THE UNIVERSITY OF CHICAGO

TWO PROBLEMS IN ENVIRONMETRICS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF STATISTICS

BY
HAE KYUNG IM

CHICAGO, ILLINOIS
DECEMBER 2005

To my parents and to Pablo

ABSTRACT

The first part of this thesis presents a new approach to emission scenario analysis using simplified Chemical Transport Models (CTM). Typically, CTM runs are highly time consuming, so obtaining fast and accurate approximations to some of their outputs is desired. We present two methods that closely approximate the ammonia wet deposition output from the Community Multiscale Air Quality (CMAQ) model by combining the CMAQ output under a base emission scenario and a few runs of a simplified model. We obtained the simpler model of CMAQ by switching off all the chemistry and only allowing ammonia to undergo physical transport and dry and wet deposition. This simplified version called here Tracer model runs 40 times faster than the full model. The second method uses a source receptor matrix and the base CMAQ run to get good approximations for any new emission scenario. We apply our methods to solve the inverse problem of correcting ammonia emissions based on observations.

The second part proposes a semiparametric method to estimate spectral densities of isotropic Gaussian processes with scattered data. We model the spectral density function (Fourier transform of the covariance function) with a linear combination of B-splines up to a cutoff frequency and, from this point, a truncated algebraic tail. We calculate an analytic expression for the covariance function and tackle several numerical issues that arise when calculating the likelihood. We maximize the likelihood using the simulated annealing method. We compare our method with a kernel method proposed by Hall et al.(1994) and a parametric method using Matérn model. Our simulations results show that our method outperforms the other two by several

criteria. Our method directly estimates the tail behavior of the spectral density, which has the biggest impact on interpolation properties, and takes fully into account the correlations between observations through the use of the likelihood.

ACKNOWLEDGMENTS

It has been a great privilege to work with my adviser Michael Stein for the last several years. His constant guidance, encouragement and contagious enthusiasm made possible the conclusion of this thesis. I thank my committee members Rao Kotamarthi, Vanja Dukic, and Per Mykland for their technical and moral support.

I would also like to thank the faculty, students, and staff members of the Department of Statistics at the University of Chicago for providing me with a friendly and intellectually stimulating environment where I could complete this work. In particular, I am grateful to Maoxia, Elliot, David, Ethan, and Mikyoung for their friendship and for making this journey much more pleasant. I could not leave out Mitzi and Karen from this list for helping me out innumerable times.

This work would not have been possible without the support of my husband Pablo. Thanks for all the moments we shared through so many years! I thank my parents for always encouraging me to improve myself, Laura for smiling at me every morning and making each day special, and the little one to come for making the future more hopeful.

The first part of this thesis is a joint work with my adviser and Rao Kotamarthi. Alexis Zubrow provided me with preprocessed data necessary for the first part of this work and invaluable help and tools for running CMAQ and analyzing the results. The second part of the thesis is a joint work with my adviser and Zhengyuan Zhu. Ken Wilder provided me with C code necessary for calculating hypergeometric functions accurately for which I am very grateful.

The research described herein has been funded wholly or in part by the United States Environmental Protection Agency through STAR Cooperative Agreement #R-82940201-0 to The University of Chicago, it has not been subjected to the Agency's required peer and policy review and therefore does not necessarily reflect the views of the Agency, and no official endorsement should be inferred.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Part I

A new approach to scenario analysis using simplified Chemical Transport Models

CHAPTER 1

INTRODUCTION

The current generation of regional scale chemistry transport (CTM) models are complex and require large computational resources for scenario and episode studies. One such model, CMAQ [Byun and Ching, 1999] is currently favored by the Environmental Protection Agency for the simulation of multiple pollutant concentration levels at urban and regional scales. The main goals of these CTMs are to simulate the physical and chemical processes that transport and transform gas and particulate pollutants emitted into the atmosphere and to assess the impact of changes in emissions on air quality. They are also used as an air quality management tool.

There is an increasing interest in the emission, distribution and deposition of ammonia since it primarily reacts with sulfuric acid and nitric acid in the atmosphere forming fine particulate matter ($PM_{2.5}$), designated as a criteria pollutant by the National Ambient Air Quality Standards set by the United States Environmental Protection Agency (USEPA). Positive association between high levels of particulate matter and adverse health effects has been noted in a number of recent studies [Peng et al., 2005, Pope, 2004, Daniels et al., 2000]. The emissions of ammonia into the atmosphere are highly uncertain [Bouwman et al., 1997]. Ammonia emission measurements are very difficult, indirect [Aneja et al., 2000, Roelle and Aneja, 2002], and sparse. Current USEPA's inventories [U.S. Environmental Protection Agency, 2000] are based on annual averages that are calculated by multiplying the source abundance by emission factors published in the literature [Asman et al., 1998]. Since the

primary sources of ammonia are from farm animals and seasonal agricultural practices and because emission factors are significantly affected by temperature, annual average emissions are grossly inadequate. Gilliland et al. [2003] estimated seasonal adjustments to ammonia emissions in the eastern United States by using an inverse modeling method. They considered the entire region as one source region, which means that for each month they estimated one global adjustment factor for the whole region. Their method demanded at least 3 simulations per month, and at the time of their study, each one required 2 weeks of runtime on a CRAY T3E system. These times have been reduced dramatically with newer versions of CMAQ and improved processors. As of October 2004, EPA was achieving annual continental US runs in about a week. Nevertheless, the number of runs necessary to solve inverse problems grows rapidly as we increase the number of source regions to resolve, so the main limitation of the scope of inverse problems is still the computational burden. Therefore, a fast approximation is crucial for inverse modeling purposes.

CHAPTER 2

CMAQ, TRACER AND MULTITRACER MODELS

We ran CMAQ in the Eastern United States region with 67×68 square cells of 36 by 36 km and 28 vertical layers. The CMAQ model version 4.3, with CB4 (carbon bond scheme) chemistry, organic and inorganic, aerosols and aqueous chemistry (CB4_ae3_aq in CMAQ terminology) was employed for making the model runs. The emissions were calculated using SMOKE (Sparse Matrix Operator Kernel Emissions Modeling System) with National Emission Inventory (NEI) 1996, generated by EPA, and the meteorology input to CMAQ was computed with the Fifth Generation Penn State/NCAR Mesoscale Model (MM5). The initial and boundary condition for the MM5 (V3.6) were generated from the 4 times a day NCEP(National Centers for Environmental Prediction)/Department of Energy Atmospheric Model Intercomparison Project-II (AMIP-II) Reanalysis data set [Kistler et al., 2001]. The input data has a spatial resolution of 2.5-deg by 2.5-deg in the horizontal and 17 mandatory pressure levels. The selected physics parameterizations for these runs include Dudhia's simple ice moisture scheme; Grell's basic cumulus parameterization scheme; the MRF (Medium Range Forecast) PBL (Planetary Boundary Layer) scheme and a radiation scheme based on the CRM (Climate Radiation Model) and updated every 30 minutes. The CMAQ trace gas and aerosol boundary conditions were set to nominal continental background conditions for the respective trace gases and aerosols.

Most of the runs were for the 8 day period from 26 June 1996 to 4 July 1996. We also did some additional runs for the 8 day period from 4 July 1996 to 12 July 1996

and for the 28 day period from 2 July 1996 to 30 July 1996. Ammonia wet deposition was aggregated over each period. The length of the shorter period used, 8 days, is approximately the lifetime of NH_x.

The Tracer model is a reduced version of CMAQ in which ammonia is the only species and it is treated as a tracer without any chemical interactions. Ammonia is allowed to undergo wet and dry deposition as gaseous ammonia. The full version of the CMAQ model converts emitted NH_3 to NH_4, which is partitioned into particulate ammonia (ammonium sulfate and ammonium nitrate) and removed from the atmosphere by dry and wet deposition. Ammonia dry and wet depositions are computed using the dry deposition velocities and wet deposition rates computed in the model for ammonia. The emissions and meteorology are the same as for the full version of CMAQ.

The Multitracer model is a generalization of Tracer that simultaneously calculates the transport and deposition of 100 ammonia-like tracers released from 100 different aggregated surface grid locations in the model.

Our computations were carried out on a Beowulf cluster with 4 computational nodes and one data-server. The nodes are dual AMD MP 1800+ processors with 2 gigabytes of memory, connected via gigabit Ethernet. The operating system is Debian Linux and the parallel software is Mpich 1.2.4. Under this system, the processing time for CMAQ was 35 minutes per day using 4 processors while the Tracer model would take 3.5 minutes per day on one processor (there was no gain in using more processors for the Tracer model). The Multitracer took about the same time as CMAQ using the same number of processors.

CHAPTER 3

METHODOLOGY

3.1 Predicting CMAQ

Our goal is to predict CMAQ ammonia wet deposition output under new emission scenarios making use of a reference CMAQ run and a few Tracer runs or a single Multitracer run. For this purpose we have analyzed the changes in CMAQ generated ammonia wet deposition using different emission scenarios.

We use as the reference emission field our current best estimate using EPA's inventory, which will be referred to as $\mathbf{S}_\circ$. We modified the reference field to obtain new scenarios in order to test our predictors. Three target scenarios, $\mathbf{S}_{\sigma Z}$ with $\sigma = 0.1$, 0.4, and 1, were generated by multiplying $\mathbf{S}_\circ$ by the exponential of a Gaussian random field. This field was simulated using a covariance function $\exp(-d/r)$, where d is the distance between cells and r is the range of the correlation (400km). The Gaussian random field is shown in Figure 3.1 where we can see regions of high and low values with diameter of the order of 400 km consistent with the range of the correlation. Modifying the reference emission field with this method generates a wide range of plausible scenarios one may want to test in an inverse modeling procedure. When $\sigma = 0.1$ ($\sigma = 0.4$, $\sigma = 1$) the base emission field is changed on average about 10% (50%, 300%). Depending on the location, changes were much more extreme; for example, for $\sigma = 1$ the base emission was multiplied by factors as large as 40. Another scenario, $\mathbf{S}_p$, was generated by multiplying the base emission by a plane

that takes value 1 in the northern end of the region (latitude 50.216^oN) and 0 in the southern end (latitude 27.125^oN). Lastly, $\mathbf{S}_1$ is a completely unrelated emission field that has value 1 (mols/s) in the whole domain. Figure 3.2 shows the reference emission scenario $\mathbf{S}_\circ$ and target scenarios $\mathbf{S}_{0.4Z}$, $\mathbf{S}_Z$, and $\mathbf{S}_p$.

If we scale the original emission by a constant factor α, the resulting CMAQ output is scaled by the same factor raised to the power of 0.7. We have tested this relationship with α taking values 0.5, 0.8, 1.5, 2, and 3. Table 3.1 shows the slopes of the least squares fit of CMAQ wet deposition under the scaled emission vs. the base CMAQ wet deposition and, for comparison, the values of $\alpha^{0.7}$. The last column of Table 3.1 shows the coefficient of determination R^2, which is calculated using $1 - \sum_i (y_i - \alpha^{0.7} x_i)^2 / \sum_i y_i^2$. We can see that these values are all above 0.99 except for $\alpha = 3$ in which case the R^2 is 0.979. Therefore, for constant adjustments of the emissions, as long as the adjustment is within this range of 0.5-3, a very good predictor of CMAQ ammonia wet deposition is the reference CMAQ wet deposition multiplied by the adjustment factor raised to the power of 0.7:

$$\mathbf{C}(\alpha \mathbf{S}_\circ) \simeq \alpha^{0.7} \mathbf{C}(\mathbf{S}_\circ), \tag{3.1}$$

where we denote CMAQ wet deposition under emission scenario $\mathbf{S}_x$ as $C(\mathbf{S}_x)$. Motivated by this result, we could propose the following naive predictor when the new emission $\mathbf{S}_n$ is not a simple rescaling of $\mathbf{S}_\circ$

$$\mathbf{C}^{(0)}(\mathbf{S}_n) \simeq \left(\frac{\mathbf{S}_n}{\mathbf{S}_\circ} \right)^{0.7} \mathbf{C}(\mathbf{S}_\circ) \tag{3.2}$$

This expression reduces to (3.1) when $\mathbf{S}_n$ is α times the reference emission $\mathbf{S}_\circ$.

A more useful approximation can be generated by examining the relationship

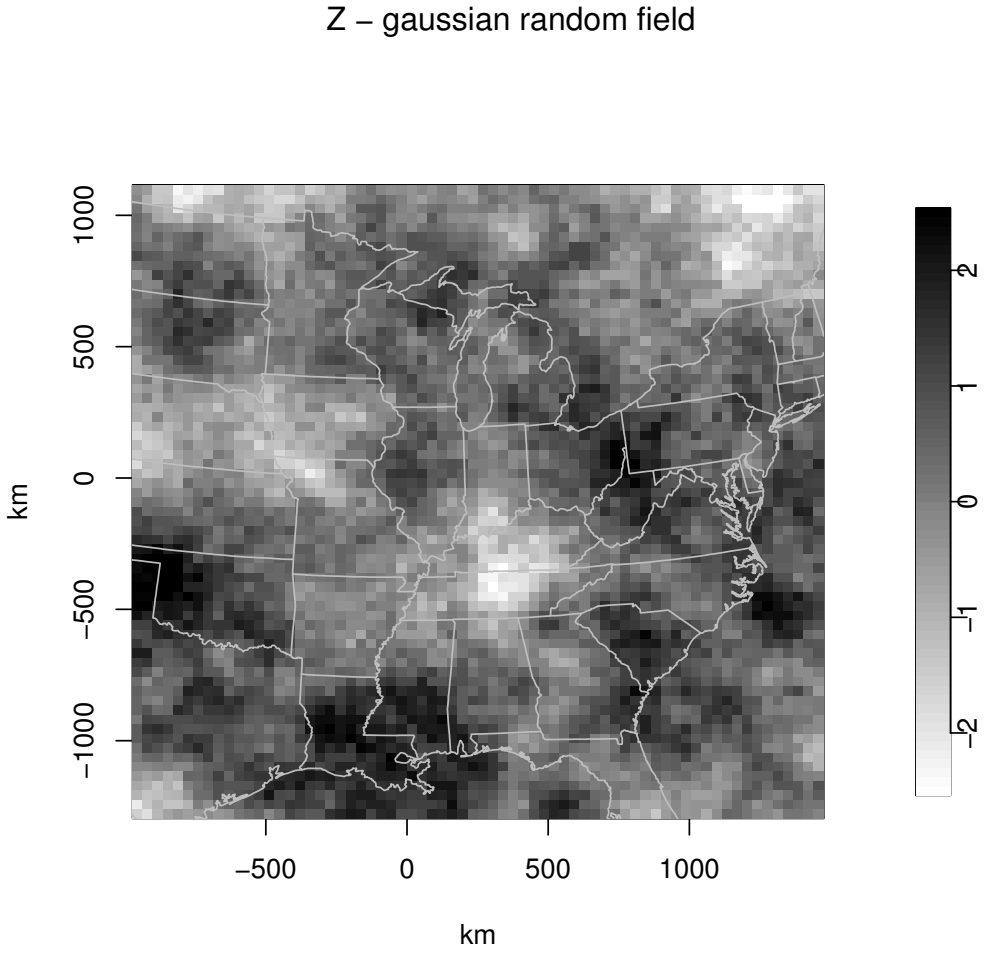

Figure 3.1: Simulated Gaussian random field with covariance function $e^{-d/r}$ where d is the distance between cells and r is the range of the correlation (400km).

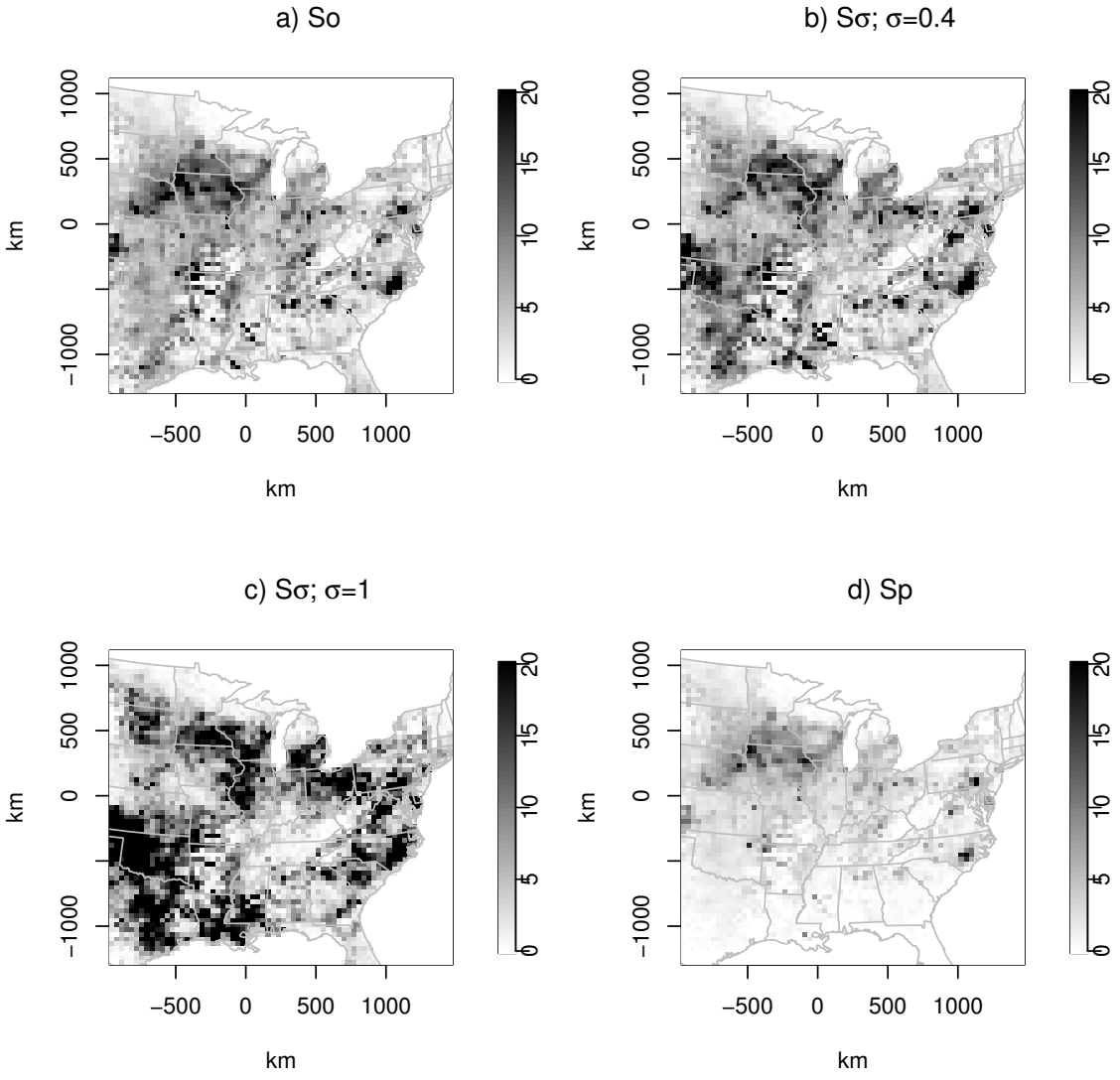

Figure 3.2: Reference emission scenario $\mathbf{S}_\circ$ and target emission scenarios $\mathbf{S}_{0.4Z}$,$\mathbf{S}_Z$, and $\mathbf{S}_p$. The range of the figure is set to 20 mols/sec for visualization purposes but the actual maximum is much larger than this value.

Table 3.1: Relationship between CMAQ wet deposition with scaled emissions and base emissions

α	$\alpha^{0.7}$	Slope	R^2
0.5	0.62	0.59	0.994
0.8	0.86	0.85	0.999
1.5	1.33	1.32	0.997
2.0	1.62	1.63	0.991
3.0	2.16	2.04	0.979

between ammonia emission source regions and the regions calculated by the model to experience significant wet deposition. Figure 3.3 shows CMAQ wet deposition vs. emissions where each cross represents a cell of the domain. The lack of correlation between the two variables indicates that ammonia is transported away from its source before being scavenged by precipitation. This result is consistent with the fact that most of CMAQ wet deposition comes from ammonium in aerosol form with sizes ranging between 0.1 and 2μm, which has residence times of the order of 4 to 7 days. Thus, a more comprehensive approximation to CMAQ wet deposition of ammonia can be generated by including the effect of transport in our predictors. One way of doing so is by substituting the Tracer model's generated wet deposition in (3.2) as a substitute for just the emission scenario's ($\mathbf{S}_\circ$, $\mathbf{S}_n$, etc.). Each Tracer calculation is an approximation of the impact of the transport history from source to sink (wet deposition in this case) of ammonia emissions, yielding an approximation for the effect of transport on calculated wet deposition. Hence our first proposed predictor of $C(S_n)$ is

$$\mathbf{C}_n^{(1)} = \left(\frac{\mathbf{t}_n}{\mathbf{t}_o}\right)^{0.7} \mathbf{C}_\circ \tag{3.3}$$

where $\mathbf{t}_n$ and $\mathbf{t}_\circ$ are the Tracer output under emission scenarios $\mathbf{S}_n$ and $\mathbf{S}_\circ$, respectively, and $\mathbf{C}_\circ = \mathbf{C}(\mathbf{S}_\circ)$ corresponds to CMAQ output under the base scenario. Notice

that the superscript with a number between parenthesis indicates that the variable is a predictor, not the actual CMAQ output, and the subscript indicates the corresponding emission scenario. Once we have a base run of CMAQ ($\mathbf{C}_\circ$) and Tracer ($\mathbf{t}_\circ$), we only need to run the Tracer model with the new emission scenario ($\mathbf{S}_n$) in order to compute this predictor. As already mentioned, this effectively reduces the computation time by a factor of 40.

We can further simplify our prediction procedure by exploiting the near linearity of the Tracer model. Since the Tracer model consists of transport, wet and dry deposition, it is very nearly linear in emissions, i.e., if we multiply emissions by α the resulting wet deposition is multiplied by α, except for some modest effect, which means that we do not need to re-run the Tracer model when the new emission is a scaling of the base emission. This almost linearity means that if we arrange the model calculated wet deposition and emissions into vectors, which will have length 4556 (67×68 cells in the domain), the tracer output can be reasonably represented as a product between a matrix $\mathbf{T}$ (as yet undetermined) and the vector of emissions $\mathbf{S}_n$:

$$\mathbf{t}(\mathbf{S}_n) \simeq \mathbf{T}\mathbf{S}_n. \tag{3.4}$$

We will refer to the matrix $\mathbf{T}$ as the transport matrix. We can calculate $\mathbf{T}$ by noticing that its i^{th} column is the wet deposition resulting from a unit emission from location i. Thus, one method to calculate $\mathbf{T}$ is by setting emissions to 1 in one location and 0 elsewhere, running the Tracer model and using the output to read off the column of $\mathbf{T}$ corresponding to that location. Repeating this procedure for all locations would allow us to construct all of $\mathbf{T}$'s columns. However, running the Tracer model 4556 times is too time consuming to be carried out in practice. We propose aggregating the 4556 locations into 100 subregions so that the transport matrix has 100 columns,

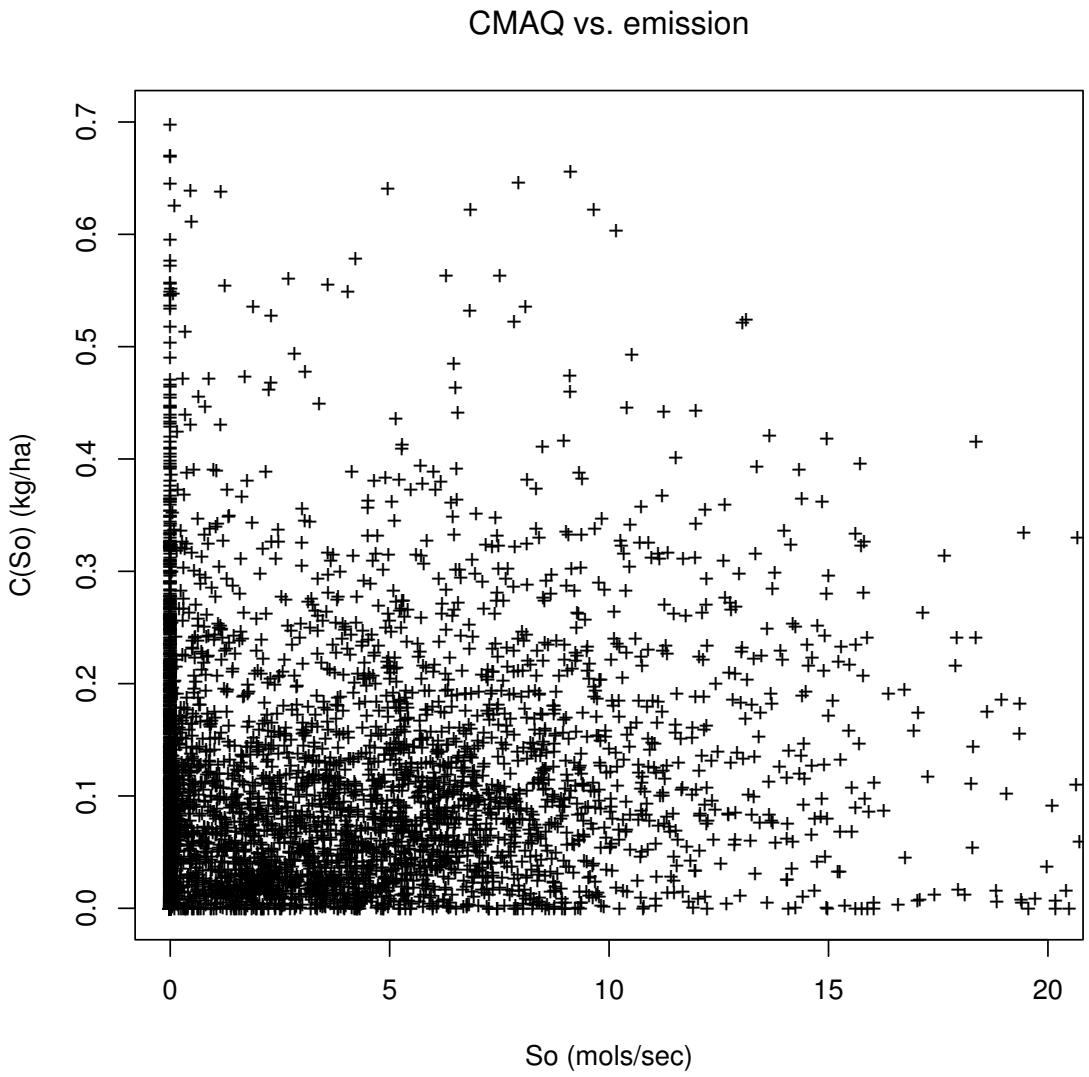

Figure 3.3: CMAQ ammonia wet deposition vs. ammonia emission.

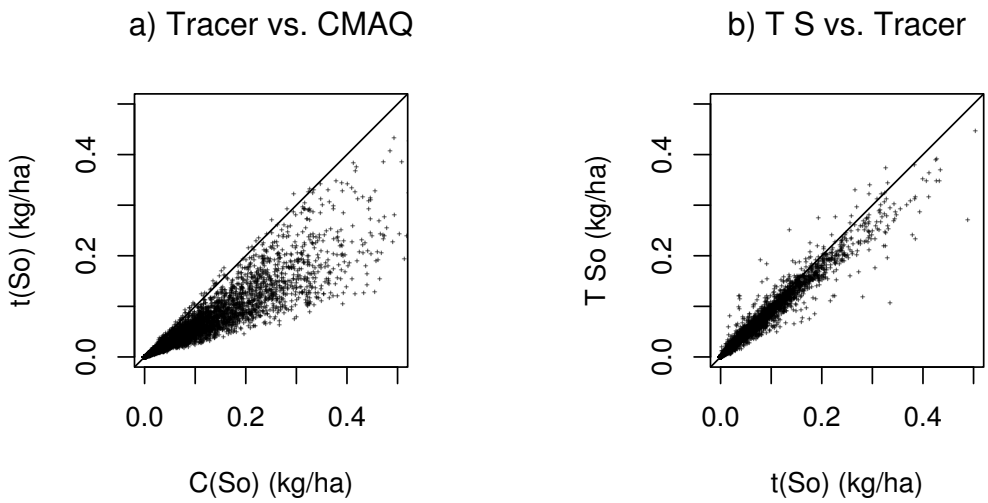

Figure 3.4: a) Tracer vs. CMAQ under $\mathbf{S}_\circ$ (the least square fit slope is 0.56 and the correlation is 0.92), b) Transport Matrix approximation vs. Tracer under $\mathbf{S}_\circ$ (the least squares fit slope is 0.88 and the correlation is 0.97).

and we could then obtain an approximate version of $\mathbf{T}$ that we denote by $\tilde{\mathbf{T}}$ by running the Tracer model only 100 times. Instead, we find $\tilde{\mathbf{T}}$ by using the Multitracer model, which runs all 100 Tracer models simultaneously. Figure 3.4b shows the Tracer model's output vs. the aggregated transport matrix times the aggregated emissions. Note that $t(\mathbf{S}_\circ) \neq \tilde{\mathbf{T}}\tilde{S}_\circ$ because of the aggregation and the small nonlinearity of the Tracer model.

The use of transport or source-receptor matrices has been used in integrated models such as RAINS [Amann et al., 2004]. The difference in approach to our method is that we use the transport matrix only as an approximation to the simplified model and further combine it with baseline full model runs to obtain a much better approximation as can be seen in the "Results" chapter.

Denoting by $\tilde{\mathbf{T}}$ and $\tilde{\mathbf{S}}$ the aggregated matrix and emission vectors, our second

proposed predictor of $\mathbf{C}(\mathbf{S}_\circ)$ is

$$\mathbf{C}_n^{(2)} = \left(\frac{\tilde{\mathbf{T}}\tilde{\mathbf{S}}_n}{\tilde{\mathbf{T}}\tilde{\mathbf{S}}_\circ}\right)^{0.7} \mathbf{C}_\circ. \tag{3.5}$$

Once we have a base run of CMAQ and one Multitracer run, we only need a matrix multiplication in order to compute this predictor for any new emission scenario.

In (3.5) we have two types of matrix/vector operations. One is the usual matrix-vector multiplication that returns a vector (as in $\tilde{\mathbf{T}}_n\tilde{\mathbf{S}}_n$) and the other one is a vector-vector multiplication that has to be interpreted as element-wise multiplication and returns a vector. We use no symbol to denote either operation but it should be clear from the arguments involved which type of multiplication is being used. Division and raising to a power should always be understood as element-wise operations.

3.2 Inverse modeling emissions

Let us assume that we have observations in every location of our domain. This is not realistic but, in practice, we can interpolate the observations, for example, by kriging [Cressie, 1993], to obtain values in each location. The inverse modeling problem consists of finding the emissions that originated this field assuming that the only source of error in the model is the emission field.

The form of our second predictor suggests a simple method to obtain a direct estimate of the unknown emission scenario. In this case, the left hand side of (3.5) is considered to be known and equal to the observed field. We can invert this relation to obtain an estimate of the unknown emission scenario $\tilde{\mathbf{S}}_{\text{obs}}$

$$\hat{\mathbf{S}}_{\text{obs}} = \tilde{\mathbf{T}}^{-1}\left[\left(\frac{\mathbf{C}_{\text{obs}}}{\mathbf{C}_\circ}\right)^{\frac{1}{0.7}}\left(\tilde{\mathbf{T}}\tilde{\mathbf{S}}_\circ\right)\right]. \tag{3.6}$$

We use the symbol $\hat{}$ to denote estimated value and $\tilde{}$ to denote aggregation. However, we do not use $\hat{}$ and $\tilde{}$ simultaneously to avoid complicated expressions.

In our case, we are using an approximate transport matrix that has been aggregated into 100 columns, so it is not a square matrix. It is common in inverse modeling problems [Enting, 2002] to encounter matrices that are not full rank. Both problems can be addressed by using a generalized inverse or pseudo-inverse. By the properties of the pseudo-inverse [Golub and Van Loan, 1996], $\hat{\mathbf{S}} = \mathbf{T}^{\#}\mathbf{X}$, where we are denoting pseudo-inverse by superscript #, is a least squares solution of the problem $\mathbf{X} = \mathbf{TS}$, i.e.,

$$||\mathbf{X} - \mathbf{T}\hat{\mathbf{S}}|| = \min_{\mathbf{S}} ||\mathbf{X} - \mathbf{T}\ \mathbf{S}||$$

where $||\cdot||^2$ is the sum of the squares of the elements. When the matrix is full rank, as is the case in this study, the least squares solution is unique.

Based on (3.6), we propose the following quick inverse modeling method: 1) run CMAQ with the best initial estimate of emissions to get $\mathbf{C}_\circ$; 2) interpolate the observations to all the locations of the domain to get $\mathbf{C}_{\text{obs}}$; 3) run Multitracer model in order to obtain the approximate transport matrix $\tilde{\mathbf{T}}$; 4) aggregate emissions to the 100 subregions used in the calculation of the transport matrix to get $\tilde{\mathbf{S}}_\circ$; 5) calculate the new aggregated emissions estimate using

$$\hat{\mathbf{S}}_n = \tilde{\mathbf{T}}^{\#}\left[\left(\frac{\mathbf{C}_{\text{obs}}}{\mathbf{C}_{\text{ref}}}\right)^{\frac{1}{0.7}}\left(\tilde{\mathbf{T}}\tilde{\mathbf{S}}_{\text{ref}}\right)\right]; \tag{3.7}$$

6) calculate adjustment factors for each region as the ratio between the aggregated emissions estimate and the aggregated reference emissions; 7) multiply the base emission field (non-aggregated) by the correction factors to get the corrected emission field.

In order to avoid the artificial discontinuities introduced by the subdivision in subregions, we interpolate (bilinearly) the adjustment factors to get a smoother field. In essence, we get 100 correction factors for the emission which is is a substantial gain from previous methods that could only be applied in practice to get one global correction factor for the whole region [Gilliland et al., 2003].

CHAPTER 4

RESULTS

4.1 CMAQ vs. Tracer

CMAQ ammonia wet deposition constitutes about half of the total deposition, whereas Tracer's ammonia wet deposition only accounts for 25% of the total. As mentioned earlier, most of the CMAQ wet deposition comes from aerosol (NH_4) and most of the dry deposition comes from gas phase (NH_3). Since there is no aerosol formation in Tracer model all the deposition is in the form of NH_3. It is reasonable then that in the Tracer model a larger fraction of ammonia is settled by dry deposition.

Figure 3.4a shows Tracer output vs. CMAQ output under the reference emission scenario $\mathbf{S}_\circ$. The correlation is relatively high (0.92) but the slope is 0.56. This bias is consistent with the proportion of wet deposition in both models. The high correlation between CMAQ output and Tracer output indicates that gas phase ammonia in the Tracer has followed a similar transport path as the aerosol in the full CMAQ. This fact supports the idea that the Tracer output incorporates the right transport information into the predictors. Although the Tracer model does not convert NH_3 into NH_4, the gas phase ammonia is dissolved in water and taken up by clouds at roughly the same rate as aerosols are. The gas phase constituents of the Tracer models do not behave chemically like ammonia but have the right transport and wet deposition properties for our purpose.

We notice that the Tracer's ammonia wet deposition is a biased approximation to full CMAQ. Nevertheless, our predictors are not affected by this bias since we

only use the ratio between Tracer (or Multitracer) model outputs under the new and base emission scenarios. Our predictors combine this imperfect approximation with the base CMAQ output in a way that exploits the information contained in each component and achieve much better approximation than just directly using Tracer output, as will be shown below.

Figure 3.4b shows Tracer output vs. the approximation obtained by multiplying the aggregated transport matrix by the aggregated emission ($\tilde{\mathbf{T}}\tilde{\mathbf{S}}_\circ$). The slope is 0.88 and the correlation is 0.97.

4.2 Predicting CMAQ

In order to assess the performance of the predictors we use three measures of performance: RMSE (root mean square error), slope and MRE (median relative error). The RMSE is the square root of the average squared difference between the predictor and the CMAQ output. The slope is calculated by using least squares regression between the predictand and the predictor, with intercept fixed to 0. Values larger (smaller) than one indicate over-prediction (under-prediction). The MRE is calculated by taking the difference between the predictor and the predictand, dividing it by the predictand and taking the median value. The RMSE penalizes large absolute errors so in general it gives more weight to larger values of predictands, whereas the MRE gives similar weights to large and small values of predictands. MRE captures errors that are usually not easily detected by looking at scatter plots, in which large relative deviation of small values are not apparent.

Figure 4.1 shows with dark crosses the predictor $\mathbf{C}^{(1)}$ from (3.3) vs. the actual CMAQ output under four different target emission scenarios $\mathbf{S}_{0.4Z}$, $\mathbf{S}_Z$, $\mathbf{S}_p$ and $\mathbf{S}_1$. We show the CMAQ output under $\mathbf{S}_\circ$ with gray circles on the same figure to give

an idea of how different the reference and target scenarios are. For $\sigma = 0.4$, which corresponds to changes of the order of 50% relative to the base emissions, the predictor $\mathbf{C}^{(1)}$ does a very good job since all the points are aligned around the unit slope line. For $\sigma = 1$ the changes relative to the base emissions are more extreme so the predictor is more dispersed around the unit slope line. Nevertheless, we get a relative error of less than 6%. For scenario $\mathbf{S}_p$ the performance is good with a median relative error of 12% and a slope of 1.08, which indicates a slight over-prediction. For $\mathbf{S}_1$, which is completely unrelated to the base emission, we see that there is some over-prediction with a slope of 1.29 but it is still substantially improved compared to the base CMAQ output. If we compare Figures 4.1b and c, the predictor under scenario $\mathbf{S}_Z$ seems to be performing much worse than under scenario $\mathbf{S}_p$. This is true if performance is measured by RMSE, which is 0.043 for the former case and 0.010 for the latter case. However, the median relative error under scenario $\mathbf{S}_Z$ is 5.7%, less than half the MRE under scenario $\mathbf{S}_p$, which is 12%.

Figure 4.2 shows the predictor $\mathbf{C}^{(2)}$ from (3.5) vs. the actual CMAQ output under the same target scenarios as in Figure 4.1. We can see that the errors are a bit larger than for predictor $\mathbf{C}^{(1)}$, but the overall performance is comparable.

Table 4.1 summarizes various measures of performance of the predictors. The first three columns of Table 4.1 show the summary of the performance of the first predictor $\mathbf{C}^{(1)}$ for scenarios $\mathbf{S}_{0.1Z}$, $\mathbf{S}_{0.4Z}$, $\mathbf{S}_Z$, $\mathbf{S}_p$, and $\mathbf{S}_1$. The following three columns correspond to predictor $\mathbf{C}^{(2)}$ and the last three columns show the difference between the CMAQ output under the target scenario and the CMAQ output under the base scenario. The latter serve as reference scales for assessing the performance of the predictors. They also serve as measures of the difference between the target and reference scenarios.

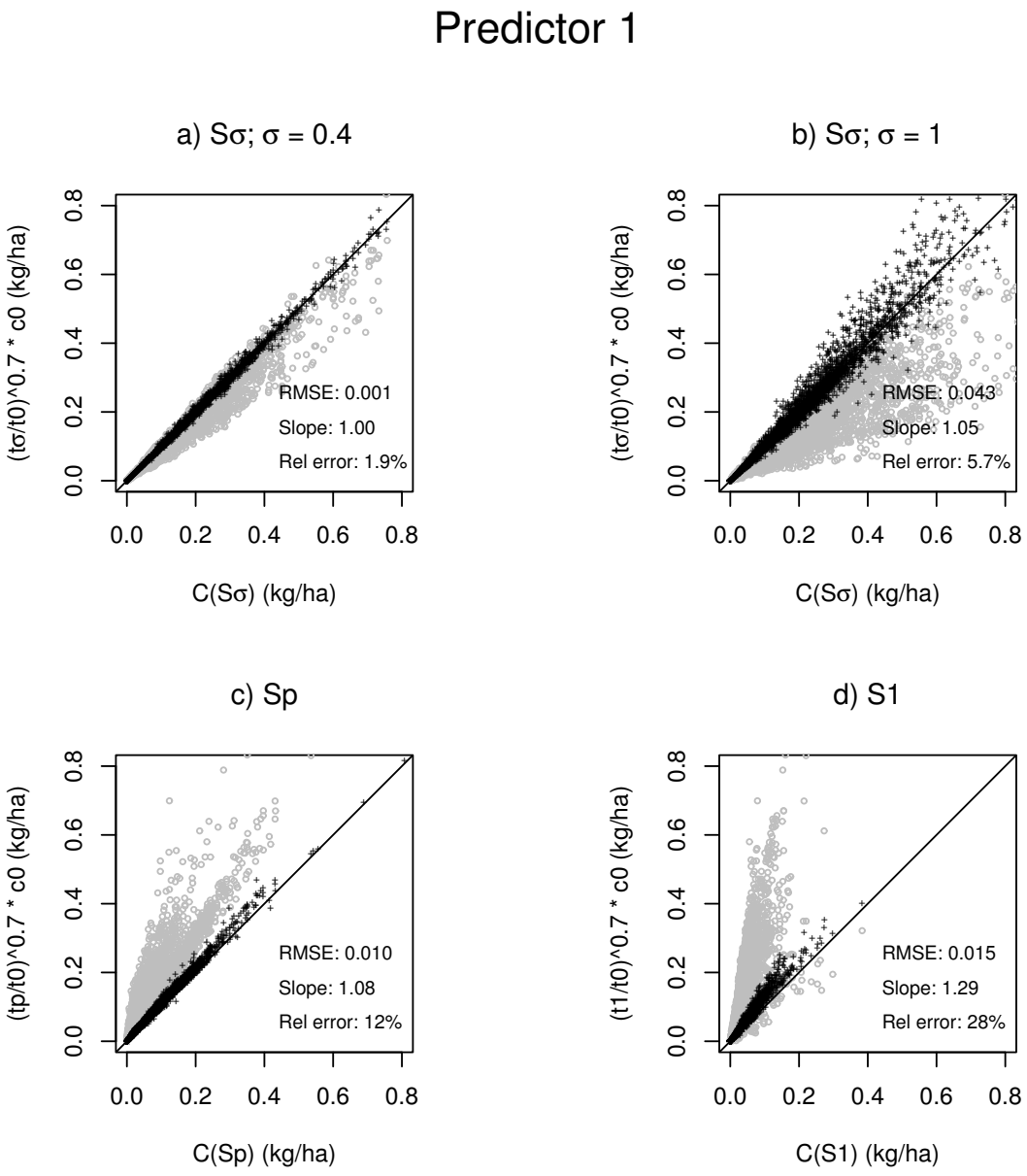

Figure 4.1: Predictor C^1 (dark +) and reference CMAQ output (gray o) vs. CMAQ output under $\mathbf{S}_{0.4Z}$, $\mathbf{S}_{1Z}$, $\mathbf{S}_p$, and $\mathbf{S}_1$.

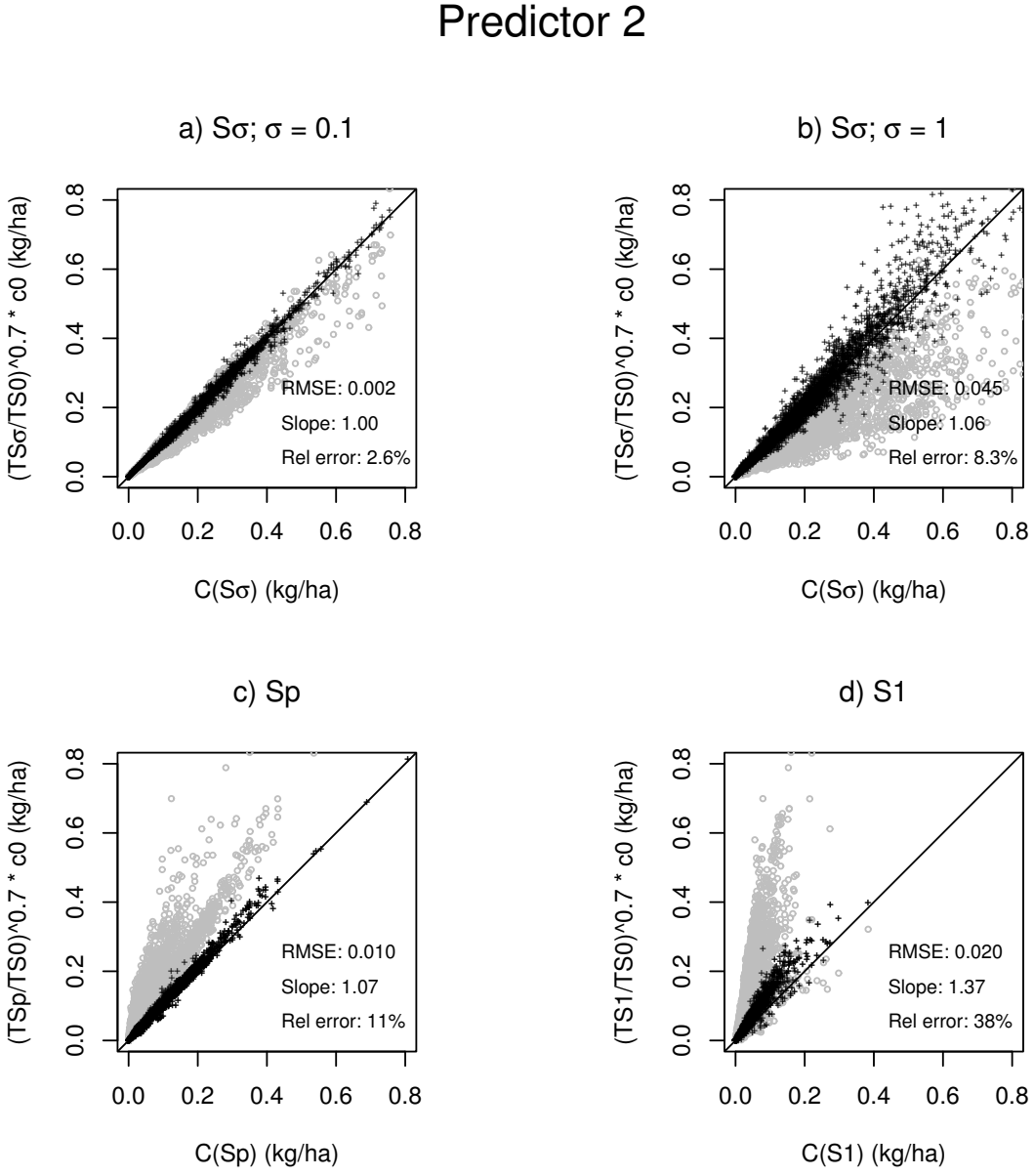

Figure 4.2: Predictor C^2 (dark +) and reference CMAQ output (gray o) vs. CMAQ output under $\mathbf{S}_{0.4Z}$, $\mathbf{S}_{1Z}$, $\mathbf{S}_p$, and $\mathbf{S}_1$.

Table 4.1: Performance measures of predictors 1, 2 and difference between reference and target CMAQ output

	predictor $C^{(1)}$			predictor $C^{(2)}$			$C(\mathbf{S}_\circ)$		
	RMSE	Slope	MRE	RMSE	Slope	MRE	RMSE	Slope	MRE
$\mathbf{S}_{\sigma(0.1)}$	0.001	1.00	0.4%	0.002	1.00	0.7%	0.006	0.98	2.6%
$\mathbf{S}_{\sigma(0.4)}$	0.007	1.00	1.9%	0.008	1.00	2.6%	0.031	0.87	11%
$\mathbf{S}_{\sigma(1)}$	0.043	1.05	5.7%	0.045	1.06	8.3%	0.108	0.61	33%
$\mathbf{S}_p$	0.010	1.08	12%	0.010	1.07	11%	0.079	1.72	75%
$\mathbf{S}_1$	0.015	1.29	28%	0.020	1.37	38%	0.122	3.25	300%

The first three rows show the performance measures of predictor $\mathbf{C}^{(1)}$ under scenarios $\mathbf{S}_{\sigma Z}$ with $\sigma = 0.1$, 0.4, and 1. The RMSE increases from 0.001 to 0.043 as σ increases from 0.1 to 1. The slopes have the ideal value of 1 for $\sigma = 0.1$, and 0.4. For $\sigma = 1$ the slope is 1.05, a slight over-prediction that can be seen in Figure 4.1b. The MRE also increases from 0.4% to 5.7%.

Since the performance of the predictor depends on the difference between the base and target emissions, it makes sense to compare the prediction error with the difference between base and target CMAQ outputs. In all 5 scenarios we see a reduction by a factor of 6 or more in the MRE of the predictor $\mathbf{C}^{(1)}$ compared to the base CMAQ output. The slopes are improved from 0.98 to 1, 0.87 to 1, 0.61 to 1.06, 1.72 to 1.07 and 3.25 to 1.37. The RMSE is improved by factors of 6.0, 4.4, 2.5, 7.9 and 8.1.

For predictor $\mathbf{C}^{(2)}$ the improvements are slightly less dramatic but overall it also performs very well. The RMSEs are improved by factors of 3.0, 3.9, 2.4, 7.9, and 6.1. The slopes improved from 0.098 to 1.00, 0.87 to 1.00, 0.61 to 1.06, 1.71 to 1.07, and 3.25 to 1.37. The MREs are improved by factors of 3.7, 4.2, 4.0, 6.8, and 7.9.

All the computations in this section and the following one were done for the period from 26 June 1996 to 4 July 1996. Some runs were done for the period from 4 July 1996 to 12 July 1996 and no substantial change in the results was found.

4.3 Inverse modeling with simulated observations

In this chapter, we consider the CMAQ outputs to be the "observed" ammonia wet deposition and the emissions to be unknown. We use (3.7) to get an estimate of the emission field. Figure 4.3 shows the estimated emissions (dark cross) and the base emissions (gray circles) vs. the target emissions for scenarios $\mathbf{S}_{\sigma Z}$ with $\sigma = 0.4$ and $\sigma = 1$, $\mathbf{S}_p$, and $\mathbf{S}_1$. In all four cases we see an improvement in the slope and dispersion of the new estimated emissions as compared to the initial emissions $\tilde{S}_\circ$. Table 4.2 shows the RMSE, slope and correlation of these estimates and the initial emission scenario compared to the target emission fields. The first 3 columns correspond to the estimated emission fields relative to the actual target emission fields and the second 3 columns correspond to the initial emission field $\mathbf{S}_\circ$ relative to the target emission fields. The RMSEs, slopes, and correlations of the estimated emission fields are consistently better than the initial field, which indicates that the inverse modeling method proposed should be useful for improving the emission fields. The RMSEs are reduced by factors of 2.3, 2.4, 1.9, 5.3, and 4.5 for the emission fields $\mathbf{S}_{0.1Z}$, $\mathbf{S}_{0.4Z}$,$\mathbf{S}_Z$, $\mathbf{S}_p$, and $\mathbf{S}_1$, respectively. In all 5 scenarios the slopes are improved substantially. The correlations are also improved from 0.99 to 1, 0.89 to 0.98, 0.53 to 0.87, 0.85 to 0.96 and 0.04 to 0.43. In this section we compare only the aggregated emission fields, i.e., we do not apply the steps 6 and 7 from the proposed method because the performance of the method is well represented by the aggregated emissions.

4.4 Inverse modeling with actual observations

We use NH_3 monthly wet deposition concentration data from the National Atmospheric Deposition Program National Trends Network (NADP) (available at http://nadp.sws.uiuc.ed
We have 63 sites within our domain with valid observations for the period of 2 July

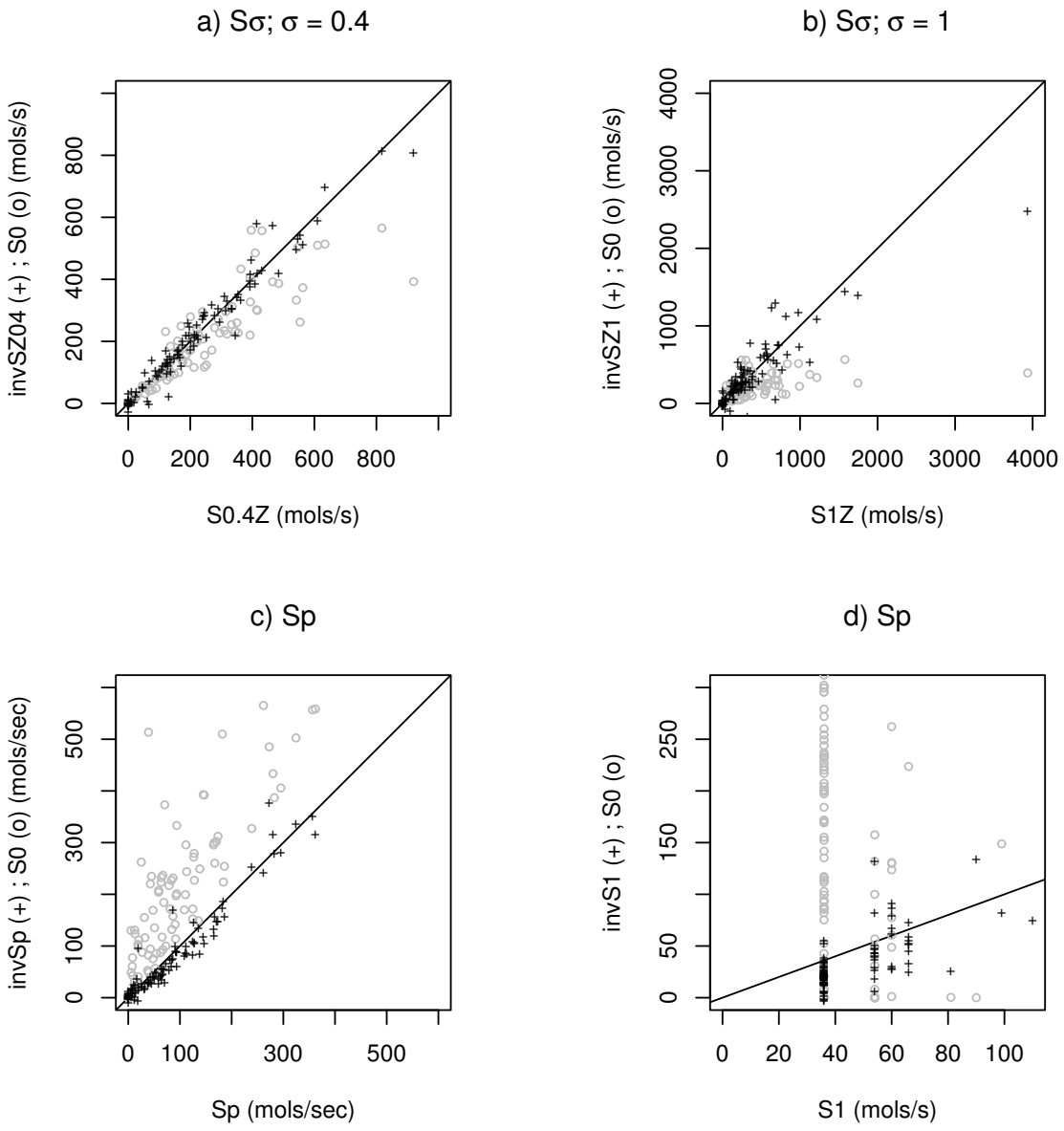

Figure 4.3: Estimated emissions (dark +) vs. actual emissions for $\mathbf{S}_{0.4Z}$, $\mathbf{S}_{1Z}$, $\mathbf{S}_p$, and $\mathbf{S}_1$. Initial emission $\mathbf{S}_\circ$ is included for comparison (gray o).

Table 4.2: Summary statistics of estimated emission fields and base emission field compared to the target field.

	$\hat{S}_n$ vs. $\mathbf{S}_n$			$\mathbf{S}_\circ$ vs. $\mathbf{S}_n$		
	RMSE	Slope	Corr	RMSE	Slope	Corr
$\mathbf{S}_{\sigma(0.1)}$	8.2	1.00	1.00	18.6	0.97	0.99
$\mathbf{S}_{\sigma(0.4)}$	39.0	0.99	0.98	92.2	0.78	0.89
$\mathbf{S}_{\sigma(1)}$	242.6	0.80	0.87	465.5	0.28	0.53
$\mathbf{S}_p$	23.8	0.95	0.96	127.4	1.78	0.85
$\mathbf{S}_y 1$	43.5	0.77	0.43	197.4	3.50	0.04

1996 to 30 July 1996. We have run CMAQ for the same period using the base emission $\mathbf{S}_\circ$ and aggregated the ammonia wet deposition for the whole period. We have also run the Multitracer model in order to obtain the transport matrix for the same period of time. Figure 4.4 shows the observed wet deposition vs. the CMAQ wet deposition at these sites. The straight line has the least squares regression slope, which is 0.29. The correlation between observations and CMAQ is only 0.27 and the modeled depositions are generally much higher than the observed depositions. This poor agreement suggests that a method based on assuming that emission errors are the main source of the discrepancies between CMAQ and observations may have problems. Nevertheless, as an illustration of our method, we apply the inverse modeling scheme, keeping in mind that one would want to reduce other sources of error (e.g., meteorology, especially rainfall amount) before one could reliably use these estimated adjustments in practice.

We use concentrations instead of depositions because the latter is less sensitive to errors in precipitation [Gilliland et al., 2003, Stein et al., 1993, Styer and Stein, 1992]. This has the same effect as adjusting the wet deposition by the ratio of actual and model precipitation as done by Yarwood et al. [2003].

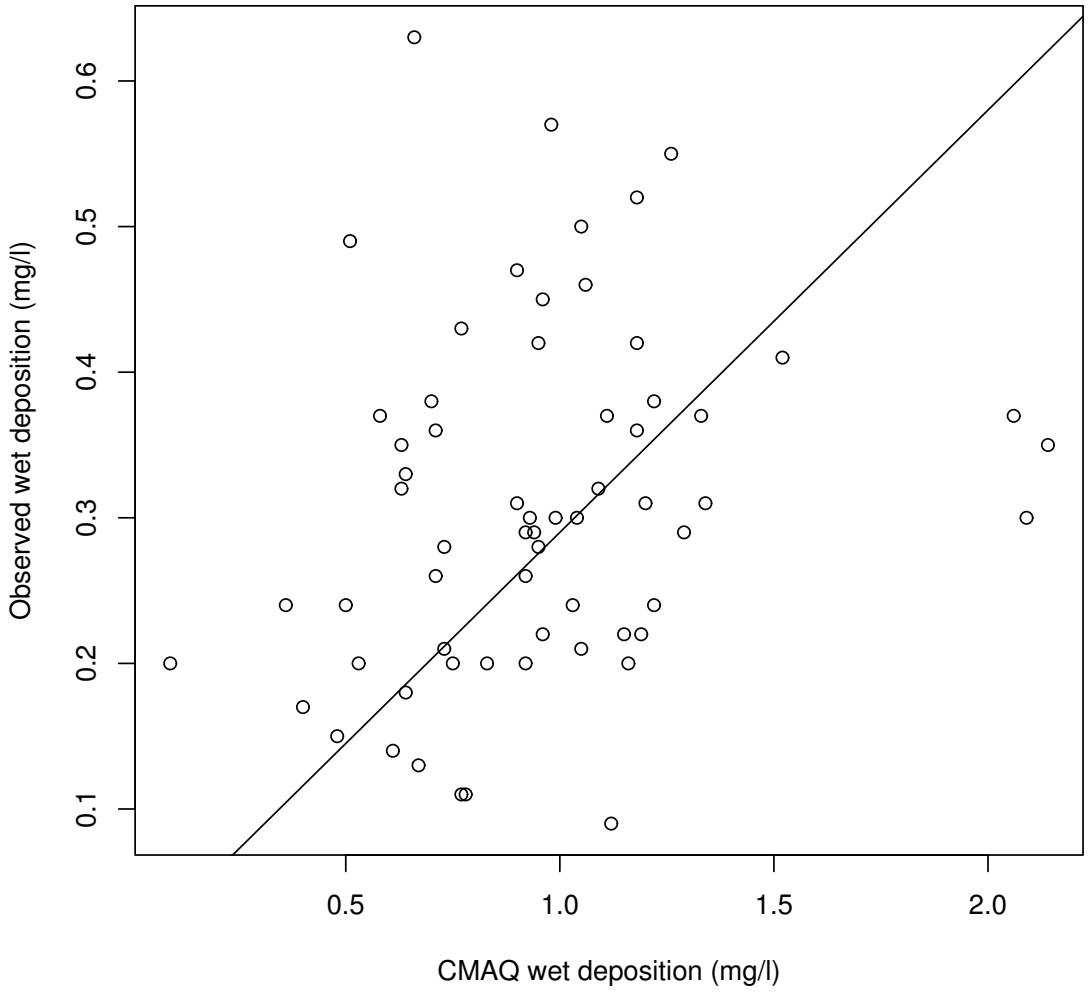

Figure 4.4: Observed ammonia wet deposition concentration (NADP) vs. CMAQ wet deposition concentration for the period from 2 July 1996 to 30 July 1996. The slope of the straight line is 0.29.

In order to interpolate the observations we modeled the observation as a constant mean plus a stationary and isotropic Gaussian random field characterized by a covariance function in the Matérn class [Stein, 1999] plus independent observation errors. We estimated the parameters of the observed process by maximizing the restricted likelihood [Stein, 1999].

When we calculated the estimated emissions with (3.7) we ran into problems because CMAQ wet concentration, which appears in the denominator, can take values that are very close to zero. A quick fix of the method was to interpolate the (log of the) ratio between the observations and CMAQ output instead of the observation itself. An additional benefit of this approach is that the log of the ratio between observed and CMAQ concentration has a simpler correlation structure than the observation field itself so it is better represented by a stationary and isotropic Gaussian random field [Jun and Stein, 2004]. In this modified method, we only use the CMAQ output at the observation sites. This makes sense since we do not have enough data to get good estimates of the ratio between observation and CMAQ away from the observation sites. Figure 4.5 shows the estimated emission field vs. the base emission field and the least squares line has a slope of 0.27, which is close to the least squares fit of observation vs. CMAQ. So the the overall average correction is close to what we would expect. We did get two regions where the new emissions were negative but their magnitudes were small, less than 0.02 mols/s, and corresponded to regions where the base emissions were close to 0. Figure 4.6 shows a) the interpolated values of the ratio between observed wet deposition and CMAQ output and b) adjustment factors for emissions interpolated to the 4556 cells of the domain. The ratio between observed and CMAQ depositions show very small values in the Southeast; the model is overpredicting by a factor of 3 to 4 in this region and by a factor of at least 2 in the

remaining regions. In 10b the regions where the reference emissions were close to zero are represented in white. The adjustment field has a similar pattern to the ratio between observation and CMAQ field with a shift to the West. This is reasonable since in order to compensate for the small ratio between observation and CMAQ in the lower left corner of the figure one would need to decrease the emissions upwind.

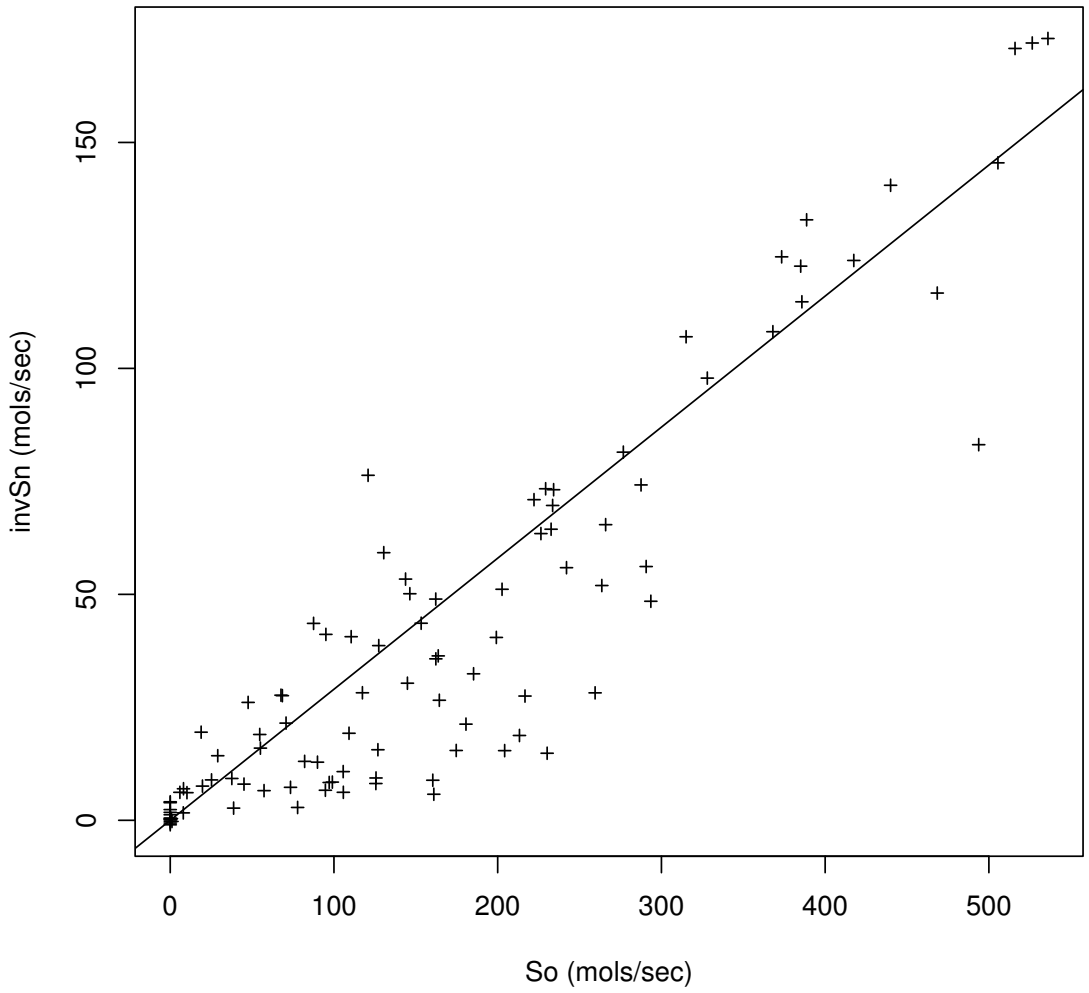

Figure 4.5: Inverse modeled ammonia emission vs. initial emission estimate $\mathbf{S}_\circ$. These are aggregated values in each of the 100 subregions. The slope of the straight line is 0.27.

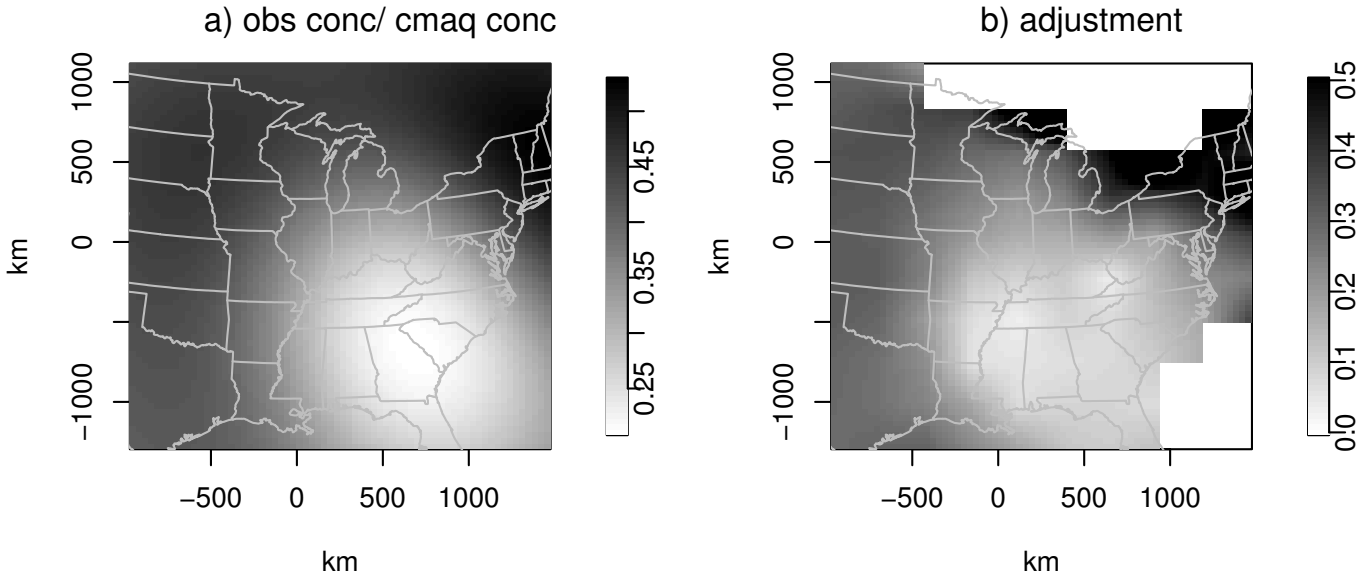

Figure 4.6: a) interpolated ratio between observed and CMAQ ammonia wet deposition concentration; b) adjustment factors interpolated bilinearly. The regions where the base emissions were close to zero are shown in white.

As far as we are aware of, there are no previous studies where emissions adjustments are calculated with spatial variation like in our study. Gilliland et al. [2003] report overall adjustment factors for the whole Eastern United States region for some of the months in 1990. Their adjustment factor for July 1990 was close to 1, whereas our adjustment factor for July 1996 is around 0.27, the least squares regression slope of the estimated emission field vs. the base emission field. This discrepancy would be smaller if we performed a 15% upward adjustment to the observed ammonia wet deposition as done by Gilliland et al. [2003] based on the comparisons of daily vs. weekly sampled ammonia wet deposition data at 4 sites in the NADP network [Butler and Likens, 1998]. Another difference between their runs and ours is that we used 1996 National Emission Inventory instead of the 2000 inventory. Further study is needed to evaluate the source of the substantial discrepancy.

CHAPTER 5

DISCUSSION

We have proposed two fast methods to approximate CMAQ ammonia wet deposition under new emission scenarios when we have CMAQ output under a reference emission scenario. The first method requires 1/40th the processing time of CMAQ. The second method requires computing a transport matrix once, which requires roughly the same time as a CMAQ run. After that, a simple matrix multiplication between the matrix and any new emission vector gives good approximations to CMAQ under new emission scenarios. These methods allow us to solve forward and inverse problems that were in practice not tractable because of the computational burden.

We have found that the performance of the predictor depends on the difference between the base and target emission fields. The closer the scenarios the better the predictors work. But even for emissions scenarios that are quite different from the base scenario, we get predictions that are close to the actual CMAQ output. This means that our predictors can be safely used for inverse modeling purposes. In order to get more accurate results we can use our predictors to quickly search in the space of emission scenarios not too far from the base case. When the new emission and base emission fields are judged to be sufficiently different, execute a new CMAQ run, use the new output as the reference CMAQ and continue the process.

Based on the form of predictor $\mathbf{C}^{(2)}$, we have proposed a new approach to inverse modeling ammonia emissions. This method requires running the Multitracer model once in order to get emission corrections for each of the 100 subregions of the domain.

This allows us to obtain spatially varying corrected emission fields with resolution of the order of 500km. We generated pseudo-data by running CMAQ under different emission fields to test our method. We found that the performance depends on the difference between the base emission and the target emission. Our method gave a substantially improved emission field compared to the base emission field.

We have applied our inverse modeling method using observed NH_3 wet deposition data. Since the match between observation and model is very poor we do not trust the actual values of the estimated emissions. Nevertheless, the relationship between the deposition and the estimated emission adjustments did show a reasonable westward shift reflecting the transport of ammonia between sources and sinks and the right overall average correction factor.

The ill-posed nature that affects most inverse problems was addressed here by two features of the method. First, the fact that we use an aggregated transport matrix forces the calculated emissions to be aggregated, and, as a consequence, smoother than what one would get with a separate adjustment for each of the 4556 cells of the domain. We aggregated the data because of resource constraints but this aggregation had the added benefit of smoothing the adjustment field. The second feature is the interpolation of the ratio between observation and CMAQ output that also has a smoothing effect. Further study on how to regularize adjustment factors is needed.

A forward problem that can be addressed with large reduction in computation time is the comparison between different ammonia inventories. For example, the CMU Ammonia Emission Inventory developed an alternative ammonia inventory for the continental US. One could check which of the two inventories, CMU or EPA (the one used for our runs), matches better with observations by using our second predictor. In practice, one would need to have a much better agreement between

CMAQ and observations in order to trust the result of this comparison. This may be the case if the aggregation period is much longer than a few weeks. We will pursue this work in the future.

Extension of our results to wet deposition of other species should be possible as long as the nonlinearity is not severe and one chooses the right temporal aggregation scale. A good preliminary test of whether the approximation method can be applied or not is to compare the output under two emission scenarios that only differ by a constant factor and check how much information the output under one scenario has about the output under the other scenario.

Part II

Semiparametric estimation of spectral density with irregular observations

CHAPTER 6

INTRODUCTION

Estimation of the covariance structure of physical processes observed in a finite set of locations is fundamental to understand the behavior of such processes and to interpolate those values to locations where measurements are not available. Kriging, an interpolation method widely used by the geophysical community, is based on the knowledge of the covariances between observed and interpolated locations.

We will concentrate on isotropic processes, which are invariant under rigid motions like translation, rotation and reflection. Under this assumption, the covariance of the process at two locations only depends on the distance between them so a covariance function on $\mathbb{R}^+$ fully describes the second order properties of the process. This function has to be positive definite in order to ensure that the variance of any linear combination of values of the process at various locations is positive. This constraint has led most people to restrict the estimation to parametric forms that were proven to be positive definite.

There has been some work in using nonparametric methods or a broad class of positive definite functions based on the spectral representation of covariance functions. Before describing these methods some properties of positive definite functions are in order. Bochner's theorem [Yaglom, 1987] states that a function is continuous and positive definite if and only if it is the Fourier transform of a positive bounded measure on $\mathbb{R}^d$, i.e.,

$$C(\mathbf{x}) = \int_{\mathbb{R}^d} \exp(i\mathbf{w}\mathbf{x}) F(\mathbf{dw}). \tag{6.1}$$

For isotropic processes (6.1) can be reduced to a one-dimensional integral

$$C(r) = 2^{(d-2)/2}\Gamma(d/2)\int_0^\infty (ru)^{-(d-2)/2} J_{(d-2)/2}(ru)dG(u) \tag{6.2}$$

where $G(u) = \int_{|\mathbf{w}|<\mathbf{u}} dF(\mathbf{dw})$ is a bounded positive measure on $\mathbb{R}$, $\Gamma(\cdot)$ is the Gamma function, and $J_\nu(\cdot)$ is the Bessel function of the first kind of order ν [Abramowitz and Stegun, 1965].

Shapiro and Botha [1991] proposed using a finite discrete measure with nodes placed at $t_1, ..., t_n$ so that the integral in (6.2) is reduced to a finite sum:

$$\tilde{C}(r) = \sum_{j=1}^{m} p_j \Omega_d(t_j h), \tag{6.3}$$

where the p_j's are positive and $\Omega_d(x) = \left(\frac{2}{x}\right)^{(d-2)/2} \Gamma\left(\frac{d}{2}\right) J_{(d-2)/2}(x)$. They use a raw covariogram estimate given by

$$\hat{C}(h) = \frac{1}{N(h)} \sum_{N(h)} (Z(\mathbf{x}_i) - \bar{Z})(Z(\mathbf{x}_j) - \bar{Z}), \tag{6.4}$$

where $Z(\mathbf{x}_i)$ are observations at site i, $\bar{Z}$ is the average of the observations, the sum runs over all pairs of observations that are approximately distance h apart, and $N(h)$ is the total number of such pairs. They estimate the values of p_j by minimizing the mean squared difference between the raw covariogram estimate at different lags and their estimator, with positivity constraint.

Genton and Gorsich [2002] follow this idea but propose using the zeros of the Bessel functions as the nodes of the discrete measure and show that their method is computationally simpler, needs fewer nodes, and does not show spurious oscillations. The problem with this choice of nodes is that these numbers are nondimensional so

it is not clear what scale should be used to translate these nodes into nodes in the frequency domain (that has dimension 1 / unit of distance). Although not totally explicit, they seem to propose using $1/r_{\max}$ as their scale. This would mean that if we added an additional observation at distance 1.5 $r_{\max}$ then the nodes would be shifted by a factor of 1.5 in the frequency domain. This behavior seems problematic.

Hall et al. [1994] propose using a kernel estimator for a preliminary covariogram estimate and, in order to ensure positive definiteness, they propose Fourier transforming the kernel estimator, setting the negative values to zero and Fourier transforming back to the spatial domain. For $d = 2$, denoting $\hat{Z}_{ij} = (Z(\mathbf{x}_i) - \bar{Z})(Z(\mathbf{x}_j) - \bar{Z})$, h_{ij} the distance between the observations Z_i and Z_j, K a kernel function (a positive symmetric probability density), and δ the bandwidth, the first step estimator of the covariogram is

$$\tilde{C}(h) = \frac{\sum_{i,j} \hat{Z}_{ij} K(\frac{h-h_{ij}}{\delta})}{\sum_{i,j} K(\frac{h-h_{ij}}{\delta})}. \tag{6.5}$$

The final estimate of the covariogram is

$$\bar{C}(h) = \int_0^\infty \left(\int_0^\infty \tilde{C}(x) x J_o(wx) dx \right)_+ w J_o(wh) dw. \tag{6.6}$$

In practice, one would need to truncate this integral to a finite domain. We will call this function the kernel estimator of the covariance function.

All three methods use the raw covariogram as the basis for estimation, which ignores the correlation between the values of the observations at different distances. Furthermore, it is well known that the high frequency properties of the spectral density determine the performance of interpolation procedures [Stein, 1999]. None of the above methods give proper consideration to the tail properties.

We propose a flexible family of models for the spectral density that is a linear

combination of B-splines of order 4 (cubic splines) up to a cutoff frequency w_t and an algebraically decaying tail from w_t to infinity. We use positive coefficients for the B-splines, which ensure positiveness of the spectral density and, as a consequence, positive definiteness of the covariance function. Assuming the process is well described by a Gaussian random field, we find the parameters that maximize the likelihood. This method estimates the tail property of the spectral density in an explicit way. It excludes exponential decay of the tail but we consider this restriction to be beneficial since such a fast decay would imply an unrealistically smooth process [Stein, 1999]. It also excludes oscillatory tails, say $w^{-\gamma}\cos^2 w$, but this behavior is generally undesirable. Additionally, through its use of likelihood, our method takes fully into account all the correlations between observations. This is, to the extent of our knowledge, the first work that uses a likelihood approach for scattered spatial data without a parametric model.

In chapter 7 we present our model and the methodology to estimate the covariance function. In chapter 8 we describe how the numerical issues that arise when calculating the likelihood were solved. In chapter 9 we describe several performance measures and compare our method with a parametric method using the Matérn model and Hall et al.'s [1994] nonparametric kernel method.

CHAPTER 7

METHODOLOGY

7.1 The splines+tail (S+T) model

We represent the spectral density as a linear combination of B-splines of order 4 (See appendix A for a brief description and references) with node sequence $(w_0, ..., w_l)$ on the interval $(0, w_t)$ for some threshold frequency w_t and after this point an algebraically decaying tail with power $-\gamma$, i.e.,

$$f(w) = \sigma^2 \sum_{i=-1}^{l+1} b_i B_i(w) \mathbb{I}(0 \leq w \leq w_t) + f_t \left(\frac{w_t}{w}\right)^{\gamma} \mathbb{I}(w > w_t), \tag{7.1}$$

The sum goes from -1 to $l+1$ in order to include all B-splines that have support on the interval $(0, w_t)$. We chose order 4 because of the flexibility that cubic splines give to represent a wide range of functions. B-splines of other orders could be used with minor adjustments. We require the spectral density to be continuous and have continuous derivative at w_t. The value of the constant f_t is chosen to achieve continuity at w_t; more explicitly, $f_t = \sum_{i=-1}^{l+1} b_i B_i(w_t)$. The coefficients of the B-splines are constrained to be positive except for b_{l+1} which is chosen so that the derivative of $f(w)$ is continuous at w_t. It is shown in Appendix C that the function is still positive. Restricting the coefficients to be positive is a simple way of ensuring positivity of the function. The B-spline coefficients closely follow the function they represent as the number of nodes increases so they eventually will become positive for a positive

twice differentiable function. This suggests that the positivity condition is not too restrictive.

Let us briefly describe the Matérn model in order to compare it to our model. This class is considered to be a sensible model for a wide range of processes arising in environmental problems [Stein, 1999, Handcock and Wallis, 1994]. With only three easily interpretable parameters (σ^2, ρ and ν), the Matérn class allows great flexibility in the type of processes it can represent. σ^2 is simply the variance of the process at a given location, ρ is a measure of the range at which the correlation becomes negligible and ν is a measure of the smoothness of the process. A process with smoothness parameter ν is $\lceil \nu \rceil$ - 1 times (mean square) differentiable, where $\lceil \cdot \rceil$ represents ceiling of the number. The spectral density of the Matérn class has the form

$$f(w) = \frac{\sigma^2 \lambda(\rho, \nu)}{(\frac{4\nu}{\rho^2} + w^2)^{\nu + d/2}} \tag{7.2}$$

with $\lambda(\rho, \nu) = \frac{\Gamma(\nu + d/2)}{\pi^{d/2}\Gamma(\nu)} (\frac{2\sqrt{\nu}}{\rho})^{2\nu}$ such that the variance, $C(0)$, is σ^2.

At high frequencies both the Matérn and our model show the same behavior $1/w^\gamma$ with $\gamma = 2\nu + d$. In what follows, we will use ν as the parameter for the S+T model.

7.2 Estimation method

Likelihood

We assume the observations come from realizations of a Gaussian random field whose value at location x is of the form

$$Z(\mathbf{x}) = \mathbf{m}(\mathbf{x})^T \beta + \epsilon(\mathbf{x}) \tag{7.3}$$

where $\mathbf{m}(\mathbf{x})$ is a known vector valued function, β is a vector of unknown coefficients and ϵ has mean 0 with covariance function $C(\epsilon(\mathbf{x}), \epsilon(\mathbf{y})) = C_\theta(|\mathbf{x}-\mathbf{y}|)$. θ is the vector of unknown parameters of the covariance function. Assuming we have observations $\mathbf{Z} = (Z(\mathbf{x}_1), ..., Z(\mathbf{x}_n))^T$ and denoting $\mathbf{M} = (\mathbf{m}(\mathbf{x}_1)...\mathbf{m}(\mathbf{x}_n))^T$, the loglikelihood has the form (ignoring an additive constant)

$$-\frac{1}{2}\log|\det\mathbf{\Sigma}_\theta| - \frac{1}{2}(\mathbf{Z}-\mathbf{M}\beta)^T\mathbf{\Sigma}_\theta^{-1}(\mathbf{Z}-\mathbf{M}\beta) \tag{7.4}$$

where $\mathbf{\Sigma}_{ij} = C_\theta(|\mathbf{x}_i - \mathbf{x}_j|)$.

Denoting $\mathbf{W} = \mathbf{M}^T\mathbf{\Sigma}_\theta^{-1}\mathbf{M}$, the restricted likelihood [Stein, 1999, McCullagh and Nelder, 1989] is (ignoring an additive constant)

$$-\frac{1}{2}\log|\det\mathbf{\Sigma}_\theta| - \frac{1}{2}\log|\det\mathbf{W}| - \frac{1}{2}\mathbf{Z}^T(\mathbf{\Sigma}_\theta^{-1} - \mathbf{\Sigma}_\theta^{-1}\mathbf{M}\mathbf{W}^{-1}\mathbf{M}^T\mathbf{\Sigma}_\theta^{-1})\mathbf{Z}. \tag{7.5}$$

Covariance function

The covariance function is calculated from the spectral density $f(w)$ by applying (6.2) for $d = 2$:

$$C(r) = 2\pi\int_0^\infty wJ_o(rw)f(w)dw. \tag{7.6}$$

For f in the S+T family we can calculate the integral in (7.6) analytically. The transform in (7.6) is called the Hankel transform of order 0. The Hankel transform of B-splines requires calculating two Bessel functions of the first kind of orders 1 and 2 and two Struve functions of orders 1 and 2 [Abramowitz and Stegun, 1965] for each node. So the computation of this part is straightforward albeit moderately time consuming. See Appendix B for details. The Hankel transform of the truncated tail

is

$$\begin{aligned}
&\int_{w_t}^{\infty} w^{1-\gamma} J_o(wr)dw \\
=& r^{\gamma-2} \int_{w_t r}^{\infty} u^{1-\gamma} J_o(u)du \\
=& r^{\gamma-2} \left(\frac{\gamma\Gamma(-\gamma/2)}{2^{\gamma}\Gamma(\gamma/2)} + \frac{(w_t r)^{2-\gamma} {}_1F_2(1-\gamma/2; 1, 2-\gamma/2; -(rw_t)^2/4)}{\gamma-2} \right)
\end{aligned} \tag{7.7}$$

where ${}_1F_2(a; b, c; z)$ is a generalized hypergeometric function with series representation $\sum_{k=0}^{\infty} \frac{(a)_k}{(b)_k (c)_k} \frac{z^k}{k!}$. Here $(\cdot)_k$ represents the Pochhammer's symbol [Abramowitz and Stegun, 1965], which is defined by $(z)_0 = 1$ and $(z)_k = z(z+1)(z+2)...(z+n-1) = \frac{\Gamma(z+n)}{\Gamma(z)}$. For $b = 1$ and $c = a+1$ the series reduces to ${}_1F_2(a; 1, a+1; z) = \sum_{k=0}^{\infty} \frac{a}{a+k} \frac{z^k}{k!^2}$.

Estimating the parameters

The parameters of our model are smoothness (ν), sill (σ^2), cutoff frequency (w_t), and the coefficients of B-splines (b_i for $i = 0, ..., l$). These are estimated by maximizing the likelihood when the mean is known or the restricted likelihood otherwise. The coefficients b_{-1} and b_{l+1} determine the derivatives of the function at the end points. We chose b_{-1} to equal b_1 and b_{l+1} to be such that the derivative of the function is continuous at w_t (See Appendix C).

As explained in chapter 8, the need for high computational speed forced us to discretize some of the parameters (smoothness and cutoff frequencies) so the usual continuous optimization routines are not applicable. Because of this and the potentially large number of parameters involved, we decided to use the simulated annealing method to maximize the likelihood.

We propose estimating the covariance functions using different numbers of nodes $(l + 1)$ for the spectral density. The number of nodes is selected using the Akaike

Information Criterion (AIC). AIC selects the model with the smallest value for $-2\text{loglikelihood} - 2k$ with k being the number of parameters in the model. Other penalizations should be considered. With the AIC, one additional node in the model should be compensated by an increase of at least one unit of loglikelihood in order to accept the larger model. In the simulation study presented in chapter 9 , we fixed the number of nodes for each estimating model after a few testing simulations. In the future, we will estimate the number of nodes for each simulation.

At the moment, we have restricted the B-splines to have uniformly distributed nodes. So given the number of nodes $(l+1)$ and the cutoff frequency (w_t), we place the nodes at locations iw_t/l for $i = -1, 0, ..., l, l+1$.

CHAPTER 8

NUMERICAL IMPLEMENTATION

8.1 The Hankel Transform

Several features make the Hankel transform of the tail (7.7) numerically hard to compute. First, there is no easy way of evaluating this hypergeometric function accurately without resorting to summing a large number of terms of its series expansion, which can lead to severe numerical errors. This problem was addressed by using arbitrary precision arithmetic libraries. Second, the Γ function is infinite when the argument is a negative integer and we have no reason to exclude negative integer values for $1 - \gamma/2 = -\nu$. However the divergence of the Γ function is compensated by the divergence of one of the terms of the series expansion of the hypergeometric function. Third, the first term ($\frac{\gamma\Gamma(-\gamma/2)}{2^\gamma\Gamma(\gamma/2)}$) is the limit of the second term ($\frac{(w_t r)^{2-\gamma}{}_1F_2(1-\gamma/2;1,2-\gamma/2;-(rw_t)^2/4)}{\gamma-2}$) as the rw_t goes to infinity so in the case where rw_t is large we need to take the difference of two very similar numbers.

The problem when ν is an integer can be solved by using an asymptotic expansion of the Γ function when the argument is close to a negative integer and subtracting it from the series expansion of the hypergeometric function. Only one term in each series expansion diverges as the arguments approaches a negative integer. So we get a modified series expansion for the difference between the two terms which can be computed in the same fashion as the hypergeometric function, i.e., by adding the series until convergence is achieved and using arbitrary precision libraries to avoid

numerical errors. When $\gamma/2-1=n$ is a positive integer, the final expression for the tail integral is

$$\begin{aligned}&\int_{w_t}^{\infty} w^{1-\gamma} J_o(wr)dw \\ &= r^{2n}\frac{\log(2)-\log(rw_t)+\psi(n+1)}{(-4)^n n!^2}+\frac{w_t^{2n}}{2n}\sum_{k=0,k\neq n}^{\infty}\frac{-n}{-n+k}\frac{(-(rw_t)^2/4)^k}{k!^2}\end{aligned} \tag{8.1}$$

where $\psi(n)$ is the digamma function [Abramowitz and Stegun, 1965], which for positive integer arguments can be evaluated as $\sum_{k=1}^{n-1}\frac{1}{k}-\gamma_{eg}$ where $\gamma_{eq}=0.577216...$ is the Euler's constant. The details are shown in Appendix D.

The problem of subtracting two very similar numbers is solved by using an asymptotic expansion of the hypergeometric function, whose leading term is $\frac{\gamma\Gamma(-\gamma/2)(\gamma-2)}{2^\gamma\Gamma(\gamma/2)(rw_t)^{2-\gamma}}$, so we are left with an expression that directly calculates the difference. See details in Appendix E. For large values of rw_t the Hankel transform of the truncated tail is approximated by

$$\begin{aligned}&\int_{w_t}^{\infty} w^{1-\gamma} J_o(wr)dw \\ &\approx \left(\frac{\cos(rw_t)-\sin(rw_t)}{\sqrt{\pi}(rw_t)^{\gamma-\frac{1}{2}}}-\frac{\left(-15+16\,\gamma+128\,\gamma^2\right)\,(\cos(rw_t)-\sin(rw_t))}{128\,\sqrt{\pi}(rw_t)^{\gamma+\frac{3}{2}}}\right. \\ &\quad \left.+\frac{(-3+8\,\gamma)\,(\cos(rw_t)+\sin(rw_t))}{8\,\sqrt{\pi}(rw_t)^{\gamma+\frac{1}{2}}}+\dots\right) r^{\gamma-2}\end{aligned} \tag{8.2}$$

8.2 Simulated Annealing

We implemented the maximization using the simulated annealing method [Givens and Hoeting, 2005]. This method is based on the way a physical system finds its minimum energy state when it is first heated to high temperature and then cooled down slowly to zero temperature. In our problem, the energy to be minimized is the negative

of the loglikelihood function. One starts with an initial value of the parameters and calculates the energy. New values of the parameters are drawn from a proposal distribution and the new energy is calculated. If the new energy is lower than the previous one the parameters get updated with the new values. If the new energy is higher than the initial energy, the new parameters are accepted with probability $\exp(-(E_f - E_i)/T)/(1 + \exp(-(E_f - E_i)/T))$, where E_i and E_f are the initial and final energies and T is the temperature. This helps the system not to be trapped in local minima. These steps are repeated several times after which the temperature is lowered and the same procedure is followed until the temperature is close to 0.

We have noticed that the convergence depends on the starting values of the parameters, most notably on the threshold frequency w_t. We use three different starting values of w_t ($1/r_{\min}$, $1/r_{\max}$, and the average of the two) and choose the estimated parameters that have the largest likelihood. We start fitting a Matérn model and use the estimated sill (σ) and smoothness (ν) as starting values for the optimization. The initial values for the coefficients of the B-Splines are set to be constant 1 at all nodes. Since we normalize the spectral density so that it yields variance σ^2 the overall scale of this coefficients is not relevant.

The proposal distribution for the coefficients of the B-splines is a mixture of log-normals, one with mean parameter 0 and variance parameter 1 and the other one with mean centered at the initial value and variance 0.1. The proposal distribution for the sill is also a mixture of normals, one with mean given by the sample variance of the observations and variance 1 and the other one centered at the initial value and variance 0.1. These numbers were chosen so that the convergence was satisfactory. We let w_t take 100 discrete values between $1/r_{\max}$ to $1/r_{\min}$. The proposal was a mixture of two uniform distributions, one that ranges over all 100 values and the

other one centered at the previous value and a range that is 10% of the whole range. Likewise, we let the smoothness parameter take 100 discrete values between 0.05 and 5. Larger values of smoothness give rise to almost singular covariance matrices. The proposal was also a mixture of uniforms, one centered at the previous value with a range that is 10% of the total range and the other one over the whole range.

Several cooling schedules were tested. The one that gave slightly better convergence was one which updated the temperature in each step according to $T_i = \frac{T_{i-1}}{1+aT_{i-1}}$ with $a = 30$ and $T_0 = 1000$. We stopped after 10000-20000 iterations, after which no changes in parameters occurred. Each optimization took around 2-5 minutes.

8.3 Tabulation of Hankel Transforms

In order to speed up the computation, we had to resort to some further approximations and shortcuts. We calculated the covariance function for $n_r = 100$ equispaced values between $r_{\min}$ and $r_{\max}$ and interpolated using cubic spline interpolation for distances between these points. Also, we restricted the values of the threshold frequency w_t and the power of the tail γ to $n_w = 100$ and $n_\gamma = 100$ discrete values. Namely, $w_t(j) = \frac{1}{r_{\max}} + \frac{j}{n_w}\left(\frac{1}{r_{\min}} - \frac{1}{r_{\max}}\right)$ and $\gamma(j) = 2 + 2\left(0.05 + \frac{j}{n_\gamma}(5 - 0.05)\right)$.

The Hankel transform of the truncated tail $\left(t(i,j,k) = \int_{w_t(k)}^{\infty} w^{1-\gamma(j)} J_o(wr(i))dw\right)$ was tabulated into an array of dimensions $n_r \times n_\gamma \times n_w = 100 \times 100 \times 100$. The Hankel transform of piecewise polynomials of the form $\mathbb{I}([w_i, w_{i+1}))(w - w_i)^m$ for $m = 0, 1, 2,$ and 3 were tabulated into an array of dimensions $l \times 4 \times n_r \times n_w = l \times 4 \times 100 \times 100$, where l is the number of polynomial pieces used in the representation of the spectral density. In order to take advantage of this tabulation at the time of calculating the transform, we converted the splines $S(w)$ (linear combination of B-splines) into a

piecewise polynomial form

$$S(w) = \sum_{i=0}^{l} \sum_{j=0}^{3} a_{ij} \mathbb{I}([w_i, w_{i+1}))(w - w_i)^j. \tag{8.3}$$

Hence the Hankel transform was reduced to multiplying the tabulated values by the corresponding coefficients.

CHAPTER 9

SIMULATIONS

We have simulated Gaussian random fields with mean 0 and various covariance functions and estimated the spectral density using S+T family of functions. For comparison purposes, we have also estimated the Matérn parameters and the kernel estimators proposed by Hall et al. [1994]. The locations were chosen to be where the National Acid Deposition Program sites are situated. We used a total of 63 sites that are shown in Figure 9.1. The smallest distance between sites is 14 km, the largest distance is 2000 km, and the median distance is 802 km.

The models we used to simulate the data are Matérn, Bessel(1/2), S+T and a family here called polynomial Matérn, whose spectral density is the product of Matérn spectral density and a positive polynomial $(\frac{((w-u)^2+v^2)((w+u)^2+v^2)}{(a^2+w^2)^{\nu+1+2}})$. This function is positive on $\mathbb{R}^+$ so it is a valid isotropic spectral density. The spectral density of the Bessel(1/2) model is exponential in the spectral domain, i.e. the spectral density has the form $\exp(-w/w_t)$. Matérn, S+T and polynomial Matérn share the same high frequency behavior, namely, $1/\omega^\gamma$. The Bessel(1/2) model has a much faster decay and has analytic realizations of the process. We do not consider this type of behavior to be reasonable but it is included here to test the method.

In order to assess the performance of each method, we look at various statistics and compare them to the true values.

- Parameter values: When the true model and the model used to estimate the covariance functions have common parameters (for example σ^2 is common to

all models) the difference between the true and the estimated parameters is an obvious measure of performance.

- Likelihood values: The value of the likelihood also gives us an indication of how good the models are fitting the data. Although it is a bit unfair to compare methods that seek maximizing the likelihood with methods that seek to optimize other criteria, large deviations from the true likelihood should give us an idea of how good the estimated function is.

- RMSE: Comparing the distance between the true and estimated covariance function or spectral densities seems to be an obvious and esthetically pleasing way of assessing the performance of methods. However, if we are interested in estimating the covariance function for interpolation purposes, this method can be misleading. The following example from Stein [1999] illustrates this point. Suppose the true model is $\exp(-r)$. The function $\exp(-r^2/2)$ is closer in mean squared sense to the true covariance function than $\exp(-2r)/2$ is. However, interpolating with the latter function can give smaller prediction errors and dramatically better estimates of uncertainties than the squared exponential covariance function ($\exp(-r^2/2)$). We will see this effect in some of our simulation results. We calculate the RMSE in a finite range, $(0, r_{\max})$ for the covariance function and $(0, 1/\bar{r}_{\min})$ for the spectral density function, where $\bar{r}_{\min} = 50$km.

- Prediction error: In most applications, the ultimate goal of estimating the covariance functions is the prediction of the random field at unobserved locations. Thus it is most reasonable to use the mean squared error of the predictions as an indicator of goodness of fit. Following Stein [1999], we use $\frac{E_o e_1^2}{E_o e_o^2}$, where E_i indicates expectation using the i^{th} covariance function, e_i is the prediction error,

i.e., the difference between the true value of the random field and the predicted value using covariance function i. The true function is labeled o. It is easy to show that $E_o e_1^2 / E_o e_o^2 = 1 + E_o(\hat{Z}_1(x) - \hat{Z}_o(x))^2 / E_o e_o^2$. We estimate the numerator in the second term by taking the sample mean (out of 100 simulations) of the squared difference between the interpolated values with the misspecified covariance function ($\hat{Z}_i$) and the interpolated values with the true covariance function ($\hat{Z}_i$).

- Prediction variance error: A common approach to estimating the variance of the predictions is to calculate it by plugging in the estimated variance parameters ($E_1 e_1^2$). Therefore, we would like the ratio between $E_1 e_1^2$ and the actual prediction variance ($E_o e_1^2$) to be as close to 1 as possible.

Maximum Likelihood, mean = 0

Table 9.1 shows the average results of running 100 simulations for each model: Matérn, polynomial Matérn, S+T, and Bessel(1/2). Each simulation was 200 independent realizations of a Gaussian random field at 63 locations, totaling 12600 observations with the given covariance functions. The covariance matrix corresponding to this type of datasets is block diagonal, which allows us to have a large number of observations (so that the parameters can be estimated well) while keeping the computational load at a manageable level. For each simulation, the ML parameters for S+T and Matérn models were estimated. Also the kernel estimate of the covariance function was calculated. The first row of Table 9.1 shows l where $l + 1$ is the number of nodes of the S+T model. For each of the simulated models, we maximized the likelihood for a few different number of nodes and chose the one that yielded the smallest AIC value ($-2\text{loglik} - 2\#$ nodes). This was done using a few simulations. When running the 100

simulations, we fixed the number of nodes. We used 4 nodes for the Matérn model, 8 for the polynomial Matérn model and 5 for the S+T model, when the number of nodes used in the actual simulation was 5, and 4 for the Bessel(1/2) model.

The next three rows show the smoothness parameters, true and estimated, using S+T and Matérn models. When the true model is Matérn or polynomial Matérn the estimated smoothness is around 2.2, a bit smaller than the truth, which was 3. This is expected for the Matérn model since the rate of decay of the estimating tail function $1/w^{2\mu+2}$ is faster than the rate of decay of the true tail function $1/(a^2+w^2)^{\nu+1}$. We can see this more clearly by looking at the derivatives of the logs of the two tail functions:

$$-\frac{2\mu+2}{w} \approx -\frac{2\nu+2}{w} + \frac{2a^2(\nu+1)}{w(w^2+a^2)}, \tag{9.1}$$

which simplifies to

$$\mu \approx \nu - \frac{a^2(\nu+1)}{w^2+a^2}. \tag{9.2}$$

Hence the estimated power should be smaller in absolute value than the true power. The approximate sign is used since we will not be able to achieve equality for all w's. A similar argument works for the polynomial Matérn function.

When simulating with S+T model, there is no approximation of the tail so the method gives an estimated value that is very close to the truth. In the case of the Bessel(1/2) model we do not have a true parameter with which to compare. We notice that S+T and Matérn methods give similar estimates of the smoothness when the truth is Bessel(1/2).

The estimated values of the sill (σ^2) are very close to the true value of 1 for all three methods except for the Matérn method when the true model is S+T, which is 1.15 with standard deviation 0.03, and the Kernel method when the true is Bessel(1/2), which is 0.96 with standard deviation 0.01. The S+T method gives 1.00 with error

0.02 when the true model is Matérn, polynomial Matérn and S+T. When the true model is Bessel(1/2), we get 1.00 with standard deviation 0.01.

The next two blocks of rows show the cut frequency w_t and the inverse range. For the S+T model, the estimated value of w_t is 0.0100 with standard deviation 0.0008. The true parameter is 0.0094 which is within one standard deviation of the estimated value. For the Matérn model both the true and estimated inverse range parameters are 0.0094. The standard deviation of the estimated value is 0.0003.

The RMSE between the true and estimated covariance functions are shown next. The S+T method shows smaller values by factors of 3.9, 2.9, 2.9 and 1.9 compared to the Kernel method, when the true models are Matérn, polynomial Matérn, S+T, and Bessel(1/2), respectively. It also gives smaller RMSE compared to Matérn method. The RMSE were calculated by taking the difference between the covariance function values of 100 equispaced points in the range 0 to $r_{\max}$.

The RMSEs between the true and estimated spectral densities are shown next. The S+T method shows smaller values by factors of 3.8, 1.5, 1.7, and 1.8 compared to the Kernel method. It also gives smaller RMSE compared to Matérn method except when the true model is Matérn. The RMSE were calculated by taking the difference between the spectral density function values at 100 equispaced points between 0 and $1/\bar{r}_{\min}$, where $\bar{r}_{\min} = 50$km.

Table 9.2 shows the prediction performance of each method. The values of the mse are median values for 100 sites located on a lattice that covers the observed region. We use the interquartile range (IQR), the difference between the third and the first quartiles, as a measure of variability. The values presented in this table are the mses scaled by the error of the best linear predictor minus 1. The mse of the predictions using the S+T models are only around 0.1% larger than the best linear predictor

for all four simulated models. When compared to the Kernel method, our method is better by factors of 273, 35, 1509, and 2 when the true models are Matérn, polynomial Matérn, S+T and Bessel(1/2), respectively. When compared to the Matérn method, our method does worse by a factor of 12 when the truth is Matérn, but our method gives better result by factors of 16, 85, and 4 when the true models are polynomial Matérn, S+T, and Bessel(1/2), respectively. In absolute terms, the Kernel method gives prediction errors of the order of 3.3, 0.4, 75.4 and 0.2 percents higher than the best linear predictor while the Matérn method gives errors of the order of 0.1, 1.8, 4 and 0.4 percents higher than the best linear predictors.

The prediction variance errors are calculated as the median absolute difference between 1 and the ratio of estimated variance with the misspecified model and the sample variance (out of 100 simulations) for the 100 sites described in the previous paragraph. One could also look at the squared root of the median of $\log^2(\frac{E_i e_i^2}{E_o e_o^2})$ but there was not much difference from the approach taken here. Our method gives predicted variances that differ about half a percent from the actual, which is better than the other two methods except for Matérn when the truth is Matérn. The Kernel method predicted variances that differed from the actual variance by 48%, 23%, 143% and 5% for true models Matérn, polynomial Matérn, S+T, and Bessel(1/2). The Matérn method's estimated variances were 0.3%, 0.9%, 8.3%, and 1.2% when the true models were Matérn, polynomial Matérn, S+T, and Bessel(1/2), respectively.

Tables 9.3 and 9.4 show the results of simulations with smaller smoothness parameter ($\nu = 0.50$). The estimated smoothness parameters are 0.43(0.04), 0.46(0.08), and 0.50(0.01). There is also a small underestimation although not as large as when the true parameter was 3.00. The Kernel method slightly underestimates the sill (σ^2) when the true models are Matérn, 0.96(0.01), and polynomial Matérn, 0.94(0.01). The

RMSEs of the covariance functions estimated using S+T model are 0.9, 2.5, and 7.3 times the RMSEs estimated using Matérn model when the true models are Matérn, polynomial Matérn, and S+T. When compared to the Kernel method the improvement factors are 5.8, 9.6, and 3.0. The ratios of the RMSEs of the spectral densities are 0.6, 1.8, and 6.5 when compared to Matérn method and 3.9, 4.6, and 1.7 when compared to the Kernel method. The mses of the predictions are also around 0.1% for our method. The factors by which our method outperforms the Kernel method is not as dramatic as with the smoother process in Table 9.1 but are still quite large: 7.6, 11.4, and 18.8 for true models Matérn, polynomial Matérn, and S+T. Relative to the Matérn method our method has smaller prediction errors for the polynomial Matérn and S+T models by factors of 3.5, and 15.5. When the true model is Matérn, using the Matérn model to estimate the parameters gives better predictions by a factor of 3 compared to our method, as expected. Again, the improvement is less dramatic than with a smoother process. In absolute terms, the Matérn method has prediction errors that range between 0.02% and 0.6%. The Kernel method has prediction errors ranging from 0.46% to 1.82%. So for practical purposes the prediction errors are quite small for all three methods unlike with the smoother process where the Kernel method errors could be quite high: about 75% or even higher depending of the particular location. The errors in the variance of predictions are of the order of 0.3% for our method, much lower than for the Kernel method, which has errors ranging from 3.4% to 9.8%. The Matérn method has errors ranging from 0.14% (when the true model is Matérn) to 1.86% when the true model is S+T.

Restricted Likelihood, estimated mean

Table 9.5 and 9.6 show the results of simulations when the mean is not known, so REML was used. The results are quite similar to the case when the mean is considered to be known (Tables 9.1 and 9.2). The RMSEs of the covariance functions and spectral densities are slightly larger than the known mean case. The improvement relative to the Kernel method is also a bit smaller but still significant, more than a factor of 2 for the RMSE of the covariance function and factors ranging from 1.1 to 2.4 for the RMSE of the spectral density. The prediction errors are around 0.2%, twice as much as the known mean case but still very small. The factors by which our method outperforms the Kernel method are also similar to the known mean case, 175, 30, 1437, and 2 when the true models are Matérn, polynomial Matérn, S+T, and Bessel(1/2). The errors in the estimated variance of the predictors are slightly higher when estimating the mean. The factors by which our method outperforms the other methods are all close to the known mean case.

		Matérn	polMatérn	S+T	Bessel(1/2)
	true			4	
l	S+T	3	7	4	3
	true	3.00	3.00	3.00	–
ν	S+T	2.19 (0.09)	2.21 (0.19)	3.01 (0.05)	1.56 (0.29)
	Mat	2.98 (0.09)	10.00 (2.39)	7.47 (0.31)	1.63 (0.18)
	true	1.00	1.00	1.00	1.00
	S+T	1.00 (0.02)	1.00 (0.02)	1.00 (0.02)	1.00 (0.01)
σ^2	Mat	1.00 (0.02)	1.00 (0.02)	1.15 (0.03)	1.00 (0.01)
	Ker	0.99 (0.02)	0.98 (0.02)	0.99 (0.02)	0.96 (0.01)
	true	–	–	0.0094	–
w_t	S+T	0.0129 (0.0010)	0.0187 (0.0018)	0.0100 (0.0008)	0.0318 (0.0033)
		0.0094			
inv.range	Mat	0.0094 (0.0003)	0.0354 (0.0046)	0.0153 (0.0005)	0.0410 (0.1957)
	S+T	0.0053 (0.0034)	0.0075 (0.0038)	0.0074 (0.0027)	0.0058 (0.0013)
RMSE	Mat	0.0074 (0.0050)	0.0536 (0.0003)	0.1239 (0.0078)	0.0104 (0.0160)
Cov	Ker	0.0205 (0.0066)	0.0217 (0.0047)	0.0208 (0.0060)	0.0112 (0.0015)
	S+T	165 (122)	301 (222)	236 (90)	61 (14)
RMSE	Mat	74 (55)	1305 (367)	2688 (132)	100 (46)
Spect.	Ker	634 (344)	466 (277)	411 (163)	110 (51)
	S+T	0	0	0	0
loglik	Mat	7 (6)	-179 (21)	-292 (24)	-7 (5)
	Ker	-1154 (189)	-260 (35)	-3197 (275)	-28 (9)

Table 9.1: Summary of simulation results for $\nu = 3.00$, $\sigma^2 = 1.00$, and inverse range(or w_t) $= 0.0094$. Maximum likelihood estimation. Average estimates from 100 simulations are shown. Each simulation consisted of 12600 observations (200 replications of 63 spatially correlated observations) Standard deviations are shown in parenthesis. Columns correspond to true models and rows correspond to estimating methods: ML using S+T, ML using Matérn, and kernel.

		Matérn	polMatérn	S+T	Bessel(1/2)
$E_o e_1^2/E_o e_o^2 - 1$	S+T	0.0012 (0.0007)	0.0011 (0.0005)	0.0005 (0.0007)	0.0010 (0.0006)
$E_o e_2^2/E_o e_o^2 - 1$	Mat	0.0001 (0.0001)	0.0177 (0.0316)	0.0424 (0.0317)	0.0040 (0.0063)
$E_o e_3^2/E_o e_o^2 - 1$	Ker	0.3280 (0.7190)	0.0380 (0.0860)	0.7543 (1.5019)	0.0023 (0.0102)
$\lvert E_1 e_1^2/E_o e_1^2 - 1\rvert$	S+T	0.0057 (0.0184)	0.0033 (0.0054)	0.0021 (0.0026)	0.0043 (0.0147)
$\lvert E_2 e_2^2/E_o e_2^2 - 1\rvert$	Mat	0.0030 (0.0062)	0.0085 (0.0199)	0.0829 (0.1373)	0.0121 (0.0177)
$\lvert E_3 e_3^2/E_o e_3^2 - 1\rvert$	Ker	0.4811 (0.6812)	0.2310 (0.4317)	1.4337 (2.7026)	0.0543 (0.1417)

Table 9.2: Simulations as in Table 9.1. $\frac{E_o e_i^2}{E_o e_o^2}$ represents the prediction error using the misspecified covariance scaled by the best linear predictor's error variance. We show the median over 100 prediction locations of $\frac{E_o e_i^2}{E_o e_o^2} - 1$. $\frac{E_i e_i^2}{E_o e_i^2}$ represents the prediction error variance calculated with the misspecified covariance scaled by the actual prediction error variance. We show the median over 100 prediction locations of the absolute value of $\frac{E_i e_i^2}{E_o e_i^2} - 1$.

		Matérn	polMatérn	S+T
	true	–	–	4
l	S+T	3	7	4
	true	0.50	0.50	0.50
ν	S+T	0.43 (0.04)	0.46 (0.08)	0.50 (0.01)
	Mat	0.49 (0.03)	0.78 (0.12)	0.67 (0.02)
	true	1.00	1.00	1.00
	S+T	1.00 (0.01)	1.00 (0.01)	1.00 (0.02)
σ^2	Mat	1.00 (0.01)	1.00 (0.01)	1.01 (0.02)
	Ker	0.96 (0.01)	0.94 (0.01)	0.98 (0.02)
	true	–	–	0.0094
w_t	S+T	0.0189 (0.0033)	0.0356 (0.0088)	0.0088 (0.0013)
	true	0.0094	0.0094	–
inv.range	Mat	0.0094 (0.0006)	0.0347 (0.0039)	0.0068 (0.0002)
	S+T	0.0046 (0.0019)	0.0060 (0.0022)	0.0054 (0.0019)
RMSE	Mat	0.0042 (0.0027)	0.0148 (0.0020)	0.0394 (0.0025)
Cov	Ker	0.0267 (0.0025)	0.0574 (0.0022)	0.0160 (0.0018)
	S+T	44 (34)	28 (8)	186 (79)
RMSE	Mat	28 (21)	52 (7)	1213 (116)
Spect	Ker	173 (75)	132 (25)	306 (149)
	S+T	0	0	0
loglik	Mat	-1 (2)	-25 (7)	-79 (12)
	Ker	-388 (54)	-606 (70)	-261 (45)

Table 9.3: Summary of simulation results for $\nu = 0.50$, $\sigma^2 = 1.00$, and inverse range(or w_t) $= 0.0094$. Average estimates from 100 simulations are shown. Each simulation consisted of 12600 observations (63 spatially correlated and 200 independent replicates) Standard deviations are shown in parenthesis. Columns correspond to true models and rows correspond to estimating methods: ML using S+T, ML using Matérn, and kernel.

		Matérn	polMatérn	S+T
$E_o e_1^2 / E_o e_o^2 - 1$	S+T	0.0006 (0.0003)	0.0016 (0.0011)	0.0004 (0.0003)
$E_o e_2^2 / E_o e_o^2 - 1$	Mat	0.0002 (0.0001)	0.0056 (0.0056)	0.0062 (0.0100)
$E_o e_3^2 / E_o e_o^2 - 1$	Ker	0.0046 (0.0434)	0.0182 (0.1137)	0.0075 (0.0219)
$E_1 e_1^2 / E_o e_1^2 - 1$	S+T	0.0025 (0.0025)	0.0032 (0.0036)	0.0022 (0.0021)
$E_2 e_2^2 / E_o e_2^2 - 1$	Mat	0.0014 (0.0032)	0.0016 (0.0131)	0.0186 (0.0228)
$E_3 e_3^2 / E_o e_3^2 - 1$	Ker	0.0774 (0.2885)	0.0978 (0.3239)	0.0347 (0.2189)

Table 9.4: Simulations as in Table 9.3. $\frac{E_o e_i^2}{E_o e_o^2}$ represents the prediction error using the misspecified covariance scaled by the best linear predictor's error variance. We show the median over 100 prediction locations of $\frac{E_o e_i^2}{E_o e_o^2} - 1$. $\frac{E_i e_i^2}{E_o e_i^2}$ represents the prediction error variance calculated with the misspecified covariance scaled by the actual prediction error variance. We show the median over 100 prediction locations of the absolute value of $\frac{E_i e_i^2}{E_o e_i^2} - 1$.

		Matérn	polMatérn	S+T	Bessel(1/2)
	true	–	–	4	–
l	S+T	3	7	4	3
	true	3.00	3.00	3.00	
ν	S+T	2.09 (0.10)	2.19 (0.22)	3.01 (0.06)	1.46 (0.22
	Mat	2.98 (0.09)	22.18 (13.34)	6.93 (0.27)	1.70 (0.18
	true	1.00	1.00	1.00	1.00
	S+T	0.99 (0.03)	0.99 (0.02)	0.99 (0.02)	1.00 (0.01)
σ^2	Mat	1.00 (0.02)	0.95 (0.02)	1.22 (0.03)	0.99 (0.01)
	Ker	0.99 (0.02)	0.98 (0.02)	0.99 (0.02)	0.96 (0.01)
	true	–	–	0.0094	–
w_t	S+T	0.0116 (0.0011)	0.0188 (0.0023)	0.0103 (0.0010)	0.0303 (0.0030)
	true	0.0094	–	–	–
inv.range	Mat	0.0094 (0.0003)	0.0541 (0.0141)	0.0144 (0.0004)	0.0221 (0.0016)
	S+T	0.0079 (0.0054)	0.0084 (0.0049)	0.0093 (0.0030)	0.0057 (0.0014)
RMSE	Mat	0.0078 (0.0056)	0.0560 (0.0012)	0.1479 (0.0095)	0.0094 (0.0017)
Cov	Ker	0.0205 (0.0066)	0.0215 (0.0046)	0.0208 (0.0060)	0.0112 (0.0015)
	S+T	242 (210)	220 (161)	355 (103)	62 (14)
RMSE	Mat	74 (57)	1525 (118)	2803 (154)	98 (15)
Spect	Ker	570 (309)	313 (178)	406 (167)	113 (53)
	S+T	0	0	0	0 (75)
likeli	Mat	9 (7)	-118 (14)	-233 (21)	-5 (4)
-true	Ker	-1028 (189)	-109 (37)	-2943 (265)	224 (14)

Table 9.5: Summary of simulation results for $\nu = 3.00$, $\sigma^2 = 1.00$, and inverse range(or w_t) = 0.0094. Restricted Maximum Likelihood. Average estimates from 100 simulations are shown. Each simulation consisted of 12600 observations (63 spatially correlated and 200 independent replicates) Standard deviations are shown in parenthesis. Columns correspond to true models and rows correspond to estimating methods: REML using S+T, REML using Matérn, and kernel.

		Matérn	polMatérn	S+T	Bessel(1/2)
$E_o e_1^2 / E_o e_o^2 - 1$	S+T	0.0020 (0.0013)	0.0013 (0.0005)	0.0005 (0.0008)	0.0009 (0.0006)
$E_o e_2^2 / E_o e_o^2 - 1$	Mat	0.0001 (0.0001)	0.0171 (0.0215)	0.0405 (0.0305)	0.0014 (0.0008)
$E_o e_3^2 / E_o e_o^2 - 1$	Ker	0.3489 (0.7377)	0.0385 (0.0860)	0.7535 (1.5340)	0.0022 (0.0103)
$E_1 e_1^2 / E_o e_1^2 - 1$	S+T	0.0110 (0.0276)	0.0044 (0.0040)	0.0024 (0.0028)	0.0041 (0.0133)
$E_2 e_2^2 / E_o e_2^2 - 1$	Mat	0.0032 (0.0062)	0.0145 (0.0280)	0.0934 (0.1529)	0.0114 (0.0158)
$E_3 e_3^2 / E_o e_3^2 - 1$	Ker	0.4897 (0.6749)	0.2322 (0.4321)	1.4505 (2.5591)	0.0542 (0.1430)

Table 9.6: Simulations as in Table 9.5. $\frac{E_o e_i^2}{E_o e_o^2}$ represents the prediction error using the misspecified covariance scaled by the best linear predictor's error variance. We show the median over 100 prediction locations of $\frac{E_o e_i^2}{E_o e_o^2} - 1$. $\frac{E_i e_i^2}{E_o e_i^2}$ represents the prediction error variance calculated with the misspecified covariance scaled by the actual prediction error variance. We show the median over 100 prediction locations of the absolute value of $\frac{E_i e_i^2}{E_o e_i^2} - 1$.

Figures 9.2-9.5 show the covariance function (left) and the spectral density (right) for the Matérn, polynomial Matérn, S+T models, and Bessel(1/2) models. These figures show one typical simulation from each model. As suggested by the RMSE values from Table 9.1, the MLE covariance functions (dots) using our model are much closer to the true covariance functions (solid line) than the kernel estimators (cross) of the covariance functions. The kernel estimator of the covariance function becomes highly wiggly for large distances, mainly because there are fewer pairs of observations that contribute to this region. In order to calculate the spectral density we followed the ad hoc solution proposed by Hall et al. [1994]. From some point T_1 use a straight line that goes from the value of the estimator at T_1 to zero at some other distance T_2. We chose $T_1 = 1500$km and $T_2 = 2000$km. Figure 9.2 (left), which corresponds to Matérn class, shows that the MLE (dots) using our model very closely fits the true spectral function. The spectral function (right) estimated with the kernel method is wiggly, although it follows to some extent the true spectral density. Figure 9.3 corresponds to the polynomial Matérn class. Both estimates have similar shapes and, except near the origin, they are very close to the truth. The right plots of Figures 9.4 and 9.5 also show that both estimates are close to the true spectral density except near the origin. In all cases, ML estimates using our model are closer to the true function.

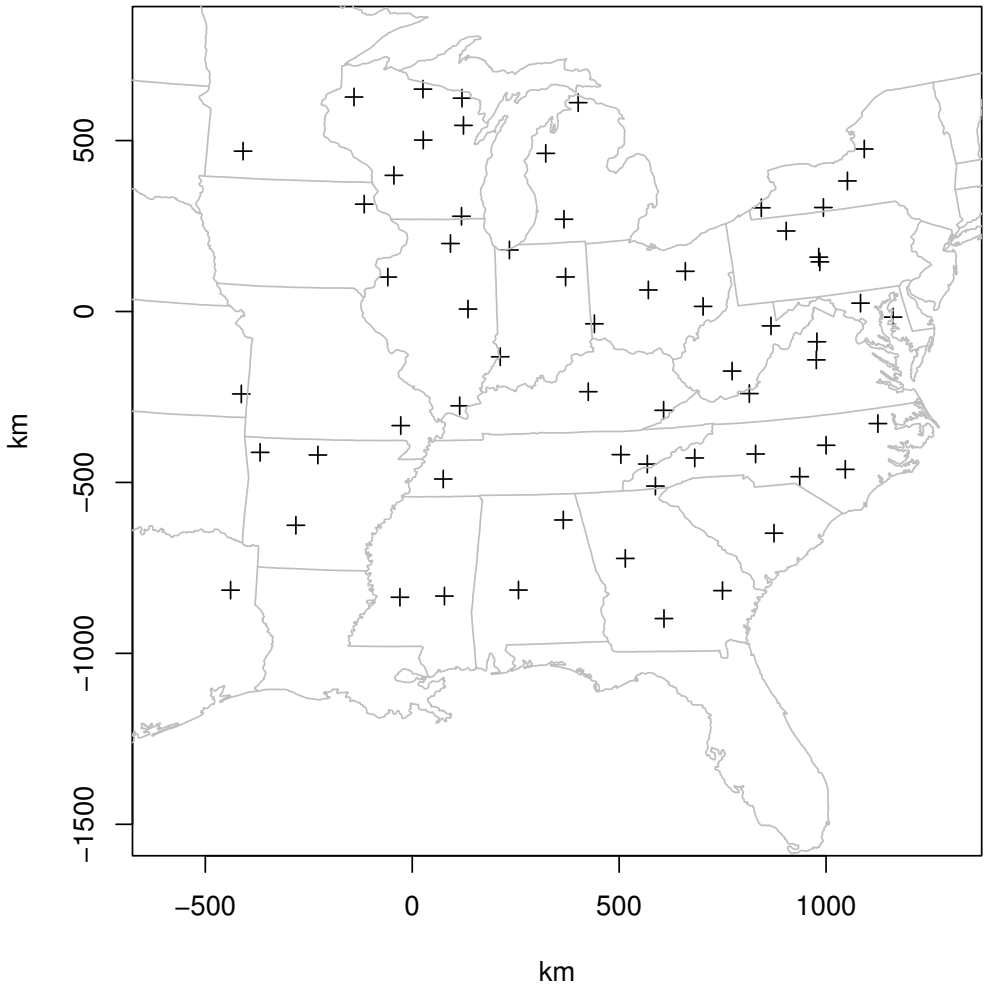

Figure 9.1: NADP monitoring sites used for simulations

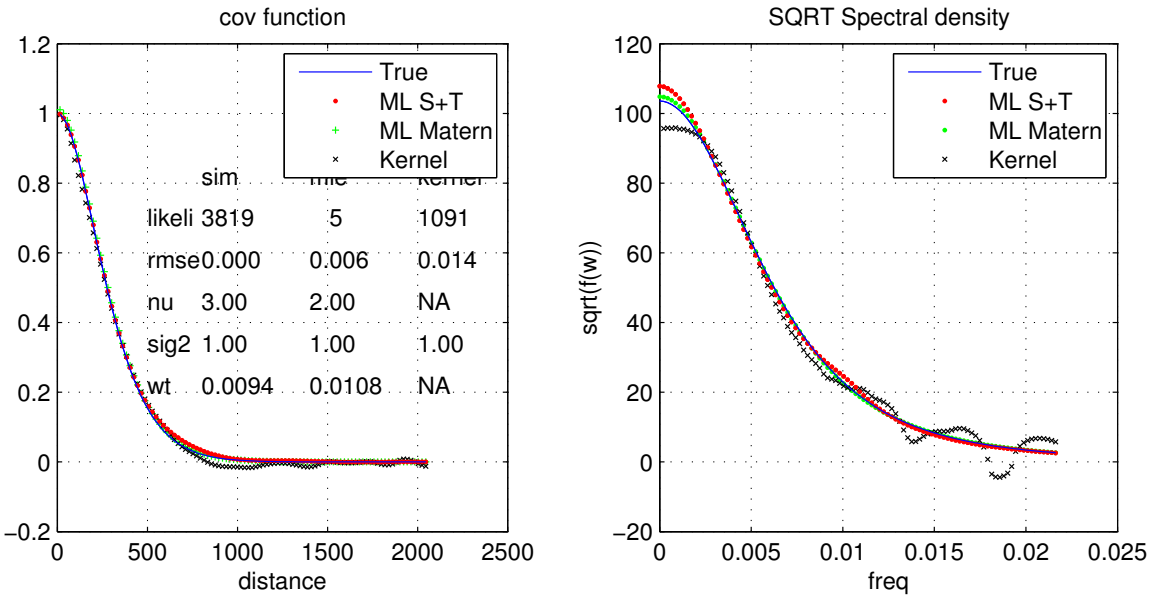

Figure 9.2: True and estimated a) covariance function and b) spectral density. The true model is Matérn with $\nu = 3$, $\sigma^2 = 1.00$, and inverse range $= 0.0094$

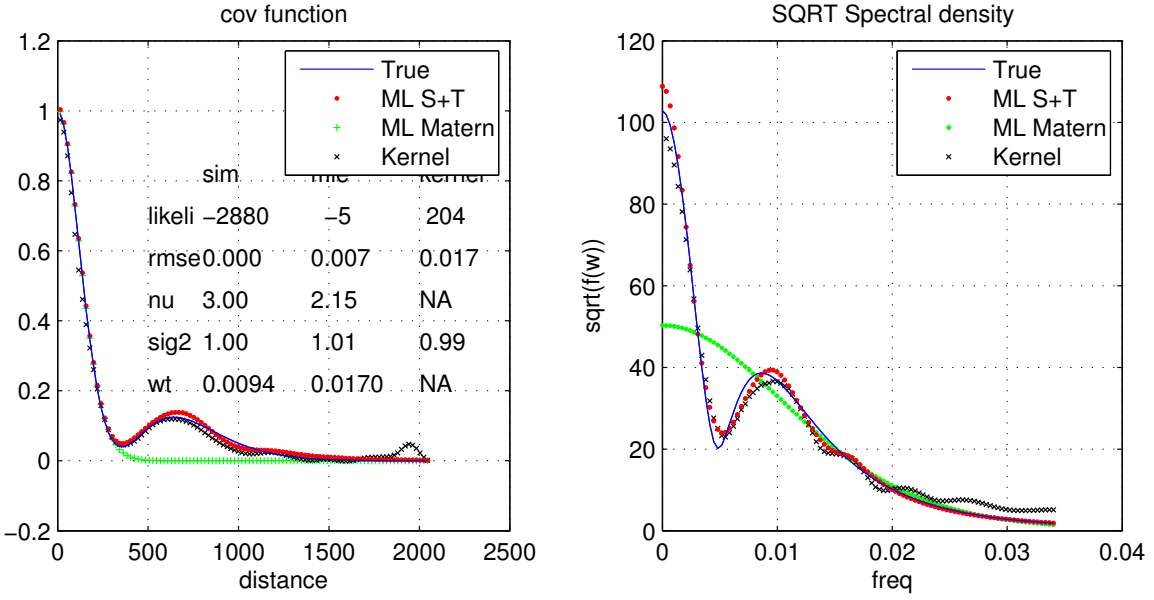

Figure 9.3: True and estimated a) covariance function and b) spectral density. The true model is polynomial Matérn with $\nu = 3$, $\sigma^2 = 1.00$, inverse range (or w_T) $= 0.0094$, $u = 0.0047$, and $v = 0.000009$.

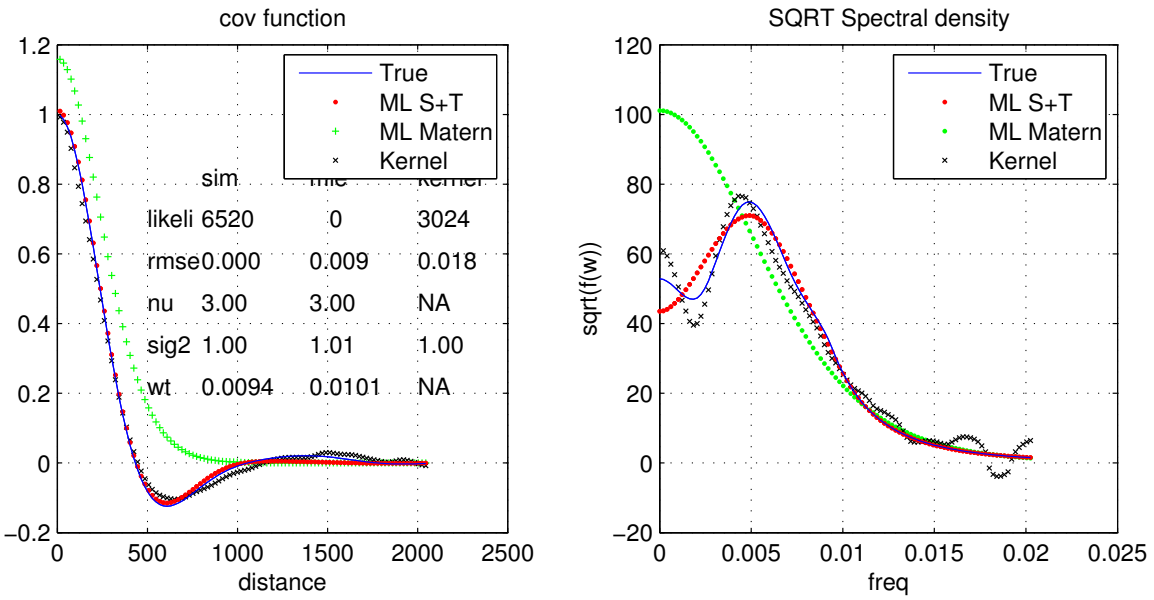

Figure 9.4: True and estimated a) covariance function and b) spectral density. The true model is S+T with $\nu = 3$, $\sigma^2 = 1.00$, $w_t = 0.0094$, and coefficients $\mathbf{b} = (1, 0.2, 2, 0.6, 0.4)$

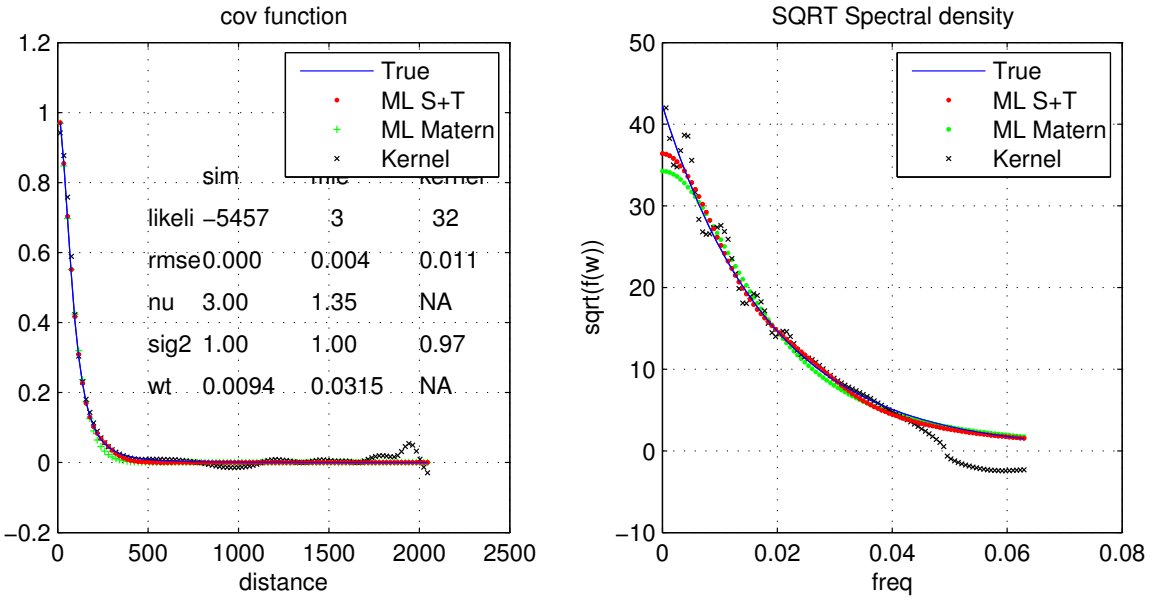

Figure 9.5: True and estimated a) covariance function and b) spectral density. The true model is Bessel(1/2) with $\sigma^2 = 1.00$ and inverse range $= 0.0094$

CHAPTER 10

SUMMARY AND DISCUSSION

We propose a flexible class of spectral density functions whose parameters can be estimated using ML or REML methods. We have calculated explicit expressions of the Hankel transform of the spectral density and tackled several numerical issues arising during the computation of the covariance function. We applied simulated annealing method to maximize both the likelihood (when the mean is known) and restricted likelihood.

We simulated observations with Matérn, polynomial Matérn, S+T, and Bessel(1/2) spectral densities. Each model was simulated 100 times and average estimated values and mses were presented. We showed that our method (S+T) outperforms the Kernel method in terms of estimated sill, RMSEs of the covariance functions, RMSE of the spectral densities, the likelihood values (this is expected since we are maximizing the likelihood but the large differences seen may be indicative of poor performance), mse of the predictions and errors in the estimated variances of the predictions. Our method also outperforms the Matérn method (parametric using Matérn as estimating model) when the true model is not Matérn by all performance criteria just listed.

The mse of predictions and the errors of the estimated variances of the predictions are the most relevant measures of performance when our ultimate goal is interpolation to locations where there are no observations. With this criteria, the Matérn method outperforms the Kernel method although it has larger RMSE values of the covariance function and the spectral density then the Kernel method. The reason for the better

prediction properties of the Matérn method is that the tail properties of the spectral function play a fundamental role in the prediction. Our method directly estimates the tail property, just like Matérn method does but it also offers more flexibility for the lower frequencies, which improves the predictions. Our method outperforms the kernel method even when the true model is Bessel(1/2), which has an exponential tail whereas our method assumes an algebraic tail.

We have performed simulations with smaller number of replicates (20 and 1 instead of 200) per simulation. Using the prediction performance criteria, we found that our method outperforms the Matérn method when the number of replicates is 20 but not so when we only have 1 replicate of the spatial process. So our method should be applied when we have a large amount of data. We did not compare the performance properties of the Kernel method because the estimated covariance functions were not positive definite. In future work, we will truncate the negative parts of the spectral densities estimated by the Kernel method and transform back to the spatial domain. We need to develop efficient methods to perform this transformation. Then we should be able to compare the prediction performance of this method to ours. Since the convergence rate of the truncated estimates are smaller than the untruncated estimates according to Hall et al. [1994], we are quite confident that our method will do a better job at predictions.

In the future, we will allow the spacing of the nodes to be estimated from the data. One possible way of doing this in an efficient way is by utilizing the information contained in the Kernel estimator. The Kernel method fails to capture the tail behavior of the spectral density but it does seem to have useful information about the mid frequency shape of the function.

APPENDIX A

B-SPLINES

The following is a summary of the properties of B-splines relevant for this work. For a more extensive description see de Boor [2001].

A piecewise polynomial of order k with strictly increasing break (node) sequence ξ is a function of the form

$$\sum_j \mathbb{I}[\xi_j, \xi_{j+1}) p^{(k)}(x) \tag{A.1}$$

where $p^{(k)}(x)$ is a polynomial of degree $k-1$ or smaller. The set of all piecewise polynomial functions of order k with break sequence ξ is denoted $\Pi_{<k,\xi}$.

B-splines are defined in terms of a non-decreasing knot sequence $\mathbf{t} = (t_j)$. The j^{th} B-spline of order 1 for knot sequence $\mathbf{t}$ is the indicator function of the interval $[t_j, t_{j+1})$:

$$B_{j,1} := B_{j,1,\mathbf{t}} = \mathbb{I}([t_j, t_{j+1})). \tag{A.2}$$

If $t_j = t_{j+1}$, $B_j = 0$. The j^{th} B-spline of order $k > 1$ is defined by the following recurrence relation

$$B_{jk} := B_{j,k,\mathbf{t}} := w_{jk} B_{j,k-1} + (1 - w_{j+1,k}) B_{j+1,k-1} \tag{A.3}$$

with $w_{jk} := w_{j,k,\mathbf{t}} := \frac{x - t_j}{t_{j+1} - t_j}$. B_{jk} is a piecewise polynomial function with break sequence $t_k, ..., t_{j+k}$. It is positive on $[t_j, t_{j+k}]$ and 0 outside this interval. B-splines of order k with knot sequence $\mathbf{t}$ span the space of piecewise polynomial functions

of order k with break sequence ξ and continuity conditions on the breaks given by the multiplicity of the knots. More specifically, the sum of the number of continuity conditions at a break ξ_j and the number of repeated knots at ξ_j equals the order k.

For the uniform knot sequence $t = (..., -\Delta, 0, \Delta, 2\Delta, ...)$ the corresponding B-splines are

$$B_{j,k,\mathbf{t}}(x) = \sum_{r=0}^{k} \frac{(-1)^{k-r}}{(k-1)!} \binom{k}{r} (r - x/\Delta + j)_+^{k-1} \tag{A.4}$$

In particular, for $k = 4$, $j = 0$, and $\mathbf{t} = \mathbb{Z}$ and $\Delta = 1$

$$B^1_{0,4,\mathbb{Z}}(x) = \begin{cases} x^3/6 & \text{if } 0 \le x < 1; \\ (-3x^3 + 12x^2 - 12x + 4)/6 & \text{if } 1 \le x < 2; \\ (3x^3 - 24x^2 + 60x - 44)/6 & \text{if } 2 \le x < 3; \\ (-x^3 + 12x^2 - 48x + 64)/6 & \text{if } 3 \le x < 4; \end{cases} \tag{A.5}$$

For arbitrary Δ and $j \neq 0$ the B-splines are obtained from (A.5) by translating the argument by $j\Delta$ and scaling it by $1/\Delta$, i.e., $B_{j,4,\mathbb{Z}}(x) = B^1_{0,4,\mathbb{Z}}(\frac{x - j\Delta}{\Delta})$.

APPENDIX B

HANKEL TRANSFORM OF POLYNOMIALS

The Hankel Transforms of piecewise polynomials of the form $(w-c)^m$ are given by

$$\int_{a'}^{b'} (w-k')^m w J_o(wr)dw = r^{-m-2} \int_{a'r}^{b'r} u(u-k'r)^m J_o(u)du, \tag{B.1}$$

which for $a = a'r$, $b = b'r$, $k = k'r$, and $m = 0$, 1 ,2, and 3 are

$$\begin{aligned}
\int_a^b uJ_o(u)du &= -aJ_1(a) + bJ_1(b) \\
\int_a^b (u-k)uJ_o(u)du &= J_1(a)(-a^2 + a(k + (\pi H_o(a))/2)) \\
&\quad + J_1(b)(b^2 + b(-k - (\pi H_o(b))/2)) \\
&\quad - (a\pi J_o(a)H_1(a))/2 + (b\pi J_o(b)H_1(b))/2 \\
\int_a^b (u-k)^2 u^2 J_o(u)du &= J_1(a)(-a^3 + 2a^2k + a(4 - k^2 - k\pi H_o(a))) \\
&\quad + J_1(b)(b^3 - 2b^2k + b(-4 + k^2 + k\pi H_o(b))) \\
&\quad + J_o(a)(-2a^2 + ak\pi H_1(a)) + J_o(b)(2b^2 - bk\pi H_1(b)) \\
\int_a^b (u-k)^3 uJ_o(u)du &= J_1(a)(-a^4 + 3a^3k + a^2(9 - 3k^2) \\
&\quad + a(-12k + k^3 + (-9/2 + (3k^2)/2)\pi H_o(a))) \\
&\quad + J_1(b)(b^4 - 3b^3k + b^2(-9 + 3k^2) \\
&\quad + b(12k - k^3 + (9/2 - (3k^2)/2)\pi H_o(b))) \\
&\quad + J_o(a)(-3a^3 + 6a^2k + a(9/2 - (3k^2)/2)\pi H_1(a)) \\
&\quad + J_o(b)(3b^3 - 6b^2k + b(-9/2 + (3k^2)/2)\pi H_1(b))
\end{aligned}$$

where $J_\nu(\cdot)$ are Bessel functions of the first kind of order ν and $H_\nu(\cdot)$ are Struve functions of order ν [Abramowitz and Stegun, 1965]

APPENDIX C

CONTINUITY OF DERIVATIVE

For uniform knots with spacing Δ the values of the spectral density on (w_{n-1}, w_n) where $w_n = \Delta n = w_t$ is

$$\begin{aligned} f(w) =& f_n B_{n-2}(\frac{w - w_{n-2}}{\Delta}) + f_{n-1} B_{n-3}(\frac{w - w_{n-3}}{\Delta}) \\ &+ f_{n-2} B_{n-4}(\frac{w - w_{n-4}}{\Delta}) + f_{n+1} B_{n+1}(\frac{w - w_{n+1}}{\Delta}) \end{aligned} \tag{C.1}$$

so

$$\begin{aligned} f'(w) =& f_n B'_{n-2}(\frac{w - w_{n-2}}{\Delta}) + f_{n-1} B'_{n-3}(\frac{w - w_{n-3}}{\Delta}) \\ &+ f_{n-2} B'_{n-4}(\frac{w - w_{n-4}}{\Delta}) + f_{n+1} B'_{n+1}(\frac{w - w_{n+1}}{\Delta}). \end{aligned} \tag{C.2}$$

At $w = w_t$

$$f'(w_t) = -\frac{n f_{n-1}}{2 w_n} + \frac{n f_{n+1}}{2 w_t}. \tag{C.3}$$

If we want $f'(w)$ to be continuous at w_t we need

$$f'(w_t) = -\frac{-\gamma f_t}{w_t} \tag{C.4}$$

where $f_t = \frac{1}{6} f_{n-1} + \frac{2}{3} f_n + \frac{1}{6} f_{n+1}$. Equating (C.3) and (C.4) we get

$$f_{n+1} = \frac{3n - \gamma}{3n + \gamma} f_{n-1} - \frac{4\gamma}{3n + \gamma} f_n. \tag{C.5}$$

Positivity of the spectral density

The values of the $(n+1)^{\text{th}}$ coefficient can be negative so we need to check whether the spectral density is still positive on the interval $(0, w_t)$. Because of the linearity of the derivative it is enough to consider separately the cases where $f_n = 0$ and $f_{n-1} = 0$. If we get positive spectral density in each case the sum of the two will also result in a positive function. Also, since the support of the B-spline with coefficient f_{n+1} (B_{n-1}) is (w_{n-1}, w_{n+3}) we only need to worry about the interval (w_{n-1}, w_n).

Case $f_{n-1} = 0$

Each of the terms in (C.1) is positive except possibly for the term corresponding to f_{n+1}, so

$$f(w) \geq f_n B_{n-2}(2) + f_{n+1} B_{B_{n-1}} := g(w). \tag{C.6}$$

Using the corresponding piecewise polynomial in each case and setting $x = \frac{w - w_{n-2}}{\Delta}$ we get

$$\begin{aligned} g(x) &= \frac{f_n}{6}\left(-3x^3 + 12x^2 - 12x + 4 - \frac{4\gamma}{3n+\gamma}(x-1)^3\right) \\ &= \frac{f_n}{6}\left(\frac{2\gamma+3n}{2(3n+\gamma)}(2-x)(4-10x+7x^2) + \frac{3n}{2(3n+\gamma)}x^3\right) \end{aligned} \tag{C.7}$$

For $w \in (w_{n-1}, w_n)$, $x \in (1, 2)$ so one only needs to check that the quadratic polynomial in the first term is positive in the interval of interest. The roots of this polynomial are not real ($\frac{1}{7}(5 \pm 3\imath)$) so if it is positive for some x it will be positive for all $x \in \mathbb{R}$. In fact, its value for $x = 0$ is 4.

Case $f_n = 0$

We have

$$f(w) \geq f_{n-1}B_{n-3}(x) + f_{n+1}B_{n-1}(x) := g(w) \tag{C.8}$$

with $x = \frac{w-w_{n-3}}{\Delta}$ so that $x \in (2,3)$ if $w \in (w_{n-1}, w_n)$. I will show that the function $g(w)$ is decreasing and that the value at $x = 3$ is positive.

$$g(x) = \frac{f_n}{6}\left(3x^3 - 24x^2 + 60x - 44 + \frac{3n-\gamma}{gn+\gamma}(x-2)^3\right) \tag{C.9}$$

$g(3) = f_n\frac{6n}{3n+\gamma} > 0.$

$$g'(x) = 6f_n\frac{(x-2)(\gamma(x-4)+6n(x-3))}{3n+\gamma} \leq 0 \tag{C.10}$$

for $x \in (2,3)$. Thus, g is non-increasing and $g(3) \geq 0$ so $g(x) > 0$ for $x \in (2,3)$

APPENDIX D

TAIL INTEGRAL WITH INTEGER SMOOTHNESS

Using the series representation of the hypergeometric function, the integral of the tail given in (7.7) can be written as

$$\int_{s_t}^{\infty} u^{1-\gamma} J_o(u) du = \frac{(-\gamma/2)\Gamma(-\gamma/2)}{2^{\gamma-1}\Gamma(-\gamma/2)} + \frac{s_t^{2-\gamma}}{\gamma-2} \sum_{k=0}^{\infty} \frac{-\gamma/2+1}{-\gamma/2+1+k} \frac{(-st^2/4)^k}{k!^2} \quad \text{(D.1)}$$

Let $-\gamma/2+1 = -n+\delta$ with $n \in \mathbb{N}$. As δ goes to zero the first term in (D.1) and the n^{th} term in the second term of (D.1) go to infinity but the total contribution of the diverging terms is finite. This can be shown by using the asymptotic expansion of the gamma function when the argument is close to a negative integer. We reorder the terms in (D.1) to make explicit the two terms that diverge when $\gamma/2$ is integer valued:

$$\begin{aligned} \int_{s_t}^{\infty} u^{1-\gamma} J_o(u) du = & \frac{\Gamma(-n+\delta)}{2^{\gamma-1}\Gamma(n+1-\delta)} + \frac{s_t^{2-\gamma}}{\gamma-2} \frac{-n+\delta}{\delta} \frac{(-st^2/4)^n}{n!^2} + \\ & + \frac{s_t^{2-\gamma}}{\gamma-2} \sum_{k=0, k\neq n}^{\infty} \frac{-\gamma/2+1}{-\gamma/2+1+k} \frac{(-st^2/4)^k}{k!^2}. \end{aligned} \quad \text{(D.2)}$$

Using $\Gamma(-n+\delta) = \frac{(-1)^n}{n!\delta} + \frac{(-1)^n \psi(n+1)}{n!} + O(\delta)$ [Wolfram Research,Inc., 2001c], where $\psi(n+1)$ is the digamma function [Wolfram Research,Inc., 2001b], we get

$$\int_{s_t}^{\infty} u^{1-\gamma} J_o(u) du = -\frac{2^{-1-2\,n}\, {s_t}^{2\,(\delta-n)} \left(-{s_t}^2\right)^n}{\delta\, n!^2} - \frac{(-1)^n\, 2^{-1+2\,\delta-2\,n}}{\delta\, n!\, \Gamma(1-\delta+n)} + \frac{(-1)^n\, 2^{-1+2\,\delta-2\,n}\, \psi(0, 1+n)}{n!\, \Gamma(1-\delta+n)} + \sum_{k=0, k\neq n}^{\infty} ... + O(\delta) \tag{D.3}$$

using Taylor expansion in δ and letting δ go to zero we get

$$\frac{\log(2) - \log(s_t) + \psi(n+1)}{(-4)^n n!^2} + \frac{s_t^{2-\gamma}}{\gamma - 2} \sum_{k=0, k\neq n}^{\infty} \frac{-\gamma/2 + 1}{-\gamma/2 + 1 + k} \frac{(-s_t^2/4)^k}{k!^2} \tag{D.4}$$

APPENDIX E

ASYMPTOTIC EXPANSION OF TAIL

We use the following asymptotic expansion of ${}_1F_2$ for large z to find an approximate expression for the truncated tail integral.

$$
\begin{aligned}
&{}_1F_2(a_1; b_1, b_2; z) \\
&\approx \frac{\Gamma(b_1)\,\Gamma(b_2)\,(-z)^{a_1}}{\Gamma(-a_1+b_1)\,\Gamma(-a_1+b_2)} \left(1 + \frac{a_1\ (1+a_1-b_1)\ (1+a_1-b_2)}{z}\right. \\
&\left. + \frac{a_1\ (1+a_1)\ (1+a_1-b_1)\ (2+a_1-b_1)\ (1+a_1-b_2)\ (2+a_1-b_2)}{2\,z^2} + \dots\right) \\
&+ \frac{(-z)^{\chi}\,\Gamma(b_1)\,\Gamma(b_2)}{2\sqrt{\pi}\,\Gamma(a_1)} \left(\cos\left(2\sqrt{-z} + \pi\chi\right)\ \left(1 + \frac{d_2}{z} + \dots\right)\right. \\
&\left. + \sin\left(2\sqrt{-z} + \pi\chi\right)\ \left(\frac{d_1}{\sqrt{-z}} + \dots\right)\right)
\end{aligned}
\tag{E.1}
$$

for large z [Wolfram Research,Inc., 2001a] and

$$\chi = \frac{1}{2}\left(\frac{1}{2} + a_1 - b_1 - b_2\right)$$

$$d_1 = \frac{1}{16}\left(-3 + 12\,{a_1}^2 - 4\,{b_1}^2 + 8\,b_2 - 4\,{b_2}^2 + 8\,b_1\ (1+b_2) - 8\,a_1\ (1+b_1+b_2)\right)$$

$$
\begin{aligned}
d_2 = \frac{1}{512}&\left(-15 + 144\,{a_1}^4 + 16\,{b_1}^4 + 16\,b_2 + 56\,{b_2}^2 - 64\,{b_2}^3 + 16\,{b_2}^4\right. \\
&-64\,{b_1}^3\ (1+b_2) - 64\,{a_1}^3\ (7 + 3\,b_1 + 3\,b_2) + 8\,{b_1}^2\ \left(7 + 8\,b_2 + 12\,{b_2}^2\right) \\
&+16\,b_1\ \left(1 + 25\,b_2 + 4\,{b_2}^2 - 4\,{b_2}^3\right) \\
&-8\,{a_1}^2\ \left(-43 + 4\,{b_1}^2 - 72\,b_2 + 4\,{b_2}^2 - 8\,b_1\ (9 + 5\,b_2)\right) \\
&\left.+16\,a_1\ \left(-1 + 4\,{b_1}^3 - 25\,b_2 - 4\,{b_2}^2 + 4\,{b_2}^3 - 4\,{b_1}^2\ (1+b_2) - b_1\ \left(25 + 40\,b_2 + 4\,{b_2}^2\right)\right)\right)
\end{aligned}
$$

Substitute this expression for ${}_1F_2$ in (7.7) and get

$$\begin{aligned}
&\int_{w_t}^{\infty} w^{1-\gamma} J_o(wr)dw \\
&= r^{\gamma-2} \int_{rw_t}^{\infty} u^{1-\gamma} J_o(u)du \\
&\approx\ r^{\gamma-2} \left(-\frac{\left(-15+16\,\gamma+128\,\gamma^2\right)\ (\cos(rw_t)-\sin(rw_t))}{128\,\sqrt{\pi}(rw_t)^{\gamma+\frac{3}{2}}} \right. \\
&\quad \left. + \frac{\cos(rw_t)-\sin(rw_t)}{\sqrt{\pi}(rw_t)^{\gamma-\frac{1}{2}}} + \frac{(-3+8\,\gamma)\ (\cos(rw_t)+\sin(rw_t))}{8\,\sqrt{\pi}(rw_t)^{\gamma+\frac{1}{2}}} + \dots \right).
\end{aligned}$$

BIBLIOGRAPHY

M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions*. Dover, New York, 9 edition, 1965.

M. Amann, J. Cofala, C. Heyes, Z. Klimont, R. Mechler, M. Posch, and W. Schöpp. The rains model. documentation of the model approach for the rains peer review 2004. http://www.iiasa.ac.at/rains/review/review-full.pdf, 2004.

V. P. Aneja, J. P. Chauhan, and J. T. Walker. Characterization of atmospheric ammonia emissions from swine waste storage and treatment lagoons. *Journal of Geophysical Research*, 105(D9):11535–11545, May 2000.

W. A. H. Asman, P. Cellier, S. Genermont, N. J. Hutchings, and S. G. Sommer. Ammonia emission research: from emission factors to process descriptions. *Eurotrac Newsletter*, 20, 1998.

A. F. Bouwman, D. S. Lee, W. A. H. Asman, F. J. Dentener, K. W. V. D. Hoek, and J. G. J. Olivier. A global high-resolution emission inventory for ammonia. *Global Biogeochemical Cycles*, 11(4):561–587, December 1997.

T. J. Butler and G. E. Likens. Weekly and daily precipitation chemistry network comparisons in the eastern u.s.: Nadp/ntn vs. map3s/airmon. *Atmospheric Environment*, 32:3749–3765, 1998.

D. W. Byun and J. K. S. Ching, editors. *Science Algorithms of the EPA Models-3 Community Multiscale Air QuaLity Model (CMAQ) Modeling System, EPA/600/R-99/030*. Off. of Res. Dev., Research Triangle Park, N.C., 1999.

N. Cressie. *Statistics for Spatial Data*. Wiley, 1993.

M. J. Daniels, F. Dominici, J. M. Samet, and S. L. Zeger. Estimating particulate matter-mortality dose-response curves and threshold levels: An analysis of daily time-series for the 20 largest us cities. *American Journal of Epidemiology*, 152(5): 397–406, 2000.

C. de Boor. *A practical guide to splines*. Springer, 2001.

I. G. Enting. *Inverse Problems in Atmospheric Constituent Transport*. Cambridge University Press, 2002.

M. G. Genton and D. J. Gorsich. Nonparametric variogram and covariogram estimation with fourier-bessel matrices. *Computational statistics & Data Analysis*, 41: 47–57, 2002.

A. Gilliland, R. L. Dennis, S. J. Roselle, and T. E. Pierce. Seasonal NH_3 emission estimates for the eastern United States. *Journal of Geophysical Research*, 108(D15), 2003. doi: 10.1029/2002JD003063.

G. H. Givens and J. A. Hoeting. *Computational Statistics*. Wiley, 2005.

G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, third edition, 1996.

P. Hall, N. Fisher, and B. Hoffmann. On the nonparametric estimation of covariance functions. *Annals of Statistics*, 22:2115–2134, 1994.

M. S. Handcock and J. R. Wallis. An approach to statistical spatial temporal modeling of meteorological fields. *Journal of the American Statistical Association*, 89:368–378, 1994.

M. Jun and M. L. Stein. Statistical comparison of observed and cmaq modeled daily sulfate levels. *Atmospheric Environment*, pages 4427–4436, 2004.

R. Kistler, E. Kalnay, W. Collins, S. Saha, G. White, J. Woollen, M. Chelliah, W. Ebisuzaki, M. Kanamitsu, V. Kousky, H. van den Dool, R. Jenne, and M. Fiorino. Ncep-ncar 50-year reanalysis: monthly mean cd rom and documentation. *Bull. Amer. Meteor. Soc.*, 82:247–267, 2001.

P. McCullagh and J. A. Nelder. *Generalized Linear Models*. Chapman and Hall, London, second edition, 1989.

R. D. Peng, F. Dominici, R. Pastor-Barriuso, S. L. Zeger, and J. M. Samet. Seasonal analyses of air pollution and mortality in 100 us cities. *American Journal of Epidemiology*, 161(6):585–594, 2005.

C. A. Pope. Air pollution and health - good news and bad. *The New England Journal of Medicine*, 351:1132–1134, 2004.

P. A. Roelle and V. P. Aneja. Characterization of ammonia emissions from soils in the upper coastal plain, north carolina. *Atmospheric Environment*, 36:1087–1097, 2002.

A. Shapiro and J. D. Botha. Variogram fitting with a genral class of conditionally nonnegative definite functions. *Computational Statistics & Data Analysis*, 11:87–96, 1991.

M. L. Stein. *Statistical Interpolation of Spatial Data: Some Theory for Kriging*. Springer, 1999.

M. L. Stein, X. Shen, and P. E. Styer. Application of a simple regression model to acid rain data. *Can. J. Stat.*, pages 331–346, 1993.

P. E. Styer and M. L. Stein. Acid deposition models for detecting the effect of changes in emissions: An exploratory investigation utilizaing meteorological variables. *Atmospheric Environment, Part A*, 26:3019–3028, 1992.

U.S. Environmental Protection Agency. *National air pollution trends:1990-1998, EPA 454/R-00-003*. Office of Air Quality Planning and Standards, Research Triangle Park, N.C., 2000.

Wolfram Research,Inc. Generalized hypergeometric function ${}_1F_2$. `http://functions.wolfram.com/HypergeometricFunctions/`, 2001a.

Wolfram Research,Inc. Polygamma function. `http://functions.wolfram.com/GammaBetaErf/PolyGamma`, 2001b.

Wolfram Research,Inc. Gamma function. `http://functions.wolfram.com/Gamma`, 2001c.

A. M. Yaglom. *Correlation Theory of Stationary and Related Random Functions*, volume 1. Springer-Verlag, New York, 1987.

G. Yarwood, S. Lau, Y. Jia, P. Karamchandani, and K. Vijayaraghavan. Modeling atmospheric mercury chemistry and deposition with camx for a 2002 annual simulation. 2003.