not clear when we should fit the data with a separable or nonseparable spatial-temporal model. In the next chapter, we will propose an empirical test for separability to better understand space time dependence. Moreover, we will present a new nonseparable nonstationary spatial-temporal model, which is a mixture of local uncorrelated stationary spatial-temporal processes. Compared to the MCSTK, this new model is simultaneously defined everywhere. And we estimate it in a Bayesian way.
Chapter 3

Nonseparable and nonstationary spatial-temporal models

3.1 An empirical test for separability

Separability is a common assumption to avoid many of the problems of space-time modeling. As we mentioned earlier, this subclass of separable spatial-temporal processes has several advantages, including rapid fitting and simple extensions of many techniques developed and successfully used in time series analysis and geostatistics. However, in real applications spatial-temporal processes are rarely stationary and separable. Therefore, we propose an empirical test for separability to better understand space-time dependency.

For the (isotropic) stationary space-time process \( \{Z(x, t) : x \in D \subset \mathbb{R}^d, t \in T \subset \mathbb{R} \} \), define

\[
R = \frac{\text{cov}(Z(x_1, t_1), Z(x_3, t_3))}{\text{cov}(Z(x_2, t_2), Z(x_4, t_4))} = \frac{C(|x_1 - x_3|, |t_1 - t_3|)}{C(|x_2 - x_4|, |t_2 - t_4|)},
\]

where \( x_1, x_2, x_3, \) and \( x_4, \in D, t_1, t_2, t_3, \) and \( t_4 \in T \). \( R \) is the ratio of space-time covariance, a function of spatial lags, \(|x_1 - x_3|\) and \(|x_2 - x_4|\), and temporal lags, \(|t_1 - t_3|\) and \(|t_2 - t_4|\). To simplify the notation, let \( h_1 = |x_1 - x_3|, h_2 = |x_2 - x_4|, u_1 = |t_1 - t_3|, \) and \( u_2 = |t_2 - t_4|, \) then \( R = C(h_1, u_1)/C(h_2, u_2) \).
Define
\[ R_S(u) = \frac{C(h_1, u)}{C(h_2, u)} \]
and
\[ R_T(h) = \frac{C(h, u_1)}{C(h, u_2)}. \]

If the spatial-temporal covariance \( C(\cdot) \) is separable, then for fixed \( h_1 \) and \( h_2 \), \( R_S \) in (3.1) can be written as
\[ R_S(u) = \frac{C_S(h_1)C_T(u)}{C_S(h_2)C_T(u)} = \frac{C_S(h_1)}{C_S(h_2)}. \]

Therefore, \( R_S(u) \) is constant for all possible \( u \). This means the pair of time series separated by spatial distance \( h_1 \) and the pair of time series separated by \( h_2 \) have the same structure. Similarly, if the covariance is separable, then for fixed \( u_1 \) and \( u_2 \),
\[ R_T(h) = \frac{C_S(h)C_T(u_1)}{C_S(h)C_T(u_2)} = \frac{C_T(u_1)}{C_T(u_2)}. \]

Consequently, \( R_T(h) \) is constant for all possible \( h \). This means the pair of spatial processes separated by temporal distance \( u_1 \) and the pair of spatial processes separated by \( u_2 \) have the same pattern.

It turns out that the test for separability is equivalent to testing if \( R_S(u) \) is a constant for fixed \( h_1 \) and \( h_2 \), and \( R_T(h) \) is a constant for fixed \( u_1 \) and \( u_2 \). That is to consider
\[ H_0 : \text{ For fixed } h_1 \text{ and } h_2, \forall u, R_S(u) = \text{ constant}, \]
and for fixed \( u_1 \) and \( u_2 \), \( \forall h, R_T(h) = \text{ constant}. \)

\[ vs \ H_1 : \text{ For fixed } h_1 \text{ and } h_2, \forall u, R_S(u) \neq \text{ constant}, \]
or for fixed \( u_1 \) and \( u_2 \), \( \forall h, R_T(h) \neq \text{ constant}. \)

Here we introduce an empirical method to perform the test for separability. The procedure is as follows.

1. Consider \( R_S(u) \) to be constant for fixed \( h_1 \) and \( h_2 \). This is equivalent to asking if the time series at different locations has the same structure over space.
(1) Find a pair of time series which are separated by the distance $h_1$, and compute the covariance of these two time series at all possible temporal lags, say $\tilde{C}(h_1, u_0), \ldots, \tilde{C}(h_1, u_k)$. $u_0, \ldots, u_k$ are all possible temporal lags.

(2) Find two time series which are separated by $h_2$ and compute the covariance between them, $\tilde{C}(h_2, u_0), \ldots, \tilde{C}(h_2, u_k)$.

(3) Take the ratio of the covariances from step (1) and step (2). $\hat{R}_S(u_i) = \frac{\tilde{C}(h_1, u_i)}{\tilde{C}(h_2, u_i)}$, $i = 0, \ldots, k$.

(4) Repeat step (1)-(3) for all possible $n$ pairs of time series pairs which are separated by the distance $h_1$ and $h_2$. Therefore, at each possible temporal lag $u_i$, we have $R_{S1}(u_i), \ldots, R_{Sn}(u_i)$, and the empirical distribution of $R_S(u_i)$ is obtained.

(5) Define $m_S(u_i) = \text{median}(R_{S1}(u_i), \ldots, R_{Sn}(u_i))$. Compare the medians $m_S(u_0), \ldots, m_S(u_k)$ using the empirical distributions from step (4). If they are significantly different, then we conclude that $R_S(u)$ is not a constant for fixed $h_1$ and $h_2$.

2. Consider if $R_T(h)$ to be constant for fixed $u_1$ and $u_2$. This is equivalent to testing if the spatial process has the same pattern over time.

(1) Find a pair of spatial processes which are separated by the temporal lag $u_1$, and compute the covariance of these two spatial processes at all possible spatial lags, say $\tilde{C}(h_0, u_1), \ldots, \tilde{C}(h_i, u_1)$, where $h_0, \ldots, h_i$ are all possible spatial lags.

(2) Find two spatial processes which are separated by $u_2$ and compute the covariance between them, $\tilde{C}(h_0, u_2), \ldots, \tilde{C}(h_i, u_2)$.

(3) Take the ratio of the covariances from step (1) and step (2). $\hat{R}_T(h_i) = \frac{\tilde{C}(h_0, u_1)}{\tilde{C}(h_i, u_2)}$, $i = 0, \ldots, l$.

(4) Repeat step (1)-(3) for all possible $m$ pairs of spatial processes which are separated by the temporal lags $u_1$ and $u_2$. At each possible spatial lag $h_i$, we have $R_{T1}(h_i), \ldots, R_{Tm}(h_i)$, and the empirical distribution of $R_T(h_i)$ is obtained.

(5) Define $m_T(h_i) = \text{median}(R_{T1}(h_i), \ldots, R_{Tm}(h_i))$. Compare the medians $m_T(h_1), \ldots, m_T(h_l)$ using the empirical distributions from (4). If they are significantly different, then we say that $R_T(h)$ is not a constant for fixed $u_1$ and $u_2$. 
If $R_S(u)$ is not a constant for fixed $h_1$ and $h_2$, or $R_T(h)$ is not a constant for fixed $u_1$ and $u_2$, then we reject the null hypothesis of separability, and say the spatial-temporal process $Z(x, t)$ is nonseparable.

At the step 1(1) and 2(1), the sample covariance is calculated for all possible temporal/spatial lags. We know that the sample covariance for large lag is a very poor estimate for the covariance at that lag, so we only consider the sample covariances associated with small lags, for example, consider the first five spatial and temporal lags. The empirical distributions obtained at 1(4) and 2(4) may have a very large range and the shapes of the empirical distributions may be skewed. So we use medians in the step 1(5) and 2(5), since median is less affected by the above factors.

A simulation study for the empirical separability test was carried out. The data for spatial-temporal process $Z(x, t)$ were generated using different separable/nonseparable spatial-temporal covariances, where the $x$ are 32 equally spaced locations over a straight line and the $t$ are 32 equally spaced time points. For each covariance structure, $N=1000$ data sets are simulated, and the empirical separability test is performed for each data set. The empirical test is carried out based on the sample covariance ratio for the first five temporal and spatial lags. First, consider if $R_S(u)$ is constant for fixed $h_1 = 0$ and $h_2 = 4$. At each temporal lag, $u_i$, $i = 1, \cdots, 5$, we obtain an empirical distribution, a median $m_s(u_i)$ and a 95% credible interval. Then compare the medians, $m_s(u_1), \cdots, m_s(u_5)$. If they are significantly different, then $R_S(u)$ is not constant for fixed $h_1$ and $h_2$. Similarly, we consider if $R_T(h)$ is constant for fixed $u_1 = 0$ and $u_2 = 4$. If $R_S(u)$ is not constant or $R_T(h)$ is not constant, then we reject the null hypothesis of separability. The power function is then calculated as:

$$\text{power} = \frac{\text{number of rejections of separability}}{N}.$$  

If the data sets are generated from a separable covariance, we expect the power to be small, near 0; on the other hand, we expect power to be close to 1 when the data sets are generated from a nonseparable covariance.

Let $\Sigma$ be a generic notation for spatial-temporal covariance, $\Sigma_S$ be a purely spatial covariance and $\Sigma_T$ be a temporal only covariance. The separable spatial-temporal covariance
The models used for simulation are as follows. Both $\Sigma_S$ and $\Sigma_T$ are Matérn type covariances,

$$C(r) = \frac{\sigma}{2^{\nu-1} \Gamma(\nu)} \left( \frac{2^{1/2 r_T}}{\alpha} \right) K_\nu \left( \frac{2^{1/2 r_T}}{\alpha} \right), \quad (3.3)$$

where $K$ is a modified Bessel function. The parameter $\sigma$ is the variance of the process, $\alpha$ represents the autocorrelation range, and $\nu$ measures the degree of smoothness associated with spatial/temporal process. The notation $M(\sigma, \alpha, \nu)$ represents a Matérn covariance function. The models for $\Sigma_s$ and $\Sigma_T$ in this simulation study are listed in the Table 3.1.

For nonseparable spatial-temporal covariance, we also use the Matérn type covariance function in which time is treated as a component of space, although this is not realistic because the units for space and time are usually different. Besides, the nonseparable covariance functions proposed by 5) and 11) are used in simulation study too. The models for nonseparable covariance are in the Table 3.2.

<table>
<thead>
<tr>
<th>$H_0$</th>
<th>$\Sigma_S$</th>
<th>$\Sigma_T$</th>
<th>power (Type I error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>$M(1, 4, 1.5)$</td>
<td>$M(1, 4, 1.5)$</td>
<td>0.291</td>
</tr>
<tr>
<td>S2</td>
<td>$M(1, 4, 2.5)$</td>
<td>$M(1, 4, 2.5)$</td>
<td>0.049</td>
</tr>
<tr>
<td>S3</td>
<td>$M(1, 4, 1.5)$</td>
<td>$M(1, 8, 1.5)$</td>
<td>0.090</td>
</tr>
<tr>
<td>S4</td>
<td>$M(1, 4, 2.5)$</td>
<td>$M(1, 8, 2.5)$</td>
<td>0.040</td>
</tr>
<tr>
<td>S5</td>
<td>$M(1, 4, 2.5)$</td>
<td>$M(1, 8, 1.5)$</td>
<td>0.053</td>
</tr>
<tr>
<td>S6</td>
<td>$M(1, 4, 1.5)$</td>
<td>$M(1, 8, 2.5)$</td>
<td>0.070</td>
</tr>
</tbody>
</table>

Table 3.1: Separable covariance models.

<table>
<thead>
<tr>
<th>$H_1$</th>
<th>$\Sigma$</th>
<th>power</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>$M(1, 4, 0.5)$</td>
<td>0.996</td>
</tr>
<tr>
<td>N2</td>
<td>$\frac{1}{(u^2+1)} \exp(-\frac{h^2}{u^2+1})$</td>
<td>0.874</td>
</tr>
<tr>
<td>N3</td>
<td>$\frac{1}{\sqrt{u+1}} \exp(-\frac{h^2}{u+1})$</td>
<td>0.965</td>
</tr>
<tr>
<td>N4</td>
<td>$\frac{1}{\sqrt{</td>
<td>u</td>
</tr>
</tbody>
</table>

Table 3.2: Nonseparable covariance models.

The simulation results show that the power is near to 0.05 for separable covari-
ances except for S1, and the power is close to 1 when nonseparable covariances are used. The contour plot for the separable covariance model S1 is shown in Figure 3.1. For small spatial and temporal distance, the shape of S1 is quite similar to a nonseparable covariance, for example the nonseparable covariance model N1 in Figure 3.2. This similarity leads to a large power (Type I error) for the test for separable covariance S1. The contour plots for nonseparable covariance models, N2 and N3, are shown in Figure 3.3 and 3.4. They have similar shapes, but nonseparable covariance N3 has a larger temporal range. By temporal range, we mean the autocorrelation decay by projecting spatial-temporal covariance on temporal domain. For the model N3, it has a stronger space-time dependency compared to the model N2. Therefore, as shown in Table 3.2, the power of the test for model N3 is larger than N2, but the power of the test for model N2 is still acceptable.

The spatial block bootstrap method might help us to get a finer empirical distribution of the ratio. But the bootstrap blocks should be large enough to preserve the shape of spatial-temporal dependency. This empirical separability test can help us to answer the question, which type of covariance, separable or nonseparable, is appropriate for a stationary spatial-temporal process. Moreover, we can get a better idea about if the spatial structure changes over time, and if the temporal structure changes over space.

In next section, we will propose a new nonseparable stationary covariance model.
Figure 3.1: The contour plot for separable covariance $S_1$.

Figure 3.2: The contour plot for nonseparable covariance $N_1$. 
Figure 3.3: The contour plot for nonseparable covariance N2.

Figure 3.4: The contour plot for nonseparable covariance N3.
3.2 A new class of nonseparable stationary covariance models

In previous section, we introduced an empirical test for separability. If the result of the test suggests separability, then the existing spatial/temporal covariance models can be employed directly. Otherwise, we need spatial-temporal models to take into account the nonseparability. Nonseparable spatial-temporal models have been proposed by 5), 11), and 31). We introduce a new class of nonseparable stationary covariance structures, and they are used to construct new nonseparable nonstationary models later.

The stationary spatial-temporal process \( \{ Z(x, t) : x \in D \subset \mathbb{R}^d, t \in T \subset \mathbb{R} \} \) can be represented in the spectral domain, instead of space-time domain, which is always interpreted as the superposition of sine and cosine waves of different frequencies \((\omega, \tau)\), where \( \omega \) is \( d \)-dimensional spatial frequency and \( \tau \) is temporal frequency. If \( Z(x, t) \) is a stationary random field with spatial-temporal covariance \( C(x, t) \), then we can represent the process in the form of the following Fourier-Stieltjes integral (see 36) for example):

\[
Z(x, t) = \int_{\mathbb{R}^d} \int_{\mathbb{R}} \exp(i\omega^T x + i\tau t) dY(\omega, \tau).
\]  

(3.4)

where \( Y \) is a random function that has uncorrelated increments with complex symmetry except for the constraint, \( dY(\omega, \tau) = dY^c(-\omega, -\tau) \), needed to ensure \( Z(x, t) \) is real-valued. \( Y^c \) denotes the conjugate of \( Y \). Using the spectral representation of \( Z \) and proceeding formally,

\[
C(x, t) = \int_{\mathbb{R}^d} \int_{\mathbb{R}} \exp(i\omega^T x + i\tau t) F(d\omega, d\tau),
\]  

(3.5)

where the function \( F \) is a positive finite measure and is called the spectral measure or spectrum for \( Z \). The spectral measure \( F \) is the mean square value of the process \( Y \),

\[
E\{|Y(\omega, \tau)|^2\} = F(\omega, \tau).
\]

It is easy to see that for any finite positive measure \( F \), the function given in (3.5) is positive.
definite, since for all finite \( n \), all \( s_1, \ldots, s_n \in \mathbb{R}^d \times \mathbb{R} \), and real \( c_1, \ldots, c_n \),

\[
\sum_{j,k=1}^{n} c_j c_k C(s_j - s_k) = \sum_{j,k=1}^{n} c_j c_k \int_{\mathbb{R}^{d+1}} \exp\{i\psi^T (s_j - s_k)\} F(d\psi)
\]

\[
= \int_{\mathbb{R}^{d+1}} \left| \sum_{j=1}^{n} c_j \exp(i\psi^T s_j) \right|^2 F(d\psi)
\]

\[
\geq 0,
\]

where \( s = (x, t) \) and \( \psi = (\omega, \tau) \) are used to simplify the notation.

**Theorem 3.2.1 (Bochner’s Theorem).** A complex-valued function \( K \) on \( \mathbb{R}^d \) is
the autocovariance function for a weakly stationary mean square continuous complex-valued random field on \( \mathbb{R}^d \) if and only if it can be represented as

\[
K(x) = \int_{\mathbb{R}^d} \exp\{i\omega^T x\} F(d\omega),
\]

where \( F \) is a positive finite measure.

Bochner’s theorem gives the relationship between the covariance and the spectrum for a stationary process. If \( F \) has a density with respect to Lebesgue measure, it is the spectral density, \( f \), which is the Fourier transform of the spatial-temporal covariance function:

\[
f(\omega, \tau) = \frac{1}{(2\pi)^{d+1}} \int_{\mathbb{R}^d} \int_{\mathbb{R}} \exp(-i\omega^T x - i\tau t) C(x, t) dx dt,
\]

and the corresponding covariance function is given by

\[
C(x, t) = \int_{\mathbb{R}^d} \int_{\mathbb{R}} \exp(i\omega^T x + i\tau t) f(\omega, \tau) d\omega d\tau.
\]

For example, the Matérn spatial spectral density is given by

\[
f(\omega) = \gamma (\alpha^2 + |\omega|^2)^{-\nu-d/2},
\]

and its corresponding Matérn spatial covariance function is

\[
C(x) = \frac{\pi^{d/2} \gamma}{2^{\nu-1} \Gamma(\nu + d/2) \alpha^{2\nu}} (\alpha |x|)^\nu K_\nu(\alpha |x|),
\]

where \( K_\nu(\alpha |x|) \) is a modified Bessel function (see 12, for example). This expression of the Matérn spatial covariance is slightly different with the one in (3.3) due to different parameterizations.
When \( f(\omega, \tau) = f^{(1)}(\omega)f^{(2)}(\tau) \), we obtain

\[
\text{cov}\{Z(x_1, t_1), Z(x_2, t_2)\} = C(x_1 - x_2, t_1 - t_2)
\]
\[
= \int_{\mathbb{R}^d} \exp\{i\omega^T(x_1 - x_2)\}f^{(1)}(\omega)(f^{(1)}(\omega))^c d\omega
\]
\[
\times \int_{\mathbb{R}^1} \exp\{i\tau(t_1 - t_2)\}f^{(2)}(\tau)(f^{(2)}(\tau))^c d\tau
\]
\[
= C^{(1)}(x_1 - x_2)C^{(2)}(t_1 - t_2),
\]
which means the corresponding spatial-temporal covariance is separable.

We propose the following spatial-temporal spectral density, that has a separable model as a particular case,

\[
f(\omega, \tau) = \gamma(\alpha^2 \beta^2 + \beta^2 |\omega|^2 + \alpha^2 \tau^2 + \epsilon|\omega|^2 \tau^2)^{-\nu}, \tag{3.8}
\]

where \( \gamma, \alpha \) and \( \beta \) are positive, \( \nu > \frac{d+1}{2} \) and \( \epsilon \in [0, 1] \). The function in (3.8) is a valid spectral density. First, \( f(\omega, \tau) > 0 \) everywhere. Second, \( f(\omega, \tau) \leq \gamma(\alpha^2 \beta^2 + \beta^2 |\omega|^2 + \alpha^2 \tau^2)^{-\nu}, \) and

\[
\int_{\mathbb{R}^d} \int_{\mathbb{R}} \exp(i\omega^T x + i\tau t)\gamma(\alpha^2 \beta^2 + \beta^2 |\omega|^2 + \alpha^2 \tau^2)^{-\nu} d\omega d\tau
\]
\[
= \frac{\pi^{d+1}}{2^{\nu-\frac{d+1}{2}} \Gamma(\nu)\alpha^{2\nu-d} \beta^{2\nu-1}} \left( \alpha \sqrt{\frac{\beta^2}{\alpha}} \right)^{\nu-\frac{d+1}{2}} \times
\]
\[
K_{\nu-\frac{d+1}{2}} \left( \alpha \sqrt{\frac{\beta^2}{\alpha}} \right), \tag{3.9}
\]

Therefore, \( \int_{\mathbb{R}^d} \int_{\mathbb{R}} \exp(i\omega^T x + i\tau t)f(\omega, \tau)d\omega d\tau \) exists.

When \( \epsilon = 1 \), the equation (3.8) can be written as

\[
f(\omega, \tau) = \gamma(\alpha^2 \beta^2 + \beta^2 |\omega|^2 + \alpha^2 \tau^2 + |\omega|^2 \tau^2)^{-\nu}
\]
\[
= \gamma(\alpha^2 + |\omega|^2)^{-\nu}(\beta^2 + \tau^2)^{-\nu}.
\]

Therefore the corresponding spatial-temporal covariance is separable. Moreover, in the expression of this covariance, both the spatial component and the temporal component are the Matérn type covariances. When \( \gamma = \alpha = \beta = d = 1 \) and \( \nu = 3/2 \), a contour plot of corresponding separable spatial-temporal covariance is given in Figure 3.5. From the plot, there are ridges along the lines where spatial lag is 0 and temporal lag is 0.
When $\epsilon = 0$,
\[ f(\omega, \tau) = \gamma(\alpha^2 \beta^2 + \beta^2 |\omega|^2 + \alpha^2 \tau^2)^{-\nu}. \] (3.10)

The function in (3.10) is an extension of the traditional Matérn spectral density. It treats time as an additional component of space, but it does have a different rate of decay. In the spectral density (3.10), the parameter $\alpha^{-1}$ explains the rate of decay of the spatial correlation. For the temporal correlation, the rate of decay is explained by the parameter $\beta^{-1}$. $\gamma$ is a scale parameter. The parameter $\nu$ measures the degree of smoothness of the process $Z$. The higher value of $\nu$, the smoother the process $Z$ will be. The corresponding spatial-temporal covariance is given by (3.9), which is a Matérn type covariance. Following the parameterization suggested by (16), we have
\[ C(x, t) = \frac{\sigma^2}{2^{\nu-1} \Gamma(\nu)} \left( \frac{\|(x, \rho t)\|}{r} \right) K_\nu \left( \frac{\|(x, \rho t)\|}{r} \right), \] (3.11)
where $\sigma^2$, $r$, $\rho$ and $\nu$ are all positive. $\| \cdot \|$ denotes the Euclidean distance. In the representation (3.11), the parameter $r$ measures how the correlation decays with distance; generally this parameter is called range. The parameter $\sigma^2$ is the variance of the process $Z$, i.e, $\sigma^2 = \text{var}(Z(x,t))$, where the covariance parameter $\sigma^2$ is usually referred to as sill. The parameter $\nu > 0$ has the same interpretation as it had in the spectral representation. It measures the degree of smoothness of the process $Z$. The parameter $\rho$, which is new, is a scale factor to take into account the change of units between the spatial and temporal domains. Therefore, this parametric model for $C$ corresponds to a $d+1$ dimensional Matérn type covariance with an extra parameter $\rho$, which can be explained as a conversion factor between the units in the space and time domains. When $\gamma = \alpha = \beta = d = 1$ and $\nu = 3/2$, a contour plot of corresponding separable spatial-temporal covariance is given in Figure 3.6. It has a very smooth surface.

When $\epsilon \in (0, 1)$, we are not able to write down the exact expression of spatial-temporal covariance, which corresponds to the spectral density in (3.8). But we can calculate this analytically, since

$$C(x,t) = \int_{\mathbb{R}^d} \int_{\mathbb{R}} \exp(i\omega^T x + i\tau t) f(\omega, \tau) d\omega d\tau$$

$$= \int_{\mathbb{R}} \exp(i\tau t) g(x, \tau) d\tau,$$

(3.12)

where $g(x, \tau) = \int_{\mathbb{R}^d} \exp(i\omega^T x) f(\omega, \tau) d\omega$. The function $g(x, \tau)$ is available from the integration, therefore $C(x, t)$ can be computed by numerically carrying out a one-dimensional Fourier transformation of $g$. This can be quickly approximated using fast Fourier transform. A separate transform needs to be done for every value of $(x, t)$ of interest, but this is feasible. The expression of $g$ is given by

$$g(x, \tau) = \frac{\pi^{d/2} \gamma}{2^{\nu - \frac{d}{2} - 1} \Gamma(\nu)} (\beta^2 + \epsilon \tau^2)^{-\nu} \left( \frac{|x|}{\theta(\tau)} \right)^{\nu - \frac{d}{2}} K_{\nu - \frac{d}{2}}(\theta(\tau)|x|),$$

where $\theta(\tau) = \sqrt{\frac{\alpha^2 (\beta^2 + \epsilon \tau^2)}{\beta^2 + \epsilon \tau^2}}$. When $\gamma = \alpha = \beta = d = 1$ and $\nu = 3/2$, contour plots of corresponding separable spatial-temporal covariances with $\epsilon = 0.1, 0.2, 0.3, 0.5, 0.7, 0.8, 0.9$ are given in Figure 3.7. The ridge is getting more obviously when $\epsilon$ is getting large, say close to 1. In fact, when $\epsilon$ is close to 1, $\epsilon|\omega|^2 \tau^2$ is close to $|\omega|^2 \tau^2$. Therefore the spectral density in (3.8) is turning close to a separable form from a nonseparable form as $\epsilon \to 1$. 


Figure 3.6: The contour plot for a nonseparable spatial-temporal covariance.

In summary, the new class spectral density in (3.8) is nonseparable for $0 \leq \epsilon < 1$, and separable for $\epsilon = 1$. Therefore, the parameter $\epsilon$ plays a role for separability. It controls the interaction between the spatial component and the temporal component. When $\epsilon$ equals 0 and 1, there are exact forms for the corresponding spatial-temporal covariances. Otherwise, the corresponding spatial-temporal covariance has to be computed numerically. Note that the degree of smoothness is same for the spatial component and the temporal component. Therefore a more general class is proposed to allow different degree of smoothness for space and time,

$$f(\omega, \tau) = \gamma \left\{ c_1 (a_1^2 + |\omega|^2)^{\alpha_1} + c_2 (a_2^2 + \tau^2)^{\alpha_2} + \epsilon (a_3^2 + |\omega|^2 \tau^2)^{\alpha_3} \right\}^{-\nu}, \quad (3.13)$$

where $a_1, a_2, a_3, \alpha_1, \alpha_2, \alpha_3$ and $c_1, c_2$ are positive; $\epsilon \in [0, 1]$ and $\frac{d}{\alpha_1 \nu} + \frac{1}{\alpha_3 \nu} < 2$. This is a valid spectral density (see the proof in Appendix). The spectral density in (3.8) is a special
Figure 3.7: Contour plots for some nonseparable spatial-temporal covariances.
case of this more general class. If \( \alpha_1 = \alpha_3 = 1 \) and \( \frac{d}{\nu} + \frac{1}{\alpha_2 \nu} < 2 \) in (3.13), we have

\[
\int_{\mathbb{R}^d} \exp(i\omega^T x)f(\omega, \tau) d\omega
\]

\[
= \frac{\pi^{d/2} \gamma}{2^{\nu-\frac{d}{2}-1} \Gamma(\nu)} (c_1 + c_3 \tau^2)^{-\nu} \left( \frac{|x|}{\rho(\tau)} \right)^{\nu-\frac{d}{2}} K_{\nu-\frac{d}{2}}(\rho(\tau)|x|),
\]

(3.14)

where \( \rho(\tau) = \left( \frac{c_2(a_2^2 + \tau^2)/a_2^2 + c_1 a_1^2 + c_3 a_3^2}{c_1 + c_3 \tau^2} \right)^{1/2} \). So the corresponding spatial-temporal covariance \( C(x, t) \) can be quickly approximated by using the fast Fourier transformation of (3.14).

### 3.3 New models for nonseparable and nonstationary processes

In the previous section, we introduced the spectral representation of stationary spatial-temporal processes, and proposed a new class of nonseparable stationary spectral densities. Now we generalize the idea to nonstationary spatial-temporal processes.

Here, we define a generalized class for spatial-temporal nonseparability and nonstationarity. In this class, the spectral representation itself and the corresponding spectral distribution function (or spectral density) can change slowly in space and time. Let \( Z(x, t) \) be a general space-time process, \( x \in D \subseteq \mathbb{R}^d \) and \( t \in T \subseteq \mathbb{R} \), we use the following representation

\[
Z(x, t) = \int_{\mathbb{R}^d} \int_{\mathbb{R}} \exp(i\omega^T x + i\tau t) dY_{x,t}(\omega, \tau).
\]

(3.15)

We are going to assume throughout this subsection that the mean of the process is zero. We will define a model for the mean in the next section. We capture the potential lack of stationarity by allowing the spectral process \( Y \) associated with \( Z \) to change in space and time.

The general spectral representation proposed in (3.15) provides an ideal framework to model complex space-time dependent structures. Next, we discuss two flexible spatial-temporal models that correspond to processes with this type of spectral representation.
3.3.1 Mixture of local spectrums (Spectral model)

A particular case of the general representation in (3.15) is when the lack of stationarity is explained by allowing the amplitude of the spectral process $Y$ to be space-time dependent. This means

$$Z(x, t) = \int_{\mathbb{R}^d} \int_{\mathbb{R}} \exp(i\omega^T x + i\tau t)\phi_{x,t}(\omega, \tau)dY_0(\omega, \tau)$$

(3.16)

where $\phi_{x,t}(\omega, \tau)$ is a space-time dependent amplitude function, and $Y_0$ is a space-time invariant Wiener process, which satisfies the relation

$$E[Y_0(\omega, \tau)Y_0^c(\omega', \tau')] = \delta(\omega - \omega')\delta(\tau - \tau')$$

where $\delta$ is the delta Dirichlet function. This is a space-time version of the evolutionary spectrum presented by 26). We assume the functions $\phi_{x,t}(\omega, \tau)$ satisfy the condition

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}} |\phi_{x,t}(\omega, \tau)|^2d\omega d\tau < \infty$$

(3.17)

for all $x$ and $t$. The functions $\phi_{x,t}(\omega, \tau)$ must also satisfy $\phi_{x,t}(\omega, \tau) = \phi_{x,t}^c(-\omega, -\tau)$ to ensure $Z(x, t)$ is real-valued.

Then, it is easy to see that the covariance function, $C$, of the process $Z(x, t)$ is given by the formula

$$\text{cov}\{Z(x_1, t_1), Z(x_2, t_2)\} = C(x_1, t_1; x_2, t_2)$$

$$= \int_{\mathbb{R}^d} \int_{\mathbb{R}} \exp\{i\omega^T(x_1 - x_2)\} \exp\{i\tau(t_1 - t_2)\} \phi_{x_1,t_1}(\omega, \tau)\phi_{x_2,t_2}^c(\omega, \tau)d\omega d\tau$$

(3.18)

In particular

$$\text{var}\{Z(x, t)\} = C(x, t; x, t) = \int_{\mathbb{R}^d} \int_{\mathbb{R}} |\phi_{x,t}(\omega, \tau)|^2d\omega d\tau$$

(3.19)

so that condition (3.17) is necessary for the variance of $Z(x, t)$ to be finite at all $x$ and $t$, i.e. for the existence of a covariance function $C(x_1, t_1; x_2, t_2)$.

The representation in (3.16) may be interpreted as a representation of the process $Z$ in the form of a superposition of sinusoidal oscillations with different frequencies $\psi = (\omega, \tau)$ and random amplitudes $\phi_{x,t}(\omega, \tau)$ varying over space and time, i.e. $Z(x, t)$ is a
superposition of amplitude modulated random oscillations. According to this interpretation, (3.19) describes the distribution of the “total power” of the process $Z(x, t)$ at location $x$ and time $t$ over the frequencies $(\omega, \tau)$, hence the contribution from the frequency $(\omega, \tau)$ is $|\phi_{x,t}(\omega, \tau)|^2 d\omega d\tau$. Therefore the function $F_{x,t}(\omega, \tau)$ defined by the relation

$$dF_{x,t}(\omega, \tau) = |\phi_{x,t}(\omega, \tau)|^2 d\omega d\tau$$

(3.20)

will be called the spatial-temporal spectral distribution function of the process $Z$, and $f_{x,t}(\omega, \tau) = |\phi_{x,t}(\omega, \tau)|^2$ is the spatial-temporal spectral density of $Z$.

There exist different representations of the form (3.15) for a spatial-temporal process $Z$, each representation is based on a different family of $\phi_s(\psi)$ functions, where we write $s = (x, t)$ and $\psi = (\omega, \tau)$ to simplify the notation. This problem is similar to the selection of a basis for a vector space. Apart from that, it would not be physically meaningful to interpret $\psi$ as the frequency in all cases. In the physical theory of oscillations the function $A_s(\psi) = \phi_s(\psi)e^{is^T\psi}$ is said to describe the amplitude modulated oscillation of frequency $\psi$ only if the “amplitude” $\phi_s(\psi)$ is slowly varying compared to $e^{is^T\psi}$ function, i.e. if the Fourier transform of $\phi_s(\psi)$ as a function of $s$ includes mainly frequencies much lower than $\psi$. It is even often assumed that this transform must be concentrated in a neighborhood of zero frequency. We restrict the permissible variability of the function $\phi_s(\psi)$ of $s$ by considering only functions $\phi_s(\psi)$ that admit a generalized Fourier representation

$$\phi_s(\psi) = \int_{\mathbb{R}^{d+1}} e^{is^T\sigma} dH(\sigma)$$

(3.21)

with $|dH(\sigma)|$ having its maximum at $\sigma = 0$ for any fixed $\psi$. This condition guarantees that the Fourier transform of $\phi_s(\psi)$, as a function of $s$, includes mainly frequencies much lower than any $\psi$, as has been suggested by Priestley in the time series context. Since $\phi_s(\psi)$ is a slowly varying function of space and time, it is clear that the process $Z$ may be regarded as being “approximately stationary” within subregions in our spatial-temporal domain. If, however, we examine the behavior of $Z$ within two subregions which are sufficiently far apart, we could find that although $Z$ is practically stationary in both subregions, the spectral distribution function of the two “portions” of $Z$ will, in general, be different (i.e., the spectral distribution of the power of $Z$ varies on space and time). Since the functions $\phi_s(\psi) = 1$ clearly satisfy the conditions to be imposed on $\phi_s(\psi)$, the representation (3.15)
certainly includes all the spatial-temporal stationary processes having a finite variance.

A particular case is when $\phi$ is a separable function of spatial frequency and temporal frequency, i.e.

$$\phi_s(\psi) = \phi^{(1)}_x(\omega)\phi^{(2)}_t(\tau).$$

(3.22)

When $\phi$ is of the form (3.22) the spatial-temporal process is separable, since

$$\text{cov}\{Z(x_1, t_1), Z(x_2, t_2)\}$$

$$= \int_{\mathbb{R}^{d+1}} \exp\{i\omega^T(x_1 - x_2)\} \exp\{i\tau^T(t_1 - t_2)\} \phi^{(1)}_{x_1}(\omega)\phi^{(2)}_{t_1}(\tau)\phi^{(1)}_{x_2}(\omega)^c\phi^{(2)}_{t_2}(\tau)^c d\omega d\tau$$

$$= \int_{\mathbb{R}^d} \exp\{i\omega^T(x_1 - x_2)\} \phi^{(1)}_{x_1}(\omega)(\phi^{(1)}_{x_2}(\omega))^c \int_{\mathbb{R}} \exp\{i\tau^T(t_1 - t_2)\} \phi^{(2)}_{t_1}(\tau)(\phi^{(2)}_{t_2}(\tau))^c d\tau$$

$$= C^{(1)}(x_1, x_2)C^{(2)}(t_1, t_2).$$

(3.23)

where $C^{(1)}$ and $C^{(2)}$ are spatial and temporal covariance functions.

We propose a more general model for $\phi$ that has the separable model in (3.22) as a particular case. We model $\phi$ as a mixture of local spectral (amplitude) functions,

$$\phi_s(\psi) = \sum_{j=1}^k K(s - s_j) \phi_{s_j}(\psi)$$

(3.24)

where each $\phi_{s_j}(\psi)$ function explains the spatial-temporal structure of $Z$ in a neighborhood of $s_j$. $K(s - s_i)$ is a kernel function.

Locally (in a neighborhood of $s_j$), we use the nonseparable parametric model proposed in (3.8) for $\phi_{s_j}$, that has a separable model as a particular case. We can write $\phi_{s_j}(\omega, \tau)$ as

$$\phi_{s_j}(\omega, \tau) = \gamma_j(\alpha_j^2\beta_j^2 + \beta_j^2\|\omega\|^2 + \alpha_j^2|\tau|^2 + \epsilon_j\|\omega\|^2|\tau|^2)^{-(\nu_j+d+1)/2}.$$

(3.25)

Again, the parameter $\alpha_j^{-1}$ explains the rate of decay of the spatial correlation component. For the temporal correlation, the rate of decay is explained by $\beta_j^{-1}$, and $\gamma_j$ is a scale parameter. The parameter $\nu_j > 0$ measures the degree of smoothness of the process $Z$ at $s_j$, the higher the value of $\nu_j$ the smoother $Z$ would be. If $\epsilon_j = 0$ we have a $d+1$ dimensional
Matérn type model with different spatial and temporal ranges, which takes into account the change in units from the spatial to the temporal domain. If \( \epsilon_j = 1 \) we have a separable model,

\[
\phi_{s_j}(\psi) = \gamma_j (\alpha_j^2 + \|\omega\|^2)^{-(\nu_j + d)/2} (\beta_j^2 + |\tau|^2)^{-(\nu_j + d)/2}.
\]

Then, the corresponding (local) covariance is separable (as in (3.23)).

Hence, the spectral model for a nonseparable and nonstationary spatial-temporal process \( Z(s) = Z(x, t) \) is given by

\[
Z(s) = \int_{\mathbb{R}^{d+1}} \exp(i\psi^T s) \left( \sum_{j=1}^{k} K(s - s_j)\phi_{s_j}(\psi) \right) dY_0(\psi), \tag{3.26}
\]

where \( \phi_{s_j}(\psi) = \phi_{s_j}(\omega, \tau) \) which is defined in (3.25). The value of \( k \) is the number of local spectral functions, and it can be chosen using scientific knowledge. \( s_j \) corresponds to the center of the subregion which gives the \( j \)-th local spectral. \( K(s - s_j) \) represents the contribution from the \( j \)-th local spectral. If the kernel function \( K \) is a separable function of space and time, and \( \epsilon_j = 1 \) for \( j = 1, \cdots, k \), the process \( Z(s) \) is separable and nonstationary.

### 3.3.2 Mixture of local space-time models (Spatial-temporal domain)

The representation for a general space-time process given in (3.26) gives

\[
Z(s) = \sum_{j=1}^{k} \left\{ \int_{\mathbb{R}^{d+1}} \exp(i\psi^T s)K(s - s_j)\phi_{s_j}(\psi)dY_0(\psi) \right\}
\]

\[
= \sum_{j=1}^{k} \left\{ K(s - s_j)\int_{\mathbb{R}^{d+1}} \exp(i\psi^T s)\phi_{s_j}(\psi)dY_0(\psi) \right\}
\]

where \( s = (x, t) \) and \( \psi = (\omega, \tau) \). It corresponds to a mixture of space-time stationary processes

\[
Z(s) = \sum_{j=1}^{k} K(s - s_j)Z_j^*(s) \tag{3.27}
\]

where each \( Z_j^* \), given by \( \int_{\mathbb{R}^{d+1}} \exp(i\psi^T s)\phi_{s_j}(\psi)dY_0(\psi) \), is a stationary process, and \( K \) is the kernel function in (3.26). The local stationary processes in (3.27) are correlated to each other, and the cross-covariance between \( Z_j^* \) and \( Z_l^* \) is given by

\[
\text{cov}\{Z_j^*(s_1), Z_l^*(s_2)\} = \int_{\mathbb{R}^{d+1}} \exp(i\psi^T (s_1 - s_2))\phi_{s_j}(\psi)\phi_{s_l}(\psi)d\psi. \tag{3.28}
\]
Another particular case of the general representation in (3.15) is when the spectral process \( Y_s(\psi) = Y_{x,t}(\omega, \tau) \) is modeled as a mixture of orthogonal stationary spectral processes \( Y_i \), with spectral density \( f_i \) for \( i = 1, \ldots, k \), that explain the space-time dependence structure in subregions of stationarity \( S_1, \ldots, S_k \),

\[
Y_s(\psi) = \sum_{i=1}^{k} K(s - s_i)Y_i(\psi)
\]  

(3.29)

where each \( Y_i \) explains the spatial-temporal structure of \( Z \) in a stationary subregion \( S_i \) with centroid \( s_i \). Subregion \( S_i \), \( i = 1, \ldots, k \), is a partition of spatial-temporal domain of interest, \( D \times T \). By partition, we mean \( \bigcup_{i=1}^{k} S_i = D \times T \) and \( S_i \cap S_j = \emptyset \) for \( \forall i \neq j \). Thus, we can write the process \( Z \) in terms of uncorrelated processes \( Z_i \), for \( i = 1, \ldots, k \):

\[
Z(x, t) = Z(s) = \int_{\mathbb{R}^{d+1}} \exp(i\psi^T s) d\left\{ \sum_{i=1}^{k} K(s - s_i)Y_i(\psi) \right\}
\]

\[
= \sum_{i=1}^{k} \left\{ K(s - s_i) \int_{\mathbb{R}^{d+1}} \exp(i\psi^T s) dY_i(\psi) \right\}
\]

\[
= \sum_{i=1}^{k} K(s - s_i)Z_i(s).
\]  

(3.30)

For each \( i, i = 1, \ldots, k \), \( Z_i \) is the stationary spatial-temporal process associated with \( Y_i \). The difference between this representation and (3.27) is that the local stationary processes in (3.27) are correlated to each other, and in (3.30) they are uncorrelated.

\( K(s - s_i) \)'s are kernel functions over space and time. They can be interpreted as weights that depend on the distance from subregion \( S_i \) (or center of subregion \( s_i \)) to the point \( s \). These weights act as windows or filters that give more importance to the process \( Z_i(s) \) for \( s \in S_i \) and less importance to the point \( s \) not in subregion \( S_i \). Assuming that \( K(s - s_i) \approx 1 \) when \( s \in S_i \) and \( K(s - s_j) \approx 0 \) when \( s \in S_j \) and \( i \neq j \), the overall process \( Z(s) \) at \( s \in S_i \) is approximated by the stationary process \( Z_i(s) \).

If a space-time process is stationary in time but nonstationary in space, then we
can rewrite (3.30) as
\[ Z(x, t) = \sum_{i=1}^{k} K(x - x_i)Z_i(x, t). \]  \hspace{1cm} (3.31)

In this case, \( Z_i \) is a local stationary spatial-temporal process within subregion \( D_i \times T \). \( \{D_1 \times T, \cdots, D_k \times T\} \) is a partition of the spatial-temporal domain \( D \times T \). \( K(x - x_i) \) is a weight centered at \( x_i \), the centroid of \( D_i \).

Similarly, if a space-time process is stationary in space but nonstationary in time, we can have the following representation for (3.30),
\[ Z(x, t) = \sum_{i=1}^{k} K(t - t_i)Z_i(x, t). \]  \hspace{1cm} (3.32)

Here, \( Z_i \) is a local stationary spatial-temporal process within subregion \( D \times T_i \). \( \{D \times T_1, \cdots, D \times T_k\} \) is a partition of spatial-temporal domain \( D \times T \). \( K(t - t_i) \) is a weight function centered at \( t_i \) the centroid of \( T_i \).

The models in (3.31) and (3.32) are two special cases for the general model in (3.30). The value of \( k \), the number of stationary subregions, can be selected by using Akaike Information Criterion (AIC) or Schwarz Bayesian Information Criterion (BIC). Also some scientific information can help us to determine the value of \( k \).

Given a process \( Z \) with the spectrum given by (3.29), the corresponding covariance function for a process is
\[ \text{cov}\{Z(s_1), Z(s_2)\} = \sum_{i=1}^{k} K(s_1 - s_i)K(s_2 - s_i)C_i(s_1 - s_2) \]  \hspace{1cm} (3.33)

where \( s_i = (x_i, t_i) \), and each \( C_i \) is a stationary covariance (corresponding to the spectral process \( Y_i \)) that explains the space-time dependency in a stationary subregion \( S_i \).

The proposed parametric model in (3.11) can be employed for \( C_i \),
\[ C_i(x, t) = \frac{\sigma_i^2}{2^{\nu_i-1}\Gamma(\nu_i)} \left( \frac{\|x - \rho_i t\|}{r_i} \right)^{\nu_i-1} K_{\nu_i} \left( \frac{\|x - \rho_i t\|}{r_i} \right). \]  \hspace{1cm} (3.34)

This parametric model for \( C_i \) corresponds to a \( d + 1 \) dimensional Matérn type covariance with an extra parameter \( \rho_i \) that can be interpreted as a conversion factor between the
units in the space and time domains. A more general model for \( C_i \) would be given by the Fourier transform of \( f_i \) in (3.8), as discussed in previous section.

We can also represent \( Z(x, t) \) in (3.30) as a convolution of independent stationary spatial-temporal processes,

\[
Z(s) = \int_{\mathbb{R}^{d+1}} K(s - u)Z_{\theta(u)}(s)du. \tag{3.35}
\]

The stochastic integral (3.35) is defined as a limit (in mean square) of approximating Riemann sums (6). In representation (3.35), \( K(\cdot) \) is a kernel function over space and time. \( \theta(u) \) is the parameter for the local stationary process \( Z_{\theta(u)} \). In model (3.35), if the kernel function \( K(\cdot) \) is sharply peaked and \( \theta(u) \) varies slowly over space and time, then for \( s = (x, t) \) near \( u \), the process \( Z \) looks like a stationary process with parameter \( \theta(u) \). On the other hand, because \( \theta(u) \) may vary significantly over the spatial-temporal domain, it also allows significant nonstationary. Kernel functions can be represented as a quadratic weight function with compact support. The shape of the process \( Z \) is represented by the parameter \( \theta(u) \), which reflects the lack of stationarity. Thus we need to choose the bandwidth of weight function as small as possible to preserve the shape. The bandwidth might change with location/time. The corresponding covariance for a process in (3.35) is

\[
cov\{Z(s_1), Z(s_2)\} = \int_{\mathbb{R}^{d+1}} K(s_1 - u)K(s_2 - u)C_{\theta(u)}(s_1 - s_2)du.
\]

### 3.4 Spatial-temporal Trend

In the previous section we have introduced some new models for nonseparable and nonstationary spatial-temporal processes with zero mean. In this section we will present spatial-temporal models for the large scale structure, i.e. trend, using covariates with dynamic coefficients.

In general, a spatial-temporal process is given by

\[
Z(x, t) = \mu(x, t) + \epsilon(x, t)
\]
where \( x \in D \subseteq \mathbb{R}^d \) and \( t \in T \subseteq \mathbb{R} \). The function \( \mu(x, t) \) represents the trend surface. The residual term \( \epsilon(x, t) \) is a zero mean space-time correlated error that explains the spatial-temporal short scale structure. Statistical models proposed in the previous section, for example, the model in (3.31), can be used for \( \epsilon(x, t) \).

Here, we represent the large scale structure (e.g., trend surface) of \( Z \) using a space-time dynamic statistical model:

\[
\mu(x, t) = \sum_{i=1}^{m} \beta_i(x, t)f_i(x, t),
\]

where \( \{f_i\}_i \) are \( m \) covariates (e.g. sine and cosines and geographic data) of interest with coefficients \( \beta_i \) that vary in space and time.

We model the dynamic coefficients \( \beta_i \) using a hierarchical model in terms of an overall time component \( \gamma_{i,t} \) and a space-time process \( \gamma_i(x, t) \),

\[
\beta_i(x, t) = \gamma_{i,t} + \gamma_i(x, t)
\]

where

\[
\gamma_{i,t} - \mu_i = \rho_{1i}(\gamma_{i,t-1} - \mu_i) + u(t),
\]

and

\[
\gamma_i(x, t) - \mu_i(x) = \rho_{2i}[\gamma_i(x, t-1) - \mu_i(x)] + \eta(x, t).
\]

\( \eta \) and \( u \) are independent white noises. Here we present both \( \gamma_{i,t} \) and \( \gamma_i(x, t) \) in the form of a first-order autoregression model. More general forms can be used for them.

By letting the coefficients \( \beta_i \) to be random, the correlations among the coefficients are added to the final (marginal) models. For the identifiability, one possible solution is to model the process \( \eta \) as a mean-zero spatial-temporal process and treat the residual term \( \epsilon \) as a white noise. The other possible solution is to simplify the dynamic model for \( \beta_i \) and let the residual term \( \epsilon \) to take into account the spatial-temporal structure.

Particularly we simplify the model for the large scale structure by representing \( \beta_i \) as two parts. One part is an overall effect, and the other part is the variation within each
subregion. The subregions come from the model for \(\epsilon(x,t)\). Therefore we represent the trend surface as

\[
\mu(x,t) = \sum_{i=1}^{m}[\gamma_i + \gamma_{ij}]f_i(x,t), \quad \text{if } x \in S_j,
\]

where \(S_j, j = 1, \ldots, k\), are subregions defined in the model of \(\epsilon(x,t)\).

Hence, a simple version of the model for \(\{Z(x,t) : x \in D \subset \mathbb{R}^d, t \in T \subset \mathbb{R}\}\) is given by

\[
Z(x,t) = \mu(x,t) + \epsilon(x,t)
\]

with

\[
\mu(x,t) = \sum_{i=1}^{m}[\gamma_i + \gamma_{ij}]f_i(x,t), \quad \text{if } x \in D_j
\]

and

\[
\epsilon(x,t) = \sum_{i=1}^{k}K(x - x_i)Z_i(x,t).
\]

\(Z_i\) is a zero-mean stationary process, which explains the spatial-temporal structure within subregion \(D_i \times T\). \(\{D_1 \times T, \ldots, D_k \times T\}\) is a partition of spatial-temporal domain \(D \times T\). \(K(x - x_i)\) is a weight function centered at \(x_i\), the centroid of \(D_i\). \(\{f_i(x,t)\}_i\) are \(m\) known covariates.

In this chapter, we proposed a new class for nonseparable and nonstationary spatial-temporal processes in general. It consists of two parts. The first part is the trend surface, which explains the large scale spatial-temporal structure. We model it as a function of covariates with dynamic coefficients varying in space and time. The second part is the residual term, which explains the small scale spatial-temporal structure. For the residual term, we model this nonstationary process as a mixture of local orthogonal stationary spatial-temporal processes. The empirical separability test can be used to better understand the covariance structure of local stationary process, and proposed nonseparable covariances are available for the local nonseparable stationary processes.
Chapter 4

Combining Spatial-Temporal Data

In many circumstances, especially in environmental applications, data from different sources always have different spatial/temporal resolutions. For example, the observed wind speed represents the wind speed at a specific location, but the wind speed from MM5 model output is for a grid cell, so it is an average over a grid cell. Therefore, we need a method to combine the data with different spatial/temporal resolutions. Bayesian methods are ideal for this problem. It is intuitive to model the data in terms of the underlying truth, because there are measurement errors for observed data and different sources of errors and bias for numerical model output. Consequently all the models can be estimated in a Bayesian way. 8) suggest a general modeling framework which is shown in Figure 4.1.

Throughout this Chapter, we assume the data from different sources have different spatial resolution. If they also have different temporal resolution, the similar methodology will tackle the problem.

4.1 Statistical models for disparate spatial-temporal data

Here we consider data from two different sources, one is point observations and the other is numerical model output. As we mentioned, we do not consider the observed
Figure 4.1: The general modeling framework.
data as “ground truth”, because there is measurement error. Thus we assume there is an unobserved underlying spatial-temporal field $Z(x, t)$, where $Z(x, t)$ measures the quantity of interest at point $(x, t)$. We have an observation $\tilde{Z}(x, t)$ at $(x, t)$, for example, the observed wind speed at 10am on July 21, 2002 at the Ocean City monitoring site, and we assume that

$$\tilde{Z}(x, t) = Z(x, t) + e(x, t), \quad (4.1)$$

where $e(x, t) \sim N(0, \sigma_e^2)$ represents the measurement error at point $(x, t)$. The process $e(x, t)$ is independent of $Z(x, t)$.

The true underlying process $Z(x, t)$ is a spatial-temporal process with a nonstationary nonseparable covariance. Therefore, the nonseparable and nonstationary spatial-temporal model proposed in previous chapter can be used to model the true process $Z(x, t)$ with mean $\mu(x, t)$ and covariance $\text{cov}\{Z(x_1, t_1), Z(x_2, t_2)\} = C(x_1, x_2; t_1, t_2)$.

We model the numerical model output, $\tilde{Z}(x, t)$, e.g., wind speed from MM5 model output, as follows:

$$\tilde{Z}(x, t) = a(x, t) + b(x, t)Z(x, t) + \delta(x, t). \quad (4.2)$$

Here the parameter function $a(x, t)$ measures the additive bias in the numerical model, which might change with location and time, and the parameter function $b(x, t)$ measures the multiplicative bias in the numerical model. The process $\delta(x, t) \sim N(0, \sigma_\delta^2)$ explains the random deviation with respect to the underlying true wind process $Z(x, t)$. The process $\delta(x, t)$ is independent of $Z(x, t)$ and $e(x, t)$, where $e(x, t)$ is the error term for observed data.

Since the output from the numerical model (e.g., MM5) are not point measurements, but are average estimates in grid cells $B_1, \ldots, B_m$ that cover the spatial domain, $D$, we have:

$$\tilde{Z}(B_i, t) = \frac{1}{|B_i|} \int_{B_i} a(x, t)dx + \frac{1}{|B_i|} \int_{B_i} b(x, t)Z(x, t)dx + \frac{1}{|B_i|} \int_{B_i} \delta(x, t)dx \quad (4.3)$$

for $i = 1, \ldots, m$.

We assume that the function $a(x, t)$ is a polynomial based on point $(x, t)$ with a vector of coefficients, $a_0$, and $b(x, t)$ is an unknown constant term $b$. Then, we can rewrite
model (4.3) as:
\[
\tilde{Z}(B_i, t) = \frac{1}{|B_i|} \int_{B_i} a(x, t) dx + \frac{b}{|B_i|} \int_{B_i} Z(x, t) dx + \frac{1}{|B_i|} \int_{B_i} \delta(x, t) dx. \tag{4.4}
\]

### 4.2 Change of support

Change of support occurs when we combine the data with different spatial resolutions. Suppose we have point measurements \( \tilde{Z} \) at \( \{ (x_i, t_i), i = 1, \cdots, n \} \), and numerical model output \( \tilde{Z} \) at \( \{ (B_j, t_j), j = 1, \cdots, m \} \).

The covariance for the grid cell averages is defined as:
\[
\text{cov}\{ Z(B_i, t_i), Z(B_j, t_j) \} = \frac{\int_{B_i} \int_{B_j} C(v_1, v_2; t_i, t_j) dv_1 dv_2}{|B_i||B_j|}. \tag{4.5}
\]

If \( B_i = x_i \) (a point), the covariance is defined as:
\[
\text{cov}\{ Z(x_i, t_i), Z(B_j, t_j) \} = \frac{\int_{B_j} C(x_i, v; t_i, t_j) dv}{|B_j|}. \tag{4.6}
\]

Therefore the covariance for numerical model output is
\[
\text{cov}\{ \tilde{Z}(B_i, t_i), \tilde{Z}(B_j, t_j) \} = \frac{b^2 \int_{B_i} \int_{B_j} C(v_1, v_2; t_i, t_j) dv_1 dv_2}{|B_i||B_j|} + \frac{\sigma^2}{|B_i|} 1\{i=j\}.
\]

The covariance between observation and numerical model output is
\[
\text{cov}\{ \tilde{Z}(x_i, t_i), \tilde{Z}(B_j, t_j) \} = \frac{b \int_{B_j} C(x_i, v; t_i, t_j) dv}{|B_j|}.
\]
4.3 Statistical assessment of numerical model performance

The objective comparison between numerical model output and observed data provides a mean for assessing numerical model performance. But it is not reasonable to compare them directly, since the numerical model output usually has a different spatial/temporal scale compared to the observed data. Therefore we statistically assess the numerical model performance by comparing the observed values with their posterior predictive distributions given the numerical model output.

Let $\hat{Z} = (\hat{Z}(x_1, t_1), \ldots, \hat{Z}(x_n, t_n))^T$ and $\tilde{Z} = (\tilde{Z}(B_1, t_1), \ldots, \tilde{Z}(B_m, t_m))^T$. For the numerical model evaluation, we simulate the observed values from the numerical model output given the numerical model output, which is to simulate from the following posterior distribution:

$$P(\hat{Z} | \tilde{Z}, a_0 = 0, b = 1).$$

Now we deduce the joint distribution of $\hat{Z}$ and $\tilde{Z}$ conditioning on the parameters $\theta$ in models (4.1) and (4.4).

$$\begin{pmatrix} \hat{Z} \\ \tilde{Z} \end{pmatrix} \sim N \left( \begin{pmatrix} \hat{\mu} \\ \tilde{\mu} \\ a + b \tilde{\mu} \end{pmatrix}, \begin{pmatrix} \hat{\Sigma} & \hat{\Sigma} & \hat{\Sigma} \\ \hat{\Sigma} & \hat{\Sigma} & \hat{\Sigma} \\ \hat{\Sigma} & \hat{\Sigma} & \hat{\Sigma} \end{pmatrix} \right),$$

where

$$\hat{\mu} = (\mu(x_1, t_1), \ldots, \mu(x_n, t_n))^T,$$

$$\tilde{\mu} = (\int_{B_1} a(v, t_1) dv, \ldots, \int_{B_m} a(v, t_m) dv)^T,$$

and

$$\hat{\mu} = (\int_{B_1} \mu(v, t_1) dv, \ldots, \int_{B_m} \mu(v, t_m) dv)^T.$$
For $i, j = 1, \ldots, n,$

$$
\Sigma_{ij} = \text{cov}\{\hat{Z}(x_i, t_i), \hat{Z}(x_j, t_j)\}
= C(x_i, x_j; t_i, t_j) + 1_{(i=j)} \sigma_e^2;
$$

for $i = 1, \ldots n$ and $j = 1, \ldots m,$

$$
\Sigma_{i,n+j} = \text{cov}\{\hat{Z}(x_i, t_i), \hat{Z}(B_j, t_j)\}
= b \int_{B_j} C(x_i, v; t_i, t_j) d\nu
= \frac{b \int_{B_j} C(x_i, v; t_i, t_j) d\nu}{|B_j|};
$$

for $i, j = 1, \ldots, m,$

$$
\Sigma_{n+i,n+j} = \text{cov}\{\hat{Z}(B_i, t_i), \hat{Z}(B_j, t_j)\}
= \frac{b^2 \int_{B_i} \int_{B_j} C(v_1, v_2; t_i, t_j) d\nu_1 d\nu_2}{|B_i||B_j|} + \frac{\sigma_e^2}{|B_i|} 1_{(i=j)}.
$$

We also use the following result from multi-variate analysis (21). If

$$
X_1
X_2
\sim N \left( \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right),
$$

then the conditional distribution of $X_1$ given $X_2, [X_1|X_2],$ is normal with mean

$$
\mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (X_2 - \mu_2),
$$

and variance

$$
\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}.
$$

Therefore, the conditional distribution of $\hat{Z}$ given $\tilde{Z}$ is normal with mean

$$
\hat{\mu} + \hat{\Sigma} \hat{\Sigma}^{-1} (\tilde{Z} - \tilde{\alpha} - b \tilde{\mu}),
$$

and variance

$$
\hat{\Sigma} - \hat{\Sigma} \hat{\Sigma}^{-1} \hat{\Sigma}.
$$

Hence,

$$
[\hat{Z}|\tilde{Z}, a_0 = 0, b = 1] \sim \text{Normal}(\hat{\mu} + \hat{\Sigma} \hat{\Sigma}^{-1} (\tilde{Z} - \tilde{\mu}), \hat{\Sigma} - \hat{\Sigma} \hat{\Sigma}^{-1} \hat{\Sigma}).
$$
4.4 Bayesian spatial-temporal prediction

Furthermore, our objective is to predict the quantity of interest $Z$ at location $x_0$ and time $t_0$ given the data, $\hat{Z}$ and $\tilde{Z}$, e.g., to obtain wind speed fields given observed wind data and MM5 model output. Thus we need the conditional distribution of $Z(x_0, t_0)$ given the data.

This is straightforward by applying the result of $[X_1 | X_2]$ in previous section with $X_1 = Z(x_0, t_0)$ and $X_2 = Z = (\hat{Z}, \tilde{Z})^T$. Then

$$\Sigma_{11} = \text{cov}\{Z(x_0, t_0), Z(x_0, t_0)\} = C(x_0, x_0; t_0, t_0) = \sigma_0^2,$$

$$\Sigma_{22} = \Sigma,$$

and

$$\Sigma_{21}^T = \Sigma_{12} = \text{cov}\{Z(x_0, t_0), Z\} = \tau,$$

where $\tau$ is a $(n + m)$ vector with components. For $i = 1, \ldots, n$,

$$\tau_i = \text{cov}\{Z(x_0, t_0), \tilde{Z}(x_i, t_i)\} = C(x_0, x_i; t_0, t_i);$$

and for $j = 1, \ldots, m$,

$$\tau_{n+j} = \text{cov}\{Z(x_0, t_0), \hat{Z}(B_j, t_j)\} = \int_{B_j} b \int_{B_j} C(x_0, v; t_0, t_j) dv.$$

Therefore, the conditional distribution of $Z(x_0, t_0)$ given $(\hat{Z}, \tilde{Z})$ and all parameters $\theta$, i.e. $P(Z(x_0, t_0)|Z, \theta)$, is normal with mean

$$\mu(x_0, t_0) + \tau^T \Sigma^{-1} (Z - \mu)^T,$$

where

$$\mu = (\hat{\mu}, \tilde{\mu} + b\tilde{\mu})^T.$$
and variance

\[ \sigma_0^2 - \tau^T \Sigma^{-1} \tau. \]

Hence, the posterior predictive distribution of \( Z(x_0, t_0) \) given the observations \( Z = (\hat{Z}, \tilde{Z})^T \) is

\[ P(Z(x_0, t_0)|Z) \propto \int P(Z(x_0, t_0)|Z, \theta)P(\theta|Z)d\theta, \tag{4.9} \]

where \( \theta \) is a collection of all parameters.

The Gibbs sampling approach is used to simulate \( N \) values of \( \theta \) from the posterior distribution of the vector parameter \( \theta \). Thus the posterior predictive distribution is approximated by

\[ P(Z(x_0, t_0)|Z) = \frac{1}{N} \sum_{i=1}^{N} P(Z(x_0, t_0)|Z, \theta^{(i)}). \tag{4.10} \]

As defined, \( \theta \) is a collection of all parameters, \( \theta = (\theta_1, \theta_2, \theta_3) \). \( \theta_1 \) are the parameters for point observations given underlying true process, and \( \theta_2 \) are the parameters for numerical model output given underlying true process. \( \theta_3 \) is a collection of all parameters used to model underlying true process. The posterior distribution, \( [\theta|\hat{Z}, \tilde{Z}] \) is

\[ [\theta|\hat{Z}, \tilde{Z}] \propto \int [\hat{Z}, \tilde{Z}, Z|\theta]|\theta|dZ. \]

We know that

\[ [\hat{Z}, \tilde{Z}, Z|\theta] = [\hat{Z}][\tilde{Z}, Z, \theta][\hat{Z}|Z, \theta][Z|\theta] \]

\[ = [\hat{Z}|Z, \theta][\tilde{Z}, Z, \theta][Z|\theta] \]

\[ = [\hat{Z}|Z, \theta_1][\tilde{Z}, Z, \theta_2][Z|\theta_3] \]

Therefore, multiple-stage Gibbs sampling approach is used here. We alternate between the parameter that measure the lack of stationary \( \theta_3 \) (Stage 1), and the parameters that assess the measurement error in the observed data \( \theta_1 \) (Stage 2) and the bias in the numerical model output \( \theta_2 \) (Stage 3). The algorithm is as follows.
Gibbs sampling: Stage 1

We obtain the conditional posterior for the parameters $\theta_3$ conditioning on the values of $Z$ that are updated in stage 4. The posterior of $\theta_3$ will be completely specified once we define the prior for $\theta_3$, because we have

$$[Z|\theta_3] \sim \text{Gaussian},$$

where the brackets $[\cdot]$ are used to denote densities.

The underlying spatial-temporal process $Z$ is model as two parts, as proposed in Chapter 3. The trend surface is the model in (3.37) with $g$ known covariates. The residual term is represented as (3.30). Assume the kernel function $K$ is known and the number of subregions, $k$, is known too. After performing the empirical separability test within each subregion, we know that nonseparable covariances are appropriate for subregions $S_1, \cdots, S_l$, and the covariances in model (3.34) are used for them. For separable covariances, we fit both spatial and temporal covariances with the traditional Matérn model, which is given in (3.3). Therefore, $\theta_3 = (\beta_1, \beta_{21}(s), \cdots, \beta_{2g}(s), \sigma^2(s), r_s(s), r_t(s), \rho(s), \nu_s(s), \nu_t(s))$. We assume the parameters are independent of each other. The parameter $\beta_1$ is a $g$-vector parameter for the overall effect in the mean model. The parameters $\{\beta_{2i}(s)\}_{i}$ are the parameters which explains the variation in mean model. The parameter $\sigma^2(s)$ represents the variance. The parameter $r_s(s)$ explains the spatial range, and $r_t(s)$ is temporal range for separable covariance function. $\rho(s)$ is the conversion factor to take into account the change of units between space and time. It is the new parameter in our nonseparable covariance model. For separable covariance, the smoothness parameter for temporal covariance is $\nu_t(s)$. $\nu_s(s)$ is the smoothness parameter for a nonseparable covariance and the spatial smoothness parameter for separable covariance. All the parameters in $\theta_3$ change over space except $\beta_1$.

We model the parameter, which changes over space, as a spatial process with mean and the Matérn spatial covariance. For example, we model $\sigma^2(s)$ as

$$\log(\sigma^2(s)) = a_{\sigma} + b_{\sigma}(s) + \epsilon_{\sigma}(s),$$

where $\epsilon_{\sigma}(s)$ is a zero mean spatial process with the Matérn covariance, which has parameters $(\text{sill}_\sigma, \text{range}_\sigma, \nu_\sigma)$. To model $\log(\sigma^2(s))$ is to ensure the value of $\sigma^2$ is
positive. The priors for the parameters in the model of \(
\log(\sigma^2(s))\) are specified as in Table 4.1. The priors for \(\{\beta_2(s)\}_i, r_s(s), r_t(s), \rho(s), \nu_s(s)\) and \(\nu_t(s)\) are specified in a manner similar to \(\sigma^2(s)\). For \(\beta_1\), it is a \(g\)-vector of unknown parameters with a flat prior.

Therefore, the conditional posterior of \(\theta_3\) is proportional to \([Z|\theta_3][\theta_3]\).

**Gibbs sampling: Stage 2**

We obtain the conditional posterior distributions of \(\theta_1\). \(\theta_1\) is the parameter for measurement error of observed data, \(\sigma^2_e\). The posterior of \(\sigma^2_e\) given the values of \(\hat{Z}\) and \(Z\) can easily be obtained by regression because we know

\[
\hat{Z}(x, t) = Z(x, t) + e(x, t),
\]

where \(\sigma^2_e\) is the variance of \(e(x, t)\), and \(Z(x, t)\) is independent of \(e(x, t)\). We get

\[
[\hat{Z}(x, t)|Z(x, t), \sigma^2_e] \sim \text{Normal}(Z(x, t), \sigma^2_e).
\]

Then, the posterior of \(\sigma^2_e\) is proportional to

\[
[\hat{Z}(x_1, t_1), \ldots, \hat{Z}(x_n, t_n)|Z(x_1, t_1), \ldots, Z(x_n, t_n), \sigma^2_e][\sigma^2_e],
\]

where

\[
[\sigma^2_e] \sim \text{IG}(c, 2).
\]

**Gibbs sampling: Stage 3**

<table>
<thead>
<tr>
<th>Parameter in (\sigma^2(s))</th>
<th>Prior</th>
<th>Hyperprior</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a_\sigma)</td>
<td>Uniform/Normal((\mu_a, \sigma_a))</td>
<td>(\mu_a) is Normal</td>
</tr>
<tr>
<td>(b_\sigma)</td>
<td>Uniform/Normal((\mu_b, \sigma_b))</td>
<td>(\mu_b) is Normal</td>
</tr>
<tr>
<td>(sill_\sigma)</td>
<td>(1/sill_\sigma)</td>
<td></td>
</tr>
<tr>
<td>(range_\sigma)</td>
<td>(\text{IG}(a, 2))</td>
<td></td>
</tr>
<tr>
<td>(\nu_\sigma)</td>
<td>(\text{IG}(b, 2))</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1: Priors for the parameters in \(\sigma^2(s)\)
Similarly, the posterior of $\theta_2$, which includes $a_0$, $b$, and $\sigma_\delta^2$, given the values of $\tilde{Z}$ and $Z$, can easily be computed because we have

$$\tilde{Z}(B_j, t_j) = \int_{B_j} a(v, t_j)dv + b \int_{B_j} Z(v, t_j)dv + \int_{B_j} \delta(v, t_j)dv,$$

where $\sigma_\delta^2$ is the variance of $\delta(v, t)$, and $Z(x, t)$ is independent of $\delta(x, t)$. We get

$$[\tilde{Z}(B_j, t_j)|Z(B_j, t_j), a_0, b, \sigma_\delta^2] \sim \text{Normal}\left(\int_{B_j} a(v, t_j)dv + bZ(B_j, t_j), \frac{\sigma_\delta^2}{|B_j|}\right),$$

where $Z(B_j, t_j) = \int_{B_j} Z(v, t_j)dv$. Then, the posterior of $a_0$, $b$, and $\sigma_\delta^2$ is proportional to

$$[\tilde{Z}(B_1, t_1), \ldots, \tilde{Z}(B_m, t_m)|Z(B_1, t_1), \ldots, Z(B_m, t_m), a_0, b, \sigma_\delta^2][a_0, b, \sigma_\delta^2],$$

where

$$[a_0, b, \sigma_\delta^2] = [a_0][b][\sigma_\delta^2],$$

$$[a_0] \sim \text{Normal},$$

$$[b] \sim \text{Normal}$$

and

$$[\sigma_\delta^2] \sim \text{IG}(d, 2).$$

The hyperprior can be specified for the parameters in the priors of $a_0$ and $b$.

**Gibbs sampling: Stage 4**

We simulate values of $Z$, the underlying true wind values, at the $(x_i, t_i), i = 1, \ldots, n$, where we have point observations for $\hat{Z}$, and at $(B_j, t_j), j = 1, \ldots, m$, where we have numerical model output $\tilde{Z}$, conditioning on the parameters $\theta_3$ (update in Stage 1) and $\hat{Z}, \tilde{Z}$. The simulated values at $m$ grid cells, $(B_j, t_j), j = 1, \ldots, m$, are obtained by simulating values of $Z$ at sample locations within each grid cell. Then $Z(B_j, t_j)$ is approximated by

$$Z(B_j, t_j) \propto L^{-1} \sum_{l=1}^{L} Z(v_{jl}, t_j),$$

where $v_{j1}, \ldots, v_{jL}$ are sample locations in $B_j$. 
In this algorithm, we fix the number of subregion $k$ and assume the kernel function is known. In fact, we can treat both $k$ and the band-width of the kernel function as unknown parameters, and estimate them using a Poisson prior for $k$ and a non-informative prior for the band-width. The empirical test for separability has been used to determine the spatial-temporal structure for each subregion. Instead, we can specify the local spatial-temporal covariance function in the general form, which corresponds to the Fourier transform of (3.8). Then we obtain the posterior distribution of $\epsilon_i$ for each subregion. A simple version is to consider $\epsilon_i = 0$ and 1 only. When $\epsilon_i = 0$, the local spatial-temporal covariance is nonseparable and it is separable when $\epsilon_i = 1$. 
Chapter 5

Application

In this chapter we will apply the methodology proposed in previous chapters to the Chesapeake Bay wind data. The goals are to understand the spatial temporal structure of wind fields over the Bay; to statistically assess the MM5 model performance; and to obtain a more reliable wind field mapping by combining numerical model output with observations.

5.1 Description of the data

Both observed and MM5 model output wind fields were used in this study. The observed data were collected by a suite of meteorological stations owned by Weatherflow, Inc (see Figure 5.1 ) and provided by Jay Titlow of Weatherflow, Inc. The MM5 data were furnished by John McHenry at Baron’s Advanced Meteorological Systems.

As introduced in Chapter 1, the observed wind data were collected at the locations shown in Figure 5.1. The anemometers were location from 9 to 18 m above ground level, depending on the location (see Table 1.1). Adjustment of the wind speed values to the standard 10 m height was accomplished using Monin-Obukhov similarity theory (2). This adjustment had little affect on the wind speed values given the small distances involved. The adjustment procedure is given in the Appendix.
The MM5 model output fields were generated on a 15-km Arakawa C grid (see Figure 5.2). This grid spacing was selected because it is similar to the finest mesh produced by operational models run at the National Centers for Environmental Prediction (NCEP), and based on discussions with Jay Titlow at Weatherflow, Inc. Wind speeds at the 10m level were used in the analysis.

The analysis of wind speed on July 21, 2002 is presented.

5.2 Exploratory analysis

In this analysis, data from 21 July 2002 have been examined in detail for the full 24-hr period. The complicated flow patterns over the region during this time are evident
in Figure 5.3, which shows MM5 model forecast winds. The arrows indicate the direction from which the winds are coming, while the length of the stem indicates wind speed in meters per second. The plots were done at 3-hr intervals. A streamline (streamlines are lines that are everywhere tangent to the instantaneous wind vector) analysis of the flow fields shows alternating areas of confluence (areas where the streamlines tend to come together) and divergence (areas where the streamlines tend to spread apart). Confluence may or may not be associated with mass convergence, while divergence may or may not be associated with mass divergence depending on how the wind speed changes in these zones (25).

An easterly (winds from the east, southeast, and northeast) wind component tends to dominate over the region for this time period (at least in the MM5 forecast). The main exception occurs at 3AM in the western part of the study area where the flow is from the west and southwest. The 3AM plot clearly shows an area of confluence on the west side of the Bay. At 6AM this area of confluence has been replaced by an area of divergence over
Figure 5.3: Wind field map from MM5 model forecast.
the center of the Bay. Diffuence seems to persist over the Bay for the rest of the period. There is little evidence in these plots to suggest that MM5 was capturing the sea breeze circulation, which observations show to be present.

5.3 Spatial-temporal structure of wind speed

We model the observed wind speed and wind speed from MM5 in terms of an unobserved underlying true process, \( \{Z(x, t) : x \in D \subset \mathbb{R}^2, t \in T \subset \mathbb{R}\} \). The spatial domain \( D \) is the area shown in Figure 5.2. The model for the observed wind speed is the model given by (4.1) and the model for MM5 model output is (4.4). The underlying true wind process is a nonseparable nonstationary process, which is in the form of (3.38). We use this model because the time series of wind speed residuals at a location is homogeneous after we remove the trend surface. The trend is modeled as functions of the known covariates, which are sine and cosine functions with respect to different periods (which is 1/frequency). The kernel function \( K \) is defined as \( K(u - s_i) = \frac{1}{h_i^2} K_0(u - \frac{u - s_i}{h_i}) \), where \( K_0(u) \) is the quadratic weight function

\[
K_0(u) = \frac{3}{4}(1 - u_1^2)_+ + \frac{3}{4}(1 - u_2^2)_+,
\]

for \( u = (u_1, u_2) \). The location \( s_i \) is the centroid of \( i \)-th subregion. The band-width is defined as the half of maximum distance for the \( i \)-th subregion.

5.3.1 Nonstationarity for wind speed

As a first attempt to dealing with the nonstationarity inherent in these kind of environmental data, we divide the spatial domain into two broad categories: land and water. From an atmospheric boundary layer standpoint, this partition appears to be a reasonable way to approach the problem of nonstationarity.

Using the \( u \) wind component, \( v \) wind component and wind speed from model output at noon on July 21, 2002, the K-means cluster procedure is used to further subdivide these two regions to help identify domains of stationarity. We iterate this process until
the AIC criteria suggests that there is no significant improvement in the estimation of the parameters. The optimal number of clusters suggested by the AIC criteria is five. The five subregions are shown in Figure 5.4. This final regional arrangement of clusters appears reasonable considering atmospheric and oceanic processes that are occurring in the boundary layer on this day. It is reasonable to assume that there will be some changes in the configuration of these regions as time passes and the boundary layer structure changes. However, an examination of this issue indicated that the clusters remained reasonable stable from one time period to the other.

Since there are 48 variables from MM5 model output, we also tried to do principle component analysis first, and then to cluster based on the principal components. The results are similar as the subregions shown in Figure 5.4, but it is not clear how to explain it in terms of principal components.

![Subregions of stationarity](image-url)

Figure 5.4: Subregions of stationarity.
5.3.2 Empirical test for separability

The empirical test for separability was performed within each subregion. At this stage, we carry out the test for separability based on the residual of wind speed by removing the trend with the covariates. We consider the ratio for the first four temporal/spatial lags. In Figure 5.5, for each subregion it shows the median of $R_S(u)$ with 95% credible interval (C.I.) at temporal lag $u$. The plot in Figure 5.5 is for fixed $h_1 = 0km$ and $h_2 = 15km$. There is no significant difference in $R_S(u)$, therefore the time series of wind speed residuals has similar structure over space within each subregion. Similarly, Figure 5.6, shows the median of $R_T(h)$ with 95% C.I. at spatial lag $h$ for each subregion. This is for fixed $u_1 = 0hour$ and $h_2 = 3hour$. There is no significant different in $R_T(u)$ except for subregion 4. So the spatial process of wind speed has same pattern over time within each subregion except for subregion 4. The result from the empirical separability test implies that all subregions are separable except for subregion 4.

The tests for stationarity, separability and isotropy within each subregion, which are using the spectral method proposed by 7), are implemented too. Stationarity holds within each subregion. Separability holds within all subregions except subregion 4. This result agrees with the result from the empirical test. Moreover subregion 1 and subregion 5 are anisotropic, which means the spatial covariance depends not only on distance but also direction. A linear transformation is used to transform the coordinates $(x, y)$ so that the spatial covariance is isotropic. The new coordinates $(x', y') = (x, y)RT$, where $R$ is a rotation matrix and $T$ is a shrinking matrix. $R$ and $T$ are defined as following:

$$
R = \begin{pmatrix}
\cos A & \sin A \\
-\sin A & \cos A 
\end{pmatrix}
$$
and

$$
T = \begin{pmatrix}
1 & 0 \\
0 & \frac{1}{R}
\end{pmatrix},
$$

where $A$ describes the angle of rotation and $R$ is used to stretch the coordinates.

After we studied nonstationarity and nonseparability, we estimated the spatial-temporal structure using the algorithm proposed in Chapter 4. The priors used for $R$ and $T$ are uniform distributions over $[0, \pi]$ and $[1, 5]$. 
Figure 5.5: The time series of wind speed has similar structure over space within each subregion.
Figure 5.6: The spatial process of wind speed has same pattern over time within each subregion except for subregion 4.
5.3.3 Spatial-temporal trend

The space-time dynamic statistical model for trend, which is proposed in Chapter 3, is simplified as two parts. One is overall temporal trend and it does not change with location. The other part is variation within each subregion that varies from subregion to subregion. The covariates used here are sine and cosine functions with respect to different time periods (which is 1/frequency). The periods of 24 hours and 12 hours are used to capture diurnal and half-diurnal cycles. The posterior means for these two parts of spatial-temporal trend are shown in the Figure 5.7.

Figure 5.7: Spatial-temporal trend for wind speed.

Figure 5.7 shows the overall mean wind field by hour for July 21, 2002. From Figure 5.7 it appears that from a time standpoint the highest wind speed over the region as a whole occurs just before sunset, while the minimum wind speed on that day occurs several hours before sunrise. The variation for subregions 1 and 3 are generally negative overtime,
while subregions 4 and 5 are generally positive. Subregion 2 moves back and forth between positive and negative values. The large positive values in subregion 5 indicate the wind speeds in that subregion are higher than the average during the mid-day hours. There are a number of meteorological processes going on in the boundary layer that could account for these differences between regions. One of these centers on the differences in terrain characteristics and roughness length (a measure of the aerodynamic roughness of a surface over which a fluid is flowing (2)), which could differ considerably from a land surface to a water surface. Since region 5 is primarily a water-based subregion, the roughness lengths should be less than over a land surfaces provided the water surface is relatively smooth. The large mid-day negative values for subregion 1 (and to a lesser extent subregion 2) are consistent with the complex terrain and the associated high roughness lengths for this part of Virginia.

5.3.4 Spatial-temporal covariance structure

First we draw empirical covariance graphs for each subregion (see Figure 5.8). It is clear that there are ridges in the empirical covariance plots for subregion 1, 2, 3 and 5. Ridges are usually shown in a separable covariance structure. On the other hand, the empirical covariance plot for subregion 4 is smoother than others. It also suggests a non-separable covariance structure for subregion 4.

As mentioned in Chapter 4, we also estimate $\epsilon_i$ in the Bayesian algorithm to indicate if the spatial-temporal process within each subregion is separable or not. Here we only consider $\epsilon_i = 0$ and 1. For separable spatial-temporal covariance, the Matérn covariance is used for spatial covariance and the exponential model is used for temporal covariance. The 3-d Matérn covariance in (3.11) is applied for nonseparable spatial-temporal covariance cases.

The posterior distribution for $\epsilon_i$ is given in Table 5.1. Remember that $\epsilon_i = 0$ indicates a nonseparable covariance for the $i$th subregion, and $\epsilon_i = 1$ implies a separable covariance for the $i$th subregion. The posterior distributions of $\epsilon_i$ strongly support that the separability holds for each subregion except for subregion 4. This result is the same as the result from the empirical separability test.
Figure 5.8: Empirical covariance.

The posterior densities for sill parameters of each subregion are shown in Figure 5.9. These regional differences (see Figure 5.4) are also reflected in the posterior distributions for the sill parameter. The larger sill value associated with the Chesapeake Bay sub-domain is more than likely related to the broad geographical region that the cluster encompasses. The largest mean of the posterior distributions was found for subregion 4, which encompasses a large variety of land and water surface. These large variations in the surface roughness characteristics would contribute to large variations in wind speeds across the region. In addition, a major contribution to the wind speed variations can be attributed to stability differences which exist between land and water surfaces due in large extent to the differential heating experienced by these surfaces. Subregion 4 is the least homogeneous of the five subregions. Large spatial variations in wind field are to be expected. The second
most spatially diverse subregion is subregion 1. Its sill value reflects this diversity. Figure 5.9 indicates that the sill parameters are significantly different, which is also the evidence for the nonstationarity over the spatial domain.

The posterior distribution for the spatial range shown in Figure 5.10 indicates that subregion 5 has the highest posterior mean. Again, this result is consistent with the nature of the subregion. One would expect strong spatial continuity in this subregion. The posterior mean is lowest for subregion 1 where the spatial continuity is weak. Subregions 2-4 also show weak spatial continuity.

The spatial-temporal covariance model (3.11) is used for subregion 4, which is non-separable. Figure 5.11 shows the posterior density of conversion parameter ($\rho$). This scale parameter takes into account the change of units between the spatial and temporal domains.

The smoothness parameter shown in Figure 5.12 does not change much from subregion to subregion.

The posterior means of covariance parameters are listed in Table 5.2.

The estimated Bayesian covariance plots are shown in Figure 5.13.

<table>
<thead>
<tr>
<th>Subregion</th>
<th>$P(\epsilon_i = 0)$</th>
<th>$P(\epsilon_i = 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subregion 1</td>
<td>7.62e-48</td>
<td>1</td>
</tr>
<tr>
<td>Subregion 2</td>
<td>1.03e-91</td>
<td>1</td>
</tr>
<tr>
<td>Subregion 3</td>
<td>3.76e-94</td>
<td>1</td>
</tr>
<tr>
<td>Subregion 4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subregion 5</td>
<td>2.56e-150</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5.1: The posterior distribution for $\epsilon_i$. 
5.4 Statistical assessment of MM5 performance

For MM5 model evaluation, we simulate observed wind values given MM5 model output. A plot for model evaluation at noon on July 21, 2002 is shown in Figure 5.14. The simulated wind speed is in the horizontal axis. The observed wind speed is in the vertical axis. The dot and short line indicate the observed value and 95% credible interval for simulated value. If the model works perfectly, the simulated value should be close to observation and should fall within the band, which is observation ± 2 standard error. But in Figure 5.14, more than half of the comparisons are out of the band, which suggests that the MM5 model does not work well.

Also the MM5 model evaluation was done at 3pm (see Figure 5.15) and 6pm (see
5.5 Wind field mapping

The original MM5 output wind speed map at noon on July 21, 2002 is shown in Figure 5.17. The color of the image represents the wind speed forecast from MM5. The value plotted on top of the image is the observed wind speed. Figure 5.17 indicates that there is a big difference between the MM5 output and the observed data.
For the prediction of wind fields, we simulate values of wind speed from the posterior predictive distribution of the true underlying wind process given the model output and the observations. Figure 5.18 shows the improved wind map made by combining model output and observed wind data at 12pm on July 21, 2002.

In Figure 5.18, the color of the background is the mean of posterior predictive distribution. The improved map agrees better with observed data. In order to quantify this improvement, $MSE = \frac{1}{n} \sum_{i=1}^{n} (\hat{r}_i - r_{oi})^2$, is calculated. The value of $n$ is the number of observations. The value of $r_{oi}$ is the observed wind speed at location $i$. The value of $\hat{r}_i$ is the prediction of wind speed at location $i$. This prediction is obtained by leaving the $i$th observation out of the estimation and prediction procedure. This is in a manner of “cross validation”. Compared to a $MSE=10.32 \ m^2/s^2$ from MM5 forecast, the improved wind
Figure 5.12: Posterior for smoothness parameter.

map has a MSE=$2.45 \, \text{m}^2/\text{s}^2$. Similarly, we simulated the improved wind map at 3pm (see Figure 5.20 ) and 6pm on July 21, 2002 (see Figure 5.22 ). The results are given by the Table 5.3.

We also calculated MSEs from the 3D-Var technique, which is three dimensional variational assimilation (20). It is a popular technique among meteorologist to improve the numerical model output. Assume the improved wind speed obtained by 3D-Var to be $\mathbf{Z}$, then

$$\mathbf{Z} = \hat{\mathbf{Z}} + W[\hat{\mathbf{Z}} - H(\hat{\mathbf{Z}})],$$

where $\hat{\mathbf{Z}}$ is the wind speed from MM5 model output and $\hat{\mathbf{Z}}$ is the observed wind speed. The matrix $W$ is an optimal weight matrix given by $BHT(R + HBHT)^{-1}$. The matrix $B$ is an error covariance matrix for MM5 model output, the matrix $R$ is an error covariance
\begin{table}[h]
\centering
\begin{tabular}{lcccccc}
\hline
Process & sill & spatial range & temporal range & $\nu$ & A & R & $\rho$
\hline
subregion 1 & 0.1615 & 14.44 & 2.02 & 0.70 & 2.78 & 1.11 & \\
subregion 2 & 0.1275 & 17.62 & 1.64 & 0.88 & & & \\
subregion 3 & 0.1200 & 16.24 & 1.92 & 0.99 & & & \\
subregion 4 & 0.3073 & 16.61 & 1.42 & 13.24 & & & \\
subregion 5 & 0.0944 & 25.64 & 1.73 & 0.58 & 1.18 & 1.17 & \\
\hline
\end{tabular}
\caption{The posterior means for covariance parameters.}
\end{table}

Figure 5.13: Bayesian covariance.
Figure 5.14: MM5 model evaluation at 12pm on July 21, 2002.
Figure 5.15: MM5 model evaluation at 3pm on July 21, 2002.
Figure 5.16: MM5 model evaluation at 6pm on July 21, 2002.
matrix for observations, and $H$ is an operator to interpolate MM5 model output to the observational locations. It is the solution to find the optimal analysis field (e.g., wind speed) that minimize a cost function, where the cost function is defined as the distance between the analysis field and the background (e.g., the wind speed from MM5 model output). This result can be interpreted as that the analysis is obtained by adding to the background the observational increment weighted by optimal weight. The larger the background error variance, the larger the correction to the background. The results are in Table 5.3.

<table>
<thead>
<tr>
<th></th>
<th>MM5 ($m^2/s^2$)</th>
<th>New method ($m^2/s^2$)</th>
<th>3D-Var ($m^2/s^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>noon</td>
<td>10.32</td>
<td>2.45</td>
<td>9.176</td>
</tr>
<tr>
<td>3pm</td>
<td>9.623</td>
<td>8.588</td>
<td>6.445</td>
</tr>
<tr>
<td>6pm</td>
<td>8.941</td>
<td>4.726</td>
<td>5.158</td>
</tr>
</tbody>
</table>

Table 5.3: The comparison of MSEs.

By combining MM5 output with observed wind data, we can obtain more reliable wind field maps. At noon and 6pm, our new method performs better than the 3D-Var technique. But it is not clear to us yet why the MSE obtained by the 3D-Var technique is smaller than the MSE obtained by our new method.
Figure 5.17: Original MM5 output wind speed at 12pm on July 21, 2002.
Figure 5.18: Improved wind speed by combing data at 12pm on July 21, 2002
Figure 5.19: Original MM5 output wind speed at 3pm on July 21, 2002.
Figure 5.20: Improved wind speed by combing data at 3pm on July 21, 2002
Figure 5.21: Original MM5 output wind speed at 6pm on July 21, 2002.
Figure 5.22: Improved wind speed by combing data at 6pm on July 21, 2002
Chapter 6

Summary and Future Work

The contributions of this research are: an empirical test for separability which can be used to understand space-time dependency; a new class of nonseparable stationary spatial-temporal covariance functions; a new class of spatial-temporal models which allow for the lack of stationarity, isotropy and do not assume separability; and a space-time framework for combining different sources of data. In the new class of nonseparable stationary spatial-temporal covariance model, the parameter \( \epsilon \) plays an important role to identify the separability. As part of the future work, we will study the property of \( \epsilon \). The parameter \( \epsilon \) can be estimated using the spectral maximum likelihood method. Therefore, a formal test for separability can be constructed.

We applied the methodology proposed in this research work to the wind data from Chesapeake Bay region. By modeling disparate wind data in terms of an underlying true wind process, which is a nonseparable nonstationary spatial-temporal process, we can obtain improved wind field maps. The improvement obtained in mean square error when the MM5 model output and the observed data are combined is very good. However, these results are based on a very limited time period. While the techniques presented in this research have great promise, both further testing on a large data set and comparison with existing techniques will show their true worth. So far it is not clear to us when is better to use our proposed method instead of the 3D-Var technique.
As noted earlier, the subregions were derived based on the noon data for one day. Even under weakly forced synoptic flow regimes, these subregions of stationary can be expected to shift from hour to hour. One alternative might be to group hours together based on the diurnal restructuring of the atmospheric boundary layer. Using this technique, the subregions may be more stable from day to day. In addition the day-to-day stability of the covariance function should be examined. Under weakly forced flow conditions these functions may show enough stability to be useful in the forecast mode.

From a meteorological viewpoint, our proposed technique could be served as a post-processing algorithm which would be run on meteorological model output fields. In addition, it could use to improve the assimilation of the surface wind field observations into the meteorological model initial field.

The meteorological community has been working on the data assimilation problem for many years. This research has produced some very sophisticated methodologies. Kalnay (2003) points out that early in the history of the the development of numerical weather prediction techniques it was found that a complete first guess estimate of the state of the atmosphere was required in addition to the traditional set of observations. The first guess field should be the meteorologist’s best estimate of the state of the atmosphere before incorporating the observations. In a statistical sense, this is similar to the the specification of a prior in Bayesian analysis. Currently short-range forecasts serve as the first guess field in what has come to be called the “analysis cycle”. According to Kalnay, the analysis cycle is an intermittent data assimilation procedure which typically uses a 6-hr cycle performed four times a day. In areas where there are many observing sites, the observed data have the most influence on the model initial field, while in areas where observational data are scarce the model first guess field is important since it can build on the observational data from data rich areas. The procedure outlined in this research provides a powerful technique for combining the first guess field with the available observations.
Chapter 7

Appendix

7.1 The adjustment of wind fields from different height

The observed wind fields have different heights with wind fields from MM5 model output. We adjust the observed wind data to the 10m level based on the stability condition (2). First we search the MM5 grid cells and find the grid cells in which contain the locations of monitoring sites. Therefore the information from the MM5 model output for those grid cells can be applied to the corresponding monitoring sites. Then for the wind speed at a monitoring site, first calculate the stability parameter defined as $h/L$, where $h$ is boundary layer height and $L$ is Monin-Obukhov length. Both $h$ and $L$ come from MM5 model output. Then the adjustment is as follows,

$$u_2 = \begin{cases} 
    u_1 \frac{\ln \left( \frac{Z_2}{Z_0} \right) + 5 \frac{Z_2}{L}}{\ln \left( \frac{Z_1}{Z_0} \right) + 5 \frac{Z_1}{L}} & \text{if } \frac{h}{L} > 1 \\
    u_1 \frac{\ln \left( \frac{Z_2}{Z_0} \right)}{\ln \left( \frac{Z_1}{Z_0} \right)} & \text{if } -1 \leq \frac{h}{L} \leq 1 \\
    u_1 \frac{\ln \left( \frac{Z_2}{Z_0} \right) - \psi_M \left( \frac{Z_2}{L} \right)}{\ln \left( \frac{Z_1}{Z_0} \right) - \psi_M \left( \frac{Z_1}{L} \right)} & \text{if } \frac{h}{L} < -1
\end{cases}$$
where \( u_2 \) is the adjusted wind, \( u_1 \) is observed wind, \( Z_1 \) is the sensor height, \( Z_2 \) is 10 meters, \( Z_0 \) is the roughness length of the sensor and

\[
\Psi_M(x) = \ln \left( \frac{1 + y^2}{2} \right) \left( \frac{1 + y}{2} \right)^2 - 2 \tan^{-1}(y) + \frac{\pi}{2}
\]

where \( y = (1 - 15x)^{1/4} \), \( x = Z_2/L \) or \( Z_1/L \).

### 7.2 The function in (3.13) is a valid spectral density

A more general nonseparable stationary spectral density was proposed in (3.13), which has the following form,

\[
f(\omega, \tau) = \gamma \left( c_1(\alpha_1 + |\omega|^2)^{\alpha_1} + c_2(\alpha_2 + \tau^2)^{\alpha_2} + \epsilon(\alpha_3 + |\omega|^2 \tau^2)^{\alpha_3} \right)^{-\nu},
\]

where \( \alpha_1, \alpha_2, \alpha_3 \) and \( c_1, c_2 \) are positive; \( \epsilon \in [0,1] \) and \( \frac{\beta_1}{\omega_1^\nu} + \frac{1}{\omega_2^\nu} < 2 \).

This is a valid spectral density. First of all, under the above conditions, \( f(\omega, \tau) > 0 \).

Let

\[
g(\omega, \tau) = \gamma \left( c_1(\alpha_1 + |\omega|^2)^{\alpha_1} + c_2(\alpha_2 + \tau^2)^{\alpha_2} \right)^{-\nu},
\]

where \( \alpha_1, \alpha_2, \alpha_3 \) and \( c_1, c_2 \) are positive, and \( \frac{\beta_1}{\omega_1^\nu} + \frac{1}{\omega_2^\nu} < 2 \). Then \( f(\omega, \tau) \leq g(\omega, \tau) \).

But \( g(\omega, \tau) \) is integrable, hence the function proposed in (3.13) is a valid spectral density.

The following theorem proved by 31) can be used directly to prove the function \( g(\omega, \tau) \) in (7.2) is integrable.

**Theorem.** Suppose \( f(\omega_1, \omega_2) = \{g_1(|\omega_1|^2) + g_1(|\omega_1|^2)\}^{-\nu} \), where \( \nu > 0 \), \( f \) is bounded, \( g_1 \) and \( g_2 \) are infinitely differentiable functions on \([0,\infty)\) and there exist \( \alpha_1, \alpha_2 \) positive such that \( g_j^{(k)}(t) \ll (1 + t)^{\alpha_j - k} \) for \( j = 1, 2 \) and all nonnegative integers \( k \). A necessary and sufficient condition for \( f \) to be integrable is \( d_1/(\alpha_1^\nu) + d_2/(\alpha_2^\nu) < 2 \), where \( \omega_j \in \mathbb{R}^j, j = 1, 2 \). If \( f \) is integrable, its Fourier transform \( K(x) \) possesses all partial derivatives of all orders for all \( x \neq 0 \).
Bibliography


