

Construction of an Improved Bayesian Clutter Suppression Model for Gas Detection

P.G. Heasler
K.K. Anderson
J.L. Hylden

October 2002



Prepared for the U.S. Department of Energy
under Contract DE-AC06-76RL01830

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.** Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC06-76RL01830



This document was printed on recycled paper.

Construction of an Improved Bayesian Clutter Suppression Model for Gas Detection

P.G. Heasler
K.K. Anderson
J.L. Hylden

October 2002

Prepared for
the U.S. Department of Energy
under Contract DE-AC06-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99352

Contents

1	Introduction	1
2	Description of Problem and Relevant Regression Models	1
2.1	Bayesian Regression Model	3
2.1.1	NLMPD Algorithm	5
2.1.2	Form of the Regression Priors	5
2.2	Matched Filter Model	6
3	Methodology for Empirically Producing Clutter Prior Distributions	8
3.1	Simple Normal Model for Clutter	8
3.2	Improved Prior for Clutter	9
4	A Simulated Example	10
5	Comparison of Bayesian Clutter Suppression Model to Matched Filter	13
6	Conclusions	19
7	Glossary of Terms	21
8	References	22

1 Introduction

This technical report describes a nonlinear Bayesian Regression model that can be used to estimate effluent concentrations from IR hyperspectral data. As the title implies, the model is constructed to account for background clutter more effectively than current “matched filter” estimators. Although the main objective is to account for background clutter, which is the dominant source of variability in IR data, the model could easily be extended to allow for uncertainties in the atmosphere. The term, “clutter,” refers to the variations that occur in the image spectra because emissivity and background temperature change from pixel to pixel. The Bayesian regression model utilizes a more complete description of background clutter to obtain better estimates. The description is in terms of a “prior distribution” on background radiance.

The estimates are produced from a nonlinear Bayesian regression model with the use of an iterative technique that produces the *maximum posterior density* (MPD) point estimates, [1], for the desired effluent concentrations. We call it the nonlinear maximum posterior density (NLMPD) estimator. This estimator is computationally fast (comparable in speed to matched filter estimators), but it does not produce the full posterior, as a Markov Chain Monte Carlo estimator would. Consequently, one cannot use the posterior distribution to produce the optimal RMSE estimator (which is the posterior mean), and one cannot use the posterior to quantify uncertainty.

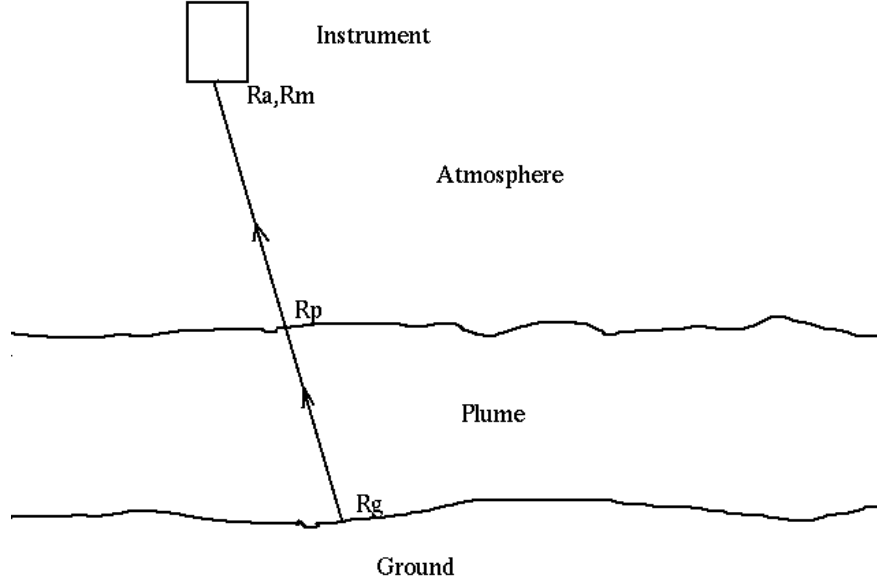
To quantify uncertainty, we have developed an asymptotic approximation to the posterior covariance matrix. As with the maximum likelihood approximation for covariance, the approximation will be “good” when the posterior resembles a normal distribution. As one will see, the asymptotic approximation seems to work acceptably well, for the effluent estimation problem.

2 Description of Problem and Relevant Regression Models

The general problem of interest is to determine the concentration of chemical effluents from an industrial site using passive IR hyper-spectral images. For each chemical effluent of interest, we would like to produce a concentration map; in other words, the estimation algorithm should produce a concentration estimate for each pixel in the image, using the IR spectrum associated with that pixel. The nonlinear Bayesian model that accomplishes this is based on a simplified physics model that describes how the IR spectra originate. The physics model is a simple three-layer transmission model illustrated in Figure 1. The “three layers” referred to are the ground, the plume, and the atmosphere. 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 “plume” is the layer of atmosphere closest to the ground that may contain the chemical effluents of interest.

Let ν represent the wavenumber (cm^{-1}) of the electro-magnetic radiation, and let $R_g(\nu)$, $R_p(\nu)$, and $R_a(\nu)$ represent the spectrum leaving the ground, plume, and atmosphere, re-

Figure 1: Three Layer Physics Model for Passive IR Spectra



spectively. Then the following formulas relate the three to each other;

$$R_p(\nu) = \tau_p(\nu)R_g(\nu) + (1 - \tau_p(\nu))\mathcal{B}(\nu, T_p) \quad (1)$$

and,

$$R_a(\nu) = \tau_a(\nu)R_p(\nu) + (1 - \tau_a(\nu))\mathcal{B}(\nu, T_a) \quad (2)$$

In the above formulas, $\mathcal{B}(\nu, T)$ represents the Planck black-body function, τ_a the atmospheric transmissivity, τ_p the plume transmissivity, T_p is plume temperature, and T_a is atmospheric temperature. The plume transmissivity, τ_p is related to the chemical effluent's concentrations, \mathcal{C}_i (ppmM) by Beer's law;

$$\tau_p(\nu) = \exp\left(-\sum_{i=1}^P A_i(\nu)\mathcal{C}_i\right) \quad (3)$$

where $A_i(\nu)$ is the known absorbance spectra for effluent i .

The resulting spectrum, R_a , represents the IR spectra presented to the instrument, but this is not the spectrum that the instrument observes; The instrument alters the resolution of the spectrum and noise is added. So to get the observed spectra, $R_{obs}(\nu)$, another step must be added to the model;

$$R_{obs}(\nu_j) = \Psi * R_a(\nu_j) + e_j \quad (4)$$

where Ψ represents an appropriate convolution filter, $\nu_j, j = 1, 2, 3 \dots n$ represent the discrete wavenumbers that the instrument records, and e_j is the instrument noise associated with wavenumber j . It is assumed that the errors are independent, unbiased, and $Var(e_j)$ is known.

Even though the physics model presented above can provide a highly accurate description of reality, it is, nevertheless, only an approximation of real IR data. It is therefore important to enumerate the simplifications that the model rests upon;

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 atmosphere can be approximated with one layer: Equation 2 describes a one layer model for the atmosphere. A more exact description of the atmosphere would divide it into several layers. In our one-layer atmosphere, τ_a would represent the average transmissivity, and T_a the average temperature.

The down-welling atmospheric radiance can be ignored: The atmosphere also radiates in a downward direction, and this radiation can be reflected by the ground. This should be a relatively insignificant term.

There are no correlations or biases in the instrument errors: Periodic instrument calibrations can introduce correlations into the errors.

2.1 Bayesian Regression Model

The general form of the regression model is, therefore;

$$R_{obs}(\nu_j) = F(\nu_j; \mathcal{C}, \eta) + e_j \quad (5)$$

where $F(\nu; \mathcal{C}, \eta)$ is a function that represents the result $\Psi * R_a(\nu)$ from the 3-layer model outlined previously. In order to calculate the result $\Psi * R_a(\nu)$, one must supply values for the following parameters; \mathcal{C} a vector of concentrations of the potential plume effluents, and η , a vector of all other parameters in the model. Thus η should determine the temperatures T_p , and T_a , the ground radiance R_g , and the atmospheric transmissivity, τ_a . The notation $F(\nu; \mathcal{C}, \eta)$ is constructed to emphasize the fact that the observed spectrum, R_{obs} , is a function of two different set of parameters; The concentrations \mathcal{C} , that we desire to estimate from the data and know little about, and the *nuisance parameters*, η , for which some auxiliary information exists.

Some information concerning the nuisance parameters is typically present in the observed spectrum, R_{obs} , however the spectrum alone is insufficient to estimate all parameters; To produce a solvable regression model, one needs to add some additional information concerning the nuisance parameters. The typical solution to this problem has been to determine an estimate of the η from auxiliary information and plug the estimate into $F(\nu; \mathcal{C}, \eta)$, so that the solution becomes unique. This solution does not account for the uncertainties in the estimate of η , and therefore has two deficiencies; (1) the uncertainties calculated for the \mathcal{C} are too optimistic, and (2) the estimator is not optimal.

A better solution to the problem is to use the auxiliary information to produce a *prior distribution* for η , because this distribution will correctly incorporate uncertainty for η into

the regression model. Thus one constructs the *prior distribution*, $p(\eta)$, with the mean of this distribution representing the “best estimate” for η and its standard deviations representing the estimate uncertainty. The prior, $p(\eta)$ would typically be considered an “informative” prior; The form of this prior *will*, of course affect the regression solution for the concentration, \mathcal{C} .

The “primary” nuisance parameters in Equations 1 and 2 are the temperatures T_p , T_a and the vectors R_g , and τ_a , so one could use these as the nuisance parameter vector ($\eta = (R_g, \tau_a, T_p, T_a)$). A strategy such as this, greatly inflates the number of unknown parameters that the regression has to deal with, because both R_g and τ_a can be of larger dimension than the data, R_{obs} . A better strategy is to express both R_g and τ_a in terms of “secondary nuisance parameters” that have lower dimension *and* a particularly simple prior distribution. In other words, we identify a transformation that expresses R_g in terms of parameter vector β_g , and expresses τ_a in terms of β_a , so that the nuisance parameter vector employed is;

$$\eta = (\beta_g, \beta_a, T_p, T_a) \quad (6)$$

To produce a regression model that is entirely Bayesian, we must also supply a prior for the concentrations, \mathcal{C} . However, unlike the η there will usually be no prior information (outside of some obvious bounding information) that we will want to incorporate into this prior. It is therefore natural to use a uniform “non-informative” prior on the \mathcal{C} . Therefore, the prior on each concentration, \mathcal{C}_i should be uniform over the interval between zero to the maximum burden for that effluent.

Bayesian regression constructs parameter estimates and associated uncertainties from the posterior distribution, thus for our specific model, one must calculate statistics associated with the posterior $f(\mathcal{C}, \eta | R_{obs})$. The most desirable Bayesian estimates for the unknown parameters would be the mean of this posterior distribution, although the median and mode are also used. (Using minimum RMSE as an optimality criteria, one finds that the mean of the distribution is the best estimate.) The posterior is defined as;

$$f(\mathcal{C}, \eta | R_{obs}) = \frac{f(R_{obs} | \mathcal{C}, \eta) p(\mathcal{C}, \eta)}{\int f(R_{obs} | \mathcal{C}, \eta) p(\mathcal{C}, \eta) d\mathcal{C} d\eta} \quad (7)$$

with the conditional distribution $f(R_{obs} | \mathcal{C}, \eta)$ defined by the regression model so that

$$f(R_{obs} | \mathcal{C}, \eta) = \prod_{i=1}^N \phi(R_{obs}(\nu_i); F(\nu_i; \mathcal{C}, \eta), Var(e_i)) \quad (8)$$

The conditional mean and covariance of the concentrations \mathcal{C} are defined by;

$$\hat{\mathcal{C}} = E(\mathcal{C} | R_{obs}) = \int \mathcal{C} f(\mathcal{C}, \eta | R_{obs}) d\mathcal{C} d\eta \quad (9)$$

and

$$Cov(\mathcal{C} | R_{obs}) = \int (\mathcal{C} - \hat{\mathcal{C}})(\mathcal{C} - \hat{\mathcal{C}})^T f(\mathcal{C}, \eta | R_{obs}) d\mathcal{C} d\eta \quad (10)$$

We desire to use the conditional expectation and covariance as estimates for \mathcal{C} and the associated uncertainty (the conditional covariance is actually the the RMSE for the estimate).

However, there is no quick, closed-form solution for doing this. Monte Carlo methods (i.e. Markov Chain Monte Carlo simulations) do exist to perform the integrations, but they are thousands of times slower than current estimators (matched filters). An objective of this work was to develop an algorithm that quickly produced an approximation to the desired Bayesian solution.

2.1.1 NLMPD Algorithm

Our “quick approximation” to the conditional expectation presented in Equation 9 is to use the “maximum posterior density” estimate in place of the conditional mean. The maximum posterior density is simply the mode of the posterior distribution, and will be a good approximation to the mean when the distribution is symmetric and unimodal. Use of the maximum posterior density is the equivalent to maximum likelihood for a classical non-linear problem, and the optimization algorithms applied to maximum likelihood can be easily adapted to the problem of finding the NPD estimate. We call such an algorithm a *non-linear maximum posterior density* (NLMPD) algorithm. Since the priors we will be applying to the regression problem will be bounded (see the next Section for more information on the prior), the maximization algorithm must deal with constraints. In other words, determination of the MPD estimate is a *non-linear programming problem*, [6] and any algorithm that will solve such a problem could be applied to obtain the estimates.

We have chosen to use a constrained version of a Levenburg-Marquardt algorithm [7] to produce the MPD solution. This is an iterative technique, and experience shows that it converges in a few steps (approximately 3 to 10) for this problem. Thus this algorithm requires roughly 3 to 10 times the computation that a matched filter estimator would, and we can consider this algorithm to be roughly of comparable speed to a matched filter.

The other component of the estimation process, calculation of its uncertainty, (as represented by the conditional covariance presented in Equation 10), is produced by calculating the *asymptotic covariance matrix*, which is the Bayesian analog of the maximum-likelihood covariance. Thus;

$$Cov(\mathcal{C}, \eta | R_{obs}) \approx [C_{prior}^{-1} + dF^T W dF]^{-1} \quad (11)$$

where C_{prior} is the covariance matrix of the prior distribution, and dF is the multivariate derivative of $F(\nu; \mathcal{C}, \eta)$ with respect to the parameter vector (\mathcal{C}, η) (and evaluated at the MDP solution). Finally, W is a diagonal weight matrix with $W_{ii} = Var(e_i)^{-1}$. As in the maximum-likelihood case, this approximation will produce good results when the posterior distribution is close to normal. As the name implies, the approximation gets better as the sample size increases (or as the weights go to infinity).

The covariance matrix approximation presented above is most unacceptable, when components of the MPD are on the distribution boundaries; In this case the distribution is quite non-Gaussian, and the asymptotic covariances are too large.

2.1.2 Form of the Regression Priors

To produce Bayesian estimates, one must supply a reasonable prior distribution for the model parameters, as represented by (\mathcal{C}, η) . The effluent concentrations, \mathcal{C} should be independent

of the nuisance parameters, η , so the combined prior can be written as a product;

$$p(\mathcal{C}, \eta) = p(\mathcal{C})p(\eta) \quad (12)$$

The prior on \mathcal{C} , $p(\mathcal{C})$, is uniform over a hyper-cube defined by the vectors $\mathbf{0}$ (lower left-hand vertex of cube) and C_{max} (upper right-hand vertex). Thus $p(\mathcal{C})$ is defined as;

$$p(\mathcal{C}) = \begin{cases} \frac{1}{\prod_i C_{max,i}} & \text{For } 0 \leq \mathcal{C}_i \leq C_{max,i} \\ 0 & \text{Otherwise} \end{cases} \quad (13)$$

This is the standard “non-informative” prior, modified by the bounding information we have about \mathcal{C} . We always have, of course, $0 \leq \mathcal{C}_i$, but one might have some arguments about the exact values to use in the upper bound, C_{max} . In that case, one can select a very large value for the components in C_{max} such as $10^6 ppmM$.

The prior on the nuisance parameters, $p(\eta)$ is more complicated because it must convey realistic information concerning the nuisance parameters. The nuisance parameter vector is defined as $\eta = (\beta_g, \beta_a, T_p, T_a)$, with β_g parameters describing the ground radiance, R_g ;

$$R_g = R_{g0} + \sum_{i=1}^{P_g} V_{gi} \beta_{gi} \quad (14)$$

with R_{g0} representing the mean of the ground radiance, V_{gi} orthonormal vectors, and each β_{gi} representing an independent normal random variable with mean zero and standard deviation of σ_{gi} . In other words, prior distribution of R_g is simply a *degenerate multivariate normal distribution*; Expressing R_g in terms of the β_g simply reduces the dimensionality of the problem, because P_g would be much smaller than the dimension of R_g . The quantities that define this prior distribution (i.e. R_{g0} , V_{gi} , and σ_{gi}) are produced by an analysis of off-plume pixels, as described in Section 2.1.2.

Since the objective of this work is to deal with the effects of clutter, we will make a simplifying assumption about the atmospheric parameters for this initial analysis; It will be assumed that the atmospheric transmissivity, τ_a and temperature, T_a are known. We will also assume that the plume temperature, T_p is known. Our intention is, of course, to relax this simplifying assumption in the next version of this algorithm.

The information for the plume temperature, T_p can be particularly critical for this estimation problem. When the plume is optically thin, it can be shown that the IR data contains no information concerning the plume temperature, T_p . In this case the supplied prior distribution for T_p supplies all information concerning the parameter.

2.2 Matched Filter Model

It is useful to describe a popular alternative to the above model, using the same notation. This popular alternative is commonly called the “matched filter” model even though it would be more accurately described as a linear regression model. The matched filter estimator is derived from the same 3-layer model presented in Equations 1 to 4. If these are combined, one obtains;

$$R_{obs} = \tau_a(1 - \tau_p)(\mathcal{B}(T_p) - R_g) + \tau_a R_g + (1 - \tau_a)\mathcal{B}(T_a) + e \quad (15)$$

The wavenumber, ν , and convolution operator, Ψ , have been dropped to simplify the notation.

Now the last term in this model, $\tau_a R_g + (1 - \tau_a)\mathcal{B}(T_a) + e$, represents the signal without the plume and we can use an evaluation of off-plume pixels can determine the mean and covariance of this. Let us suppose that an off-plume analysis provides us with the mean,

$$\mu_{off} = E(\tau_a R_g + (1 - \tau_a)\mathcal{B}(T_a) + e) \quad (16)$$

and the covariance

$$C_{off} = Cov(\tau_a R_g + (1 - \tau_a)\mathcal{B}(T_a) + e) \quad (17)$$

so that the regression model can be written as;

$$R_{obs} - \mu_{off} = \tau_a(1 - \tau_p)(\mathcal{B}(T_p) - R_g) + e' \quad (18)$$

with $Cov(e') = C_{off}$. We produce a linear regression model by substituting for $1 - \tau_p$ the approximation;

$$1 - \tau_p \approx \sum_i A_i \mathcal{C}_i \quad (19)$$

to obtain

$$R_{obs} - R_{off} = \tau_a(\mathcal{B}(T_p) - R_g)(\sum_i A_i \mathcal{C}_i) + e' \quad (20)$$

In order to use this regression model, all terms except for the concentrations (as represented by \mathcal{C}_i must be known. This implies that estimates of τ_a , R_g , and T_p all must be available. In practice, one uses a very crude estimate for the R_g ; It is commonly assumed that R_g is a Planck function at a nominal temperature. This assumption is somewhat inconsistent with the assumptions that are applied to the off-plume portion of the model, which also contains the term R_g . In fact, one of the reasons our Bayesian regression can be expected to work better than this model is because it utilizes off-plume information in a consistent way.

With this linear regression model now defined, one can produce the standard least-squares estimator for \mathcal{C} which is;

$$\hat{\mathcal{C}} = (X^T W X)^{-1} X^T W (R_{obs} - R_{off}) \quad (21)$$

with;

$$W = C_{off}^{-1}$$

$$X_{ji} = \tau_a(\nu_j)(\mathcal{B}(\nu_j, T_p) - R_g(\nu_j))A_i(\nu_j)$$

Equation 21 represents the “matched filter” estimator, which, as one can see, originates from a linear regression model.

3 Methodology for Empirically Producing Clutter Prior Distributions

In this section, we discuss how one can empirically determine the prior distribution for R_g (background radiance) using the remote-sensing image. The model we use for this study is a multivariate normal prior which is determined from off-plume spectra. This methodology is very similar in spirit to the off-plume analysis used in the match filter. We also discuss an alternative distribution that is expected to model clutter more accurately than a simple normal. It is not yet implemented, but would be the next step in algorithm development.

A general advantage of the Bayesian regression methodology is its ability to incorporate complicated information concerning the nuisance parameters (in this case R_g) into the estimation algorithm in a fairly automatic way.

3.1 Simple Normal Model for Clutter

It is assumed that one has identified a subset of image pixels that are indeed off-plume, and are representative of the image pixels otherwise. The basic strategy is to assume that the prior on R_g is multivariate normal and use the off-plume regression model to relate the mean and covariance of R_g to the mean and covariance of the observed spectra R_{obs} , which can be calculated from the pixel subset.

Since there is no plume, the regression model that describes the relationship of the observed and background spectra in this subset is;

$$R_{obs} = \tau_a R_g + (1 - \tau_a) \mathcal{B}(T_a) + e \quad (22)$$

with τ_a and $\mathcal{B}(T_a)$ considered as known quantities, as well as the variance of the error term, $Var(e)$.

The mean and covariance of these two spectra, R_g and R_{obs} are therefore related by the two formulas;

$$\mu_{obs} = \tau_a \mu_g + (1 - \tau_a) \mathcal{B}(T_a) \quad (23)$$

and

$$C_{obs} = diag(\tau_a) C_g diag(\tau_a) + diag(Var(e)) \quad (24)$$

So μ_{obs} and C_{obs} can be calculated from the observations in the identified off-plume subset and the two formulas presented above can be used to produce estimates of μ_g and C_g , which determines the prior on R_g . In other words, the prior density is;

$$p(R_g) = \phi(R_g; \mu_g, C_g) \quad (25)$$

In order to reduce the number of parameters required to describe the prior distribution, an eigen-value decomposition can be performed on C_g to produce the representation in Equation 14.

This methodology can really be classified as an *empirical Bayes approach*, (see [1]), a statistical methodology that is used to estimate priors in Bayesian problems. It should be

noted that many variants of this strategy could be employed (and that is one of its strengths). The next section discusses an empirical Bayes methodology for fitting a better distribution than a multivariate normal. However, the simple normal model for the prior is the prior used in all actual computations that this report presents.

3.2 Improved Prior for Clutter

A simple multivariate normal cannot be expected to fit the background radiance well when the background consists of highly inhomogeneous pixels. Background radiance from a particular type of pixel (for example water pixels, or asphalt pixels) can be expected to produce a distribution that is roughly normal, but when different types of pixels are mixed together, one can expect the resulting distribution to resemble a mixture of normals. So, to improve the clutter prior distribution, we propose replacing the simple normal prior with a “mixture of normals” prior.

The mixture of normals prior is defined as;

$$p(R_g) = \sum_{i=1}^M \rho_i \phi(R_g | \mu_i, C_i) \quad (26)$$

where the index i represents the different “types” of pixels in the scene. The parameter, ρ_i , describes the proportion of type i pixels in the scene, and (μ_i, C_i) define the multivariate normal distribution applicable to type i pixels. There are two methods one can use to build a mixture prior of this form, and the methodology one would favor depends upon what information is available.

Pixel Classification Available: If the pixels in the scene were classified into homogeneous background types by some auxiliary analysis, one could use this classification to define the groups. Under this set of circumstances, the off-plume pixels could be broken into M types, and the mean, μ_i , and covariance C_i would be calculated for each type with the use of Equations 23 and 24. The proportion, ρ_i , would be set to the proportion of pixels in type i . This would produce a complete mixture prior for clutter.

Pixel Classification Unavailable: If the off-plume pixels were not already classified into groups, one could use a hierarchical cluster algorithm (such as Splus’s *hclust*, [9]) to perform this task. The resulting clusters would not necessarily represent unique pixel backgrounds, but this fact would not cause problems.

Once the off-plume pixels were divided into suitable classes, one would estimate the ρ_i , μ_i , and C_i using exactly the same methods as in the previous case.

Another feature that might be valuable to incorporate into the clutter prior is some smoothness; Background spectra are known to be relatively smooth (as compared to gas spectra), so it may be valuable to have the prior reflect this. To accomplish this we might choose a class of smooth functions, such as cubic splines, to represent the R_g . We would substitute the spline model into regression model presented in Equation 22 and fit the model

to each pixel's data to produce an estimate for R_g . The mean and covariance matrix for the resulting R_g would then be determined as described previously. However the new covariance matrix would contain smoothness not present in the unsmoothed covariance matrix.

4 A Simulated Example

To evaluate the Bayesian estimator, we have constructed a simple Monte Carlo program that simulates R_{obs} according to the model described in Equations 1 to 4. Since this is the very model that is used to derive the Bayesian regression estimator (and also matched filter estimator), we would expect its performance to be good. This report will not investigate the performance of the estimator under other sets of conditions; for example when the physics-based model is more complex than the simple 3-layer model we have described.

The Monte Carlo simulates R_{obs} differently in one respect than the regression model; The Bayesian regression model assumes that R_g has a multivariate normal distribution, but the simulation assumes that this distribution is discrete (in conformance to [8]). A set of 6 different materials (chosen from the NEFDS data set) are used as backgrounds and their measured emissivities are used to construct R_g . The 6 different materials are randomly sampled to produce the backgrounds used in the simulation. The emissivities actually used are presented in Figure 2. The emissivities were chosen to be the extreme representatives from the NEFDS data set, with the objective of presenting the estimation algorithms with a difficult clutter problem. Obviously, the difference in performance between the Bayesian regression and matched filter estimator will be diminished as the backgrounds are made more homogeneous.

Figures 3 through 4 present the intermediate results of a single simulation. Figure 3 presents R_g , the background radiance and R_p , the plume radiance. As one can see, R_p closely resembles R_g , except for some very small bumps. These bumps are the "signal" that the estimation algorithm must use to produce gas concentrations. From this plot, one can see how important clutter can be to this problem; the bumps that represent the signal could easily be mistaken for variations in the background signal.

The size of the signal presented in Figure 3 depends upon the gas concentration and the temperature contrast (i.e., temperature difference between plume and ground). In the illustrated case, these are relatively small; the temperature contrast is 5C and the effluent concentrations are 10ppmM. Increasing either of these quantities would make the gas signal larger.

Figure 4 illustrates what happens to the spectrum after it travels through the atmosphere; The resulting curve, R_a , is an average of R_p and the atmospheric black-body function, $\mathcal{B}(T_a)$. The final figure, Figure 5, illustrates what instrument noise does to the signal. The instrument noise is typical for a dispersive instrument, about 0.2% of the observed spectra.

The absorbance spectra of the three gases included in the plume are illustrated in Figure 6. One gas, ammonia, has one of the strongest absorbance spectra among potential effluents. It also has an absorbance spectra that differs markedly from background emissivity spectra and thus should be easiest to estimate. The other effluent absorbance spectra are smoother, and therefore could be more easily confused with background clutter.

Figure 2: Emissivities of Materials used in the Simulation
Ground Emissivity

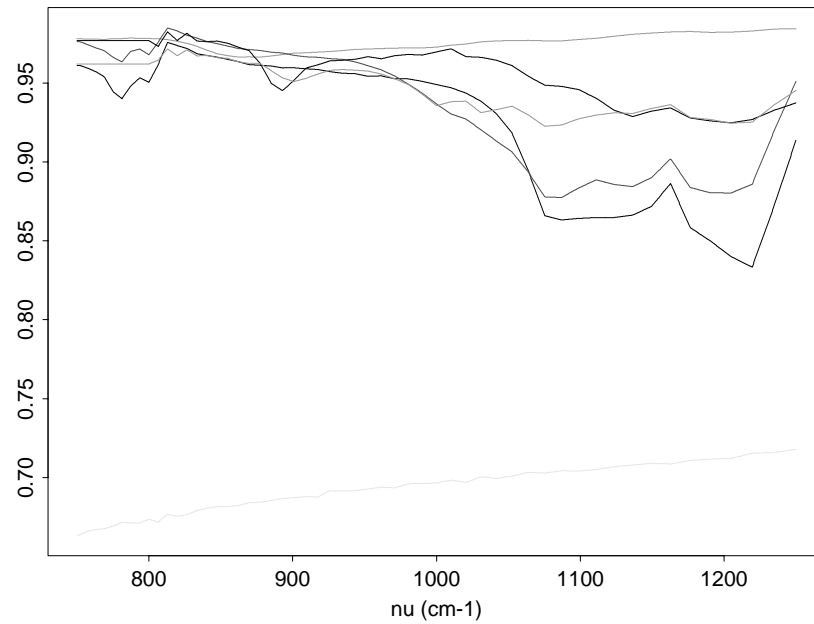


Figure 3: Simulation of R_g and R_p
Rg and Rp of Simulated Spectrum

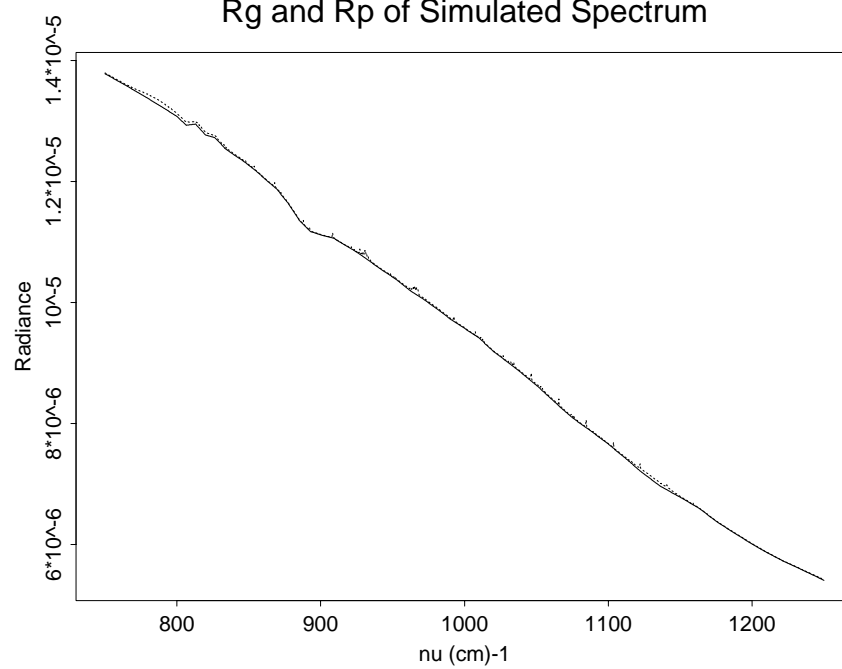


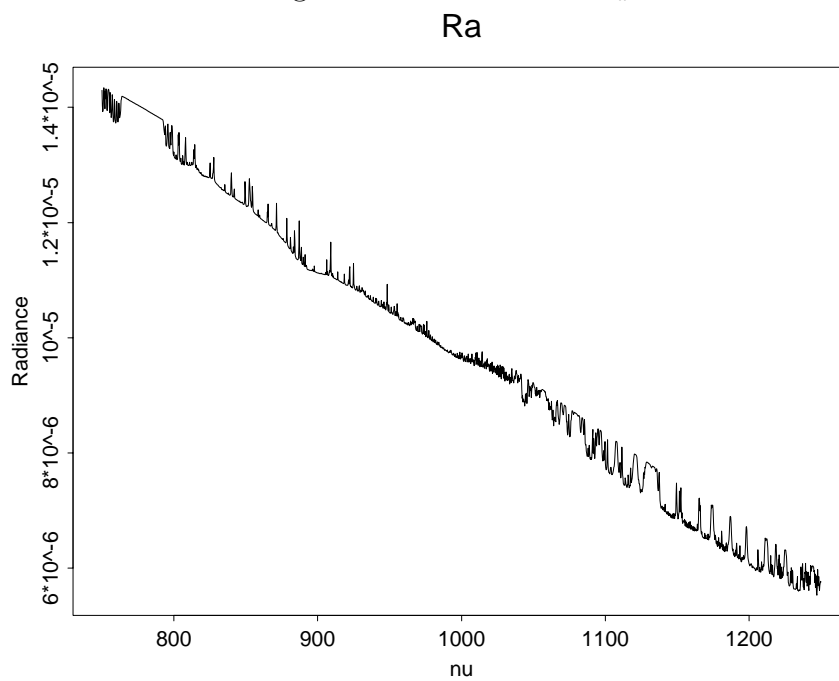
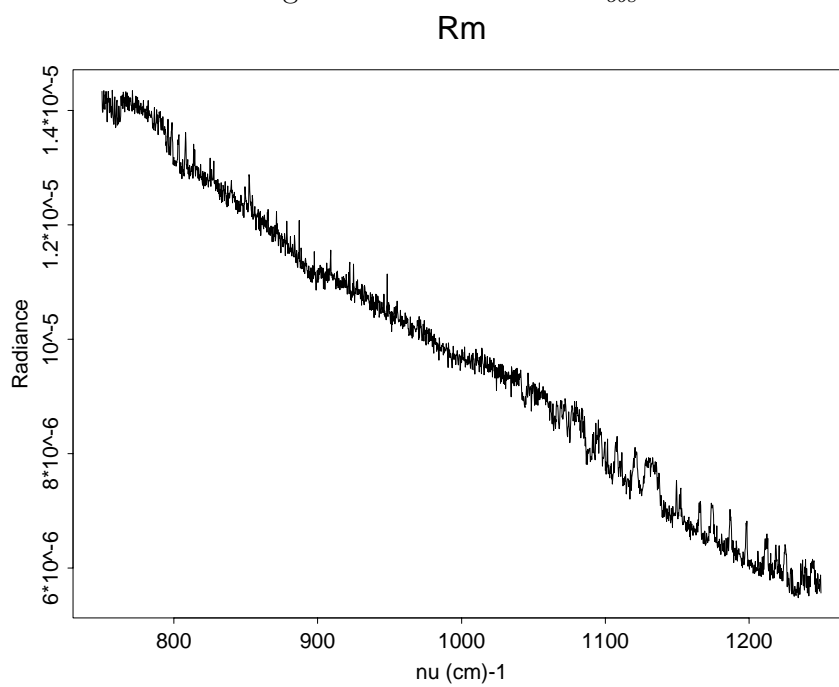
Figure 4: Simulation of R_a Figure 5: Simulation of R_{obs} 

Figure 6: Gases Included in the plume
Absorbance Spectra

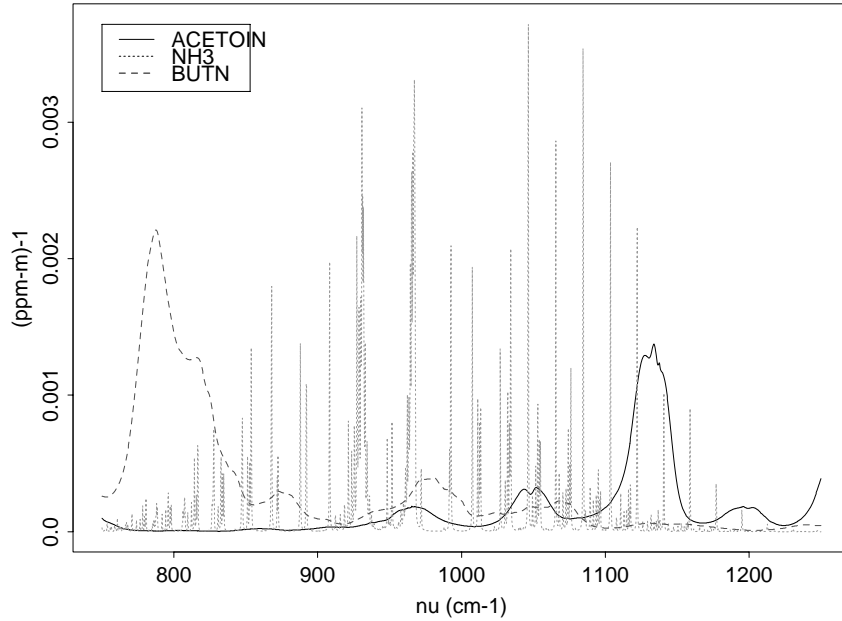


Table 1 summarizes the major details of the estimation scenario that will be used to evaluate the estimators described in this report. As one can see, this scenario focuses on background clutter as a major source of variability.

5 Comparison of Bayesian Clutter Suppression Model to Matched Filter

Approximately a thousand spectra were simulated under the scenario conditions listed above. 500 "off-plume" spectra were simulated and used to estimate (1) the prior distribution for the Bayes estimator, and (2) the mean and covariance for the matched filter. The estimators were applied to 450 spectra whose effluent gas concentrations ranged from 0ppmM to 110ppmM. The concentration of the 3 gases were varied in unison over this range.

Tables 2 to 4 present a comparison of the RMSE (root mean squared error) for the two estimators at selected concentrations. The tables also break RMSE into bias and standard error. Finally, each table presents the "estimated standard error"; This is the standard error calculated by the algorithm, and should conform to the standard deviations produced by the simulation. These estimated standard errors are an important component of an estimation algorithm. One might go so far as to maintain that an estimation algorithm that cannot produce a good description of its uncertainty is worse than useless. Of course, a standard error is a good description of uncertainty only if the estimator bias is near zero.

From the Tables, one can see that the matched filter estimated standard error grossly

Table 1: Scenario Details

Instrument Error: 0.2% of gross background radiation.

Inst. Resolution: 0.25cm^{-1} .

Temperatures: Average ground temperature (T_g): 300C. Average plume temperature, T_p , and atmosphere temperature, T_a : 302C. (So there is a 2 degree temperature contrast).

Temperature Uncertainty: $Stdev(T_g) = 0.5C$, $Stdev(T_p) = 0.1C$ and $Stdev(T_a) = 0.1C$.

Atmospheric Transmissivity, τ_a : Assumed known. Represents the standard temperate summer atmosphere.

Effluents Present: Acetoin, NH₃, and Butylnitrite.

Effluent Concentration: Concentrations vary from 0 to 110ppmM.

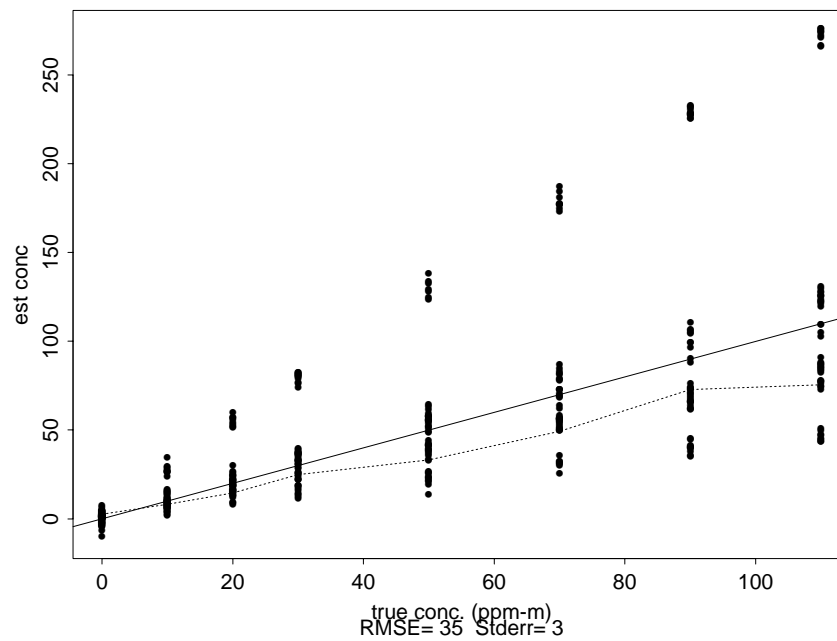
Background Emissivity: Emissivity vectors randomly chosen from 6 emissivities displayed in Figure 2.

Table 2: Acetoin Statistics for Bayes Regression vs Matched Filter

True Conc.	Bias	Std Dev	RMSE
Matched Filter	Est. SE: 3.33		
0	-0.379	2.48	2.49
10	1.076	8.16	8.15
50	3.507	33.27	33.12
90	16.305	71.58	72.72
Bayes Regr.	Est. SE: 4.21		
0	1.671	2.98	3.40
10	-0.817	4.65	4.68
50	-0.877	5.43	5.44
90	1.882	5.34	5.61

(All Values in ppmM)

Figure 7: Estimates for Acetoin
Matched Filter Results, ACETOIN



NLMPD Results, ACETOIN

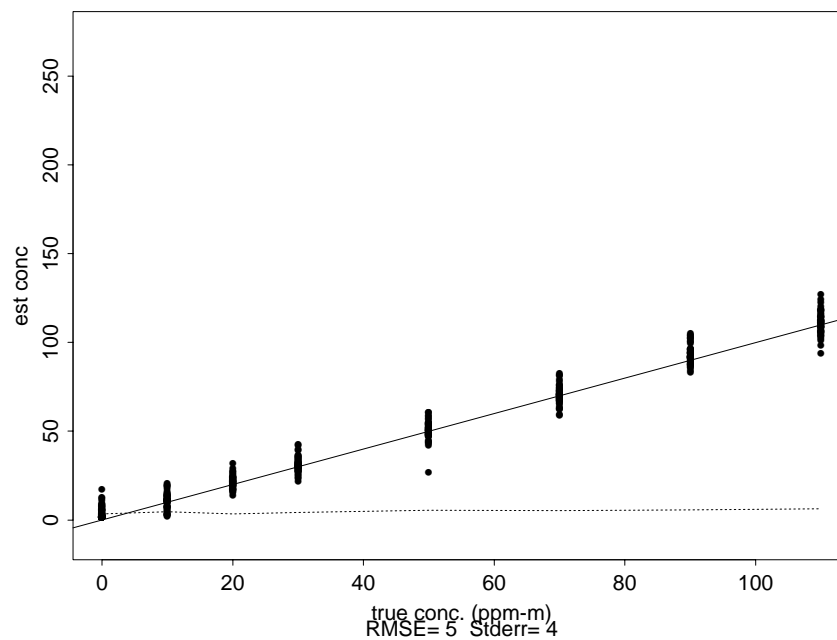


Table 3: NH3 Statistics for Bayes Regression vs Matched Filter

True Conc.	Bias	Std Dev	RMSE
Matched Filter	Est. SE: 1.40		
0	0.0230	0.959	0.954
10	-0.0677	8.029	7.948
50	-8.6668	34.548	35.282
90	-3.0725	72.332	71.671
Bayes Regr.	Est. SE: 2.21		
0	0.5746	1.02	1.16
10	0.3971	1.64	1.67
50	-0.1560	2.64	2.62
90	0.3561	2.69	2.68

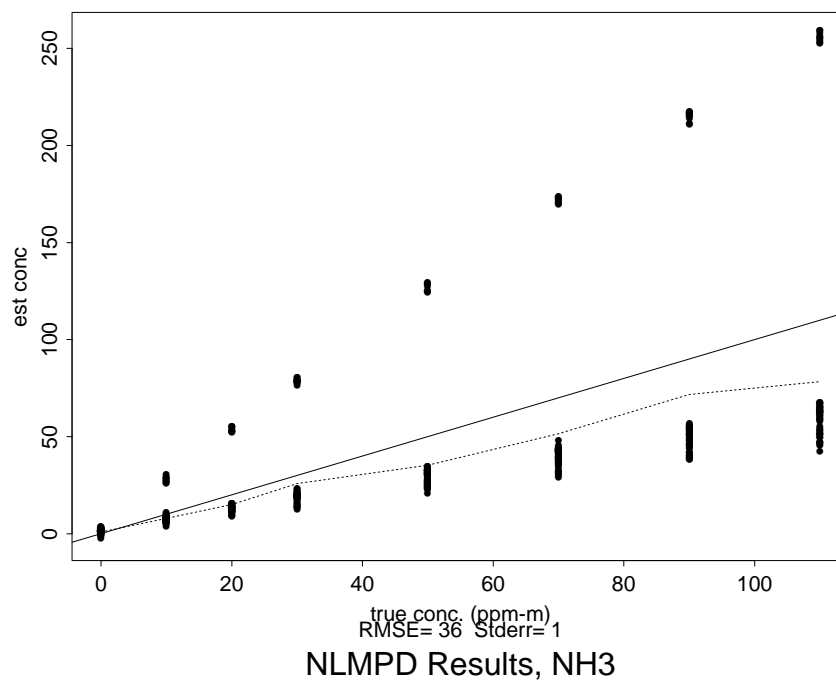
(All Values in ppmM)

Table 4: Butylnitrite Statistics for Bayes Regression vs Matched Filter

True Conc.	Bias	Std Dev	RMSE
Matched Filter	Est. SE: 2.49		
0	0.264	1.67	1.68
10	-1.027	10.54	10.49
50	-11.216	43.53	44.52
90	-1.062	93.26	92.33
Bayes Regr.	Est. SE: 3.79		
0	1.5313	2.58	2.99
10	0.1324	3.11	3.08
50	0.0737	3.93	3.89
90	-0.3439	4.82	4.78

(All Values in ppmM)

Figure 8: Estimates for NH₃
Matched Filter Results, NH₃



NLMPD Results, NH₃

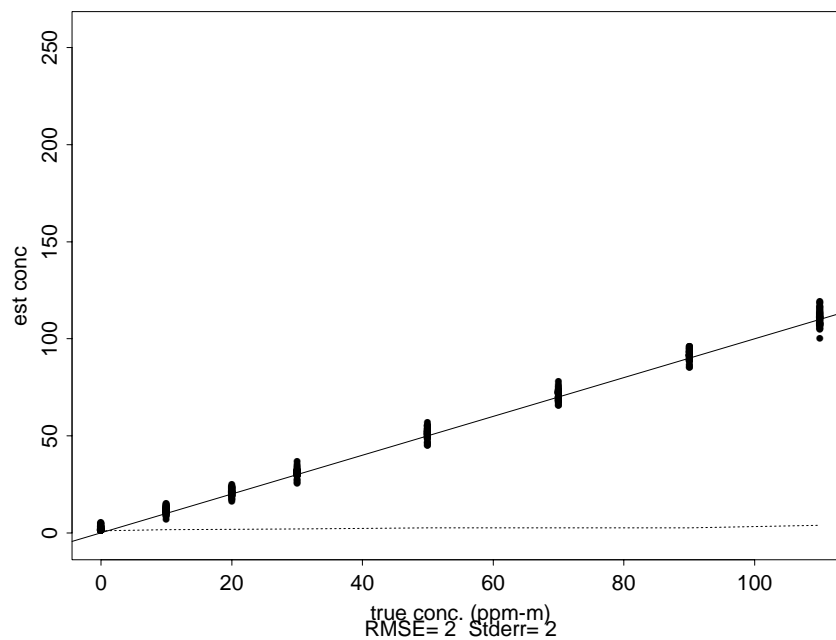
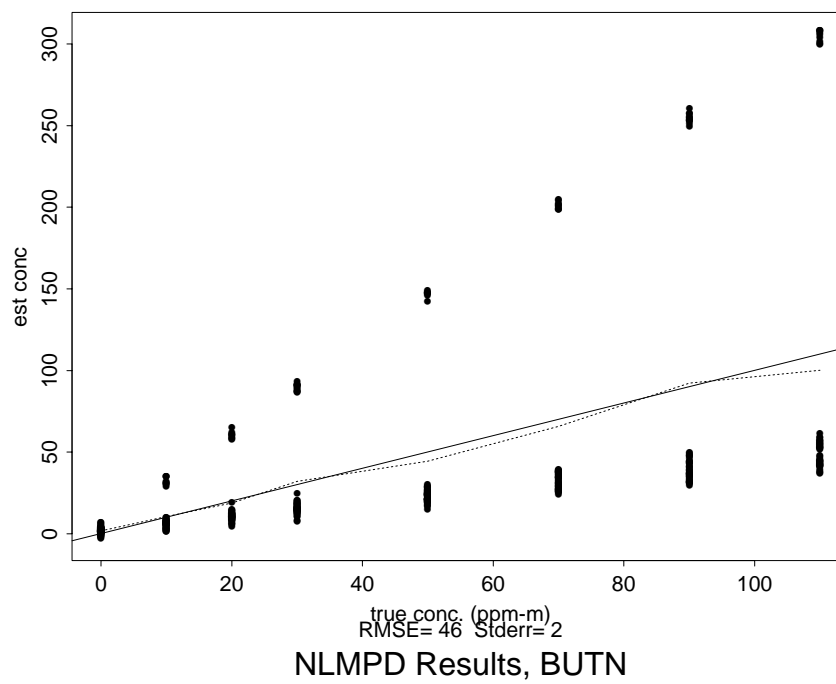
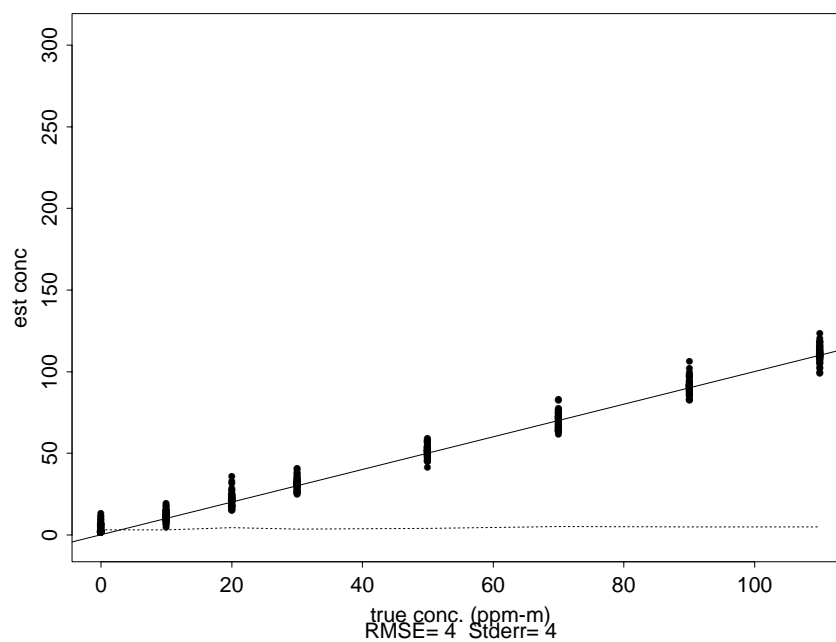


Figure 9: Estimates for Butylnitrite.
Matched Filter Results, BUTN



NLMPD Results, BUTN



underestimates error, except when the gas concentrations are zero. Also, for the matched filter, a single standard error is inappropriate, because error grows with effluent concentration. In fact a proportional error description would be most appropriate with the relative standard deviation (RSD) for the three gases being from 80% to 100%. With a RSD this large, the matched filter can function poorly even as a gas detector. For the 90ppmM simulation results, the estimated standard error is off by a factor of 20.

On the other hand, the Bayesian estimated standard error is extremely accurate. The estimated standard error differs by 20 to 30% from the simulated results. Also, the error is practically independent of effluent concentration, so a single standard error can describe uncertainty. For the 90ppmM simulations, the Bayesian RMSE is a factor of 10 lower than the matched filter results.

Also, the matched filter estimator exhibits significant bias, but the Bayesian estimator is practically bias free.

In Figures 7 through 9 the estimated values are plotted against true concentrations, to allow the reader to visually compare the error. The dotted lines in the figures represent the RMSE at each concentration.

We have run a few variants on the scenario described above to determine how certain nuisance parameter uncertainties would affect the results and found that T_p is the most critical nuisance parameter. If we assume that less is known about either the ground or atmospheric temperatures, (T_g, T_a) , Bayesian regression still behaves well because it can estimate these parameters from the data. On the other hand, the plume temperature, T_p cannot always be estimated from the data. When the gas concentrations are near zero (so that the plume is optically thin), the regression algorithm simply propagates the error in T_p into the concentration estimates, so that a correct standard error is produced.

For the larger concentrations (i.e., 100ppmM), some information about T_p can be extracted from the data, and one can be less concerned about the a-priori information supplied for T_p . In all of these cases, Bayesian regression performs better than the matched filter, however the difference in RMSE is only a factor of 5. These results are obtained when the prior standard deviations on the temperatures are set to 1C.

6 Conclusions

The Bayesian regression model developed in this report does significantly better at dealing with background clutter than current matched filter methods. Besides producing more accurate estimates, the Bayesian regression model produces uncertainty estimates that are accurate, while the matched filter does not. The Bayesian regression methodology also provides a framework that could be easily extended to include other sources of prior information and uncertainty.

Specific conclusions are;

Bayesian regression requires a prior distribution to describe clutter (i.e. background variability). This can be estimated from a suitable set of off-plume pixels.

Bayesian regression can dramatically reduce the effects of clutter; in realistic cases, by an order of magnitude.

The Bayesian gas concentration estimates are virtually unbiased.

For Bayesian regression, uncertainties are almost independent of gas concentration, while for matched filters, it is proportional to gas concentration.

The calculated uncertainty values are in good agreement with the simulated results, while those of the matched filter too small by orders of magnitude.

7 Glossary of Terms

\mathcal{C} : A vector of gas concentrations (measured in ppmM) present in the plume.

$Cov(x)$: The covariance of the random vector x .

$diag(x)$: A diagonal matrix with entries x_i on the main diagonal.

$E(x)$: The expected value of the random vector x .

$\phi(x, \mu, C)$: A multivariate normal density with mean of μ and covariance matrix of C .

$\mathcal{B}(\nu, T)$: A Planck black-body function at temperature T , in Kelvin. ν is the wavenumber in cm^{-1} .

η : This is used to represent the “nuisance parameters” in a regression model.

R_g, R_p, R_a, R_{obs} : Radiance spectra, with the subscripts indicating the origin of the spectra; g stands for “ground,” p stands for “plume,” a for “atmosphere,” and finally obs for the observed spectrum.

RMSE: Represents root mean squared error, which measures the error between the true value and an estimate; $RMSE = \sqrt{E((Est - True)^2)}$.

τ_a : Transmissivity of the atmosphere.

8 References

- [1] J.O. Berger, "Statistical Decision Theory and Bayesian Analysis," Springer Series in Statistics, N.Y., 1985.
- [2] "Non-conventional Exploitation Factors Data System," (NEFDS), 1992.
- [3] P.V. Villeneuve, A.D. Stocker, "HIRIS Algorithm Physics Model Description, Version 2.3" Space Computer Corp., 12121 Wilshire Blvd, Suite 910, Los Angeles, CA., prepared for LLNL, June 2, 2000.
- [4] C.C. Borel, W. Clodius, "Recipes for Writing Algorithms for Atmospheric Corrections and Temperature/Emissivity Separations in the Thermal Regime for a Multi-Spectral Sensor," LA-UR-01-1442, LANL, SPIE'99, Conf 4381-24, 1999.
- [5] C.C. Borel, "Iterative Retrieval of Surface Emissivity and Temperature for a Hyperspectral Sensor," First JPL Workshop on Remote Sensing of Land Surface Emissivity, May 6, 1997.
- [6] M.S. Bazaraa, H.D. Sherali, C.M. Shetty, "Nonlinear Programming, theory and algorithms," Wiley, N.Y., 1993.
- [7] Y. Bard, "Nonlinear Parameter Estimation," Academic Press, N.Y., 1974.
- [8] E.T. Scharlemann, "Modeling of passive and active remote sensing systems II," technical report, LLNL, March 16, 2000.
- [9] W.N. Venables, B.D. Ripley, "Modern Applied Statistics with S-PLUS," Springer, N.Y., 1997.