A Nonlinear and Non-Gaussian State-Space Model for Censored Air Pollution Data

Craig J. Johns Robert H. Shumway *

May 29, 2003

Abstract

Lidar technology is used to quantify airborne particulate matter less than 10 microns in diameter (PM_{10}). This spatio-temporal lidar data on PM_{10} may be subject to censoring due to detection limits. We modify a nonlinear and non-Gaussian state-space model to accommodate data subject to detection limits and outline strategies for Markov-Chain Monte Carlo estimation estimation and filtering. The methods are applied to spatio-temporal lidar measurements of dust particle concentrations.

KEYWORDS: lidar, censored data, state-space model

1 INTRODUCTION

Particulate matter less than 10 microns in aerodynamic diameter, PM_{10}, is a regulated National Ambient Air Quality Standard (NAAQS) pollutant because of the effect small particles have on visibility and human respiratory health and (U.S. E.P.A. 1997). Quantifying PM_{10} emissions from all sources within a given air quality management district is essential for developing long-term PM_{10} air quality attainment plans. One type of PM_{10},

*Craig J. Johns is Assistant Professor, Mathematics Department, University of Colorado, Denver, CO, 80217-3364 and Robert H. Shumway is Professor, University of California, Davis, CA, 95616.

fugitive dust, is generated from moving point and nonpoint sources such as activity at construction sites, natural wind erosion of exposed soil, transportation sources, and agricultural operations (Cowherd 1993).

1.1 Lidar Data

Holmén et al. (1998) proposed a remote sensing technique involving light detection and ranging, lidar, as a means for measuring PM_{10} emissions from agricultural moving point sources. In a nutshell, lidar technology uses a laser signal and telescope receiver to measure PM_{10}. The characteristics of the return signal, or back-scatter, from the laser determine the characteristics of the concentration of PM_{10} in the dust plume, (Measures, 1984). For the lidar machine that collected the data described in this paper, the backscatter signal is sampled as the sum of 25 pulses of the 50 Hz laser, giving a one-dimensional 'snapshot' or 'time slice' of the dust plume generated by the tractor approximately every 0.5 seconds. Each time slice is a spatial series of data, with individual data points evenly spaced 4.95 meters apart on the downwind side of the field along the lidar beam trajectory. Figure 1 shows a single time slice.

The return signal is transformed into data by way of a digitizer. The lidar digitizer converts electronic signal into digits that are stored by a computer. The digitizer is a scalable map from the set of all signal strengths onto a finite set of digits. In order to capture all signals without any censoring, the maximum return signal of a data collection period must be known *a priori*. Even so, if the digitizer map is spread over too large a range, then information is lost as many different signals are mapped to the same digit, a type of roundoff error. On the other hand, if the digitizer upper bound is too low, then all signals above a threshold are mapped to the largest possible digit, and censoring results. Although both result in a loss of information, technician knowledge and trial runs are used in setting this parameter, so that some middle ground is achieved. We point out this characteristic of the lidar machine as it is very likely that any data set collected by this machine, and others like it, will routinely collect data sets which contain a nontrivial proportion of censored data.

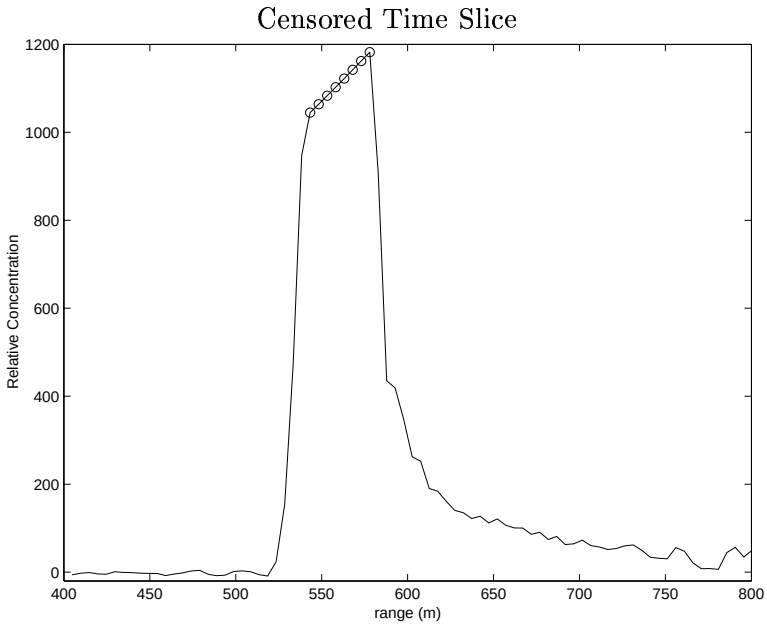

Figure 1: A single time slice which shows the relative concentration of PM_{10} over the ranges 400m to 800m. Measurements are every 4.95 meters apart along the range axis, time slices are approximately 0.5 seconds apart. Upper detection limits for values that have been censored are denoted by ∘.

Figure 1 shows a time slice where the backscatter signal was large enough to cause censoring at 8 points. The nonlinear trend in the censoring values is due to the transformation that maps the digitized values into relative concentration values. Table 1 gives a view of the extent of censoring in the data set that we describe below.

1.2 Data Collection

To collect PM_{10} emissions data from a moving point source, a lidar instrument was positioned on the downwind edge of the field where a tractor was using a disk harrow to prepare the soil after wheat harvest. Figure 2 shows a schematic diagram of the data collection. The lidar beam was pointed parallel to the field edge and roughly perpendicular to the wind direction. Repeated lidar measurements were made at a fixed elevation angle (2°) as the tractor tilled the field. The analysis herein focuses on comparing data from the six individual tractor path lines as described in Section 1.3 and corresponding to the data collection

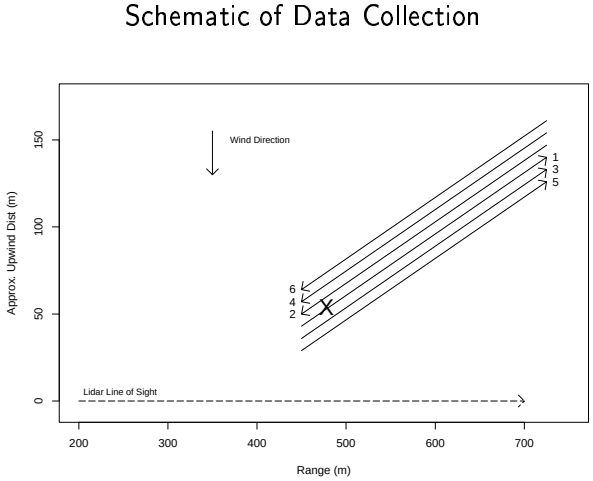

Figure 2: Schematic of the data collection showing prevailing wind direction, approximate tractor path, lidar beam trajectory and lidar instrument location. The "X" marks a tractor location that could result in data censoring due to high dust generation close to the lidar beam. The numbers 1-6 denote the sequence of tractor path lines excluding turns.

periods shown in Figures 2 and 3.

To refer to a spatial location, we use the term "range" to denote the distance from the lidar instrument to a point along the laser beam trajectory. The range corresponding to the location of the highest concentration in a time slice is the 'peak location.' Figure 1 shows a time slice with peak location at roughly 570 meters As indicated by the $\circ$ symbol plotted in the figure, the data was partially censored due to the upper detection limits of the lidar. This censoring occurred most dramatically at close ranges when the tractor was only a short distance upwind of the lidar beam line (Figure 2, "**X**") because the dust plume measured was most dense at this short upwind distance. Of the 1645 time slices collected, 924 (56.5%) contain at least one range point with a censored observation, and a total of 3531 (3.0%) censored values were recorded. While three time slices have as many as 17 range points where clipping occurred, most of the affected time slices have fewer than six censored points. The maximum number of censored points in any time slice in the lines we consider is 11.

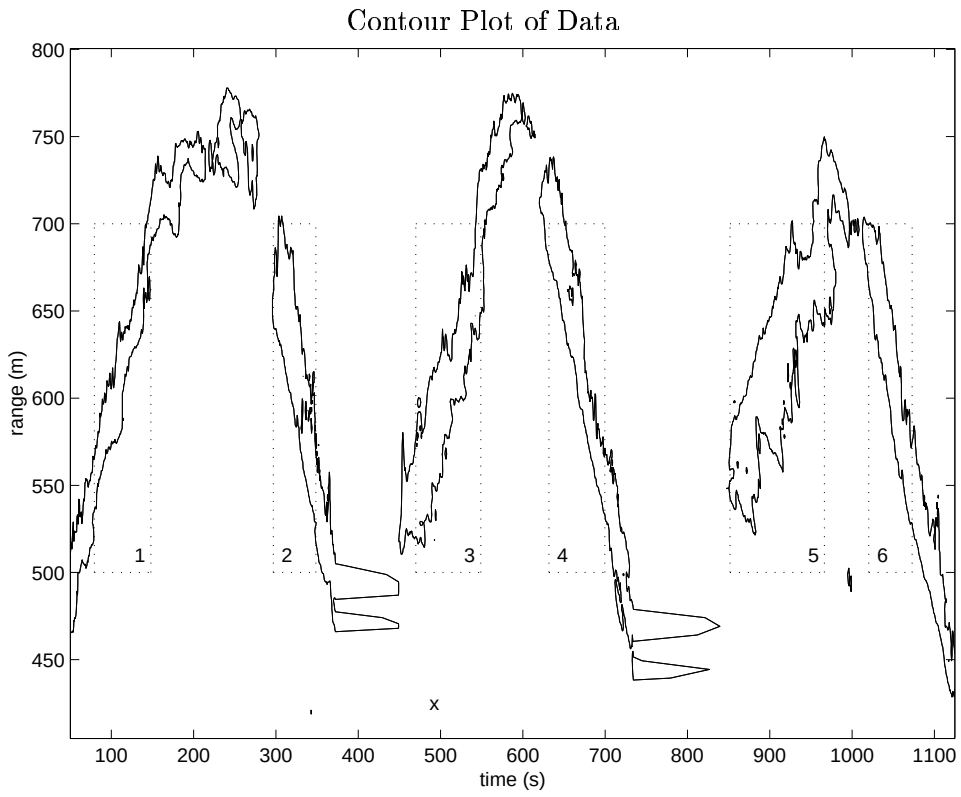

Figure 3: A single line contour plot of the data. The contour lines at a given time period show the approximate location of the dust plume. The boxed regions depict the approximate time periods for the six lines described in Section 1.3. The particular time period depicted in Figure 1 is denoted with an $\times$.

1.3 Comparable Sets

Preliminary to building a general model suitable for government agency predictions and estimations, a large number of data sets such as the one described above will need to be collected and analyzed. Prior to all this effort and cost, we wish to determine if the plume generated by a single tractor operation is changing substantially over time or space. If there is little variability in the plume characteristics over some unit smaller than a complete field operation, then there is hope that future field tests can be reduced to measurements over the new unit. In order to test whether or not the field operation is comprised of equivalent smaller units in mind, we follow Johns et al. (2001) to define comparable sets, or *lines*.

Censoring Information

Line	Duration (seconds)	No. of Time Slices	No. of Censored Time Slices	Total Censored Points
Line 1	68.7	122	58	118
Line 2	52.3	98	79	391
Line 3	79.5	140	100	407
Line 4	68.7	120	109	278
Line 5	115.4	203	141	452
Line 6	53.3	95	70	501
Total	437.9	771	557	1885

Table 1: Censoring information for the six lines described in Section 1.3 and depicted in Figures 2 and 3. The lines are roughly described as the set of time slices where the center of the dust plume is between 500 and 700 meters in range.

`line`: the data collection period corresponding to a straightaway section of the tractor path.

This definition is natural because typical tractor movement during field preparation is comprised of parallel lines and turns. The lidar signals recorded during the tractor turns had smaller concentration levels than those time slices recorded while the tractor was traveling on a straight path. Furthermore, the dust plumes generated on the turns could not be easily assigned to an individual line. Therefore, our analysis only considered time slices corresponding to the straight sections of the tractor's path.

The lines defined herein are depicted in Figure 3, with the dashed boxes as approximate boundaries for the lines. Details of the lines in terms of length of time, number of time slices and censoring are summarized in Table 1. Specific determination of the lines as given in Figure 3 and is in part constrained by the data collection sub-periods. For safety reasons, the lidar was turned off for periods of time during tractor turns near the lidar trajectory. The off periods created obvious starting points for lines 1, 3 and 5 and ending points for lines 2, 4 and 6. The other end points are determined by the location of the center of the

plume. For practical purposes, the lines are defined as the time periods when the center of the plume is between 500 and 700 meters of range. The intention of the definition of the lines is to make them as similar as possible to facilitate line to line comparisons.

2 State-Space Models

A State-Space or Dynamic Linear Model refers to a wide class of hierarchical models that have been used to model data from such diverse fields as aerospace research (Kalman, 1960), economics (Shumway and Stoffer, 1982), medicine (Jones 1984), and geosciences (Hamill and Snyder, 2002). The general idea behind this hierarchical model is that the *"state"* of a system at time t can only be observed indirectly through observations that follow a random regression model.

2.1 Traditional State-Space Models

Traditionally, the random regression is of the form: $\mathbf{y}_t = H_t\boldsymbol{\theta}_t + \mathbf{v}_t$, where H_t is a $d \times p$ matrix of known coefficients (design matrix) which converts the current unobserved stochastic $p \times 1$ state vector $\boldsymbol{\theta}_t$ into the $d \times 1$ observation vector $\mathbf{y}_t$. This form is commonly refereed to as the observation equation. The observation error vectors, $\mathbf{v}_t$ and $\mathbf{v}_{t'}$ have common covariance, Ω, and are considered independent for $t \neq t'$. The random vector of unobserved states $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_n$ is of primary interest. Classical models assume that $\boldsymbol{\theta}_t$ follows a possibly non-stationary multivariate autoregressive process of order one: $\boldsymbol{\theta}_t = G_t\boldsymbol{\theta}_{t-1} + \mathbf{w}_t$ where G_t is a known $p \times p$ matrix. This form is the state equation. Classical models further assume errors in the state and observation equations are independent over time.

With knowledge of transition matrices, covariances and the initial mean, the well known Kalman filter recursions provide the best linear estimate (Brockwell and Davis, 1987, sec 12.1) of $\boldsymbol{\theta}_s$ given the information in the data up to time t. When $s = t$ the estimation is referred to as *filtering*. For the case $s > t$, the estimation is referred to as *prediction* or *forecasting*. When a *post hoc* analysis is suitable, the Kalman recursions are used to find the best linear estimator (the Kalman *smoother*) of $\boldsymbol{\theta}_t$ given all the data $\mathbf{y}_1, \ldots \mathbf{y}_n$ from a

data collection regime.

The linearity and normality assumptions in the classical model are often too restrictive for many modeling situations and a host of extensions have been studied. Meinhold and Singpurwalla (1989) describe some non-robust properties of the Kalman filter and suggest using errors that follow Student's-t distribution. In a similar vein, Kitagawa (1987) suggests numerical integration techniques to estimate nonstationary and non-normal state-space models. Kitagawa (1987) also encourages the use of heavy tailed distributions, such as t-distributions, to model nonstationary time series. However, these techniques are not well suited to vector processes. Shumway and Stoffer (1982) extended the Kalman recursions to the case when observations are missing and give an EM algorithm solution for the linear observation equation and linear state transition case.

2.2 Nonlinear and Non-Gaussian State-Space Models

A natural extension is to relax the linear dependence of $\boldsymbol{\theta}_t$ upon its predecessor by allowing the state equation to be nonlinear: $\boldsymbol{\theta}_t = \mathbf{g}_t(\boldsymbol{\theta}_{t-1}) + \mathbf{w}_t$, where each $\mathbf{g}_t$ is known (perhaps conditional upon some information). This formulation is natural in physical systems where states are modeled as evolving according to nonlinear differential equations.

Likewise, in many physical systems, data may be related to the state by a nonlinear function, $\mathbf{h}_t$; $\mathbf{y}_t = \mathbf{h}_t(\boldsymbol{\theta}_t) + \mathbf{v}_t$. In many situations it is reasonable to model, $\mathbf{g}_t = \mathbf{g}$ and/or $\mathbf{h}_t = \mathbf{h}$, meaning that the functions in the state and observation equations are not evolving in time, except through the evolution of the states. Harvey (1989) describes the extended or linearized Kalman filter as a potential solution in the case when the state and observation equations are nonlinear. However, the linearized Kalman filter is famous for instability.

Carlin, Poulson, and Stoffer (1992) (CPS hereafter) propose removing the normal errors assumption by modeling state and observation errors as normal, conditional upon nuisance parameters. Their hierarchical Bayesian approach also allows for nonlinearity in the models. Specifically they model the state transitions by assuming conditional normality of errors in the states and observations.

$$[\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1}, \Sigma, \lambda_t] \sim N(\, \mathbf{g}_t(\boldsymbol{\theta}_{t-1}),\ \lambda_t \Sigma\,), \tag{1}$$

and the observations as

$$[\mathbf{y}_t \mid \boldsymbol{\theta}_t, \Omega, \xi_t] \sim N(\, \mathbf{h}_t(\boldsymbol{\theta}_t),\ \xi_t \Omega\,), \tag{2}$$

where λ_t and ξ_t are time varying nuisance parameters with a known distributions, $\mathbf{g}_t$ maps the previous state onto the mean space of the current state and $\mathbf{h}_t$ maps the current state onto the mean space of the observations. These conditional normal distributions, when combined with the distributions of λ_t and ξ_t can produce a wide variety of distributions for the update errors and observations, yet are well suited for estimation via Markov Chain Monte Carlo (MCMC) methods.

Let $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n)$ denote the d-vectors of data collected in a sampling regime with time index $t = 1, 2, \ldots, n$ and let $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_n)$ be the accumulation of state vectors coterminous with the data vectors. The likelihood specified in CPS is:

$$p(\boldsymbol{\Theta}, \mathbf{Y} \mid \Sigma, \Omega) \ = \ q_0(\boldsymbol{\theta}_0 \mid \mu_0, \Sigma_0) \prod_{t=1}^{n} q_t(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1}, \Sigma) \prod_{t=1}^{n} p_t(\mathbf{y}_t \mid \boldsymbol{\theta}_t,\ \Omega) \tag{3}$$

where q_t and p_t are density functions such that $q_t(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1}, \Sigma) = \int f(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1}, \lambda_t \Sigma) \pi_{1_t}(\lambda_t) d\lambda_t$, $p_t(\mathbf{y}_t \mid \boldsymbol{\theta}_t,\ \Omega) = \int f(\mathbf{y}_t \mid \boldsymbol{\theta}_t,\ \xi_t \Omega) \pi_{2_t}(\xi_t) d\xi_t$, and f denotes the multivariate normal density function. The functions $\mathbf{g}_t$ and $\mathbf{h}_t$ are considered known or known conditional upon a nuisance parameter and are suppressed in (3) for notational convenience.

2.3 Modifications for Censored Data

In many situations, technologies which are invaluable for the collection of data are hampered in their efficacy due to detection limits. An abundance of statistical literature has been devoted to the issue. (See for example, Aitkin 1981, Gleit 1985, Johns et al. 2001, Schmee and Hahn 1979, Shumway et al. 1989, and Tanner 1996). Here we modify the CPS (1992) model in order to fit nonlinear observation equations for the censored data.

In the case where some observations have been censored by a detection limit, we suggest the following modification to facilitate computations. Specifically, we assume that the distributions of the observational errors for $\mathbf{y}_t$ are conditionally independent. That is, conditional upon $\boldsymbol{\theta}_t, \xi_t$ and Ω, y_{tk} is independent of $y_{tk'}$ for $k \neq k'$. One route toward this conditional independence is to force Ω in (2) to be a diagonal matrix with zeros in the

off-diagonal elements. This assumption facilitates computations required for evaluations of marginal and conditional distributions in the Gibbs Sampler. Thus each element of the second product on the right hand side of the distribution (3) is replaced with (Gelman et al. 1995)

$$\mathcal{L}_t(\boldsymbol{\theta}_t, \mathbf{y}_t, \mathbf{D}_t, \Omega, \boldsymbol{\delta}_t, \xi_t) \;=\; \prod_{k=1}^{d} p_{tk}(y_{tk} \mid \boldsymbol{\theta}_t,\; \xi_t\omega_k)^{(1-\delta_{tk})} P_{tk}(Y_{tk} \geq D_{tk} \mid \boldsymbol{\theta}_t,\; \xi_t\omega_k)^{\delta_{tk}} \qquad (4)$$

where D_{tk} is the upper detection value for Y_{tk}, $P_{tk}(\mathcal{A})$ denotes the probability of event $\mathcal{A}$ under p_{tk} and $\delta_{tk} = 0$ if y_{tk} is observed, and 1 if y_{tk} has been censored. This modification is strictly for computational simplicity, as it eliminates the need to calculate an integral of dimension equal to the number of censored observations in $\mathbf{y}_t$. Similar modifications can be made to cover the left censoring (lower detection limits) and "left and right" censoring (upper and lower detection limits) cases.

Primary interest lies in the distribution of the states and variance components Σ and Ω, given the data: $p(\boldsymbol{\Theta},\, \Sigma,\, \Omega \;\mid\; \mathbf{Y}) \;\propto\; p(\boldsymbol{\Theta}, \mathbf{Y} \mid \Sigma, \Omega)\pi(\Sigma)\pi(\Omega)$ where $\pi(\Sigma)$ and $\pi(\Omega)$ are prior distributions for Σ and Ω respectively. To understand the characteristics of this distribution and the behavior of the states (or other parameters) MCMC methods present a feasible approach. Inferences from the model can be made by sampling from the conditional distribution $[\boldsymbol{\Theta}, \boldsymbol{\theta}_0, \Sigma, \Omega, \boldsymbol{\lambda}, \boldsymbol{\xi} \mid \mathbf{Y}, \boldsymbol{\Delta}]$, where $\boldsymbol{\Delta} = \{\delta_{tk}\}$. Because of the hierarchical nature of the model, conditional distributions useful for Gibbs sampling are easily determined and are given int the Appendix.

3 Lidar Data Model

The hierarchical idea of defining the data collected by the lidar instrument at a given point in time as a function of unknown parameters is quite natural. Seinfeld and Pandis (1998) describe atmospheric pollutant plume evolution over time using differential equations. The non-standardized Gaussian curve is an approximate solution to these differential equations and provides a basis for modeling the average datum at a given time and range bin. We consider the parameters of the non-standardized Gaussian curve the states of interest and model their behavior over time with the state-space construction. The observation equation

is given by $y_{tk} = \mathbf{h}_k(\boldsymbol{\theta}_t) + v_{tk}$ where

$$h_k(\boldsymbol{\theta}_t) = \theta_{t1} + \theta_{t2} \exp\left(-\frac{(x_k - \theta_{t4})^2}{\theta_{t3}}\right), \tag{5}$$

x_k is the k^{th} range bin and the vector of errors $\mathbf{v}_t$ has probability density function p_t as in (3), with the modification that $\Omega = \omega I$, where I is the identity matrix. The parameters indexed by $\boldsymbol{\theta}_t$ are interpreted as follows: θ_{t1} is proportional to the ambient PM_{10} level, θ_{t2} is related to the maximum concentration PM_{10} in the lidar line of sight, θ_{t3} is related to the plume width or dispersion and θ_{t4} is related to the range location of the center of the plume. Of primary interest for this data is the ratio $\tau_t = \theta_{t2}/\theta_{t3}$ which is proportional to the integrated total mass of PM_{10} in the lidar line of sight at time t.

Further, we assumed that the observation errors are normally distributed, rather than conditionally normal as in (2). This is formally accomplished by taking the prior distribution on ξ_t to be a point mass distribution. Prior distributions for the observation error variance, ω, were taken to be inverse-gamma with mean 6400 and standard deviation 320.

The state equation is modeled as a random walk with non-Gaussian transitions;

$$[\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}, \lambda_t, \Sigma] \sim \mathcal{N}(\, \boldsymbol{\theta}_{t-1},\, \lambda_t \Sigma\,) \tag{6}$$

where λ_t are the nuisance parameters which allow for non-Gaussian transition. The λ_t will be relatively large when state-to-state transitions are relatively large, thus λ_t can be thought of as a measure of wind variability from plume generation until the plume is measured in the lidar beam. These parameters are given identical priors; inverse-gamma with mean 1 and standard deviation 0.25.

For the prior distribution on $\boldsymbol{\theta}_0$ we use a normal distribution with mean $\hat{\boldsymbol{\theta}}_0$ and variance-covariance matrix $\mathcal{S}_0$ where $\hat{\boldsymbol{\theta}}_0$ is found by maximizing the likelihood function for the data in the time period just previous to the first time period of interest, and $\mathcal{S}_0$ is the corresponding approximate asymptotic variance.

The Σ parameter models the conditional variance for the state transitions. As these may be affected by changes in wind direction, we use an inverse-Wishart prior with 20 degrees of freedom. The centering matrix, $\mathcal{V}$, is formed so that $\mathcal{V}_{ij} = \rho_{ij} s_i s_j$, where

$$s_1 = 15 \qquad s_2 = 1000 \qquad s_3 = 3 \qquad s_4 = 2$$
$$\rho_{ii} = 1 \qquad \rho_{12} = 0 \qquad \rho_{13} = 0 \qquad \rho_{14} = 0$$
$$\rho_{23} = -0.50 \quad \rho_{24} = -0.5 \quad \rho_{34} = 0.5.$$

As previously mentioned, a set of conditional distributions useful for the Gibbs sampler are easy to write down (Appendix), but these are difficult to sample because of the nonlinearity in the observation equations. We used Metropolis-Hastings steps in the Gibbs sampler to estimate the model.

We ran the Markov-Chain for each line separately. The prior distributions described above were consistent for each line. Five chains for each line (Raftery and Lewis, 1996) were burned in for 2,000 iterations, after which $\hat{R}$ (Gelman, 1996) and visual diagnostics were used to verify that the parallel chains had converged. The means of the last 1,000 iterations of each of these preliminary chains were combined to yield starting values for the final chain that was run for 20,000 iterations. A Monte-Carlo sample of size 1000 was taken as a random sample from these 20,000 possibilities,

The chains seemed to converge quickly, as evidenced by Figure 4. The figure shows the observation variance for the last range bin against the corresponding iteration number in the Markov Chain. The five lines correspond with different starting values for the other parameters, in particular the values of $\boldsymbol{\theta}_t$ and Σ. In all cases, the starting values for λ_t were equal to 1 for all t. Visual and numeric measures of convergence for other parameters, in this and other lines, likewise show fast convergence of the Markov-Chain.

4 Results

4.1 Differences in Integrated Total Mass

The main thrust of the data collection is the hope that the lines are sufficiently homogeneous so that data from a few lines of tractor movement will be Representative for all other possible lines for a given field. Toward this end we calculated the the integrated total mass τ_{itk} (Section 3) where i denotes the line and t represents the time slice in the line and k is the MC sample index. The box-plots in Figure 5 displays the variability of the medians,

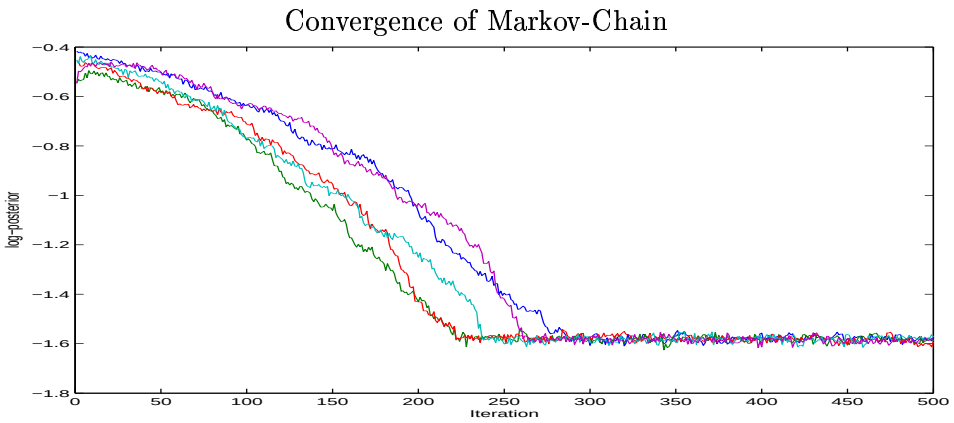

Figure 4: Evaluation of the log-posterior by Gibbs iterate for the five different initial value sets for the Markov-Chain of Line 6. The starting values were different for each of the five chains in every respect except the values for λ_t.

$\tilde{\tau}_{it\cdot}$, for the six lines. Line 2 has a large variance in the medians because of a few time slices for which the observation model (5) does not fit the data well. It is likely that Line 5 has large variance because of variability in meteorological variables described below.

We also calculated the average integrated total mass, $\bar{\tau}_{i\cdot k}$, for the lines, and from the MC-samples for each line and estimated the posterior density functions for the average integrated total mass, $\bar{\tau}_{i\cdot\cdot}$ for each of the six lines. The estimated posterior distribution densities shown in Figure 6 are sufficiently different from each other to cast serious doubt on the hypothesis that the lines are homogeneous. We confirmed this with a cell means model where a cell is interpreted as a line and replicates came from the MC sample. We compared the cell means model with restricted model that allowed for differences due to tractor-direction and that had a constant offset for upwind distance. This restricted model was significantly different from the cell-means model, ($F_{(3,5994)} > 500$, p-value $< .0001$).

5 Discussion

The results regarding Average Integrated Mass in Section 4 are slightly different than those presented in Johns et al. (2001). They found that some lines were equivalent in average integrated total mass, but the relationships they found could not be scientifically explained.

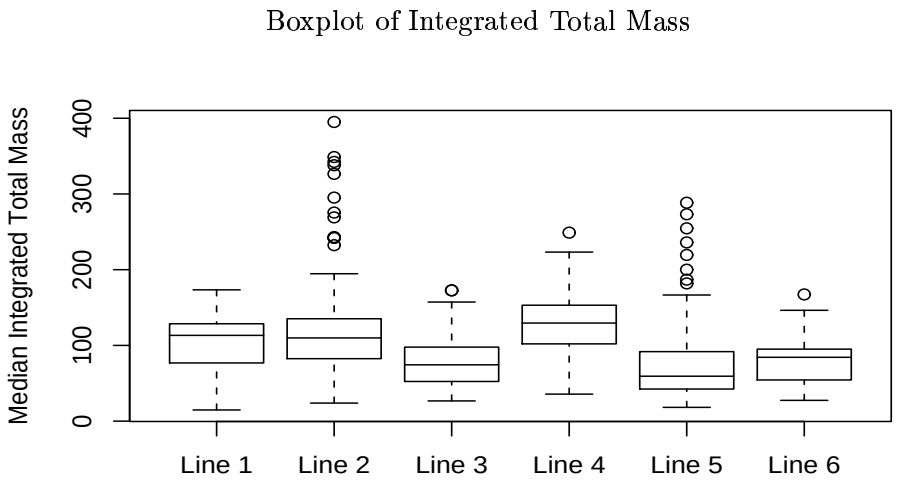

Figure 5: Boxplots of the integrated total mass values for the six lines. The "replications" in a boxplot are $\tilde{\tau}_{it}$..

Johns et al. (2001) used a simpler model that is subject to over-smoothing. Here we have described a far more flexible and non-Gaussian model.

The non-Gaussianity of the model presented here is important for the state transitions as there are time points in each line where

The information in the likelihoods, (4) using observation equation (5) seems to be great compared to the random walk updates (6) as the median values from the posterior sample are very similar to the maximum likelihood estimates. In part, this is due to the relatively vague priors in Section 3. Figure 7 shows the maximum likelihood estimates compared with the posterior medians. This suggests that the information in $\mathbf{y}_t$ about the parameter vector $\boldsymbol{\theta}_t$ is nearly independent of the information in $\mathbf{y}_{t+1}$ regarding $\boldsymbol{\theta}_{t+1}$ in the posterior distribution.

Even though the likelihood functions (4) seem to exert most of the force on the posterior distributions, there are cases involving many time slices where the likelihood function does not give reasonable estimates for θ_t. In one particular case in Line 2, the likelihood estimate for the plume maximum is nearly 20 times the estimate from the posterior. Thus, to simplify the model down to a collection of independent likelihoods is likely to subject the analysis to undue variations exacerbated by censoring. On the other hand, this problem may be

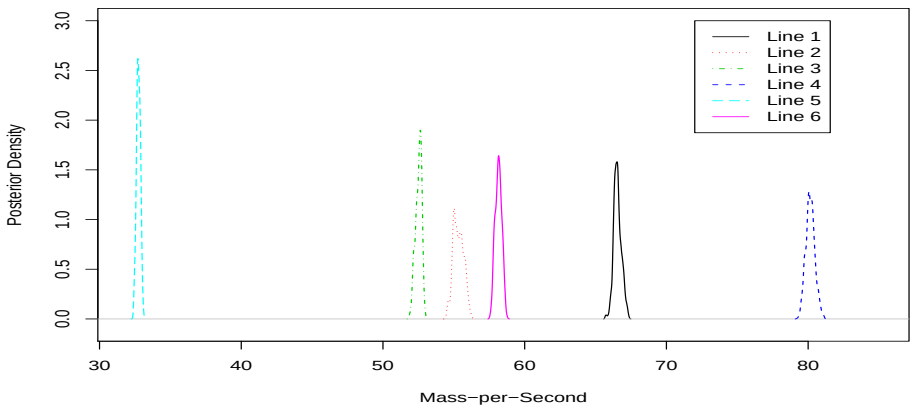

Figure 6: Estimates of the posterior densities for average total mass per second (up to a proportionality constant) based on the Monte-Carlo sample.

mitigated by better data models in (5).

The need for better models in (5) is great. The time slice depicted in Figure 1 is a typical example of how the values the right of the plume center are larger than those on the left of the plume center. Other models that are better approximations to solutions of the differential equations which model dust plume behavior (Seinfeld and Pandis, 1998) used in place of our $\mathbf{h}_t$ in (5), are needed to improve on the fitting of this data. One attempted improvement that we have tried is the sum of two non-standardized Gaussian curves. The problem with an observation model of this type is that the likelihood functions (4) frequently have multiple modes.

A further improvement is to add a model that incorporates meteorological variables such as wind speed and direction. It was noted from other data sources that during the time period corresponding with Line 5 of the data collection, the wind speed and variability in direction was much greater than during the other lines of data collection. As modeled above, the time slices in Line 5 typically have larger dispersion coefficients (θ_{t3}) and smaller maximum concentration coefficients (θ_{t2}) than in other lines.

As stated in Section 1.1, the lidar machine has an inherent high probability to censor

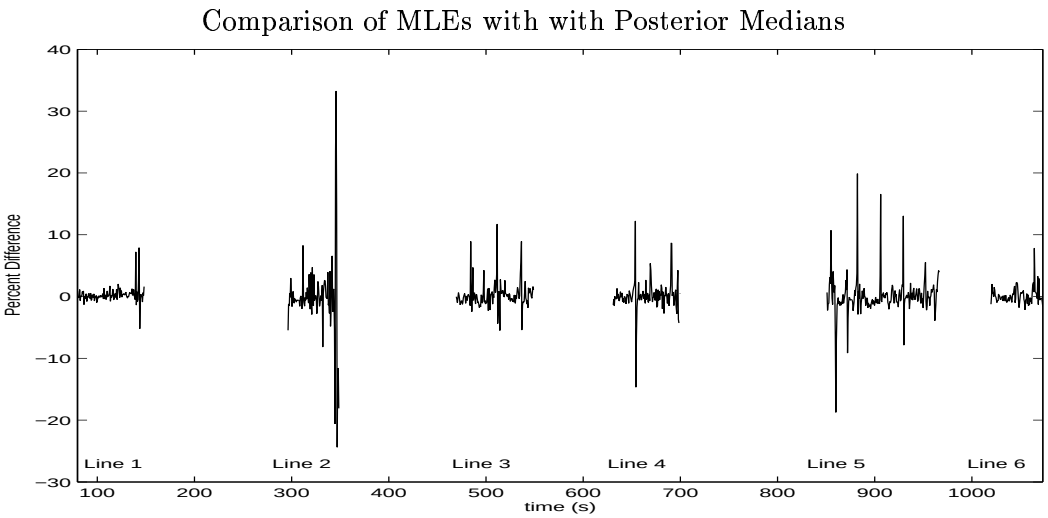

Figure 7: The percentage change in the estimates of integrated mass τ_t. The comparison is between the maximum likelihood estimate and the posterior median.

data. Thus, a major contribution of this paper is to modify current state-space methodology to provide a framework for the case for time-series observations that are subject to upper detection limits. These methods are also relevant to other spatio-temporal measurements where lower detection limits are common.

Appendix

Bayes rule creates the following simplifications in the conditional distributions under the specifications of the model given in CPS (1992).

$$
\begin{aligned}
[\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{j\neq t}, \Sigma, \boldsymbol{\lambda}, \Omega, \boldsymbol{\xi}, \mathbf{Y}, \boldsymbol{\Delta}] &\sim [\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_{t+1}, \lambda_t, \lambda_{t+1}, \Sigma, \xi_t, \Omega, \mathbf{y}_t, \boldsymbol{\delta}_t] \\
[\lambda_t \mid \boldsymbol{\Theta}, \boldsymbol{\theta}_0, \Sigma, \Omega, \lambda_{j\neq t}, \boldsymbol{\xi}, \mathbf{Y}, \boldsymbol{\Delta},] &\sim [\lambda_t \mid \boldsymbol{\theta}_t, \boldsymbol{\theta}_{t-1}, \Sigma] \\
[\xi_t \mid \boldsymbol{\Theta}, \boldsymbol{\theta}_0, \Sigma, \Omega, \boldsymbol{\lambda}, \xi_{j\neq t}, \mathbf{Y}, \boldsymbol{\Delta},] &\sim [\xi_t \mid \boldsymbol{\theta}_t, \Omega, \mathbf{y}_t, \boldsymbol{\delta}_t] \\
\text{for } 1 \leq t \leq n, & \quad \text{and} \\
[\boldsymbol{\theta}_0 \mid \boldsymbol{\theta}_{j\neq t}, \Sigma, \boldsymbol{\lambda}, \Omega, \boldsymbol{\xi}, \mathbf{Y}, \boldsymbol{\Delta}] &\sim [\boldsymbol{\theta}_0 \mid \boldsymbol{\theta}_1, \Sigma_0, \Sigma, \lambda_1] \\
[\Sigma \mid \boldsymbol{\Theta}, \boldsymbol{\theta}_0, \boldsymbol{\lambda}, \Omega, \boldsymbol{\xi}, \mathbf{Y}, \boldsymbol{\Delta}] &\sim [\Sigma \mid \boldsymbol{\Theta}, \boldsymbol{\theta}_0, \boldsymbol{\lambda}] \\
[\Omega \mid \boldsymbol{\Theta}, \boldsymbol{\theta}_0, \mathbf{Y}, \Sigma, \boldsymbol{\lambda}, \boldsymbol{\xi}, \boldsymbol{\Delta}] &\sim [\Omega \mid \boldsymbol{\Theta}, \boldsymbol{\xi}, \mathbf{Y}, \boldsymbol{\Delta}]\,.
\end{aligned}
$$

In particular, we give the conditional distributions useful for the Gibbs sampler.

The conditional distribution of $\boldsymbol{\theta}_t$ for $1 \leq t \leq n$ given data and all other parameters is data proportional to

$$p\left(\boldsymbol{\theta}_t|\cdot\right) \quad \propto \quad \mathcal{N}_{\boldsymbol{\theta}_t}(\mathbf{g}(\boldsymbol{\theta}_{t-1}), \lambda_t\Sigma) * \mathcal{N}_{\boldsymbol{\theta}_{t+1}}(\mathbf{g}(\boldsymbol{\theta}_t), \lambda_{t+1}\Sigma) * \mathcal{L}(\mathbf{y}_t, \mathbf{D}_t, \Omega_t, \boldsymbol{\delta}_t, \xi_t)$$

where $\mathcal{N}_{\mathbf{x}}(\boldsymbol{\mu}, \Sigma)$ is the density function for the multivariate normal with argument $\mathbf{x}$, mean $\boldsymbol{\mu}$ and variance matrix Σ, $\mathcal{L}_t$ is defined in (4) and $*$ denotes convolution. Note that $\mathcal{L}_t$ is similar to a frequentist likelihood for the observations at time t and that some simplification can be achieved when $\mathbf{g}_{t+1}$ is linear as the first two normal distributions can be combined. For $t = 0$;

$$p\left(\boldsymbol{\theta}_0|\cdot\right) \quad \propto \quad \mathcal{N}_{\boldsymbol{\theta}_0}(\boldsymbol{\mu}, \mathcal{S}) * \mathcal{N}_{\boldsymbol{\theta}_1}(\mathbf{g}(\boldsymbol{\theta}_0), \lambda_t\Sigma)$$

For the observational variance components, with priors $\pi(\omega_k)$

$$p\left(\omega_k|\cdot\right) \;\propto\; \pi(\omega_k) \prod_{t=1}^{n} (\xi_t\omega_k)^{(1-\delta_{tk})/2} \exp\left(-\frac{(y_{tk} - \mathbf{h}_k(\boldsymbol{\theta}_t))^2}{2\omega_k\xi_t}\right)^{(1-\delta_{tk})} \Phi\left(-\frac{(y_{tk} - \mathbf{h}_k(\boldsymbol{\theta}_t))}{\sqrt{\omega_k\xi_t}}\right)^{\delta_{tk}}$$

When these are modeled with errors that depend on time via ξ_t, the conditional distributions for ξ_t given all other parameters is proportional to

$$p(\xi_t \mid \cdot) \propto \pi(\xi_t) \prod_{k=1}^{d} (\xi_t\omega_k)^{(1-\delta_{tk})/2} \exp\left(-\frac{(y_{tk} - \mathbf{h}_k(\boldsymbol{\theta}_t))^2}{2\omega_k\xi_t}\right)^{(1-\delta_{tk})} \Phi\left(-\frac{(y_{tk} - \mathbf{h}_k(\boldsymbol{\theta}_t))}{\sqrt{\omega_k\xi_t}}\right)^{\delta_{tk}}$$

The conditional distribution of Σ given all the other parameters and the data is proportional to the product of the prior distribution and an inverse-Wishart distribution with degrees of freedom n and centering matrix $\sum_{t=1}^{n}(\boldsymbol{\theta}_t - \mathbf{g}_t(\boldsymbol{\theta}_{t-1}))(\boldsymbol{\theta}_t - \mathbf{g}_t(\boldsymbol{\theta}_{t-1}))^T/\lambda_t$. If the conjugate prior, $IW(S, \nu)$ is used, then the conditional distribution is an inverse-Wishart with centering matrix $\sum_{t=1}^{n}(\boldsymbol{\theta}_t - \mathbf{g}_t(\boldsymbol{\theta}_{t-1}))(\boldsymbol{\theta}_t - \mathbf{g}_t(\boldsymbol{\theta}_{t-1}))^T/\lambda_t + S$ and degrees of freedom $\nu + n$.

Finally, the conditional distribution of λ_t given all other parameters is the product of the prior with an inverse-gamma distribution with shape parameter $m/2$, where m is the dimension of $\boldsymbol{\theta}_t$ and scale parameter $(\boldsymbol{\theta}_t - \mathbf{g}_t(\boldsymbol{\theta}_{t-1}))^T\Sigma^{-1}(\boldsymbol{\theta}_t - \mathbf{g}_t(\boldsymbol{\theta}_{t-1}))/2$. So, for λ_t and the priors specified in Section 3, the conditional distribution is an inverse-gamma distribution with parameters $18 + 4/2$ and $17 + (\boldsymbol{\theta}_t - \mathbf{g}_t(\boldsymbol{\theta}_{t-1}))^T\Sigma^{-1}(\boldsymbol{\theta}_t - \mathbf{g}_t(\boldsymbol{\theta}_{t-1}))/2$

References

Aitkin, M. (1981), "A Note on the Regression Analysis of Censored Data," *Technometrics*, 23, 161-163.

Barndorff-Neilsen O, Halgreen C. 1977. Infinite Divisibility of the Hyperbolic and Generalized Inverse Gaussian Distributions. In *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **38:** 309–311.

Brockwell PJ and Davis RA. 1991. *Time Series: Theory and Methods*, Springer: New York, NY.

Carlin B, Polson N, Stoffer D. 1992. A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling, *Journal of the American Statistical Association.* **418:** 494–500.

Cowherd, C., Jr. (1993), In: *Aerosol Measurement: Principles, Techniques, and Applications*; Willeke, K., Baron, P. A., (eds.); Van Nostrand Reinhold; New York, NY; 640-658.

Gelman A, Carlin JB, Stern HS, Rubin DB. 1995. *Bayesian Data Analysis.* Chapman & Hall, New York, NY.

Gelman A. 1996. Inference and Monitoring Convergence. In *M*arkov Chain Monte Carlo in Practice, Gilks WR, Richardson S, Spiegelhalter DJ, (eds.); Chapman and Hall/CRC; Boca Raton, FL.

Gleit A. 1985. Estimation for Small Normal Data Sets with Detection Limits. *Environmental Science and Technology* **19:** 1206-1213.

Hamill TM, Snyder C. 2002. Using Improved Background-Error Covariances from an Ensemble Kalman Filter for Adaptive Observations. *Monthly Weather Review* **130:** 1552-1572.

Harvey AC. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*; Cambridge, Cambridge, UK.

Holmén BA, Eichinger W, Flocchini RG. 1998. Application of Elastic Lidar to PM_{10} Emissions from Agricultural Nonpoint Sources. *Environmental Science & Technology* **32:** 3068-76.

Johns CJ, Holmén BA, Niemeier DA, and Shumway RH. 2001. Nonlinear Regression for Modeling Censored One-Dimensional Concentrations Profiles of Fugitive Dust Plumes; *Journal of Agricultural, Biological and Environmental Statistics*, **6:** 99-117.

Jones RH. 1984. Fitting multivariate models to unequally spaced data. In *Time Series Analysis of Irregularly Observed Data* Parzen E, (ed.), Lectures Notes in Statistics, 25, Springer-Verlag, New York; 158-188.

Kalman RE. 1953. A New Approach to Linear Filtering and Prediction Problems. *Transactions of ASME, Journal of Basic Engineering*, **83:** 35–45.

Kitagawa G. 1987. Non-Gaussian State-Space Modeling of Nonstationary Time Series, with discussion, *Journal of the American Statistical Association* **400:** 1032–1044.

Measures, RM. 1984. *Laser Remote Sensing*; Wiley Interscience; New York, NY.

Meinhold R, Singpurwalla N. 1989. Robustification of Kalman Filter Models, *Journal of the American Statistical Society*, **84:** 479–486.

Raftery AE, Lewis SM. 1996. Implementing MCMC. In *M*arkov Chain Monte Carlo in Practice, Gilks WR, Richardson S, Spiegelhalter DJ, (eds.); Chapman and Hall/CRC; Boca Raton, FL.

Seinfeld JH, Pandis SN. 1998. *Atmospheric Chemistry and Physics.* John Wiley & Sons; New York, NY.

Schmee J, Hahn GJ. 1979. A Simple Method for Regression Analysis with Censored Data. *Technometrics*, **21:** 417-432.

Shumway RH, Azari AS, Johnson P. 1989. "Estimating Mean Concentrations Under Transformation for Environmental Data With Detection Limits," *Technometrics*, **31:** 347-356.

Shumway RH, Stoffer, DS. 1982. An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis,* **3:** 253-264.

Tanner MA. 1996. *Tools for Statistical Inference: Methods for Exploration of Posterior Distributions and Likelihood Functions.* Springer-Verlag; New York, NY.

U.S. E. P. A. (1997) *National Ambient Air Quality Standards for Particulate Matter; Final Rule.* Federal Register, Friday July 18, 1997 40 CFR Part 50, 1-102.