Sensors **2007**, *7*, 905-920

sensors
ISSN 1424-8220
© 2007 by MDPI
http://www.mdpi.org/sensors

Nonlinear Bayesian Algorithms for Gas Plume Detection and Estimation from Hyper-spectral Thermal Image Data

Patrick Heasler *, Christian Posse, Jeff Hylden and Kevin Anderson

PO Box 999, Pacific Northwest National Laboratory, Richland, WA 99352
E-mail: patrick.heasler@pnl.gov, christian.posse@pnl.gov,
E-mail: jeff.hylden@pnl.gov, kevin.anderson@pnl.gov
* Author to whom correspondence should be addressed.

Received: 12 April 2007/ Accepted: 6 June 2007/ Published: 7 June 2007

Abstract: This paper presents a nonlinear Bayesian regression algorithm for detecting and estimating gas plume content from hyper-spectral data. Remote sensing data, by its very nature, is collected under less controlled conditions than laboratory data. As a result, the physics-based model that is used to describe the relationship between the observed remote-sensing spectra, and the terrestrial (or atmospheric) parameters that are estimated is typically littered with many unknown ”nuisance” parameters. Bayesian methods are well-suited for this context as they automatically incorporate the uncertainties associated with all nuisance parameters into the error estimates of the parameters of interest. The nonlinear Bayesian regression methodology is illustrated on simulated data from a three-layer model for longwave infrared (LWIR) measurements from a passive instrument. The generated LWIR scenes contain plumes of varying intensities, and this allows estimation uncertainty and probability of detection to be quantified. The results show that this approach should permit more accurate estimation as well as a more reasonable description of estimate uncertainty. Specifically, the methodology produces a standard error that is more realistic than that produced by matched filter estimation.

Key words: plumes, bayesian, regression, MCMC, hyperspectral, LWIR, uncertainty

1. Introduction

Estimating the constituent concentrations of industrial gas plumes from hyper-spectral data has received much attention in the recent years. See, e.g., [1–6] and their detailed overview. A typical issue

when dealing with hyper-spectral data is that confounding factors such as Earth's surface emission, atmospheric absorbance, temperatures, and sensor noise must be accounted for. This results in physics-based models that are over-parameterized, with most of the parameters being nuisance parameters.

Often the nuisance parameters are not completely unknown. Some information exists, usually from previous measurements or from a generic mathematical model. Many remote sensing estimation techniques leverage this external information by (1) estimating the nuisance parameters, (2) then plugging the estimates into the physics-based formulas and (3) mathematically solving for the desired parameters. An estimation technique for the nuisance parameters that has become popular consists in first identifying a "best" scene from large look-up tables that may contain thousands or more real scenes, and then deriving from this scene the necessary parameters [7–9]. Though promising, the plug-in approach does not account for the uncertainties in the estimates of the nuisance parameters, and therefore has two deficiencies; (1) the uncertainties calculated for the parameters of interest are too optimistic, and (2) the estimator is not optimal [10].

Bayesian methods, on the other hand, are better suited for dealing with over-parameterized models. They explicitly make use of the nuisance parameter uncertainties when constructing the estimates of the parameters of interest. Moreover, under general conditions, Bayesian methods still provide consistent estimates of parameters of interest even when the number of nuisance parameters grows with the sample size [11], as is the case with the temperature-emissivity separation problem [12]. This advantage has been recently recognized in hyper-spectral data problems [13, 14].

The typical radiance models used in hyper-spectral data analyses are nonlinear in nature. State-of-the-art methods linearize these models using physics-based considerations such as focusing on optically thin plumes, assuming a known ground emissivity, ignoring the down-welling sky radiance, ignoring the nonlinear variation of ground emissivity, or/and linearizing the plume/background temperature difference.

This paper introduces efficient estimation methods that do not require such simplifications. First, a computationally fast algorithm (comparable in speed to matched filter estimators) for point estimates is described. This iterative procedure, called the Nonlinear Maximum Posterior Density (NLMPD) estimator produces the traditional Bayesian point estimate, which is the maximum posterior density value. This estimator, however, does not provide the full posterior, and, consequently, cannot always accurately quantify uncertainty. In order to produce these uncertainties, Markov Chain Monte Carlo (MCMC) algorithms are introduced [15, 16]. In order to handle (1) the typical high correlations observed between parameters of the physics-based model and (2) the fact that solutions are often at or close to the boundary of the parameter space, the standard MCMC methodology should be improved. A common solution to the first problem is to thin the output of the MCMC simulations. However, the amount of thinning required to sufficiently de-correlate the MCMC output is very significant. This paper investigates an alternative solution which uses Hamiltonian paths in the MCMC algorithms to directly produce uncorrelated outputs [17, 18]. To resolve the boundary problem, legitimate physics-based bouncing algorithms are added to the MCMC sequence. By "legitimate" we mean that the bounces conserve the equilibrium property which is central to the MCMC methodology.

The rest of this paper is organized as follows. Section 2 describes the radiance model used to illustrate

the Bayesian methodology presented in Section 3. Section 4 addresses the parameter estimation of the nonlinear Bayesian regression model. Applications on realistic simulated data are shown in Section 5.

2. Radiance Model

A complete derivation of the radiative transfer model in the thermal hyper-spectral regime is complicated, e.g., [19]. A model based on the simplified three-layer transmission model [20, 21] is sufficient to illustrate the significant aspects of our nonlinear Bayesian regression methodology. The three-layer transmission model provides a very good approximation when the temperature variation within the plume is small with respect to the temperature difference between the plume and the background. The three layers referred to are the ground, the plume and the atmosphere. The plume is the layer of atmosphere closest to the ground that may contain the chemical effluents of interest. In the model, the IR radiation originates from the ground, and while traveling through the other two layers, it is modified according to the formulas presented below. The model we use also incorporates the down-welling radiance factor. This allows one to separately estimate background emissivity from background temperature.

Let ν represent the wavenumber (cm^{-1}) of the electro-magnetic radiation, and let $L_g(\nu)$, $L_p(\nu)$, and $L_a(\nu)$ represent the spectrum leaving the ground, plume, and atmosphere. Then, the following formulas relate the three to each other:

$$L_g(\nu) = \epsilon_g(\nu)\mathcal{B}(\nu, T_g) + (1 - \epsilon_g(\nu))L_d(\nu), \tag{1}$$

$$L_p(\nu) = \tau_p(\nu)L_g(\nu) + (1 - \tau_p(\nu))\,\mathcal{B}(\nu, T_p) \tag{2}$$

and

$$L_a(\nu) = \tau_a(\nu)L_p(\nu) + L_u(\nu) \tag{3}$$

where ϵ_g represents the ground emissivity, T_g the ground temperature, and T_p the plume temperature. The atmosphere is described by the transmissivity, τ_a, the up-welling radiance, L_u, and the down-welling radiance, L_d. Meteorological data can supply values for these atmospheric quantities, therefore, they are assumed known. The expression, $\mathcal{B}(\nu, T)$ represents the Planck spectral radiance of a blackbody at temperature T. Finally, the plume transmissivity, τ_p is related to the chemical effluent's concentrations $\mathcal{C}_j$ (ppm-M) by Beer's law:

$$\tau_p(\nu) = \exp\left(-\sum_{j=1}^{J} \mathcal{A}_j(\nu)\mathcal{C}_j\right) \tag{4}$$

where $\mathcal{A}_j(\nu)$ is the known absorbance spectra for effluent j, $j = 1, \ldots, J$.

Finally instrument error adds noise, so that the observed spectra is given by

$$L_{obs}(\nu) = L_a(\nu) + E(\nu) \tag{5}$$

where $E(\nu)$ is the instrument noise associated with the wavenumber ν. It is assumed that the errors are independent, unbiased (zero mean), and their variance is known.

As mentioned earlier, the three-layer model makes some simplifications:

- *No term for solar radiation*: For this model to be appropriate, one must assume that either (1) the observations are taken at night, or (2) the contribution of solar radiation is insignificant for the IR band being used.

- *The atmospheric terms are known*: The terms τ_a, L_u, and L_d are assumed to be known. This simplification is invoked to allow us to study the dominant source of variability in this problem, which is background clutter (i.e. variability in T_g and ϵ_g). The strategy is to include uncertainty in these atmospheric terms at a later date.

- *There are no correlations or biases in the instrument errors*: A well-calibrated instrument may approximate this assumption. However, periodic instrument calibrations can introduce correlations into these errors.

The Bayesian algorithm introduced in the next section can be extended so that these restrictions can be relaxed.

3. Nonlinear Bayesian Regression Model

3.1. Bayesian Methodology

Given the observed radiation, $L_{obs}(\nu)$, the general Bayesian formulation of the remote sensing problem is the nonlinear regression model

$$L_{obs}(\nu) = F(\nu; \beta, \eta) + E(\nu) \tag{6}$$

where $F(\nu; \beta, \eta)$ describes a physics-based relationship between the observed radiation and the state parameters that describe the observed system, and the noise, or error, $E(\nu)$ is assumed to come from a Gaussian distribution with known variance $v(\nu)$. The errors are assumed independent, therefore, their covariance matrix Σ_E is diagonal, i.e., $\Sigma_E = Diag(v(\nu_1), \ldots, v(\nu_n))$, where n denotes the number of discrete wavenumbers that the instrument records. The Gaussian assumption is fairly reasonable; The error in most IR instruments is dominated by Poisson shot noise, which is well approximated by a normal distribution.

The above model is nonlinear in the sense that F is a nonlinear function of the state parameters. These parameters are divided into two vectors, β and η , to distinguish between those that we want to estimate, β, and the nuisance parameters, η, i.e., parameters of no direct interest, except for the way they complicate the task of estimating the desired parameters.

Which parameters are nuisance parameters or parameters of interest depend on the particular application. For the temperature/emissivity separation problem, β represents the background temperature and emissivity. In the plume gas concentration estimation problem β denotes the chemical effluent burdens.

No matter what the application is, a typical problem of physics-based models for hyper-spectral data is that the unknown parameters outnumber the spectral channels recorded, so if the model is used to formulate a classic (non-Bayesian) regression model, it will represent an under-determined set of equations. One could say that the spectrum is "under-sampled," but it is important to note that the problem cannot

be solved by increasing the sampling. The number of nuisance parameters increases as the sampling rate increases [12].

"Under sampling" is usually solved by including additional information about the nuisance parameters into the problem formulation. The non-Bayesian way to do this is to plug in estimates for the parameters, while the Bayesian formulation considers parameters as random variables and uses a prior distribution to supply this information. The chief advantage a prior distribution has over the plug-in approach is that estimate variabilities or uncertainties can be automatically included.

A secondary advantage is that a prior distribution provides a richer framework for describing what is known about the parameters. In the plume gas concentration estimation problem, the component of the prior that describes the temperatures and the ground emissivity, (representing the nuisance parameters) may include all information known about these parameters; For example, background emissivity must be constrained to the interval [0,1] and also be smooth across energy channels; This information can easily be built into the model prior. Also, the background typically has a strong spatial structure, and it might be very useful to include this information into the prior.

The parameters of interest must also be given a prior. If nothing is known about them, one would apply a non-informative or weakly informative prior to these parameters. However, if some information concerning them is available, their priors can be modified to include such information. For example, it is obvious that the effluent burdens reside within an interval with lower bound 0 and some maximum plausible burden.

3.2. Derivation of the Prior Information

Let $p(\beta, \eta)$ denote the prior distribution of the model parameters. Without loss of generality, the nuisance parameters can be assumed to be independent from the parameters of interest:

$$p(\beta, \eta) = p_\beta(\beta) p_\eta(\eta) \tag{7}$$

where $\eta = (T_g, T_p, \epsilon_g)$ and $\beta = \mathcal{C} = (\mathcal{C}_1, \ldots, \mathcal{C}_J)$.

The generalized Beta distribution and the truncated Gaussian distribution with large variance are commonly used to defined weakly informative priors on large intervals [15, 16]. Using the second approach, a weak or diffuse prior for the concentrations can be formulated as follows:

$$p_\beta(\beta) = C_0 \left[\prod_{j=1}^{J} I\left(\mathcal{C}_j \in [0, M_j]\right) \right] \phi_J\left(\mathcal{C}, \mu_\mathcal{C}, \Sigma_\mathcal{C}\right) \tag{8}$$

where C_0 is a normalization constant so that p_β is a true distribution, $I\left(\mathcal{C}_j \in [0, M_j]\right)$ is 1 if $0 \leq \mathcal{C}_| \leq M_j$, and 0 otherwise, and $\phi_J\left(\cdot, \mu, \Sigma\right)$ denotes the J-dimensional Gaussian distribution with mean μ and covariance Σ. Note that the normalization constant C_0 does not need to be known for the Bayesian estimation of the state parameters. By choosing M_j and $\Sigma_\mathcal{C}$ large, and $\mu_\mathcal{C} = 0$, p_β slightly favors small concentration values, because most gases are likely to have zero or near zero concentration.

External information regarding the nuisance parameters is often available. For example, T_p is typically within 2 to 3K degrees of the the atmosphere, which is known from meteorological data. A prior

for the background temperature T_g can be developed from scene brightness temperature, i.e., the conversion to brightness temperature from radiance using the inverse of Planck's function, and with use of the Nonconventional Exploitation Factors Data System (NEFDS), a government database of surface reflection parameters, e.g., [22] and the National Geospatial-Intelligence Agency (NGA).

Defining priors for the background emissivity is more challenging. The ground emissivity has two important properties that need to be considered in the model and the prior: smoothness and constraint to be in the interval $[0, 1]$. On the other hand, there is a lot of available information on ground emissivity. For example, the NEFDS database contains pre-measured surface reflection parameters for over 400 materials corresponding to a wide variety of objects ranging from camouflage to white paint. This context motivates the following approach for building a prior distribution for the ground emissivity: (1) Model the emissivity as a smooth function of the wavenumber using for example a spline function [23], (2) choose a Gaussian prior for the spline coefficients, with mean and covariance determined from the mean and covariance from the NEFDS library, and (3) map the spline function to the $[0, 1]$ interval via a nonlinear transformation such as the logistic function: $logistic(x) = (1 + \exp(-x))^{-1}$.

A simpler use of the NEFDS library for defining a prior on ϵ_g is as follows. First, consider a sample of m emissivities corresponding to a subset of m materials in the NEFDS library. Compute the mean $\bar{E}$ of this sample. Compute the singular value decomposition of the sample of centered emissivities. Let dE_k denote the k-th eigenvector. Then, one can model ϵ_g as follows:

$$\epsilon_g = \bar{E} + \lambda \sum_{k=1}^{m-1} dE_k \alpha_k \tag{9}$$

where $\alpha = (\alpha_1, \ldots, \alpha_{m-1})$ has a Gaussian distribution with mean 0 and covariance equal to the square of the diagonal matrix of eigenvalues resulting from the singular value decomposition, and λ is a constant we set to 4 in order to generate a diffuse prior. For our experiments we used $m = 6$ emissivities. These emissivities were chosen to be extreme representatives of the NEFDS library, and are displayed in Figure 1.

4. Parameter Inference

Bayesian inference about the model parameters relies on their posterior distribution. Given the data model (6) and the parameter prior distribution (7), the posterior distribution $\pi(\beta, \eta | L_{obs})$ of these parameters is given by

$$\pi(\beta, \eta | L_{obs}) = C_1 \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \frac{\left(L_{obs}(\nu_i) - F(\nu_i; \beta, \eta)\right)^2}{v(\nu_i)} \right) p(\beta, \eta) \tag{10}$$

where C_1 is a normalization constant. Two approaches are commonly used to construct parameter estimates and their associated uncertainties. One can either use the posterior mean and posterior covariance, or the posterior mode and highest posterior density interval.

4.1. NLMPD Algorithm

In order to produce the mode of the posterior (10), we chose a constrained version of a Levenburg-Marquardt iterative algorithm [24] we coined NLMPD for Nonlinear Maximum Posterior Density. Since

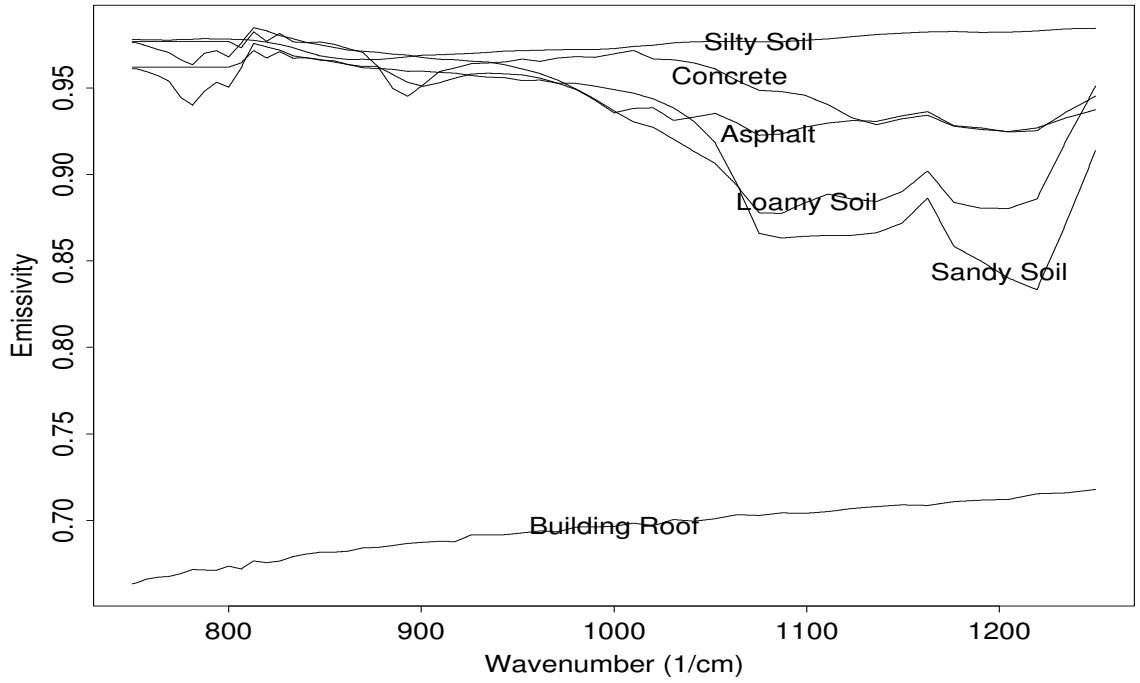

Figure 1. Emissivities of materials used in the simulations.

the priors we used are bounded, the maximization algorithm must deal with these constraints. NLMPD is very efficient for nonlinear least-squares problems and usually converged in a few steps (3 to 10) for our problems. Thus this algorithm requires roughly 3 to 10 times the computation that a matched filter estimator would, and we can consider that they are comparable in speed. NLMPD also requires the first derivatives of the minimizing function to be known. Consequently, one can approximate, at no cost, the uncertainties associated with the mode estimate with an approximation of the posterior covariance matrix given by

$$\Sigma_{\text{post}} \approx \left(\Sigma_{\text{prior}}^{-1} + dF^T \Sigma_E^{-1} dF\right)^{-1} \tag{11}$$

where Σ_{prior} is the covariance matrix of the prior distribution, and dF is the multivariate derivative of $F(\nu; \beta, \eta)$ with respect to the parameters (β, η). This approximation will produce good results when the posterior distribution is close to normal and the gas burden estimates are not highly correlated. However, when the gas burdens are highly correlated (in the posterior, not prior), the covariance approximation can be too large.

4.2. Markov Chain Monte Carlo Algorithm

Let X denote the model parameters. MCMC generates sequences of random variables $X_0, X_1, X_2, \ldots$ that form, after a sufficiently long burn-in of say T iterations, dependent random samples of the posterior distribution $\pi(X)$. Thus, any feature of the posterior distribution, including the distribution itself, can be

approximated from the dependent sample [16].

The simplest MCMC is the Metropolis-Hastings algorithm [25], which, at each time t, chose the next state X_{t+1} by first sampling a candidate point Y from a proposal distribution $q(\cdot|X_t)$. The candidate point Y is then accepted with probability

$$R = \min\left(1, \frac{\pi(Y)q(X_t|Y)}{\pi(X_t)q(Y|X_t)}\right). \tag{12}$$

If the candidate point is accepted, the next state become $X_{t+1} = Y$, otherwise the chain does not move, i.e., $X_{t+1} = X_t$. In particular, note that this algorithm does not require computation of the normalization constant C_1 in (10). In most applications, C_1 cannot be reliably computed as it requires a high-dimensional integration.

Remarkably, the proposal distribution q can have any form. It may also depend on the current point X_t. However, the rate of convergence of the chain to the posterior distribution, as quantified by the *effective sample size* of the generated sequence and its *mixing* will depend on how close q is to π. The effective sample size is given by the length of the sequence minus the number of rejections in (12) while the quality of the mixing is measured by the auto-correlations found in the sequence. Low auto-correlations (i.e., large mixing) produce faster convergence for the posterior estimates. We have found that an efficient proposal q for our problem is a multivariate Gaussian, centered at the current point in the sequence and with covariance equal to the covariance estimate provided by NLMPD divided by 4.

Due to the typical high-correlations observed between the effluent concentrations within the plume, the mixing remains in many cases very low. This forces us to generate very long MCMC sequences and thin them. Thinning consists of subsampling from the sequences by selecting every N-th observation. In other words, in order to produce a reliable sample of size say 5000, one need to generate first an MCMC sequence of length $5000 \times N$ and then extract every N-th observation. As typical values of N in our context is in the order of 100, generating such MCMC becomes costly.

Recently, Hamiltonian paths have been proposed to both increase the mixing in the MCMC sequence and reduce the number of rejections [17]. The Hamiltonian MCMC is built upon the basic principle of Hamiltonian mechanics. Let the function U be such that $\pi(X) \propto \exp(-U(X))$. The model parameters X can be regarded as a position vector and $U(X)$ the potential energy function. Introducing a fictitious momentum vector $P = (p_1, \ldots, p_d)$ and mass vector $M = (m_1, \ldots, m_d)$, where d denotes the number of model parameters, one can define the kinetic energy function $K(P) = \frac{1}{2}\sum_{i=1}^{d} p_i^2/m_i$. The total energy is then

$$H(X, P) = U(X) + K(P). \tag{13}$$

Consequently, if one can sample (X, P) from the distribution $\pi(X, P) \propto \exp(-H(X, P))$, then the marginal distribution of X is exactly the target distribution $\pi(X)$.

Hamiltonian dynamics allows one to move along trajectories of constant H, taking large jumps in the parameter space. The Hamiltonian algorithm alternates between picking a new momentum vector and following such trajectories. Each iteration starts with a generation of a new momentum according to a multivariate uncorrelated Gaussian distribution with (diagonal) covariance D. Then a trajectory that maintains a constant H is approximated by a discretized version, called leapfrog technique, which

consists of iterating the following steps h times:

$$P(t+\delta/2) = P(t) - \frac{\delta}{2}\left.\frac{\partial U}{\partial X}\right|_{X(t)} \quad (14)$$

$$X(t+\delta) = X(t) + \frac{\delta}{M}P(t+\delta/2) \quad (15)$$

$$P(t+\delta) = P(t+\delta/2) - \frac{\delta}{2}\left.\frac{\partial U}{\partial X}\right|_{X(t+\delta)} \quad (16)$$

where t denotes the t-th iteration of Hamiltonian MCMC, $X(t) = X_t$, $P(t)$ denotes the momentum at the beginning of the leapfrog step, and δ represents a small increment.

Because the leapfrog technique is only an approximation, H may vary along these paths. A Metropolis step is then use to re-establish the equilibrium. The state at the end of the Hamiltonian path $X_{t+1} = X(t+h\delta)$ is accepted with probability $R = \min(1, \exp(-H(X_{t+1}, P_{t+1}) + H(X_t, P_t))$ where $P_t = P(t)$ and $P_{t+1} = P(t+h\delta)$.

The efficiency of the Hamiltonian algorithm depends on the choices for the δ, h, M, and D. δ must be small enough to ensure that H does not change much on the approximated paths (to minimize the rejections in the Metropolis step), but it must be large enough to not slow down too much the computational time. $\Delta = h\delta$ must be large enough to ensure good mixing. M and D must be chosen to ensure that the Hamiltonian MCMC samples adequately cover the target distribution π. For our problems, we found that $\Delta = 0.05$, $h = 50$, M equal to the inverse of the diagonal elements of the covariance matrix estimate given by NLMPD, and $D = Diag(M^{0.5})$ produce the desired sequences.

Finally, in our context one can often expect to find parameter estimates on the distribution boundaries. For example, some gases that are part of the model are not present in the plume or are in very small concentrations. MCMC techniques, Metropolis and Hamiltonian, have difficulties generating legitimate candidates that are within the boundaries of the parameter space, leading to many rejections before a suitable candidate can be proposed. To solve this problem, we present a bouncing algorithm that conserves the equilibrium property central to the MCMC algorithm.

Let V and W denote the current and next states in the MCMC sequence. For the standard MCMC algorithm, $V = X_t$ and $W = Y$, while for the Hamiltonian MCMC, $V = X(t+g\delta)$ and $W = X(t+(g+1)\delta))$ with $g < h$. Note that one can write $W = V + \Delta V$, The idea is that if W lies outside the distribution boundaries, we replace it with a legitimate candidate W^* in the same direction as W, i.e., $W^* = V + \rho\Delta V$. The algorithm mimics the physics of a ball leaving the position given by V in the direction of W, bouncing back on its path whenever it hits the boundary wall and stopping when it has traveled the distance ΔV. Note that given the boundary locations, the ending point may be between the points V and W or somewhere in the opposite direction of W, i.e., ρ may be negative.

The bouncing algorithm determines ρ as follows. Let $nv = |\Delta V|$ be the norm of ΔV, and $A = \Delta V/nv$. Let U and L denote the upper and lower bounds of the parameter space. Compute the following quantities:

1. $a_u \leftarrow (U - V)/A$ and $a_l \leftarrow (L - V)/A$.

2. $b_1 \leftarrow \max\{p \min\{a_l, a_u\}\}$

3. $b_2 \leftarrow \min\{p\max\{a_l, a_u\}\}$

4. $b \leftarrow b_2 - b_1$

5. $\rho \leftarrow (nv - b_1)$ modulo $2b$

6. if $\rho > b$ then $\rho \leftarrow 2b - \rho$

7. $\rho \leftarrow (\rho + b_1)/nv$

where $p\max\{a, b\}$ denotes the maximum taken on the vector components ($p\max\{a, b\}_i = \max\{a_i, b_i\}$), while $p\min\{a, b\}$ denotes the minimum.

5. Applications

5.1. NLBR vs. Matched Filter

To test the methodology, we simulated spectra with known gas burdens and compared the NLBR approach with a matched filter, perhaps the most popular algorithm currently in use for gas plume detection. We considered the case of a plume with 3 non-correlated gases — NH3, Ethylene and Butyl Acetate — at various concentration levels: 0, 10, 20, 30, 50, 70, 90, and 110 ppm-M. For each level of concentration, 100 spectra were generated. For each simulation, the background emissivity was randomly selected from the 6 emissivities shown in Figure 1.

Figure 2 compares estimates of NH3 obtained with NLMPD and the matched filter. The figure plots the true burden of NH3 (horizontal axis) versus the estimates (vertical axis). Consequently, a plot for a perfect algorithm would have all the estimates on the identity line. One can see that the matched filter estimator displays much more variability than the NLBR estimator (except when the gas burden is 0). In fact, the matched-filter plot contains points that look like outliers. These "outliers" are associated with the simulations that picked an atypical background emissivity (the bottom one in Figure 1). In the linearization that the matched filter algorithm uses, an average emissivity is plugged in, which results the observed "outliers." In the NLMPD algorithm the variability in emissivity seen in Figure 1 is accounted for in the prior.

Because of this problem, the matched-filter estimates are biased, and the actual error in the estimates (matched-filter Root Mean Squared Error RMSE=36ppm-M) is much greater than the error predicted by the algorithm (matched-filter predicted standard error=1ppm-M). On the other hand, these problems are absent from the NLBR estimator; It has no detectable bias, and the observed and predicted estimation error, given by (11), are the same (RMSE=2ppm-M, average predicted standard error=2ppm-M).

5.2. Gas Detection

The fact that the NLBR algorithm produces realistic standard errors has great benefit when the estimates are used for detection. The usual statistic to test the presence or absence (0 ppm-M) of a gas is the t-statistic, i.e., the algorithm estimate divided by the algorithm predicted standard error given by (11).

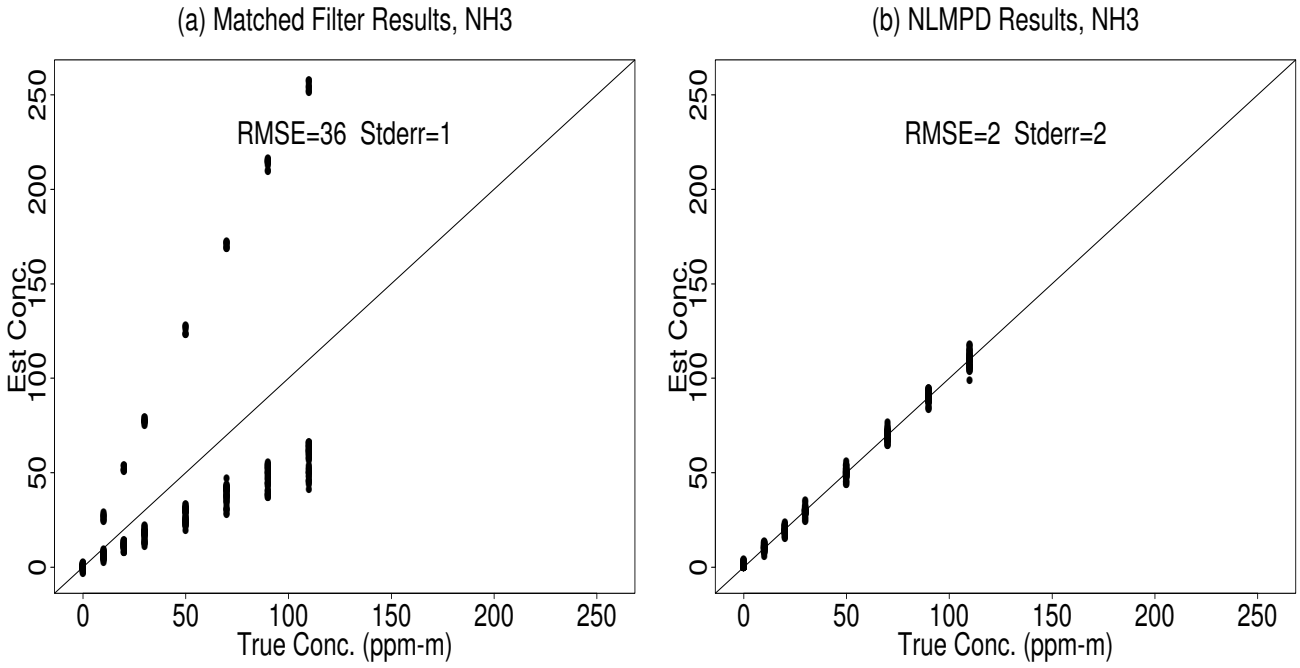

Figure 2. Matched-Filter Estimator Compared to NLBR Estimator.

Under the "null" hypothesis of no gas burden, the t-statistic should approximately follow a truncated Gaussian distribution (truncated because burdens are constrained to be non-negative).

Figure 3 presents the distributions of the t-statistic for NH3 and Ethylene, calculated from 1000 spectra simulated as in the previous example at concentration level 0 ppm-M. The distributions are presented in the form of Q-normal plots, and therefore, the t-statistic values should ideally line up on a straight line except for the truncated values. As one can see from the plots, the null distributions conform closely to the theoretical ideal.

This means that detection thresholds for NLBR estimates and the associated false call rate can be theoretically calculated and results will correspond to theory. Figure 4 presents probability of detection (POD) curves calculated from 2000 spectra, the above 1000 spectra and 200 spectra for each of the concentration levels 50, 100, 150, 200 and 250 ppm-M. Detection occurs when the t-statistic is larger than 3, a threshold that should give about a one in a thousand chance of a false call. In the figure, POD is plotted as a function of true gas concentration, so POD(0) is the false call rate. For NH3, the observed false call rate is in fact 1/1000, for Ethylene the false call rate is 3/1000, very good agreement with theory.

Fits of the NLBR model to actual hyperspectral cubes have shown that the performance shown in Figure 4 is more optomistic than that experienced on real data. For example, the residuals from real data fits are typically from 2 to 5 times larger than theory would predict, indicating that un-modeled sources of variability exist in the data. One obvious source is the atmospheric terms in the model, which are currently assumed to be perfectly known. The next step in our modeling strategy will be to construct a prior that adequately reflects variability in these terms.

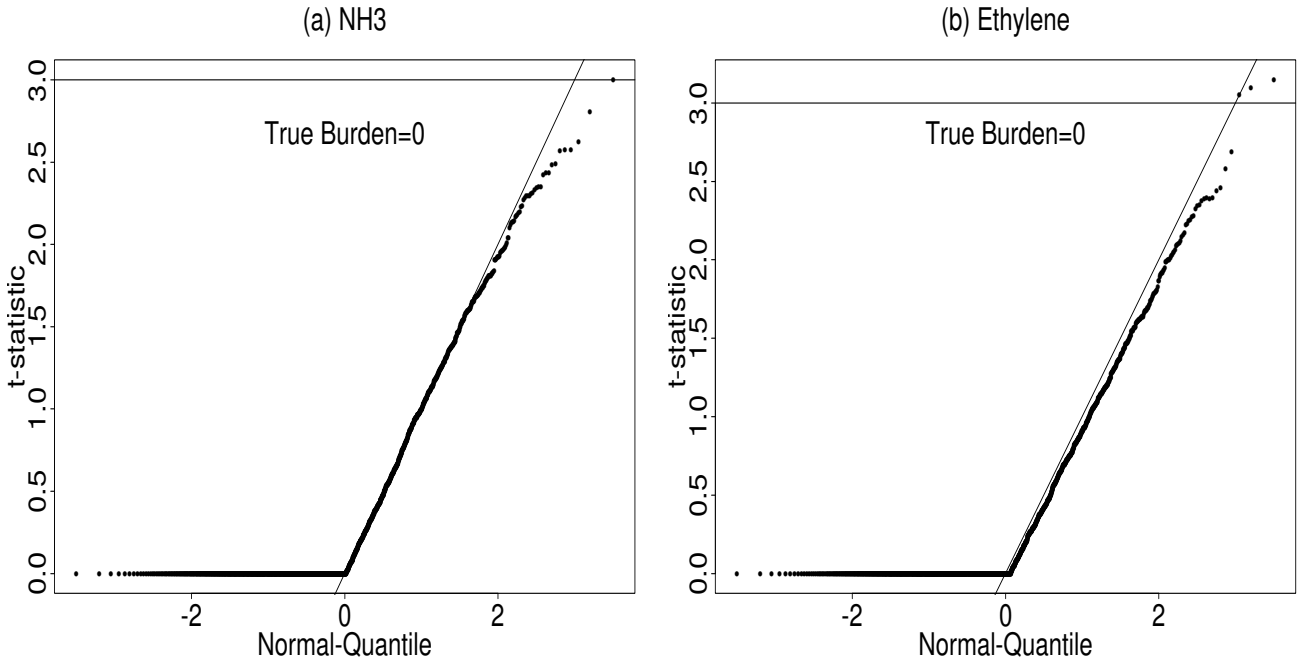

Figure 3. Q-normal Plots of the T-statistic for NH3 and Ethylene, When the True Gas Burden is 0.

5.3. Highly Correlated Gases

The results presented up to this point utilize the "quick" Bayesian solution provided by the NLMPD approach. Figure 5 presents a case for which the complete posterior, as provided by MCMC algorithm is very useful. For this example, 11 highly correlated gases were placed in the plume. The gas correlations are so large that gas burden estimation would be an ill-conditioned problem, if an unconstrained linear regression model were employed. (The principal constraint that makes this problem solvable is that all gas burdens must be non-negative). With the aid of this constraint, the Bayesian regression algorithm transforms an ill-conditioned problem to one with a very informative solution. Unfortunately, the posterior distribution for this problem is poorly approximated by a normal, and the NLMPD approach cannot provide reliable estimates. On the other hand, the MCMC algorithm can calculate the full posterior and provide critical insight on what is happening.

The two histograms in Figure 5 represent the posterior distribution for Butyl Acetate, one of the gases in the 11-gas plume. The first posterior distribution represents the estimated burden of Butyl Acetate when the true burden for each of the 11 gases is 0 ppm-M. For this case, the MCMC algorithm estimates the true burden with great accuracy even though large correlations exist between gases. In fact, a 95% highest probability density interval is (0,4.5) ppm-M, quite a small interval. Furthermore, the MCMC output reveals that the posterior for Butyl Acetate has an exponential shape with a modal value of 0, which is far from normal.

The second posterior distribution illustrates the problem that occurs when the plume contains a large amount of the gases (100 ppm-M for each gas). In this case, the shape of the distribution is dramatically different; between 0 and 100 ppm-M it is roughly uniform, and becomes roughly normal above 100 ppm-M (or more appropriately half-normal). The uniform portion below 100 ppm-M is produced by the

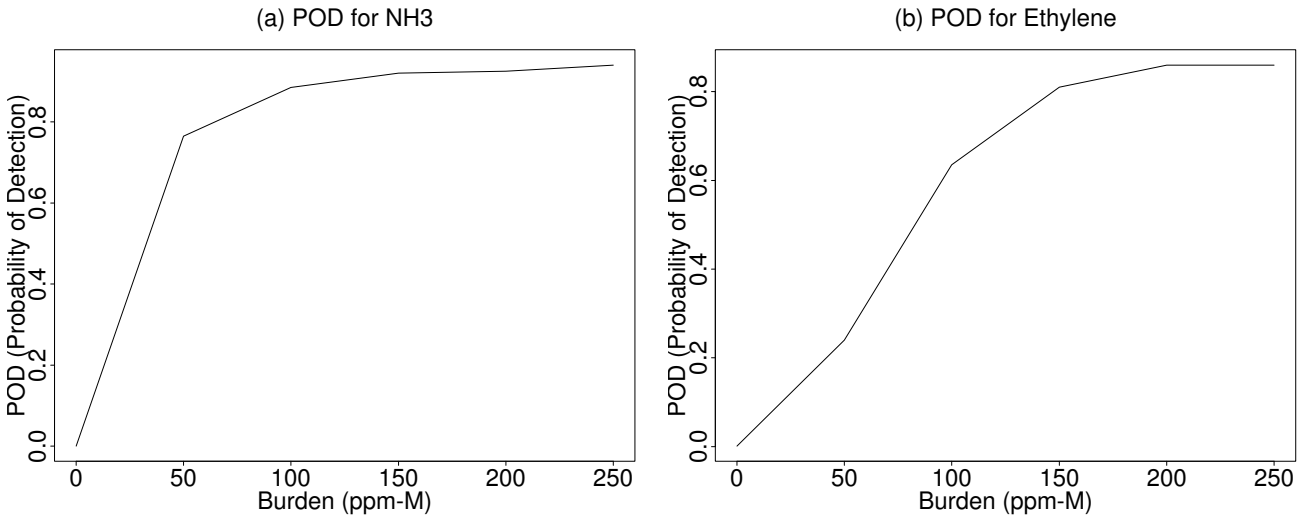

Figure 4. POD Curves for NH3 and Ethylene.

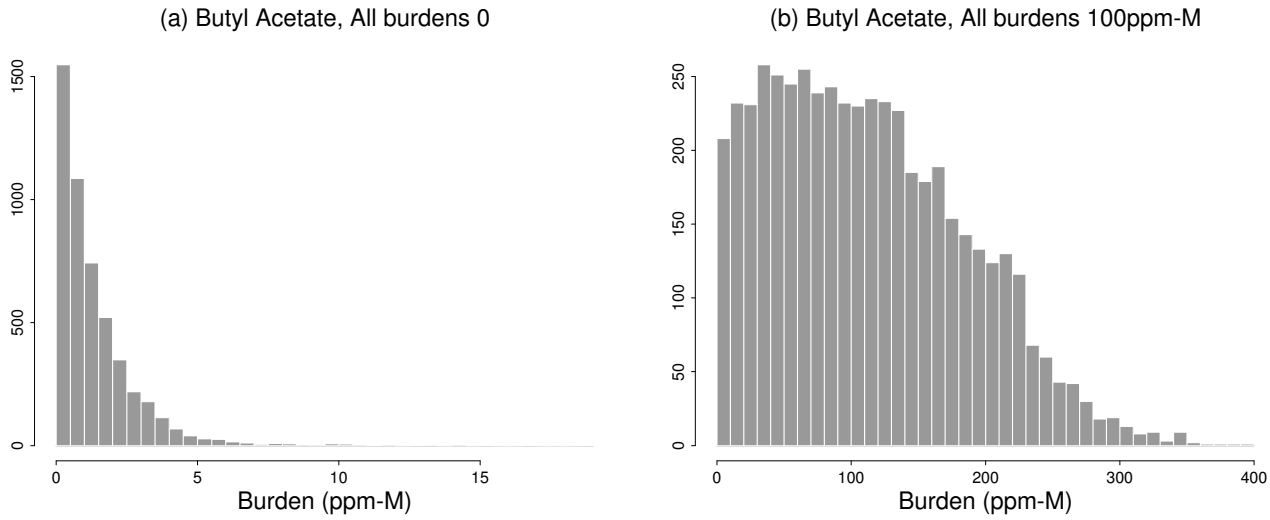

Figure 5. MCMC Derived Posterior for a Plume with 11 Highly Correlated Gases.

fact that the Butyl Acetate signal is equivalent to a linear combination of other gases. The normal half of the distribution is caused by the constraints on the linear combinations (burdens are non-negative), which become active at higher-burdens. Note that the exact nature of the correlations in the plume gases can create posteriors with quite a complex shape; It is possible to have several modes in the posterior. As one can see, the MCMC can be very valuable for evaluating plumes that may contain several gases.

The above histograms were constructed from a sample of 5000 data points generated by the Hamiltonian MCMC, after a burn-in phase of size 1000. Similar outputs could be obtained with the "traditional" MCMC algorithm described in Section 4.2, using a thinning of size $N = 100$. In other words, the second MCMC algorithm had to generate first a sequence of length 600,000, the first 100,000 being used for the burn-in phase, and then extract each $100'th$ observation to build the final sample. Because of this very high thinning, the "slow" Hamiltonian MCMC turned out to be faster by a factor of 2 than the traditional MCMC algorithm. Note that in less correlated experiments as the examples discussed earlier, the two

algorithms were found to be comparable in speed.

6. Summary and Conclusions

The nonlinear Bayesian regression shows definite promise for producing better estimates from hyper-spectral data. To be most useful, NLBR should not be considered to be a specific algorithm, derived from a specific model, but a general framework for producing estimates that can be tailored to the problem at hand.

The effectiveness of these regression models depends heavily on the prior information supplied to them. The methodology does require that realistic prior distributions be constructed on the nuisance parameters. However, the fact that uncertainties can be incorporated into the prior distributions in a straightforward manner means that the Bayesian methodology incorporates these uncertainties in a fairly automatic way. In fact, one of the strongest motivations for using a Bayesian models may be because its statements of estimate uncertainty are more believable than those from other methodologies.

Our simulations involving NLBR have produced results that correspond with theory. The application of the present NLBR model to real hyper-spectral data have produced less optimal performance; It is obvious that real hyper-spectral data contains more variability than our NLBR currently describes, most prominently, atmospheric variability [26]. However, the flexibility of the NLBR framework should allow us to include these other sources of variability without any dramatic change in the general approach.

Acknowledgments

The authors would like to acknowledge the helpful advice of Herb Fry, Tom Burr, Bernie Foy, and Brian McVey of Los Alamos National Laboratory. This work was supported by the Computational and Information Science Initiative at the US Department of Energy's Pacific Northwest National Laboratory, Richland WA 99352. Pacific Northwest National Laboratory is operated by Battelle Memorial Institute for the US Department of Energy under Contract DE-AC05-76RL01830.

References

1. Funk, C. C.; Theiler, J.; Roberts, D. A.; .; Borel, C. C. Clustering to improve matched filter detection of weak gas plumes in hyperspectral thermal imagery. *IEEE Trans. Geosci. and Remote Sensing* **2001**, *39*, 1410–1420.
2. Young, S. J. Detection and quantification of gases in industrial-stack plumes using thermal-infrared hyperspectral imaging. *Aerospace Report ATR-2002(8407)-1* **2002**.
3. Messinger, D. Gaseous plume detection in hyperspectral images: a comparison of methods. In *Proceedings of SPIE'04, series 5424*, pages 592–603, **2004**.
4. ODonnell, E.; Messinger, D.; Salvaggio, C.; Schott, J. Identification and detection of gaseous effluents from hyperspectral imagery using invariant algorithms. In *Proceedings of SPIE'04, series 5425*, **2004**.
5. Theiler, J.; Foy, B.; Fraser, A. Characterizing non-gaussian clutter and detecting weak gaseous plumes in hyperspectral imagery. In *Proceedings of SPIE'05, series 5806*, pages 182–193, **2005**.

6. Pogorzala, D. *Gas plume species identification in LWIR hyperspectral imagery by regression analyses*. PhD thesis, Rochester Institute of Technology, **2005**.
7. Hernandez-Baquero, E. D.; Schott, J. R. Atmospheric compensation for surface temperature and emissivity separation. *Proceedings of the SPIE'00 Aerosense, Orlando, FL* **2000**.
8. Borel, C. C. Recipes for writing algorithms for atmospheric corrections and temperature/emissivity separations in the thermal regime for a multi-spectral sensor. *Proceedings of the SPIE'01 Aerosense, Orlando, FL* **2001**.
9. Borel, C. C. Artemiss - an algorithm to retrieve temperature and emissivity from hyper-spectral thermal image data. *Proceedings of the 28th Annual GOMATech Conference, Hyperspectral Imaging Session, Tampa, FL* **2003**.
10. Yang, Z. L.; Tse, Y. K.; Bai, Z. D. On the asymptotic effect of substituting estimators for nuisance parameters in inferential statistics. Technical report, Singapore Management University, School of Economics and Social Sciences, **2003**.
11. Lancaster, T. Orthogonal parameters and panel data. *Review of Economic Studies* **2000**, *69*, 647–666.
12. Realmuto, V. J. Separating the effects of temperature and emissivity: Emissivity spectrum normalization. *Proceedings of the 2nd TIMS Workshop, JPL Publications* **1990**, *99-55*, 31–35.
13. Milman, A. S. *Mathematical Principles of Remote Sensing: Making Inferences from Noisy Data*. Taylor & Francis, New York, first edition, **2000**.
14. Burr, T.; McVey, B.; Sander, E. Chemical identification using bayesian model selection. In *Proceedings of 2002 Spring Research Conference on Statistics in Industry and Technology*, **2002**.
15. Gelman, A.; Carlin, J. B.; Stern, H. S.; Rubin, D. B. *Bayesian Data Analysis*. Chapman & Hall, London, **1995**.
16. Gilks, W. R.; Richardson, S.; Spiegelhalter, D. J. *Markov Chain Monte Carlo*. Chapman & Hall, London, **1996**.
17. Hanson, K. M. Markov chain monte carlo posterior sampling with the hamiltonian method. In Sonka, M.; Hanson, K. M., editors, *Proceedings of SPIE Vol. 4322, Medical Imaging: Image Processing*, **2001**.
18. Alder, B.; Wainwright, T. Studies in molecular dynamics i, general method. *Journal of Chemical Physics* **1959**, *31(2)*, 459–466.
19. Berk, A.; Acharya, P. K.; Bernstein, L. S.; Anderson, G. P.; Chetwynd, J. H.; Hoke, M. L. Reformulation of the modtran band model for higher spectral resolution. In Shen, S.; Descour, M., editors, *Proceedings of SPIE Vol. 4049, Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI*, **2000**.
20. Beer, R. *Remote Sensing by Fourier Transform Spectrometry*. Wiley Interscience, New York, **1991**.
21. Flanigan, D. F. Prediction of the limits of detection of hazardous vapors by passive infrared with the use of modtran. *Applied Optics* **1996**, *35*, 6090–6098.
22. Westlund, H. B.; Meyer, G. W. A brdf database employing the beard-maxwell reflection model. In *Proceedings of Graphics Interface 2002*, **2002**.
23. Bartels, R. H.; Barsky, B. A.; Beatty, J. C. *An Introduction to Splines for Use in Computer Graphics*

and Geometric Modelling. Morgan Kaufman, Los Altos, CA, **1987**.

24. Marquardt, D. M. An algorithm for least squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431–441.
25. Hastings, W. K. Monte carlo sampling methods using markov chains and their applications. *Biometrika* **1970**, pages 97–109.
26. Miller, B.; Messinger, D. The effects of atmospheric compensation upon gaseous plume signatures. In *Proceedings of SPIE'05, series 5806*, **2005**.

© 2007 by MDPI (http://www.mdpi.org). Reproduction is permitted for noncommercial purposes.