ABSTRACT

BRANTLEY, HALLEY LEE. Quantile Regression and Hierarchical Models for Near-Source Air Quality Data. (Under the direction of Joseph Guinness and Montserrat Fuentes).

Industrial facilities such as refineries and chemical plants emit hazardous air pollutants from a variety of sources. Some emission sources, such as stacks, have been well characterized through emissions tests and engineering calculations, but other sources including equipment leaks, process malfunctions, and waste ponds are more difficult to measure and model due to their temporal variability. A variety of new measurement and monitoring approaches have been developed to locate and quantify these types of sources. In this dissertation, we present a collection of statistical methods and models for analyzing these new sources of data. Chapter 1 provides an introduction to these methods and some helpful background.

In Chapter 2, we derive the properties and demonstrate the desirability of a model-based method for estimating the spatially-varying effects of covariates on the quantile function. By modeling the quantile function as a combination of I-spline basis functions and Pareto tail distributions, we allow for flexible parametric modeling of the extremes while preserving non-parametric flexibility in the center of the distribution. We further establish that the model guarantees the desired degree of differentiability in the density function and enables the estimation of non-stationary covariance functions dependent on the predictors. We demonstrate through a simulation study that the proposed method produces more efficient estimates of the effects of predictors than other methods, particularly in distributions with heavy tails. To illustrate the utility of the model we apply it to measurements of benzene collected around an oil refinery to determine the effect of an emission source within the refinery on the distribution of the fence line measurements.

In Chapter 3, we address the problem of estimating smoothly varying baseline trends in time series data. This problem arises in several applications, including chemistry, macroeconomics, and medicine; however, our study is motivated by the analysis of data from low cost air quality sensors. Our methods extend the quantile trend filtering framework to allow the estimation of multiple quantile trends simultaneously while ensuring that the quantiles do not cross. To handle the computational challenge posed by very long time series, we propose a parallelizable alternating direction method of moments.
ADMM) algorithm. The ADMM algorithm enables the estimation of trends in a piecewise manner, both reducing the computation time and extending the limits of the method to larger data sizes. We also address smoothing parameter selection and propose a modified criterion based on the extended Bayesian Information Criterion. Through simulation studies and our motivating application to low cost air quality sensor data, we demonstrate that our model provides better quantile trend estimates than existing methods and improves signal classification of low-cost air quality sensor output.

In the final Chapter, we propose a hierarchical Bayesian model for the de-trended sensor measurements to estimate both the location and strength of multiple temporally variable emission sources within a facility. Co-located 3D wind measurements are used as inputs into a backward Lagrangian Stochastic dispersion model, which simulates particle trajectories backward in time. The backward particle trajectories are transformed into predictors that can be used in a functional linear regression. Our simulation study illustrates the ability of our model to accurately characterize sources under different meteorological conditions and source profiles. We apply our model to sensor measurements collected at the perimeter of a facility in Louisville, KY and identify several sources active during the four days analyzed.
Quantile Regression and Hierarchical Models for Near-Source Air Quality Data

by

Halley Lee Brantley

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

Statistics

Raleigh, North Carolina

2019

APPROVED BY:

________________________________________________________________________
Joseph Guinness  Montserrat Fuentes
Co-chair of Advisory Committee  Co-chair of Advisory Committee

________________________________________________________________________
Eric Chi  Brian Reich

________________________________________________________________________
H. Christopher Frey
DEDICATION

To my wonderful husband, Suchit Mehrotra.
BIOGRAPHY

The author was born and raised Fort Worth, TX. She moved to North Carolina to study Chemistry at Davidson College and to Durham to get her Masters in Environmental Management from Duke University. She discovered her love of statistics while working for the EPA analyzing air quality data and is excited to start the next chapter of her life as a data scientist at United Health Group.
ACKNOWLEDGEMENTS

I would not have completed this dissertation without the endless support of my parents, Rickey and Ellen Brantley, and I am so thankful for my inspiring and encouraging siblings Ben, Hannah, and Meg. I am also eternally grateful to my NC family, Drs. Genevieve Lipp and Laura Carlson, for always being there to celebrate or commiserate and for helping me to remember the important things in life. I would also like to thank my EPA mentor Eben Thoma for his data, expertise, and inspiration, and my advisor, Dr. Montse Fuentes, and committee members, Dr. Eric Chi, Dr. Brian Reich, and Dr. Chris Frey for their knowledge, advice, guidance, and support. Finally, I want to express my immense gratitude to my advisor Dr. Joseph Guinness for all of his patience, thoughtfulness, persistence, encouragement, and help throughout the writing of this dissertation.
# TABLE OF CONTENTS

LIST OF TABLES ................................................. vii

LIST OF FIGURES ............................................. viii

Chapter 1 INTRODUCTION ....................................... 1

Chapter 2 SMOOTH DENSITY SPATIAL QUANTILE REGRESSION .... 5
  2.1 Introduction ........................................... 5
  2.2 Proposed Model ......................................... 7
  2.3 Model Properties ....................................... 10
    2.3.1 Validity Of Quantile Function ..................... 10
    2.3.2 Continuity and Differentiability .................. 10
    2.3.3 Expectations and Covariance ...................... 12
  2.4 Simulation Study ....................................... 13
  2.5 Application ........................................... 17
    2.5.1 Data ............................................... 17
  2.6 Discussion ............................................ 21

Chapter 3 BASELINE DRIFT ESTIMATION FOR AIR QUALITY DATA
  USING QUANTILE TREND FILTERING ......................... 23
  3.1 Introduction ........................................... 23
  3.2 Baseline Trend Estimation ............................. 26
    3.2.1 Background ......................................... 26
    3.2.2 Quantile Trend Filtering ......................... 28
    3.2.3 ADMM for Big Data ................................ 29
  3.3 Model Selection ....................................... 36
    3.3.1 Validation ......................................... 36
    3.3.2 Information Criteria ............................... 37
  3.4 Simulation Studies .................................... 38
    3.4.1 Estimating Quantiles .............................. 39
    3.4.2 Peak Detection ..................................... 40
  3.5 Analysis of Air Quality Data ........................... 45
    3.5.1 Short series of air quality measurements ....... 45
    3.5.2 Long series of air quality measurements ....... 47
  3.6 Conclusion and Discussion ................................ 49
  3.7 Acknowledgments ....................................... 51
  3.8 Appendix ............................................... 51

Chapter 4 CHARACTERIZING TEMPORALLY VARYING EMISSION
  SOURCES USING BAYESIAN MODELING ....................... 52
LIST OF TABLES

Table 2.1 True parameter functions by design used in the simulation study. The location is given as \( s = (s_1, s_2) \). \( \Phi^{-1}(\tau) \) represents the quantile function of the standard normal evaluated at \( \tau \) and \( Q_{Pareto} \) represents the quantile function of the Pareto distribution with the given parameters....................... 14

Table 2.2 Comparison of fitted \( \beta(\tau) \) functions \( \tau = (.05, .06, ..., .94, .95) \). COV represents the coverage of the 95% credible (IQR and GAUS) or confidence interval (NCQR). .................. 15

Table 2.3 Comparison of fitted \( \beta(\tau) \) functions \( \tau = (.950, .951, ..., .994, .995) \). COV represents the coverage of the 95% credible (IQR and GAUS) or confidence interval (NCQR). .................. 16

Table 2.4 Comparison of mean estimated log scores .................. 16

Table 2.5 Estimated log-scores for training and validation data by method. .................. 20

Table 4.1 Comparison of simulation results by design, number of bases, and grid size, SE is the standard error for the 50 simulated datasets. ... 66

Table 4.2 Comparison of in-sample and LOO elpd for June 17-19 by node, grid size and number of bases. .................. 68
LIST OF FIGURES

Figure 2.1  Example set of normalized B-spline (left) and corresponding I-spline (right) basis functions. Dotted vertical lines indicate knot locations. ................................................................. 9

Figure 2.2  Examples of quantile function (top row) and corresponding density function (bottom row) constructed using different bases. .................. 10

Figure 2.3  Benzene measurements by time and location. Source locations, e₁ and e₂, are shown in black. Points have been jittered slightly to improve visibility. ................................................................. 18

Figure 2.4  Wind roses for different seasons. ......................................................... 19

Figure 2.5  Estimated predictor effect by quantile and location. .......................... 21

Figure 2.6  Spatial variation in the effect of the predictors on the 95th quantile of fenceline benzene measurements. ................................. 22

Figure 3.1  Example of 3 co-located SPod PID sensor readings over a 24 hour period. ........................................................................................................... 24

Figure 3.2  Examples of trend filtering solutions (red) and 15th quantile trend filtering solution (blue). Standard trend filtering performs well in the no-signal case (a) but struggles to distinguish between the slowly varying trend and the rapidly-varying signal (b). The quantile trend is not affected by the signal and provides an estimate of the baseline. ................................................................. 27

Figure 3.3  Window boundaries and trends fit separately in each window. Each window’s trend estimate is plotted in a different line type. ............. 30

Figure 3.4  Trend fit with our ADMM algorithm with 3 windows (converged in 7 iterations), compared to trend from simultaneous fit. .............. 31

Figure 3.5  Timing experiments comparing quantile trend filtering with varying numbers of windows by data size. ................................................................. 36

Figure 3.6  Simulated data with true quantile trends. ........................................... 39

Figure 3.7  RMSE by design, method, quantile and data size. Points and error bars represent mean RMSE ± twice the standard error. ................. 41

Figure 3.8  RMSE by method, quantile, and data size for peaks design. ............ 42

Figure 3.9  Example signal classification using threshold. Red indicates true signal (yᵢ − θᵢ > 0.5), blue indicates observations classified as signal, i.e. values greater than 1.2 after baseline removal using detrend_eBIC. ................................................................. 44

Figure 3.10 Class averaged accuracy by threshold, data size, and method (1 is best 0.5 is worst). ................................................................. 44
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.11</td>
<td>Estimated 5\textsuperscript{th} quantile trends for SPods a, b, and c, using \texttt{qsreg} and \texttt{detrendr}. SPod c contains some missing values that were interpolated before the trends were estimated. \texttt{qsreg} is more influenced by the signal component apparent on both nodes resulting in under-smoothing.</td>
</tr>
<tr>
<td>3.12</td>
<td>Rugplot showing locations of signal after baseline removal using the \texttt{detrendr} estimate of 5th quantile. Horizontal dashed lines represent the thresholds calculated using the 90\textsuperscript{th}, 95\textsuperscript{th}, and 99\textsuperscript{th} quantiles of the detrended data. The 95\textsuperscript{th} quantile (black) was used to classify the signal shown as vertical lines at the bottom of the plot.</td>
</tr>
<tr>
<td>3.13</td>
<td>Variation of Information between sensors by method (color), quantile (columns) and threshold (shape) for two hour time period.</td>
</tr>
<tr>
<td>3.14</td>
<td>Variation of Information between sensors by method (color), quantile (columns) and threshold (shape) for full day.</td>
</tr>
<tr>
<td>3.15</td>
<td>SPod a versus SPod c before and after de-trending with \texttt{qsreg} and \texttt{detrendr}.</td>
</tr>
<tr>
<td>3.16</td>
<td>Variation of Information (VI) for \texttt{detrendr} and \texttt{qsreg} by quantile trend and threshold. Each point represents a full day of data. The dashed line represents y=x. In most cases \texttt{detrendr} results in a lower VI than \texttt{qsreg}.</td>
</tr>
<tr>
<td>4.1</td>
<td>Co-located measurements from SPod A and B over time and compared to each other.</td>
</tr>
<tr>
<td>4.2</td>
<td>Wind roses by measurement day, direction indicates source of wind. The majority of the wind is coming from the south/southwest. Wind speed values are in m/s.</td>
</tr>
<tr>
<td>4.3</td>
<td>Polar plots of PID signal binned by wind direction and speed.</td>
</tr>
<tr>
<td>4.4</td>
<td>50 calculated back-trajectories during very stable (a) and very unstable (b) atmospheric conditions.</td>
</tr>
<tr>
<td>4.5</td>
<td>(a) Source locations for simulation designs, the triangle represents the sensor location and the color represents the average value of trajectory weights in each cell. (b) Source strengths over time used for simulation designs.</td>
</tr>
<tr>
<td>4.6</td>
<td>Proportion of backward trajectories that pass through source cell 1 and source cell 2 for 100m (left) grid and 200m grid (right).</td>
</tr>
<tr>
<td>4.7</td>
<td>For each of the three designs, several simulated responses are shown.</td>
</tr>
<tr>
<td>4.8</td>
<td>Estimates of source strength for designs 1, 2, and 3 (left to right). Bold lines indicate true source strength, dashed lines indicate mean of simulation estimates.</td>
</tr>
<tr>
<td>4.9</td>
<td>Estimates of mean source strength for design 3 by location.</td>
</tr>
<tr>
<td>4.10</td>
<td>SPod A and B measurements compared to the predicted values using 100 m grid cells and 9 basis functions.</td>
</tr>
</tbody>
</table>
Figure 4.11 Observed versus predicted for both SPods using 100 m grid cells and 9 basis functions.

Figure 4.12 Estimated source locations (left) and temporal profiles (right) from SPod A, using 100 m grid and 10 bases. Labeled values correspond to the grid cell IDs. Left: colors represent integrated source strengths. Right: red lines represent the medians of the posterior samples of the source strengths for each grid cell. The shaded areas represent the 80% credible intervals and are only shown for the labeled sources.

Figure 4.13 Estimated source locations (left) and temporal profiles (right) from SPod B, using 100 m grid and 10 bases. Labeled values correspond to the grid cell IDs. Left: colors represent integrated source strengths. Right: red lines represent the medians of the posterior samples of the source strengths for each grid cell. The shaded areas represent the 80% credible intervals and are only shown for the labeled sources.
Modeling near-source air quality measurements can present a variety of challenges including complex spatial and temporal correlation and non-normal distributions. Often the statistical quantities of interest for pollutant concentrations are not the mean values but the quantiles. Understanding the effects of predictors on the upper quantiles of pollutant concentrations can be more important than understanding their effects on the mean, when detecting emission sources and reducing health impacts. To accurately estimate smoothly varying background concentrations or sensor drift, the lower quantiles are of the most interest. Even in cases when the mean is the quantity of interest, the correlation structure present in both the response and predictors can present a challenge for modeling. This chapter provides an introduction to, and a summary of, our contributions to the analysis of near-source air quality measurements.

Since Koenker and Bassett published the seminal paper on quantile regression in 1978 [Koenker and Bassett Jr, 1978], an immense body of literature, both frequentist and Bayesian, has been developed. The quantile function for a random variable $Y$ with
a right-continuous distribution function $F(y) = Pr(Y \leq y)$ is defined on $\tau \in [0, 1]$ as

$$Q_Y(\tau) = F^{-1}(\tau) = \inf\{y | F(y) \geq \tau\} \quad (1.1)$$

In that first paper, Koenker and Bassett exploited an alternative definition of an ordinary sample quantile: given a sample $\{y_i : i = 1, \ldots, n\}$, the $\tau$th sample quantile $\hat{Q}_Y(\tau)$ can be defined as any solution to the minimization problem

$$\min_{b \in \mathbb{R}} \left[ \sum_{\{i : y_i \geq b\}} \tau|y_i - b| + \sum_{\{i : y_i < b\}} (1 - \tau)|y_i - b| \right] = \min_{b \in \mathbb{R}} \sum_{i=1}^{n} \rho_\tau(y_i - b) \quad (1.2)$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the check loss function. Quantile regression resulted from the replacement of $b$ with $X\beta$, where $X \in \mathbb{R}^{n \times p}$ is a matrix of predictors and $\beta \in \mathbb{R}^p$ quantifies the effects of the predictors on the $\tau$th quantile. Because the problem can be framed as a linear program, it can be solved efficiently using standard linear programming algorithms [Koenker and d’Orey, 1987, 1994].

Yu and Moyeed [2001] introduced the idea of Bayesian quantile regression employing the asymmetric Laplace distribution (ASL) as the working likelihood. Assuming $y_i = x_i^T \beta(\tau) + \epsilon_i$ for $i = 1, \ldots, n$, and $\epsilon_i \sim ASL(0, \sigma, \tau)$ with $\tau \in (0, 1)$, the density of $\epsilon_i$ is given by:

$$f(\epsilon_i | 0, \sigma, \tau) = \frac{\tau(1 - \tau)}{\sigma} \exp \left\{ -\frac{\rho_\tau(\epsilon_i)}{\sigma} \right\}. \quad (1.3)$$

Maximizing this likelihood with respect to $\beta$ is equivalent to the minimization problem in (1.2), but if the observed response is not distributed according to the ASL the posterior inference is invalid.

The above methods were designed to estimate the effect of predictors on a single quantile level. If the effect on multiple quantiles is desired, the standard methods can be applied separately for each quantile level of interest. A well-known issue with this approach is that it does not ensure a quantile function of $Y$ that is monotonic in $\tau$ for all possible covariate values. It is possible that the predicted value of the 85-th quantile of $Y|X$ is higher than the predicted value of the 90-th quantile which contradicts the definition of quantile. Since the quantile estimates are expected to converge to the true quantile function as sample size increases, the estimated quantiles are more likely to cross
in regions with a paucity of data like the upper and lower tails of a distribution.

Reich et al. [2012] proposed a model for the entire quantile function that could incorporate multiple predictors by using basis functions:

\[
Q(\tau|x_1, \ldots, x_P) = \sum_{m=1}^{M} \sum_{p=1}^{P} \theta_{m,p} B_m(\tau) x_p
\]

where \(B_m\) are the basis functions, \(\theta_{m,p}\) are the coefficients and \(x_p\) are observed predictors. A significant advantage to modeling the entire quantile function is that an estimate of the true likelihood function is obtained, along with valid credible intervals for the parameters. Reich et al. [2012] limited the covariate domain to \([0, 1]^p\) and used B-spline basis functions, constraining \(\theta_{m,p} \geq \theta_{m-1,p}\) to ensure monotonicity. In subsequent papers, the spline basis functions were replaced with piecewise parametric basis functions [Reich and Smith, 2013, Reich, 2012]. Benefits of the piecewise parametric basis functions include a natural centering on the heteroskedastic parametric model and a closed form for the density, while disadvantages include a lack of flexibility compared to spline functions and discontinuity in the density functions.

Zhou et al. [2011] proposed modeling the center of the quantile function using integrated piecewise polynomial spline (I-spline) basis functions and the tails using the quantile function of the Generalized Pareto Distribution (GPD). Using a parametric distribution for the tails rather than the more flexible splines can prevent over-fitting and enable the estimation of unbounded densities. Smith et al. [2015] expanded the I-spline regression model by allowing the effect of covariates to vary by quantile and estimating the shape parameters of the tails. In Chapter 2, we establish the conditions under which this model guarantees the desired degree of differentiability in the density function. Gaussian process priors are used for the coefficients of the basis functions in order to allow the effects of the predictors to vary across space, and we demonstrate that the resulting covariance of the response is non-stationary and dependent on the predictors. We demonstrate the utility of the model by estimating the effects of emission sources on the distribution of benzene measurements collected around the perimeter of an oil refinery.

While the methods described above constrain the relationship between a quantile and predictor to be linear, a number of non-parametric quantile regression methods have also been developed. Koenker et al. [1994] proposed a method for estimating quantile
smoothing splines by minimizing

$$\minimize_g \sum_i \rho_r(y_i - g(x_i)) + \lambda \int |g''(x)|^p dx,$$

where $\lambda$ is a nonnegative regularization parameter, and $g(x)$ is a smooth quantile function. They focused on the case where $p = 1$. Nychka et al. [1995] and Oh et al. [2011] focused on $p = 2$ and devised an algorithm for quickly solving the optimization problem along with a generalized cross-validation method for choosing the regularization parameter. Kim et al. [2009] advanced $\ell_1$-trend filtering as an alternative to smoothing splines, and suggested $\ell_1$-quantile-trend-filtering as an extension but did not explore it further. In Chapter 3, we extend the quantile trend filtering framework to model multiple quantiles simultaneously, ensuring non-crossing through constraints to ensure validity and improve trend estimates. We also develop a parallelizable ADMM algorithm to extend the method to larger series and reduce computation time. Finally, we propose a modified criterion for selecting the regularization parameter which improves estimates of the baseline drift in low-cost air quality sensors.

Our contribution in the final chapter (Ch. 4) is a model combining the de-trended low-cost sensor data from Chapter 3 with 3D wind measurements, to estimate both the location and strength of multiple temporally variable emission sources. We propose a Bayesian hierarchical model for the distribution of the measurements as a function of unknown source strengths. We use informative priors to ensure the source estimates are smooth, non-negative functions and to stabilize the model variance by shrinking the estimates toward zero. This model enables the estimation of multiple temporally varying source strengths from a series of concentration measurements collected at a single location, which, to our knowledge, no previously published methods have done.
2.1 Introduction

Quantile regression offers an important alternative to traditional mean regression for problems where the interest lies not in the center of the distribution but in some other aspect. Since the first quantile regression paper was published by Koenker and Bassett Jr [1978], an immense body of literature has been developed and is reviewed in Koenker [2005]. Yu and Moyeed [2001] proposed a form of Bayesian quantile regression employing the Asymmetric Laplace Distribution (ASL) as the working likelihood, due to its similarity to the check loss function used by Koenker and Bassett Jr [1978]. Both of these approaches perform separate analyses for each quantile level of interest. When quantiles are estimated separately, there is no guarantee of a valid non-decreasing quantile function. There are several approaches to address this issue. The first one is a two-stage method: in the first-stage the quantiles are fit separately using one of the above methods, and in the
second stage the estimates are smoothed to ensure monotonicity. This approach has been taken by a variety of authors including Neocleous and Portnoy [2008], Rodrigues and Fan [2016], and Reich et al. [2012] who used it as a more computationally efficient Bayesian spatial method. Bondell et al. [2010] embed a constraint that ensures monotonicity into the minimization problem, while Cai and Jiang [2015] use prior specifications to ensure constraints in the Bayesian framework.

The final approach, which we will adopt and extend, is to model the entire quantile function jointly using basis functions. This is the approach taken by Reich et al. [2012] and others [Reich, 2012, Smith et al., 2015] and is more naturally implemented using a Bayesian framework. Regardless of the approach taken, ensuring monotonicity requires either some form of distributional assumption, or constraints on the quantile regression coefficients and the parameter space of the predictors. Cai and Jiang [2015] demonstrated that when predictors are constrained to be positive, the quantile function is monotonic for every possible predictor value if and only if the basis functions are monotonic. This is the approach taken by Zhou et al. [Zhou et al., 2011, 2012] who first proposed the I-spline quantile regression model whose properties we derive in this paper.

As in mean regression, a method of incorporating spatial correlation into quantile regression is to model spatially-varying parameters using Gaussian process priors. Lum and Gelfand [2012] use the ASL for the likelihood and incorporate spatial correlation by modeling the error as a function of a Gaussian process and an independent and identically distributed exponential random variable. For large datasets they propose an asymmetric Laplace predictive process, extending the method introduced by Banerjee et al. [2008]. However, the use of the ASL does not allow for valid posterior inference because it does not represent the true likelihood of the observations. Yang and He [2015] combined spatial priors with their Bayesian empirical likelihood approach for modeling the conditional quantiles in the presence of both predictors and spatial correlation, but their method only allows for effects to be estimated at a small fixed number of quantile levels. Several previous methods of modeling a spatially varying conditional quantile function using basis functions have also been advanced [Reich et al., 2012, Reich, 2012].

We consider the model first proposed by Zhou et al. (2011, 2012) where the quantile function is modeled as a combination of I-splines and the Generalized Pareto Distribution (GPD). The GPD is used to model the tails because it has been shown to be the natural choice for exceedances over a threshold [Davison and Smith, 1990] and provides flexibility
as a result of the shape parameter which controls boundedness and the existence of moments. A full description of the I-spline quantile regression model for both independent and spatially correlated data is given in Section 2. In this paper, we formulate the conditions under which the resulting density has the desired degree of differentiability and derive the marginal expectations and spatial covariances which can be non-stationary (Section 3). Our simulation studies demonstrate that ensuring a smooth density can lead to more accurate effect estimates and predictive distributions, compared with methods that do not ensure differentiability (Section 4). We apply the method to benzene measurements from a petrochemical facility to determine the effects of emission sources on concentrations (Section 5).

### 2.2 Proposed Model

We model the quantile function of the stochastic process $Y(s)$ as a linear combination of the predictors:

$$Q(\tau | s, x(s)) = \beta_0(\tau, s) + \sum_{p=1}^{P} x_p(s) \beta_p(\tau, s),$$

(2.1)

where $x(s) = (x_1(s), ..., x_P(s)) \in \mathbb{R}_+^P$, is the vector of predictors observed at location $s$, $\beta_0(\tau, s)$ is the quantile function at location $s$ when all predictors are 0, and $\beta_p(\tau, s)$ is the effect of predictor $p$ on quantile level $\tau$ at location $s$. We further follow the approach of Zhou et al. [2011] and model $\beta(\tau, s)$ as a linear combination of I-spline basis functions in the center of the distribution. We denote the $m^{th}$ I-spline basis function evaluated at $\tau$ as $I_m(\tau)$ and define the constant basis function $I_0(\tau) = 1$ for all $\tau$. While I-splines allow for a large degree of flexibility in the center of the distribution, unbounded distributions cannot be estimated using I-splines with a finite number of knots. To solve this issue we use the quantile function of the GPD to model the relationship of the covariate(s) to the process in the tails of the distribution. The model for $\beta_p(\tau, s)$ can then be expressed as
\[ \beta_p(\tau, s) = \begin{cases} 
\theta_0, p(s) - \frac{\sigma_{L, p}(s)}{\alpha_L(s)} \left( \frac{\tau}{\tau_L} \right)^{-\alpha_L(s)} - 1 & \tau < \tau_L \\
\sum_{m=0}^{M} \theta_{m, p}(s) I_m(\tau) + \left[ \sum_{m=0}^{M} \theta_{m, p}(s) \right] + \frac{\sigma_{U, p}(s)}{\alpha_U(s)} \left[ \left( \frac{1 - \tau}{1 - \tau_U} \right)^{-\alpha_U(s)} - 1 \right] & \tau_L \leq \tau \leq \tau_U \\
\sum_{m=0}^{M} \theta_{m, p}(s) & \tau > \tau_U, 
\end{cases} \]

(2.2)

where \( \tau_L \) and \( \tau_U \) are the thresholds between the tails and the center of the distribution, \( \theta_0, p \) is the location parameter at the lower tail, and \( \theta_{m, p}(s) \) represents the coefficient of the \( m \)th I-spline basis function and \( p \)th predictor at location \( s \). I-splines are monotonic polynomials formed by integrating normalized B-splines (Fig. 1) [Ramsay, 1988]. They are defined on a sequence of knots \( \{ \tau_0 = ... = \tau_k < ... < \tau_{M+1} = ... = \tau_{M+1+k} \} \), where \( k \) represents the degree of the polynomial and \( M \) is the number of non-constant basis functions.

The GPD has three parameters: the shape parameter \( \alpha \), the scale parameter \( \sigma \), and a location parameter \( \mu \). In our parameterization, the location parameter of the lower tail is equal to \( \theta_0, p(s) \) and the location parameter of the upper tail is equal to \( \sum_{m=0}^{M} \theta_{m, p}(s) \) to ensure the quantile function is continuous. We denote the shape parameters of the lower and upper tails as \( \alpha_L(s) \) and \( \alpha_U(s) \), respectively, and the scale parameters as \( \sigma_{L, p}(s) \) and \( \sigma_{U, p}(s) \). We require the shape parameter to be constant across predictors in order to ensure that the density in the tails follows a parametric distribution. The scale parameters vary by both predictor and location and allow the predictors to affect the tails differently. When \( \alpha < 0 \), the support of GPD is also bounded above, otherwise the domain is unbounded above. The case when \( \alpha = 0 \) is interpreted as the limit when \( \alpha \to 0 \), i.e. \( \frac{\sigma_{U, p}(s)}{\alpha_U(s)} \left( \frac{1 - \tau}{1 - \tau_U} \right)^{-\alpha_U(s)} - 1 \) is replaced with \( -\sigma_{U, p}(s) \log \left( \frac{1 - \tau}{1 - \tau_U} \right) \). The expectation exists if \( \alpha \) is less than 1, and the variance exists if \( \alpha \) is less than \( 1/2 \).

This model formulation ensures a quantile function that is continuous and differentiable at all but a finite number of points. We can thus exploit the result of Tokdar and Kadane (2012) who demonstrated that a differentiable and invertible quantile function corresponds with the density

\[ f(y) = \frac{1}{Q'(Q^{-1}(y))}. \]

(2.3)

To ensure the quantile function is monotonic we introduce latent parameters with Gaussian process priors, \( \theta^*_{m, p} \sim \mathcal{GP}(\mu^*_{m, p}, \Sigma^*_{m, p}) \) and define \( \theta_0, p(s) = \theta^*_{0, p}(s) \) and \( \theta_{m, p}(s) = \)
Figure 2.1 Example set of normalized B-spline (left) and corresponding I-spline (right) basis functions. Dotted vertical lines indicate knot locations.

\[ \exp \theta_{m,p}(s) \text{ for } m > 0. \] By using this formulation the resulting \( \theta_{m,p}(s) \) are modeled as log Gaussian processes. No constraints are placed on \( \theta_{0,p} \) which allows predictors to have a negative effect on the response.

The model formulation has many advantages including the ability to allow the effect of each predictor to vary by quantile level and by spatial location while guaranteeing a valid quantile function. It can also accommodate a variety of tail distributions including both bounded and unbounded tails. Furthermore, we show in Section 3 that we can guarantee the degree of differentiability of the corresponding density function.

Reich [2012] proposed a similar model, constructing the quantile function using parametric Gaussian basis functions. While the parametric basis functions allow for straightforward evaluation of the density, they do not guarantee a differentiable quantile function, which results in a non-continuous density function (Fig. 2). We show through both simulation and applied data analysis that constraining the density to be continuous and differentiable can result in better parameter estimates and out-of-sample scores.
2.3 Model Properties

2.3.1 Validity Of Quantile Function

Assuming an I-spline order $k > 1$, the proposed quantile function is continuous everywhere and is differentiable for all values of $\tau \in (0,1)$ except $\tau_L$ and $\tau_U$. Thus, a necessary and sufficient constraint to ensure a valid quantile function is $Q'(\tau) \geq 0$ for all $\tau$ at which the derivative exists. For all values of $\tau$ such that $\tau_L < \tau < \tau_U$, $Q'(\tau) = \sum_{m=1}^{M} B_m(\tau) \sum_{p=1}^{P} \theta_{m,p} x_p$. By definition, $B_0(\tau) = 0$ for all $\tau$ and $B_m(\tau) \geq 0$ for all $m$ and $\tau$. Without loss of generality, we will henceforth assume that the predictors are all non-negative, i.e. $x \in \mathbb{R}_+^P$, therefore a sufficient constraint to ensure a valid quantile function is $\theta_{m,p} \geq 0$ for all $p$ and $m > 0$. If $\sigma_{L,p} > 0$ for any $p$ and $\tau \leq \tau_L$, $Q'(\tau) = \sigma_{L,p} x_p \tau^{-\alpha_L-1} \tau_L^{\alpha_L} > 0$. Similarly if $\sigma_{U,p} > 0$ for any $p$, $Q'(\tau) > 0$ when $\tau \geq \tau_U$.

2.3.2 Continuity and Differentiability

In many cases, such as the application described below, it is desirable to ensure that the density is continuous and smooth. Proposition 1 establishes the conditions for continuity
of the density function.

**Proposition 1.** Let $Y$ have a quantile function as defined in (2.1) and (2.2) with $\sigma_{L,p} > 0$ for at least one $p$, then the density of $Y$ is continuous at $Q(\tau_L|x, \Theta)$ for any $x \in \mathbb{R}_+^P$ if and only if

$$\theta_{1,p} = \frac{\sigma_{L,p}}{\tau_L I_1'(\tau_L)},$$

for all $p$. Similarly, given $\sigma_{U,p} > 0$ for at least one $p$, the density of $Y$ is continuous at $Q(\tau_U|x, \Theta)$ if and only if

$$\theta_{M,p} = \frac{\sigma_{U,p}}{(1 - \tau_U)I_M'(\tau_U)}.$$  \hfill (2.5)

Having clarified the conditions for a continuous density, which can be viewed as 0th order differentiability, Theorem 1 proceeds to establish the conditions for $q$th order differentiability of the density function of $Y$.

**Theorem 1.** Let $Y$ have a quantile function as defined in (2.1) and (2.2) with an I-spline basis order greater than $q + 1$ and a density that is continuous and $(q - 1)^{th}$ order differentiable at $Q(\tau_L)$. If $\alpha_L$ is constrained so that Eq. A.1 does not result in $\theta_{q+1,p} < 0$, then $Y$ has a density that is $q^{th}$-order differentiable at $Q(\tau_L)$ for any $x \in \mathbb{R}_+^P$ if and only if

$$\theta_{q+1,p} = \frac{1}{I_{q+1}^{(q+1)}(\tau_L)} \left\{ \frac{-\sigma_{L,p}}{\alpha_L I_{q+1}^{(q+1)}(\tau_L)} (-\alpha_L - q)_{q+1} - \sum_{m=1}^{q} \theta_{m,p} I_{m}^{(q+1)}(\tau_L) \right\}$$

where $I_{q+1}^{(q+1)}(\tau_L)$ is the $(q + 1)$th order derivative of the $(q + 1)$th I-spline basis function, $(-\alpha_L - q)_{q+1} = \prod_{j=0}^{q}(-\alpha_L - j)$.

The conditions which guarantee differentiability at $\tau_U$ are similar and are given in the supplemental information. Combined with the positivity constraint on the $\theta$s, these results imply that the shape parameters have an upper bound that is a function of the knot placement. Ensuring a density that is first order differentiable results in the possible values for $\alpha_L$ being bounded above by $-1 - \tau_L I_{1}^{(2)}(\tau_L) I_{1}^{(1)}(\tau_L)$. This bound is a function of $I_{1}^{(1)}(\tau_L)$ and $I_{1}^{(2)}(\tau_L)$ which are functions of the first two knot locations. We can still model any tail behavior provided the outermost knots are placed sufficiently close.
2.3.3 Expectations and Covariance

While our models allow for flexible non-Gaussian distributions, sometimes the first two moments are of interest (e.g., for best linear unbiased prediction). We now elaborate on the various types of covariance structure that can be estimated using the proposed model. We model the covariances of the latent parameters \( \theta_{m,p} \) using covariance function \( C \) such that \( \text{Cov}[\theta_{m,p}(s), \theta_{m,p}(s')] = \eta_{m,p}^2 C(s, s') \) and \( \text{Var}[\theta_{m,p}(s)] = \eta_{m,p}^2 + \lambda_{m,p}^2 \). Consequently, the expectation of \( \theta_{m,p} \) can be expressed as

\[
E[\theta_{m,p}] = \mu_{m,p} = \exp[\mu_{m,p}^* + (\eta_{m,p}^2 + \lambda_{m,p}^2)/2]
\]

and the covariance of \( \theta_{m,p} \) is

\[
\Sigma_{m,p}(s, s') = \text{Cov}[\theta_{m,p}(s), \theta_{m,p}(s')] = \mu_{m,p}^2 (\exp[\eta_{m,p}^2 C(s, s')] - 1). \tag{2.7}
\]

In this section we describe the covariance of the case when \( \tau_L = 0 \) and \( \tau_U = 1 \), we elaborate on other cases in the supplementary material. Under these conditions, the conditional expectation of \( Y(s)|\Theta(s), x(s) \) is

\[
E[Y(s)|\Theta(s), x(s)] = \int_0^1 Q_Y[\tau|\Theta(s), x(s)]d\tau = \sum_m \sum_p \theta_{m,p}(s)x_{p,t}(s)G_m, \tag{2.8}
\]

where \( G_m = \int_0^1 I_m(\tau)d\tau \). We further marginalize over the log Gaussian processes \( \theta_{m,p}(s) \), with mean \( \mu_{m,p} \) and covariance \( \Sigma_{m,p} \), to obtain the expectation and covariance of \( Y(s) \)

\[
E[Y_t(s)|x(s)] = \sum_m \sum_p \mu_{m,p}(s)x_{p,t}(s)G_m. \tag{2.9}
\]

\[
\text{Cov}[Y(s), Y(s')|x(s), x(s')] = \sum_m \sum_p G_m^2 x_p(s)x_p(s')[\Sigma_{m,p}(s, s')] \tag{2.10}
\]

Through this simple case we can see that the covariance is dependent on the values of the predictors in addition to the covariance functions of the latent parameters. This dependence on the predictors can result in non-stationary covariances if \( x_p \) vary across space, even if \( C(s, s') \) is stationary.
2.4 Simulation Study

Our simulation studies demonstrate the superior efficiency of the proposed I-spline quantile regression method (IQR) using four designs from data generating models that are not in the proposed model class (Table 2.1). The designs include cases with both light tails (D1 and D3) and heavy tails (D2 and D4), and with (D3 and D4) and without (D1 and D2) spatial correlation. The designs illustrate the flexibility of the proposed method compared with previously established methods.

For each design the observed response is indexed as \( y_t(s_i) \), where \( t \in \{1, ..., n\} \) indexes the observations at a given location \( s_i \) with \( i \in \{1, ..., S\} \). The predictor vector \( x_{1,t} \) is generated by sampling from a uniform random variable in D1 and D2. In D3 and D4, \( z_t \) is generated by sampling from a Gaussian process with mean 0, and exponential covariance with range 1 and \( x_{1,t} = \Phi^{-1}(z_t) \), where \( \Phi^{-1}(\tau) \) is the quantile function of the standard normal. The predictor \( x_{2,t} \) is generated by sampling from a uniform random variable in all designs. The observed response is generated by drawing an independent random uniform variable \( u_t(s_i) \) and setting:

\[
y_t(s_i) = \beta_0(u_t(s_i), s_i) + \beta_1(u_t(s_i), s_i)x_{1,t}(s_i) + \beta_2(u_t(s_i), s_i)x_{2,t}(s_i). \tag{2.11}
\]

In all designs we assume multiple observations are obtained for each location. For each design we simulate \( B = 50 \) independent datasets. In D1 and D2 we simulate 1000 observations per dataset, assuming all observations are from a single location and thus independent. In D3 and D4 we use \( S = 16 \) locations evenly spaced on a unit square and simulate 100 observations per site for a total of 1600 observations per dataset. For each of the datasets we randomly assign 10% of the data to be used as validation data for the out-of-sample calculations and use the other 90% as training data. Computational details including a description of the Markov Chain Monte Carlo (MCMC) algorithm and prior specifications are included in the supplementary material.

We compare the estimates from the proposed model (IQR) with those from the model using parametric Gaussian basis functions (GAUS) proposed by Reich [2012] and the non-crossing quantile regression estimates (NCQR) proposed by Bondell et al. [2010]. For the IQR and GAUS methods the estimates of \( \beta(\tau, s) \) represent the means of the corresponding posterior samples. For the NCQR method the estimates of \( \beta(\tau, s) \) are obtained by minimizing the check loss function combined with the non-crossing constraint. The GAUS
Table 2.1 True parameter functions by design used in the simulation study. The location is given as \( s = (s_1, s_2) \), \( \Phi^{-1}(\tau) \) represents the quantile function of the standard normal evaluated at \( \tau \) and \( Q_{\text{Pareto}} \) represents the quantile function of the Pareto distribution with the given parameters.

<table>
<thead>
<tr>
<th>Design</th>
<th>Function 1</th>
<th>Function 2</th>
<th>Function 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>0.1( \Phi^{-1}(\tau) )</td>
<td>0.3( \tau )</td>
<td>( Q_{\text{Pareto}}(\tau, \alpha = -0.2, \mu = 0, \sigma = 0.1) )</td>
</tr>
<tr>
<td>D2</td>
<td>0.1( \Phi^{-1}(\tau) )</td>
<td>0.3( \tau )</td>
<td>( Q_{\text{Pareto}}(\tau, \alpha = 0.3, \mu = 0, \sigma = 0.3) )</td>
</tr>
<tr>
<td>D3</td>
<td>( 0.05 + 0.2s_1 s_2 \Phi^{-1}(\tau) )</td>
<td>( 0.3e^{s_2} + 0.2\tau )</td>
<td>( Q_{\text{Pareto}}(\tau, \alpha = -0.1, \mu = 0, \sigma = 0.1) )</td>
</tr>
<tr>
<td>D4</td>
<td>( 0.05 + 0.2s_1 s_2 \Phi^{-1}(\tau) )</td>
<td>( 0.3e^{s_2} + 0.2\tau )</td>
<td>( Q_{\text{Pareto}}(\tau, \alpha = 0.4s_1, \mu = 0.3, \sigma = 0.4) )</td>
</tr>
</tbody>
</table>

model allows for spatially varying coefficients and spatial correlation while the NCQR method assumes independent and identically distributed samples.

We index the quantile levels at which the methods are compared by \( j \in 1, ..., J \). For each quantile level, \( \tau_j \), and simulated dataset replicate, \( b \in \{1, ..., B\} \), the estimated coefficients \( \hat{\beta}_p(\tau_j, s_i) \), were compared using root mean integrated square error (RMISE). The RMISE for simulated dataset \( b \) was calculated for a given \( \beta_p \) and sequence \( \tau_1, ..., \tau_J \):

\[
RMISE(\beta_p)(b) = \sqrt{\frac{1}{S} \sum_{i=1}^{S} \sum_{j=1}^{J} \delta_j \left[ \hat{\beta}_p(\tau_j, s_i)(b) - \beta_p(\tau_j, s_i) \right]^2}
\]  

(2.12)

where \( \delta_j = \tau_j - \tau_{j-1} \). The means and standard errors of the RMISEs as well as the coverage of the 95% confidence (NCQR) or credible (IQR and GAUS) intervals were then calculated for each method and design (Table 2.2).

Both IQR and the GAUS method produce density estimates. The NCQR method does not estimate the entire quantile function and therefore can not be used to create a density estimate without substantial additional calculation. To evaluate the predictive densities we use the log score, which is the logarithm of the predicted density evaluated at the training and validation data. This is a strictly proper scoring rule [Gneiting and Raftery, 2007]. We calculate the log score for each observation as the log of the posterior mean of the predictive density evaluated at the observation. The total log score for each dataset is calculated as the mean of the log scores for the individual observations. The mean and standard error by simulation design are calculated using the total log score values of the 50 simulated datasets.

We compare all three methods using \( \tau = \{0.05, 0.06, ..., 0.94, 0.95\} \). Four non-constant basis functions per predictor were used in both the IQR and GAUS methods. The results
Table 2.2 Comparison of fitted $\beta(\tau)$ functions $\tau = \{.05, .06, ..., .94, .95\}$. COV represents the coverage of the 95% credible (IQR and GAUS) or confidence interval (NCQR).

<table>
<thead>
<tr>
<th></th>
<th>$\beta_0$</th>
<th></th>
<th>$\beta_1$</th>
<th></th>
<th>$\beta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMISE</td>
<td>SE</td>
<td>COV</td>
<td>RMISE</td>
<td>SE</td>
</tr>
<tr>
<td>D1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IQR</td>
<td>0.014</td>
<td>0.001</td>
<td>0.92</td>
<td>0.027</td>
<td>0.001</td>
</tr>
<tr>
<td>GAUS</td>
<td>0.016</td>
<td>0.001</td>
<td>0.92</td>
<td>0.022</td>
<td>0.002</td>
</tr>
<tr>
<td>NCQR</td>
<td>0.017</td>
<td>0.001</td>
<td>0.96</td>
<td>0.025</td>
<td>0.001</td>
</tr>
<tr>
<td>D2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IQR</td>
<td>0.019</td>
<td>0.001</td>
<td>0.93</td>
<td>0.035</td>
<td>0.002</td>
</tr>
<tr>
<td>GAUS</td>
<td>0.038</td>
<td>0.005</td>
<td>0.83</td>
<td>0.065</td>
<td>0.009</td>
</tr>
<tr>
<td>NCQR</td>
<td>0.025</td>
<td>0.001</td>
<td>0.98</td>
<td>0.045</td>
<td>0.002</td>
</tr>
<tr>
<td>D3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IQR</td>
<td>0.029</td>
<td>0.001</td>
<td>0.95</td>
<td>0.050</td>
<td>0.002</td>
</tr>
<tr>
<td>GAUS</td>
<td>0.027</td>
<td>0.001</td>
<td>0.97</td>
<td>0.046</td>
<td>0.001</td>
</tr>
<tr>
<td>NCQR</td>
<td>0.050</td>
<td>0.000</td>
<td>0.64</td>
<td>0.201</td>
<td>0.001</td>
</tr>
<tr>
<td>D4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IQR</td>
<td>0.038</td>
<td>0.001</td>
<td>0.94</td>
<td>0.062</td>
<td>0.002</td>
</tr>
<tr>
<td>GAUS</td>
<td>0.094</td>
<td>0.047</td>
<td>0.94</td>
<td>0.104</td>
<td>0.034</td>
</tr>
<tr>
<td>NCQR</td>
<td>0.054</td>
<td>0.001</td>
<td>0.75</td>
<td>0.197</td>
<td>0.001</td>
</tr>
</tbody>
</table>

given in Table 2.2 demonstrate that while the 3 methods perform similarly for D1 (independent, light tails), the IQR method performs substantially better than GAUS in the heavy-tailed designs (D2 and D4) and substantially better than NCQR in the spatially varying designs (D3 and D4). Compared to the nominal coverage rate of 0.95, the IQR method has good coverage for all of the designs, with the lowest coverage being 0.88 for $\beta_1$ in D1. GAUS had poor coverage for D2, while NCQR had poor coverage for D3 and D4.

Unlike the NCQR method, both our method and the GAUS method assume parametric forms for the tails and so can be used to estimate parameter effects on extreme quantiles. We compare the parameter estimates for these two methods evaluated at $\tau = \{0.950, 0.951, ..., 0.994, 0.995\}$ in Table 2.3. Our method performs better in all cases except D1 $\beta_1$, which is a linear function of $\tau$.

The results of the log-score comparisons are consistent with the parameter estimates (Table 2.4). However, the GAUS method consistently produces higher log-scores in-sample than the IQR method. Because the likelihood is not constrained to be continuous
**Table 2.3** Comparison of fitted $\beta(\tau)$ functions $\tau = (.950, .951, ..., .994, .995)$. COV represents the coverage of the 95% credible (IQR and GAUS) or confidence interval (NCQR).

<table>
<thead>
<tr>
<th></th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMISE</td>
<td>SE</td>
<td>COV</td>
</tr>
<tr>
<td>D1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IQR</td>
<td>0.0047</td>
<td>0.0004</td>
<td>0.96</td>
</tr>
<tr>
<td>GAUS</td>
<td>0.0051</td>
<td>0.0005</td>
<td>0.98</td>
</tr>
<tr>
<td>D2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IQR</td>
<td>0.0139</td>
<td>0.0014</td>
<td>0.95</td>
</tr>
<tr>
<td>GAUS</td>
<td>0.0266</td>
<td>0.0054</td>
<td>0.78</td>
</tr>
<tr>
<td>D3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IQR</td>
<td>0.0094</td>
<td>0.0003</td>
<td>0.97</td>
</tr>
<tr>
<td>GAUS</td>
<td>0.0119</td>
<td>0.0004</td>
<td>0.95</td>
</tr>
<tr>
<td>D4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IQR</td>
<td>0.0196</td>
<td>0.0012</td>
<td>0.96</td>
</tr>
<tr>
<td>GAUS</td>
<td>0.0802</td>
<td>0.0476</td>
<td>0.94</td>
</tr>
</tbody>
</table>

In the GAUS method, very large likelihood values can be obtained for the in-sample observations (Fig. 2.2). In the heavy-tailed designs the IQR method results in higher out-of-sample log-scores.

**Table 2.4** Comparison of mean estimated log scores

<table>
<thead>
<tr>
<th></th>
<th>In-sample Mean</th>
<th>SE</th>
<th>Out-of-sample Mean</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IQR</td>
<td>0.339</td>
<td>0.003</td>
<td>0.315</td>
<td>0.008</td>
</tr>
<tr>
<td>GAUS</td>
<td>0.356</td>
<td>0.003</td>
<td>0.322</td>
<td>0.008</td>
</tr>
<tr>
<td>D2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IQR</td>
<td>−0.223</td>
<td>0.004</td>
<td>−0.254</td>
<td>0.017</td>
</tr>
<tr>
<td>GAUS</td>
<td>−0.219</td>
<td>0.005</td>
<td>−0.288</td>
<td>0.022</td>
</tr>
<tr>
<td>D3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IQR</td>
<td>0.476</td>
<td>0.003</td>
<td>0.418</td>
<td>0.010</td>
</tr>
<tr>
<td>GAUS</td>
<td>0.536</td>
<td>0.003</td>
<td>0.419</td>
<td>0.009</td>
</tr>
<tr>
<td>D4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IQR</td>
<td>−0.191</td>
<td>0.004</td>
<td>−0.238</td>
<td>0.012</td>
</tr>
<tr>
<td>GAUS</td>
<td>−0.126</td>
<td>0.006</td>
<td>−0.287</td>
<td>0.022</td>
</tr>
</tbody>
</table>
2.5 Application

2.5.1 Data

An amendment to the U.S. National Emission Standards for Hazardous Air Pollutants for petroleum refineries requires the use of two-week time-integrated passive samplers at specified intervals around the facility fence line to establish levels of benzene in the air [EPA, 2014]. The utility of fence line measurements as a method of controlling emissions is contingent on their distributions being dependent on nearby sources within the facility. To evaluate the efficacy of passive samplers in monitoring benzene emissions from petroleum refineries, researchers from US EPA Office of Research and Development conducted a year-long field study in collaboration with Flint Hills Resources in Corpus Christi, TX [Thoma et al., 2011]. Preliminary analyses found that under consistent wind conditions, downwind concentrations were statistically higher than upwind concentrations [Thoma et al., 2011]. More sophisticated modeling should be able to shed light on the contributions of individual sources to the concentrations observed at the fence line. Modeling these concentrations requires an extra level of complexity because near-source air pollutant measurements typically exhibit strong spatial correlation along with non-stationary and non-Gaussian distributions even after transformation. Both the spatial covariance and the distribution of the pollutant concentrations can vary as a function of wind and emission source location. Accurately modeling the entire distribution and spatial structure of the pollutant concentrations should improve inference concerning the strengths of known sources. Additionally, due to the stochastic nature of dispersion and variation in background pollutant concentration levels, the effect of a specific source on the pollutant distribution may not be detected through mean regression. Of particular concern both for exposure and compliance evaluation are the source effects on the upper tail of the distribution, in particular the 95th percentile.

The measurements used in this study were collected between Dec 3, 2008 and Dec 2, 2009 around the Flint Hills West Refinery [Thoma et al., 2011]. The samplers were attached to the boundary fence around the facility approximately 1.5 m above the ground at 15 locations (Fig. 2.3). In addition, one sampler (633) was deployed at a nearby Texas Commission on Environmental Quality (TCEQ) continuous air monitoring station (CAMS). A total of 406 two-week time-integrated benzene concentration measurements collected over the course of the year were used in the analysis. Hourly temperature, wind
speed and direction were also measured at TCEQ CAMS 633.

The concentrations exhibited both spatial and temporal trends (Fig. 2.3). In particular, the variance increased dramatically during the summer months. The highest concentrations were observed on the northern edge of the refinery (sites 360, 20, and 50) while the lowest concentrations were observed on the southern edge (sites 250, 633, and 270). The increase in variance can partly be explained by meteorology (Fig. 2.4). During the summer the winds consistently blow from the southeast, while during the rest of the year they are more evenly distributed.

![Map of benzene measurements](image)

**Figure 2.3** Benzene measurements by time and location. Source locations, e_1 and e_2, are shown in black. Points have been jittered slightly to improve visibility.

A visual analysis of the concentrations and wind roses for the hourly measurements at each time period suggested that the concentrations were correlated with a source within the refinery. Two probable emission source locations e_1 and e_2 were selected using the reported emission inventory. To determine the effect of the emission sources on the distribution of the benzene concentration, we denote the t^{th} observed value of the benzene concentration at site s_i as y_i(s_t) where i = 1, ..., 16 and t = 1, ..., 26 and will model the quantile function of Y using equation 2.1 and 2.2. Our full model includes an intercept and three predictors: transport from source 1, transport from source 2, and temperature.

The predictors that represent transport from a source are calculated from the observed hourly wind vectors and relative spatial locations of the source and measurement. The
The $t^{th}$ observed value of the transport from source 1 to location $s_i$ is defined as

$$x_{1,t}(s_i) = \sum_{h=1}^{336} \left\{ \max \left( \frac{\mathbf{w}_{t,h} \cdot (\mathbf{e}_1 - s_i)}{||\mathbf{e}_1 - s_i||}, 0 \right) \right\}$$

(2.13)

where $\mathbf{e}_1$ is the location of emission source 1, and $s_i$ is the measurement location. Each hourly wind vector, $\mathbf{w}_{t,h}$, for the two-week period with $h = 1, \ldots, 336$ was transformed into the same coordinate system and projected onto the vector from the source to the measurement ($\mathbf{e}_1 - s_i$). Assuming a constant emission source, the resulting scalar quantity represents the amount of pollutant transported from $\mathbf{e}_1$ to $s_i$, ignoring the effects of vertical dispersion. When the wind is blowing from $s_i$ toward $\mathbf{e}_1$, transport from $\mathbf{e}_1$ will be negative. However, due to finite, small background concentrations, the integrated benzene concentration will remain the same rather than decreasing under these conditions. Therefore the maximum of the transport from $\mathbf{e}_1$ and zero was taken before taking the sum over $h$ in period $t$ (2.13). The transport from source 2 was calculated similarly.

We use 10-fold cross-validation to determine the most appropriate model for the benzene concentrations. Using each fold as a validation data set, the in-sample and out-of-sample log-score was calculated using both the proposed IQR method and the GAUS method proposed by Reich [2012] for each combination of predictors (Table 5). An expo-
Table 2.5 Estimated log-scores for training and validation data by method.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>In-sample</th>
<th></th>
<th>Out-of-sample</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IQR</td>
<td>GAUS</td>
<td>IQR</td>
<td>GAUS</td>
</tr>
<tr>
<td></td>
<td>Mean SE</td>
<td>Mean SE</td>
<td>Mean SE</td>
<td>Mean SE</td>
</tr>
<tr>
<td>None</td>
<td>−0.29 0.01</td>
<td>−0.16 0.01</td>
<td>−0.42 0.07</td>
<td>−0.45 0.09</td>
</tr>
<tr>
<td>Source 1</td>
<td>0.04 0.01</td>
<td>0.49 0.01</td>
<td>−0.13 0.07</td>
<td>−0.18 0.07</td>
</tr>
<tr>
<td>Source 2</td>
<td>0.04 0.01</td>
<td>0.47 0.01</td>
<td>−0.14 0.08</td>
<td>−0.21 0.08</td>
</tr>
<tr>
<td>Temperature</td>
<td>0.10 0.01</td>
<td>0.57 0.02</td>
<td>−0.07 0.08</td>
<td>−0.25 0.08</td>
</tr>
<tr>
<td>Source 1 + Source 2</td>
<td>0.21 0.01</td>
<td>0.81 0.02</td>
<td>−0.02 0.08</td>
<td>−0.22 0.06</td>
</tr>
<tr>
<td>Source 1 + Temp</td>
<td>0.30 0.01</td>
<td>1.00 0.03</td>
<td>0.08 0.08</td>
<td>−0.19 0.09</td>
</tr>
<tr>
<td>Source 2 + Temp</td>
<td>0.27 0.01</td>
<td>0.88 0.02</td>
<td>0.06 0.08</td>
<td>−0.24 0.09</td>
</tr>
<tr>
<td>All</td>
<td>0.39 0.01</td>
<td>0.95 0.04</td>
<td>0.13 0.09</td>
<td>−0.39 0.08</td>
</tr>
</tbody>
</table>

The model was fit to the entire dataset to determine the effects of the sources and temperature on the distribution of benzene at the fence line. We plot the coefficients by quantile level and location in Fig. 2.5. We can see that the base distribution does not vary as much by location as the effects of the sources and temperature. The effects of the sources on the quantiles of the concentrations range from positive to negative, with the majority of the source effects being positive. The negative effects could be due to the fact that these sources may not have been constant over the course of the entire year. If wind from a given source corresponded to time points when the source was not emitting it could result in a negative effect on the concentrations. As can be seen in Fig. 2.6, the
effect of source 1 on the 95th quantile is large and positive for the sites on the northern edge of the refinery and some sites along the southern edge of the refinery. The northern sites were also the locations where the highest concentrations were observed. The effect of source 2 on the 95th quantile was smaller overall and varied by site with positive effects observed on the background site and sites on the northern edge of the refinery (Fig. 2.6). Temperature also had a strong positive effect on concentrations on the northern edge of the refinery indicating the possibility of another emission source during the summer near the northern edge of the refinery that was not accounted for.

Figure 2.5 Estimated predictor effect by quantile and location.

2.6 Discussion

We have derived the properties and demonstrated the utility of a method for spatial quantile regression that allows for spatially-varying coefficients and flexible tail distributions. By modeling the entire quantile function we exploit the flexibility of non-parametric basis functions in the center of the distribution and the constraints of parametric tails in the areas of the distribution where data is sparse. We have shown the conditions under which the model guarantees a smooth density function with the desired degrees of differentiability and enables the estimation of a non-stationary covariance that is dependent on the
predictors. Through both simulations and an application to fence line benzene concentrations we have demonstrated the utility of ensuring a smooth density function with parametric tails and the flexibility and accuracy of the method compared to previous work.

While the model doesn’t currently account for temporal correlation in the response variable, a non-linear function of time could easily be incorporated as a predictor using the current framework. Additionally, temporal correlation could be accounted for by adjusting the priors of the coefficients or incorporating a copula. A multivariate extension for modeling multiple pollutants simultaneously could also be developed through the use of multivariate spatial priors.

Supplementary Materials

Proofs of Proposition 1 and Theorem 1 as well as computing details are contained in Appendix A.

Disclaimer The views expressed in this publication are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency. Acknowledgements The authors would like to thank Oak Ridge Institute for Science and Education for the fellowship funding that supported this work.
3.1 Introduction

In the last decade, low cost and portable air quality sensors have enjoyed dramatically increased usage. These sensors can provide an un-calibrated measure of a variety of pollutants in near real time, but deriving meaningful information from sensor data remains a challenge [Snyder et al., 2013]. The "SPod" is a low-cost sensor currently being investigated by researchers at the U.S. Environmental Protection Agency to detect volatile organic compound (VOC) emissions from industrial facilities [Thoma et al., 2016]. Due to changes in temperature and relative humidity the output signal exhibits a slowly varying baseline drift on the order of minutes to hours. Figure 3.1 provides an example of measurements from three SPod sensors co-located at the border of an industrial facility.
Figure 3.1 Example of 3 co-located SPod PID sensor readings over a 24 hour period.

All of the sensors respond to the pollutant signal, which is illustrated by the three sharp transient spikes at 11:32, 14:10, and 16:03. However, the baseline drift varies from one sensor to another, obscuring the detection of the peaks that alert the intrusion of pollutants. We show later that by estimating the baseline drift in each sensor and removing it from the observed signals, peaks can be reliably detected from concordant residual signals from a collection of SPods using a simple data-driven thresholding strategy. Thus, accurately demixing a noisy observed time series into a slowly varying component and a transient component can lead to greatly improved and simplified downstream analysis.

While this work is motivated by the analysis of data from low cost air quality sensors, the problem of demixing noisy time series into trends and transients is ubiquitous across many fields of study. In a wide range of applications that spans chemistry [Ning et al., 2014], macroeconomics [Yamada, 2017], environmental science [Brantley et al., 2014], and medical sciences [Pettersson et al., 2013, Marandi and Sabzpoushan, 2015], scalar functions of time $y(t)$ are observed and assumed to be a superposition of an underlying slowly varying baseline trend $\theta(t)$, other more rapidly varying components $s(t)$, and noise. In practice, $y(t)$ is observed at discrete time points $t_1, \ldots, t_n$, and we model the vector of samples $y_i = y(t_i)$ as

$$y = \theta + s + \varepsilon,$$

where $\theta_i = \theta(t_i)$, $s_i = s(t_i)$, and $\varepsilon \in \mathbb{R}^n$ is a vector of uncorrelated noise. For notational simplicity, for the rest of the paper, we assume that the time points take on the values
\( t_i = i \), but it is straightforward to generalize to an arbitrary grid of time points.

In some applications, the slowly varying component \( \theta \) is the signal of interest, and the transient component \( s \) is a vector of nuisance parameters. In our air quality application, the roles of \( \theta \) and \( s \) are reversed; \( s \) represents the signal of interest and \( \theta \) represents a baseline drift that obscures the identification of the important transient events encoded in \( s \).

To tackle demixing problems, we introduce a scalable baseline estimation framework by building on \( \ell_1 \)-trend filtering, a relatively new nonparametric estimation framework. Our contributions are three-fold.

- Kim et al. [2009] proposed using the check function as a possible extension of \( \ell_1 \)-trend filtering but did not investigate it further. Here, we develop the basic \( \ell_1 \)-quantile-trend-filtering framework and extend it to model multiple quantiles simultaneously with non-crossing constraints to ensure validity and improve trend estimates.

- To reduce computation time and extend the method to long time series, we develop a parallelizable ADMM algorithm. The algorithm proceeds by splitting the time domain into overlapping windows, fitting the model separately for each of the windows and reconciling estimates from the overlapping intervals.

- Finally, we propose a modified criterion for performing model selection.

In the rest of the paper, we detail our quantile trend filtering algorithms (Section 3.2) as well as how to choose the smoothing parameter (Section 3.3). We demonstrate through simulation studies that our proposed model provides better or comparable estimates of non-parametric quantile trends than existing methods (Section 3.4). We further show that quantile trend filtering is a more effective method of drift removal for low-cost air quality sensors and results in improved signal classification compared to quantile smoothing splines (Section 3.5). Finally, we discuss potential extensions of quantile trend filtering (Section 3.6).
3.2 Baseline Trend Estimation

3.2.1 Background

Kim et al. [2009] originally proposed $\ell_1$-trend filtering to estimate trends with piecewise polynomial functions, assuming that the observed time series $y$ consists of a trend $\theta$ plus uncorrelated noise $\varepsilon$, namely $y = \theta + \varepsilon$. The estimated trend is the solution to the following convex optimization problem

$$\min_{\theta} \frac{1}{2} \|y - \theta\|_2^2 + \lambda \|D^{(k+1)}\theta\|_1,$$

where $\lambda$ is a nonnegative regularization parameter, and the matrix $D^{(k+1)} \in \mathbb{R}^{(n-k-1) \times n}$ is the discrete difference operator of order $k+1$. To understand the purpose of penalizing the 1-norm of $D^{(k+1)}\theta$ consider the difference operator when $k = 0$.

$$D^{(1)} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}.$$  

Thus, $\|D^{(1)}\theta\|_1 = \sum_{i=1}^{n-1} |\theta_i - \theta_{i+1}|$, which is known as the total variation denoising penalty in one dimension in the signal processing literature [Rudin et al., 1992] or the fused lasso penalty in the statistics literature [Tibshirani et al., 2005]. The penalty term incentivizes solutions which are piecewise constant. For $k \geq 1$, the difference operator $D^{(k+1)} \in \mathbb{R}^{(n-k-1) \times n}$ is defined recursively as follows

$$D^{(k+1)} = D^{(1)}D^{(k)}.$$  

Penalizing the 1-norm of the vector $D^{(k+1)}\theta$ produces estimates of $\theta$ that are piecewise polynomials of order $k$.

Tibshirani [2014] proved that with a judicious choice of $\lambda$ the trend filtering estimate converges to the true underlying function at the minimax rate for functions whose $k$th derivative is of bounded variation and showed that trend filtering is locally adaptive
Figure 3.2 Examples of trend filtering solutions (red) and 15th quantile trending filtering solution (blue). Standard trend filtering performs well in the no-signal case (a) but struggles to distinguish between the slowly varying trend and the rapidly-varying signal (b). The quantile trend is not affected by the signal and provides an estimate of the baseline.

When the time series consists of only the trend and random noise, which is illustrated in Figure 4.3a. As noted earlier, in some applications, such as the air quality monitoring problem considered in this paper, the data contain a rapidly varying signal in addition to the slowly varying trend and noise. Figure 4.3b shows that standard trend filtering is not designed to distinguish between the slowly varying trend and the rapidly-varying signal, as the smooth component estimate $\theta$ is biased towards the peaks of the transient components.

To account for the presence of transient components in the observed time series $y$, we propose quantile trend filtering Figure 4.3b. To estimate the trend in the $\tau$th quantile, we solve the convex optimization problem

$$
\min_{\theta} \rho_{\tau}(y - \theta) + \lambda \|D^{(k+1)}\theta\|_1,
$$

(3.1)

where $\rho_{\tau}(r)$ is the check function

$$
\rho_{\tau}(r) = \sum_{i=1}^{n} r_i(\tau - 1(r_i < 0)),
$$

(3.2)

and $1(A)$ is 1 if its input $A$ is true and 0 otherwise. Note that we do not explicitly model $s$. Rather, we focus on estimating $\theta$. We then estimate $s + \varepsilon$ as the difference $y - \theta$. 

27
Before elaborating on how we compute our proposed $\ell_1$-quantile trend filtering estimator, we discuss similarities and differences between our proposed estimator and existing quantile trend estimators.

**Relationship to Prior Work**

In this application, as well as those described in Ning et al. [2014], Marandi and Sabzpanesh [2015], and Pettersson et al. [2013], the goal is to estimate the trend in the baseline not the mean. We can define the trend in the baseline as the trend in a low quantile of the data. A variety of methods for estimating quantile trends have already been proposed. Koenker and Bassett [1978] were the first to propose substituting the sum-of-squares term with the check function (3.2) to estimate a conditional quantile instead of the conditional mean. Later, Koenker et al. [1994] proposed quantile trend filtering with $k = 1$ producing quantile trends that are piecewise linear, but they did not consider extensions to higher order differences. Rather than using the $\ell_1$-norm to penalize the discrete differences, Nychka et al. [1995] used the smoothing spline penalty based on the square of the $\ell_2$-norm:

$$\sum_{i=1}^{n} \rho_\tau(y(t_i) - \theta(t_i)) + \lambda \int \theta''(t)^2 dt,$$

where $\theta(t)$ is a smooth function of time and $\lambda$ is a tuning parameter that controls the degree of smoothing. Oh et al. [2011] proposed an algorithm for solving the quantile smoothing spline problem by approximating the check function with a differentiable function. Racine and Li [2017] propose a method for estimating quantile trends that does employ the check function. In their method, the response is constrained to follow a location scale model and the conditional quantiles are estimated by combining Gaussian quantile functions with a kernel smoother and solving a local-linear least squares problem.

### 3.2.2 Quantile Trend Filtering

We combine the ideas of quantile regression and trend filtering. For a single quantile level $\tau$, the quantile trend filtering problem is given in (3.1). As with classic quantile regression, the quantile trend filtering problem is a linear program which can be solved by a number of methods. We want to estimate multiple quantiles simultaneously and to
ensure that our quantile estimates are valid by enforcing the constraint that if \( \tau_2 > \tau_1 \) then \( Q(\tau_2) \geq Q(\tau_1) \) where \( Q \) is the quantile function of \( y \). Even if a single quantile is ultimately desired, ensuring non-crossing allows information from nearby quantiles to be used to improve the estimates as we will see in the peak detection experiments in Section 3.4.2. Given quantiles \( \tau_1 < \tau_2 < \ldots < \tau_J \), the optimization problem becomes

\[
\min_{\Theta \in C} \sum_{j=1}^{J} \left[ \rho_{\tau_j} (y - \theta_j) + \lambda_j \| D^{(k+1)} \theta_j \|_1 \right] 
\]  

(3.3)

where \( \Theta = \begin{pmatrix} \theta_1 & \theta_2 & \ldots & \theta_J \end{pmatrix} \in \mathbb{R}^{n \times J} \) is a matrix whose \( j \)th column corresponds to the \( j \)th quantile signal \( \theta_j \) and the set \( C = \{ \Theta \in \mathbb{R}^{n \times J} : \theta_{ij} \leq \theta_{ij'} \text{ for } j \leq j' \} \) encodes the non-crossing quantile constraints. The additional non-crossing constraints are linear inequalities involving the parameters, so the non-crossing quantile trends can still be estimated by a number of available linear programming solvers. We allow for the possibility that the degree of smoothness in the trends varies by quantile by allowing the smoothing parameter to vary with quantile as well. In the rest of this paper, we use \( k = 2 \) to produce piecewise quadratic polynomials and report numerical results using the commercial solver Gurobi [Gurobi Optimization, 2018] and its R package implementation. However, we could easily substitute a free solver such as the Rglpk package by Theussl and Hornik [2017].

3.2.3 ADMM for Big Data

As the size of the data increases, computation time becomes prohibitive. In our application to air quality sensor data, measurements are recorded every second resulting in 86,400 observations per day. This number of observations is already too large to use with currently available R packages used for estimating quantile trends [Douglas Nychka et al., 2017, Koenker, 2016]. To our knowledge, no one has addressed the problem of finding smooth quantile trends of series that are too large to be processed simultaneously. We propose a divide-and-conquer approach via an ADMM algorithm for solving large problems in a piecewise fashion.
Figure 3.3 Window boundaries and trends fit separately in each window. Each window’s trend estimate is plotted in a different line type.

**Formulation**

To decrease computation time and extend our method to larger problems, we divide our observed series $y_i$ with $i = 1, \ldots, n$ into $W$ overlapping windows of observations, defining the vector of sequential elements indexed from $l_w$ to $u_w$ as $y^{(w)} = \{y_{l_w}, y_{l_w+1}, \ldots, y_{u_w-1}, y_{u_w}\}$, with

$$1 = l_1 < l_2 < u_1 < l_3 < u_2 < l_4 < u_3 < \cdots < u_W = n.$$

We define $n_w = u_w - l_w + 1$ so that $y^{(w)} \in \mathbb{R}^{n_w}$. Figure 3.3 shows an example of 1200 observations being mapped into three equally sized overlapping windows of observations. While the overlapping trend estimates between $l_2$ and $u_2$ do not vary dramatically, the difference is more pronounced in the trend in the 5th quantile between $l_3$ and $u_2$. Thus, we need a way of enforcing estimates to be identical in the overlapping regions.

Given quantiles $\tau_1 < \cdots < \tau_J$, we introduce dummy variables $\theta_j^{(w)} \in \mathbb{R}^{n_w}$ as the value of the $\tau_j$th quantile trend in window $w$. We then "stitch" together the $W$ quantile trend estimates into consensus over the overlapping regions by introducing the constraint $\theta_{ij}^{(w)} = \theta_{i-l_w+1,j}$ for $i = 1, \ldots, n_w$ and for all $j$. Let $\Theta^{(w)}$ be the matrix whose $j$th column is $\theta_j^{(w)}$. Then we can write these constraints more concisely as $\Theta^{(w)} = U^{(w)}\Theta$, where $U^{(w)} \in \{0, 1\}^{n_w \times n}$ is a matrix that selects rows of $\Theta$ corresponding to the $w$th window,
Figure 3.4 Trend fit with our ADMM algorithm with 3 windows (converged in 7 iterations), compared to trend from simultaneous fit.

\[ \mathbf{U}^{(w)} = \begin{pmatrix} \mathbf{e}_{1}^T \\ \vdots \\ \mathbf{e}_{W}^T \end{pmatrix}, \]

where \( \mathbf{e}_{i} \in \mathbb{R}^{n} \) denotes the \( i \)th standard basis vector. Furthermore, let \( \iota_{C} \) denote the indicator function of the non-crossing quantile constraint, namely \( \iota_{C}(\Theta) \) is zero if \( \Theta \in C \) and infinity otherwise. Our windowed quantile trend optimization problem can then be written as

\[
\begin{aligned}
\minimize_{w=1}^{W} \sum_{w=1}^{W} \left\{ \sum_{j=1}^{J} \left[ \rho_{T_j} (\mathbf{y}^{(w)} - \mathbf{\theta}_{j}^{(w)}) + \lambda_{j} \| \mathbf{D}^{(k+1)} \mathbf{\theta}_{j}^{(w)} \|_1 \right] + \iota_{C}(\mathbf{\Theta}^{(w)}) \right\} \\
\text{subject to} \quad \mathbf{\Theta}^{(w)} = \mathbf{U}^{(w)} \mathbf{\Theta} \quad \text{for} \quad w = 1, \ldots, W.
\end{aligned}
\] (3.4)

The solution to (3.4) is not identical to the solution to (3.3) because of double counting of the overlapping sections. The solutions are very close, however, and the differences are essentially immaterial concerning downstream analysis. Figure 3.4 provides an illustration of the trends estimated using multiple windows compared with the trends estimated using a single window; estimates using multiple and single windows are nearly indistinguishable.
Algorithm

The ADMM algorithm [Gabay and Mercier, 1975, Glowinski and Marroco, 1975] is described in greater detail by Boyd et al. [2011], but we briefly review how it can be used to iteratively solve the following equality constrained optimization problem which is a more general form of (3.4).

\[
\begin{align*}
\text{minimize} \quad & f(\phi) + g(\tilde{\phi}) \\
\text{subject to} \quad & A\phi + B\tilde{\phi} = c.
\end{align*}
\] (3.5)

Recall that finding the minimizer to an equality constrained optimization problem is equivalent to the identifying the saddle point of the Lagrangian function associated with the problem (3.5). ADMM seeks the saddle point of a related function called the augmented Lagrangian,

\[
L_{\gamma}(\phi, \tilde{\phi}, \omega) = f(\phi) + g(\tilde{\phi}) + \langle \omega, c - A\phi - B\tilde{\phi} \rangle + \frac{\gamma}{2} ||c - A\phi - B\tilde{\phi}||^2,
\]

where the dual variable $\omega$ is a vector of Lagrange multipliers and $\gamma$ is a nonnegative tuning parameter. When $\gamma = 0$, the augmented Lagrangian coincides with the ordinary Lagrangian.

ADMM minimizes the augmented Lagrangian one block of variables at a time before updating the dual variable $\omega$. This yields the following sequence of updates at the $(m+1)^{th}$ ADMM iteration

\[
\begin{align*}
\phi_{m+1} &= \arg\min_{\phi} L_{\gamma}(\phi, \phi_m, \omega_m) \\
\tilde{\phi}_{m+1} &= \arg\min_{\tilde{\phi}} L_{\gamma}(\phi_{m+1}, \tilde{\phi}, \omega_m) \\
\omega_{m+1} &= \omega_m + \gamma(c - A\phi_{m+1} - B\tilde{\phi}_{m+1}).
\end{align*}
\] (3.6)

Returning to our constrained windows problem giving in (3.4), let $\Omega^{(w)}$ denote the Lagrange multiplier matrix for the $w$th consensus constraint, namely $\Theta^{(w)} = U^{(w)}\Theta$, and let $\omega_j^{(w)}$ denote its $j$th column.
The augmented Lagrangian is given by

\[ L(\Theta, \{\Theta^{(w)}\}, \{\Omega^{(w)}\}) = \sum_{w=1}^{W} L_w(\Theta, \Theta^{(w)}, \Omega^{(w)}), \]

where

\[ L_w(\Theta, \Theta^{(w)}, \Omega^{(w)}) = \sum_{j=1}^{J} \left[ \rho_j (y^{(w)}_i - \theta^{(w)}_j) + \lambda_j \|D^{(k+1)}\theta^{(w)}_j\|_1 \right. \]

\[ + (\theta^{(w)}_j - U^{(w)} \theta^{(w)})^T \omega^{(w)}_j + \frac{\gamma}{2} \|\theta^{(w)}_j - U^{(w)} \theta^{(w)}\|_2^2 \left. \right] + \iota_C(\Theta^{(w)}), \]

where \( \gamma \) is a positive tuning parameter.

The ADMM algorithm alternates between updating the consensus variable \( \Theta \), the window variables \( \{\Theta^{(w)}\} \), and the Lagrange multipliers \( \{\Omega^{(w)}\} \). At the \((m+1)\)th iteration, we perform the following sequence of updates

\[
\Theta_{m+1} = \arg\min_{\Theta} L(\Theta, \{\Theta^{(w)}_m\}, \{\Omega^{(w)}_m\})
\]

\[
\Theta^{(w)}_{m+1} = \arg\min_{\{\Theta^{(w)}\}} L(\Theta_{m+1}, \{\Theta^{(w)}\}, \{\Omega^{(w)}_m\})
\]

**Updating \( \Theta \):** Some algebra shows that, defining \( i_w = i - l_w + 1 \), updating the consensus variable step is computed as follows.

\[
\theta_{ij} = \begin{cases} 
\frac{1}{2} \left( \theta^{(w-1)}_{i_{w-1}j} + \theta^{(w)}_{i_{w}j} \right) - \frac{1}{2\gamma} \left( \omega^{(w-1)}_{i_{w-1}j} + \omega^{(w)}_{i_{w}j} \right) & \text{if } l_w \leq i \leq u_{w-1}, \\
\theta^{(w)}_{i_{w}j} & \text{if } u_{w-1} < i \leq l_{w+1}, \\
\frac{1}{2} \left( \theta^{(w)}_{i_{w}j} + \theta^{(w+1)}_{i_{w+1}j} \right) - \frac{1}{2\gamma} \left( \omega^{(w)}_{i_{w}j} + \omega^{(w+1)}_{i_{w+1}j} \right) & \text{if } l_{w+1} < i \leq u_w.
\end{cases}
\]  

(3.7)

The consensus update (3.7) is rather intuitive. We essentially average the trend estimates in overlapping sections of the windows, subject to some adjustment by the Lagrange multipliers, and leave the trend estimates in non-overlapping sections of the windows untouched. For notational ease, we write the consensus update (3.7) compactly as
Algorithm 1 ADMM algorithm for quantile trend filtering with windows

Define $D = D^{(k+1)}$.

**initialize:**

for $w = 1, \ldots, W$ do
  $\Theta_0^{(w)} \leftarrow \arg \min_{\Theta^{(w)} \in C} \sum_{j=1}^J \rho_{\tau_j}(y^{(w)} - \theta_j^{(w)}) + \lambda \|D\theta_j^{(w)}\|_1$

  $\Omega_0^{(w)} \leftarrow 0$

end for

$m \leftarrow 0$

repeat

 $\Theta_{m+1} \leftarrow \psi(\{\Theta_m^{(w)}\}, \{\Omega_m^{(w)}\})$

for $w = 1, \ldots, W$ do

  $\Theta_m^{(w)} \leftarrow \arg \min_{\Theta^{(w)}} \mathcal{L}_w(\Theta_{m+1}^{(w)}, \Theta^{(w)}, \Omega_m^{(w)})$

  $\Omega_m^{(w)} \leftarrow \Omega_m^{(w)} + \gamma(\Theta_{m+1}^{(w)} - U^{(w)}\Theta_m^{(w)})$

end for

$m \leftarrow m + 1$

until convergence

return $\Theta_m$

$\Theta = \psi(\{\Theta^{(w)}\}, \{\Omega^{(w)}\})$.

**Updating $\{\Theta^{(w)}\}$:** We then estimate the trend separately in each window, which can be done in parallel, while penalizing the differences in the overlapping pieces of the trends as outlined in Algorithm 1. The use of the Augmented Lagrangian converts the problem of solving a potentially large linear program into a solving a collection of smaller quadratic programs. The gurobi R package [Gurobi Optimization, 2018] can solve quadratic programs in addition to linear programs, but we can also use the free R package quadprog [Weingessel and Turlach, 2013].

Algorithm 1 has the following convergence guarantees.

**Proposition 2.** Let $\{(\Theta_m^{(w)}), \Theta_m\}$ denote the $m$th collection of iterates generated by Algorithm 1. Then (i) $\|\Theta_m^{(w)} - U^{(w)}\Theta_m\|_F \to 0$ and (ii) $p_m \to p^*$, where $p^*$ is the optimal objective function value of (3.4) and $p_m$ is the objective function value of (3.4) evaluated at $\{(\Theta_m^{(w)}), \Theta_m\}$.

The proof of Proposition 2 is a straightforward application of the convergence result presented in Section 3.2 of Boyd et al. [2011].
To terminate our algorithm, we use the stopping criteria described by Boyd et al. [2011]. The criteria are based on the primal and dual residuals, which represent the residuals for primal and dual feasibility, respectively. The primal residual at the \( m \)th iteration,

\[
r_{\text{primal}}^m = \sqrt{\sum_{w=1}^{W} \| \Theta_m^{(w)} - U_m^{(w)} \Omega_m \|^2_F},
\]

represents the difference between the trend values in the windows and the consensus trend value. The dual residual at the \( m \)th iteration,

\[
r_{\text{dual}}^m = \gamma \sqrt{\sum_{w=1}^{W} \| \Theta_m - \Theta_{m-1} \|^2_F},
\]

represents the change in the consensus variable from one iterate to the next. The algorithm is stopped when

\[
\begin{align*}
 r_{\text{primal}}^m &< \epsilon_{\text{abs}} \sqrt{nJ} + \epsilon_{\text{rel}} \max_w \left[ \max \left\{ \| \Theta_m^{(w)} \|_F, \| \Theta_m \|_F \right\} \right] \\
 r_{\text{dual}}^m &< \epsilon_{\text{abs}} \sqrt{nJ} + \epsilon_{\text{rel}} \sum_{w=1}^{W} \| \Omega_m^{(w)} \|^2_F.
\end{align*}
\] (3.8)

The quantile trend filtering problem for a single window is a linear program with \( N \times J \) parameters (number of observations by number of quantiles), which can be solved in computational time proportional to \((NJ)^3\). Consequently, solving a large problem using Algorithm 1 should require less computational time than solving (3.3), even if the sub-problems are solved sequentially. We demonstrate the advantages of Algorithm 1 through timing experiments (Figure 3.5). For each data size, 25 datasets were simulated using the peaks simulation design described below. Trends for the fifth, tenth, and fifteenth quantiles were fit simultaneously using \( \lambda_j = n/5 \) for all \( j \). We use from one to four windows for each data size with an overlap of 500. Algorithm 1 was the stopped when (3.8) was satisfied, defining \( \epsilon_{\text{abs}} = 0.01 \) and \( \epsilon_{\text{rel}} = 0.001 \). Figure 3.5 shows that using 4 windows
Timing experiments comparing quantile trend filtering with varying numbers of windows by data size.

instead of one on data sizes of 55,000 provides a factor of 3 decrease in computation time. The timing experiments were conducted on an Intel Xeon based Linux cluster using two processor cores.

3.3 Model Selection

An important practical issue in baseline estimation is the choice of the regularization parameter $\lambda$, which controls the degree of smoothness in $\theta$. In this section, we introduce four methods for choosing $\lambda$. The first is a validation based approach; the latter three are based on information criteria. Each of the criteria we compare is calculated for a single quantile ($\tau_j$). Rather than combine results over quantiles, we allow the value of $\lambda$ to vary by quantile resulting in $\lambda = \{\lambda_1, ..., \lambda_J\}$. To choose the best value for each $\lambda_j$, we first estimate all of the quantile trends using $\lambda_j = \lambda$ for all $j$ over a grid of values for $\lambda$. We then determine the $\lambda_j$ that maximizes the criteria chosen evaluated using $\tau_j$. Finally, we re-estimate the non-crossing trends with the optimal values for $\lambda_j$. A more thorough approach would involve fitting the model on a $J$ dimensional grid of values for $\lambda$ but this is computationally infeasible.

3.3.1 Validation

Our method can easily handle missing data by defining the check loss function to output 0 for missing values. Specifically, we use the following modified function $\tilde{\rho}_r$ in place of
the \( \rho_r \) function given in (3.2)

\[
\tilde{\rho}_r(r) = \sum_{i \notin \mathcal{V}} r_i (\tau - 1 (r_i < 0)),
\]

where \( \mathcal{V} \) is a held-out validation subset of \( \{1, \ldots, n\} \) and solve the problem

\[
\min_{\Theta \in \mathcal{C}} \sum_{j=1}^{J} \left[ \tilde{\rho}_{\tau_j}(y - \hat{\theta}_j + \lambda_j \|D^{(k+1)}\hat{\theta}_j\|_1) \right],
\]

which can be solved via Algorithm 1 with trivial modification to the quadratic program sub-problems. For each quantile level \( j \), we select the \( \lambda_j \) that minimizes the hold-out prediction error quantified by \( \tilde{\rho}_{\tau_j}(y - \hat{\theta}_j(\lambda_j)) \) where \( \hat{\theta}_j(\lambda_j) \) is the solution to (3.10) and

\[
\tilde{\rho}_r(r) = \sum_{i \in \mathcal{V}} r_i (\tau - 1 (r_i < 0)).
\]

### 3.3.2 Information Criteria

Koenker et al. [1994] addressed the choice of regularization parameter by proposing the Schwarz criterion for the selection of \( \lambda \)

\[
\text{SIC}(p_\lambda, \tau_j) = \log \left[ \frac{1}{n} \rho_{\tau_j}(y - \theta_j) \right] + \frac{1}{2n} p_\lambda \log n.
\]

where \( p_\lambda = \sum_i 1(y_i = \hat{\theta}_i) \) is the number of non-interpolated points, which can be thought of as active knots. Equivalently, \( p_\lambda \) can be substituted with the number of non-zero components in \( D^{(k+1)}\hat{\theta}_j \) which we denote \( \nu \) and have found to be more numerically stable. The SIC is based on the traditional Bayesian Information Criterion (BIC) which is given by

\[
\text{BIC}(\nu) = -2 \log(L(\hat{\theta})) + \nu \log n
\]

where \( L \) is the likelihood function. If we take the approach used in Bayesian quantile regression [Yu and Moyeed, 2001], and view minimizing the check function as maximizing the asymmetric Laplace likelihood,
\[
L(y | \theta) = \left[ \frac{\tau^n(1 - \tau)}{\sigma} \right]^n \exp \left\{ -\rho \left( \frac{y - \theta}{\sigma} \right) \right\},
\]
we can compute the BIC as
\[
\text{BIC}(\nu, \tau_j) = \frac{2}{\sigma} \rho \tau_j (y - \hat{\theta}_j) + \nu \log n
\]

where \( \hat{\theta} \) is the estimated trend, and \( \nu \) is the number of non-zero elements of \( D^{(k+1)} \hat{\theta} \).

We can choose any \( \sigma > 0 \) and have found empirically that \( \sigma = \frac{1 - |1 - 2\tau|}{2} \) produces stable estimates.

Chen and Chen [2008] proposed the extended Bayesian Information Criteria (eBIC), specifically designed for large parameter spaces.

\[
\text{eBIC}_\gamma(\nu) = -2 \log(L(\hat{\theta})) + \nu \log n + 2\gamma \log \left( \frac{P}{\nu} \right), \quad \gamma \in [0, 1]
\]

where \( P \) is the total number of possible parameters and \( \nu \) is the number of non-zero parameters included in given mod. We used this criteria with \( \gamma = 1 \), and \( P = n - k - 1 \).

In the simulation study, we compare the performance of the SIC, scaled eBIC (with \( \sigma \) defined above), and validation methods.

### 3.4 Simulation Studies

We conduct two simulation studies to compare the performance of our quantile trend filtering method and regularization parameter selection criteria with previously published methods. The first study compares the method’s ability to estimate quantiles when the observed series consists of a smooth trend plus independent error, but does not contain transient components. The second study is based on our application and compares the method’s ability to estimate baseline trends and enable peak detection when the time series contains a non-negative, transient signal in addition to the trend and random component.

We compare three criteria for choosing the smoothing parameter for quantile trend filtering: \( \lambda \) chosen using SIC (3.11) (\text{detrendr\_SIC}); \( \lambda \) chosen using the validation method with the validation set consisting of every 5th observation (\text{detrendr\_valid}); and \( \lambda \) cho-
Figure 3.6 Simulated data with true quantile trends.

For the second study, we also examine the effect of the non-crossing quantile constraint by estimating the quantile trends separately and choosing $\lambda$ using eBIC (detrendr_eBIC). We do not include detrendr_Xing in the first study because the difference in quantiles is large enough that we would not expect the non-crossing constraint to make a difference.

We also compare the performance of our quantile trend filtering method with three previously published methods, none of which guarantee non-crossing quantiles:

- npqv: The local linear quantile method (quantile-ll) described in Racine and Li [2017]. Code was obtained from the author.

- qsreg: Quantile smoothing splines described in Oh et al. [2004] and Nychka et al. [1995] and implemented in the fields R package [Douglas Nychka et al., 2017]. The regularization parameter was chosen using generalized cross-validation.

- rqss: Quantile trend filtering with $k = 1$ available in the quantreg package and described in Koenker et al. [1994]. The regularization parameter is chosen using a grid search and minimizing the SIC (3.11) as described in Koenker et al. [1994].

3.4.1 Estimating Quantiles

To compare performance in estimating quantile trends in the absence of a signal component, three simulation designs from Racine and Li [2017] were considered. For all designs
\[ t = 1, \ldots, n, \ x(t) = t/n, \text{ and the response } y \text{ was generated as} \]
\[ y(t) = \sin(2\pi x(t)) + \epsilon(t) \]

The errors were simulated as independent draws from the following distributions:

- **Gaussian**: \( \epsilon(x(t)) \sim N\left(0, \left(\frac{1+x(t)^2}{4}\right)^2\right) \)
- **Beta**: \( \epsilon(x(t)) \sim \text{Beta}(1, 11 - 10x(t)) \)
- **Mixed normal**: \( \epsilon(x(t)) \) is simulated from a mixture of \( N(-1, 1) \) and \( N(1, 1) \) with mixing probability \( x(t) \).

The true quantile trends and an example simulated data set is show in Figure 3.6. One hundred datasets were generated of sizes 300, 500 and 1000.

Quantile trends were estimated for \( \tau = \{0.05, 0.25, 0.5, 0.75, 0.95\} \) and the root mean squared error was calculated as \( \text{RMSE}(\tau_j) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{\theta}_{ij} - \theta_{ij})^2} \), where \( \theta_{ij} \) is the true value of the \( \tau_j \)th quantile of \( y \) at \( t = i \). Figure 3.7 shows the mean RMSE plus or minus twice the standard error for each method, quantile level, and sample size. In all three designs the proposed \texttt{detrend} methods are either better than or comparable to existing methods. Overall the \texttt{detrend_eBIC} performs best. In the mixed normal design, specifically, our methods have lower RMSEs for the 5th and 95th quantiles. The \texttt{npqw} method performs particularly poorly on the mixed normal design due to the violation of the assumption that the data come from a scale-location model.

### 3.4.2 Peak Detection

The second simulation design is closely motivated by our air quality analysis problem. We assume that the measured data can be represented by
\[ y(t_i) = \theta(t_i) + s(t_i) + \epsilon_i, \]
where \( t_i = i \) for \( i = 1, \ldots, n \), \( \theta(t) \) is the drift component that varies smoothly over time, \( s(t) \) is the true signal at time \( t \), and \( \epsilon_i \) are i.i.d. errors distributed as \( N(0, 0.25^2) \). We generate \( \theta(t) \) using cubic natural spline basis functions with degrees of freedom sampled from a Poisson distribution with mean parameter equal to \( n/100 \), and coefficients drawn
Figure 3.7 RMSE by design, method, quantile and data size. Points and error bars represent mean RMSE ± twice the standard error.
from an exponential distribution with rate 1. The true signal function $s(t)$ is assumed to be zero with peaks generated using the Gaussian density function. The number of peaks is sampled from a binomial distribution with size equal to $n$ and probability equal to 0.005 with location parameters uniformly distributed between 1 and $n - 1$ and bandwidths uniformly distributed between 2 and 12. The simulated peaks were multiplied by a factor that was randomly drawn from a normal distribution with mean 20 and standard deviation of 4. An example dataset with 4 signal peaks is shown in Figure 3.9. One hundred datasets were generated for each $n \in \{500, 1000, 2000, 4000\}$.

We compare the ability of the methods to estimate the true quantiles of $y(t_i) - s(t_i)$ for $\tau \in \{0.01, 0.05, 0.1\}$ and calculate the RMSE (Figure 3.8). In this simulation study, our proposed method `detrend_eBIC` method substantially outperforms the others. The `detrend_Xing` method, which is the `detrend_eBIC` method fit without the non-crossing constraints, performs similarly for larger quantiles and larger datasets. However, `detrend_Xing` produces significantly worse estimates for more extreme quantiles ($\tau = 0.01$) and smaller data sets ($\tau = 0.05$ and $n = 500$). These results indicate that even when a single quantile is of interest, simultaneously fitting nearby quantiles and utilizing the non-crossing constraint can improve estimates when data is sparse by using information from nearby quantiles. The `qsreg` method is comparable to the `detrend_eBIC` method on the larger datasets, but its performance deteriorates as the data size shrinks. The `npqw` and `rqss` methods both perform poorly on this design.

While minimizing RMSE is desirable in general, in our application, the primary met-
ric of success is accurately classifying observations $y_i$ into signal present or absent. To evaluate the accuracy of our method compared to other methods we define true signal as any time point when the simulated peak value is greater than 0.5. We compare three different quantiles for the baseline estimation and four different thresholds for classifying the signal after subtracting the estimated baseline from the observations. Figure 3.9 illustrates the observations classified as signal after subtracting the baseline trend compared to the "true" signal.

To compare the resulting signal classifications, we calculate the class averaged accuracy (CAA), which is defined as

$$\text{CAA} = \frac{1}{2} \left[ \frac{\sum_{i=1}^{n} \mathbb{1}[\delta_i = 1 \cap \hat{\delta}_i = 1]}{\sum_{i=1}^{n} \mathbb{1}[\delta_i = 1]} + \frac{\sum_{i=1}^{n} \mathbb{1}[\delta_i = 0 \cap \hat{\delta}_i = 0]}{\sum_{i=1}^{n} \mathbb{1}[\delta_i = 0]} \right].$$

where $\delta_i \in \{0, 1\}$ is the vector of true signal classifications and $\hat{\delta}_i \in \{0, 1\}$ is the vector of estimated signal classifications, namely $\hat{\delta}_i = \mathbb{1}(y_i - \hat{\theta}_i > 0.5)$. We use this metric because our classes tend to be very imbalanced with many more zeros than ones. The CAA metric should give a score close to 0.5 both for random guessing and also for trivial classifiers such as $\hat{\delta}_i = 0$ for all $i$.

Our `detrend_eBIC` method results in the largest CAA values (Figure 3.10) in addition to the smallest RMSE values (Figure 3.8). While `qsreg` was competitive with our method in some cases, in the majority of cases the largest CAA values for each threshold were produced using the `detrend_eBIC` method with the 1st or 5th quantiles.
Figure 3.9 Example signal classification using threshold. Red indicates true signal \((y_i - \theta_i > 0.5)\), blue indicates observations classified as signal, i.e. values greater than 1.2 after baseline removal using \texttt{detrend\_eBIC}.

Figure 3.10 Class averaged accuracy by threshold, data size, and method (1 is best 0.5 is worst).
3.5 Analysis of Air Quality Data

The low-cost "SPod" air quality sensors output a time series that includes a slowly varying baseline, the sensor response to pollutants, and high frequency random noise. These sensors record measurements every second and are used to monitor pollutant concentrations at the perimeter of industrial facilities. Time points with high concentrations are identified and compared with concurrent wind direction and speed. Ideally, three co-located and time aligned sensors (as shown in Figure 3.1) responding to a pollutant plume would result in the same signal classification after baseline trend removal and proper threshold choice. We first illustrate the difference between our detrend_eBIC method, hereafter referred to as detrend_eBIC, and qsreg using data from 13:10 to 15:10 from Figure 3.1 (Section 3.5.1). We then compare the methods on the complete day shown in Figure 3.1, estimating trends by applying the qsreg method to 2 hour non-overlapping windows of the data and Algorithm 1 to the entire day. We focus on this day because data from three sensors was available. Finally, we examine an entire week of measurements from two co-located sensors (Section 3.5.2).

3.5.1 Short series of air quality measurements

We compare our detrend_eBIC method with the qsreg method on a two-hour subset of one-second SPod data (n=7200) both to facilitate visualization and because the qsreg method cannot handle all 24 hours simultaneously. We estimate the baseline trend using $\tau = \{0.01, 0.05, 0.1\}$ and compare three thresholds for classifying the signal. The thresholds are calculated as the 90\(^{th}\), 95\(^{th}\), and 99\(^{th}\) quantiles of the de-trended series for each SPod. If there is signal present in the dataset, values above these thresholds should occur simultaneously on all three SPods. We do not use class-averaged accuracy to compare the signal classifications because we do not have a reference value to define as the "true" signal. Instead, we compute the variation of information (VI) which compares the similarity between two classifications. Given the signal classifications for SPods a and
Figure 3.11 Estimated 5th quantile trends for SPods a, b, and c, using `qsreg` and `detrendr`. SPod c contains some missing values that were interpolated before the trends were estimated. `qsreg` is more influenced by the signal component apparent on both nodes resulting in undersmoothing.
b, \( \delta_i^{(a)} \in \{0, 1\} \) and \( \delta_i^{(b)} \in \{0, 1\} \), for \( i \in \{1, \ldots, n\} \) the VI is defined as:

\[
    r_{jk} = \frac{1}{n} \sum_i \mathbb{1}(\delta_i^{(a)} = j \cap \delta_i^{(b)} = k)
\]

\[
    \text{VI}(a, b) = -\sum_{j,k} r_{jk} \left[ \log \left( \frac{1}{n} \sum_i \mathbb{1}(\delta_i^{(a)} = j) \right) + \log \left( \frac{1}{n} \sum_i \mathbb{1}(\delta_i^{(b)} = k) \right) \right]
\]

where \((j,k) \in \{(0,0), (0,1), (1,0), (1,1)\}\). The VI is a distance metric for measuring similarity of classifications and will be 0 if the classifications are identical and increase as the classifications become more different.

Figure 3.11 shows the estimated 5th quantile trends from each method for each SPod. The \texttt{detrendr} method results in a smoother baseline estimate while the \texttt{qsreg} method absorbs more of the peaks obscuring some of the signal. Figure 3.12 shows the series after subtracting the \texttt{detrendr} estimate of the 5th quantile and classifying the signal using the 95th quantile of the detrended data. The 90th and 99th thresholds are also shown for comparison in blue and orange, respectively. The largest peaks at 14:10 are easily identified as signal, but good baseline estimates also enable proper classification of the smaller peaks like the one at 15:12. The under-smoothing of the \texttt{qsreg} method results in less similar signal classifications and higher VI values for the 90th and 95th quantile thresholds (Figure 3.13). However, when the 99th threshold is used only the highest observations are classified as signal and the baseline estimation method isn’t as important (Figure 3.13).

### 3.5.2 Long series of air quality measurements

Algorithm 1 was used to remove the baseline drift from the full day of data (Figure 3.1) consisting of 86,400 observations per SPod and compared to the series detrended using the \texttt{qsreg} trends estimated using non-overlapping 2 hour windows. As in the shorter illustration, the \texttt{detrendr} method results in generally lower VI scores than the \texttt{qsreg} method (Figure 3.14). The \texttt{detrendr} method also results in better correlation in the detrended series as is illustrated in (Figure 3.15). The Spearman correlation coefficients for SPods a and b, SPods a and c, and SPods b and c after removing the 5th quantile trend using \texttt{detrendr} were 0.37, 0.75, and 0.43, compared with 0.07, 0.24, and 0.16 using \texttt{qsreg}. The noise variance was higher for SPod b than for SPods a and c resulting in
Figure 3.12 Rugplot showing locations of signal after baseline removal using the `detrendr` estimate of 5th quantile. Horizontal dashed lines represent the thresholds calculated using the 90th, 95th, and 99th quantiles of the detrended data. The 95th quantile (black) was used to classify the signal shown as vertical lines at the bottom of the plot.

Figure 3.13 Variation of Information between sensors by method (color), quantile (columns) and threshold (shape) for two hour time period.
lower correlation and higher VI values for the ab and bc metrics compared with ac.

Finally, we estimated the quantile trends for 7 days of measurements of two co-located SPods. Figure 3.16 demonstrates the improvement in classification similarity when using `detrendr`. Each point represents a day of measurements and all points that fall below the dashed line have more similar classifications using `detrendr` compared to `qsreg`. The improvement of `detrendr` over `qsreg` is more severe at lower thresholds. This indicates that `detrendr` gives greater agreement on signal classification when the methods are tuned to deliver positive classifications more frequently.

### 3.6 Conclusion and Discussion

We have expanded the quantile trend filtering method by implementing a non-crossing constraint, a new algorithm for processing large series, and proposing a modified criteria for smoothing parameter selection. Furthermore, we have demonstrated the utility of quantile trend filtering in both simulations and applied settings. Our ADMM algorithm for large series both reduces the computing time and allows trends to be estimated on series that cannot be estimated simultaneously while our scaled extended BIC criterion.
Figure 3.15 SPod a versus SPod c before and after de-trending with qsreg and detrendr.

Figure 3.16 Variation of Information (VI) for detrendr and qsreg by quantile trend and threshold. Each point represents a full day of data. The dashed line represents y=x. In most cases detrendr results in a lower VI than qsreg.
was shown to provide better estimates of quantile trends in series with and without a signal component.

In our application to low cost air quality sensor data, we have shown that the baseline drift in low cost air quality sensors can be removed through estimating quantile trends, but the data size was too large for existing methods to be computationally feasible. While \texttt{qsreg} cannot feasibly handle more than a few hour windows of data, our new methods were able to process 24 hours simultaneously and deliver signal classifications that were more consistent between the two sensors for a week of data (168 hours).

In the future, quantile trend filtering could be extended to observations measured at non-uniform spacing by incorporating the distance in covariate spacing into the differencing matrix. It could also be extended to estimate smooth spatial trends by a similar adjustment to the differencing matrix based on spatial distances between observations.

3.7 Acknowledgments

The authors greatly appreciate Eben Thoma at the US Environmental Protection Agency for providing the SPod datasets. The authors are also thankful for the Oak Ridge Institute of Science Fellowship that partially supported this work. Plots were created using \texttt{ggplot2} [Wickham, 2016].

3.8 Appendix

R-package \texttt{detrendr} containing code to perform the methods described in this chapter is available at \url{https://github.com/halleybrantley/detrendr}. 
4.1 Introduction

Air pollutant emissions from industrial facilities can originate from known, well characterized sources such as stacks or more complex, possibly unknown sources such as equipment leaks, process startups, malfunctions, and waste ponds. Many of these emissions are located at ground level and can result in near-source exposure and odor nuisance issues. These emissions are also often temporally variable or randomly occurring making them difficult to locate and characterize [Thoma et al., 2019]. To obtain information on the location and strength of sources within facilities, high time resolution measurements of air pollutants and meteorology can be collected using a variety of sensors and instruments ranging in cost and precision [Snyder et al., 2013].
The fields of pollution source assessment and pollution source apportionment are both dedicated to identifying and characterizing air pollution sources. Pollution source apportionment methods, however, consist primarily of matrix decomposition techniques and can be limited by their reliance on simultaneous measurements of multiple pollutants and known source profiles [Hopke, 2016]. Consequently, source apportionment methods cannot be used with sensors that provide a single concentration time series rather than pollutant composition data. The most common methods for combining pollutant concentrations with wind information are graphical, and include pollution roses implemented in the openair R package [Carslaw and Ropkins, 2012], and plots of the Potential Source Contribution Function (PSCF) and the Conditional Probability Function (CFP) [Pekney et al., 2006]. Both the PSCF and CFP plots rely on backward trajectory models. Several more complex statistical models for concentrations have been proposed, either using corresponding wind measurements directly [Henry et al., 2002] or forward dispersion model outputs from a fixed number of source locations [Williams et al., 2011, Humphries et al., 2012, Brereton et al., 2018]. These models all assume constant source strength over time and either require an array of sensors or wind information [Humphries et al., 2012, Brereton et al., 2018] or only provide estimates of the direction of a source and not the location or source strength [Williams et al., 2011] and are thus not suitable for use with a single near-source sensor.

Backward trajectory modeling was developed to identify the locations of pollutant sources by combining measurements of pollutant concentrations and meteorological conditions. The Hybrid Single-Particle Lagrangian Integrated Trajectory model (HYSPLIT) developed by the National Oceanic and Atmospheric Administration (NOAA) dates back to 1949 when the U.S. Weather Bureau was tasked with identifying the source of radioactive debris from the first Soviet atomic test. Substantial progress in meteorological modeling has been made since the first trajectories were calculated by hand. However, backward trajectory modeling has remained relevant and improvements include the incorporation of a Lagrangian dispersion component in the backward trajectory calculations [Stein et al., 2015].

A forward Lagrangian stochastic model is the most natural means of modeling dispersion in the atmospheric boundary layer and operates by simulating thousands of tracer particles at a source and recording their locations as they travel downwind [Flesch et al., 2004]. Backward-in-time trajectory models with dispersion can be calculated in a similar
manner by following particles backward in time from a receptor and reversing the sign of the "damping" terms in the dispersion equations. While this model has some limitations, the possible loss of accuracy from relying on backward calculations from a single measurement location is counterbalanced by the dramatic increase in efficiency compared to forward calculations from multiple unknown source locations. While the HYSPLIT model was developed to model long range pollutant transport on the order of hours to days, the frequently used software "WindTrax" [Crenna et al., 2008] applies similar principles on the order of seconds and is more appropriate for near-source applications. The details of the surface-layer backward Lagrangian Stochastic model (bLS) implemented in "Windtrax" are presented by Flesch et al. [2004].

To the best of our knowledge, all of the previous approaches have assumed constant source strengths over the period of interest. This is a significant limitation, since many sources of fugitive emissions are known to be temporally variable. We propose a new Bayesian hierarchical model that enables the estimation of source strengths at multiple locations that vary over time using backward trajectories calculated using the backwards Lagrangian Stochastic (bLS) dispersion model. While we use the bLS trajectories here, our model can be used to combine pollutant measurements with backward or forward trajectories from any appropriate dispersion model.

4.2 Methods

4.2.1 Data Description and Exploratory Analysis

The near-source sensor data modeled here was collected as part of the Rubbertown Next Generation Emissions Measurement Demonstration project described in Thoma et al. [2019]. Our focus is on the analysis of the time-resolved, non-speciating photoionization detector (PID) sensors deployed near industrial facilities in Louisville, Kentucky. The EPA SPod sensor combines pollutant concentration information from a high sensitivity 10.6 eV PID (PN 045-014, Baseline-Mocon, Inc. or PN MiniPID2-HS, Ion Science Inc.) with time-synchronized wind vector and sonic temperature measurements from an ultrasonic anemometer (PN 81000V, R.M. Young, Inc.). The SPod PIDs produce an uncalibrated volatile organic compound detection signal in mV, digitized to 16 bits and reported as "counts" (cts). The PID values shown in the figures all represent 1000s of
Figure 4.1 Co-located measurements from SPod A and B over time and compared to each other.

The SPod sensors also exhibit baseline drift due to changes in temperature and relative humidity. The incorporation of a polymide strip heater (PN HK6903, Minco) running at 30°C into the SPod design reduced the amount of baseline drift but did not eliminate it entirely. Two SPod sensors were co-located in order to facilitate the decomposition of the series into baseline trend and signal components.

In this chapter we examine four days of data (June 17-20) during which there was a known emission, confirmed by other nearby measurements [Thoma et al., 2019]. The baseline drift was estimated using the detrendr method described in Ch. 3 and the detrended series were aggregated into 10 min averages, the interval required to calculate the wind statistics for the dispersion model. After de-trending and aggregating the Spearman correlation between the two SPods was 0.87. The measurements for each of the SPods are shown in Figure 4.1.

Each of the SPods also recorded 3D wind vectors and sonic temperature at 1 Hz. The Spearman correlations between the 10 min measurements of the $u$ (from the west) and $v$ (from the south) components of the wind vector were both greater than 0.99, while the correlation between the $w$ (vertical) components was 0.94. The Pearson correlation
coefficients were similar. Wind roses by day are shown in Figure 4.2. The wind is predominantly from the south and southwest with occasional wind from the northeast and southeast. The wind speed ranged from 0.08 to 4.6 m/s. As a preliminary analysis, polar plots of the PID signal were created by binning the wind direction and speed, and averaging the PID measurements in each bin by day and SPod using the openair R package [Carslaw and Ropkins, 2012] (Figure 4.3). On June 17th, the high signal seems to correspond with low wind speeds but the direction is not clear. On June 18th through June 19th, there seems to be possible evidence of two sources, one to the southwest and one to the northwest. While these plots can provide useful, quick summaries of the relationship between wind measurements and concentration measurements, a more sophisticated model for the sensor measurements that incorporates an established dispersion model can provide much more information on the source locations and strengths, especially for sources that may vary over time.

4.2.2 Back-trajectory Calculations

The "WindTrax" model has been implemented in the R package bLSmodelR by Häni et al. [2018]. The bLS model assumes horizontally homogeneous and vertically inhomogeneous, Gaussian turbulence. The vertical profiles of the wind speed and the turbulence statistics are based on Monin-Obhukov Similarity Theory (MOST). The high frequency (1 Hz) sonic anemometer measurements of the 3D wind components and the sonic temperature were collected at a height of 2.3 m above ground level and processed as described in Häni et al. [2018]. We outline the processing steps here. For each 10-min interval the
Figure 4.3 Polar plots of PID signal binned by wind direction and speed.
coordinate system of the wind vector measurements was rotated three times so that one of the axes points in the direction of the average wind velocity. The resulting vector components $u$, $v$ and $w$ represent the along-wind, cross-wind, and vertical velocities, respectively. The rotations were calculated so that $\bar{v} = \bar{w} = v'w' = 0$, for the rotated measurements in each 10 minute interval. The overbar represents the temporal average and the prime denotes the deviation from the average. From the rotated vectors, the wind vector variances and covariances and the heat flux defined as the covariance of the vertical wind component and the sonic temperature were calculated for each 10 min interval. The MOST parameters: the friction velocity ($u_*$) and the Obhukov length ($L$) were also calculated from the rotated vectors using

$$u_* = \sqrt{-\bar{u}'w'}$$

$$L = -\frac{u_*^3T_v}{\kappa g(w'T_v)}$$

where $T_v$ is the sonic temperature in Kelvins, $\kappa$ is the von Karman constant equal to 0.4, and $g$ is the gravitational constant equal to 9.8 m/s$^2$. These statistics were then used as

**Figure 4.4** 50 calculated back-trajectories during very stable (a) and very unstable (b) atmospheric conditions.
inputs into the backward trajectory function from \texttt{bLSmodelR} [Häni et al., 2018].

Example backward trajectories are shown in Figure 4.4. Unstable conditions are defined as those in which $-100 < L < 0$, and are typical during the daytime. Unstable conditions result in much more dispersion and uncertainty around the source location (Figure 4.4b). Stable conditions are defined as $0 < L < 10$, are more common at night, and result in more predictable transport of sources (Figure 4.4a).

This bLS trajectory model was originally developed for area sources at ground level and only particles that intersect the ground were considered when calculating estimated source strength. However, emissions from industrial facilities may occur at a number of heights so we consider all particle trajectories below 50 m rather than only those that intersect the ground. To identify the source location we divide the area around the sensor location into grid cells. In this chapter, we consider 100 m and 200 m grid cells centered at the measurement location and extending 800 m in each cardinal direction. For each 10 min set of meteorological parameters we calculated 500 backward trajectories and calculated the proportion of the trajectories that pass through each grid cell.

### 4.2.3 Model Framework

Defining $Y(t)$ as the sensor measurement at time $t$, and $Q_k(t)$ as the source strength of grid cell $k$ at time $t$, we model

$$Y(t) = \beta_0 + \sum_k W_k(t)Q_k(t) + \epsilon(t)$$  \hspace{1cm} (4.1)

$$\epsilon(t) \sim N(\rho \epsilon(t-1), \sigma^2)$$  \hspace{1cm} (4.2)

where $W_k(t)$ is the proportion of particle back trajectories originating from the sensor location at time $t$ that cross grid cell $k$ at a height less than 50 m. The quantity $W_k(t)Q_k(t)$ is meant to capture the amount of source material that is transported from grid cell $k$ to the sensor. We assume $Q_k(t)$ is non-negative and varies smoothly over time. We define $\theta(t) = (\theta_1(t), \ldots, \theta_M(t))^T$ as a fixed B-spline basis made up of $M$ functions evaluated at $t$ and model $Q_k(t) = \theta(t)^T \beta_k$, where $\beta_k$ is the $M \times 1$ vector of spline coefficients for the
$k$th grid cell. For the source effects in (4.1) evaluated at $t$, substitution yields

$$
\sum_k W_k(t)Q_k(t) = \sum_k W_k(t)[\theta(t)^T\beta_k]
$$

(4.3)

$$
= \sum_k [W_k(t)\theta(t)]^T\beta_k
$$

(4.4)

$$
= \sum_k W^*_k(t)^T\beta_k
$$

(4.5)

where $W^*_k(t) \in \mathbb{R}^M$. By using B-spline basis functions which are by definition non-negative we can ensure non-negative source estimates by using log multivariate normal priors on $\beta_k$ that penalize both the values and the first order differences.

$$
\log(\beta_k) \sim \mathcal{N}(\mu_0, (\gamma_k I + \lambda_k D^TD)^{-1})
$$

(4.6)

where $D$ is the first order difference operator:

$$
D = \begin{pmatrix}
-1 & 1 & 0 & \cdots & 0 & 0 \\
0 & -1 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & -1 & 1
\end{pmatrix}
$$

By penalizing the differences in the basis coefficients we stabilize the estimates of the source at time points when we don’t have information due to the wind direction. For both the simulation study and application we set $\mu_0 = -6$ and use the following priors for the other parameters:

$$
\gamma_k \sim \text{Gamma}(0.1, 25)
$$

$$
\lambda_k \sim \text{Gamma}(100, 25)
$$

$$
\rho \sim \text{Gamma}(5, 100)
$$

$$
\beta_0 \sim \mathcal{N}(0, 0.02^2)
$$

$$
\sigma \sim \begin{cases} 
1 & \sigma > 0 \\
0 & \sigma \leq 0
\end{cases}
$$
The prior for $\sigma$ is improper. We implement the model using the \texttt{rstan} package [Stan Development Team, 2018].

### 4.3 Simulation Study

To demonstrate the effectiveness of our modeling framework under differing source characteristics, we evaluate our estimates on simulated data. We simulate data using the backward trajectory weights from June 19th when the wind was predominantly from the southwest (Figure 4.2). A total of $n = 144$ observations of 10 min back trajectories were used, and 3 designs were considered. For each design we consider the weights for both the 100 m and 200 m grids described above. Only grid cells with non-zero weights for at least 20% of the observations were considered as predictors when fitting the model. Two source locations were selected (Figure 4.5a), and temporal source profiles were defined using 8 B-spline basis functions (Figure 4.5b). The source strengths were estimated using 6 and 10 basis functions to compare the effects of underestimating versus overestimating the smoothness of the source function. Source 1 had a higher mean weight value than source 2, indicating that the wind more frequently blows from source 1 toward the sensor than from source 2. Additionally, the backward trajectory weights for source 1 were non-zero during the early and late parts of the day, while the weights for source 2 were only non-zero during the middle of the day (Figure 4.6). The three designs were defined as follows:

- D1: Source 1 is non-zero, Source 2 is zero.
- D2: Source 2 is non-zero, Source 1 is zero.
- D3: Both Source 1 and 2 are non-zero.

Fifty datasets were simulated for each grid size and design using Eq. (4.1) with $\rho = 0.2$, $\sigma = 0.1$, $\beta_0 = 0.05$. Example simulated sensor responses for each design are shown in Figure 4.7. The effect of source 1 is clear at the beginning of the time period while the effect of source 2 occurs between time points 50 and 100.

Figure 4.8 shows all of the estimated source functions from the 50 simulated datasets along with the mean of the estimated functions and the true source function. For almost all of the simulated datasets our model correctly identifies the location and temporal
Figure 4.5 (a) Source locations for simulation designs, the triangle represents the sensor location and the color represents the average value of trajectory weights in each cell. (b) Source strengths over time used for simulation designs.
Figure 4.6 Proportion of backward trajectories that pass through source cell 1 and source cell 2 for 100m (left) grid and 200m grid (right).

Profile of source 1. Overall, the estimates for the 200 m grid tend to be better than those from the 100 m grid, likely due to increased correlation between the 100 m grid weights. Increasing the number of basis functions improves the estimates of source 1 using the 200 m grid but doesn’t improve the estimates for the 100 m grid. For source 2, the model is generally able to correctly identify the location and the shape between time points 50 and 100, but it overestimates the source strength between time points 1 to 50, and 125 to 144. Using more basis functions improves the estimate of the peak at time 80, but also exacerbates the overestimates. Because there are no trajectories that pass through the source location during this time period, there are no observations to indicate that the source has decreased. Additionally, especially for the 100 m grid, there are a number of cases where source 2 is incorrectly attributed to other grid cells. This issue is even more prevalent for design 3. Figure 4.9 shows the locations of grid cells with sources incorrectly estimated as non-zero. The colors represent the mean estimated source strength across the 50 simulated datasets. The most likely grid cells to have incorrect non-zero estimates are those adjacent to the true source and along the vector connecting the source to the sensor location. The weights in these adjacent cells are the most highly correlated with the weights in the true source cell. Even so, in the majority of cases the true source was correctly identified. For design 3, using the 100 m grid and 10 bases, the mean of the estimates of the true sources were 3.8 and 1.6, which are an order of magnitude larger.
than the mean of the estimates for the next highest grid cell, 0.1.

In addition to the visual comparison of the source estimates, we also calculate the root mean squared error of the source estimates and the leave-one-out expected log pointwise predictive density estimate (elpd_{loo}) [Vehtari et al., 2017] (Table 4.1). The expected log pointwise predictive density (elpd) is an in-sample metric of the predictive power of the model and represents the sum of the log probabilities of observing the data used to fit the model. Higher values of elpd represent higher probabilities of observing the data given our model. We also want to evaluate our models on their ability to predict out-of-sample observations or observations that were not used to fit the model. Rather than use k-fold cross validation which is computationally intensive, we compare the models using the estimate of leave-one-out elpd proposed by Vehtari et al. [2017]. The leave-one-out (LOO) expected log pointwise predictive density is defined as

\[
elpd_{loo} = \sum_{i=1}^{n} \log p(y_i|y_{-i}) \tag{4.7}\n\]

where \(p(y_i|y_{-i})\) is the posterior predictive density of \(y_i\) given the other values of \(y\). Vehtari et al. [2017] demonstrate that stable estimates of elpd_{loo} can be obtained from a single model fit using importance sampling. We define \(\hat{elpd}_{loo}\) as the importance-sampled estimate of elpd_{loo}, and calculate it using the loo R package [Vehtari et al., 2018].

The RMSE and elpd_{loo} were calculated for each of the 50 simulated datasets and the means and standard errors are reported in Table 4.1. As seen in Figure 4.8 increasing the number of bases for design 1 with the 100 m grid does not reduce the RMSE or
Figure 4.8 Estimates of source strength for designs 1, 2, and 3 (left to right). Bold lines indicate true source strength, dashed lines indicate mean of simulation estimates.
**Figure 4.9** Estimates of mean source strength for design 3 by location.

**Table 4.1** Comparison of simulation results by design, number of bases, and grid size, SE is the standard error for the 50 simulated datasets.

<table>
<thead>
<tr>
<th>Design</th>
<th>Bases</th>
<th>Grid</th>
<th>elpd_{100}</th>
<th>SE</th>
<th>RMSE</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>100</td>
<td>119.2</td>
<td>1.3</td>
<td>0.027</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>100</td>
<td>120.0</td>
<td>1.3</td>
<td>0.029</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>200</td>
<td>118.9</td>
<td>1.6</td>
<td>0.035</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>200</td>
<td>120.0</td>
<td>1.6</td>
<td>0.028</td>
<td>0.002</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>100</td>
<td>120.6</td>
<td>1.2</td>
<td>0.104</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>100</td>
<td>121.7</td>
<td>1.3</td>
<td>0.119</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>200</td>
<td>121.7</td>
<td>1.2</td>
<td>0.144</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>200</td>
<td>122.7</td>
<td>1.2</td>
<td>0.193</td>
<td>0.011</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>100</td>
<td>118.2</td>
<td>1.4</td>
<td>0.121</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>100</td>
<td>120.0</td>
<td>1.4</td>
<td>0.133</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>200</td>
<td>120.1</td>
<td>1.5</td>
<td>0.160</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>200</td>
<td>121.6</td>
<td>1.4</td>
<td>0.198</td>
<td>0.010</td>
</tr>
</tbody>
</table>
significantly change the \(\hat{elpd}_{loo}\); however, it does reduce the RMSE and increase the \(\hat{elpd}_{loo}\) for the 200 m grid. For designs 2 and 3 the RMSE values for the 200 m grid are larger than for the 100 m grid, likely a result of a smaller number of sources with true values of 0. The \(\hat{elpd}_{loo}\) values, on the other hand are greater for the 200 m grid, although not significantly. The results demonstrate that the choice of grid size and number of bases do not have a large impact on the estimates as long as the correlation between weights in adjacent grid cells is not too large.

4.4 Results

We fit our model to the four days of data shown in Figure 4.1. The SPods are modeled separately in order to compare the resulting source estimates. All of the source strength estimates shown are relative estimates, because the SPods do not provide calibrated concentration data. We fit the model separately to each day to reduce computation time. For each day of data and each node, we compared grids of 100 m and 200 m and 8, 9, and 10 basis functions using the estimated in-sample expected pointwise predicted density (\(\hat{elpd}\)) and estimated leave-one-out elpd (\(\hat{elpd}_{loo}\)) in Table 4.2. Higher values of \(\hat{elpd}\) represent higher predicted probabilities of the data used to fit the model, while higher values of \(\hat{elpd}_{loo}\) represent higher predicted probabilities for new data. The \(\hat{elpd}\) values for SPod B are generally lower than those for SPod A, due to the higher overall variance in the measurements of SPod B as seen in Figure 4.10. As expected, the in-sample elpd generally increases with more basis functions and more predictors (smaller grid size). On 6/17 and 6/20, for both SPods the \(\hat{elpd}_{loo}\) is higher for the 200 m grid than the 100 m grid, while the opposite is true for 6/19. On 6/18 the 100 m grid produces higher \(\hat{elpd}_{loo}\) values for SPod A but not SPod B. The best choice of number of basis functions varies by both day and SPod, but on many of the days the differences are negligible.

Figure 4.10 and Figure 4.11 provide an example of the predicted values from the models with 100 m grid cells and 9 basis functions compared to the observed values for both SPods. The predicted values explain 46% and 49% of the variation in SPods A and B, respectively. It can be seen in Figure 4.10 that the large peaks are generally well predicted by the model, but the model fails to capture the more rapid transitions on June 17 and June 18.

The estimated source strengths vary between days and within days, and the 80%
Table 4.2 Comparison of in-sample and LOO elpd for June 17-19 by node, grid size and number of bases.

<table>
<thead>
<tr>
<th>Day</th>
<th>Bases</th>
<th>SPod A</th>
<th></th>
<th></th>
<th></th>
<th>SPod B</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>eLPD</td>
<td>eLPD</td>
<td>eLPD\textsubscript{loo}</td>
<td>eLPD\textsubscript{loo}</td>
<td></td>
<td>eLPD</td>
<td>eLPD</td>
<td>eLPD\textsubscript{loo}</td>
</tr>
<tr>
<td>Grid</td>
<td>100</td>
<td>200</td>
<td>100</td>
<td>200</td>
<td></td>
<td>100</td>
<td>200</td>
<td>100</td>
<td>200</td>
</tr>
<tr>
<td>6/17</td>
<td>8</td>
<td>93.7</td>
<td>92.1</td>
<td>45</td>
<td>50.7</td>
<td>-0.2</td>
<td>-2.8</td>
<td>-20.7</td>
<td>-17.9</td>
</tr>
<tr>
<td>6/17</td>
<td>9</td>
<td>94.5</td>
<td>93.4</td>
<td>44.6</td>
<td>54.8</td>
<td>0.5</td>
<td>-1.9</td>
<td>-20.7</td>
<td>-18</td>
</tr>
<tr>
<td>6/17</td>
<td>10</td>
<td>94.9</td>
<td>94.4</td>
<td>47</td>
<td>56.2</td>
<td>1.9</td>
<td>-0.9</td>
<td>-21.3</td>
<td>-17</td>
</tr>
<tr>
<td>6/18</td>
<td>8</td>
<td>174.7</td>
<td>165.5</td>
<td>129.6</td>
<td>111.4</td>
<td>53.2</td>
<td>52.7</td>
<td>12.4</td>
<td>21.8</td>
</tr>
<tr>
<td>6/18</td>
<td>9</td>
<td>176.8</td>
<td>166.3</td>
<td>131.5</td>
<td>113.9</td>
<td>54</td>
<td>53.5</td>
<td>15.3</td>
<td>24</td>
</tr>
<tr>
<td>6/18</td>
<td>10</td>
<td>177.5</td>
<td>170.1</td>
<td>128.8</td>
<td>117</td>
<td>57.4</td>
<td>56.5</td>
<td>13.9</td>
<td>24.1</td>
</tr>
<tr>
<td>6/19</td>
<td>8</td>
<td>246.2</td>
<td>233</td>
<td>226</td>
<td>217.9</td>
<td>162.2</td>
<td>157.6</td>
<td>144.7</td>
<td>143.3</td>
</tr>
<tr>
<td>6/19</td>
<td>9</td>
<td>250.1</td>
<td>239.2</td>
<td>231.4</td>
<td>222.9</td>
<td>163.4</td>
<td>158.2</td>
<td>145.9</td>
<td>142.3</td>
</tr>
<tr>
<td>6/19</td>
<td>10</td>
<td>250.4</td>
<td>241.7</td>
<td>229.3</td>
<td>224.3</td>
<td>165.6</td>
<td>160.7</td>
<td>148.4</td>
<td>145</td>
</tr>
<tr>
<td>6/20</td>
<td>8</td>
<td>276.5</td>
<td>275.1</td>
<td>255.3</td>
<td>257.7</td>
<td>174.2</td>
<td>174.2</td>
<td>155</td>
<td>155.6</td>
</tr>
<tr>
<td>6/20</td>
<td>9</td>
<td>276.5</td>
<td>275.6</td>
<td>255</td>
<td>258.2</td>
<td>175.8</td>
<td>175.6</td>
<td>154.9</td>
<td>156.7</td>
</tr>
<tr>
<td>6/20</td>
<td>10</td>
<td>280.6</td>
<td>287.8</td>
<td>259.4</td>
<td>259.9</td>
<td>177</td>
<td>177.4</td>
<td>156.9</td>
<td>157.1</td>
</tr>
</tbody>
</table>

credible intervals are large, often including zero. The source strengths estimated using the 100m grid and 10 basis functions shown are shown in Figure 4.12 (SPod A) and Figure 4.13 (SPod B). The strongest source occurred on the afternoon of June 19th at grid cell 159 to the southwest of the sensor location within the facility being monitored and the source estimates from the two SPods resulted in similar temporal profiles. On June 20th, there is evidence from both SPods of a relatively constant source at grid cell 143 although the strength of the source is smaller than the one observed on June 19th. On the other days, there is some discrepancy between the estimates from the two SPods. On June 17th, both estimates indicate that the emission source is south of the sensor but the distance from the measurement location varies. The discrepancies between the source estimates on June 17th and 18th may be due to a source to the south or southeast that is outside the domain considered. The goal of this field study was to monitor the facilities located between the sensor and the Ohio River in the northwest corner of the map, so the sources to the southeast were not of primary interest.
**Figure 4.10** SPod A and B measurements compared to the predicted values using 100 m grid cells and 9 basis functions.

**Figure 4.11** Observed versus predicted for both SPods using 100 m grid cells and 9 basis functions.
Figure 4.12  Estimated source locations (left) and temporal profiles (right) from SPod A, using 100 m grid and 10 bases. Labeled values correspond to the grid cell IDs. Left: colors represent integrated source strengths. Right: red lines represent the medians of the posterior samples of the source strengths for each grid cell. The shaded areas represent the 80% credible intervals and are only shown for the labeled sources.
Figure 4.13 Estimated source locations (left) and temporal profiles (right) from SPod B, using 100 m grid and 10 bases. Labeled values correspond to the grid cell IDs. Left: colors represent integrated source strengths. Right: red lines represent the medians of the posterior samples of the source strengths for each grid cell. The shaded areas represent the 80% credible intervals and are only shown for the labeled sources.
4.5 Discussion and Conclusion

In this chapter, we address the problem of estimating temporally and spatially varying emission sources using a single near-source air pollutant sensor. We use high frequency wind measurements to calculate backward particle trajectories and propose a Bayesian hierarchical model for the sensor measurements as a function of the backward trajectories. We show through a simulation study that our model is able to accurately estimate source strengths and locations given favorable meteorological conditions.

We analyze four days of measurements from two SPod sensors located on the perimeter of an industrial facility in Louisville, Kentucky. Multiple sources emitting over varying time spans were detected within the four days, with the strongest source occurring during the afternoon of June 19th, southwest of the sensor location. Due to the use of only one sensor the credible intervals of the source estimates are large.

A variety of extensions could be implemented to improve source estimates. The addition of more sensors along the opposite edge of the facility would enable triangulation and reduce uncertainty. Estimates might also be improved by incorporating the output from a forward dispersion model, with source locations chosen based on the results from the backward trajectory model. Additionally, a forward dispersion model that accounts for the wind flow around physical structures in the domain could further refine our results. Finally, modeling multiple sensors simultaneously using a multivariate framework would allow information to be shared between SPods and could lead to smaller credible intervals for our parameters.
REFERENCES


A.1 Computing Details

We estimate the parameters through MCMC sampling using the R programming language. We draw 25,000 MCMC samples and discard the first 5,000 as burn-in and monitor convergence using trace plots of the deviance as well as several representative parameters. Having ensured a continuous, differentiable and increasing quantile function, we calculate the likelihood by inverting the quantile function. Our formulation of the likelihood can be inverted analytically when $\tau$ is less than $\tau_L$ or greater than $\tau_U$. When $\tau_L < \tau < \tau_U$ the quantile function is a polynomial in $\tau$ and can be inverted either by finding the roots of the cubic polynomial (for I-splines with order 3) or using Halley’s root finding algorithm [Hansen and Patrick, 1976], then the likelihood is calculated using (3.2).

To ensure the degree of smoothness at the thresholds matches the degree of smoothness
at the internal knots we sample $\theta_{1,p}$, $\theta_{M,p}$, $\alpha_L$ and $\alpha_U$, and define $\theta_{2,p}$, $\theta_{M-1,p}$, $\sigma_{L,p}$ and $\sigma_{U,p}$ in accordance with Proposition 1 and Theorem 1. The latent parameters $\theta_{m,p}^*$ were given Gaussian process priors with mean $\mu_{m,p}^*$ and spatial variance $\eta_{m,p}^2$ and the nugget variance was fixed: $\lambda_{m,p}^2 = 0.01(\eta_{m,p}^2)$. The prior parameters were given conjugate hyper-priors. For the simulation study $\mu_{m,p}^* \sim N(-3, 1)$ and $\eta_{m,p}^{-2} \sim \text{Gamma}(0.5, 0.005)$, except for those corresponding to the unconstrained parameters $\mu_0^* \sim N(0, 10)$. For the application $\eta_{m,p}^{-2} \sim \text{Gamma}(0.1, 0.1)$ was used. For both the simulation study and application an exponential covariance function was used with a range fixed at 0.5. The prior parameters were updated using a Gibbs update from the full conditionals. The latent parameters were updated using Metropolis-Hastings with the step size tuned to have an acceptance rate between 0.3 and 0.7.

Latent parameters were also used to sample the tail shape parameters with $\alpha = \log(\alpha^* - 0.4) - \log(0.4 - \alpha^*)$ and $\alpha^* \sim \mathcal{GP}(\mu_\alpha, \eta_{\alpha}^2)$. This ensured that the range of $\alpha$ was constrained to $(-0.4, 0.4)$. The upper bound prevents invalid quantile functions and the lower bound helps ensure convergence. Conjugate hyper-priors were used in both the simulation study and application with $\mu_\alpha \sim N(0, 1)$ and $\eta_{\alpha}^{-2} \sim \text{Gamma}(2, 0.1)$. The prior parameters are updated by using a Gibbs Sampling step. Theorem 1 is used to update the corresponding $\theta$s to ensure the differentiability of the quantile functions.

The tail scale parameters, $\sigma_{p}$, were updated using Proposition 1. The knots for the I-splines were evenly spaced between 0 and 1.

For the GAUS method, in the simulation study we draw 25,000 MCMC samples and discard the first 5,000 as burn-in and monitor the trace plot of the deviance for convergence. In the application, the GAUS method converged more slowly, to adjust 35,000 samples were drawn and 15,000 were discarded as burn-in. A fixed rate of 0.5 for the exponential covariance was also used for the GAUS method. In both the simulation study and application all parameters were given conjugate hyper-priors with $N(0, 100)$ used as the hyperprior for the means of the Gaussian processes. The hyper-priors for the spatial precision are $\text{Gamma}(0.5, 0.005)$ in the simulation study and $\text{Gamma}(0.1, 0.1)$ in the application.
A.2 Differentiability at $\tau_U$

Let $Y$ have a quantile function as defined in (2.1) and (2.2) with an I-spline basis order greater than $q + 1$ and a density that is continuous and $(q - 1)^{th}$ order differentiable at $Q(\tau_U)$. If $\alpha_U$ is constrained so that Eq. A.1 does not result in $\theta_{M-q,p} < 0$, then $Y$ has a density that is $q^{th}$-order differentiable at $Q(\tau_U)$ for any $x \in \mathbb{R}^+$ if and only if

$$\theta_{M-q,p} = \frac{1}{I_{M-q}^{(q+1)}(\tau_U)} \left\{ \frac{\sigma_{U,p}}{\alpha_U(\tau_U - 1)^{q+1}} \left( -\alpha_U - q \right)_{q+1} - \sum_{m=M-q+1}^{M} \theta_{m,p} I_{m}^{(q)}(\tau_U) \right\} \quad (A.1)$$

where $I_{M-q}^{(q+1)}(\tau_U)$ is the $(q + 1)^{th}$ order derivative of the $(M - q)^{th}$ I-spline basis function, $(-\alpha_U - q)_{q+1} = \prod_{j=0}^{q}(-\alpha_U - j)$.

A.3 Proofs

Proof of Proposition 1

Proof. We will only prove the case for the lower tail, the proof for the upper tail is equivalent. Given the assumptions we have already shown $Q'(\tau) > 0$ for all $\tau$. Thus the density exists and is given by $f(y) = \frac{1}{Q(Q^{-1}(y))}$. The density is continuous at $\tau_L$ if and only if $Q'(\tau)$ is continuous at $\tau_L$. By definition only a single I-spline basis function has a non-zero derivative at $\tau_L$: $I_{1}^{(1)}(\tau_L)$. Therefore, the following is a necessary and sufficient condition for a continuous density for any value of $x_p \geq 0$:

$$Q'(\tau_L) = \sum_{p=1}^{P} \sigma_{L,p} x_p \tau_L^{-1} = \sum_{m=1}^{M} \sum_{p=1}^{P} \theta_{m,p} x_p I_{m}^{(1)}(\tau_L) = \sum_{p=1}^{P} \theta_{1,p} x_p I_{1}^{(1)}(\tau_L) \quad (A.2)$$

Now since $x_p \geq 0$ is defined arbitrarily, take $x_p = 0$ for all $p \neq 1$, that is $x_p = 0$ for all predictors other than the intercept term, $x_1 = 1$. Then (A.2) is equivalent to

$$\theta_{1,1} = \frac{\sigma_{L,1}}{\tau_L * I_{1}^{(1)}(\tau_L)} \quad (A.3)$$

Similarly, for $p > 1$ take $x_p \neq 0$ for some $q$ and $x_p = 0$ for all $p > 1$ and $p \neq q$. Then (A.2) and (A.3) are equivalent to
\[ \theta_{1,q} = \frac{\sigma_{L,q}}{\tau_L * I_1(\tau_L)}. \]  

(A.4)

Hence we have proved Proposition 1.

Proof of Theorem 1

Proof. Let \( Y \) have a quantile functions as defined in Eq. 2.1 (main text) with an I-spline basis order greater than \( q + 1 \) and a density that is \( (q - 1) \)th order differentiable. Given \( \tau \leq \tau_L \),

\[ \beta_p(\tau) = \theta_{1,p} - \frac{\sigma_{L,p}}{\alpha_L} \left( \frac{\tau}{\tau_L} \right)^{-\alpha_L} - 1 \]

and \( Q(\tau) = \sum_{p=1}^{P} x_p^\beta_p(\tau) \). The density of \( Y \) is \( q \)th order differentiable if and only if \( Q(\tau) \) is \( (q + 1) \)th order differentiable. By definition \( Q(\tau) \) is \( (q + 1) \)th order differentiable at all points except \( \tau_L \) and \( \tau_U \). \( Q(\tau) \) is \( (q + 1) \)th order differentiable at \( \tau_L \) if and only if,

\[ \sum_{p=1}^{P} x_p \sum_{m=0}^{M} \theta_{m,p} I_{m}^{(q+1)}(\tau_L) = \sum_{p=1}^{P} x_p \beta_p^{(q+1)}(\tau_L) \]  

(A.5)

\[ I_{m}^{(q+1)}(\tau_L) = 0 \] for \( m = 0 \) and \( m > q + 1 \) so eq. A.5 is equivalent to

\[ \sum_{p=1}^{P} x_p \sum_{m=1}^{q+1} \theta_{m,p} I_{m}^{(q+1)}(\tau_L) = \sum_{p=1}^{P} x_p \beta_p^{(q+1)}(\tau_L) \]  

(A.6)

Because \( \beta(\tau) \) is a polynomial in \( \tau \) we can write

\[ \beta_p^{(q+1)}(\tau) = \frac{-\sigma_{L,p}^{\alpha_L}}{\alpha_L} \tau^{-\alpha_L-q-1} \prod_{j=0}^{q} (-\alpha_L - j) \]  

(A.7)

\[ \beta^{(q+1)}(\tau_L) = \frac{-\sigma_{L,p}}{\alpha_L \tau_{L+1}^{(-\alpha_L - q)_{q+1}}} \]  

(A.8)

Now since \( x_p \geq 0 \) is defined arbitrarily, take \( x_p = 0 \) for all \( p \neq 1 \), that is \( x_p = 0 \) for all predictors other than the intercept term, \( x_1 = 1 \). We further start with the case with \( q = 1 \). By proposition 1, \( \theta_{1,p} \) can be written as a function of \( \sigma_{L,p} \) and eq. A.6 is satisfied if and only if
\[ \theta_{1,1}I_1^{(2)}(\tau_L) + \theta_{2,1}I_2^{(2)}(\tau_L) = \sigma_L \tau_L^{-2}(-\alpha_L - 1) \]  
(A.9)

\[ \theta_{2,1} = \frac{1}{I_2^{(2)}(\tau_L)} \left[ \sigma_L \tau_L^{-2}(-\alpha_L - 1) - \theta_{1,1}I_1^{(2)}(\tau_L) \right] \]  
(A.10)

Similarly, for \( p > 1 \) take \( x_p \neq 0 \) for some \( q \) and \( x_p = 0 \) for all \( p > 1 \) and \( p \neq q \). Then (A.6) and (A.9) are equivalent to

\[ \theta_{2,p} = \frac{1}{I_2^{(2)}(\tau_L)} \left[ \sigma_{L,p} \tau_L^{-2}(-\alpha_L - 1) - \theta_{1,p}I_1^{(2)}(\tau_L) \right] \]  
(A.11)

We have thus shown that we can ensure 1st order differentiability of the density function of \( Y \) at \( Q(\tau_L) \) by constraining \( \theta_{1,p} \) and \( \theta_{2,p} \) as functions of \( \sigma_{L,p} \) and \( \alpha_L \) for all \( p \). Returning to the more general case, given a density of \( Y \) that is \((q-1)\text{th}\) order differentiable, we again take \( x_p = 0 \) for all \( p \neq 1 \). Then the density of \( Y \) is \( q\text{th} \) order differentiable if and only if

\[ \sum_{m=1}^{q+1} \theta_{m,1}I_m^{(q+1)}(\tau_L) = -\frac{\sigma_{L,1}}{\alpha_L \tau_{q+1}(-\alpha_L - q)} \]  
(A.12)

\[ \theta_{q+1,1} = \frac{1}{I_{q+1}^{(q+1)}(\tau_L)} \left[ -\frac{\sigma_{L,1}}{\alpha_L \tau_L}(-\alpha_L - q) \tau_{q+1} - \sum_{m=1}^{q} \theta_{m,1}I_m^{(q+1)}(\tau_L) \right] \]  
(A.13)

Similarly, for \( p > 1 \) take \( x_p \neq 0 \) for some \( q \) and \( x_p = 0 \) for all \( p > 1 \) and \( p \neq q \). Then (A.6) and (A.12) are equivalent to

\[ \theta_{q+1,p} = \frac{1}{I_{q+1}^{(q+1)}(\tau_L)} \left[ -\frac{\sigma_{L,p}}{\alpha_L \tau_{q+1}(-\alpha_L - q)} \tau_{q+1} - \sum_{m=1}^{q} \theta_{m,p}I_m^{(q+1)}(\tau_L) \right] \]  
(A.14)

We have thus proved Theorem 1.
A.4 Expectation and Covariance

\[
E[Y_t(s)|\Theta(s), X_t(s)] = \int_0^1 Q_Y[\tau|\Theta(s), X_t(s)]d\tau
\]
\[
= \sum_m \sum_p \theta_{m,p}(s)x_{p,t}(s)G_m
\]
\[
+ \sum_p \left( \tau_L \theta_{1,p}(s)x_{p,t}(s) + (1 - \tau_U)x_{p,t}(s) \sum_m \theta_{m,p}(s) + \frac{\sigma_{L,p}(s)x_{p,t}(s)\tau_L}{\alpha_L(s) - 1} + \frac{(1 - \tau_U)\sigma_{U,p}(s)x_{p,t}(s)}{1 - \alpha_U(s)} \right)
\]

where \( G_m = \int_{\tau_L}^{\tau_U} I_m(\tau)d\tau \). As the last two terms approach zero, the distribution of the marginal expectation of \( Y \) becomes a linear combination of the \( \theta_{m,p} \) which have log normal distributions.

The marginal expectation of \( Y(s) \), marginalizing over \( \theta_{m,p} \) and \( \sigma \) is

\[
E[Y_t(s)|X_t(s)] = \sum_m \sum_p \mu_{m,p}x_{p,t}(s)G_m
\]
\[
+ \sum_p \left( \tau_L \mu_{1,p}x_{p,t}(s) + (1 - \tau_U)x_{p,t}(s) \sum_m \mu_{m,p} \right) +
\]
\[
+ \sum_p \left( \tau_L^2 \mu_{2,p}I_2(\tau_L)x_{p,t}(s)E \left[ \frac{1}{\alpha_L - 1} \right] + (1 - \tau_U)^2 \mu_{M,p}I_1(\tau_U)x_{p,t}(s)E \left[ \frac{1}{1 - \alpha_U} \right] \right)
\]

Based on the model structure, given \( \Theta, Y_t(s) \) and \( Y_t(s') \) are independent. Thus the conditional covariance is zero and

\[
E[Y_t(s)Y_t(s')|X_t(s)\Theta(s)] = E[Y_t(s)|X_t(s)\Theta(s)]E[Y_t(s')|X_t(s')\Theta(s')]
\]
\[
E[Y_t(s)Y_t(s')|X_t(s)] = E_\Theta\left[ E[Y_t(s)|X_t(s), \Theta(s)] E[Y_t(s')|X_t(s'), \Theta(s')] \right] \\
= \sum_m G^2_{m} x_{p,t}(s)x_{p,t}(s')[\Sigma_{m,p}(s, s') + \mu^2_{m,p}] \\
+ \sum_m \sum_p \{x_{p,t}(s)G_{m}\mu_{m,p}(s)\} \sum_{(l,k) \neq (m,p)} \{x_{k,t}(s')G_{l}\mu_{l,k}\}
\]