

A Bayesian approach to identification of gaseous effluents in passive LWIR imagery

Shawn Higbee^a, David Messinger^b, Yolande Tra^b, Joseph Voelkel^b, and Lawrence Chilton^c

^aAir Force Institute of Technology, Rochester, New York, 14623, USA;

^bRochester Institute of Technology, Rochester, New York, 14623, USA;

^c Pacific Northwest National Laboratory, Richland, WA 99352, USA

ABSTRACT

Typically a regression approach is applied in order to identify the constituents present in a hyperspectral image, and the task of species identification amounts to choosing the best regression model. Common model selection approaches (stepwise and criterion based methods) have well known multiple comparisons problems, and they do not allow the user to control the experiment-wise error rate, or allow the user to include scene-specific knowledge in the inference process.

A Bayesian model selection technique called Gibbs Variable Selection (GVS) that better handles these issues is presented and implemented via Markov chain monte carlo (MCMC). GVS can be used to simultaneously conduct inference on the optical path depth and probability of inclusion in a pixel for a each species in a library. This method flexibly accommodates an analyst's prior knowledge of the species present in a scene, as well as mixtures of species of any arbitrary complexity. A series of automated diagnostic measures are developed to monitor convergence of the Markov chains without operator intervention. This method is compared against traditional regression approaches for model selection and results from LWIR data from the Airborne Hyperspectral Imager (AHI) are presented. Finally, the applicability of this identification framework to a variety of scenarios such as persistent surveillance is discussed.

Keywords: long wave infrared, Bayesian inference, Markov chain monte carlo, Gibbs sampling, Gibbs Variable Selection

1. INTRODUCTION

In the portions of the electromagnetic (EM) spectrum roughly from $8 \mu m$ to $12 \mu m$ there is a “window” of relatively high atmospheric transmission. This window is often called the long-wave infrared (LWIR) and it allows the observation of chemical phenomenology of gases that relate to the rotation and vibration of their bonds such as when they selectively absorb incident radiation or self-emit radiation in specific wavelengths as determined by the bond structure of the gas. One can make laboratory measurements of the absorption or emission properties of gas species at very high resolutions¹ for later use in identifying a sample of an unknown gas.

In a remote sensing context, where one does not have the ability to directly capture and analyze the gas of interest, one can attempt to observe the sample *in situ* in a scene through hyperspectral imaging (HSI) sensors. These sensors produce an “image cube” that has spatial dimensions (x, y) and a spectral dimension λ . Thus every observation in the image cube has a unique index $\mathbf{L}_{x,y,\lambda}$ such that $x \in [1, 2, \dots, X]$, $y \in [1, 2, \dots, Y]$, and $\lambda \in [1, 2, \dots, \Lambda]$ and any spatial location in the image cube has a corresponding $(\Lambda \times 1)$ column vector of measurements $\mathbf{L}_{x,y}$ for that pixel. Note that all vectors are assumed to be column vectors unless otherwise specified and all variables that are vectors or matrices are in boldface.

These image cubes and the sensors that generate them are created not only as a means of detecting the presence of a gas, but also identifying its makeup by comparison to a library of known spectra. The phenomenology of such situations involves the chemical makeup of the gas, the radiometric properties of the scene underneath the gas, and the atmosphere through which passively generated LWIR energy must propagate in order to be observed by a remote sensor.

The radiometric properties of the scene uniquely contribute to the challenge of this problem,² because it is the combination of the physical properties of the background and gaseous plume (which is a function of the temperature difference between the two) that contribute to the observed spectral signature from the plume pixel.

Frequently some sort of regression methodology is applied in order to estimate the quantity of constituents in a plume, often measured only to a multiplicative constant (*i.e.*, one can only report the relative proportions of each gas). This leads to the problem that when one has a large library of reference measurements, several different competing mixtures of gases may fit the data (more or less) equally well. This problem is termed one of “model uncertainty” and is a key factor in being able to determine the composition of the gas in a plume. Physically the models are mutually exclusive (only one can be correct) and thus one might wish to make a probabilistic statement about the probability of any given model, or lacking that, the probability that a predictor belongs in a model.

This paper demonstrates a novel methodology that may be applied to the problem of choosing an appropriate model (and thus the correct mixture) of gases in the plume. The sections below will describe in detail the radiometry and physical assumptions of the physical problem, the use of a Bayesian model selection procedure, and how it can be applied to the effluent identification problem, and finally results from an experiment using the method will be presented.

2. BACKGROUND

The basic phenomenology that gaseous effluent detection and identification relies upon in passive LWIR imagery is the interaction between EM radiation and the dipole moment of gases’ molecular bonds.³ Strong interactions between EM radiation and a gas occur when an oscillating dipole moment (be it rotational or vibrational) of the gas has the same frequency as the EM radiation interacting with it. One can use a spectrometer to precisely measure the absorption spectra of known gases¹ for the purposes of creating an absorption spectra library for the purpose of comparison for other unknown species.

For most practical scenarios in remote sensing the observed radiance \mathbf{L} in question will have a large contribution from the background (soil, vegetation, buildings, etc.) behind or underneath the gas. That radiation of energy can be modeled as a black body as given by

$$\mathbf{B}(T) = \frac{2\pi hc^2}{\lambda^5 (e^{\frac{hc}{\lambda kT}} - 1)} \cdot 10^{-6} \quad [Wm^{-2}sr^{-1}\mu m]. \quad (1)$$

In Equation 1, k is Boltzman’s constant ($1.38 \times 10^{-23} JK^{-1}$), c is the speed of light, h is Plank’s constant, and T is the temperature of the object in degrees Kelvin. The units of radiance (\mathbf{L}) are watts per meter squared per steradian per micron. Note that once temperature T is specified the user can retrieve via simple calculation, the radiance such an object will radiate for any wavelength λ . How much EM energy any object will radiate is determined by the element-wise multiplication of the material’s emissivity ξ and a blackbody of the appropriate temperature $\xi BB(T)$

There are two other sources of radiance that are important (also $\Lambda \times 1$ vectors): “upwelled radiance” (\mathbf{L}_u) is the self-emitted radiance of the air-column through which the image is taken, and “downwelled radiance” (\mathbf{L}_d) is the self-emitted radiance of the warm atmosphere down onto an object and then reflected back towards the sensor. All of the radiance from a point on the ground, either self emitted or reflected downwelling, then must pass through the atmosphere τ_a which is not perfectly transparent in the LWIR and amounts to another element-wise multiplication by a $\Lambda \times 1$ vector on its path to the sensor.

Additionally, if there is a gaseous plume present in the scene, it will act as an additional “filter” over the ground leaving radiance τ_{plume} and act as an emitter of radiance itself ($\mathbf{B}(T_{plume})$) both of which are also $\Lambda \times 1$ vectors. Note that the plume and the ground are at different temperatures, and have different emissivities, etc.

These terms are combined into the equation for sensor-reaching radiance (Equation 2)⁴ for every pixel in a scene. For ease of understanding, first consider a single non-plume pixel, whose sensor-reaching radiance is given by the element-wise addition/multiplication of

$$\mathbf{L}_{(x,y)} = [\boldsymbol{\xi}_s \cdot \mathbf{B}_s + (1 - \boldsymbol{\xi}_s) \cdot \mathbf{L}_d] \cdot \boldsymbol{\tau}_a + \mathbf{L}_u \quad (2)$$

where $\boldsymbol{\xi}_s, \mathbf{L}_d, \boldsymbol{\tau}_a, \mathbf{L}_u$ are the surface emissivity, downwelling radiance, atmospheric transmission, and upwelling radiance respectively. Note that the $\mathbf{L}_u, \mathbf{L}_d$ and $\boldsymbol{\tau}_a$ terms are treated as constants over the entire scene.

Now consider the more complicated case where a plume is present, by simply adding another two terms into Equation 2 to create Equation 3. This amounts to propagating the ground-leaving terms through the plume $\boldsymbol{\tau}_{plume}$ (again via element-wise multiplication) and adding in the self-radiance of the plume itself $\boldsymbol{\xi}_{plume} \mathbf{B}(\lambda, T_{plume})$. The on-plume pixel radiance is now written as

$$\mathbf{L} = [(\boldsymbol{\xi}_s \mathbf{B}(T_s) + (1 - \boldsymbol{\xi}_s) \mathbf{L}_d) \boldsymbol{\tau}_p + \boldsymbol{\xi}_p \mathbf{B}(T_p)] \boldsymbol{\tau}_{atm} + \mathbf{L}_u. \quad (3)$$

Assuming the plume obeys the Beer-Lambert Law, one can apply a thin gas approximation, ignoring the dependence of the absorption spectra on temperature, and approximate $\boldsymbol{\tau}_p(\lambda) \approx 1 - \boldsymbol{\xi}_p \approx 1 - \sum \mathbf{a}_l C_l$ by making use of the first two terms of the Taylor's series expansion of $\boldsymbol{\xi}_p$. Each \mathbf{a}_l is a vector of the l^{th} constituent gas absorption spectra (which are known from standard laboratory measurements and downsampled to match the spectral resolution of the other terms) and C_l are their concentration path lengths (CPL). Here one assumes that the gas is well mixed and the CPL's are really the integral of concentrations of each constituent gas along the path created by the projection of the detector onto the ground (note: the summation is over the gas(es) which are present, and the integral for the path distance)

$$C_l = \int_{plume}^{sensor} \sum_{l=1}^q c_l \, dpath. \quad (4)$$

Because one does not know the thickness of the gas layer (or its density distribution) in practice it is only possible to conduct inference on the CPL of the gas(es) present and not on the absolute magnitude of gas present. In other words, $C_l \propto$ concentration of the set of l gases (c_1, c_2, \dots, c_l) and in order to do exact inference on c_l one would need to know the plume's optical path distribution ($\int dpath$ in Equation 4).

With the above substitutions and expanding the terms from Equation 3, we generate Equation 5. Equation 5 is practically important for this problem because it lets us predict whether the gas plume will absorb some of the energy from the ground behind it, or emit additional energy with it. This absorption/emission phenomena is dependent on the sign of the center term, which is sometimes called the "contrast term" ($[\mathbf{B}(T_p) - \boldsymbol{\xi}_s \mathbf{B}(T_s) - (1 - \boldsymbol{\xi}_s) \mathbf{L}_d]$). Qualitatively the black body for the plume and the ground "compete" with one another in this term, and if the contrast term is net positive, the plume will be in emission ("adding spikes to the spectra") and if it is net negative the plume will be in absorption ("cutting notches out of the spectra"). Exactly where those spikes/notches are, and their magnitude is governed by the weighted sum of their absorption spectra (\mathbf{a}_l multiplied by their concentration path lengths C_l).

$$\mathbf{L} = [\boldsymbol{\xi}_s \mathbf{B}(T_s) + (1 - \boldsymbol{\xi}_s) \mathbf{L}_d + \sum \mathbf{a}_l C_l [\mathbf{B}(T_p) - \boldsymbol{\xi}_s \mathbf{B}(T_s) - (1 - \boldsymbol{\xi}_s) \mathbf{L}_d]] \boldsymbol{\tau}_{atm} + \mathbf{L}_u \quad (5)$$

It is important to note that when the contrast term is net zero, the plume is neither in emission nor absorption and thus does not modify the spectral signature of the underlying scene. Thus it is only when this contrast term is (measurably) non-zero and the physical assumptions regarding the behavior of a thin plume hold, that one can observe the optical manifestation of the gas plume.

3. METHODS

A recently developed technique known as Optimized Land Surface Temperature and Emissivity Retrieval (OLSTER)⁵ uses an iterative gradient decent search, guided by a variety of physical constraints to estimate the scene-wide $\boldsymbol{\tau}_{atm}, \mathbf{L}_d$, and \mathbf{L}_u in addition to $\boldsymbol{\xi}_s$ and T_s for every pixel. The algorithm makes simultaneous use of several well-known atmospheric models including MODTRAN (MODerate resolution atmospheric TRANsmis-sion)⁶ and ISAC (In Scene Atmospheric Compensation) along with empirical constraints on certain parameters,

for example ξ_s is constrained to be a “smooth” function, in order to obtain the previously mentioned parameter estimates.

Practically, the ability to do temperature/emissivity separation is tremendously important for several reasons. First, the material specific part of the background $\xi_s \mathbf{B}(T_s)$ is only the ξ_s term. Thus it enables remote sensing to be applied directly to the problem of material identification. The reason it is most valuable for this research is that it effectively enables the process of “taking apart the hypercube.” Recall that the output of the sensor was the left-hand side of Equation 2, and OLSTER gives us the ability to individually obtain point estimates for all of the terms on the right-hand side of that equation. Further, given a library of known gas absorption spectra, and a known sensor response function, one can predict what the sensor would see, for a given scene, with any arbitrary plume inserted.

This enables the creation of very realistic data products that the authors refer to as “semi-synthetic” data, which are created in large part from real data, with only an artificially inserted plume, using point estimates of the remaining terms in Equation 3. One can then explore interesting scenarios such as applying various spatial plume models, performing sequential inference, and examining the background-dependency of plume identification.

The sensor used to generate these data is the Airborne Hyperspectral Imager (AHI) that was built by the University of Hawaii that collects a nominal 256 bands in the 7.5 μm to 11.5 μm wavelength region.⁷ Because the signal to noise ratio (SNR) was deemed poor for the data in hand, band summation was performed in order to create 50 output bands (nominal band centers range from 8.1 μm to 11.5 μm). The scenes used for this research comes from an EPA study from 2003 where AHI imagery was collected of a chemical refinery in Texas. The imagery includes a mixture of agricultural and industrial land use that includes a wide variety of natural and man-made materials.

The gas library used for reference was obtained from the Department of Energy’s Pacific Northwest National Labs (PNNL) vapor-phase infrared spectra database. These spectra were collected through Fourier Transform InfraRed (FTIR) spectroscopy from 1.6 μm to 16 μm . The authors used the 25°C library of spectra as the difference between the down sampled libraries was trivial. One important artifact of working with a spectral library, is the issue of multicollinearity within the library, which is made dramatically worse by the process of downsampling the spectra to 50 bands.

There are two common approaches to measuring the degree of multicollinearity in a matrix, its condition number and variance inflation factors (VIF). The condition number is a single metric, defined as the ratio of the largest to the smallest singular values of a matrix, that is useful in approximating the relative error experienced in a linear system of equations from perturbations in the inputs. Alternatively, VIFs are computed for the design matrix \mathbf{X} by regressing each column of \mathbf{X} against all of the other columns, then transforming the R^2 values for each of the regressions via

$$VIF_i = \frac{1}{1 - R_i^2}. \quad (6)$$

General rules for VIFs being too high are values over 5, in which case the β_i with the highest VIF_i should be removed and the process repeated until all of the VIFs are below the threshold.

The tremendous impact that the spectral downsampling of the data has on the regression process can be seen by comparing these two metrics for the 256 band library and the 50 band library. The original 31 gas library at 256 band spectral resolution has a condition number of 8×10^6 with all of the VIFs under 25 and all but two under 10. In comparison the library that was downsampled to match the 50 band resolution of the AHI data has a condition number of 1×10^7 and a maximum VIF of 146 while over half of the library has VIFs over 10.

In order to reduce the multicollinearity of the gas library, the authors used a two stage approach. First VIFs were sequentially computed, removing gases with a $VIF_i > 5$ and then an eigenvector approach via the singular value decomposition (SVD) was applied to the remainder of the library. The library \mathbf{X} was decomposed into

$$\mathbf{X} = \mathbf{U}\mathbf{W}\mathbf{V}' \quad (7)$$

and by visual inspection of the *last* columns of \mathbf{V}' in comparison to the percentage of variability attributed to each right singular vector \mathbf{W} , one can identify the combinations of gases that are most linearly degenerate. This process is accomplished in the same manner, but with the opposite intent as principal components analysis. By noting the gases which have a large contribution to the last few columns of \mathbf{V}' it is possible to determine which gases contribute to the most linearly degenerate subspace of the eigenvectors that span $\mathbf{X}'\mathbf{X}$. Through this two stage process, the library was thinned from 31 to 15 gases and the condition number decreased to 4×10^2 . It is important to note that this dramatic thinning of the library would not have been necessary if image data of a higher spectral resolution were available.

Often the problem of gas plume identification is cast in the form of an ordinary least squares (OLS) linear regression problem. This can be accomplished by noting that in Equation 5 the absorption spectra are linear with respect to the plume CPL coefficients C_l , arranged as columns to form \mathbf{X} in the canonical regression equation

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (8)$$

where \mathbf{Y} is a $\Lambda \times 1$ “response” vector and is made up of observed radiance values, and \mathbf{X} is known as the “design matrix” (typically) $\Lambda \times (L + 1)$ where $l \in [1 \dots L]$. The error process $\boldsymbol{\epsilon}$, accounts for the discrepancy between the predicted and the actual values of \mathbf{Y} . The $\boldsymbol{\beta}$ is a $(l + 1) \times 1$ vector of coefficients that one generally wishes to estimate. In the gas identification problem the columns of the design matrix will be made up of known gas absorption spectra that have been appropriately rebinned.

The ordinary least squares (OLS) estimates for $\boldsymbol{\beta}$ are given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}. \quad (9)$$

These fitted regression coefficients are then used as the concentration path length estimates for a plume pixel.

The fundamental challenge of this application area is to simultaneously cope with the physical issues of various materials being present under the plume and the unknown plume temperature along with the mathematical issues of selecting the correct regression model for a given pixel. Succeeding in this model selection task is equivalent to correctly identifying the gas. A variety of different approaches for these combined tasks can be found in the literature.

Using fully-synthetic imagery Pogorzala, et al.⁸ developed the use of non-plume pixels to develop a set of endmembers which could be used to form a linear basis for the materials in the scene (excluding the gas). Then using stepwise regression via an F-test he was able to simultaneously estimate the mixture of endmembers and the mixture of gases that best fit any given pixel of data, over an exhaustive range of temperature contrasts between the plume and the background.

Gallagher et al⁹ employ a variety of OLS and weighted least squares (WLS) methods to determine the integrated CPL C_l for plumes embedded in synthetic LWIR imagery. Generally they relied on the ability to use the difference between off-plume and on-plume pixels to avoid having to know the surface temperature T_s or surface emissivity ϵ_s , and relied on atmospheric modeling code for estimating the remaining terms in the sensor-reaching radiance equation (Equation 5).¹⁰ The surface-leaving radiance terms were modeled via a principal components technique based on the off-plume pixels.

Another regression-based approach used by Chilton¹¹ is to use Bayesian Model Averaging (BMA) to mitigate the uncertainty in which model is the correct one. The premise of BMA is that if one were to average the fitted regression coefficients over the set of all possible models, the averaged value can be shown to be optimal in terms of minimizing prediction error for future observations.

For many problems the full set of possible models is huge, considering that there are l gases available for regressing against, there are 2^l possible models to fit, which is impractical for even a modest size gas library. There are a number of fast and computationally cheaper approximations used to execute BMA, most notably through the use of the Bayesian Information Criteria (BIC).

One interesting recent Bayesian approach by Heasler, et al.¹² focused on using non-linear regression implemented via the Metropolis-Hastings algorithm. These authors treated the background-plume temperature

contrast and the surface emissivity as nuisance parameters, and utilized a “fully Bayesian” approach, which relies on creating prior distributions for scene emissivities by modeling a database of material emissivities.

Other classic approaches often include the use of Spectral Matched Filters (SMF), projection operators, or Principle Components Analysis (PCA).¹³ SMF requires that one knows the gases of interest and attempts to estimate for each pixel the probability that it contains any given gas or mixture of gases. PCA essentially relies on the gases predominating one of the projections of the data onto an eigen-plane, and a user creates some arbitrary threshold in the output scores in order to classify a set of pixels as detections.¹⁴

There has also been considerable work in application-specific scenarios¹⁵ testing the sensitivity of detection to a variety of atmospheric parameters. These studies are generally based on examining the sensitivity of SMF receiver operating curve (ROC) to degradation in atmospheric parameters.¹⁶ While these authors weren’t explicitly trying to do inference, their approaches are philosophically congruent with the Bayesian paradigm. In a Bayesian sense, one would place prior distributions on the atmospheric parameters and collect data, then using Bayes rule, use a posterior distribution to make inference on the parameters of interest. Thus, it is possible to think of their work (in a sense) as focused on better defining the priors for atmospheric parameters rather than conducting inference on a specific gas after observing data.

The methods currently applied to gas identification generally suffer from one or more of a number of overarching problems. First is their inability to make a rational statement about which mixture of gases is the correct one, when several very different mixtures of gases could be fit equally well to the data. Second, these tasks often fall into what are generally what is known as the NP-hard complexity class of problems and require exhaustive search to solve (particularly with SMFs). Further, they often don’t allow the user to make use of any prior knowledge regarding the presence or absence of a particular gas. Finally, with potential exceptions^{11, 12} these methods often do not provide a framework for risk control in any global sense. They don’t produce anything like a multiple simultaneous confidence interval around β at an experiment-wise $1 - \alpha$ confidence level.

The authors propose the following two step approach towards the gas identification problem, first a step to deal with the rank-deficiency in any given gas library and second is a Bayesian regression methodology to aid in the model selection process.

All least squares problems are hampered by the multicollinearity of the design matrix and the problem of gas absorption library’s present a unique challenge in this regard. Because every application/sensor has a unique gas library of interest and spectral resolution, it is important to thin out extremely highly correlated gases (through an approach such as the VIF/SVD method described previously) before proceeding with the regression step. At this point, we focus on a Bayesian regression and model selection method for determining which gases are present in an observed pixel.

Bayesian statistics are based on the conditional probability rule where θ is generic placeholder for some parameter of interest

$$posterior = Pr(\theta|data) = \frac{Pr(data|\theta)Pr(\theta)}{Pr(data)} \propto Pr(data|\theta)Pr(\theta) = Likelihood \times prior. \quad (10)$$

It is important to note that $Pr(\theta)$, the prior distribution of the parameter θ , represents one’s pre-data understanding of the parameter (note that distributions involving θ and x can be continuous, discrete, jointly both and/or vector-valued). After collecting data, one uses a probability model or likelihood function that the data is assumed to have been generated by, but is also a function of the unknown parameter θ , normalized by the marginal probability of the value of outcome $Pr(data)$ to create a “post data” estimate of the parameter, which is conditioned upon the data $f(\theta|x)$. The distribution of $Pr(\theta|data)$ is the posterior distribution of θ and is the basis for all probability statements about θ .¹⁷

To deal with the problem of model uncertainty, the authors propose the use of a modified version of Gibbs Variable Selection (GVS)¹⁸ for approaching the model selection aspect of this problem. GVS is based on the premise that instead of solving the usual regression model (Equation 8) one uses a modified version that contains an $l \times 1$ vector of indicator variables γ

$$\mathbf{Y} = \sum_{l=1}^q \gamma_l X_l \beta_l + \epsilon. \quad (11)$$

Then using Markov Chain Monte Carlo (MCMC) one can make Bayesian inference on the posterior distributions of $\Pr(\boldsymbol{\gamma}|data)$ and $\Pr(\boldsymbol{\beta}|data)$.¹⁹

In practice, Bayesian inference is often conducted by finding a Markov Chain (MC) whose state space is the set of $\boldsymbol{\theta}$ and then empirically the long-run behavior of that state space (after sufficient iteration) approximates the posterior.¹⁷ This MC iteration (or MCMC) is often accomplished through Gibbs sampling or the Metropolis algorithm.

GVS uses the model in Equation 11 and assumes the regression coefficients $\boldsymbol{\beta}$ are a vector of *iid* $MVN(0, \sigma^2 \mathbf{I})$ variables and the indicator variables are *iid* Bernoulli(p). The model space \mathcal{M} (set of all 2^q possible models) is then partitioned according to those variables that are in the r^{th} model under consideration M_r by $\boldsymbol{\gamma}$ and those that are not in model M_r by $\boldsymbol{\beta}_{/\boldsymbol{\gamma}}$. These then directly correspond to the appropriate sets of regression parameters $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ and $\boldsymbol{\beta}_{/\boldsymbol{\gamma}}$ that are in and not in the model (respectively). One can then write the prior distributions of the parameters as the joint probability of the $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$

$$Pr(\boldsymbol{\beta}, \boldsymbol{\gamma}) = Pr(\boldsymbol{\beta}_{\boldsymbol{\gamma}}|\boldsymbol{\gamma})Pr(\boldsymbol{\beta}_{/\boldsymbol{\gamma}}|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma})Pr(\boldsymbol{\gamma}). \quad (12)$$

The likelihood is given by

$$Pr(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}) = Pr(\mathbf{y}|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma}). \quad (13)$$

Thus the posterior distribution of the parameters is proportional to the product of the RHSs of Equations 12 and 13.

GVS relies on the indicator variables to partition the model space, such that $\Pr(\boldsymbol{\beta}_{\boldsymbol{\gamma}}|\boldsymbol{\gamma}, \mathbf{y})$ is the posterior of the model specified by the vector $\boldsymbol{\gamma}$, while $\Pr(\boldsymbol{\beta}_{/\boldsymbol{\gamma}}|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma}, \mathbf{y})$ is the conditional posterior distribution of the parameters *not* included in model $\boldsymbol{\gamma}$. In terms of model selection one can interpret $\Pr(\boldsymbol{\beta}_{\boldsymbol{\gamma}}|\boldsymbol{\gamma})$ as the prior distribution for a given model. The quantity $\Pr(\boldsymbol{\beta}_{/\boldsymbol{\gamma}}|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma})$ is commonly called a “pseudo-prior” because the parameter vector $\boldsymbol{\beta}_{/\boldsymbol{\gamma}}$ does not actually influence the posterior distributions of interest $\Pr(\boldsymbol{\beta}_{\boldsymbol{\gamma}}|\boldsymbol{\gamma}, \mathbf{y})$. It is important for the MCMC to converge in a reasonable number of iterations, that the pseudo-priors be specified so that they lie roughly in the vicinity of either the OLS estimate, or those derived from a pilot run of the MC. One advantage of GVS is that in a single MCMC run, one obtains the full posterior for a set of models whose complexity is not limited to an arbitrary number of gases.

In ordinary statistical practice, a human operator would monitor some small number of MCs and adaptively determine a reasonable number of iterations, the thinning interval, and diagnose any problems, for every single regression operation. For imaging applications, where a single image has millions of pixels, and users typically have the ability to collect many images in a short span of time, the feasibility of iterative bayesian methods is seriously limited by the processing burden involved.

The authors have implemented a modified version of GVS that takes advantage of the conjugate multivariate normal/inverse-gamma family of models to execute part of the sampling. Among the variety of MCMC sampling mechanisms, Gibbs sampling is often a simple, effective means of drawing samples from a posterior distribution, and all that is required is for the user to be able to identify the full conditional distributions of each parameter in the model (in this case $\Pr(\boldsymbol{\gamma}|\boldsymbol{\beta}, \sigma^2, data)$ and $\Pr(\boldsymbol{\beta}|\boldsymbol{\gamma}, \sigma^2, data)$). Gibbs sampling is particularly useful in the case where one can identify the conditional distributions of each parameter, but it is difficult to sample from the joint distribution of the prior and the likelihood.

A convenient (although less broadly applicable) means of conducting GVS is in the case where one limits inference to a narrow class of probability models, *i.e.* conjugate families of distributions for the regression coefficients, where the joint posterior distribution is known. Careful examination of the GVS process reveals that the sampling order (and thus the convergence for) the $\Pr(\boldsymbol{\beta}|data)$ occurs first, then the sampling from the $\Pr(\boldsymbol{\gamma}|data)$ distribution takes place, with the caveat that all $Pr(\boldsymbol{\beta}|data)$'s that aren't currently included in the model are drawn from the pseudo-priors, which are essentially samples drawn from the conjugate MVN/gamma family (even for models that don't exploit this conjugacy).

If one limits themselves to the MVN/gamma family of densities for this problem, one can sample all of the $\Pr(\beta|data)$'s at every iteration from the joint posterior, and because their posterior is known in closed form, this negates the need to wait for convergence of this part of the chain. To handle the indicator variables, the densities of $\Pr(\gamma|data)$ are still updated through Gibbs sampling, which must converge iteratively. This entire process can be achieved by:

For the conjugate MVN/Gamma sampling process proceeds, given that

$$\mathbf{Y}|\beta, \sigma^2 \sim MVN(\mathbf{X}\beta, \sigma^2\mathbf{I})$$

one draws samples from

$$Pr(\beta|\sigma^2, \mathbf{Y}) \sim MVN(\hat{\beta}, \mathbf{V}_\beta)$$

where $\hat{\beta}$ is obtained through the OLS process, via QR decomposition of \mathbf{X} and \mathbf{V}_β is obtained by

$$\mathbf{V}_\beta = \mathbf{R}^{-1}(\mathbf{R}^{-1})' = (\mathbf{X}'\mathbf{X})^{-1}.$$

Samples from the distribution of σ^2 are drawn by creating realizations from

$$s^2 = \frac{1}{\Lambda - L}(\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta})$$

and then

$$\sigma^2|Y \sim Inv - \chi^2(\Lambda - L, s^2).$$

Then at every step, the variable selection sub-routine is carried out via Gibbs sampling, as outlined in the GVS procedure. This simple modification of GVS, to sample directly from the conjugate posterior distributions for the regression coefficients (which does not require convergence), and then Gibbs sampling the indicator variables, allows for dramatically faster convergence and shorter MCs for each pixel.

Because the posterior distribution $\Pr(\gamma|data)$ is available in its empirical form, one can (in some cases) estimate the frequency with which each model is visited as the MC traverses the model space. For large models this is problematic as the model space \mathcal{M} has dimensions 2^L and in order to sufficiently characterize $\Pr(\mathcal{M}|Data)$ one would need MCs that are several orders of magnitude longer than 2^{LL} for every single pixel of interest, which can be impractical to compute.

There is a body of mathematical theory that assures that a correctly designed MC will (after infinite iteration) converge to the desired target distribution, the problem is that all practical MCMC runs are finite and even though there is assurance that the posterior distributions will "get there" (in terms of approximating the correct posteriors) there is no foolproof way to know if convergence has been reached for any finite-length MC run.¹⁷

One last detail that is important for the performance of GVS (or any regression technique with this type of gas library) is dealing with the significant disparities in peak magnitude (300 fold magnitude difference), that exist between the spectra in the gas library. One simple technique is to use standardized regression. This is accomplished by replacing \mathbf{X} with

$$\mathbf{X}^* = \frac{1}{\Lambda - 1} \frac{\mathbf{X} - \bar{\mathbf{X}}}{s_{\mathbf{X}}}$$

and \mathbf{Y} with

$$\mathbf{Y}^* = \frac{1}{\Lambda - 1} \frac{\mathbf{Y} - \bar{\mathbf{Y}}}{s_{\mathbf{Y}}}$$

in the regression process. Then the expectations of the posterior standardized regression coefficients $E[Pr(\beta^*|data)]$ are transformed back into their original units via

$$E[Pr(\beta|data)] = E[Pr(\beta^*|data)] \cdot \frac{s_{\mathbf{Y}}}{s_{\mathbf{X}}}$$

and the intercept is returned from

$$E[Pr(\beta_0|data)] = \bar{Y} - \bar{X}' E[Pr(\beta|data)].$$

The following section describes an experiment that was undertaken in order to demonstrate the efficacy of GVS for gas detection. Because OLSTER allows us to “take apart” the hypercube into its constituent radiometric terms, insert any arbitrary gas plume, and re-assemble the cube; it provides a natural framework for testing this type of algorithm.

4. RESULTS

A common framework for evaluating the performance of a target detection method is an “receiver operating characteristic” (ROC) curve. The premise of which is to plot the cumulative probability of detection against the cumulative probability of false alarm for an algorithm, which produces a curve that shows the tradeoff between detecting as many events of interest relative to the number of false alarms that a method would produce. The goal is to find algorithms that maximize the number of detections for a given false-alarm rate, which graphically equates to on average higher curves. A simple metric that can be used to compare two ROC curves is simply the integrated area under ROC curve (denoted as AUROC), which provides some gross estimation of performance and a general means for comparing several competing algorithms.

The performance evaluation of GVS applied to a gas ID problem was conducted using a series of (identical) tiled AHI hyperspectral images, which all had varying amounts of a randomly selected gaseous mixture inserted. The gases were inserted identically in each subtitle, at the same concentration and temperature, with each tile differing only in the concentration of gas used (the range tested included the concentrations: 0, 100, 1000 and 10000 ppm-m). Such a wide range of concentrations was needed, due to the extreme range of values for peak absorption in the gas library. A single band of the composite tile is shown in Figure 1.(for color image see electronic version)

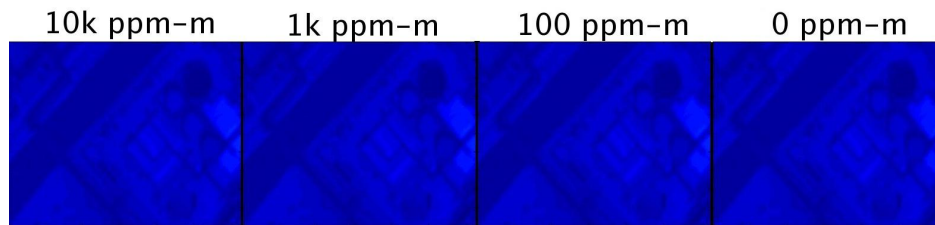


Figure 1. Tiled array of images, with varying implanted gas concentrations

The gas mixtures implanted in the data were at 330K while the average surface temperature for the scene was 309K. GVS was run with a $Pr(\gamma = .5)$ meaning that *a priori* each gas has a 50% chance of being included or excluded in the model. GVS was run for 4000 iterations per pixel, with a thinning interval of 40 and the first 500 iterations were discarded. In order to reduce run time, the image was spatially subsampled, with every 20th pixel included, in order to include a wide variety of material types, but limit the processing time to only 400 pixels per concentration level. The modified GVS procedure, in un-optimized Matlab code, takes approximately 11 seconds of processing per pixel on a dual-core 2.8 GHz Unix machine.

Once GVS was run on every pixel of the tiled images, the posterior probability of the i^{th} gas being present $Pr(\gamma_i|data)$ was gathered from each of the tiles. The expectations of the posterior densities were used to generate the point estimates of probability of inclusion for the ROC curve generation. The $Pr(\gamma_i|data)$ for each pixel in the 0 ppm-m tile was used as the “target absent” distribution and compared against each of the “target present” cases, at each concentration level. From these distributions, the AUROC was computed for each gas and is displayed in the right hand panel of Table 1.

As a means of comparison, the same tiles were processed using stepwise regression, as per the MATLAB implementation. The closest comparison (in a frequentist framework) to a posterior AUROC is obtained by

Table 1. GVS AUROCS per gas (true constituents are in boldface)

ppm-m	1 Gas			2 Gases			4 Gases		
	100	1,000	10,000	100	1,000	10,000	100	1,000	10,000
acrol	0.63	0.94	0.96	0.61	0.01	0.08	0.44	0.08	0
c6h6	0.52	0.69	0.31	0.26	0	0	0.61	0.46	0.97
dclp12	0.51	0.23	0.19	0.29	0.01	0.86	0.23	0.09	0.09
dclp13	0.46	0.32	0.08	0.62	0.04	0.86	0.68	0.14	0.84
edb	0.42	0.31	0.07	0.22	0.01	0.19	0.90	0.25	1.00
f114	0.48	0.41	0.20	0.44	0.03	0.58	0.38	1.00	1.00
f12	0.43	0.30	0.20	0.99	1.00	1.00	0.42	0.10	0.69
f218	0.46	0.23	0.19	0.97	1.00	1.00	0.54	0	0
nh3	0.45	0.09	0.18	0.28	0.01	0.73	0.35	0	0
phg	0.46	0.33	0.03	0.52	0.01	0.87	0.24	0.02	1.00
sf6	0.45	0.20	0.19	0.20	0	0.02	0.35	0	0.36
so2	0.48	0.36	0.11	0.31	0.01	0.84	0.28	0.07	0.92
tce	0.46	0.44	0.05	0.45	0.01	0.85	0.57	0.03	0.54

Table 2. Stepwise regression model frequencies (true constituents are in boldface)

ppm-m	1 Gas			2 Gases			4 Gases		
	100	1,000	10,000	100	1,000	10,000	100	1,000	10,000
acrol	0.98	1.00	1.00	0.99	0.16	0.07	0.95	0.55	0
c6h6	0.98	1.00	1.00	0.89	0.42	0	0.96	1.00	0.99
dclp12	0.02	0	0	0	0.01	0	0	0.42	0.62
dclp13	0	0	0	0	0	0	0	0	0.66
edb	0.01	0	0	0.01	0	0	0	0	0.02
f114	0.34	0.47	0.57	0.47	0.16	0.43	0.34	1.00	1.00
f12	0	0	0	0.61	1.00	1.00	0	0.08	0.87
f218	0	0	0	0.60	1.00	1.00	0	0	0
nh3	0	0	0	0	0	0.01	0	0	0.04
phg	0	0	0.01	0.02	0	0	0	0	0
sf6	0	0	0	0	0	0	0	0	0
so2	0.22	0.11	0	0.06	0	0.43	0.08	0	0.94
tce	0	0	0	0	0	0	0	0	0

looking at the observed frequency with which a gas is included in the final fitted model, at each concentration level. The authors readily admit that this is not a statistically rigorous comparison owing to the many important differences between frequentist/Bayesian inference (among others), however it provides the end-user with a comparison that can at least be used to inform the analytic process.

As shown in Tables 1 and 2 the performance of the modified GVS algorithm is roughly on par with stepwise regression when applied to this library, with these particular choices of implanted species. For both Tables 1 and 2, one would ideally like to see the boldface entries have a value of unity and zeros elsewhere. These exemplars were chosen to include a broad selection of gases from the library. Since the focus of this work is model selection and in the interest of brevity the fitted/posterior regression coefficients are not shown, although there generally is close agreement between stepwise regression and GVS.

One interesting artifact of this data generation method is that the variance observed in the fitted/posterior coefficients is not the result of measurement error (i.e. it is not the ϵ in Equations 8 or 11) because each pixel vector in the semi-synthetic data tile was assembled without error. Thus any variance in the fitted/posterior coefficients is a result of the non-unique mapping in the OLS process, which is a manifestation of the rank

deficiency of the gas library, and exists in both the Bayesian and frequentist methods. For this reason, showing a comparison of multiple simultaneous confidence intervals around the fitted β 's against credible sets around the posterior $Pr(\beta|data)$'s, isn't relevant for this dataset.

In the single gas example, GVS avoids the false alarm associated with c6h6 that stepwise regression encountered. In the two gas example, at low concentrations GVS had a much higher AUROCs than stepwise for the correct model and avoided two large false alarms (acrol and c6h6). At higher concentrations in the two gas example, GVS correctly identifies the correct gases in all pixels (AUROC for f12 and f218 = 1.00) but a number of other gases are potential false alarms with AUROCs in the .7 - .9 range. In the four gas example, stepwise had fewer false alarms than GVS at both the highest and lowest concentration levels.

Generally speaking, stepwise regression had much lower average model frequencies than did GVS, for gases that were not in the tiles. Throughout this experiment, GVS was used with "non-informative priors", which implies that the prior distributions were chosen so as to let the data play the greatest possible role in the regression/model selection process.

5. CONCLUSIONS

From an analyst's perspective, the chief advantage of using a Bayesian technique such as GVS is that there is room in the technique for applying scene-specific knowledge in the analysis process. For example, an analyst having some practical knowledge of what gases are more likely to be present in a scene might specify some prior structure for $Pr(\gamma)$ that places higher probability of inclusion on certain gases. Additionally, there is additional ground to be gained in tuning the prior distributions for $Pr(\gamma)$ for specific scenarios and gas libraries.

Furthermore, the Bayesian paradigm creates a natural framework for conducting multi-pixel inference, whereby one uses the posteriors from an adjacent pixel to influence the prior distribution of its neighbors, taking advantage of the spatially contiguous nature of many plumes. Finally, there exists the potential for the aggregation of inference in a sequential manner, whereby successive frames of data use the posterior distributions of the indicators from previous frames as the priors for analysis at some later time step.

The authors have introduced a new method of gaseous effluent identification in passive LWIR imagery. This technique (GVS) allows for simultaneous estimation of the concentration path length of each gas, as well as the probability of inclusion of each gas. This technique was demonstrated on semi-synthetic data produced from real LWIR HSI and the results of GVS were compared to stepwise regression.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support for this project from Battelle Memorial Institute and the Pacific Northwest National Labs. The views expressed in this article are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

REFERENCES

1. Sharpe, Sams, Johnson, Chu, Rhoderick, and Guenther, "Creation of 0.01 $cm^{(-1)}$ resolution, quantitative, infrared spectral libraries of gas samples," in *Proceedings of SPIE*, **4577**, Proceedings of SPIE, 2001.
2. M. T. Stephen Walsh, Larry Chilton and C. Metoyer, "Effect of the temperature-emissivity contrast on the chemical signal for gas plume detection using thermal image data," *Sensors* **8**, pp. 6471-6483, 2008.
3. J. M. Brown, *Molecular Spectroscopy*, Oxford Science Publications, 1998.
4. J. Schott, *Remote Sensing: The Image Chain Approach*, Oxford University Press, Oxford, NY, 2007.
5. M. Boonmee, J. R. Schott, and D. W. Messinger, "Land surface temperature and emissivity retrieval from thermal infrared hyperspectral imagery," *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XII - Proceedings of SPIE* **6233**, 2006.
6. A. Berk, G. Anderson, L. Bernstein, P. Acharya, H. Dothe, M. Matthew, S. Adler-Golden, Chetwynd, S. Richtsmeier, B. Pukall, C. Allred, L. Jeong, and M. Hoke, "Modtran4 radiative transfer modeling for atmospheric correction," *SPIE*, 1999.

7. P. Lucey, T. Williams, M. Mignard, J. Julian, D. Kokbun, G. Allen, D. Hampton, W. Schaff, M. Schlagen, E. Winter, W. Kendall, A. Stocker, K. Horton, and A. Bowman, "Ahi: An airborne long wave infrared hyperspectral imager," in *SPIE Conference on Airborne Reconnaissance XXII*, 1998.
8. D. Pogorzala, D. Messinger, C. Salvaggio, and J. Schott, "Gas plume species identification in airborne lwir imagery using constrained stepwise regression analyses," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI*, **5806**, SPIE, SPIE, 2005.
9. N. Gallagher, B. Wise, and D. Sheen, "Estimation of trace vapor concentration-pathlength in plumes for remote sensing applications from hyperspectral images," *Analytica chimica ACTA* , 2003.
10. N. Gallagher, B. Wise, and D. Sheen, "Error analysis for estimation of trace vapor concentration pathlength in stack plumes," *Applied Spectroscopy* , 2003.
11. L. Chilton, "Model and method for gas identification." Personal communication, 2006.
12. P. Heasler, C. Posse, J. Hylden, and K. Anderson, "Nonlinear bayesian algorithms for gas plume detection and estimation from hyper-spectral thermal image data," *Sensors* , 2007.
13. D. Messinger, "Gaseous plume detection in hyperspectral images: a comparison of methods," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery X*, S. S. Shen and P. E. Lewis, eds., *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery X* , SPIE, 2004.
14. D. Messinger, C. Salvaggio, and N. Sinisgalli, "Detection of gaseous effluents from airborne lwir hyperspectral imagery usin physics-based signatures," *International Journal of High Speed Electronics and Systems* **20**(10), pp. 1–12, 2006.
15. D. Sheen, N. Gallagher, S. Sharpe, K. Anderson, and J. Schultz, "Impact of background and atmospheric variability on infrared hyperspectral chemical detection sensitivity," *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX* , 2003.
16. S. T. Kacendar, D. F. Gill, J. A. Lelii, J. Foreman, and C. B. Batrone, "Spectral effluent detection sensitivity study," in *SPIE Conference on Algorithms for Multispectral and Hyperspectral Imagery IV*, 1998.
17. A. Gelman, J. Carlin, H. Stern, and D. Rubin, *Bayesian Data Analysis*, Chapman and Hall, 2 ed., 2004.
18. P. Dellaportas, J. Forester, and I. Ntzoufras, "On bayesian model and variable selection using mcmc," *Statistics and Computing* **12**, 2002.
19. I. Ntzoufras, "Gibbs variable selection using bugs," *Journal of Statistical Software*, **7**(7), 2002.