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

by

Shawn Higbee

B.S. Iowa State University, 2002

M.S. University of Massachusetts Lowell, 2005

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Chester F. Carlson Center for Imaging Science
Rochester Institute of Technology

September 2, 2009

Signature of the Author _____

Accepted by _____
          Coordinator, Ph.D. Degree Program   Date

CHESTER F. CARLSON CENTER FOR IMAGING SCIENCE

ROCHESTER INSTITUTE OF TECHNOLOGY

ROCHESTER, NEW YORK

<u>CERTIFICATE OF APPROVAL</u>

Ph.D. DEGREE DISSERTATION

The Ph.D. Degree Dissertation of Shawn Higbee
has been examined and approved by the
dissertation committee as satisfactory for the
dissertation required for the
Ph.D. degree in Imaging Science

_____

Dr. David Messinger, dissertation Advisor

_____

Dr. Yolande Tra

_____

Dr. Carl Salvaggio

_____

Dr. Joseph Voelkel

_____

Date

DISSERTATION RELEASE PERMISSION

ROCHESTER INSTITUTE OF TECHNOLOGY

CHESTER F. CARLSON CENTER FOR IMAGING SCIENCE

Title of Dissertation:

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

I, Shawn Higbee, hereby grant permission to Wallace Memorial Library of R.I.T. to reproduce my thesis in whole or in part. Any reproduction will not be for commercial use or profit.

Signature _____

Date

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

by

Shawn Higbee

Submitted to the
Chester F. Carlson Center for Imaging Science
in partial fulfillment of the requirements
for the Doctor of Philosophy Degree
at the Rochester Institute of Technology

Abstract

Typically a regression approach is applied in order to identify the gaseous 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 modified version of GVS with fast convergence properties that is tailored to unsupervised use in hyperspectral image analysis will be presented. Additionally a series of automated diagnostic measures have been developed to monitor convergence of the MCMC with minimal operator intervention. Finally, the applicability of aggregating inference from adjacent pixels will be discussed.

This method is compared against stepwise regression for model selection and results from LWIR data from the Airborne Hyperspectral Imager (AHI) are presented. Finally, the applicability of this method to operational scenarios and various sensors will be discussed.

Acknowledgements

The most important discovery of the last three years of my life is not contained within these pages. That discovery is ipseity.
Thanks to all of those who have played a role, large or small, in its revelation.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the portions of the electromagnetic (EM) spectrum that ranges in wavelength roughly from 8 μm to 12 μm 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 resolutions [1] 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 (a, b) and a spectral dimension c. Thus every observation in the image cube has a unique index $\boldsymbol{L}_{a,b,c}$ such that $a \in [1, 2, \cdots, A]$, $b \in [1, 2, \cdots, B]$, and $c \in [1, 2, \cdots, C]$ and any spatial location in the image cube has a corresponding $(c \times 1)$ column vector of measurements $\boldsymbol{L}_{a,b,\cdot}$ for that pixel. Note that all vectors are assumed to to be column vectors unless otherwise specified and all variables that are vectors and matrices are in boldface. For the matrix A the symbol A' is the matrix transpose and A^{-1} is the matrix inverse such that $A^{-1}A$ is the identity matrix I.

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, because it is the combination of the physical properties of the background and gaseous plume (specifically the temperature difference between the two) that contribute to the observed spectral signature from the plume pixel.

Typically 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 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 quantitatively measure the presence/amount of gas in a plume.

Research objectives

This document describes a body of research that leverages advances in physical modeling of remotely sensed images as well as a statistical methodology that may be applied to the problem of choosing an appropriate model (and thus the correct mixture) of gases in the plume. The specific objectives that were used to guide this work include:

- Demonstrate successful regression-based plume identification and quantification through some sort of Bayesian model selection.

- Make use of real data cubes as a starting point for research, both to improve the credibility of work and include realistic estimates of the algorithm performance.

- Leverage physical models where possible (OLSTER, physical plume model, etc) to improve the identification and inference process.

- Demonstrate the potential for multi-pixel aggregation of inference of adjacent pixels in an image

This remainder of this thesis is organized in the following manner. First a background of the physical problem will be presented, including the data used, its associated processing, and current methods used for this problem in the literature. Next follows a

discussion of the methods developed for gas plume identification, including a Bayesian regression technique that can be applied on single pixels or groups of pixels. This is followed by a presentation of the results generated by this method in a series of experiments carried out on data derived from real hyperspectral imagery. Finally the thesis concludes with a discussion of the conclusions reached, ideas for future work and an appendix of supporting work.

Chapter 2

Background

This section focuses around introducing the reader to the problem of gaseous effluent identification and providing the necessary foundation of physics that enables the ensuing algorithmic problem to be understood. It begins with a discussion of the phenomenology relevant to the LWIR, the sensor used to collect the raw data, the process of temperature-emissivity separation used to generate the final data sets used throughout this thesis, the gas library that provides both the spectra used for artificial plume implantation as well as a library for matching, a discussion of existing methods in the literature, and finally the concept of operations in which these methods could be applied is discussed.

Phenomenology of gas identification

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 [2]. 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. Because this excitation happens generally at only a few discrete frequencies for a particular molecule, there are a finite number of stationary energy states in which a molecule can exist in equilibrium. The transitions to move a molecule from one energy (E_1) state to another (E_2) is governed by

$$h\nu = E_2 - E_1 \tag{2.1}$$

where h is Planck's constant $(6.6261 \times 10^{-34} Js)$ and ν is the frequency of the energy being absorbed or emitted. The fact that Equation 2.1 holds for *any* change in energy (both energy absorption and emission are included) and the existence of only discrete stationary energy states, means that the frequencies where a molecule will absorb energy are the same as those where it emits energy. For the sake of convenience these frequencies can be converted to units of wavelength via

$$\lambda = \frac{c}{\nu} \tag{2.2}$$

where c is the speed of light $(2.998 \times 10^8 m/s)$. Because the change in energy states of all molecules is governed by Equation 2.1, one can use a spectrometer to precisely measure the absorption spectra of known gases [1]. As mentioned previously it is the *interaction* of EM radiation with a molecule that determines whether a molecule will absorb some of the radiation that is incident upon it or emit energy. For most practical scenarios in remote sensing the observed radiance 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 by a radiometric object known as a "black body."

Across the EM spectrum, frequent reference is made to the concept of a "black body" radiator (BB), which is a somewhat theoretical construct describing the distribution of EM as a function of wavelength that an ideal radiator would radiate. It is said to be ideal in the sense that it absorbs all of the energy incident upon it. The empirical equation often used to describe a BB is given by Equation 6.6.

$$\boldsymbol{B}(\lambda) = \frac{2\pi hc^2}{\lambda^5 (e^{\frac{hc}{\lambda kT}} - 1)} \cdot 10^{-6} \quad [Wm^{-2}sr^{-1}\mu m] \tag{2.3}$$

In Equation 6.6, k is Boltzman's constant $(1.381 \times 10^{-23} \; JK^{-1})$, 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 (L) are watts per meter squared per steradian per micron. Note that temperature T is the only variable in the equation, and once it is specified the user can retrieve via simple calculation, the radiance such an object will radiate for any wavelength. Tangible examples of a BB are a soldering iron, a heating coil, etc.

although any object above absolute zero degrees Kelvin radiates at least some energy in
the LWIR. How much EM energy any BB will radiate is determined by the material's
emissivity $\boldsymbol{\xi}$.

For any object at temperature T, with known radiance (given by Equation 6.6)
only some of its energy will actually propagate away from an object. The "emissivity"
$\boldsymbol{\xi}$ is a $c \times 1$ vector that specifies (as a percent) how much energy, per wavelength, will
propagate away from a material. A true theoretical black body has unit emissivity for
all wavelengths. Emissivity ξ is generally material specific (and less than one) and thus
it can be useful for determining the materials present in remotely sensed data. Thus
the radiance that an object propagates is most appropriately written as the element-
wise product of $\boldsymbol{\xi} \cdot \boldsymbol{B}(T)$ where it is influenced simultaneously by its emissivity and
its temperature. The net result of which is another $C \times 1$ vector that describes the
spectral distribution of self-radiance that actually propagates away from the object.
Throughout this document element-wise products will be denoted with a "dot" $(A \cdot B)$
while matrix products will be denoted without any symbol (AB).

Two other noteworthy sources of radiance that contribute to the sensor-reaching
radiance are (also $C \times 1$ vectors) "upwelled radiance" $(\boldsymbol{L}_u)$ and "downwelled radiance"
$(\boldsymbol{L}_d)$. Upwelled radiance $(\boldsymbol{L}_u)$ is the self-radiance of the air-column through which the
image is taken, while downwelled radiance $(\boldsymbol{L}_d)$ 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 an object, either self emitted or reflected downwelling, then must
pass through the atmosphere $(\boldsymbol{\tau})$ (which is not perfectly transparent in the LWIR and
amounts to another element-wise multiplication by a $C \times 1$ vector) on its way 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 $(\boldsymbol{\tau}_{plume})$, and act as an emitter of
radiance itself $(\boldsymbol{B}(T_{plume}))$ both of which are also $C \times 1$ 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) [3] for a given (a, b) location (note all terms are $c \times 1$ vectors, and this equation man-
ifests itself differently for every pixel in a scene). For ease of understanding first consider
a non-plume pixel, whose radiance is given by the element-wise addition/multiplication
of

$$\boldsymbol{L}_{(a,b,\cdot)} = [\boldsymbol{\xi}_{surface,(a,b,\cdot)} \cdot \boldsymbol{B}_{surface,(a,b,\cdot)} + (\boldsymbol{1} - \boldsymbol{\xi}_{surface,(a,b,\cdot)}) \cdot \boldsymbol{L}_{downwell}] \cdot \boldsymbol{\tau_{atm}} + \boldsymbol{L}_{upwell}.$$

$$(2.4)$$

In words this could be stated as "the vector of radiance, as a function of wavelength, that reaches the sensor from a single a, b location = [radiance from the heat of the ground + reflected radiance from the sky above] * atmospheric absorption on the path to the sensor + upwelling radiance of the air column one is imaging through." Note that the $\boldsymbol{L}_{upwell}$, $\boldsymbol{L}_{downwell}$ and $\boldsymbol{\tau}_{atm}$ terms are treated as constants over the entire scene, and thus do not have spatial subscripts.

Now consider the more complicated case where a plume is present, and for brevity we will assume all quantities are $C{\times}1$ vectors and are implicitly functions of wavelength, and consider only a single pixel (thus dropping the spatial indices). By simply adding another two terms into Equation 2.4 to create Equation 6.8, 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 gas $\boldsymbol{\xi}_{plume}\boldsymbol{B}(\lambda, T_{plume})$. Note, from here forward all subscripts will be abbreviated and one can assume that the quantities are all $c \times 1$ vectors unless otherwise stated.

$$\boldsymbol{L} = [(\boldsymbol{\xi}_s \cdot \boldsymbol{B}(T_s) + (1 - \boldsymbol{\xi}_s) \cdot \boldsymbol{L}_d) \cdot \boldsymbol{\tau}_p + \boldsymbol{\xi}_p \boldsymbol{B}(T_p)] \boldsymbol{\tau}_{atm} + \boldsymbol{L}_u. \qquad (2.5)$$

Assuming the plume is uniform and optically thin one can approximate $\boldsymbol{\tau}_p(\lambda) \approx 1 - \boldsymbol{\xi}_p \approx 1 - \sum \boldsymbol{a}_l C_l$ by making use of the Taylor's series expansion of $\boldsymbol{\xi}_p$. Each $\boldsymbol{a}_n$ is a vector of the nth ($n \in [1, 2, \ldots, N]$ constituent gas absorption spectra (also $C \times 1$ vectors, which are known from standard laboratory measurements) and C_n 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 for which gas(es) are present, and the integral for the path distance)

$$C_n = \int \sum_{n=1}^{N} c_n \quad dpath. \qquad (2.6)$$

One assumes that the plume is locally "well mixed", and although the variation in thickness (relative to the distance at which the plume is observed) is small, one still

does not know the absolute thickness of the gas layer and can only conduct inference on the CPL of the gas(es) present. In other words, $C_n \propto$ concentration of the set of n gases $(c_1, c_2, \ldots, c_N)$ and in order to do exact inference on c_n one would need to know the plume's optical path thickness (*dpath* in Equation 3.11). Recall that the gases act like spectral "filters", which means that they modify the radiance that propagates through them. The key difference that will be shown below is that when a gas plume is "in emission" the amount of radiance that propagates from it can be *greater* in certain wavelengths than the ground leaving radiance alone.

With the above substitutions and expanding the terms, we generate Equation 3.12 from Equation 6.8. Equation 3.12 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" $([\boldsymbol{B}(T_p) - \boldsymbol{\xi}_s \cdot \boldsymbol{B}(T_s) - (1 - \boldsymbol{\xi}_s) \cdot \boldsymbol{L}_d])$. 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 ("add spikes to the spectra") and if it is net negative the plume will be in absorption ("cuts notches out of the spectra"). Exactly where those spikes/notches are, and their magnitude is governed by the weighted sum of their absorption spectra ($\boldsymbol{a}_n$ multiplied by their concentration path lengths C_n). This phenomena is graphically shown in Figure 2.1.

$$\boldsymbol{L} = [\boldsymbol{\xi}_s \cdot \boldsymbol{B}(T_s) + (1 - \boldsymbol{\xi}_s) \cdot \boldsymbol{L}_d + \sum \boldsymbol{a}_l C_l \cdot [\boldsymbol{B}(T_p) - \boldsymbol{\xi}_s \cdot \boldsymbol{B}(T_s) - (1 - \boldsymbol{\xi}_s) \cdot \boldsymbol{L}_d]] \cdot \boldsymbol{\tau}_{atm} + \boldsymbol{L}_u \quad (2.7)$$

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, that one can observe the optical manifestation of the gas plume. For a single on-plume pixel, this scenario is depicted in Figure 2.

The physical dispersion of the plume (relative to its release point) is also going to affect its optical properties. First, the plume will become less concentrated as it disperses. This will be dramatically affected by the atmosphere (wind speed in particular) and can be modeled by some advanced plume models. In general, one would expect an exponential decay to be a reasonable first order approximation to concentration as
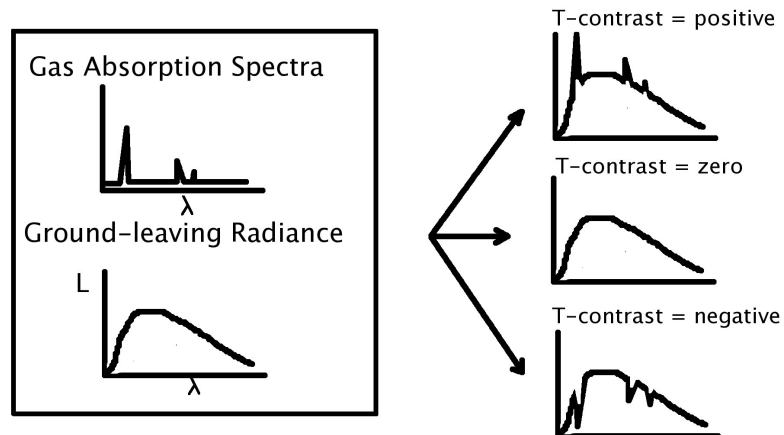
Figure 2.1: Notional impact of temperature contrast leading to emission or absorption of a gas plume
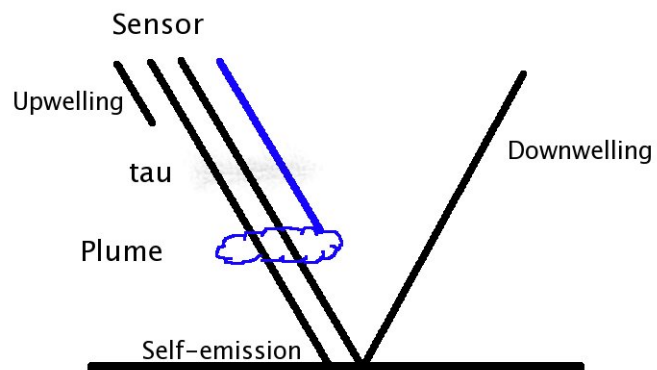
Figure 2.2: Relevant sources of sensor reaching photons in the LWIR

a function of distance from the release point simply by virtue of volumetric expansion and dispersion of the effluent (see Figure 2.3). Across the direction of travel, plumes are often modeled as bivariate normal, the covariance matrix of which is determined by the physical properties of the atmosphere.

Figure 2.3: Notional relationship between plume concentration and distance from the release point

The second physical relationship that one would expect to occur is the process of the plume's temperature approaching ambient. If the gas was hotter than ambient when it was released, this implies cooling will occur, and if the release was colder than ambient, it implies the plume will warm towards the ambient air temperature. This ambient temperature may or may not be the same as the surface temperature under the plume and is probably best estimated from meteorologic sources. Recall that it is the relative difference in temperatures (between the plume and the ground) that makes the plume absorb or emit radiance in the LWIR. The expected warming/cooling relationship is characterized in Figure 2.4.

Temperature Emissivity Separation

One important point that was alluded to, but not directly addressed in the preceding discussion is the fact that in the simple case of an off-plume pixel what actually *propagates* towards the sensor is $\boldsymbol{\xi}_s \cdot \boldsymbol{B}(T_s)$ which contains $C + 1$ unknowns (c emissivities
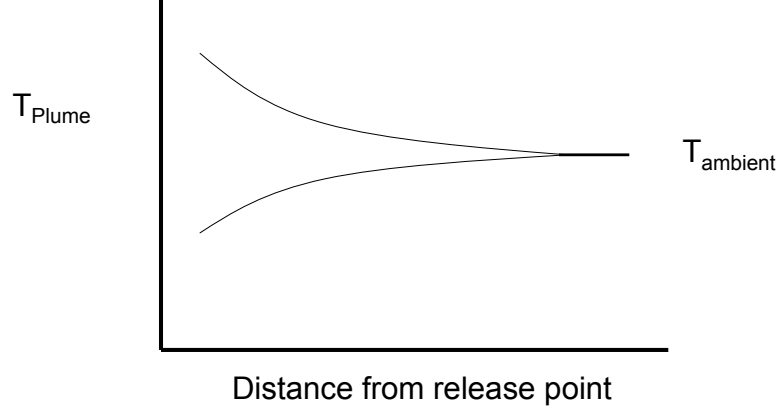
Figure 2.4: Notional relationship between plume temperature and distance from the release point

and 1 temperature). What is observed is a vector of $\boldsymbol{L}_{C \times 1}$ radiance values, thus one will always have more unknowns than measurements, and as such an "ill-posed" or "under-determined" system of equations. The general class of problems that aim to somehow use the vector of observed radiance values, often along with other physical constraints, to retrieve *both* the surface temperature (T_s) and surface emissivity (ϵ_s) is called temperature/emissivity separation or temperature/emissivity retrieval.

A recently developed technique known as Optimized Land Surface Temperature and Emissivity Retrieval (OLSTER) [4] that uses an iterative gradient decent search, guided by a variety of physical constraints to estimate the scene-wide $\boldsymbol{\tau}_{atm}$, $\boldsymbol{L}_d$, and $\boldsymbol{L}_u$ in addition to $\boldsymbol{\xi}_s$ and T_s for every pixel, was used to perform temperature/emmisivity separation for this project. The algorithm makes simultaneous use of several well-known atmospheric models including MODTRAN and ISAC along with empirical constraints on certain parameters, for example $\boldsymbol{\xi}_s$ 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 $\boldsymbol{\xi}_s \cdot \boldsymbol{B}(T_s)$ is only the $\boldsymbol{\xi}_s$ 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 is

effectively enables the process of "taking apart the hypercube." Recall that the output of the sensor was the left-hand side of Equation 2.4, and OLSTER gives us the ability to individually obtain 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 author terms "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 6.8. One can then explore interesting scenarios such as applying various spatial plume models, doing sequential inference, and examining the background-dependency of plume identification.

Examples (from an AHI image) of the data products that OLSTER outputs include image-wide vectors for $\boldsymbol{\tau}_{atm}, \boldsymbol{L}_u, \boldsymbol{L}_d$ and pixel by pixel values for T_s and $\boldsymbol{\xi}_s$. Figure 2.5 shows the scene-wide atmospheric parameters, Figure 2.6 shows the surface temperature T_s map, and Figure 2.7 shows the covariance and correlation matrices of the surface emissivity cube $\boldsymbol{\xi}_s$.

Sensor and Data

The sensor used to generate the data for this project is the Airborne Hyperspectral Imager (AHI) that was built at the University of Hawaii for the purpose of detecting buried landmines. The sensor collects a nominal 256 bands in the 7.5 μm to 11.5 μm wavelength region [5]. 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 8.094 μm to 11.53 μm). As evident in Figure 2.8 this loss of spectral resolution could have a significant impact on ones ability identify the fine spectra features of some gas species. In general, the difficultly in identifying a gas with a practical spectrometer (one that has finite spectral resolution) is simply dependent on the width of the spectral features of the gas(es) in question.

The hyperspectral image (HSI) data used for this research comes from an EPA study from 2003 where AHI imagery was collected (with limited ground truth) of a chemical refinery in Texas. The imagery includes day and night scenes, scenes with plumes present, and mixture of agricultural and industrial land use (the temperature
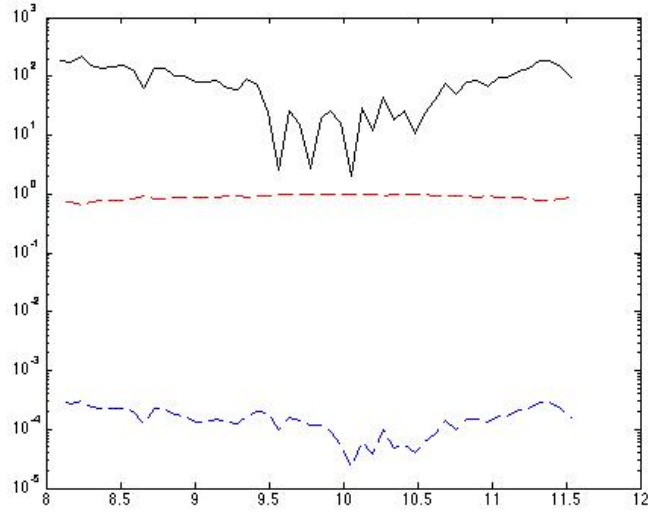
Figure 2.5: OLSTER outputs versus wavelength (μm): $\boldsymbol{L}_d$(dashed blue), $\boldsymbol{\tau}_{atm}$ (dashed red), and $\boldsymbol{L}_u$ (solid black) with AHI spectral resolution, note log scale
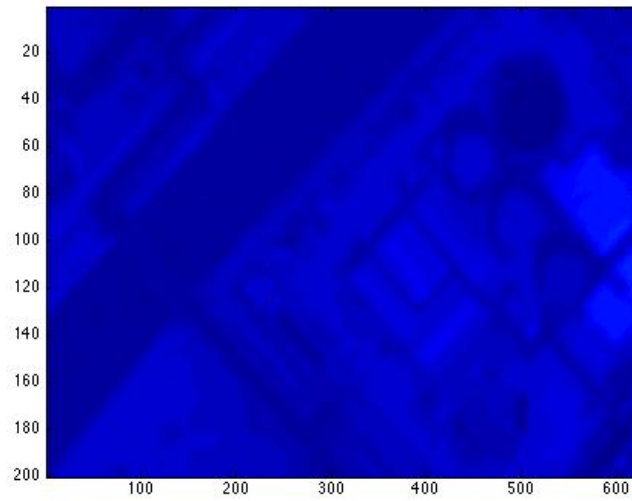
Figure 2.6: OLSTER surface temperature T_s output
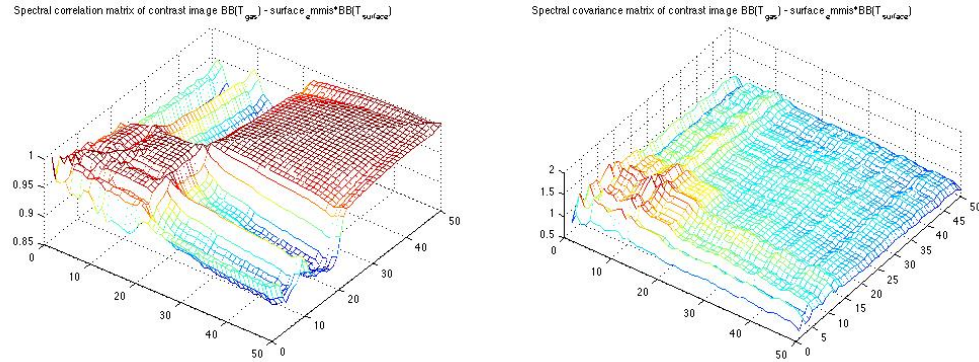
Figure 2.7: OLSTER surface emissivity $\boldsymbol{\xi}_s$ correlation matrix (left) and covariance matrix (right)

map of one daytime scene is shown in Figure 2.6).

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. This library of spectra is available for gases at 5°, 25°, and 50° C. Although there are subtle differences between the spectra of a given gas at different temperatures, the difference is trivial when the gas spectra are rebinned to match the down-sampled AHI bin centers (50 bands). This low sensitivity to gas temperature is one of the few benefits of working with down-sampled data. The author used the $25°C$ library of spectra as the difference between the down sampled libraries was trivial.

Any practical optical instrument will act like a low-pass filter on an incoming signal and will often display non-uniformities spatially in the field of view as well as spectrally. While completely modeling such effects is often necessary for optical designers, it is not always practical for end-users of the data. As a first-order approximation for the output of an optical system, one can assume that the "point spread function" of the sensor is a three-dimensional Dirac delta function (spatially as well as spectrally), and the sensor is a linear and shift-invariant system. This implies that one can ignore spatial, and spectral non-uniformities of the instrument and not use a (more realistic) convolution method of rebinning high-resolution spectra that would take into account how a practical sensor would "blur" a point-source input.

This means that a RECT(c) function was used for rebinning the high-resolution spectra to approximate what the spectra would look like if observed by the AHI sensor (which was also rebinned to 50 bands). Ultimately this amounts to nothing more than local averaging around the desired bin centers for the 50 output bands. An example with gas "acrol05" is shown in Figure 2.8. Note how the fine spectral features at 8.6 μm, 10.1 μm, and 10.4 μm are dramatically reduced by the spectral averaging proceedure. This blurring also impacts the difference between spectra, which determines one's ability to differentiate gases. Despite this loss in resolution, some gases are still fairly distinct from one another at this low spectral resolution (see Figure 2.9). One additional 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.
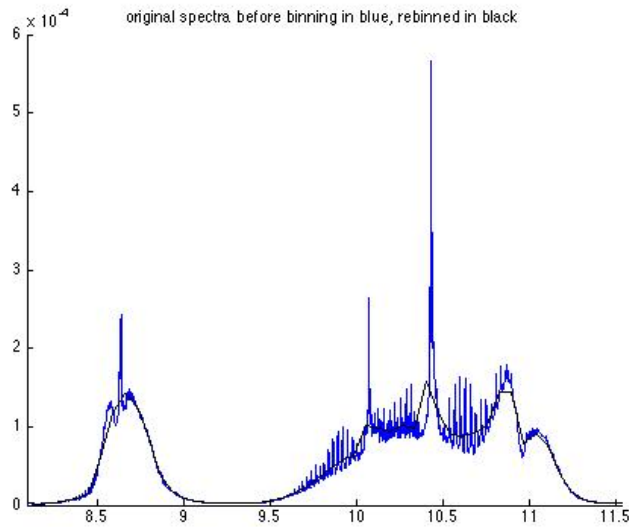
Figure 2.8: Input (blue) and rebinned (black) library spectra for acrol05 (AHI 50 bands)

Alternate Methodology in the Literature

Several methods exist for detecting and identifying a gas plume in an LWIR image. They generally fall into two categories, either some sort of regression method or a
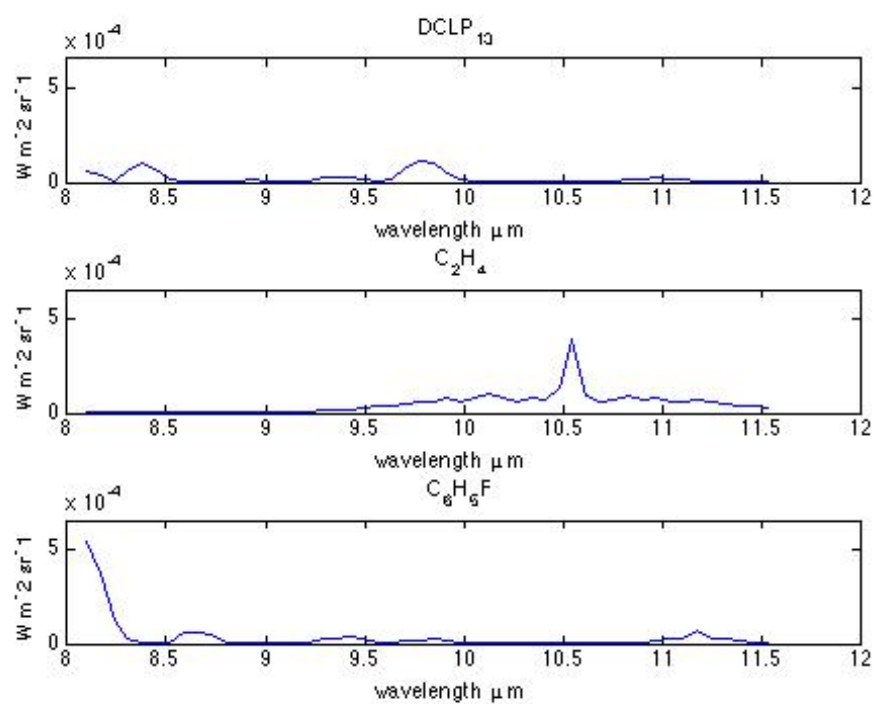
Figure 2.9: Three absorption spectra from the author's gas library at the downsampled AHI spectral resolution, potted with common vertical scales

matched filter.

Pogorzala [6] proposed 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. This was carried out on fully-synthetic data, which allows easy access to ground truth for the plume constituents, plume spatial extent, as well as the ability to fully control the sources present in the data.

Stepwise regression is a method of model selection that belongs to the class of frequentist methods (as opposed to Bayesian methods). Recall that in the canonical regression model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

the vector of observed $\boldsymbol{Y}$ values are a linear combination of a deterministic design matrix $\boldsymbol{X}$ (dimensions $C \times N$) and deterministic vector of $\boldsymbol{\beta}$s (the fact that these coefficients are fixed unknown constants is the essential point) and a random process $\boldsymbol{\epsilon}$. Then when using ordinary least squares to estimate $\hat{\boldsymbol{\beta}}$ using the normal equations

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

the estimate $\hat{\boldsymbol{\beta}}$ is a linear combination of a deterministic $\boldsymbol{X}$ and a random $\boldsymbol{Y}$, thus the estimates for $\hat{\boldsymbol{\beta}}$s are themselves random quantities.

Because the fitted regression coefficients are random one generally tests each of them to see if they are statistically significant from zero (often done with a t-test), while the entire regression model is tested via an F test

$$F^*_{N,C-N} = \frac{MSR}{MSE}.$$

Furthermore one can measure the ability of a predictor X_k to improve the fit of the model to the data relative to a model X_{k_0} that does not contain X_k (as judged by the L2 norm $||\hat{y_k} - \hat{y_{k_0}}||_2$) by computing the ratio of the "extra" variability (ME_{extra}) described by adding the kth term relative to the estimate of random error (MSE)

$$F_{1,C-N} = \frac{MS_{extra}}{MSE}$$

and comparing it to critical values from an F distribution.

Under the assumption that the error process involved is independently and identically distributed (denoted by either *iid* or $\sim$) $N(0, \sigma^2)$ the F^* can be compared to some critical value, chosen from at a user-specified confidence level, to determine if a covariate should be added to (or retained in) the model. As one adds predictors that contribute meaningfully to the fit of the model, the MSR will generally increase and the MSE will generally decrease.

One can then start with no (or some or all of the) predictors in the model and cyclically add (or subtract) predictors, according to whether or not their calculated F^* value is large (or small) enough. Then once a new predictor is added (or dropped), all the remaining predictors are tested to see if they still contribute enough to contribute to the model fit (as judged by F^*) to justify keeping in the model. After exhaustive search, one can obtain the final model.

The limitations of this approach depend primarily on the regression methodology used. Stepwise regression via an F-test is well known to be sensitive to multi-collinearity in the data. More importantly, any arbitrarily chosen F$(1 - \alpha)$ statistic rarely provides reliable performance in terms of risk control, and does not correspond to anything like a $1 - \alpha$ experiment-wise error rate. Furthermore, there is the additional problem of many simultaneous tests, leading to over-confidence in the result of the experiment.

While these points are features of stepwise-regression in particular, there is still the overarching problem among all regression methods used in this application that many competing models may fit the data nearly equally well. Thus the choice among models is very important for the end user, but in stepwise regression there is no mechanism for mitigating the obvious fact that only one of them is physcially correct, while several competing models may fit the data (nearly) equally well. In essence, stepwise methods, while straightforward to implement and flexible with respect to the size of the gas library, are notorious for finding models that have high R^2 statistics but are greedy with respect to the inclusion of parameters.

Gallagher et al [7] [8] employ a variety of ordinary least squares (OLS) and weighted least squares (WLS) methods to determine the integrated CPL C_n 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 3.12). The surface leaving radiance terms were modeled via a principle components technique based on the off-plume pixels. The authors show success with identifying some species; however, their ability to determine the integrated CPL C_n was limited primarily by their inability to characterize the background radiance $\boldsymbol{\xi}_s \cdot \boldsymbol{B}(T_s)$, followed by sensor noise, and uncertainty in plume temperature T_p.

Another regression-based approach used by Chilton [9] 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. This amounts to computing

$$pr(\boldsymbol{\beta}|Data) = \sum_{k=1}^{K} Pr(\boldsymbol{\beta}|\mathcal{M}_k, Data)pr(\mathcal{M}_k|Data) \tag{2.8}$$

which is the weighted average of the regression coefficients given the data. This method also allows the computation of explicit solutions for the posterior mean $E[\boldsymbol{\beta}|Data]$ and posterior variance $Var[\boldsymbol{\beta}|Data]$ of the parameter vector $\boldsymbol{\beta}$.

For many problems the full set of possible models is huge, considering that there are N gases available for regressing against, there are 2^N 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) [10]. BIC is an approximation to what are known as Bayes factors (BF), which are the ratio (for a pair of competing models)

$$BF = \frac{Pr(\mathcal{M}_0|Data)}{Pr(\mathcal{M}_1|Data)}. \tag{2.9}$$

A BF greater than unity indicates evidence for $\mathcal{M}_0$ over $\mathcal{M}_1$.

This can be combined with "branch and bound" techniques that find optimum subsets of (full) models by ignoring subsets of models that do not have the potential to improve the fit to the data. Many of these computational shortcuts rely on "vague" or

"uninformative" priors for the regression coefficients, which on the surface might seem to be advantageous from the perspective of trying to be objective in ones inference, but it prevents the legitimate use of prior knowledge in scenarios where the user might wish to incorporate prior scene knowledge or some sort of informed expert opinion about the regression parameters. If one is willing to live with these tradeoffs (and make certain distributional assumptions), there are a number of closed-form solutions for the posterior mean and variance of $\boldsymbol{\beta}$ [11], furthermore, there exists widely available routines for the automated implementation of BMA.

One interesting recent Bayesian approach by Heasler *et al.* [12] focused on using non-linear Bayesian regression implemented via the Metropolis-Hastings algorithm [13]. These authors treated the background-plume temperature contrast and the surface emissivity as nuisance parameters, and utilized a non-linear regression scheme because Equation 3.12 is non-linear with respect to these parameters because of the blackbody functions. The authors chose to use a "full Bayesian" approach, which relies on creating prior distributions for scene emissivities by modeling a database of material emissivities, which may or may not represent the emissivity in any particular scene, not to mention the challenges of faithfully representing the scene as a mixture of those materials. While this method is the most comprehensive in its approach to physically modeling the problem of gaseous effluent identification, it still has the basic problem of model uncertainty.

Ultimately they show some success with their approach, however they (arguably) did not make full use of the Bayesian paradigm, in that for a given scene there is opportunity to leverage scene-specific prior information. An obvious choice for the source of this scene-specific information could be physical models such as OLSTER. Even if one harbored some suspicions about the validity of the point estimates that a model (such as MODTRAN a commonly used atmospheric compensation code) produced, it would still be a fairly simple matter to incorporate those point estimates as a "point mass" in a prior structure, combined with some uncertainty as defined by a subject matter expert. Thus allowing the data to have reasonable latitude in shaping the posterior estimates. The Appendix contains a more in-depth discussion of this process. Suffice it to say that there is room in the Bayesian framework for a user to make a subjective judgement regarding their pre-data understanding of a parameter and for practical reasons they may wish to use some sort of *ad hoc* instantiation of this.

There has also been considerable work in application-specific scenarios [14] [15] testing the sensitivity of detection to a variety of atmospheric parameters. These studies are generally based on examining the spectral matched filter (SMF) receiver operating curve (ROC) sensitivity to degradation in atmospheric parameters. 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.

Other classic approaches often include the use of SMFs, projection operators, or Principle Components Analysis (PCA) [16] [17]. SMFs require 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 judging the output as a detection. SMF's often use some sort of scaled quadratic detector, a common one is known as Constrained Energy Minimization (CEM)

$$CEM = \frac{(\boldsymbol{x} - \boldsymbol{t})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}})}{(\boldsymbol{x} - \bar{\boldsymbol{x}})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}})}. \tag{2.10}$$

In the CEM detector $\boldsymbol{x}$ is a pixel vector under test, $\boldsymbol{t}$ is the known target spectrum one is searching for, $\boldsymbol{\Sigma}^{-1}$ is the inverse of the image-wide spectral covariance matrix, and $\bar{\boldsymbol{x}}$ is the image-wide mean vector. This yields a scalar for every pixel which is notionally proportional to the probability that the target $\boldsymbol{t}$ exists in that pixel. One simply creates a threshold for the CEM values and assigns $\forall CEM \geq 1 - \alpha$ to a "target" class, which in most cases includes a fair number of false detections and fails to detect some legitimate target pixels.

Once the set of likely gas plume pixels is selected out of the image, one now has the task of trying to determine the mixture of gas(es) that make up that plume (ignoring false detections and missed detections for the moment). This can only be done by exhaustively using different $\boldsymbol{t}$ vectors above and seeing which ones produce the highest return for a given pixel, but this is tedious, computationally intensive, and isn't practical in the case where there are significant numbers of gas species that are mixed together in the plume. For this problem we currently have a library of 31 gas spectra and would

want to ultimately demonstrate a method on at least three or four gas mixtures, note $\binom{30}{3} = 4060$ and $\binom{30}{4} = 27{,}405$, which would require roughly $\mathcal{O}(\texttt{gases in library!})$ SMF runs, for every single pixel.

A recent approach by Salvador et al [18] makes use of the wavelet packet transform in a detection framework by trying to find a wavelet basis for projecting a pixel vector against, that allows maximum separability of a gas spectra of interest and the background it is observed with. The novel feature of this approach is the multi-resolution and translation-invariant nature of wavelets. This approach still requires exhaustive testing against all gases in a library, as well as the custom development of the optimal wavelet decomposition for each gas, which is likely to be sensor-specific. Ultimately, the use of wavelet-based detection methods still have the same general set of advantages and disadvantages as SMF-based methods.

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. Practically this means that exhaustive search is required in order to determine if a gas (or mixture of gases is present) and there generally aren't any shortcuts such as recursive algorithms that can be used to narrow the search space. 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 [9] [12], these methods often do not provide a framework for risk control in any global sense. That is to say, they don't produce anything like a multiple simultaneous confidence interval around $\boldsymbol{\beta}$ at a uniform confidence level $1 - \alpha$.

Concept of Operations

The problem of gaseous effluent identification includes two different concepts of operation (CONOPS) both persistent surveillance and single look scenarios. Both cases can be practically extended to remove the action of an operator in the loop, but for the development of this approach, the author assumes that a human will generally be specifying which gases are most likely in the scene, providing human interpretation of some sort of detection results and providing the cueing necessary to reduce the need to

process every pixel in an image. This is also consistent with the current development state of OLSTER. Even though OLSTER is semi automated, it does not exist in a hardened form and its use currently requires some level of operator interaction/oversight. As such, this approach focuses on the process of plume identification, and relatively little will be done to contribute to the sizable body of literature on target detection. The method that is proposed will be broadly applicable to most data sets and nothing prevents its unsupervised application in the absence of an operator in the loop.

Chapter 3

Methods

This section describes a novel approach developed by the author for gaseous plume identification. First an in-depth background on the use of multiple linear regression is presented, followed by an approach that was developed to deal with the rank-deficiency of the gas library and spectral resolution of the data available for this research. The discussion then proceeds to introduce the broad class of Bayesian methods, including Markov Chain Monte Carlo (MCMC) and Gibbs sampling techniques. The remainder of this section will focus around a model selection technique called Gibbs Variable Selection (GVS) and a variant of this technique that the author developed for unsupervised application to imaging problems. Finally the behavior of this new variant of GVS will described, including how it can be tuned to a certain data set, extensions to multi-pixel inference, and finally a description of how its performance is measured is presented.

Multiple Linear Regression

Often the problem of gas plume identification is cast in the form of regression equations. This can be accomplished by noting that in Equation 3.12, once OLSTER has been applied to the non-plume pixels, the "contrast term" is linear with respect to the plume CPL coefficients C_n. By inserting the point estimates for all of the radiometric terms Equation 3.12 can be recast in the canonical form of a linear regression model

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}, \tag{3.1}$$

where Y is a $C \times 1$ "response" vector and is made up of observed radiance values, and X is known as the "design matrix" (typically) $C \times (N+1)$ whose columns are made up of the "predictors" or "covariates" that are assumed to be known without error and can be used to predict the Y values. ϵ is an error process that accounts for the discrepancy between the predicted and the actual values of Y. The β is a $(N+1) \times 1$ vector of coefficients that one generally wishes to estimate. In order to estimate an intercept it is necessary to include a column of ones as the leading column in X and thus the leading entry of β is the intercept. 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 β are given by

$$\hat{\beta} = (X'X)^{-1}X'Y, \tag{3.2}$$

where X' denotes the matrix transpose of X. The OLS estimates are theoretically and practically important because under the assumption that the error process in Equation 3.1 is *iid* $N(0, \sigma^2)$ then the OLS estimates are the maximum likelihood estimates (MLE) for β, are unbiased and are asymptotically efficient, they are also minimum-variance unbiased estimators (MVUE) or best among linearly unbiased estimators (BLUE). These are important results in that it implies that no unbiased estimator with a lower variance can be found. It should also be noted that practically the estimates obtained via OLS are sensitive to (near) rank-deficiency of $X'X$ in Equation 3.2, the result being that the standard errors of the estimated regression coefficients grow dramatically [19]. This implies that in problems where some degree of rank-deficiency (or multi-collinearity) exists, one might be better off to tolerate a biased estimate in return for a method that offers lower standard error.

There are a number of reasons (both practical and theoretic) for directly dealing with the issue of multi-collinearity (near rank-deficiency or rank-deficiency) of the columns of the design matrix. From a statistics point of view, there is the above mentioned impact on the variability of fitted regression coefficients. From a linear algebra point of view the motivation is to reduce the error (in a point estimate sense) of the estimated regression coefficients. Generally speaking, all regression methods will benefit from operating with a more-nearly-orthogonal (i.e. less rank-deficient) design matrix. The Bayesian regression method presented in the Methods section is just as sensitive to

this rank deficiency, as frequentist methods such as OLS. This motives the discussion of ways to measure and mitigate multi-collinearity.

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. The condition number is based on the singular value decomposition (SVD) where any matrix $\boldsymbol{X}$ of arbitrary dimensions $C \times N$ and $\boldsymbol{X} \in \mathcal{C}^{C \times N}$ is decomposed into

$$X = UWV^{'} \tag{3.3}$$

where

$$\begin{cases} U \in \mathcal{C}^{C \times C} & \text{is unitary \& contains eigen vectors of } XX^{'} \\ V \in \mathcal{C}^{N \times N} & \text{is unitary \& contains eigen vectors of } X^{'}X \\ W \in \mathcal{R}^{C \times N} & \text{is diagonal.} \end{cases}$$

and the condition number κ is defined as the ratio of the largest (first) singular value to the Nth (smallest) singular value of the matrix $\boldsymbol{X}$

$$\kappa = \frac{\sigma_1}{\sigma_N}. \tag{3.4}$$

Alternatively, VIFs are computed for the design matrix $\boldsymbol{X}$ by regressing each column of $\boldsymbol{X}$ against all of the other columns (except the first column, whose values are all unity), then transforming the R^2 values for each of the regressions via

$$VIF_n = \frac{1}{1 - R_n^2}. \tag{3.5}$$

General rules for VIFs being too high are values over a threshold such as 5 or 10. In that case the β_n with the highest VIF_n should be removed and the process repeated until all of the VIFs are below the threshold.

The 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 30 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
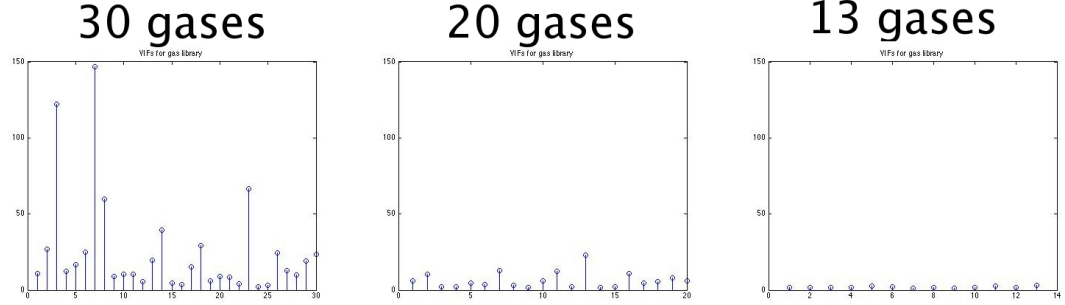
Figure 3.1: VIFs of gas library undergoing thinning (full, partially thinned, and fully thinned - left to right)

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. This indicates that even though there wasn't a substantial penalty in terms of the error of the point estimates of the OLS process (as judged by the condition number) the variability of the fitted regression coefficients is much more sensitive to the increase in rank-deficiency brought on by downsampling. This is evidenced by the substantial increase in the maximum VIF between the two libraries.

In order to reduce the multicollinearity of the gas library, the author developed a two stage approach that combines $VIFs$ and PCA. 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 remainder of the library $\boldsymbol{X}$ was decomposed via the SVD as previously described

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{W}\boldsymbol{V}^{'} \tag{3.6}$$

and by visual inspection of the *last* columns of $\boldsymbol{V}^{'}$ in comparison to the percentage of variability attributed to each right singular vector $\boldsymbol{W}$, 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 is typically done. By noting the gases which have a large contribution to the last few columns of $\boldsymbol{V}^{'}$ it is possible to determine which gases contribute to the most linearly degenerate
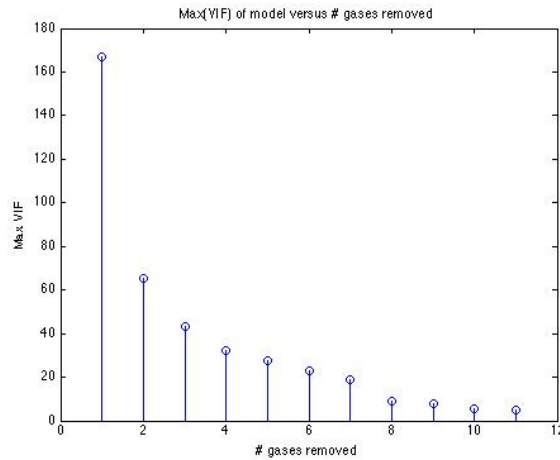
Figure 3.2: max(VIFs) versus number of gases removed

subspace of the eigenvectors that span $\boldsymbol{X}'\boldsymbol{X}$. The key point is that even though the right singular vectors describe linear combinations of the gas library, the last few linear combinations are frequently dominated by just a few gases. This means that these gases are the primary contributors to the most linearly degenerate dimension spanned by the covariance matrix of the gas library, and removing them will have the greatest impact on the condition number of the gas library.

Using VIFs as a sensitive measure of the degree of rank deficiency, one can track the change in rank-deficiency by examining the decrease in VIFs through each step in the thinning process. This decrease in VIFs between the 3 libraries (full, partially thinned via VIFs, and fully thinned via VIFs and PCA) is shown in Figure 3.1.

The reason this two stage approach was employed was because even after the VIF step alone, the condition number was still sizeable, and there were convergence issues with the iterative Bayesian method that will be introduced shortly. This second step produced a much more nearly orthogonal library, which performed well with the method that will be presented, as well as standard methods such as stepwise regression. Through this two stage process, the library was thinned from 31 to 20 gases via the VIF approach, and from 20 to 13 gases through the PCA process. Through these steps, the condition number decreased to 4×10^2 in the final library. The change in max(VIF) as a function of gases being removed can be seen in Figure 3.2 The change in singular
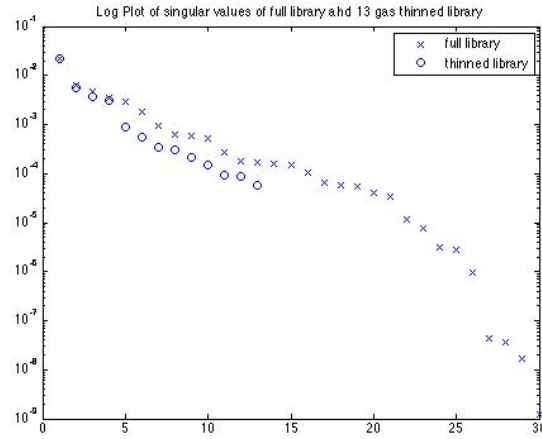
Figure 3.3: Log singular values of the full and 13 gas library

values between the full and fully-thinned singular values are show in Figure 3.3.

The importance of this thinning procedure also goes beyond this particular problem of gas identification, as it could be applied to problems spectral un-mixing or image classification. Generally speaking, any least-squares application is going to have the issue of multi-collinearity and this method provides a way to control both the variance of the regression estimates (through the VIF's) as well as the magnitude of relative error in the point estimates of the regression estimates themselves (through the PCA process). Finally, 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.

Regression applied to gas identification

If we consider the overall problem of determining which gases are in an observed pixel, this corresponds to the statistical problem of model uncertainty. Model uncertainty refers to the problem where many competing (and mutually exclusive) models could fit the data nearly equally well, and thus in addition to estimating the regression coefficients one also needs to estimate what is the correct model.

This highlights several important issues that will arise in the gas identification problem.

1. There is a clear multiple comparisons problem in that one desires to make simultaneous probability statements about the various fitted coefficients ($\boldsymbol{\beta}$).

2. For any given model the user needs access to the probability that the stated model (subject to certain assumptions) actually generated the data, and one generally prefers the model that ranks highest.

3. In addition to knowing the estimated amount of gas present, one would generally like to know that probability that any of that gas is present. That is to say, one might like to have independent judgments regarding the presence/absence, and quantity of a gas.

4. In practice observed gas plumes are likely to be larger than a single pixel, and observed in more than just a single image. Thus one might wish to include some sort of spatial or spectral aggregation in a probability model that would allow for the inclusion of these known phenomena.

Bayesian Inference and Gibbs Variable Selection

Recall that Bayesian inference differs from classical (frequentist) inference in that parameters in a statistical model (the $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ in Equation 3.8) are no longer thought of as fixed unknown points, where the goal of inference is to create point or interval estimates as near as possible to that unknown value or that contain the parameter value (respectively). Bayesian inference relies on representing parameters as probability distributions, the relative shape (or width) of which is used to represent how much uncertainty there is in one's understanding of that parameter.

Bayesian statistics are based on the conditional probability rule where $\boldsymbol{\theta}$ is some vector of parameters of interest and *data* are observations, yielding

$$posterior = f(\theta|data) = \frac{f(data|\theta)f(\theta)}{f(data)} \propto f(data|\theta)f(\theta) = Likelihood \times prior. \quad (3.7)$$

It is important to note that $f(\theta)$, the prior distribution of the parameter θ, represents one's pre-data understanding of the parameter (note that distributions involving θ and *data* 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 $f(x)$ to create a "post data" estimate of the parameter, which is conditioned upon the data $f(\theta|data)$. The distribution of $f(\theta|data)$ is the posterior distribution of θ and is the basis for all probability statements about θ [13].

Because the specification of prior distributions is, at least to some degree, somewhat arbitrary, it is important to recognize the "tradeoff" that takes place between the prior, likelihood, and the posterior. First, one should recognize that the only possible values that a posterior $f(\theta|data)$ can contain, must have non-zero probability on that same interval as the prior $f(\theta)$. Intuitively this makes sense, as if a subject matter expert had said "there is no way that the parameter θ is less than three" and places correspondingly zero probability on $x \leq 3$. Thus it is not conceivable (if one believed the expert) that after observing data, the posterior would imply that $x = 2.8$ was a possible value. Without loss of generality, $\boldsymbol{\theta}$ will be assumed to be vector valued unless noted otherwise.

If for example, one wished to conduct inference on the mean μ of a univariate random process and used the prior $\mu \sim iid\, N(0, \sigma = 10)$ and the likelihood for the data was $f(data|\mu) \sim iid\, N(5, \sigma = 1)$, that is to say that the actual process generating the data had a $\mu = 5$ when ones prior belief was that the most likely value for the mean was zero. One would expect the central tendency for the posterior $f(\mu|data)$ to fall somewhere in between 0 and 5. As an aside, there is a similar compromise made between the variance of the prior and the likelihood. In the days of pencil and paper calculus, these relationships (and the reliance on closed form solutions) were central to the practice of Bayesian methods, but currently they can be thought of as supplying intuition about the expected outcome of the inference.

It is important to point out that there exists the opportunity for a user to create *ad hoc* prior distributions that combine some sort of point mass located at a known estimate for a parameter, in addition to a larger more diffuse area of probability support. This informs the inference process, but leaves the opportunity for the data to suggest a posterior value other than the one of highest *a priori* mass. Because the choice of a prior distribution is inherently subjective, this is a rational way of balancing some limited amount of knowledge about a parameter while being open to other values of the parameter. The downside of this sort of approach is that "non-standard" hybrid

densities such as the ones just described, are often only tractable with the Metropolis or Metropolis-Hastings algorithm, which is computationally simple and generic, but is also somewhat inefficient in traversing the sample space.

Bayes rule (Equation 3.7) states the probability relationship that can be used to calculate this combined "data plus prior belief" probability, but practically it is not immediately obvious how one is supposed to actually execute the operations necessary to compute a posterior in anything but the simplest of discrete probability scenarios. Initially it might seem appealing to try and solve for the closed form of the posterior and simulate from it, but this is not practical in many cases. In practice, what is done is to find 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. This MC iteration (or MCMC) is often accomplished through Gibbs sampling or the Metropolis algorithm. See the Appendix for a more detailed description of why MCs can be used for approximating posterior distributions and how they are designed.

Gibbs sampling or "successive substitution" sampling [13] is the practice whereby for any parameter vector $\boldsymbol{\Theta}$ one has the starting point $h(\boldsymbol{\theta}) \propto f(data|\boldsymbol{\theta})f(\boldsymbol{\theta})$ where $h(\boldsymbol{\theta})$ is the distribution one hopes will approximate the posterior $f(\boldsymbol{\theta}|data)$. Let the starting vector be denoted by $\boldsymbol{\theta}^0$ where the superscript indicates the iteration number M

$$\boldsymbol{\theta}^0 = (\theta_1^0, \theta_2^0, \ldots, \theta_N^0).$$

Each full iteration (for iterations $m = 1, \ldots, M$ there are p sub-iterations and the subscript denotes the sub-iterations ($p = 1, \ldots, P$) one for each parameter in the chain. Thus on each iteration, there are p sub-iterations where one successively samples every single parameter from the conditional distribution $f(\boldsymbol{\Theta})$ (the order of sampling generally is arbitrary), while leaving the other parameters fixed from the previous iteration:

1. $\boldsymbol{\theta}_1^1$ from $h(\cdot, \theta_2^0, ..., \theta_P^0)$

2. $\boldsymbol{\theta}_2^1$ from $h(\theta_1^1, \cdot, \theta_3^0, ..., \theta_P^0)$

3. $\boldsymbol{\theta}_3^1$ from $h(\theta_1^1, \theta_2^1, \cdot, ..., \theta_P^0)$

   $\vdots$

4. $\boldsymbol{\theta}_m^1$ from $h(\theta_1^1, \theta_2^1, \theta_3^1, ..., \cdot)$

$$\vdots$$

5. $\boldsymbol{\theta}_M^2$ from $h(\cdot, \theta_2^1, \theta_3^1, ..., \theta_P^1)$

Thus on the mth iteration one uses the previous iteration's values for all parameters, except the one that is being randomly drawn. Given the "memoryless" property of MC's, the current state of a MC should only depend on its immediate predecessor. The goal of this entire endeavor is that after a sufficient number of iterations M, that $h(\boldsymbol{\theta})$ the distribution of $\boldsymbol{\theta}$ is a close approximation of the posterior $f(\boldsymbol{\theta}|data)$.

Practically, to achieve memoryless behavior one must discard a sufficient number values at the beginning of each MC so as the mitigate the possibility that the starting values of the chain are impacting the chain's behavior. This process is known as "burn in" and often somewhere between 10 to 1,000 iterations are typically discarded. Although Gibbs sampling is simple to implement and very effective, it does rely on the user having access to the full conditional distributions for all of the parameters that are the subject of inference.

Some last general notes about Bayesian inference, before moving on to specific uses of the paradigm include:

- Bayesian inference is inherently subjective. With different priors, one can observe the same data and form different posteriors for a given parameter.

- The posterior generally represents a sort of compromise between the prior and the data. The mean of the posterior will generally be between the mean of prior and the maximum likelihood estimate from the data and the variance of the posterior will generally be smaller than the variance of the prior, because the data provides information on the parameter.

- The support of the posterior is bounded by the support of the prior. This means if the user places exactly zero prior probability on parts of $\boldsymbol{R}^n$ for a parameter, then the posterior can not be non-zero in those places

- There is often interest in reducing ones dependence upon, or sensitivity to, specific values of observed data by making the prior distribution "flat", "diffuse", "vague", "robust", or "non-informative." Often this can be achieved in some sense by using (depending on the parameterization) a prior that does not seem to prefer

any specific value of $\boldsymbol{\theta}$. Examples include Uniform(-∞, ∞) or Normal with large variance.

- It is important to recognize beforehand that one one should try to use "proper" prior distributions if possible, *i.e.* distributions that sum to unity. For example, the distribution $Uniform(-\inf, \inf)$ is not proper, but the $Uniform(-1000, 1000)$ is. If proper priors are not available, improper ones can still be used in many cases, but then the posteriors will need to be normalized to have unit area.

Having introduced the Bayesian approach to statistical inference, the discussion now turns to the application specific problem of not only using Bayesian methods to solve a regression problem, but also assist in the process of choosing a regression model.

Model Selection via Gibbs Variable Selection

The author proposes the use of a Bayesian model selection technique known as Gibbs Variable Selection (GVS) [20] [21] for approaching the simultaneous problems of regression and model selection. GVS is based on the premise that instead of solving the usual regression model (Equation 3.1) that one uses a modified version that contains an $N \times 1$ γ vector of indicator variables

$$\boldsymbol{Y} = \sum_{n=1}^{N} \gamma_n X_n \beta_n + \epsilon. \tag{3.8}$$

Then using Markov Chain Monte Carlo (MCMC) one can make Bayesian inference on both $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$.

GVS applied to multiple liner regression uses the model in Equation 3.8 and assumes the regression coefficients $\boldsymbol{\beta}$ are a vector of *iid* $MVN(0, \sigma^2 \boldsymbol{I})$ variables and the indicator variables are *iid* Bernoulli(p). Then model space $\mathcal{M}$ (the set of all 2^N possible models) is then partitioned according to those variables that are **in** the rth model under consideration M_r by $\boldsymbol{\gamma}$ and those that are not in model M_r by $/\boldsymbol{\gamma}$. These then directly correspond to the appropriate sets of regression parameters $\boldsymbol{\beta}_\gamma$ and $\boldsymbol{\beta}_{/\gamma}$ 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 $\boldsymbol{\gamma}$

$$f(\boldsymbol{\beta}, \boldsymbol{\gamma}) = f(\boldsymbol{\beta}_\gamma | \boldsymbol{\gamma}) f(\boldsymbol{\beta}_{/\gamma} | \boldsymbol{\beta}_\gamma, \boldsymbol{\gamma}) f(\boldsymbol{\gamma}). \tag{3.9}$$

The likelihood for this model is given by

$$f(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\gamma}) = f(\boldsymbol{y}|\boldsymbol{\beta}_{\gamma},\boldsymbol{\gamma}). \tag{3.10}$$

Thus the posterior distribution of the parameters is proportional to the product of the RHSs of Equations 3.9 and 3.10.

GVS relies on the indicator variables to partition the model space, such that $f(\boldsymbol{\beta}_{\gamma}|\boldsymbol{\gamma},\boldsymbol{y})$ is the posterior of the model specified by the vector $\boldsymbol{\gamma}$, while $f(\boldsymbol{\beta}_{/\gamma}|\boldsymbol{\beta}_{\gamma},\boldsymbol{\gamma},\boldsymbol{y})$ is the conditional prior distribution of the parameters *not* included in model γ. In terms of model selection one can interpret $f(\boldsymbol{\beta}_{\gamma}|\boldsymbol{\gamma})$ as the prior distribution for a given model, as compared with $f(\boldsymbol{\beta}_{/\gamma}|\boldsymbol{\beta}_{\gamma},\boldsymbol{\gamma})$, which is commonly called a "pseudo-prior" as the parameter vector $\boldsymbol{\beta}_{/\gamma}$ does not actually influence the posterior distributions of interest $f(\boldsymbol{\beta}_{\gamma}|\boldsymbol{\gamma},\boldsymbol{y})$.

GVS makes use of Gibbs sampling for the estimation of the posterior densities of the indicator variables γ via

$$f(\gamma_n|\boldsymbol{Y}) \sim Bern(\frac{O_n}{1+O_n})$$

and the un-scaled probability of inclusion (for the nth gas) is given by the odds ratio

$$O_n = \frac{f(y|\beta,\gamma_n = 1,/\gamma)}{f(y|\beta,\gamma_n = 0,/\gamma)}\frac{f(\beta|\gamma_n = 1,/\gamma)}{f(\beta|\gamma_n = 0,/\gamma)}\frac{f(\gamma_n = 1,/\gamma)}{f(\gamma_n = 0,/\gamma)}.$$

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 *all* models whose complexity is not limited to an arbitrary number of gases. An example of GVS applied to a simple regression problem is included in the Appendix.

In this case we assume a priori independence, which leads to a product form for the densities $\prod_{}^{N} Pr(\boldsymbol{\gamma}|\boldsymbol{\beta},\sigma^2,data)$ and $\prod_{}^{N} Pr(\boldsymbol{\beta}|\boldsymbol{\gamma},\sigma^2,data)$; allowing us to sample from their marginal densities one realization at a time.

Gibbs sampling is particularly useful in the case where one can identify the conditional distributions of each parameter, conditioned on all the other parameters, 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 $\Pr(\beta|data)$ occurs first, then the sampling from the $\Pr(\gamma|data)$ distribution takes place, with the caveat that all $Pr(\beta|data)$'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).

In an approach developed by the author, if one limits themselves to the MVN/inverse-gamma family of densities for this problem, one can sample all of the $\Pr(\boldsymbol{\beta}|data)$'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 modified GVS process (termed M-GVS) was implemented using standardized regression (without an intercept term), via the process described below.

For the conjugate MVN/inverse-gamma sampling process proceeds, given that

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

one draws samples from

$$Pr(\boldsymbol{\beta}|\sigma^2, \boldsymbol{Y}) \sim MVN(\hat{\boldsymbol{\beta}}, \boldsymbol{V_\beta})$$

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

$$\boldsymbol{V_\beta} = \boldsymbol{R}^{-1}(\boldsymbol{R}^{-1})' = (\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

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

$$s^2 = \frac{1}{C-N}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$$

and then

$$\sigma^2|Y \sim Inv - \chi^2(C - N, s^2)$$

Then at every step, the variable selection sub-routine is carried out via Gibbs sampling whereby samples from the posterior indicator variables are drawn via

$$f(\gamma_n|\boldsymbol{Y}) \;\sim\; Bern(\frac{O_n}{1+O_n})$$

and the un-scaled probability of inclusion is given by the odds ratio

$$O_n = \frac{f(y|\beta,\gamma_n=1,\gamma_{/n})}{f(y|\beta,\gamma_n=0,\gamma_{/n})}\frac{f(\beta|\gamma_n=1,\gamma_{/n})}{f(\beta|\gamma_n=0,\gamma_{/n})}\frac{f(\gamma_n=1,\gamma_{/n})}{f(\gamma_n=0,\gamma_{/n})}.$$

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.

Computational Details of Gibbs Variable Selection

M-GVS is also simple to implement as it functions as two nested "for" loops, the outer one encompassing the regression part, while the inner loop is the variable selection piece. For iterations $m = [1, \ldots, M]$ and covariates $n = [1, \ldots, N]$ and using Matlab notation for pseudo-code

> for m = 1:M
> Conjugate sampling of β_n
> $\vdots$
>
>   for n = 1:N
>   Gibbs sampling of γ_n
> $\vdots$
>
>   end
> end.

Table 3.1: M-GVS code architecture

The outer loop works by drawing samples from

$$Pr(\boldsymbol{\beta}|\sigma^2,\boldsymbol{Y}) \sim MVN(\hat{\boldsymbol{\beta}},\boldsymbol{V_\beta})$$

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

$$\boldsymbol{V_\beta} = \boldsymbol{R}^{-1}(\boldsymbol{R}^{-1})' = (\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

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

$$s^2 = \frac{1}{C-N}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$$

and then

$$\sigma^2|Y \sim Inv - \chi^2(C - N, s^2)$$

At the heart of the variable selection sub-loop are two equations where we update the (n-th) indicator variable by

$$\gamma(n) = Bernoulli(\frac{O_n}{1 + O_n}) \tag{3.11}$$

where

$$O_j = \frac{f(y|\beta, \gamma_n = 1, \gamma_{/n})}{f(y|\beta, \gamma_n = 0, \gamma_{/n})} \frac{f(\beta|\gamma_n = 1, \gamma_{/n})}{f(\beta|\gamma_n = 0, \gamma_{/n})} \frac{f(\gamma_n = 1, \gamma_{/n})}{f(\gamma_n = 0, \gamma_{/n})}. \tag{3.12}$$

In Equation 3.12 the numerator terms are for the current model with the nth gas turned "on" ($\gamma_n = 1$) and all the denominator terms are the same model, but with the nth gas turned "off" ($\gamma_n = 0$).

Pr(data | current β and γ)	Pr(β | current γ)	Prior for γ
Pr(data | current β and γ)	Pr(β | current γ)	Prior for γ

Table 3.2: Model selection terms

which for the sake of simplicity, the terms in Equation 3.12 will be temporarily referred to as [A,B,C,D,E,F] according to the convention in Table 3.

A	B	C
D	E	F

Table 3.3: Notation for model selection terms

The basic challenge to the implementation of M-GVS is in computing the terms A,B,D, and E which are evaluated from 50-dimensional MVN distributions. These

terms can (depending on SNR) commonly range anywhere from e^1 to e^{-265} and often evaluate numerically as zero. The problem is that the smallest real positive uniquely identifiable constant in Matlab (obtained from the function realmin) is roughly e^{-307} and the denominator (most frequently, but also the numerator) of the variable selection piece frequently became computationally indistinguishable from zero. It is also noteworthy that these values returned by evaluating the MVN densities are often below machine epsilon, which for a double-precision float in Matlab is $2e^{-16}$.

The author devised the following strategy for dealing with the numerical issues in computing Equation 3.12. The basic framework is as follows:

1. Note that our end use for the quantity $\frac{O_n}{1+O_n}$ is simply a parameter estimate for the posterior inclusion probability $\hat{p}$ that gets fed into the $Bernoulli(\hat{p})$ distribution at every sub-iteration.

2. The function $\hat{p} = \frac{O_n}{1+O_n}$ will have (for all practical purposes), three regions:

    (a) Where $\hat{p} \approx 0$ (it is within some ϵ-neighborhood of zero), and γ_n has a high probability of being zero on the next iteration. This is a case where the terms A,B,D, and E may be difficult to evaluate.

    (b) Where $\hat{p}$ is not appreciably close to zero or unity. This is the region where it will be well behaved.

    (c) Where $\hat{p} \approx 1$ (it is within some ϵ-neighborhood of unity), and γ_n has a high probability of being unity on the next iteration. This is a case where the terms A,B,D, and E may be difficult to evaluate.

3. With some algebraic re-arrangement, see that

$$\hat{p} = \frac{O_n}{1 + O_n}$$

can be re-arranged into

$$\frac{-\hat{p}}{1 - \hat{p}} = O_n = \frac{ABC}{DEF}.$$

Taking the natural log of both the RHS and LHS we get

$$ln[\frac{-\hat{p}}{1 - \hat{p}}] = ln[\frac{ABC}{DEF}] = ln(A) + ln(B) + ln(C) - ln(D) - ln(E) - ln(F).$$

The dominant terms of which (in all degenerate cases) are A,B,D, and E and we approximate the RHS of the preceding equation with

$$ln[\frac{-\hat{p}}{1-\hat{p}}] \approx ln(A) + ln(B) - ln(D) - ln(E).$$

4. Since the functional form of terms A,B,D, and E are MVN densities, with the constants canceling before the logarithms were taken, this can be represented as

$$ln[\frac{-\hat{p}}{1-\hat{p}}] \approx -.5(Y-\gamma X\beta)'(\sigma^2 I)^{-1}(Y-\gamma X\beta)_A + -.5(\beta_{OLS}-\gamma\beta)'(\sigma^2 V_\beta)^{-1}(\beta_{OLS}-\gamma\beta)_B$$

$$-(-.5)(Y-\gamma X\beta)'(\sigma^2 I)^{-1}(Y-\gamma X\beta)_D - (-.5)(\beta_{OLS}-\gamma\beta)'(\sigma^2 V_\beta)^{-1}(\beta_{OLS}-\gamma\beta)_E$$

where the subscript β_{OLS} denotes the ordinary least squares estimate of β (which in this case is taken as the mean of β). Note that in terms A and B that γ has one additional unity value than in the denominator, in the nth position, where in terms D and E the $\gamma_n = 0$ was set to zero, and all of the other indicators where left in their current state.

5. Given the above equation, we can analytically determine regions on the LHS that are approximately unity or zero, and thus establish bounds on the exponents on the RHS.

Before continuing, it is important to point out that the point of this exercise was to create a representation for $\hat{p}$ that eliminated the need to compute O_n for numerically intractable cases. These cases can be identified through careful re-arrangement of the problem, making use of the fact that the dominant terms on the RHS of the preceding equation are of an exponential form.

We take advantage of the three regions mentioned previously, by defining a heuristic where based on some small value for p^* ($p^* = .0001$ was used in all of the following examples)

$$\hat{p} =$$
$$\begin{cases} 1 & ln(A) + ln(B) - ln(D) - ln(E) > ln[\frac{-(1-p^*)}{1-(1-p^*)}] \\ 0 & ln(A) + ln(B) - ln(D) - ln(E) < ln[\frac{-p^*}{1-p^*}] \\ \text{compute } \hat{p} \text{ directly via } \frac{O_n}{1+O_n} & else \end{cases}$$

Limitations of Bayesian Inference and Gibbs Variable Selection

Bayesian inference is not immune from the need to do general regression diagnostics and the need to assess the fit of a model in the traditional linear modeling sense. In this regard, GVS is equally subject to user-error as frequentist methods. This will always be a problem, particularly when statistical methods are applied *en masse* to blocks of data that are so large that only the most cursory of model diagnostics can be performed, and the concept of operations requires minimal user interaction with the data. This reliance on automated methods of analysis, in the absence of custom (and time consuming) modeling and method development, are a common challenge in arenas such as this where statistical methodology is applied in the physical sciences.

Despite its strengths there are a number of drawbacks to using GVS for model selection. First is the dependence on a design matrix that is (reasonably) orthogonal, which was acknowledged as a limitation by its inventors [21]. For the development of this work, particularly given the low-spectral resolution dataset available to train on, the author chose to focus on thinning the library instead of finding ways to harden GVS against multi-collinearity.

Because the absorption spectra for chemically similar structures tend to be similar, one would expect that the modes of excitation of their molecular bonds to be somewhat similar. A simple example is the family of hydrocarbons that exhibits what is is commonly called a "benzene ring" such as: toluene, xylene, phenol, etc. Any multi-colinearity in the design matrix will cause poor behavior of the MCs. Ultimately the solution to this problem might require a more flexible form of MCMC sampling such as Reverse Jump Metropolis Hastings.

In addition to poor behavior in the presence of multi-collinearity, other problems with GVS include the general perils of executing MCMC-based inference. The most pressing of these is the potential for never reaching convergence in a practical run length, and the absence of any absolute assurance that a chain has reached convergence. The quality of the inference made in this situation would be suspect, and the user might not be fully aware that this is the case. Obviously the use of diagnostic statistics is somewhat of a safeguard against this, but all MCMC runs are finite and if a chain does not satisfactorily pass a battery of tests, or only partially passes the operator still has to make a subjective decision as to what they should do with the posteriors in front of them. Furthermore, the tests themselves are only sample statistics, with their

own distributional properties and there is no way to ensure that passing a test is both necessary and sufficient for establishing convergence.

Furthermore, assuming one is comfortable with the implementation of MCMC, there is always the general concern of lack of robustness of the prior distributions or the risks associated with their mis-specification. One could argue that there is only one true prior, reflecting the subjective opinion(s) of the user, and there does not need to be any apology after the fact if a posterior distribution doesn't behave as expected across a wide range of values of the parameter in question. This must be weighed against the desire for wide-spread (and often unsupervised) application of the method, and the need for some robustness in cases where the user needs to be protected from some level of naiveté.

Ideally one would like for the data to play a maximal role in the inference process (conditioned on the parameters of interest), while the user-supplied prior play a minor role in the inference. This is to say that one would hope to get nearly the same answer for a given dataset, if the user had selected a different prior than the one they had supplied. A study was conducted to determine the robustness properties of M-GVS, and determine the inherent properties of the method.

Using the (16x6) design matrix below (as a numerically well-behaved surrogate for the gas library)

```
-1    -1     1    -1     1     1
 1    -1    -1    -1    -1     1
-1     1    -1    -1     1    -1
 1     1     1    -1    -1    -1
-1    -1     1     1    -1    -1
 1    -1    -1     1     1    -1
-1     1    -1     1    -1     1
 1     1     1     1     1     1
-1    -1     1    -1     1     1
 1    -1    -1    -1    -1     1
-1     1    -1    -1     1    -1
 1     1     1    -1    -1    -1
-1    -1     1     1    -1    -1
 1    -1    -1     1     1    -1
```

```
-1      1      -1      1      -1      1
 1      1       1      1       1      1
```

we create data using the recipe:

$$Y = X\beta + \epsilon \tag{3.13}$$

where

$$\epsilon \sim N(0,1).$$

The value chosen for β_i is a specified value for SNR, all the others were zero. So if we wanted Y with an SNR of 10, with the 3rd gas, the β would be $[0,0,10,0\dots]$. The gases that are truly in the model are denoted by γ, while the out-of-model gases are denoted by $/\gamma$. The indices of the gases that were implanted were randomly selected without replacement for each experiment. We are interested in measuring 3 things:

1. The mean posterior indicators of the in-model gases (to qualitatively see if they approach unity)

2. The mean posterior indicators of the out-of-model gases (to qualitatively see if they approach zero)

3. The metric for judging the overall efficacy of M-GVS will be the ratio of the mean posterior indicators of the in-model gases to the out-of-model gases.

$$Q = \frac{E[Pr(\gamma|data)]}{E[Pr(/\gamma|data)]} \tag{3.14}$$

The above scheme was tested at SNRs in the range [0, 1, 5, 10, 15, 25, 50, 150, 1 000, 10 000], 25 times (experiments) at each parameter combination, with each MCMC being run for 2,000 iterations with no thinning and burn in of 200. A range of uniform/fixed prior inclusion probabilities of was tested [.01, .1, .3, .5]. The order of combinations in Figures 3.4, 3.7, 3.10, 3.13, 3.16, and 3.19 below utilizes a different line for each $Pr(\gamma)$ (the highest line is always the smallest value of inclusion probability), and the ordering of the horizontal axis is left-to-right increasing SNR.

For all other plots the ordering is (from left to right on the horizontal axis) is the ordered pairs of:
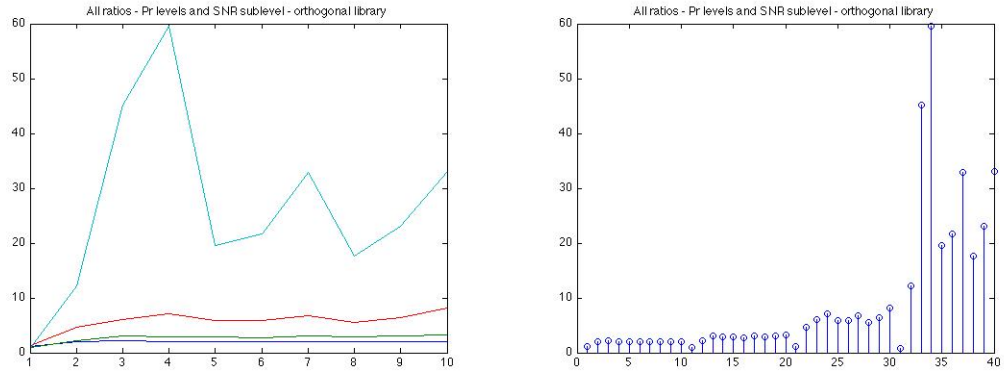
Figure 3.4: Q as a function of SNR and $\Pr(\gamma)$ - single gas, orthogonal library

1. $\Pr(\gamma) = .5$, SNR $= 0$

2. $\Pr(\gamma) = .5$, SNR $= 1$

3. $\Pr(\gamma) = .5$, SNR $= 5$

4. $\Pr(\gamma) = .5$, SNR $= 10$

5. $\Pr(\gamma) = .5$, SNR $= 15$

6. $\Pr(\gamma) = .5$, SNR $= 25$

7. $\Pr(\gamma) = .5$, SNR $= 50$

8. $\Pr(\gamma) = .5$, SNR $= 150$

9. $\Pr(\gamma) = .5$, SNR $= 1{,}000$

10. $\Pr(\gamma) = .5$, SNR $= 10{,}000$

11. $\Pr(\gamma) = .3$, SNR $= 0$

12. $\Pr(\gamma) = .3$, SNR $= 1$
    $\vdots$
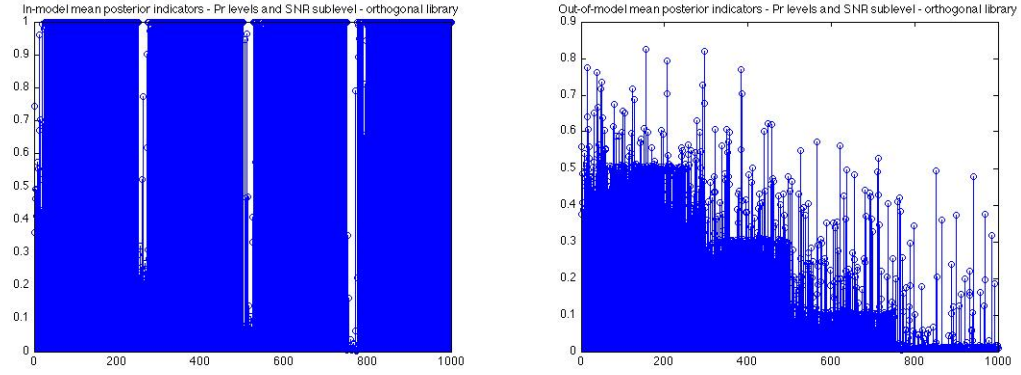
Figure 3.5: In model posterior indicators (left) and out of model posterior indicators (right) as a function of SNR and $\Pr(\gamma)$ - single gas, orthogonal library
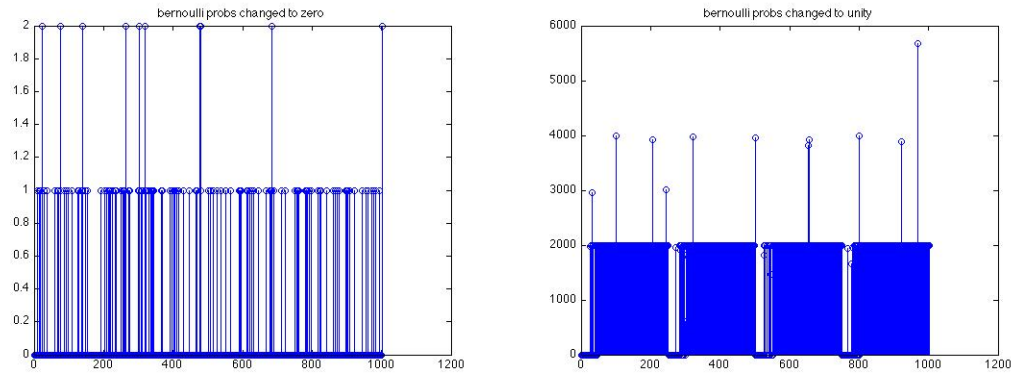
Figure 3.6: Count of times a Bernoulli probability was set to zero or one - single gas, orthogonal library
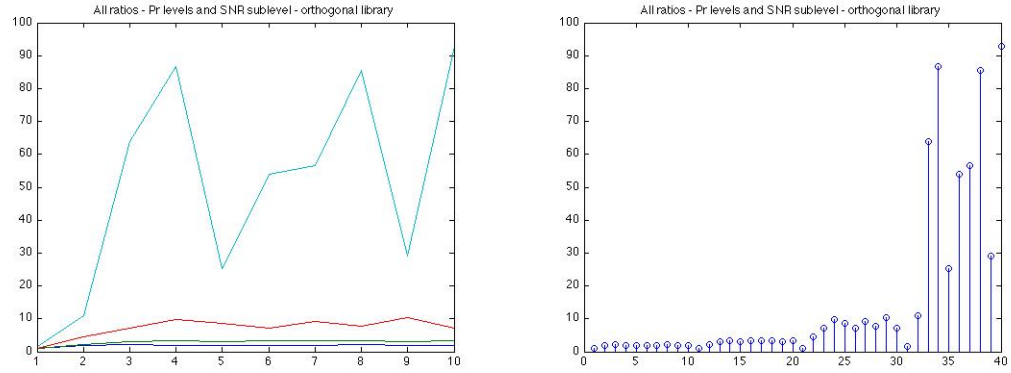
Figure 3.7: Q as a function of SNR and $\Pr(\gamma)$ - three gases, orthogonal library
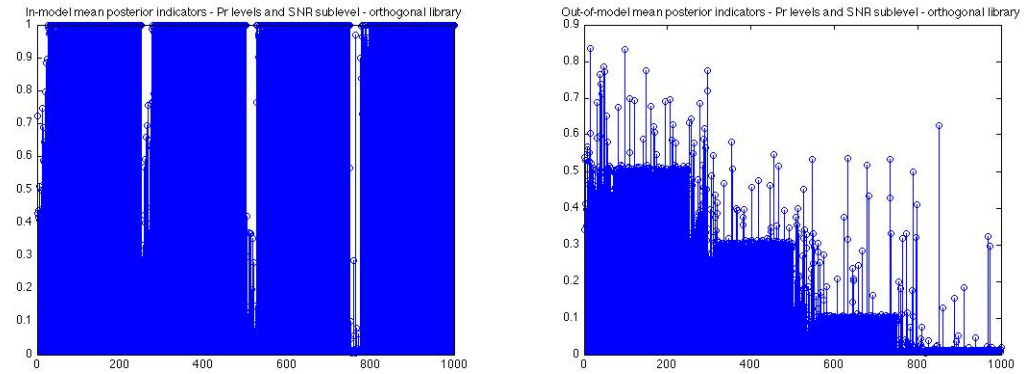
Figure 3.8: In model posterior indicators (left) and out of model posterior indicators (right) as a function of SNR and $\Pr(\gamma)$ - three gases, orthogonal library
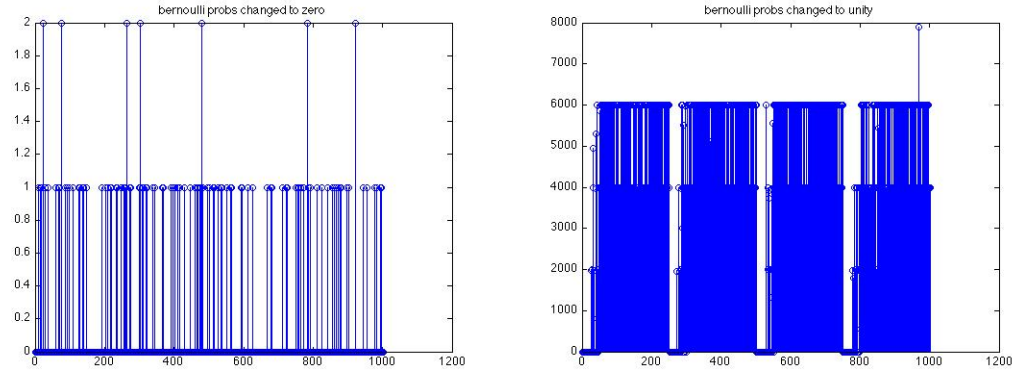
Figure 3.9: Count of times a Bernoulli probability was set to zero or one - three gases, orthogonal library
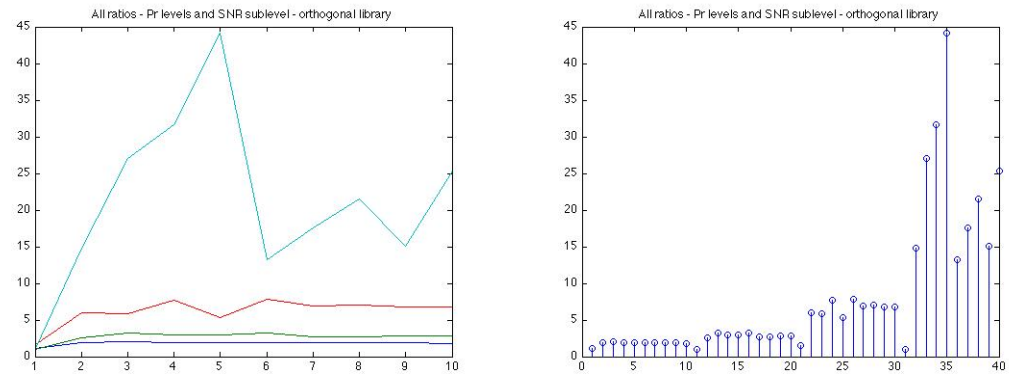
Figure 3.10: Q as a function of SNR and $\Pr(\gamma)$ - single gas, orthogonal library, standardized regression
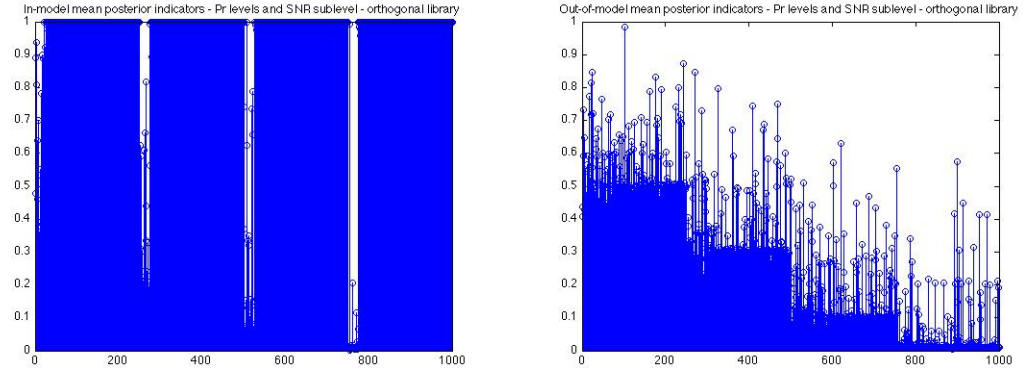
Figure 3.11: In model posterior indicators (left) and out of model posterior indicators (right) as a function of SNR and $Pr(\gamma)$ - single gas, orthogonal library, standardized regression
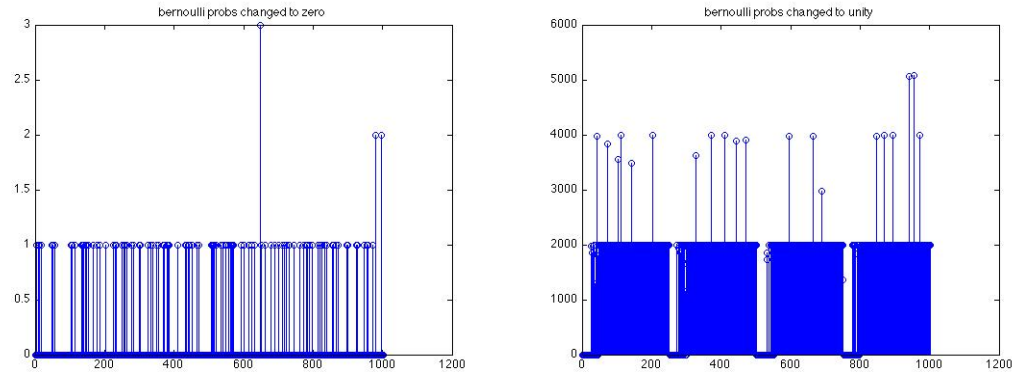
Figure 3.12: Count of times a Bernoulli probability was set to zero or one - single gas, orthogonal library, standardized regression
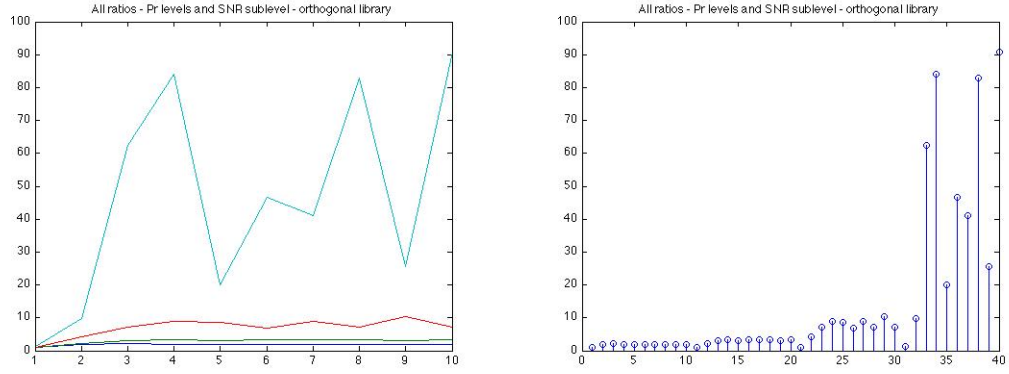
Figure 3.13: Q as a function of SNR and $\Pr(\gamma)$ - three gases, orthogonal library, standardized regression
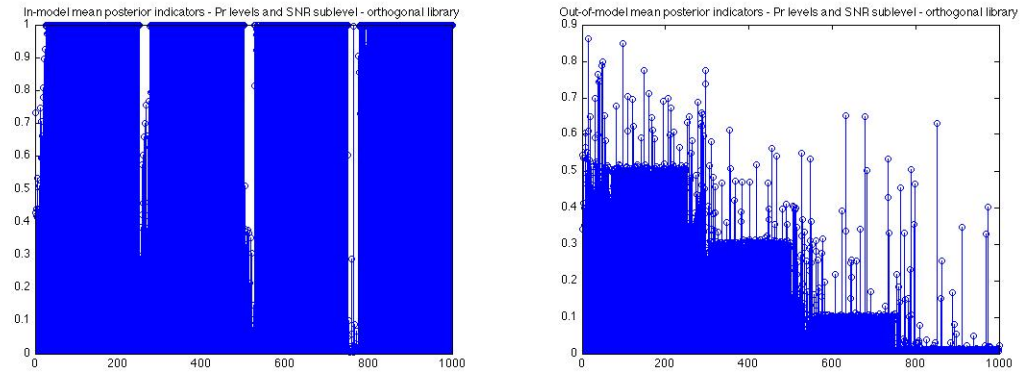
Figure 3.14: In model posterior indicators (left) and out of model posterior indicators (right) as a function of SNR and $\Pr(\gamma)$ - three gases, orthogonal library, standardized regression
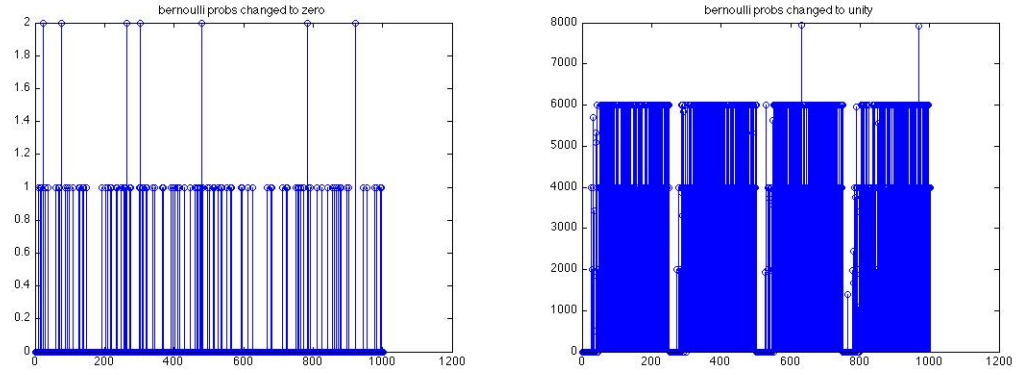
Figure 3.15: Count of times a Bernoulli probability was set to zero or one - three gases, orthogonal library, standardized regression
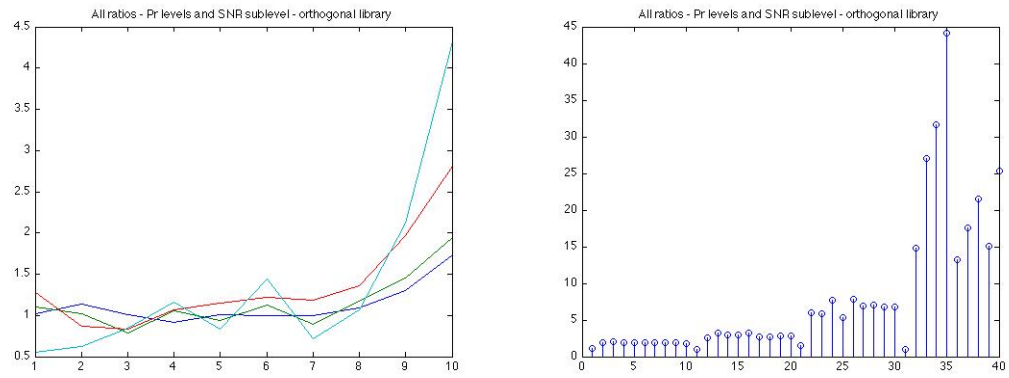
Figure 3.16: Q as a function of SNR and $\Pr(\gamma)$ - single gas, real gas library, standardized regression
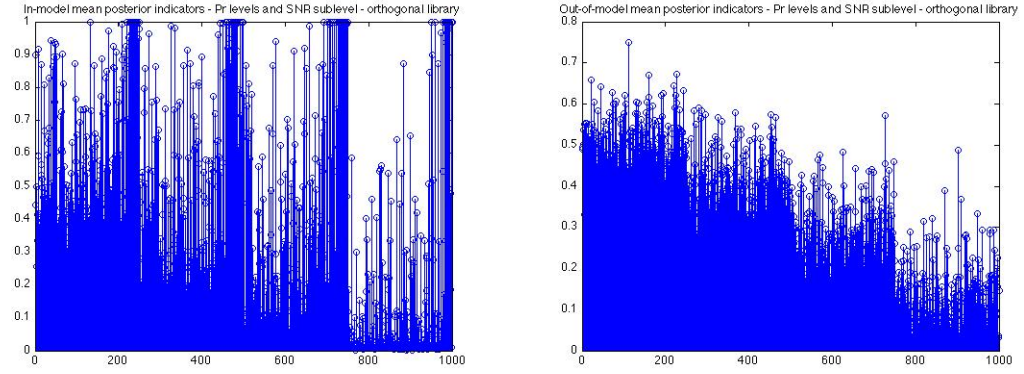
Figure 3.17: In model posterior indicators (left) and out of model posterior indicators (right) as a function of SNR and $\Pr(\gamma)$ - single gas, real gas library, standardized regression
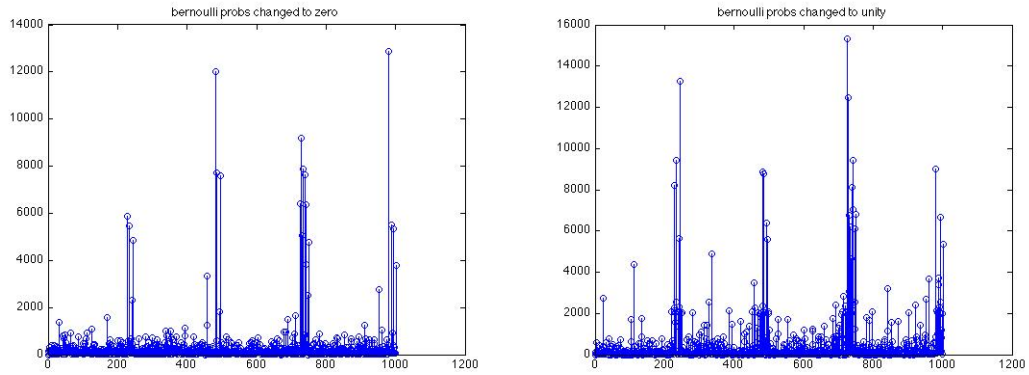
Figure 3.18: Count of times a Bernoulli probability was set to zero or one - single gas, real gas library, standardized regression
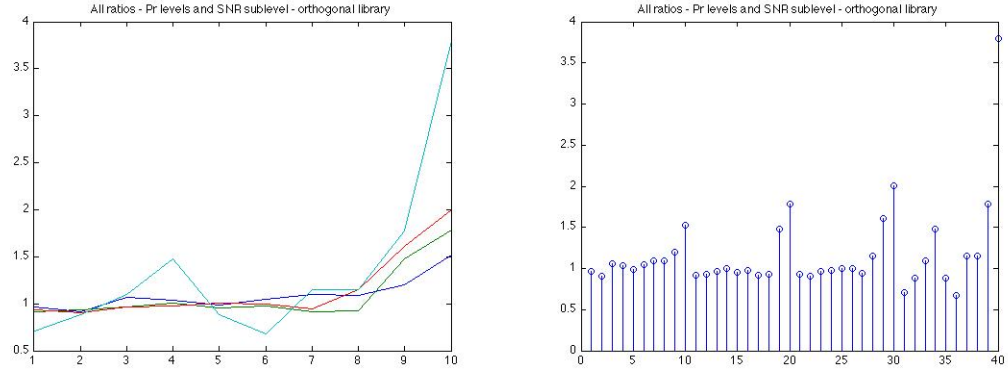
Figure 3.19: Q as a function of SNR and $\Pr(\gamma)$ - three gases, real gas library, standardized regression
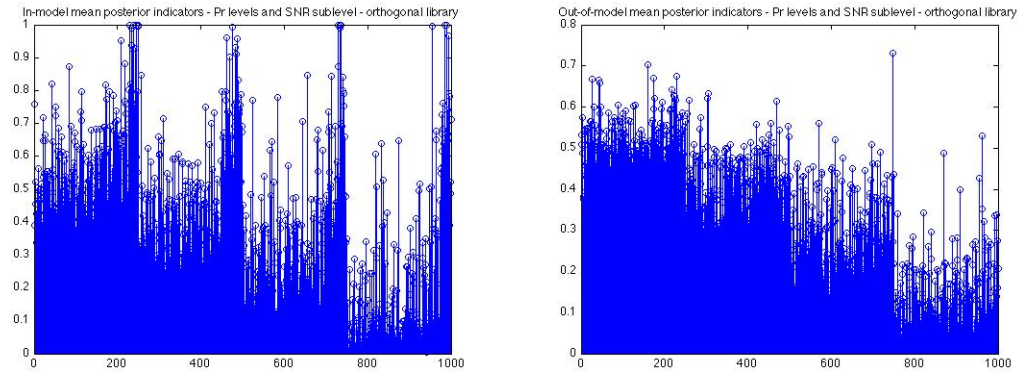
Figure 3.20: In model posterior indicators (left) and out of model posterior indicators (right) as a function of SNR and $\Pr(\gamma)$ - three gases, real gas library, standardized regression
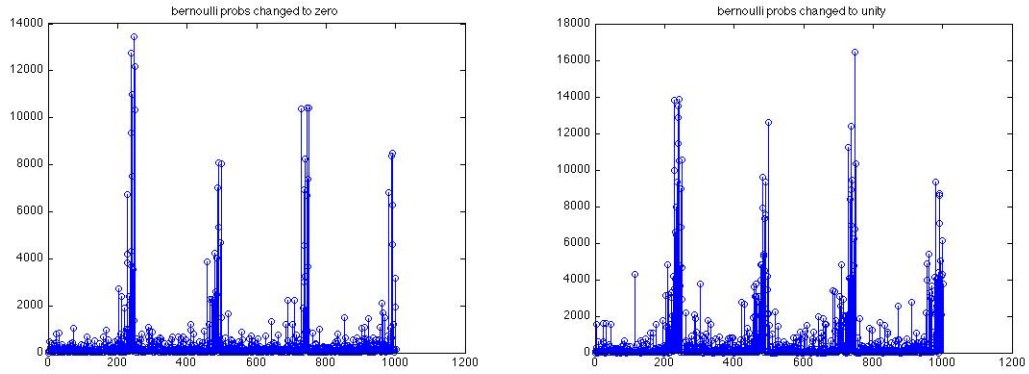
Figure 3.21: Count of times a Bernoulli probability was set to zero or one - three gases, real gas library, standardized regression

for each experiment

Generally M-GVS produces a Q metric that is generally monotone increasing with increasing SNR within a degree of variability that governed by the user-specified $Pr(\gamma)$. That is to say the pale blue lines (corresponding to $Pr(\gamma) = 0.01$ in Figures 3.4, 3.7, 3.10, and 3.13 have a much higher Q, but are also much more variable. There is also the clear trend that the Q metric tends to be $1/Pr(\gamma)$. This is concurrent with the observation that the out-of-model posterior inclusion probabilities tend to simply return the user-specified prior inclusion probabilities.

This leads to the trend that when Q is used as a metric for success, small values for $Pr(\gamma)$ tend to appear to do better, not because the numerator showed any greater "success" in identifying the correct covariate, but rather because of the non-robustness of the denominator simply returning the prior. That is not to say that this phenomenon can not be usefully leveraged in an application, but it must be made clear to the user that while M-GVS is robust for the in-model posterior inclusion probabilities (M-GVS reliably returns unity at any appreciable SNR) there is an inherent non-robustness in the out-of-model inclusion probabilities. Also, the standardized regression results are consistent with the unstandardized results for the orthogonal matrix.

The "notches" that are most noticeable in the pale blue lines in Figures 3.4, 3.7, 3.10, and 3.13 is a result of having the columns of X all of the same magnitude and orthogonal. This causes the effect where any one gas being dropped out causes a

dramatic shift in the evaluated values of the likelihoods (terms A,B,D,E from before). This is because as the SNR goes up, the likelihoods become more sharply peaked, and the impact of dropping out just one column of X in any of terms A,B,C, and D (which is what happens at every sub-iteration) causes a discontinuity in the metric Q. This phenomena is consistent for the orthogonal library, and its impact is less at lower values of $Pr(\gamma)$. The notching behavior also appears to be be consistent across combinations of SNR and within levels of $Pr(\gamma)$.

When the experiment was repeated for the real gas library behaves similarly to the orthogonal one in the sense that lower $Pr(\gamma)$ values tend to perform better than higher ones. The real library also behaves in a monotone-increasing manner (as judged by Q) with increasing SNR. Finally the lower prior probabilities tend to produce more variable estimates of Q. It is important to point out that the reason the gas library requires such large SNRs (relative to the orthogonal matrix) is because it contains very small values, as the library-wide average 1.5×10^{-4}.

It is unclear what causes the lack of clearly obvious "notches" in Figures 3.19 while they seem to be evident in Figure 3.16 (its arguable if this is in the noise though). Qualitatively, this could be due to the greater degree of "mixing" going on with 3 gases instead of one, averaging out the impact of the difference in magnitude of the columns of the library. Additionally, this might account for Figures 3.18 and 3.21 behaving less like step functions, and more like random draws from a variety of different mixtures of likelihoods.

The conclusion of this experiment, is that M-GVS does have robustness issues for out-of-model posterior inclusion probabilities that a user must be aware of. Generally it motivates the use of lower prior inclusion probabilities (rather than higher ones), not because it improves the ability of the algorithm to identify the correct model, but rather because it tends to reduce the posterior inclusion probabilities of the out-of-model gases. Furthermore, it validates the behavior of standardized regression with M-GVS, and varying levels of model complexity (one and 3 gases).

Another experiment was conducted to determine how low it was feasible to set the prior inclusion probabilities to take advantage of their known non-robustness for out-of-model gases. The experiment was repeated for both the orthogonal library and the real gas library, over a SNR range of [0, 1, 10, 150, 1000, 10 000, 100 000, 1 000 000], with a prior inclusion probability range of [0.000001, 0.00001, 0.0001, 0.0001, 0.001, 0.01,

Table 3.4: Q metric for orthogonal matrix as a function of SNR and prior inclusion probability

	SNR			
$Pr(\gamma)$	0	1	10	150
0.3	1.107	2.126	2.423	2.239
0.2	1.038	3.946	3.838	4.889
0.1	1.740	7.906	10.053	10.872
0.01	3.670	14.806	57.536	109.802
0.001	4.948	41.841	92.323	104.081
0.0001	0.004	35.316	18.778	Inf
0.00001	$6.2e11$	$2.9e14$	2876.800	16.162
0.000001	0.000	13462.000	Inf	Inf

	SNR			
$Pr(\gamma)$	1,000	10,000	100,000	1,000,000
0.3	3.646	3.214	3.384	3.300
0.2	3.729	3.935	5.436	5.169
0.1	9.673	4.205	4.824	8.185
0.01	98.521	15.727	67.215	68.823
0.001	14.384	81.174	103.482	62.268
0.0001	282.039	4230.588	7991.111	42.937
0.00001	1057.647	Inf	Inf	Inf
0.000001	Inf	Inf	Inf	Inf

0.1, 0.2, 0.3], and 2 gases inserted. The results of this second experiment are shown in Tables 3.4 and 3.5. While there is clearly latitude for choosing different values of $Pr(\gamma)$, particularly if one knew the SNR that they were going to be operating in, the author chose to use $Pr(\gamma) = 0.001$ for a fixed prior inclusion probability. Other values were considered in the range 0.0001 to 0.001, but anecdotal testing showed little difference between the two.

Informed Inference with M-GVS

The last property of M-GVS that it is of interest to validate, particularly in light of the very low prior probability of inclusion suggested above, is the ability to use M-GVS in an informed manner. Assuming an informed user may designate certain gases as

Table 3.5: Q metric for real gas library as a function of SNR and prior inclusion probability

$Pr(\gamma)$	SNR			
	0	1	10	150
0.3	1.136	1.136	0.982	1.105
0.2	0.717	0.972	1.473	1.348
0.1	1.017	0.629	0.670	2.210
0.01	2.481	0.905	0.949	0.615
0.001	1.563	1.087	2.709	0.624
0.0001	1.542	4.429	0.349	6.168
0.00001	3.263	0.795	3.781	1.505
0.000001	0.510	0.175	2.771	0.140

$Pr(\gamma)$	SNR			
	1,000	10,000	100,000	1,000,000
0.3	1.177	1.891	2.315	1.825
0.2	1.150	1.979	1.601	2.252
0.1	0.915	2.618	2.591	2.100
0.01	2.414	3.386	4.213	4.086
0.001	2.271	7.407	3.505	3.977
0.0001	1.424	10.105	4.960	2.461
0.00001	0.767	2.735	4.945	9.125
0.000001	5.367	9.336	3.466	3.233

a priori more likely than others in a scene, it is instructive to measure outcomes of the three cases where the user is: correctly informed, wrongly informed, or uniformed about the constituents in the plume.

We will assume that an uninformed user will supply a fixed uniform prior inclusion probability for all gases, while a correctly informed user will supply a higher probability for the gases that are truly in the image and/or a lower one for gases that are not in an image. Finally, we define a wrongly-informed user as one that supplies the converse, and for the sake of simplicity, we will not consider mixed cases.

Success in this set of experiments will be measured in the same way, via the Q metric, which is useful for validation studies where the true model is known, and this tool can be applied to the different cases of performance measurement in informed inference. An experiment was run where a fixed prior inclusion probability of $Pr(\gamma) = 0.001$ was compared with an informed user supplying $Pr(\gamma)_\gamma = 0.01$ for the in-model gases and $Pr(\gamma)_{/\gamma} = 0.0001$ for the out-of-model gases, and a wrongly informed user supplying the converse. The results of which are shown in Tables 3.6 and 3.7.

This informed user experiment shows a number of important trends. First, is that it demonstrates the basic phenomenology of an informed user providing correct information about the inclusion probabilities, and there being an improvement in identification performance. These data also show that a correctly informed user will be able to detect a gas at a lower SNR than an un-informed (or wrongly informed) user will. Finally, this experiment speaks to the issue of robustness, and shows that even a user who is wrong, at some (higher) SNR, the posteriors will reflect the true model and not the user's prior belief. Its important to note that this example of robustness happens over only after an order of magnitude or more increase in SNR, so this is not a fail-safe example of robustness.

It has been shown that M-GVS is not "robust" in the conventional statistical sense, where one would hope that the data would dominate the inference process, and the user-specified prior would have minimal role, particularly in the cases where the user is wrong. This is not to say that M-GVS can not be used in an informative manner, just that there are risks to miss-specification of the priors, and generally one should have good reason before arbitrarily supplying a non-uniform prior. These caveats being what they are, there is clearly a practical benefit from supplying the correct prior, and

Table 3.6: Q metric for orthogonal library for three notional users

User	SNR					
	0	1	5	10	15	25
Informed	1.92	38.47	63.50	4878.64	1581.54	555.14
Non-informed	0.26	17.04	28.20	56.26	72.81	32.15
Wrong	0.21	7.50	23.48	24.24	34.31	12.66

User	SNR				
	50	150	$1,000$	$10,000$	$100,000$
Informed	22.65	1881.31	7779.46	4202.04	7380.51
Non-informed	34.54	187.09	619.68	295.07	95.08
Wrong	28.85	28.87	70.70	56.40	18.52

Table 3.7: Q metric for real gas library for three notional users

User	SNR					
	0	1	5	10	15	25
Informed	2.11	2.89	2.53	2.56	5.45	2.52
Non-informed	0.75	1.23	0.60	1.61	1.22	0.38
Wrong	0.21	0.34	0.14	0.51	0.50	0.35

User	50	150	$1,000$	$10,000$	$100,000$
Informed	2.83	4.48	4.50	4.84	3.79
Non-informed	0.67	1.75	2.71	4.55	2.70
Wrong	0.30	0.52	1.20	2.95	2.40

this motivates the cautious application of an informed prior structure for a spatial aggregation scheme that will be introduced shortly.

Clearly there is additional room for optimizing this prior inclusion probability tradespace, as the optimal choice of this parameter is going to be SNR, gas library, and spectral resolution-specific, but for the purposes of this thesis and the following work, the author used the uninformed value of $Pr(\gamma)^* = 0.001$ unless otherwise specified. This choice is driven by the above trade studies, where M-GVS generally achieves a workable level of separability for most SNRs.

The application which the author has made use of an informative prior structure is in the application of M-GVS to multiple-pixel plumes. These are instances where one exploits the spatially contiguous nature of real plumes and "chains together" inference in a spatial manner, allowing the results from neighboring pixels influence the inference step on subsequent pixels. This application, and the details of its development, are the subject of a future section.

MCMC Convergence Diagnostics

Once a MCMC run has iterated sufficiently, which depending on the problem could in practice take anywhere from 10 to 10^{10} iterations, one can apply a variety of diagnostic techniques to the chain. The issue of how many iterations to collect (assuming the chain has converged) is only a function of the quality with which a finite set of values can be used to approximate the underlying pdfs/pmfs. 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.

Generally one applies a number of tests to the chain to help ensure convergence to the target distribution. The first is simply running several chains, with vastly different starting values simultaneously, and checking to see if they converged to (nearly) the same values. This guards against the possibility of the chain getting "stuck" in an isolated pocket of probability. Next, most MCMC algorithms are most effective when each successive draw from the posterior is more or less independent from the previous one. This is equivalent to saying that the parameter space is most easily traversed when

there is relatively low serial autocorrelation between iterations. This can be practically achieved in two ways, first by ensuring that the columns of the design matrix are orthogonal and then by thinning the chain. Thinning the chain is simply retaining only every 10th or 20th value which has a substantial computational burden because all 20 of the iterations need to be computed to generate a single output value. However it provides for better quality inference and helps ensure chain convergence. Often the user will decide what thinning interval to use based a visual examination of the individual chain autocorrelation functions.

Here autocorrelation functions are defined as the average correlation between data values x that are $h = 0, 1, 2 \ldots M$ steps of size δ separated. It is canonically estimated using

$$ACF(h) = \frac{1}{Var(x - \mu)} E[(x_0 - \mu)(x_{0+h\delta} - \mu)]. \tag{3.15}$$

In practice the use of the Fast Fourier Transform (FFT, and its inverse transform IFFT) is used as it is substantially faster than computing the expectation in Equation 3.15. For any function $f(x)$, whose Fourier transform pair is $F(\zeta)$, the sample autocorrelation function (ACF) is estimated by

$$ACF(h) = IFFT \left[\left\| \frac{F(\zeta) \cdot F(\zeta)}{||F(\zeta)||_2 ||F(\zeta)||_2} \right\| \right]. \tag{3.16}$$

Readers should note that there is an important difference in the statistics use of the term ACF (as defined in Equations 3.15 and 3.16) and the electrical engineering (EE) use of the term (they would call it the preceding equations the "normalized ACF"). Furthermore, what the EE community calls the ACF, the statistics community calls the "auto-covariance function." The author intends the meaning implied by the equations above for all in all references to an ACF, where the maximal value is unity and the value of the ACF evaluated at $h = 0$ is unity.

A variety of statistical tests exist for testing whether or not the chain has converged [22]. Popular tests include the Gelman-Rubin statistic which indicates the potential for improving inference if iteration is continued. The Heidelberger-Welch statistic tests for both stationarity of the MC and its convergence, the latter of which compares early parts of the MC to later parts (if the chain has converged they should be identical). Further, the Geweke convergence statistic can also be used to detect differences in the

early versus late parts of a MC. Additionally one can test for cross-correlation between chains qualitatively using plots or quantitatively using a multivariate version of the Gelman-Rubin statistic.

Due to the fact that all of the data used in testing M-GVS is semi-synthetic and the true gas mixture being inserted is known in all cases, only single chain inference was used throughout this work as a means to limit the resources required by the problem. Because under M-GVS the posterior distribution of the regression coefficients is known before the MCMC begins, one can sample from it directly. This means that there is no need to wait for (or monitor) its convergence, and thus attention can be focused on monitoring the convergence of the indicator variables. If this method were put to use in a real scenario, one would ideally want to run several parallel chains per pixel and implement additional chain-to-chain diagnostics.

The two main challenges in monitoring the indicator variables is that many of the common diagnostic measures described above were generally designed for continuous random variables, and the large number of iterates, chains, and pixels to be monitored for even a single image. For example, the original AHI image that this work is based on has roughly 125,000 pixels in it, and it is just one image among dozens collected on that day, among dozens of days of collection, for a relatively small field campaign. This motivates the development of a somewhat custom monitoring scheme that places an emphasis on sensitivity, potentially at the expense of specificity, and can provide the user with macro-scale feedback on how well the chains are performing.

For this thinned library, there are 13 gases (columns of $\boldsymbol{X}$), one would generally be interested in monitoring each chain for three things: "drift" in the fraction of times a gas is included, an increase in volatility of the MC for each indicator, and finally the parameter to parameter correlations should stabilize. Each of these properties is monitored in a retrospective fashion, for cumulative fractions (in units of 10% of the chain length) at a time. So each statistic is computed first on the last 10% of the chain, then on the last 20% of the chain, and so forth. The statistic that is reported is the fraction of the chain, that passes the tests, which is the first cumulative window that fails the battery of tests at a specified failure rate. A notional depiction of this process is shown in Figure 3.22.

This is somewhat unorthodox in practice, but it is intended to answer a slightly different question than traditional MC monitoring, which is only focused on answering

the "are we there yet" question of convergence and is nearly always done with an operator's intervention in every single chain. In imaging applications, there is almost always going to be a constrained computation budget and the need for minimal operator interaction (unlike many statistics problems where an operator inspects each chain manually), an operator must decide in advance of the analysis how much computational resources will be made available for a problem, and due to the volume of data to be processed, there little opportunity to go back and further process an individual chain (unless it is of particularly high importance to the user). Also the operator will be keenly interested in knowing by how much they could reduce their computation load for a given data set if it seems to be converging quickly. This method of cumulative forward-sliding windows also has the potential for using the convergence diagnostic to adaptively determine the burn-in to be discarded, as well as deciding on the fly if iteration can be stopped ahead of schedule. These are topics the author has not fully explored, but the potential for their development is clear.

The Geweke convergence diagnostic [23] is useful for testing for a drift in the mean of each individual MC. It is usually implemented by comparing the first 10% of a chain to the last 50% of a window, by standardizing each of them by the time-series estimate of standard deviation from the auto-covariance function (ACVF) at lag zero ($h = 0$) as defined by

$$ACVF(h) = IFFT\left[\left|\left|F(\zeta) \cdot F(\zeta)\right|\right|\right].$$

(3.17)

If the chain is a mean zero, second-order stationary time-series, the sub-window means will be approximately $\sim N(0, Var(x - \mu))$ and one can compute a one-way ANOVA procedure to test for the difference in the mean at the beginning and end of each window. Its important to note that the ACVF estimate of σ is important (instead of the standard formula for standard deviation) because of the clear biasing effect of serial correlation between successive values in the chain. In the author's implementation, the number of chains that fail the one-way ANOVA ($p = 0.05$) test between the first 10% and last 50% of the cumulative forward-sliding window is recorded.

Measuring the chain volatility is a challenge due to the binary nature of the inclusion posteriors, which motivates taking the second derivative of the chains (approximated through adjacent second differencing), thus increasing the dynamic range of the statistic from 0 to unity to -2 to +2. The L2 norm of the second derivatives is taken as a summary
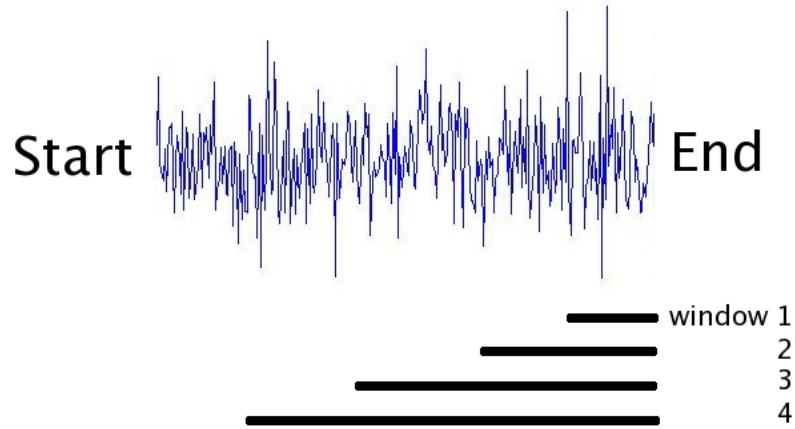
Figure 3.22: Process of cumulative forward-sliding windows being computed

method of measuring the magnitude of the acceleration change in second derivatives. This statistic is computed for non-overlapping sliding windows over the quantiles of the MCs, and a one-way ANOVA test is computed (with p = 0.05) to test if there is a difference between the first third of the windows and the last third of the windows. The number of test failures in each cumulative forward-sliding window is recorded.

Finally, the author developed a convergence monitoring scheme for the correlation between the posteriors of the indicator variables. For each cumulative forward-sliding window the correlation matrix is computed between the indicator variables (with machine epsilon times a $N(0,1)$ random variable added to any column that is all zeros or all ones). Then for the upper-triangular portion of the ten resulting correlation matrices, a second-order polynomial is fit to the correlation between every pair-wise combination of gases in the stack of correlation matrices. Then a 95% confidence interval is created for the slope and quadratic terms (ignoring the intercept term) and the fraction of those confidence intervals that does not contain zero is recorded. It is important to note that this test is designed to allow for a non-zero correlation between the posterior inclusion probabilities to exist for any pair-wise combination of gases (ie. an intercept term in the diagnostic regression) but would hopefully catch a change in those correlations between the cumulative forward-sliding quantiles of the MCs.

An overall number of expected errors is computed by multiplying 0.05 times the

number of test statistics recorded. Note that the correlation-stability diagnostic is measured on the entire stack at once (even though the correlation matrices are computed per cumulative window), while the Geweke and second-derivative test are computed window by window. The number of expected (or "allowable") errors first has the number of counts for the correlation test subtracted, and then in a similar forward-sliding window fashion, the cumulative number of Geweke and second-derivative test failure statistics are subtracted and when that number goes below zero, the corresponding quantile is reported to the user. This creates the possibility that the correlation test alone could cause the entire ensemble of tests to fail, even if the Geweke and L2 diagnostics do not indicate a failure to converge. The percentage point reported to the user notionally indicates the fraction of the chain, after which point it has likely converged, but the chain values before that point likely have not. Because this metric in some sense indicates the degree to which the MCMC process could be accelerated (through reduced iteration number) the author terms the output metric an "acceleration factor."

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 $\boldsymbol{X}$ with

$$\boldsymbol{X}^* = \frac{1}{N-1}\frac{\boldsymbol{X}-\bar{\boldsymbol{X}}}{s_{\boldsymbol{X}}}$$

and $\boldsymbol{Y}$ with

$$\boldsymbol{Y}^* = \frac{1}{N-1}\frac{\boldsymbol{Y}-\bar{\boldsymbol{Y}}}{s_{\boldsymbol{Y}}}$$

in the regression process. Note that the denominators $s_{\boldsymbol{X}}$ and $s_{\boldsymbol{Y}}$ are vectors, which is an element-wise division, producing $\boldsymbol{X}^*$ with the same dimensions as $\boldsymbol{X}$ and $\boldsymbol{Y}^*$ with the same dimensions as $\boldsymbol{Y}$. Then the expectations of the posterior standardized regression coefficients $E[Pr(\boldsymbol{\beta}^*|data)]$ are transformed back into their original units via

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

and the intercept is returned from

$$E[Pr(\boldsymbol{\beta}_0|data)] = \bar{\boldsymbol{Y}} - \bar{\boldsymbol{X}}' E[Pr(\boldsymbol{\beta}|data)].$$

Standardized regression was used for all M-GVS and SAM-GVS steps and unstandardized regression was used only for the convergence diagnostics.

Spatially Aggregated M-GVS (SAM-GVS)

Recall that the Bayesian paradigm allows for the user to provide a subjective opinion about the value of a parameter (in our case, perhaps about the probability of the existence of a particular gas in a scene), and most of the discussion until this point has proceeded with the intent that M-GVS proceed without the benefit of an informed user nor the benefit of the spatial structure inherent in most practical images. Given that gaseous effluent plumes are likely to be spatially contiguous (regardless of their shape), this motivates the development of a method that can leverage this behavior to increase the potential for identifying the gases in a plume. The natural choice for "pooling" inference spatially is to somehow use the posterior distribution from adjacent pixels to influence the prior distribution used for the M-GVS procedure.

These interests are what led to the development of what the author terms spatially aggregated M-GVS (SAM-GVS), which is the set of logic developed to flexibly pool the posteriors from adjacent pixels and allow them to be used to inform the regression step at the next pixel. The desired behavior of SAM-GVS is that it should

- Roughly maintain the average probability between the E(aggregate prior) and the E(optimal uniformative prior)

- Through the process of creating an informative prior, the relative separability should remain approximately constant

- SAM-GVS should never perform worse than the optimal uniformative prior $Pr(\gamma)^*$, which is to say in the absence of any better information (ie. without any separable neighbors), the algorithm should have a way to default back to the optimal uninformative priors (or analyst-supplied informed prior)

- In the process of balancing between informed and un-informed cases, there should be a smooth transition between the naive and informed cases, and furthermore there should not be any saturation effects from an adjacent posterior completely overwhelming a successive prior.

- Primarily motived by the interest in never saturating, but also in order to more closely mimic the behavior of a human operator, non-linear behavior is likely desirable in the SAM-GVS process.

The concept of "separability" as the fundamentally most useful quantity that M-GVS provides, from the perspective of a human analyst. That is to say that the absolute magnitude of the posterior indicator variables is probably less important to an analyst than the relative magnitudes of the posterior indicators. The author recognizes that there are many ways that this separability could be measured, but through experience with these data, the definition of separability that will be used (a scalar quantity for a single pixel) is

$$sep_n = \frac{Pr(\gamma_n|data)}{E_{0-33p}[Pr(\gamma|data)]}. \tag{3.18}$$

In words this is the ratio of the nth posterior indicator value, to the mean of the lowest $1/3$ of the posterior indicators for the same pixel. The metric in Equation 3.19 is well suited to cases where the correct model is known. An alternative metric of separability used later in this manuscript (where one does not necessarily know in advance which gas is the correct one) is

$$sep = \frac{max(Pr(\gamma|data))}{E_{0-33p}[Pr(\gamma|data)]}. \tag{3.19}$$

Starting with the metric for separability in Equation 3.19, one can compute this metric for each of the adjacent neighbors for which M-GVS (or SAM-GVS) has been accomplished. SAM-GVS then proceeds by comparing the maximum value among and within each neighboring pixel's $(max_{spatially}(max_{spectrally}(sep_{neighbors})))$ separabilty, with the intent that choosing to use the most separable pixel is most likely to increase the sensitivity of inference in adjacent pixels. This separability is then input into a piecewise non-linear weighting function, computed for each pixel

$$\phi = \begin{cases} 0 & \text{if } sep_n \leq 1 \\ 0.475 \cdot sep_n - 0.475 & \text{if } 1 < sep_n \leq 2 \\ \frac{1}{1+exp(-sep_n+2)}(.95 \cdot Pr(\gamma)^*) + Pr(\gamma)^* & \text{if } 2 < sep_n. \end{cases} \tag{3.20}$$

This non-linearly weighted mixing fraction ϕ is shown in Figure 3.23 and then used
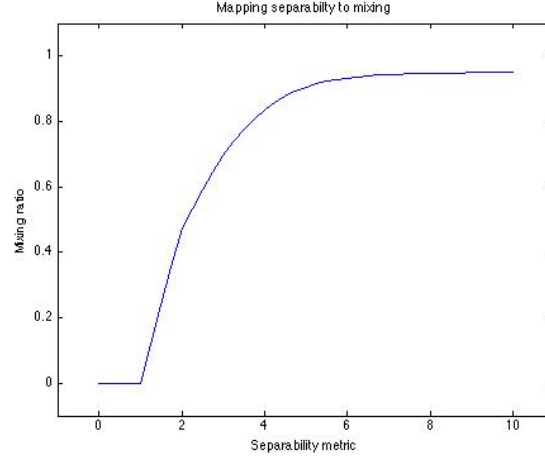
Figure 3.23: Non-linear mixing curve used to map separability (sep_n) to mixing ratio ϕ

to linearly mix a new weighted prior (note the indexing is over wavelength) by

$$\hat{Pr}(\gamma) = (1 - \phi)Pr(\gamma)^* + \phi\left(\frac{Pr(\gamma|data)}{||Pr(\gamma|data)||} \times .95 \times sep \times \pi\right). \qquad (3.21)$$

In this case π is defined as

$$\pi \equiv \frac{1}{1 + exp(-[max(Pr(\gamma|data)^*) \times 7] + 2)}$$

A plot of this curve is included in Figure 3. The effect of this penalty term is that in cases where the most separable adjacent posterior $Pr(\gamma|data)^*$ is also very small, it is penalized in the mixing function by π.
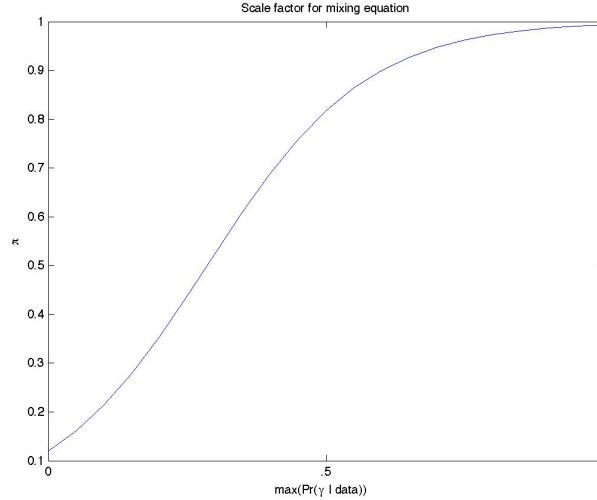
Because there is the potential to create values greater than unity, any $\hat{Pr}(\gamma) > .99$ has every wavelength re-scaled so that its peak magnitude is 0.95.

The mixing function in Equation 3.21 is designed such that when there is low separability (sep) for a particular wavelength, then new prior will tend to default to the user-specified prior, while wavelengths that are highly separable from their neighbors will tend to take on the shape of the most separable posterior (via the unit vector $\frac{Pr(\gamma_c|data)}{||Pr(\gamma|data)||}$), weighted by the separability of that wavelength, of the most separable adjacent posterior.

In use, one steps over the image in a left-to-right and top-to-bottom (or some arbitrary choice of stepping order) and for each pixel SAM-GVS computes the mixed prior $\hat{Pr}(\gamma)$ (assuming that the neighboring pixels have been visited) and uses M-GVS and the user-supplied prior $Pr(\gamma)^*$ if no adjacent pixels have yet been visited. With this scheme, at most, any pixel will have 4 neighbors (potentially) inform its inference. This scenario is shown in Figure 3.24 where for example, when evaluating the pixel with the "x" in it, the shaded blue pixels will have potential to contribute to its inference.

Performance assessment

In problems such as land cover classification, it is common to use a "confusion matrix" to represent the accuracy produced by a supervised classification method. Such a matrix reports the number (or simply relative frequency) with which a material is correctly classified as itself. The reported number is aggregated over the entire image (or at
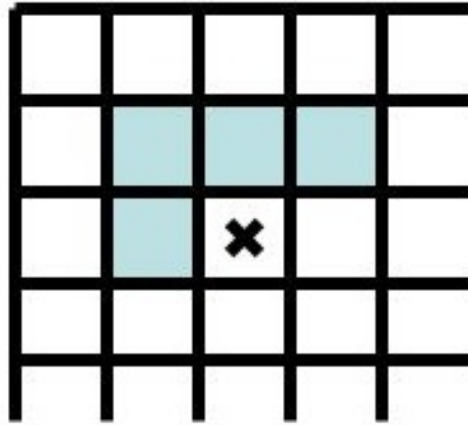
Figure 3.24: Pixel-level depiction of pixels available for spatial aggregation

least the training set) and what is reported is a statistic that is supposed to represent the image-wide performance of the classification scheme. This works well given that in this example, classification is a binary decision (there is no statement that "maybe this pixel belongs in a certain class") but it is not applicable to a scenario where there is only a probability given that a particular gas species is present or not.

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 closest comparison (in a frequentist framework) of stepwise regression to a posterior AUROC is obtained by looking at the observed frequency with which a gas is included in the final fitted model. The authors readily admit that this is not a statistically rigorous comparison owing to the many important differences between fre-

quentist/Bayesian inference (among others), however it provides the end-user with a comparison that can at least be used to inform the analytic process.

Finally, even despite its limitations, the computation of AUROCs is further limited by the sample support available for study. In scenarios where one wishes to look at a spatially non-uniform plume, implanted in an image, particularly in the context of spatial aggregation (SAM-GVS), the best available method for comparing SAM-GVS to M-GVS is to visually compare the probability of inclusion planes for a gas of interest. This is due to the very small sample sizes, and the potential impact for spatial non-uniformity in the image to distort the results.

Chapter 4

Results

This section describes a series of results obtained from semi-sythetic imagery, in an effort to test a variety of different scenarios. The first is an experiment designed to provide some benchmark of M-GVS performance relative to stepwise regression, as a common and broadly understood technique. Second is a test of SAM-GVS on spatially varying plumes, relative to how M-GVS performed on those plumes without any spatial aggregation.

M-GVS Results

The performance evaluation of M-GVS applied to a gas ID problem was conducted using a series of identical sub-sampled, tiled AHI hyperspectral images, which all had varying amounts of a randomly selected gaseous mixture inserted. The gases were inserted identically in each subtile, 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 4.1.

The gas mixtures implanted in the data were at 330K while the average surface temperature for the scene was 309K. M-GVS was run with a $\Pr(\gamma = .0001)$ in order to take advantage of the known non-robustness for out-of-model gases. M-GVS was run for 1000 iterations per pixel, without thinning and the first 100 iterations were discarded. In order to reduce run time, the image was spatially subsampled, with
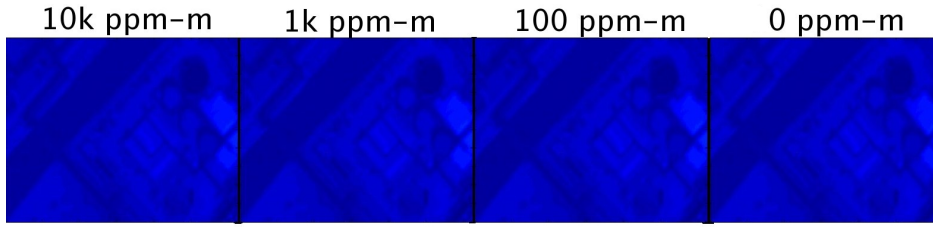
Figure 4.1: Tiled array of images, with varying implanted gas concentrations

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 19 seconds of processing per pixel on a dual-core 2.8 GHz Unix machine.

Once M-GVS was run on every pixel of the tiled images, the posterior probability of the nth gas being present $\Pr(\gamma_n|data)$ 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 $\Pr(\gamma_n|data)$ 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 4.1. As mentioned previously, the best means of comparison available is to process the same tiles using stepwise regression (as per the MATLAB implementation) as a benchmark for performance.

Generally speaking, stepwise regression had much lower average model frequencies than did M-GVS, for gases that were not in the tiles. For the 2 gas and 4 gas mixtures, M-GVS generally showed a greater sensitivity at lower concentrations than stepwise regression did. At the highest concentrations stepwise regression and M-GVS performed comparably for the in-model gases.

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 3.1 or 3.8) because each pixel vector in the semi-synthetic data tile was assembled without error. Thus any variance in the fitted/posterior coefficients is

Table 4.1: M-GVS AUROCS per gas (true constituents are in boldface)

	1 Gas			2 Gases			4 Gases		
ppm-m	100	1,000	10,000	100	1,000	10,000	100	1,000	10,000
acrol	**0.550**	**0.851**	**0.941**	0.592	0.540	0.223	0.460	0.172	0.153
c6h6	0.478	0.402	0.338	0.384	0.136	0.053	**0.570**	**0.828**	**0.918**
dclp12	0.461	0.540	0.656	0.498	0.591	0.642	0.379	0.340	0.410
dclp13	0.416	0.399	0.580	0.575	0.911	0.919	**0.446**	**0.751**	**0.847**
edb	0.431	0.491	0.605	0.434	0.418	0.418	0.506	0.856	0.910
f114	0.467	0.524	0.648	0.510	0.486	0.387	**0.322**	**0.989**	**1.000**
f12	0.345	0.475	0.833	**0.884**	**0.999**	**1.000**	0.320	0.414	0.520
f218	0.262	0.164	0.674	**0.910**	**1.000**	**1.000**	0.317	0.289	0.420
nh3	0.464	0.540	0.621	0.472	0.413	0.496	0.416	0.213	0.139
phg	0.344	0.321	0.388	0.449	0.664	0.737	0.219	0.578	0.646
sf6	0.394	0.441	0.649	0.360	0.448	0.578	0.363	0.395	0.652
so2	0.467	0.381	0.453	0.338	0.745	0.909	**0.306**	**0.612**	**0.903**
tce	0.423	0.509	0.722	0.439	0.635	0.686	0.416	0.508	0.734

Table 4.2: Stepwise regression model frequencies (true constituents are in boldface)

	1 Gas			2 Gases			4 Gases		
ppm-m	100	1,000	10,000	100	1,000	10,000	100	1,000	10,000
acrol	**0.957**	**0.970**	**0.998**	0.936	0.175	0.061	0.955	0.580	0.000
c6h6	0.970	0.968	0.998	0.898	0.514	0.007	**0.977**	**1.000**	**0.993**
dclp12	0.030	0.011	0.000	0.000	0.002	0.000	0.000	0.357	0.600
dclp13	0.000	0.000	0.000	0.000	0.000	0.000	**0.000**	**0.002**	**0.532**
edb	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.027
f114	0.289	0.407	0.557	0.414	0.191	0.416	**0.275**	**0.980**	**0.998**
f12	0.000	0.000	0.020	**0.595**	**0.995**	**1.000**	0.000	0.084	0.818
f218	0.000	0.000	0.000	**0.582**	**0.998**	**1.000**	0.000	0.000	0.000
nh3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.059
phg	0.002	0.002	0.000	0.009	0.000	0.000	0.000	0.000	0.000
sf6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
so2	0.248	0.134	0.000	0.091	0.000	0.443	**0.107**	**0.018**	**0.939**
tce	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Figure 4.2: M-GVS AUROCS (left) and stepwise regression observed model frequencies (right) - 1 gas, index 1
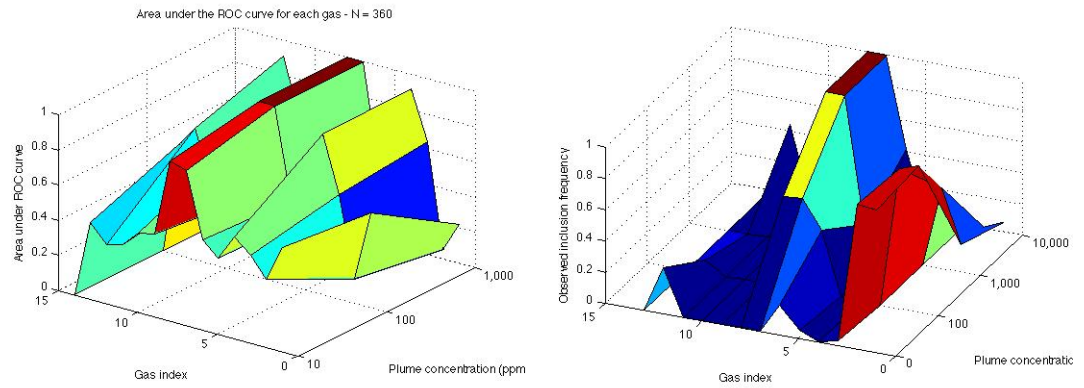
Figure 4.3: M-GVS AUROCS (left) and stepwise regression observed model frequencies (right) - 2 gases, indices 7 and 8
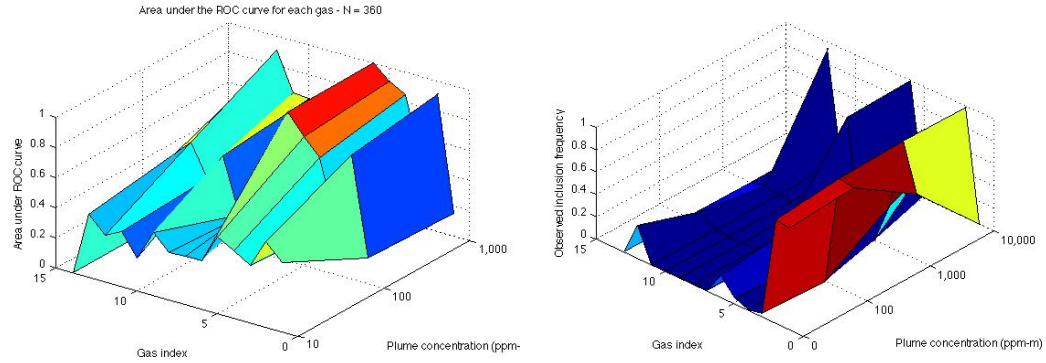
Figure 4.4: M-GVS AUROCS (left) and stepwise regression observed model frequencies (right) - 4 gases, indices 2, 4, 6, and 12

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, M-GVS avoids the false alarm associated with c6h6 that stepwise regression encountered, however at higher concentrations M-GVS produced several large mean posterior indicators (roughly 0.6 - 0.7) for some out-of-model gases. In the two gas example, at low concentrations GVS had a much higher AUROCs than stepwise model frequencies 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 M-GVS at both the highest and lowest concentration levels. For the four gas case M-GVS found dclp13 and so2 at much lower concentrations than did stepwise regression.

Generally speaking, stepwise regression had much lower average model frequencies than did GVS, for gases that were not in the tiles. Stepwise regression was a more specific tool for assessing which gases were in the model, while M-GVS was generally a more sensitive tool and found gases at lower concentrations.

Table 4.3: Spatial plume weighting template

	0.05	
0.10	0.15	0.10
0.10	0.20	0.10
0.15	0.50	0.15
0.10	0.70	0.10
0.20	0.80	0.20
	1.00	
	1.00	

SAM-GVS Results

The data used for the SAM-GVS experiment is generated in the same tiled fashion as in the M-GVS experiments (same gases and maximum concentrations per subtile), with the exception that instead of inserting a uniform layer of gas in every pixel, a spatial concentration template was utilized in order to create a reasonably realistic spatially varying concentration of gas that was implanted in each tile. The actual spatial weights are shown in Figure 4.

Due to resource constraints, the AHI image was subsampled down to 10×10 sub-tiles for this experiment, while still containing a broad sampling of materials throughout the image, to create the subtiles. An image of both the spatial template and a band from the template applied to an actual image is shown in Figure 4.

As mentioned previously, there was not a large enough sample size of each concentration level of pixels to form a ROC curve for the SAM-GVS test, which motivates the use of a more ad-hoc comparison, of simply comparing the mean posterior indicator variables from SAM-GVS to those of M-GVS.

What follows is a series of tables that show the actual posterior inclusion probabilities for the spatially varying plume, for each gas, for each tile. The tables are organized as follows:

- 1gas

    - acrol

        * M-GVS Table 4.4
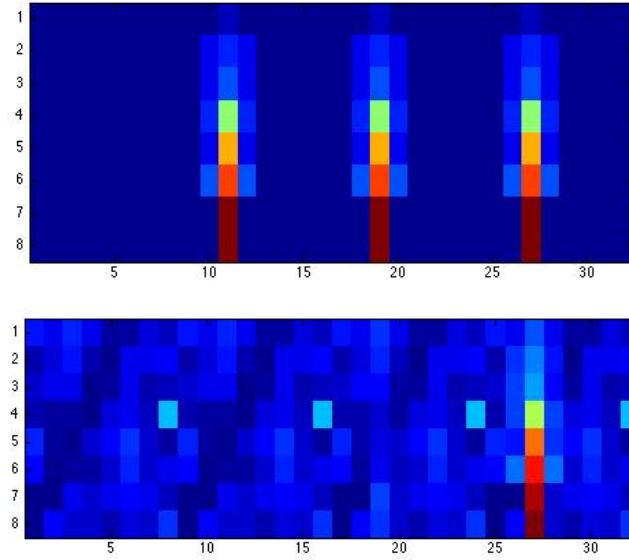
Figure 4.5: Tiled array of spatially-varying implanted images (spatial template - top pane, actual image plane - bottom pane)

                       * SAM-GVS Table 4.5

- 2 gases

  - f12

    * M-GVS Table 4.6
    * SAM-GVS Table 4.7

  - f218

    * M-GVS Table 4.8
    * SAM-GVS Table 4.9

- 4 gases

  - c6h6

    * M-GVS Table 4.10
    * SAM-GVS Table 4.11

- dclp13
    * M-GVS Table 4.12
    * SAM-GVS Table 4.13
- f114
    * M-GVS Table 4.14
    * SAM-GVS Table 4.15
- so2
    * M-GVS Table 4.16
    * SAM-GVS Table 4.17

Generally-speaking Tables 4.4 through 4.17 demonstrate the M-GVS is a relatively sensitive tool for finding spatially-varying plumes in an image, but that its main shortcoming is its lack of specificity. This is a conclusion that echoes the findings of the uniform-implantation tile experiment. SAM-GVS tends to function best as a false-detection-suppression tool because of its tendency to reduce the off-plume posterior indicators.

Because the pixels are the same in all of the tiles, it allows us to see the spatial impact of the varying radiometric terms underneath the plume. Frequently a M-GVS/SAM-GVS pair of images will have a common false-alarm, which suggests the aliasing of a particular gas in the library, with particular emissivity vectors in the image.

Table 4.4: Single gas M-GVS mean posterior indicator variables - acrol

0 ppm-m tile							
0.937	0.509	0.524	0.927	0.012	0.861	0.972	1.000
0.185	0.670	0.152	0.023	0.765	0.581	1.000	0.071
0.055	0.014	0.087	0.278	0.236	1.000	0.019	0.648
0.925	0.141	0.692	0.288	1.000	0.016	0.944	0.508
0.147	0.021	0.567	0.333	0.964	0.387	0.471	0.998
0.398	0.244	0.513	1.000	0.766	0.004	0.056	1.000
0.428	0.157	0.291	0.212	0.917	0.059	1.000	0.367
0.722	0.156	0.795	0.232	0.833	1.000	0.979	0.998

100 ppm-m tile							
0.111	0.980	0.014	0.862	**0.419**	0.240	0.989	1.000
0.192	0.573	0.452	**0.887**	**0.842**	**0.364**	1.000	0.682
0.864	0.027	0.764	**0.424**	**0.830**	**0.953**	0.796	0.645
1.000	0.643	0.210	**0.889**	**1.000**	**0.249**	0.863	0.346
0.029	0.066	0.721	**0.574**	**0.999**	**0.366**	0.030	0.390
0.170	0.204	0.501	**1.000**	**0.330**	**0.047**	0.455	0.991
0.096	0.171	0.564	0.246	**0.888**	0.349	0.833	0.960
0.143	0.597	0.195	0.202	**0.027**	1.000	0.952	0.997

1,000 ppm-m tile							
0.838	0.585	0.804	0.880	**0.912**	0.422	0.999	1.000
0.249	0.984	0.232	**0.469**	**0.862**	**0.578**	0.996	0.267
0.779	0.139	0.590	**0.964**	**0.781**	**1.000**	0.326	0.997
0.984	0.113	0.909	**0.245**	**1.000**	**0.384**	0.100	0.783
0.188	0.120	0.449	**0.338**	**1.000**	**0.009**	0.002	0.766
0.841	0.307	0.082	**1.000**	**1.000**	**0.006**	0.295	0.997
0.719	0.308	0.388	0.594	**0.994**	0.031	0.994	1.000
0.525	0.209	0.930	0.442	**0.991**	1.000	0.844	0.953

10,000 ppm-m tile							
0.229	0.299	0.584	0.553	**0.977**	0.046	0.997	0.998
0.098	0.990	0.014	**1.000**	**1.000**	**1.000**	0.999	0.547
0.993	0.072	0.377	**1.000**	**1.000**	**1.000**	0.809	0.278
0.963	0.168	0.543	**1.000**	**1.000**	**1.000**	0.786	0.815
0.113	0.132	0.486	**1.000**	**1.000**	**0.674**	0.514	0.974
0.099	0.049	0.037	**1.000**	**1.000**	**0.989**	0.073	0.832
0.756	0.300	0.532	0.044	**1.000**	0.001	0.977	0.981
0.626	0.092	0.370	0.487	**1.000**	0.997	0.224	0.923

Table 4.5: Single gas SAM-GVS mean posterior indicator variables - acrol

			0 ppm-m tile				
0.301	0.018	0.019	0.980	0.000	0.939	1.000	0.999
0.000	0.000	0.000	0.001	0.001	0.000	0.997	0.997
0.000	0.000	0.287	0.002	0.000	1.000	0.000	0.000
0.105	0.002	0.004	0.003	0.924	0.000	0.016	0.018
0.000	0.022	0.007	1.000	0.905	0.000	0.004	0.921
0.023	1.000	0.872	0.970	0.571	0.006	0.000	0.990
0.344	0.018	0.083	0.000	0.594	0.908	0.984	0.844
0.622	0.699	0.363	0.463	0.803	1.000	0.989	0.548

			100 ppm-m tile				
0.001	0.071	0.000	0.012	**0.000**	0.026	0.960	1.000
0.000	0.139	0.000	**0.019**	**0.987**	**0.181**	1.000	0.003
0.030	0.000	0.070	**0.000**	**0.000**	**0.795**	0.000	0.000
0.742	0.427	0.000	**0.000**	**1.000**	**0.041**	0.143	0.000
0.000	0.000	0.000	**0.000**	**1.000**	**0.000**	0.009	0.567
0.148	0.161	0.004	**0.992**	**0.285**	**0.000**	0.002	0.953
0.501	0.028	0.095	0.833	**0.656**	0.637	0.971	0.457
0.761	0.357	0.140	0.059	**0.270**	1.000	0.077	0.967

			1,000 ppm-m tile				
0.002	0.988	0.073	0.018	**0.052**	0.000	1.000	1.000
0.000	0.053	0.000	**0.986**	**0.142**	**0.088**	1.000	0.006
0.167	0.000	0.174	**0.010**	**0.026**	**0.808**	0.022	0.135
0.082	0.002	0.764	**0.006**	**0.988**	**0.704**	0.103	0.000
0.001	0.001	0.283	**0.911**	**1.000**	**0.001**	0.000	0.693
0.073	0.000	0.001	**1.000**	**0.436**	**0.012**	0.000	0.725
0.000	0.018	0.000	0.019	**1.000**	0.904	0.992	0.246
0.201	0.033	0.102	0.621	**0.992**	1.000	0.214	0.303

			10,000 ppm-m tile				
0.082	0.949	0.000	0.009	**0.977**	0.029	1.000	0.993
0.000	0.000	0.000	**1.000**	**1.000**	**0.677**	1.000	0.000
0.024	0.001	0.034	**1.000**	**0.991**	**1.000**	0.980	0.554
0.877	0.000	0.934	**0.672**	**1.000**	**0.949**	0.734	0.038
0.003	0.632	0.061	**0.704**	**1.000**	**0.440**	0.794	0.986
0.058	0.028	0.000	**1.000**	**1.000**	**0.989**	0.000	0.972
0.000	0.000	0.073	0.049	**1.000**	0.000	1.000	0.472
0.760	0.848	0.561	0.047	**1.000**	0.945	0.673	0.892

Table 4.6: Two gas M-GVS mean posterior indicator variables - freon12

0 ppm-m tile							
0.028	0.001	0.017	0.014	0.002	0.002	0.111	0.001
0.002	0.021	0.001	0.002	0.001	0.003	0.323	0.002
0.009	0.013	0.001	0.007	0.042	0.306	0.002	0.000
0.009	0.003	0.001	0.002	0.017	0.006	0.001	0.010
0.002	0.001	0.004	0.002	0.016	0.006	0.007	0.000
0.000	0.006	0.000	0.224	0.002	0.020	0.003	0.012
0.020	0.002	0.010	0.009	0.002	0.004	0.239	0.023
0.001	0.001	0.001	0.001	0.003	0.533	0.000	0.000

100 ppm-m tile							
0.000	0.009	0.007	0.001	**0.002**	0.000	0.043	0.313
0.001	0.059	0.001	**0.002**	**0.039**	**0.001**	0.010	0.000
0.001	0.002	0.000	**0.225**	**0.000**	**0.000**	0.002	0.023
0.019	0.001	0.014	**0.034**	**0.004**	**0.000**	0.001	0.009
0.044	0.003	0.044	**0.006**	0.499	**0.001**	0.006	0.017
0.000	0.000	0.007	**0.012**	0.129	**0.002**	0.008	0.003
0.000	0.000	0.001	0.002	**0.003**	0.000	0.244	0.182
0.008	0.011	0.000	0.028	**0.600**	0.001	0.000	0.002

1,000 ppm-m tile							
0.001	0.043	0.001	0.003	**0.935**	0.001	0.002	0.756
0.001	0.006	0.006	**0.997**	**1.000**	**0.947**	0.096	0.002
0.007	0.002	0.001	**0.997**	**0.788**	**0.099**	0.007	0.004
0.002	0.001	0.001	**0.528**	**1.000**	**0.994**	0.001	0.003
0.001	0.000	0.000	**0.989**	**1.000**	**0.062**	0.006	0.182
0.000	0.004	0.001	**1.000**	**1.000**	**0.851**	0.030	0.037
0.020	0.002	0.010	0.000	**1.000**	0.000	0.000	0.033
0.006	0.001	0.000	0.056	**1.000**	0.106	0.070	0.001

10,000 ppm-m tile							
0.009	0.007	0.006	0.001	**1.000**	0.001	0.000	0.001
0.009	0.000	0.001	**1.000**	**1.000**	**1.000**	0.001	0.014
0.001	0.001	0.042	**1.000**	**1.000**	**1.000**	0.001	0.006
0.002	0.000	0.002	**1.000**	**1.000**	**1.000**	0.003	0.004
0.002	0.009	0.001	**1.000**	**1.000**	**1.000**	0.006	0.001
0.001	0.000	0.002	**1.000**	**1.000**	**1.000**	0.001	0.180
0.002	0.003	0.000	0.004	**1.000**	0.004	0.017	0.028
0.000	0.012	0.001	0.017	**1.000**	0.001	0.095	0.069

Table 4.7: Two gas SAM-GVS mean posterior indicator variables - freon12

0 ppm-m tile							
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.291	0.000	0.000	0.002	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.003	0.000	0.154	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.003	0.013	0.001	0.000	0.029	0.003	0.000	0.000

100 ppm-m tile							
0.000	0.000	0.000	0.000	**0.000**	0.000	0.000	0.250
0.000	0.000	0.000	**0.000**	**0.000**	**0.000**	0.000	0.000
0.000	0.000	0.000	**0.000**	**0.000**	**0.000**	0.000	0.000
0.000	0.000	0.000	**0.000**	**0.007**	**0.000**	0.000	0.000
0.000	0.000	0.010	**0.000**	**0.000**	**0.000**	0.000	0.000
0.000	0.000	0.000	**0.000**	**0.001**	**0.000**	0.000	0.000
0.000	0.000	0.000	0.013	**0.059**	0.000	0.349	0.000
0.001	0.001	0.002	0.001	**0.522**	0.429	0.001	0.051

1,000 ppm-m tile							
0.000	0.000	0.703	0.835	**0.890**	0.000	0.000	0.000
0.000	0.000	0.000	**0.211**	**0.998**	**0.218**	0.012	0.000
0.000	0.000	0.000	**1.000**	**1.000**	**0.007**	0.000	0.000
0.000	0.000	0.000	**1.000**	**1.000**	**0.746**	0.000	0.000
0.000	0.000	0.000	**0.999**	**1.000**	**0.000**	0.000	0.000
0.093	0.000	0.000	**1.000**	**1.000**	**0.021**	0.000	0.000
0.162	0.000	0.000	0.765	**1.000**	0.112	0.003	0.000
0.002	0.001	0.007	0.001	**1.000**	0.515	0.004	0.034

10,000 ppm-m tile							
0.001	0.000	0.585	0.741	**1.000**	0.000	0.000	0.000
0.000	0.000	0.794	**1.000**	**1.000**	**1.000**	0.120	0.000
0.000	0.000	0.845	**1.000**	**1.000**	**1.000**	0.000	0.000
0.000	0.000	0.806	**1.000**	**1.000**	**1.000**	0.086	0.000
0.000	0.000	0.772	**1.000**	**1.000**	**1.000**	0.034	0.000
0.000	0.000	0.000	**1.000**	**1.000**	**1.000**	0.000	0.000
0.010	0.000	0.000	0.693	**1.000**	0.277	0.000	0.000
0.006	0.003	0.002	0.000	**1.000**	0.430	0.004	0.049

Table 4.8: Two gas M-GVS mean posterior indicator variables - freon218

0 ppm-m tile							
0.002	0.050	0.003	0.012	0.033	0.056	0.003	0.008
0.002	0.000	0.000	0.002	0.003	0.001	0.151	0.002
0.001	0.000	0.022	0.010	0.014	0.057	0.000	0.009
0.003	0.012	0.004	0.001	0.000	0.002	0.000	0.000
0.002	0.002	0.046	0.006	0.001	0.000	0.007	0.010
0.004	0.003	0.000	0.028	0.000	0.010	0.000	0.002
0.000	0.003	0.002	0.001	0.048	0.000	0.003	0.023
0.061	0.000	0.000	0.002	0.000	0.254	0.002	0.004

100 ppm-m tile							
0.013	0.002	0.000	0.032	**0.001**	0.000	0.001	0.007
0.001	0.001	0.000	**0.008**	**0.000**	**0.001**	0.000	0.004
0.006	0.002	0.001	**0.016**	**0.003**	**0.001**	0.000	0.009
0.026	0.001	0.002	**0.006**	**0.001**	**0.001**	0.003	0.006
0.002	0.101	0.004	**0.021**	**0.022**	**0.011**	0.012	0.032
0.130	0.001	0.001	**0.000**	**0.060**	**0.001**	0.000	0.006
0.018	0.002	0.001	0.014	**0.006**	0.000	0.055	0.011
0.000	0.000	0.002	0.001	**0.537**	0.001	0.011	0.001

1,000 ppm-m tile							
0.033	0.000	0.000	0.041	**0.288**	0.002	0.000	0.420
0.002	0.001	0.001	**0.930**	**1.000**	**0.705**	0.002	0.001
0.000	0.013	0.019	**0.996**	**0.999**	**0.031**	0.002	0.053
0.172	0.023	0.001	**0.860**	**1.000**	**0.989**	0.034	0.002
0.003	0.055	0.022	**0.474**	**1.000**	**0.006**	0.001	0.001
0.109	0.012	0.000	**0.999**	**1.000**	**0.695**	0.000	0.001
0.002	0.001	0.002	0.000	**1.000**	0.001	0.000	0.093
0.000	0.006	0.002	0.002	**1.000**	0.145	0.003	0.003

10,000 ppm-m tile							
0.093	0.000	0.002	0.001	**1.000**	0.000	0.000	0.001
0.000	0.000	0.000	**1.000**	**1.000**	**1.000**	0.016	0.003
0.000	0.003	0.001	**1.000**	**1.000**	**1.000**	0.000	0.186
0.037	0.006	0.001	**1.000**	**1.000**	**1.000**	0.001	0.000
0.000	0.000	0.001	**1.000**	**1.000**	**1.000**	0.000	0.001
0.014	0.001	0.002	**1.000**	**1.000**	**1.000**	0.001	0.179
0.002	0.000	0.002	0.037	**1.000**	0.001	0.033	0.000
0.028	0.009	0.001	0.002	**1.000**	0.000	0.001	0.002

Table 4.9: Two gas SAM-GVS mean posterior indicator variables - freon218

0 ppm-m tile							
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.065	0.000	0.000	0.000	0.240	0.000	0.000	0.000
0.775	0.000	0.000	0.772	0.303	0.868	0.012	0.509
0.000	0.836	0.047	0.741	0.000	0.000	0.000	0.000
0.865	0.000	0.848	0.000	0.000	0.000	0.000	0.001
0.000	0.000	0.000	0.230	0.000	0.000	0.000	0.000
0.000	0.029	0.062	0.495	0.000	0.780	0.075	0.071
0.006	0.003	0.013	0.000	0.000	0.001	0.001	0.001

100 ppm-m tile							
0.000	0.000	0.000	0.000	**0.000**	0.000	0.000	0.000
0.000	0.000	0.641	**0.000**	**0.000**	**0.000**	0.505	0.000
0.000	0.000	0.000	**0.000**	**0.000**	**0.000**	0.000	0.000
0.000	0.000	0.000	**0.772**	**0.475**	**0.000**	0.000	0.000
0.000	0.000	0.000	**0.888**	**0.229**	**0.512**	0.000	0.000
0.000	0.168	0.353	**0.012**	**0.000**	**0.000**	0.034	0.034
0.000	0.076	0.834	0.000	**0.000**	0.000	0.130	0.202
0.004	0.001	0.000	0.001	**0.409**	0.000	0.013	0.003

1,000 ppm-m tile							
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	**1.000**	**1.000**	**0.894**	0.000	0.763
0.013	0.357	0.000	**0.281**	**0.999**	**0.000**	0.000	0.000
0.000	0.063	0.000	**0.657**	**1.000**	**0.038**	0.000	0.000
0.000	0.083	0.000	**0.000**	**1.000**	**0.000**	0.000	0.000
0.000	0.854	0.000	**0.667**	**1.000**	**0.077**	0.000	0.009
0.000	0.102	0.006	0.000	**1.000**	0.000	0.000	0.529
0.002	0.000	0.004	0.006	**1.000**	0.119	0.000	0.001

10,000 ppm-m tile							
0.001	0.000	0.000	0.000	**1.000**	0.000	0.000	0.000
0.000	0.402	0.000	**1.000**	**1.000**	**1.000**	0.000	0.000
0.000	0.835	0.000	**1.000**	**1.000**	**1.000**	0.000	0.000
0.000	0.838	0.000	**1.000**	**1.000**	**1.000**	0.000	0.000
0.000	0.911	0.000	**1.000**	**1.000**	**1.000**	0.000	0.041
0.000	0.853	0.000	**1.000**	**1.000**	**1.000**	0.024	0.067
0.000	0.791	0.055	0.000	**1.000**	0.000	0.000	0.014
0.031	0.003	0.001	0.001	**1.000**	0.042	0.000	0.002

Table 4.10: Four gas M-GVS mean posterior indicator variables - c6h6

0 ppm-m tile							
0.146	0.208	0.867	0.943	0.423	0.290	1.000	0.997
0.631	0.165	0.224	0.751	0.784	0.472	0.745	0.888
0.248	0.741	0.356	0.830	0.384	0.276	0.956	0.079
0.993	0.601	0.557	0.487	0.982	0.065	0.210	0.135
0.795	0.196	0.235	0.671	1.000	0.477	0.133	0.121
0.527	0.627	0.562	0.814	0.481	0.402	0.478	0.067
0.514	0.743	0.101	0.043	0.471	0.699	0.891	0.968
0.289	0.384	0.591	0.984	0.860	0.770	0.228	0.758

100 ppm-m tile							
0.907	0.512	0.387	0.330	**0.179**	0.486	0.998	0.968
0.686	0.691	0.727	**0.647**	**0.195**	**0.445**	0.256	0.535
0.954	0.512	0.483	**0.523**	**0.297**	**0.834**	0.592	0.007
0.984	0.249	0.070	**0.563**	**0.301**	**0.620**	0.071	0.935
0.971	0.897	0.486	**0.825**	**0.971**	**0.739**	0.241	0.768
0.398	0.637	0.362	**0.340**	**0.181**	**0.839**	0.098	0.425
0.455	0.039	0.934	0.330	**0.438**	0.758	0.068	0.240
0.067	0.515	0.489	0.577	**0.492**	0.425	0.684	0.940

1,000 ppm-m tile							
0.981	0.674	0.889	0.552	**0.558**	0.556	0.970	0.997
0.664	0.989	0.157	**0.188**	**0.996**	**0.912**	1.000	0.354
0.979	0.083	0.068	**0.982**	**0.304**	**1.000**	0.380	0.250
0.999	0.348	0.874	**0.478**	**0.951**	**0.581**	0.526	0.257
0.912	0.224	0.907	**0.350**	**0.993**	**0.335**	0.263	0.991
0.824	0.715	0.644	**0.999**	**0.994**	**0.877**	0.389	0.943
0.947	0.814	0.571	0.759	**0.438**	0.357	0.783	0.927
0.745	0.344	0.256	0.296	**0.997**	0.699	0.287	0.978

10,000 ppm-m tile							
0.931	0.437	0.548	0.614	**0.972**	0.814	0.747	0.986
0.436	0.633	0.260	**0.956**	**0.862**	**0.730**	0.097	0.113
0.548	0.566	0.329	**0.719**	**0.994**	**0.996**	0.160	0.081
0.800	0.689	0.721	**0.495**	**0.855**	**0.918**	0.238	0.923
0.473	0.625	0.301	**0.927**	**0.816**	**0.633**	0.329	0.582
0.854	0.085	0.073	**0.997**	**0.998**	**0.998**	0.341	0.871
0.781	0.251	0.484	0.669	**0.999**	0.504	0.273	0.057
0.922	0.016	0.265	0.313	**0.937**	0.613	0.901	0.760

Table 4.11: Four gas SAM-GVS mean posterior indicator variables - c6h6

			0 ppm-m tile				
0.971	0.000	0.003	0.059	0.093	0.012	0.145	0.983
0.867	0.491	0.001	0.327	0.007	0.047	0.093	0.961
0.998	0.877	0.000	0.400	0.032	0.004	0.024	0.968
0.993	0.993	0.000	0.594	0.000	0.003	0.000	0.980
0.843	0.988	0.721	0.966	0.000	0.016	0.413	0.734
0.940	0.960	0.868	1.000	0.001	0.363	0.091	0.912
0.984	0.562	0.994	0.998	0.079	0.079	0.325	0.004
0.775	0.228	0.551	0.720	0.570	0.049	0.751	0.953

			100 ppm-m tile				
0.997	0.000	0.335	0.918	**0.003**	0.002	0.000	1.000
0.991	0.621	0.000	**0.073**	**0.433**	**0.182**	0.264	0.000
1.000	0.044	0.001	**0.000**	**0.115**	**0.000**	0.003	0.000
0.999	0.997	0.001	**0.003**	**0.046**	**0.107**	0.010	0.008
0.959	0.964	0.014	**0.066**	**0.818**	**0.027**	0.093	0.021
0.529	0.814	0.000	**0.200**	**0.000**	**0.000**	0.000	0.116
0.000	0.224	0.973	0.004	**0.558**	0.000	0.998	0.024
0.910	0.050	0.348	0.709	**0.653**	0.623	0.501	0.469

			1,000 ppm-m tile				
0.820	0.925	0.996	0.972	**0.000**	0.009	0.004	0.999
0.006	0.364	0.970	**0.990**	**0.345**	**0.000**	0.046	0.013
0.065	0.007	0.892	**0.996**	**0.133**	**0.993**	0.070	0.000
0.742	0.011	0.562	**0.997**	**0.644**	**0.100**	0.222	0.042
0.947	0.016	0.007	**0.833**	**0.941**	**0.053**	0.000	0.000
0.980	0.966	0.020	**0.987**	**0.137**	**0.014**	0.002	0.626
0.971	0.996	0.000	0.980	**0.983**	0.097	0.439	0.999
0.306	0.654	0.380	0.865	**0.861**	0.988	0.555	0.991

			10,000 ppm-m tile				
0.978	0.954	0.891	0.997	**0.998**	0.012	0.276	0.080
0.027	0.749	0.980	**0.998**	**1.000**	**0.000**	0.192	0.041
0.891	0.014	0.947	**0.386**	**0.047**	**0.998**	0.004	0.002
1.000	0.865	0.974	**0.195**	**0.949**	**0.824**	0.000	0.356
1.000	1.000	1.000	**0.343**	**1.000**	**0.007**	0.036	0.000
0.963	0.996	0.853	**0.003**	**0.710**	**0.929**	0.002	0.000
0.913	0.948	0.883	0.014	**0.997**	0.181	0.907	0.993
0.446	0.222	0.148	0.687	**1.000**	0.008	0.244	0.990

Table 4.12: Four gas M-GVS mean posterior indicator variables - dclp13

0 ppm-m tile							
0.037	0.002	0.002	0.003	0.002	0.000	0.001	0.036
0.001	0.001	0.000	0.000	0.007	0.011	0.019	0.008
0.009	0.003	0.008	0.000	0.009	0.016	0.001	0.089
0.020	0.009	0.002	0.102	0.003	0.003	0.003	0.022
0.002	0.000	0.013	0.000	0.023	0.111	0.039	0.001
0.016	0.007	0.004	0.128	0.083	0.002	0.000	0.012
0.014	0.000	0.000	0.001	0.000	0.047	0.016	0.002
0.000	0.036	0.006	0.122	0.000	0.111	0.016	0.003

100 ppm-m tile							
0.014	0.017	0.001	0.004	**0.006**	0.001	0.000	0.301
0.000	0.009	0.002	**0.010**	**0.002**	**0.068**	0.055	0.003
0.002	0.002	0.003	**0.004**	**0.002**	**0.004**	0.061	0.006
0.009	0.028	0.002	**0.006**	**0.022**	**0.024**	0.016	0.003
0.146	0.001	0.009	**0.000**	**0.033**	**0.029**	0.000	0.010
0.068	0.117	0.003	**0.008**	**0.003**	**0.073**	0.001	0.030
0.014	0.001	0.020	0.016	**0.003**	0.004	0.008	0.003
0.000	0.102	0.002	0.000	**0.087**	0.023	0.079	0.000

1,000 ppm-m tile							
0.030	0.041	0.000	0.033	**0.004**	0.002	0.000	0.080
0.031	0.023	0.071	**0.004**	**0.275**	**0.033**	0.003	0.001
0.006	0.000	0.001	**0.004**	**0.003**	**0.001**	0.003	0.004
0.030	0.004	0.017	**0.011**	**0.016**	**0.148**	0.012	0.000
0.029	0.001	0.001	**0.006**	**0.021**	**0.002**	0.039	0.071
0.001	0.002	0.020	**0.006**	**0.435**	**0.078**	0.004	0.001
0.000	0.147	0.036	0.001	**0.001**	0.003	0.016	0.053
0.138	0.009	0.001	0.001	**0.412**	0.115	0.004	0.008

10,000 ppm-m tile							
0.003	0.010	0.011	0.034	**0.347**	0.036	0.047	0.247
0.001	0.197	0.002	**0.219**	**0.009**	**0.010**	0.006	0.004
0.001	0.001	0.003	**0.059**	**0.011**	**0.009**	0.000	0.002
0.007	0.004	0.037	**0.003**	**0.009**	**0.177**	0.000	0.011
0.041	0.002	0.001	**0.364**	**0.125**	**0.019**	0.007	0.013
0.101	0.001	0.000	**0.002**	**0.120**	**0.303**	0.057	0.001
0.017	0.016	0.011	0.004	**0.087**	0.003	0.018	0.001
0.002	0.000	0.004	0.023	**0.008**	0.110	0.021	0.004

Table 4.13: Four gas SAM-GVS mean posterior indicator variables - dclp13

			0 ppm-m tile				
0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.001
0.336	0.000	0.180	0.000	0.000	0.000	0.000	0.000
0.752	0.000	0.000	0.002	0.003	0.000	0.000	0.000
0.723	0.001	0.000	0.000	0.123	0.000	0.112	0.000
0.907	0.515	0.563	0.000	0.000	0.002	0.000	0.000
0.483	0.364	0.631	0.023	0.000	0.378	0.630	0.001
0.293	0.690	0.910	0.000	0.000	0.000	0.002	0.000
0.055	0.166	0.001	0.009	0.018	0.030	0.006	0.004

			100 ppm-m tile				
0.000	0.000	0.000	0.000	**0.000**	0.000	0.000	0.020
0.544	0.003	0.740	**0.000**	**0.000**	**0.000**	0.000	0.000
0.160	0.001	0.001	**0.000**	**0.000**	**0.000**	0.000	0.000
0.695	0.001	0.000	**0.000**	**0.000**	**0.000**	0.000	0.000
0.861	0.001	0.000	**0.000**	**0.000**	**0.004**	0.714	0.000
0.000	0.003	0.000	**0.000**	**0.000**	**0.000**	0.000	0.000
0.000	0.562	0.000	0.039	**0.872**	0.285	0.000	0.000
0.004	0.087	0.021	0.002	**0.066**	0.044	0.000	0.002

			1,000 ppm-m tile				
0.627	0.000	0.000	0.000	**0.000**	0.000	0.000	0.006
0.000	0.573	0.004	**0.000**	**0.002**	**0.000**	0.000	0.000
0.000	0.000	0.778	**0.000**	**0.003**	**0.002**	0.000	0.000
0.057	0.000	0.677	**0.007**	**0.000**	**0.000**	0.000	0.002
0.056	0.000	0.000	**0.000**	**0.021**	**0.000**	0.630	0.000
0.652	0.214	0.000	**0.002**	**0.000**	**0.000**	0.000	0.000
0.859	0.002	0.000	0.000	**0.659**	0.001	0.000	0.023
0.002	0.012	0.001	0.001	**0.024**	0.001	0.046	0.003

			10,000 ppm-m tile				
0.605	0.000	0.000	0.001	**0.002**	0.000	0.000	0.041
0.000	0.781	0.000	**0.000**	**0.000**	**0.000**	0.000	0.000
0.010	0.000	0.000	**0.716**	**0.571**	**0.146**	0.002	0.000
0.504	0.626	0.000	**0.825**	**0.232**	**0.172**	0.000	0.000
0.890	0.778	0.000	**0.654**	**0.123**	**0.000**	0.000	0.000
0.633	0.762	0.000	**0.693**	**0.494**	**0.000**	0.000	0.000
0.812	0.846	0.019	0.329	**0.078**	0.016	0.000	0.000
0.014	0.017	0.002	0.002	**0.303**	0.017	0.004	0.146

Table 4.14: Four gas M-GVS mean posterior indicator variables - f114

0 ppm-m tile							
0.518	0.030	0.020	0.625	0.407	0.307	0.682	0.225
0.009	0.427	0.008	0.293	0.279	0.032	0.066	0.032
0.046	0.268	0.022	0.143	0.765	0.152	0.246	0.022
0.072	0.387	0.044	0.019	0.389	0.096	0.613	0.086
0.313	0.191	0.546	0.158	0.652	0.017	0.380	0.448
0.022	0.242	0.525	0.004	0.022	0.300	0.058	0.129
0.393	0.435	0.089	0.061	0.469	0.075	0.287	0.508
0.010	0.000	0.368	0.425	0.326	0.710	0.011	0.393

100 ppm-m tile							
0.057	0.294	0.491	0.221	**0.357**	0.372	0.277	0.573
0.044	0.617	0.270	**0.226**	**0.277**	**0.002**	0.593	0.303
0.050	0.012	0.347	**0.339**	**0.283**	**0.027**	0.031	0.158
0.537	0.041	0.439	**0.234**	**0.029**	**0.139**	0.436	0.027
0.372	0.221	0.004	**0.009**	**0.002**	**0.022**	0.320	0.019
0.168	0.004	0.002	**0.494**	**0.047**	**0.189**	0.046	0.785
0.399	0.001	0.008	0.224	**0.050**	0.098	0.111	0.508
0.088	0.021	0.226	0.011	**0.002**	0.452	0.007	0.328

1,000 ppm-m tile							
0.616	0.256	0.422	0.536	**0.038**	0.161	0.026	0.390
0.004	0.126	0.075	**0.201**	**0.993**	**0.030**	0.295	0.206
0.577	0.237	0.596	**0.727**	**0.517**	**0.058**	0.175	0.536
0.562	0.070	0.162	**0.108**	**1.000**	**0.365**	0.161	0.438
0.172	0.236	0.017	**0.016**	**1.000**	**0.227**	0.001	0.014
0.033	0.319	0.023	**0.902**	**1.000**	**0.177**	0.021	0.041
0.572	0.214	0.092	0.070	**1.000**	0.001	0.446	0.029
0.006	0.016	0.000	0.037	**1.000**	0.789	0.419	0.028

10,000 ppm-m tile							
0.267	0.023	0.016	0.018	**1.000**	0.212	0.696	0.685
0.628	0.509	0.006	**1.000**	**1.000**	**1.000**	0.068	0.443
0.011	0.058	0.178	**1.000**	**1.000**	**1.000**	0.325	0.289
0.415	0.294	0.287	**1.000**	**1.000**	**1.000**	0.447	0.001
0.393	0.306	0.070	**1.000**	**1.000**	**1.000**	0.410	0.253
0.009	0.157	0.113	**1.000**	**1.000**	**1.000**	0.022	0.541
0.621	0.000	0.311	0.232	**1.000**	0.246	0.027	0.018
0.424	0.002	0.293	0.537	**1.000**	0.721	0.014	0.526

Table 4.15: Four gas SAM-GVS mean posterior indicator variables - f114

			0 ppm-m tile				
0.255	0.905	0.818	0.000	0.000	0.585	0.882	0.065
0.077	0.001	0.008	0.000	0.067	0.709	0.071	0.000
0.000	0.000	0.192	0.000	0.665	0.424	0.000	0.000
0.000	0.000	0.634	0.670	0.000	0.145	0.000	0.000
0.101	0.012	0.070	0.000	0.588	0.000	0.000	0.000
0.141	0.000	0.062	0.217	0.017	0.079	0.051	0.177
0.000	0.032	0.000	0.250	0.102	0.468	0.760	0.842
0.137	0.008	0.275	0.001	0.166	0.349	0.018	0.503

			100 ppm-m tile				
0.000	0.239	0.000	0.000	**0.000**	0.635	0.923	0.336
0.000	0.283	0.000	**0.014**	**0.552**	**0.755**	0.000	0.000
0.000	0.003	0.006	**0.018**	**0.000**	**0.000**	0.000	0.268
0.000	0.000	0.557	**0.952**	**0.515**	**0.000**	0.080	0.931
0.000	0.000	0.013	**0.115**	**0.030**	**0.000**	0.021	0.000
0.692	0.000	0.000	**0.747**	**0.000**	**0.749**	0.740	0.843
0.871	0.000	0.408	0.000	**0.000**	0.002	0.000	0.932
0.498	0.011	0.016	0.006	**0.031**	0.680	0.047	0.600

			1,000 ppm-m tile				
0.000	0.102	0.001	0.038	**0.000**	0.024	0.947	0.251
0.041	0.076	0.107	**0.014**	**0.987**	**0.740**	0.565	0.845
0.923	0.008	0.000	**0.016**	**0.786**	**0.149**	0.093	0.067
0.052	0.603	0.000	**0.050**	**1.000**	**0.000**	0.840	0.652
0.000	0.000	0.169	**0.000**	**1.000**	**0.002**	0.041	0.000
0.000	0.000	0.591	**0.716**	**1.000**	**0.296**	0.954	0.857
0.000	0.000	0.858	0.052	**1.000**	0.001	0.000	0.026
0.016	0.010	0.498	0.666	**1.000**	0.217	0.269	0.018

			10,000 ppm-m tile				
0.216	0.027	0.000	0.000	**1.000**	0.376	0.060	0.350
0.000	0.000	0.047	**1.000**	**1.000**	**1.000**	0.571	0.000
0.028	0.256	0.000	**1.000**	**1.000**	**1.000**	0.022	0.023
0.000	0.023	0.000	**1.000**	**1.000**	**1.000**	0.118	0.669
0.000	0.000	0.065	**1.000**	**1.000**	**1.000**	0.059	0.071
0.000	0.001	0.083	**1.000**	**1.000**	**1.000**	0.951	0.136
0.000	0.000	0.000	0.031	**1.000**	0.019	0.951	0.000
0.032	0.017	0.202	0.212	**1.000**	0.023	0.655	0.021

Table 4.16: Four gas M-GVS mean posterior indicator variables - so2

0 ppm-m tile							
0.120	0.009	0.000	0.022	0.254	0.140	0.981	1.000
0.022	0.044	0.000	0.297	0.201	0.002	0.998	0.076
0.003	0.006	0.097	0.135	0.136	0.997	0.023	0.539
0.573	0.122	0.019	0.007	0.939	0.001	0.191	0.003
0.010	0.417	0.016	0.105	0.990	0.029	0.002	0.239
0.041	0.116	0.004	0.755	0.014	0.002	0.009	0.561
0.048	0.003	0.010	0.006	0.078	0.000	0.416	0.461
0.076	0.007	0.008	0.001	0.012	0.999	0.006	0.027

100 ppm-m tile							
0.285	0.037	0.042	0.003	**0.077**	0.010	0.977	1.000
0.002	0.010	0.001	**0.029**	**0.017**	**0.006**	1.000	0.376
0.030	0.003	0.319	**0.379**	**0.032**	**0.988**	0.008	0.651
0.943	0.101	0.072	**0.006**	**0.145**	**0.007**	0.254	0.001
0.004	0.107	0.080	**0.170**	**0.759**	**0.157**	0.003	0.069
0.024	0.040	0.007	**0.997**	**0.002**	**0.044**	0.010	0.316
0.010	0.000	0.002	0.024	**0.004**	0.007	0.095	0.043
0.099	0.046	0.013	0.000	**0.000**	0.996	0.013	0.126

1,000 ppm-m tile							
0.162	0.069	0.072	0.188	**0.040**	0.349	0.947	1.000
0.001	0.011	0.010	**0.000**	**0.000**	**0.004**	0.997	0.087
0.139	0.021	0.130	**0.001**	**0.012**	**0.996**	0.006	0.835
0.815	0.059	0.154	**0.009**	**0.128**	**0.002**	0.073	0.001
0.000	0.234	0.453	**0.058**	**0.100**	**0.039**	0.055	0.640
0.136	0.082	0.009	**0.915**	**0.623**	**0.023**	0.001	0.956
0.354	0.003	0.028	0.228	**0.018**	0.001	0.606	0.143
0.098	0.002	0.002	0.011	**0.746**	0.999	0.029	0.039

10,000 ppm-m tile							
0.740	0.026	0.019	0.040	**0.428**	0.049	0.674	1.000
0.061	0.018	0.002	**0.714**	**0.010**	**0.111**	0.953	0.068
0.020	0.030	0.020	**0.016**	**0.007**	**0.098**	0.002	0.253
0.932	0.377	0.024	**0.016**	**1.000**	**0.794**	0.562	0.019
0.008	0.019	0.032	**0.648**	**1.000**	**0.078**	0.001	0.117
0.067	0.003	0.021	**0.003**	**0.996**	**0.840**	0.003	0.881
0.234	0.032	0.044	0.172	**1.000**	0.008	0.278	0.058
0.369	0.001	0.028	0.006	**0.891**	0.998	0.004	0.028

Table 4.17: Four gas SAM-GVS mean posterior indicator variables - so2

0 ppm-m tile							
0.000	0.000	0.000	0.000	0.001	0.000	0.012	0.989
0.000	0.812	0.000	0.000	0.736	0.048	0.993	0.001
0.000	0.000	0.930	0.004	0.455	0.050	0.000	0.031
0.038	0.000	0.924	0.645	0.001	0.000	0.000	0.000
0.000	0.000	0.000	0.001	0.998	0.061	0.000	0.000
0.000	0.000	0.000	0.659	0.800	0.000	0.000	0.377
0.000	0.000	0.000	0.109	0.631	0.425	0.928	0.732
0.180	0.006	0.001	0.006	0.027	0.928	0.031	0.075

100 ppm-m tile							
0.002	0.000	0.000	0.000	**0.001**	0.000	0.004	1.000
0.000	0.795	0.000	**0.000**	**0.011**	**0.109**	0.802	0.007
0.000	0.835	0.903	**0.840**	**0.661**	**0.660**	0.000	0.182
0.022	0.001	0.884	**0.020**	**0.703**	**0.839**	0.892	0.073
0.000	0.087	0.848	**0.758**	**0.247**	**0.000**	0.000	0.000
0.920	0.000	0.053	**0.998**	**0.000**	**0.664**	0.000	0.998
0.000	0.000	0.000	0.000	**0.027**	0.000	0.027	0.940
0.081	0.010	0.008	0.000	**0.004**	1.000	0.017	0.207

1,000 ppm-m tile							
0.082	0.001	0.000	0.000	**0.870**	0.000	0.324	1.000
0.832	0.000	0.000	**0.000**	**0.471**	**0.224**	0.996	0.000
0.001	0.572	0.000	**0.000**	**0.672**	**0.418**	0.000	0.004
0.003	0.882	0.000	**0.000**	**0.899**	**0.017**	0.458	0.000
0.000	0.944	0.727	**0.000**	**0.246**	**0.430**	0.000	0.000
0.000	0.000	0.725	**0.000**	**0.877**	**0.030**	0.191	0.994
0.000	0.000	0.624	0.000	**0.706**	0.012	0.000	0.006
0.048	0.000	0.043	0.026	**0.007**	1.000	0.000	0.172

10,000 ppm-m tile							
0.003	0.000	0.000	0.000	**0.000**	0.897	0.148	0.994
0.000	0.000	0.000	**0.002**	**0.008**	**0.918**	0.999	0.000
0.000	0.805	0.000	**0.000**	**0.000**	**0.032**	0.205	0.009
0.786	0.000	0.000	**0.485**	**1.000**	**0.835**	0.684	0.095
0.000	0.000	0.090	**0.128**	**1.000**	**0.862**	0.694	0.000
0.001	0.000	0.000	**0.027**	**0.882**	**0.001**	0.208	0.982
0.000	0.000	0.000	0.242	**1.000**	0.177	0.970	0.001
0.147	0.001	0.008	0.003	**0.410**	0.903	0.067	0.108

Chapter 5

Conclusions

The conclusions reached in this thesis can be broken down into three broad categories. First is a set of conclusions based on M-GVS itself, followed by a discussion of M-GVS applied to single-pixel identification problems, and finally a discussion of the utility of the multi-pixel framework. This will be followed by a discussion of the overall conclusions reached in the course of this thesis, and a discussion of what areas might merit future research.

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 $Pr(\boldsymbol{\gamma})$ that places higher probability of inclusion on certain gases. Additionally there is substantial room for "tuning" the method for application-specific scenarios in terms of the library constituents, determining the optimal uninformed $Pr(\gamma)$, and acceleration factors for a given problem.

This body of work demonstrates that M-GVS is a viable technique for gaseous effluent detection and describes the advantages and disadvantages of the method. Furthermore, the operating envelope of M-GVS was explored; as a function of model complexity, $Pr(\gamma)$, user-supplied prior information, and SNR. Finally, the potential to apply many of these "single pixel" techniques to other problems in spectral imaging was described, particularly the method of gas library thinning as applied to spectral unmixing.

This document also highlights the inherent challenges in using M-GVS, which are

properties of the method itself. In particular, its non-robustness in dealing with out-of-model gases is a substantial challenge to its widespread use in the analysis community. While there is clear practical utility to the method, there is an obvious degenerate behavior in the way that M-GVS returns the user-specified prior for out-of model posterior indicators. The extent to which an analyst can leverage this non-robust behavior to suppress out-of-model gases (false-alarms) is also a property will have to be validated for every application. The degree to which this limits its utility, and the motivation for exploring alternate methodologies is a question that will be user and application-specific.

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. This is based on the demonstrated phenomenology whereby a correctly informed prior will positively impact the observed posterior indicator variable's potential to improve detection. This potential was demonstrated on various synthetic spatially varying plumes, and with its current state of development, it shows that the current mixing functions are best suited to increasing the specificity of M-GVS. While not demonstrated here, there is latitude to develop mixing functions that are tailored to other uses (sensitivity instead of specificity) or finding plumes of a specific shape. Possibly the most interesting use of SAM-GVS would be in a post-detection framework where some sort of detection scores could also be included in the mixing functions.

Finally, the author has shown that there exists the clear 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. This is based on the same phenomenology as the spatial aggregation and is described in greater detail in the Future Work section.

While this body of work focused on a Bayesian regression method applied to an effluent identification problem, its development might be useful to a wide variety of imaging problems. Many imaging problems including target detection, landcover classification, image utility measurement, and sensor modeling often take advantage of some sort of least squares process and/or some sort of spectral library. The Bayesian paradigm and spectral library thinning methodologies demonstrated in this thesis have clear potential to contribute to those problems as well.

To summarize, the author has introduced two new methods of gaseous effluent identification in passive LWIR imagery. The first technique (M-GVS) allows for simultaneous estimation of the concentration path length of each gas, as well as the probability of inclusion of each gas. It is an iterative method that is implemented through MCMC and was developed and tuned for unsupervised application to imaging applications where rapid convergence is needed. This technique was demonstrated on semi-synthetic data produced from real LWIR HSI and the results of M-GVS were compared to stepwise regression. Additionally spatial aggregation was demonstrated via SAM-GVS and the results were compared to M-GVS.

In closing, the author would like to thank Dr David Messinger, Dr Carl Salvaggio, Dr Yolande Tra, Dr Joseph Voelkel, and Dr David Ross for their mentorship, guidance, and candid feedback in the development of this work.

Chapter 6

Future Work

This chapter describes a variety of topics encountered in the process of this research that might warrant future research. It includes additional data processing methods, statistical approaches, as well as ideas for various CONOPs that were alluded to previously.

Data Processing

One substantial limiting factor of this research was the spectral resolution of the AHI data that was the starting point of the work. Recall that the 50 band spectral resolution was a result of downsampling, due to low SNR in the data that was available. This loss of spectral resolution, and its associated numerical linear algebra complications were a substantial constraint in the development of this work.

One possible way to improve the spectral resolution of these data, while minimizing the impact of low SNR is to perform OLSTER on the rebinned data, then interpolate the results to the full spectral resolution. This is tenable because the retrieved physical parameters are often reasonably smooth functions and would in theory allow for better species identification.

Adapting GVS to Specialized CONOPS

The persistent surveillance approach is built on the assumption that one has the ability to view the same area of interest many, many times and obtain many images of that area. Assuming one has the ability to perform high-fidelity spatial registration, one

can start to build fairly sophisticated models that could describe: the variability in the $\boldsymbol{\xi}_s$ for every point on the ground, the variability in the surface temperature throughout the day, the dependence on viewing geometry, etc. Further, one could begin to incorporate atmospheric models to estimate a variety of atmospheric parameters and so forth. Practically one assumes that it is possible to get enough imagery under favorable conditions, without gas plumes in order to carry out this approach.

In the context of this research proposal, given that we lack the infrastructure to persistently observe an area, the next best alternative is to use the existing AHI image cubes that have been "taken apart" with OLSTER, insert a gas plume, and then "reassembled" using Equation 3.12. One can then spatially modify the plume to generate fairly realistic imagery that might be obtained from a persistent surveillance scenario. A description of how these plumes are modified along with examples are included in the Appendix.

The other common scenario is what the author calls the "single look" scenario, where one only has available one image cube of radiance values, and the image contains a plume that one would like to identify. Using any of a standard set of anomaly/target detection algorithms, it is conceivable that one could find and spatially bound the plume in the image. Then use OLSTER as described previously to perform temperature/emissivity separation for the non-plume portion of the image.

Under the assumption that the non-plume pixels contain a similar set of materials (and thus $\boldsymbol{\xi}_s$ vectors) it is possible to apply a standard technique known as endmember selection to identify constituent materials in the scene [17] [6]. Then attempt to find the best fitting linear combination of those materials to represent the surface materials that exist *under* the plume, followed by a regression step aimed at identifying the gas plume that is above those pixels.

The plume-ground temperature contrast will play a key role in determining the columns of the design matrix. It is possible to either exhaustively solve the regression problem for a range of potential contrast values, or use a Bayesian approach and place a prior on the temperature contrast term. One could potentially use a physical model to create an empirical distribution for this parameter.

One can imagine uses for which there would be interest in real-time identification of effluents, such as by first responders to a fire at a chemical refinery. Semi-autonomous applications such as these place significantly greater challenges on the detection phase

of the processing steps. Another practical constraint for the single look scenario is the requirement that the plume occupy (roughly) less than half of the field of view of the sensor, so that there are sufficient numbers of non-plume pixels available for OLSTER and endmember selection.

Rank-deficiency of the Gas Library

The Methods chapter of this thesis describes the process used to thin the gas library for the purposes of reducing its rank-deficiency. Two other competing strategies include some sort of partitioning scheme where the library is simply broken into two or more pieces for the regression step, or using a regression approach that deals directly with the rank-deficiency of the design matrix.

There are many ways to effectively partition the gas library for the purpose of these (or other regression-based) plume identification methods. Some possibilities include:

- Perhaps choosing the subsets of spectra to test based on trying to achieve the most uniform possible rank of $(X'X)$.

- Possibly identify up front, the *most* rank deficient $(X'X)$ that could be formed, set those spectra aside, and then run the others first. This is going to be pretty unsatisfying if some of those spectra are high priority gases to test against.

- Randomly selecting a gas spectra for inclusion into a subset, up to the point that the condition number κ hits a certain threshold, say 10^2. This could also be a useful way to apply some sort of bootstrapped confidence interval on the posterior probability of each model or performance metric.

- Perhaps selecting gas spectra for inclusion into a subset, based on their importance to the user, up to the point that the condition number κ hits a certain threshold.

- Ultimately if the regression model is run in subsets (re-using the same data vector with several different design matrices) and one then hopes to aggregate the results, there is a clear multiple hypothesis testing problem. Perhaps if one runs say ten subsets of the entire gas library, one might need to choose the top k parameters with the highest posterior probability of inclusion from each subset run, then run

those again. This has the obvious problem that one is not guaranteed to ever get to a design matrix where $X'X$ is well conditioned (in fact one as spectral resolution goes down, one is *more* likely to get a poorly conditioned $X'X$), and then there is the question of how to rationally combine inference from several separate regression runs.

In a linear algebra context, this problem amounts to finding an optimal scheme for permuting the columns of the design matrix, such that a block diagonalization of $X'X$ (with varying numbers of sub blocks) produces a set of sub blocks with uniformly low condition numbers. That is to say

$$\begin{pmatrix} [X_1] & 0 & \cdots & 0 \\ 0 & [X_2] & \cdots & 0 \\ \vdots & & & \end{pmatrix}$$

such that for some real, positive constant Ξ that $\forall \quad \kappa(X_L) \leq \Xi$ and $[X_1] \cap [X_2] \cap \cdots = \emptyset$. Once found, these sub blocks contain the indices of the optimal partitioning.

An alternate method, that might make use of off-the shelf code, would be to randomly permute the columns of X and forming small sub blocks, such that the sub blocks have the most uniformly high set of singular values, as can be obtained by permuting and taking subsets from the columns of X. Whether this sub block approach or a user-driven approach is more attractive, likely depends on whether one decides to whiten the gas library with the non-plume covariance matrix (which might necessitate recomputing for every image).

Although there are many ways to effectively partition the gas library for the purpose of these (or other regression-based) plume identification methods, an alternative way to deal with $X'X$ that is rank deficient, is to apply "ridge regression" (RR). RR is a form of Tikonhov Regularization, where instead of solving $\boldsymbol{\beta} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'Y}$ one solves

$$\boldsymbol{\beta} = (\boldsymbol{X'X} + c\boldsymbol{I})^{-1}\boldsymbol{X'Y} \tag{6.1}$$

over a range of values of c, where $\boldsymbol{I}$ is the identity matrix. For c that is "large enough" the inverse term in Equation 6.1 will become full rank and the estimate for $\boldsymbol{\beta}$ will stabilize. One generally then plots β_i versus c and picks a value where all the $\boldsymbol{\beta}$'s appear to have stabilized. RR can be thought of as "intentionally solving the wrong

problem" in order to have a problem that is better behaved. That is to say, that the fitted values for the regression coefficients from RR are biased, but will have lower variance than their unbiased counterparts from OLS regression (Equation 3.2)[19].

This bias/variance trade is often results in a dramatic reduction in the standard errors of the regression coefficients. The judgement of whether or not a practitioner is willing to live with biased estimates is subjective. An example of a "ridge trace" is shown in Figure 6.1, which is a plot of the regularization parameter c (in Equation 6.1) versus the set of all $\hat{\beta}$. This plot was obtained through the statistical computing package R's built in `lmridge` function, applied to the 31 gas library of absorption spectra. For this gas library, an appropriate value for the regularization parameter would be roughly 0.02, which is the point at which the ridge trace values stabilize (they are basically constant to the right of 0.02 in Figure 6.1).
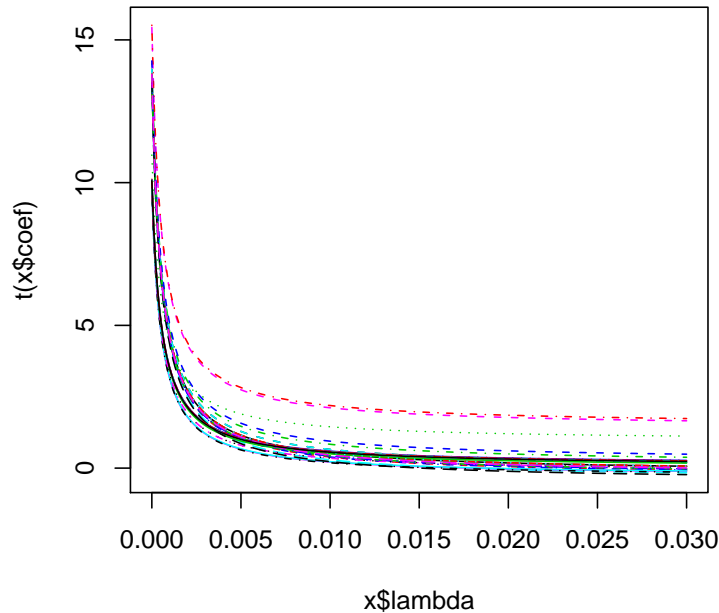
Figure 6.1: Plot of regression coefficients versus values of c

The author posits that it is an application-specific question as to whether RR is preferable to library thinning, as the answer is likely to depend greatly on the user's needs and priorities. Subjectively the difference between the two approaches can be summarized in that RR is effectively "solving the 'wrong' problem, for the sake of a numerically tractable one", while thinning the gas library amounts to clearly (and somewhat arbitrarily) eliminating certain certain solutions from being reached, with no reassurance that the part of the sample space that is discarded isn't the "interesting part." It is the author's opinion that, assuming one's application permits it thinning the library, possibly into several subsets with equally-poorly conditioned libraries that could be processed in parallel, is the best compromise between inclusiveness and numerical stability.

There are a number of theoretic (read "semi-automated") ways of choosing the optimal c value and also ridge regression is available both via Markov Chain Monte Carlo (MCMC) and OLS methods, however the problem is that the estimates produced are always biased. This may or may not be a problem depending on the application, the amount of bias, and how readily that bias can be compensated for.

In the context of model selection, biased estimates might be preferable if it aids in ones ability to select the correct model. Furthermore, in this application one does not know the optical path length (thickness) of the plume, and as such any estimated value for $\boldsymbol{\beta}$ is going to be known only up to some (unknown) multiplicative constant. Thus the penalty of biased estimates is not as onerous as it might seem, because the estimates that OLS will produce are also going to be biased as a result of the physical problem.

For the purposes of this project, the author is limited to a library of 31 total gas spectra, the condition number of $\kappa(\boldsymbol{X'X})$ is on the order of 10^5. With only a single spectra removed, this condition number generally drops to roughly 10^2 or less, thus arguing that another path is to use some hybrid of RR and library thinning.

One of the benefits of using a Bayesian approach to inference of regression coefficients is the flexibility of dealing with multiple simultaneous comparisons of parameters. In many problems, one wishes to have certain probability coverage for interval estimate, often quoted as a $(1-\alpha)$ confidence interval (CI). There are a number of ways to correct frequentist CI's if one wishes to take them "pair-wise" or conduct multiple simultaneous testing [19]. These generally amount to correction-factors by which the intervals

are widened, in order to maintain joint $(1 - \alpha)$ probability coverage. In a Bayesian context, the analog of CI's are "$(1 - \alpha)$ credible sets" (CS) or highest posterior density (HPD) intervals. Because the posterior densities of parameters are obtained empirically through MCMC simulation, one can jointly integrate out the desired simultaneous $(1 - \alpha)$ probability regions, and avoid heuristic correction factors to adjust the intervals (assuming proper priors were used). Practically this is important because the user is interested in the probabilistic inferences regarding the quantity of gases present to be simultaneously correct.

Further, because the posterior distribution $f(\boldsymbol{\theta}|x)$ 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. This (normalized) frequency is a directed estimate of the probability of any given model, and is reported along with the probability of inclusion for each indicator variable and the point or interval estimate for each regression parameter. For large models this is problematic as the model space $\mathcal{M}$ has dimensions 2^l (note $2^{31} \approx 2 \times 10^9$) and in order to sufficiently characterize $pr(\mathcal{M}|Data)$ one would need MCs that are several orders of magnitude longer than 2^9, which can be impractical to compute.

Further Extensions of M-GVS

There is an unresolved issue of the potential impact for the search order used to influence. Ideally, the order in which the image is stepped over would not influence the results, but it clearly has the potential to under the current instantiation of SAM-GVS. One simple way of mitigating this effect would be to randomize the search order, although to be of any clear benefit, the entire image would have to be re-run some large number of times in order to average out (for every pixel) the impact of search order.

One potential way to leverage the impact of search order is to use SAM-GVS in a cued manner, such that it starts at the pixel with highest detection score in an image, then propagates out from there. This gives the greatest possible benefit towards informing a weak plume pixel from the inference in one of its more concentrated neighbors.

Additionally, there exists potential for developing different mixing functions for different purposes. For example, this work has focused on identifying weak plumes with an inherent bias towards sensitivity at the expense of specificity. One could easily

develop a mixing function that favored specific gases, or spatial orientations of plumes (perhaps incorporating meteorological information), or simply one that attempted to increase the confidence level at which any gas is identified.

Further Extensions of SAM-GVS

There are a variety of ways this spatial aggregation could be implemented [24] [25]. Since the entire aim of this project is precise inference on the presence/absence of a gas and its abundance, one would generally prefer to use a full spatial statistics model, ideally one that was derived to faithfully represent the phenomena in question. There has been considerable research in the field of atmospheric science towards modeling the properties of gaseous plumes and their spread through the atmosphere [26] [27]. Many of these models are based on some form of turbulent flow or expansion methodology, and are reasonable approximations to how a gas plume on a flat surface (with controlled atmospheric inputs) would physically behave.

These models hold promise for this application in several ways. First, they provide a physical-principles based way of making quantitative inference on a whole host of plume parameters. This provides a large degree of realism, in that it allows for an exploration of ways to model the plume concentration and the plume temperature. Recall that in Equation 3.12 one practically has point estimates for every term except the plume temperature T_p and the integrated concentration path length C_l. These physical models could be used to derive a specific form for the relationships shown in Figures 2.3 and 2.4. Runs from this type of model could also be used as an input into a Bayesian model, perhaps as an initial condition, or simply a dataset on which to develop a 2-D auto-correlation function that could be used to conduct inference on adjacent pixels.

The main drawback of these models is the need for ground truth to properly initialize them, examples include: wind speed/direction at several heights in the air column, density profiles of the air column, rate of release of the effluent, etc. Their greatest utility probably lies in their use as a synthetic source of training data, from which to estimate the macro-scale auto-correlation in observed plume radiance. That fitted model could then be used as a framework one would apply to real data. This semi-parametric approach would help introduce some realism from the model, but without becoming entirely consumed by iterating a model against data that was produced by the

model, in the absence of meaningful ground truth. If full ground truth were available for the input parameters of a physical model, which they are not, conducting Bayesian inference on this model (and its associated parameters) directly would be the most appealing direction for inference. The lack of this ground truth motivates the search for some data driven approach.

An alternate approach that is data driven is the use of so called Modeling Field Theory (MFT) [28]. MFT was developed for semi-automated analysis of synthetic aperture radar (SAR) images. After mean subtraction and some sort of plume detection via SMF or PCA, one has a set of 2-D grayscale images of the filter output, the magnitude of each pixel might be roughly proportional to the plume concentration (but contains many false alarms). Essentially MFT is an iterative method for computing maximum likelihood estimation, where one randomly "seeds" the filter output image with a number of 2-D Gaussian distributions. Then allowing the location, and covariance matrices of the seeds to vary, one iteratively computes the MLE for the location (mean) and covariance of each seed. After sufficient iteration, the spatial structure of the seeds adapts so that one has an easy to use mixture of Gaussians model to use for the spatial concentration/temperature of the plume. Additionally one can include a spatially uniformly varying "noise" term to account for a portion of the background variability (random measurement noise).

By spatially summing the resultant likelihoods from MFT and setting a threshold on the minimum integrated value (and possibly some basic image morphology) one can identify the set of likelihoods that best represent the plume. It should also be noted that this approach can be used simultaneously on several image detection planes (obtained from different filters) and the results integrated in order to increase robustness. These summed likelihoods (once normalized) can be used to create parametric approximations to quantities that are roughly proportional to the filter output. For example a large filter output could be proportional to a large integrated path concentration C_l, or a large background-plume temperature contrast (recall Figures 2.3 and 2.4). As an aside, the denominator of the CEM operator is designed to make is somewhat resistant to changes in overall signal level, but this is not necessarily the case with all detectors. In either case, one can use a data-derived metric to provide rough estimation for either of these relationships or simply to provide an estimate for the spatial boundary of the plume.

Temporal Extensions of M-GVS

The greatest strength (or weakness depending on one's perspective) of Bayesian inference is that it explicitly incorporates *a priori* understanding or belief about the parameter(s) of interest. This can be thought of as the "aggregation" of understanding, as data is collected. Two interesting applications in the plume identification problem in which someone might wish to aggregate information are in some form of sequential analysis or spatial analysis. For this problem, the use of sequential analysis amounts to using multiple frames of data, each having been collected at different times. Spatial aggregation might be useful if one is anticipating multi-pixel, well-mixed plumes where one might wish to pool the estimation results from adjacent pixels and estimate adjacent pixels simultaneously. The extent to which this is tenable will likely depend on accuracy of surface emissivity estimates, if one is to address the case where a plume is above adjacent pixels containing different materials.

Sequential analysis is in a general sense, a field of statistics that makes use of a non-fixed sample size to carry out statistical testing. It can be shown [29] that instead of using a fixed sample size and a standard likelihood ratio test (LRT) to test a set of parameters at a specified $(1 - \alpha)$ error rate, one can in general devise tests that are more sensitive, at least in the sense that the same decision with the same error rate can be reached through the use of less data, by testing sequentially instead of all at once. In some applications such as remote sensing, it is simply common to collect data in a sequential manner, because of successive frames being taken of the scene. Thus it is natural (depending on the sampling frequency) to try to combine inference from consecutive frames of data.

Fortunately Bayesian analysis is a natural framework for this sequential type of testing [30]. For data collected at time steps of size Δ apart and starting time t_0 Bayes theorem is:

$$f(\boldsymbol{\theta}|y_{t_0}) \propto f(y_{t_0}|\boldsymbol{\theta})f(\boldsymbol{\theta}). \tag{6.2}$$

Then when we consider our knowledge about θ after a sequentially collected data set at some future time $t_0 + \Delta$ [30] assuming that the likelihood $f(y|\theta)$ is *iid* in both measurements,

$$f(\boldsymbol{\theta}|y_{t_0}, y_{t_0+\Delta}) \propto f(y_{t_0}, y_{t_0+\Delta}|\boldsymbol{\theta})f(\boldsymbol{\theta}) \tag{6.3}$$

$$= f(y_{t_0}|\boldsymbol{\theta})f(y_{t_0+\Delta}|\boldsymbol{\theta})f(\boldsymbol{\theta})$$

$$\propto f(y_{t_0+\Delta}|\boldsymbol{\theta})f(\boldsymbol{\theta}|y_{t_0}). \tag{6.4}$$

The point is that one can treat the posterior distribution (last term in Equation 6.4) from the first sample as the prior for the next round of data collection. Thus, "chaining together" inference is nothing more than updating one's prior distribution of the parameter $\boldsymbol{\theta}$ as additional data is collected at the time steps $t_0, t_0+\Delta, t_0+2\Delta, ..., t_0+h\Delta$ for $h = 0, 1, \ldots$ under the assumption that the data are *iid* over the period of observation. In our case the "additional data" will have to be semi-synthetic data obtained from an AHI radiance cube that is "taken apart" via OLSTER, then a plume inserted to make y_{t_0}. The plume can then be spatially modified to simulate its drift/dispersion/cooling/atmosphere/etc. and then re-observed to create $y_{t_0+h\Delta}$. In a real scenario there would be a number of other factors that would also need to be accounted for, such as change in viewing geometry, sensor noise, registration of the images, changes in sensor calibration, etc. Using semi-synthetic images circumvents these issues and represents the opportunity to explore the upper limit on the performance of this technique.

Using Physics to Inform the Regression Process

Another unique approach to including physics-based modeling results in this problem, that the author explored but ultimately abandoned, is called a radiometrically-informed design matrix (RIDM) which instead of using a leading column of ones appended to the gas library (design matrix) one appends a blackbody (BB) curve (Equation 6.6) at some temperature T as the leading column. This is useful in this application where the user has some physical knowledge of the (clearly) non-uniform underlying wavelength distribution of energy of the surface beneath the gas plume. This is best explained by thinking about the intent of least squares (in the mathematical sense) as the projection of the data (observed radiance) onto a set of basis vectors (gas library) and minimizing the L_2 norm of the residual. From physics, we know that the data (for most natural materials with high emissivities) are dominated by (near) blackbody behavior, which motivates its explicit inclusion as a basis vector instead of the **1** vector. From a statistical perspective, the motivation of the intercept term is to capture the mean of the data if all the regression coefficients were zero (i.e. for a non-plume pixel), and again, the physical default for LWIR spectral data will be a black body curve, not a vector of

ones.

The idea was discarded because it did not given any increase in the separability of gases that were detected, nor the ability to detect them at any lower concentration levels. The author attributes this phenomena to the fact that was controls the standard error of the regression coefficients is the equation

$$MSE = \frac{(Y - X\beta)'(Y - X\beta)}{n - rank(X)}.$$

Recall that the convergence of the indicator variables relies upon first having "good" estimates of the regression parameters at each iteration. Note also that the maximum *a posterior* parameter estimates with the conjugate MVN/Gamma model are identical to the MLE's obtained under the Gauss-Markov model.

The important subtlety in this is that for the scenario where one is using semi-synthetic imagery, that one is working with the linear model $Y = X\beta$ <u>not</u> $Y = X\beta + \epsilon$. This means that any contribution to MSE (for a linear model applied to this data) is not going to be driven by some error process, but rather by the rank-deficiency (assuming convergence) of $\boldsymbol{X}$ in the denominator of the formula for MSE.

Alternate Performance Assessment Methods

In the context of GVS which simultaneously estimates model probability, variable inclusion probability, and inference on a regression coefficient; the above approach lacks the subtlety required to incorporate all of these (inherently subjective) tasks. Further, when one considers the potential of spatial and sequential aggregation, the task of quantifying the performance of the identification task becomes more difficult. Recall, that part of the motivation for this entire class of problems is mitigation of the plight of "mono-numeroucis" (the faulty notion that a single statistic can completely describe the effectiveness of a filter applied to any task, on any data, under any circumstances) that comes about by simply choosing "the best" fitting model and declaring it correct. Accordingly, performance validation for model selection should be appropriately subtle.

The output of the proposed methodology includes: model probabilities (polychotomous), variable inclusion probabilities (dichotomous), and coefficient values (continuous) all of which have practical significance to the analyst. Further, it is possible to rank order the models in terms of their posterior probability (noting that in reality only

one of them is correct), but quantifying the performance of the regression equations either requires some distance measurement from a point estimate to the true value or determination of whether or not a $(1 - \alpha)$ credible interval covers the true value.

In general one would prefer an assessment scheme that can be flexible enough to measure inference on all three of these parameters and their associated probabilities $p(\hat{M}_r), p(\hat{\gamma}_l), f(\hat{\beta}_l)$ for every model and associated combination of parameters. For the task of model selection, one generally would choose the model based on the ranked order of their probabilities. Thus one can easily create order statistics for each model $p(M_{(1)}) \geq p(M_{(2)}) \geq \ldots$ such that they are ranked as the most likely, followed by the second most likely, etc. Of course one would prefer the correct model M_{k*} to have the lowest possible index, and the order statistics of the M_{k*} ought to be captured in the performance evaluation. Given practical computing constraints the computation of the model probabilities, albeit desirable, might be untenable due to the number of MCMC iterations that would be required. Using suboptimal Matlab code on a modern laptop, computing 2^{31} iterations of the MC would require approximately 24.5 days. Fortunately the model probabilities are the least important output to estimate and unless substantial computational advances are made in the time required to simulate from MCs, the author does not intend to assess this type of performance.

One simple way to quantify the performance of the indicator variable selection is the use of L_1 or L_2 vector norms to measure the distance between the true vector of indicator variables $\boldsymbol{\gamma}^*$ and the estimated $\Pr(\boldsymbol{\gamma}|x)$. Similarly these vector norms can be used to measure the distance between the posterior mode (or mean) for the regression coefficients $\max_l Pr(\boldsymbol{\beta}_l|data)$ and the true values $\boldsymbol{\beta}^*$. Alternatively, because the joint posterior distribution is available for the regression coefficients, one can integrate out the posterior density for each parameter and find the probability of the simultaneous pairwise credible intervals that contain all of the true values (assuming symmetric distributions) using

$$1 - \int \ldots \int_{-\max |\beta_l^*|}^{\max |\beta_l^*|} f(\beta_1, \beta_2, \ldots, \beta_l|x)dx. \tag{6.5}$$

This approach would have to be modified if any of the priors or likelihoods were asymmetric.

Finally, there is the area of validation of the results that is rich for exploration.

First, the question of what potential exists for using "hold one out" cross validation on the plume imagery, either hold out spatial pixels or spectral layers, to determine the spatial/spectral identification sensitivity of these MCMC-based techniques?

Alternative (albeit computationally very expensive) approaches include computing bootstrap a confidence intervals for any of the performance metrics used to quantify identification performance. This is in addition to using bootstrapping to mitigate the impact of location where a plume was implanted into an images, or the order in which SAM-GVS stepped across the pixels.

Modeling and Sensor-Design Applications of M-GVS

Using the common framework of statistical power, the author discovered the grounds for a method of performance prediction that allows the user to judge, based on an actual gas library (at a given spectral resolution, condition number, etc), the upper threshold of performance of any regression technique used for gas identification. This can then be quantified (for specific scenarios) in terms of either: minimum required SNR, minimum required contrast temperature, or minimum required CPL or optical transmission.

This method is based on the notion of statistical power, where one simulates data, and in applies (any sort of) regression method, and tests if the fitted coefficients are statistically different from zero. The point being that this allows for an analyst to be able to compare regression methodologies, perform mission tasking, or even sensor design studies, with a real gas library at a given spectral resolution.

How this would work in practice is for a given gas library (at a fixed spectral resolution) one would simulate data at a given SNR and simulate the statistical power function for every gas at that combination of concentration of gas, SNR, and spectral resolution. Then repeat this for all possible possible combinations of the parameter set and apply either a response surface methodology or a search algorithm such as gradient-descent in order to identify the maximum possible power subject to the other constraints.

An example of this methodology (using the AHI 50 band full library) is provided for the linear model with C observations and N covariates

$$Y = X\beta + \epsilon \tag{6.6}$$

where Y is $n \times 1$, X is $C \times (N+1)$, β is $(N+1) \times 1$, and ϵ is $C \times 1$.

$$\epsilon = iid\, N(0, \sigma^2) \tag{6.7}$$

The "standard frequentist" t-test for coefficients under the hypotheses:

$$Ho : \beta_j = 0$$

$$Ha : \beta_j \neq 0$$

is

$$T^* = \frac{\hat{\beta}_j - \beta_j}{\sqrt{MSE}\sqrt{(X'X)^{-1}}} \sim t_{n-p}\ under\ Ho. \tag{6.8}$$

So the naive power of this test (putting aside the multiple comparisons issue) is defined as:

$$power(\theta) = power(\sigma^2) = \text{fraction of times Ho is rejected based on test statistic } T^*$$

Generating fake data using a user-specified set of gases (predictors), one can generate data via Equation 6.6 for some range of σ^2 values. Then for M iterations, one simply needs to find the fraction of samples that for the correct predictors (the ones that were non-zero), Ho is rejected at the two-tailed 95% confidence level. So as an example, lets look at the case where we have four non-zero coefficients in the true model:

For each iteration,

1. Take $\beta_4^T = [1, 1, 1, 1, 0, \ldots, 0]$ and create several different sets of data, one for each levels of σ^2

2. Do OLS regression using the randomly generated X and Y, for each level of σ^2, collect the vector $\hat{\beta}$, but look only at the first four values (where the true $\beta = 1$)

3. Record the number of those first i estimated coefficients that reject Ho (according to Equation 6.8)

4. Add up all of the rejections of (Ho = 0), only for the variables that were really generated with $\beta = 1$, but where all predictors were in the model

Now compute power (for this one combination of parameters) by

$$Power(\sigma^2) = \frac{\text{total number of rejections for first } i \text{ variables} \,|\sigma^2}{M}.$$

This method can be extended to compensate for multiple comparisons as well, but the basic premise is that one can determine before a sensor is built, or before a mission is flown, what the power is expected to be for a given gas (or set of gases) and this might be used to guide those processes.

Physics-related Bayesian Inference

There exists the potential for the inclusion of scene-derived radiance parameters (i.e. surface emissivity $\boldsymbol{\xi}_s$ and temperature T_s) in a "full Bayesian" context in the regression model. Because it is not possible to write the full conditional densities for the parameters in the sensor-reaching radiance model, it will most likely require the use of a reverse-jump Metropolis-Hastings approach to the MCMC regression[31]. This would involve writing a prior distribution for every term except the absorption spectra in Equation 3.12 and carrying out inference on the integrated CPLs. This is admissible, up to the point that it limits ones ability to make inference on the plume temperature, because that term is no longer strictly identifiable as it is captured inside another function.

This fully Bayesian approach is probably best suited for the single pixel problem, but would in theory enable the user to conduct inference on the plume temperature as well as the CPL and probability of inclusion of each gas in the library. The obvious challenges of doing this include the need for ground truthed data (possibly from a synthetic scene generation tool such as DIRSIG) and the increased computational burden and convergence challenge of adding this layer of complexity to the regression model.

Realistic Plume Generation

For all of the SAM-GVS results presented previous, a fixed spatial plume template was utilized to define the spatial outline of the plume. There is significant potential for using a method that more closely mimics real gas plume, at least in terms of spatial

morphology. The following paragraphs describe a process whereby one could add this additional degree of realism to the plume implantation process.

The process of implanting the plumes into the semi-synthetic AHI images could proceed as follows:

- Start with the physical estimates obtained through OLSTER: $\boldsymbol{\tau}_{atm}$, $\boldsymbol{L}_u$, $\boldsymbol{L}_d$, $\boldsymbol{T}_s$, and $\boldsymbol{\xi}_s$.

- Generation of a simple empirical spatial plume template (see Figure 6.2 for one example) that has a maximum of unity and a roughly exponential intensity falloff. Because it is scaled to have a maximum value of one, it can be thought of as a dimensionless "concentration factor" and is intended to be proportional to a physical CPL C_l.

- The manipulation of the basic template via affine transformations to simulate the effects of different wind speed and direction (see Figures 6.3 and 6.4)

- Once the plume dispersion/drift have been simulated, the plume template is multiplied by the desired mixture of gas absorption spectra and then combined (via Equation 3.12) to generate what the image would have looked like to the sensor, with that plume in it.

The affine transformation uses a transformation matrix to rotate (by an angle ϕ) and scale (by ρ_x and ρ_y) an input image $[x, y, 1]^T$ into a new coordinate system $[x^*, y^*, 1]^T$. This is accomplished via

$$\begin{bmatrix} x^* \\ y^* \\ 1 \end{bmatrix} = \begin{bmatrix} \rho_x cos\phi & -sin\phi & 0 \\ \rho_y sin\phi & cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.$$

The output values of the process are then interpolated bi-linearly (using Matlab's built-in `imtransform` function) to generate the images shown in Figures 6.3 and 6.4. Note, this process requires careful user oversight, in order for the generated plumes to be of appropriate size and shape so as to appear somewhat realistic.

Although this approach is clearly ad hoc, and lacks the ability to tie in with physical input parameters (wind speed, conservation of plume mass, etc), this plume implantation procedure is an effective way to rapidly modify and insert plumes for the purpose
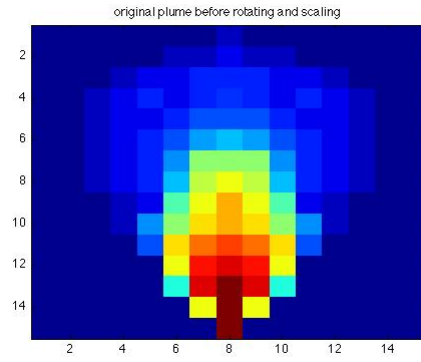
Figure 6.2: Basic plume template, max = 1
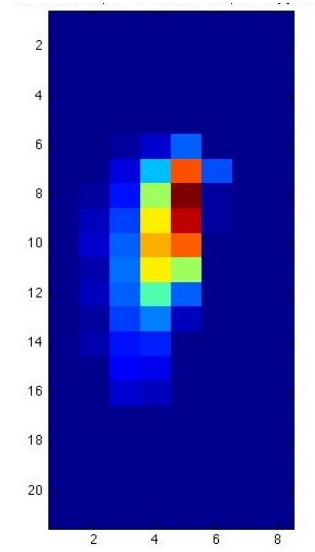
Figure 6.3: Plume template rotated for a wind direction of 020 and x,y scaling (0.3 , 1)
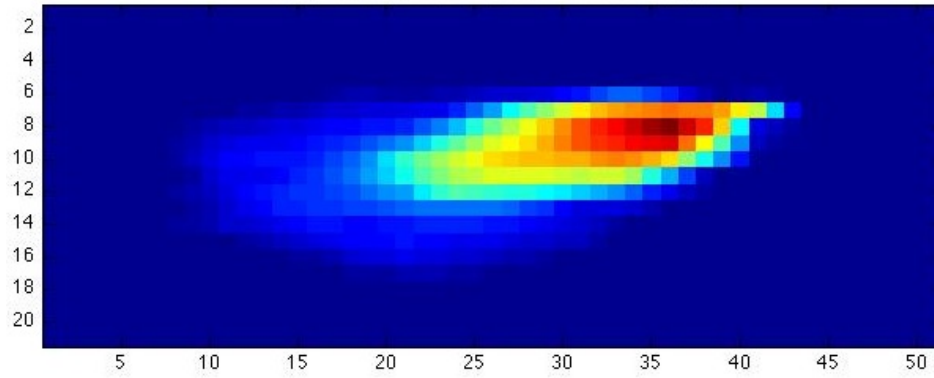
Figure 6.4: Plume template rotated for a wind direction of 060 and x,y scaling (2.5 , 1)

of algorithm development. In the end, the author elected to use a fixed plume spatial template for the sake of simplicity, but the methodology for much more complicated plume shapes was developed.

Chapter 7

Appendix

This chapter contains details on several underlying pieces of technology that are critical to this project. They include:

- A discussion of why MCMC can be used for Bayesian inference

- An example of GVS used on a simple regression problem

Background on MCMC

Recall that conducting Bayesian inference is usually accomplished not by simulating from the posterior $f(\Theta|x)$ itself, but rather from the "state space" of a MC that has the same limiting distribution as the posterior.

   The reason this works has to do with graph theory, as applied to directed acyclic graphs (DAGs). If one places at the nodes, random distributions that represent the parameter space for a Bayesian model that are conditionally dependent only on their parents (and thus conditionally *independent* of its other ancestors), then one can show [32] that a MC with those same nodes will in the limit (as sample size goes to infinity) have the same distribution of transition states as the posterior. It is this notion that provides the necessary and sufficient conditions for DAGs to have corresponding irreducible, aperiodic, and persistent MCs that allow Bayesian inference to take place. All that is in practice required is for the user to be able to draw the appropriate DAG for their problem, and be able to iterate the corresponding MC until an ergodicity is achieved for each of the parameters. In the case of ordinary multiple linear regression,

the DAG would look something like the schematic shown in Figure 7.1. In Figure 7.1 the β nodes represent the regression coefficients, μ represents the mean of the data, $\boldsymbol{X}$ the design matrix, τ the precision (variance^{-1}), and $\boldsymbol{Y}$ the observed data.
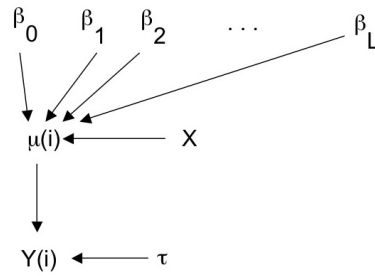
Figure 7.1: Directed acyclic graph for multiple linear regression

The case of GVS applied to regression is only slightly more complicated by the addition of the indicator variables γ_l as shown in Figure 7.2.
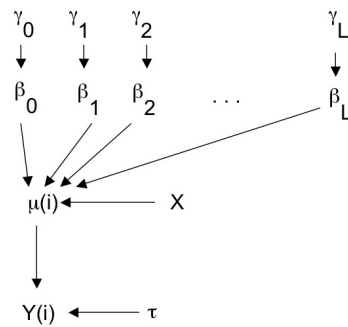
Figure 7.2: Directed acyclic graph for GVS

At its essence, using iterative techniques for approximating posterior distributions is not entirely unlike the method one would use to solve

$$x = sin(x).$$

The solution is obtained by inserting an initial estimate into one of the two terms and then one iterates until the LHS and RHS match. Alternatively one could restate the problem as solving

$$f(x) = f(x').$$

Where the right hand side (RHS) is the same as left hand side (LHS) are identical functions, except the value where the functions are being evaluated is different by a small quantity (x versus x'). The intent being that after sufficient iteration, one could make the difference between the LHS and RHS to be be negligibly small, thus achieving convergence.

Bayesian inference via MCMC makes use of this concept of iterating the LHS and RHS of an equality relationship until they are essentially equal. The only difference is that instead of a fixed constant being inserted on either side, one inserts realizations from a distribution, and the LHS/RHS convergence is in terms of the *distributions*, not specific values. This is accomplished explicitly via Bayes rule

$$f(\boldsymbol{\theta}|x) \propto f(x|\boldsymbol{\theta})f(\boldsymbol{\theta}) \tag{7.1}$$

and all that is required is sufficient iteration until the LHS distribution has stabilized. As mentioned previously, frequently one will discard some number of the starting values of the MC in order to minimize the dependence on the starting values of the chain, this is known as allowing the chain to "burn in."

GVS Example

As a practical example of how Bayesian inference is conducted, specifically demonstrating MCMC and GVS, this section shows the process (with an emphasis on how a user would interpret the results). This section is important because it illustrates a process that is often highly dependent on user interaction, while the bulk of this thesis was designed so that the absolute minimum of user-interaction is involved.

The simplest implementation of GVS is under the assumptions that the covariates $\boldsymbol{\beta}_l$ are independent and non-informative. A natural choice for this parameterization would be the use of a product of independent normal distribution with some large variance

$$\beta_j \sim \ iid \ N(0, \sigma = 10^3) \ \rightarrow \ \boldsymbol{\beta} \sim \ iid \ MVN(\mathbf{0}, \sigma\boldsymbol{I}).$$

This prior has the advantage of integrating to unity (and thus is "proper") but is effectively uniform over the majority of the parameter space of interest. Alternatively one could have used $Uniform(-\infty, \infty)$ but this prior does not have unit area and thus The initial pdfs for the β_l's are shown in Figure 7.3. Note that in statistics common convention for a normal distribution is to state its mean and either standard deviation σ or variance σ^2. In Bayesian statistics, it is common practice to use units of precision $\tau = 1/\sigma^2$, although the author has made the effort to translate all results into units of σ for ease of understanding.
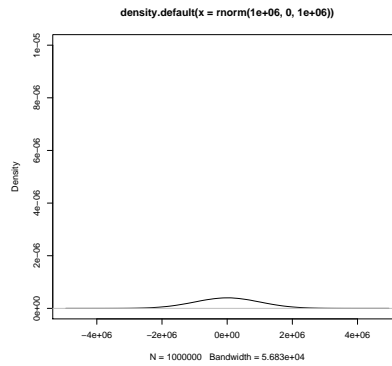
Figure 7.3: Univariate uninformative prior for β_l

The priors for each of the indicator variables γ_l are $Bernoulli(.5)$ meaning that

there is an equal chance a priori that each of the variables are included. The prior probability mass function (pmf) for the indicator variables are shown in Figure 7.4.
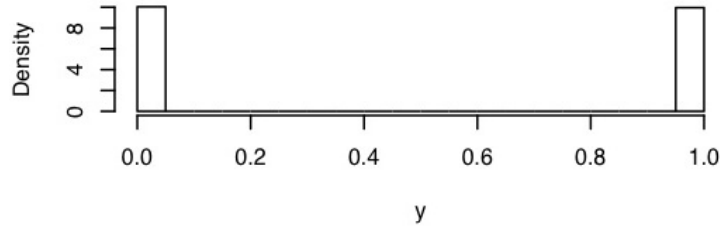
Figure 7.4: Univariate uninformative priors for γ_l

One implementation of GVS has been in an open source software environment called JAGS (Just Another Gibbs Sampler) and it implements in C the actual chain generation and sampling. To demonstrate the practical utility of this technique on a known set of data, the author generated a set of data based on a mixture of three gasses from the full gas library according to

$$y = 5 + c * gas(3) + c * gas(12) + c * gas(14) + \epsilon \quad \epsilon \sim iid \ N(0,1). \qquad (7.2)$$

This scheme allows the user to test a variety of implantation strengths c and compare the effectiveness of algorithms at correctly identifying which gases are in the plume. Graphically the linear combination of the spectra and error terms are shown in Figure 7.5.

What follows is a (lengthy) set of plots showing the output of the GVS procedure, the posterior distributions $f(\boldsymbol{\theta}|x)$. This run made use of the OLS estimates used as pseudo-priors for the variables *not* included in the model, and the design matrix was standardized. This example was generated using the first 29 of the 31 available gas spectra in order to maintain unbiased estimates and reduce the rank deficiency of $X'X$. Recall that the input was a set of 29 *iid* $N(0, \sigma = 10^3)$ distributions for the priors on each of the regression coefficients β_l and a set of 29 $Bernoulli(0.5)$ distributions as priors on the indicator variables γ_l. The set of posteriors is shown in Figures 7.22 through 7.37, these qualitatively demonstrate the output of Bayesian inference (note the true value for $\beta_3, \beta_{12}, \beta_{14} = 0.5$). Quantitatively, one can also look at the quantiles of the
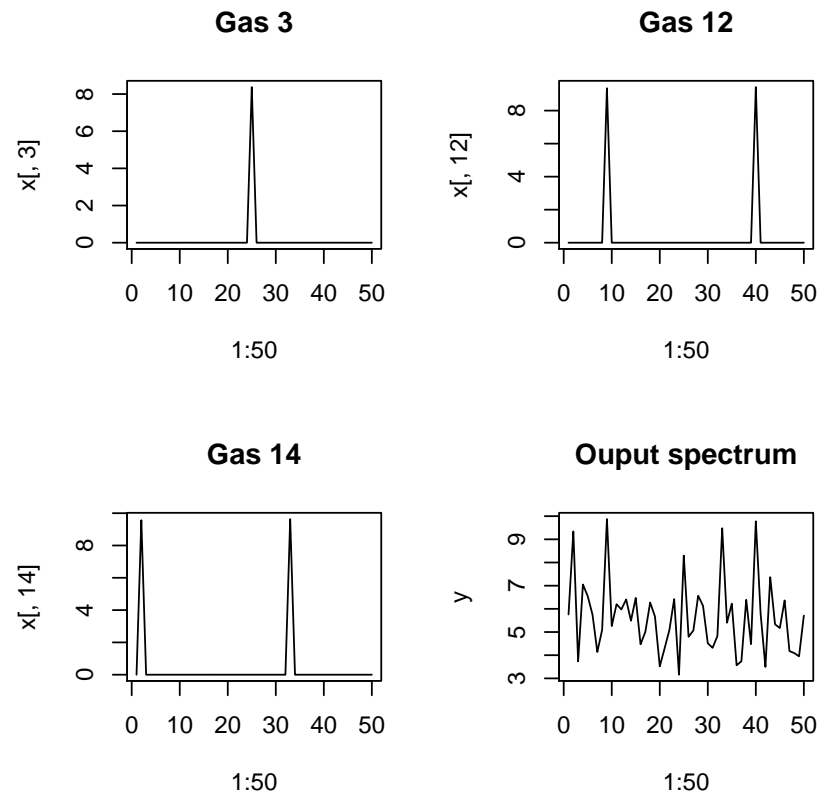
Figure 7.5: Spectra used to create test case for GVS

posteriors, highest probability density (HPD) regions, and the posterior modes (most likely value). For the sake of brevity only the summary statistics are shown for one of the four chains. Each chain had different starting parameters and was iterated 8192 times. 8192 was chosen because $2^{13} = 8192$ for optimum performance when calculating the chain autocorrelation functions (ACFs) via the Fast Fourier Transform (FFT). The $b[l]$ values are the regression coefficients, $beta[l]$ are the standardized regression coefficients, and $g[l]$ are the indicator variables. Once the posteriors are in hand, it is straightforward to compute any summary statistic of interest.

Overall it should be made clear that Bayesian inference, while possible to automate to a degree, still requires a fair amount of user interaction to monitor the process for correct operation and carefully assess its performance. In the following set of plots of MCMC output, one should qualitatively note the following:

- This is only a single chain, run on a single pixel. It should be thought of as a simple demonstration of the technique, not as an authoritative statement on GVS, Bayesian statistics, linear modeling, or synthetic data generation. Only carefully designed simulation studies, run many times over can begin to describe the long run statistical properties of any method. Just as in hyperspectral target detection, whether or not a detector is able to find a target in any single pixel is not really an important issue, rather it is the scene-wide and really the long run behavior over many scenes that is of interest.

- First look at the trace plots for each of the parameters. Upon convergence they should all be an ergodic process and be indistinguishable from uncorrelated (white) noise. This is one of the first places where it will become obvious that a chain is mixing poorly.

- The next place to look for good chain mixing is the chain autocorrelation plots. Ideally one would like to see unit auto-correlation at the origin and very low autocorrelation with all other step sizes. In this regard, this set of ACFs is as good as can be practically expected.

- The posteriors for b[3], b[12] and b[14] are more sharply peaked, and not centered at the origin. These pdfs indicate the post-data view of the value of these parameters, the interpretations being that one is more certain about the location

of the b[3], b[12] and b[14] than the other predictors, and that that location is non-zero. In this case the most likely value is around 0.5, recall that $c = 0.05$ was the true value used in this example.

- Similarly the posteriors for g[3], g[12] and g[14] are not equally distributed on the values zero and one, like the rest of the posteriors. These pmfs indicate the probability of inclusion in the model of these variables.

- The pdf for b[3] and pmf for g[3] is not as clearly indicative of this variable being non-zero and included in the model (respectively). This could be caused by a number of factors including: insufficient iteration, lack of convergence, rank deficiency of the design matrix, etc. One should also bear in mind that the results shown are the output of only one MC, with only one instantiation of the error process in Equation 7.2. Ideally one would repeat this for many simulated data vectors, each with multiple chains, and each chain having different starting values in order to truly show the potential of the technique.

- Examining posterior plots such as these is a great way to quickly observe the behavior of the chain, but actual inference is usually done with summary statistics such as the posterior mean, posterior variance, posterior quantiles, etc. An example of common summary statistics is included for the chain at the end of this Appendix.
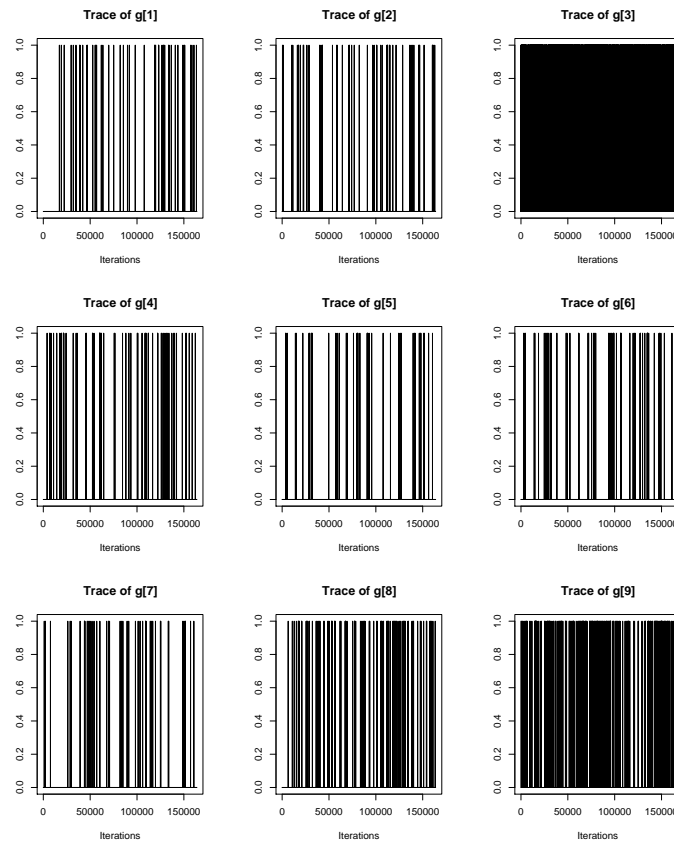
Figure 7.6: Trace plots from GVS
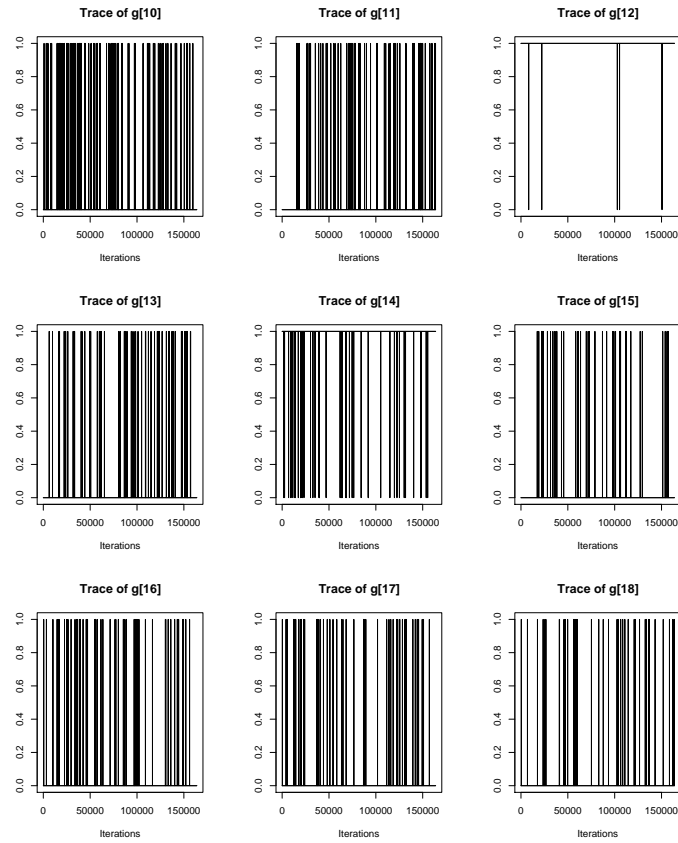
Figure 7.7: Trace plots from GVS

Trace of g[19]   Trace of g[20]   Trace of g[21]

Trace of g[22]   Trace of g[23]   Trace of g[24]

Trace of g[25]   Trace of g[26]   Trace of g[27]
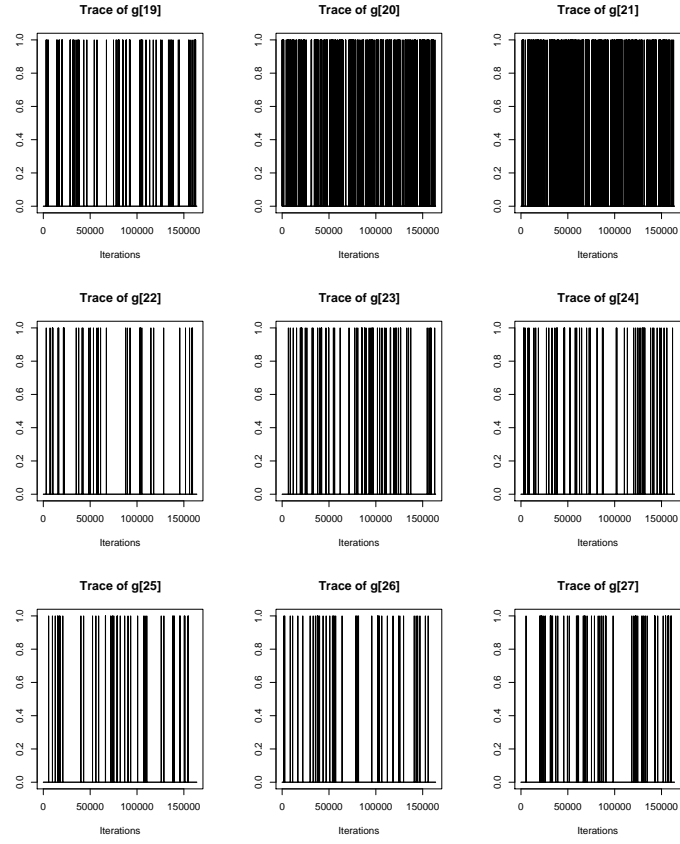
Figure 7.8: Trace plots from GVS

Trace of g[28]   Trace of g[29]
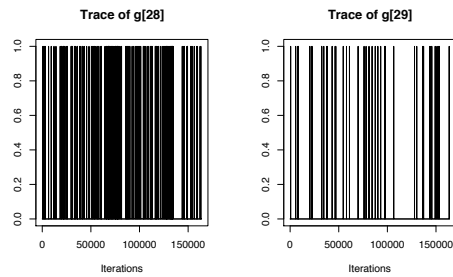
Figure 7.9: Trace plots from GVS
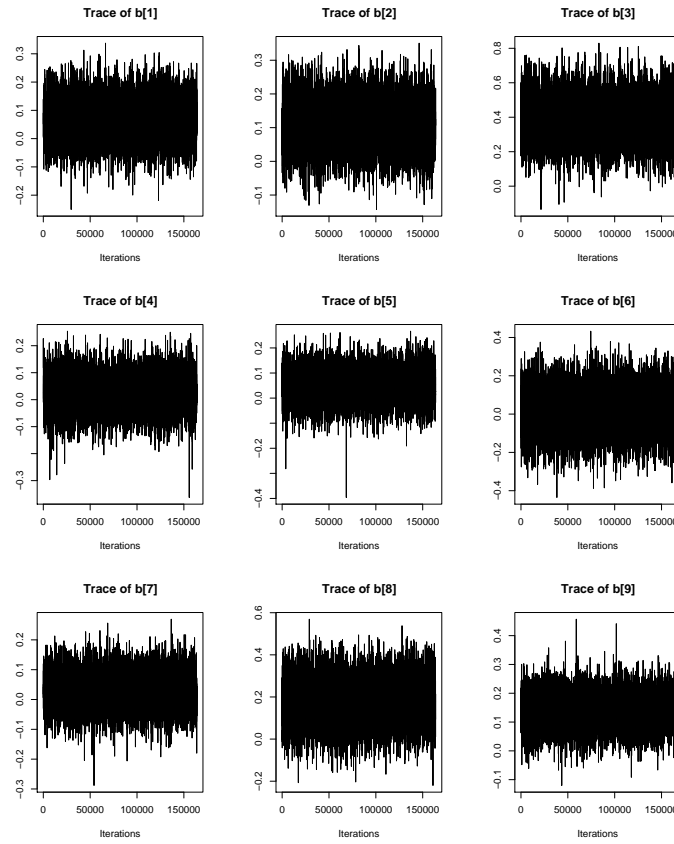
Figure 7.10: Trace plots from GVS
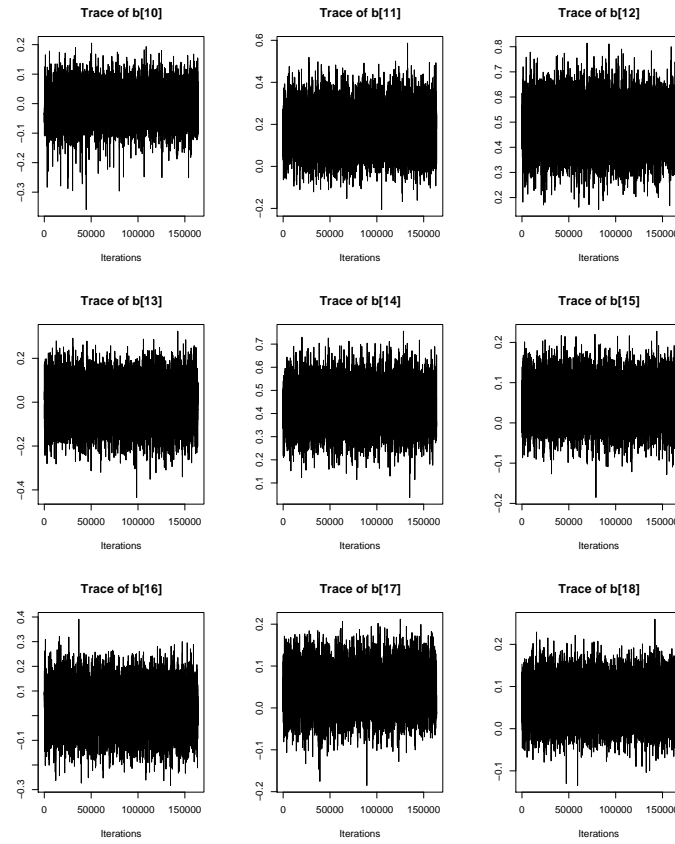
Figure 7.11: Trace plots from GVS
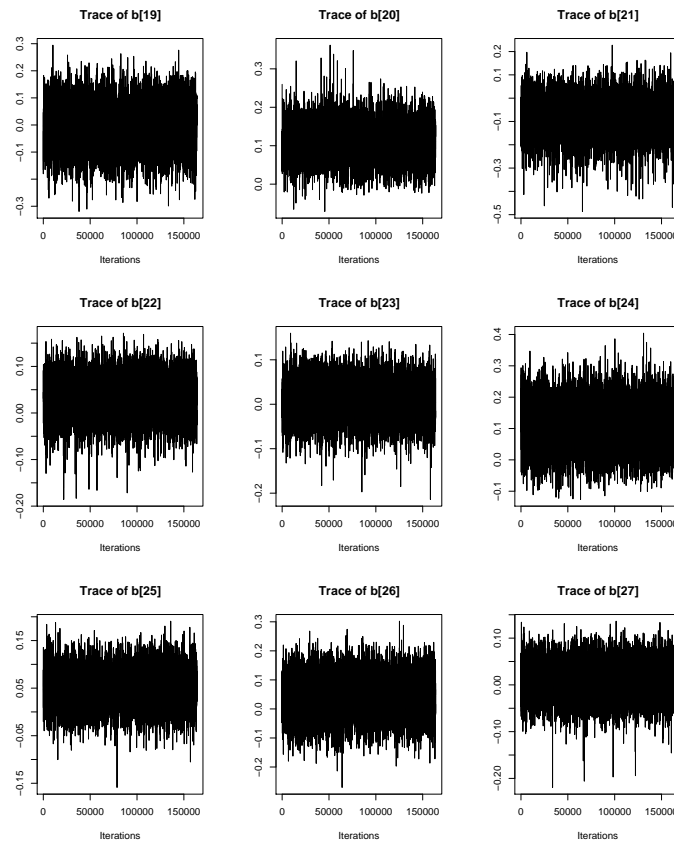
Figure 7.12: Trace plots from GVS
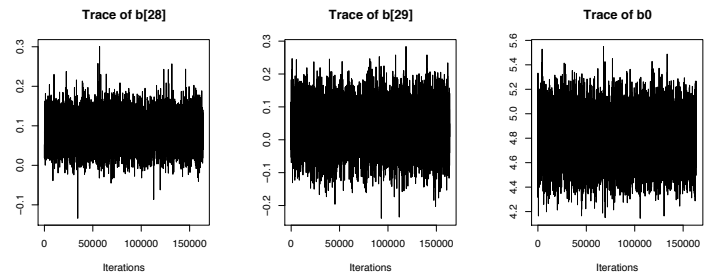
Figure 7.13: Trace plots from GVS
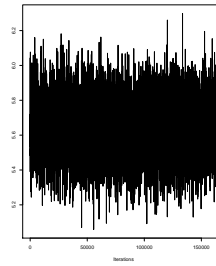
Figure 7.14: Trace plot of tau from GVS
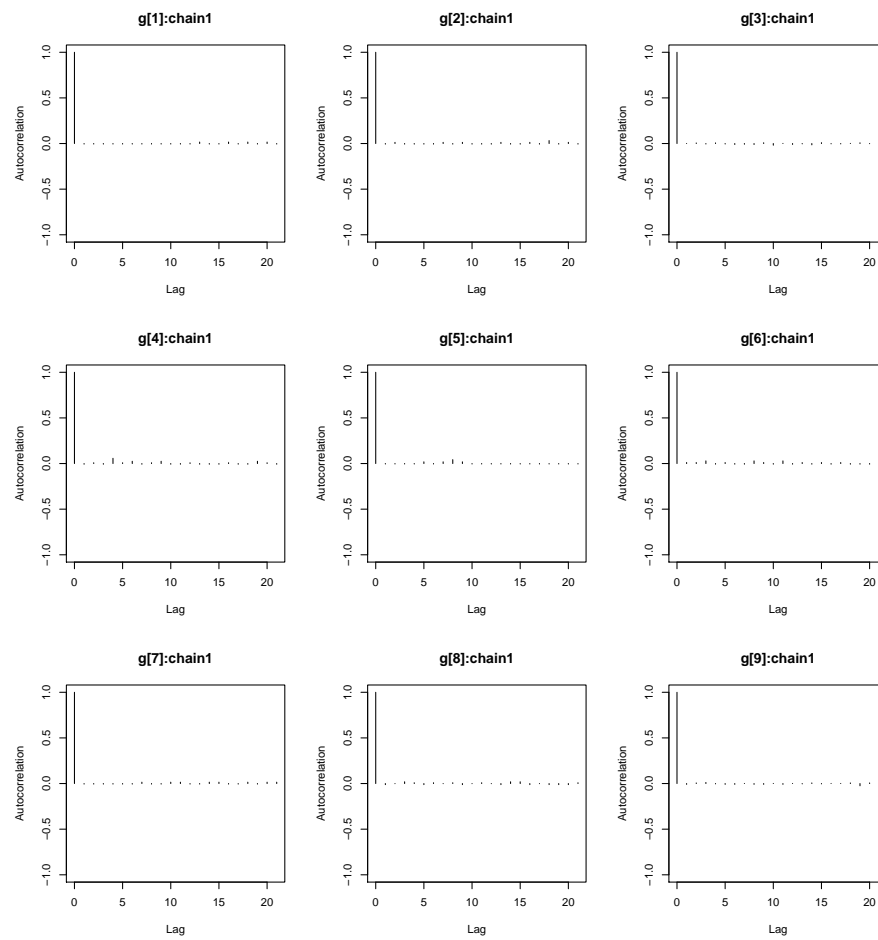
Figure 7.15: ACF plots from GVS
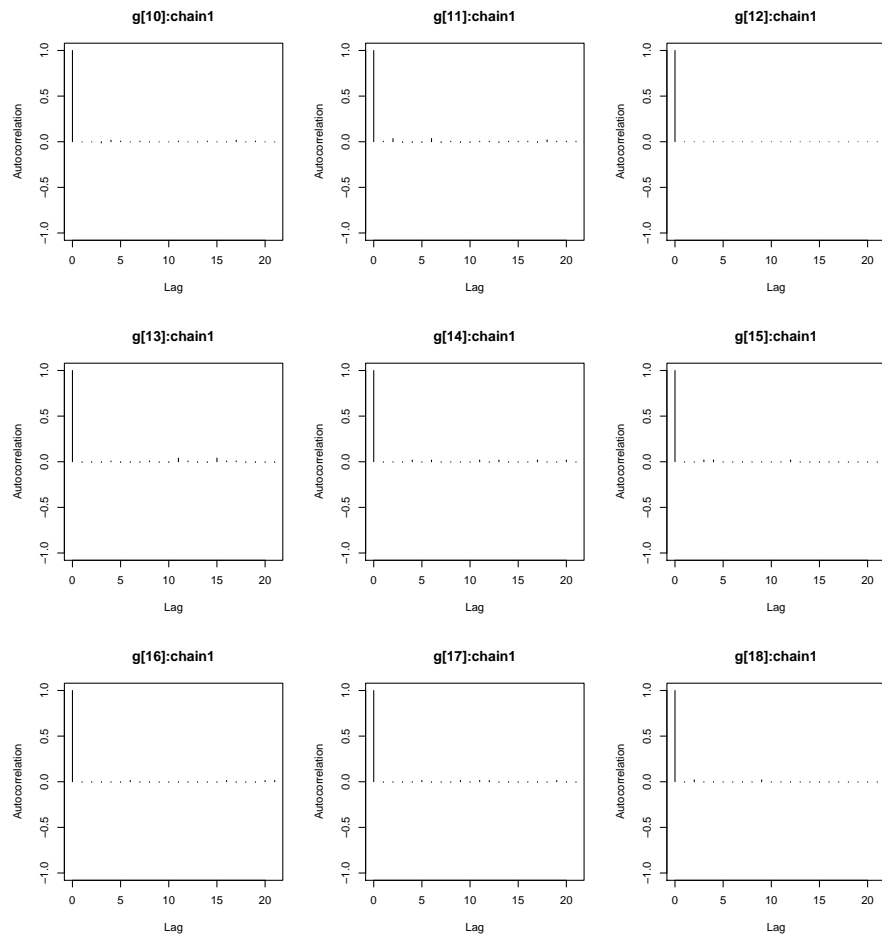
Figure 7.16: ACF plots from GVS
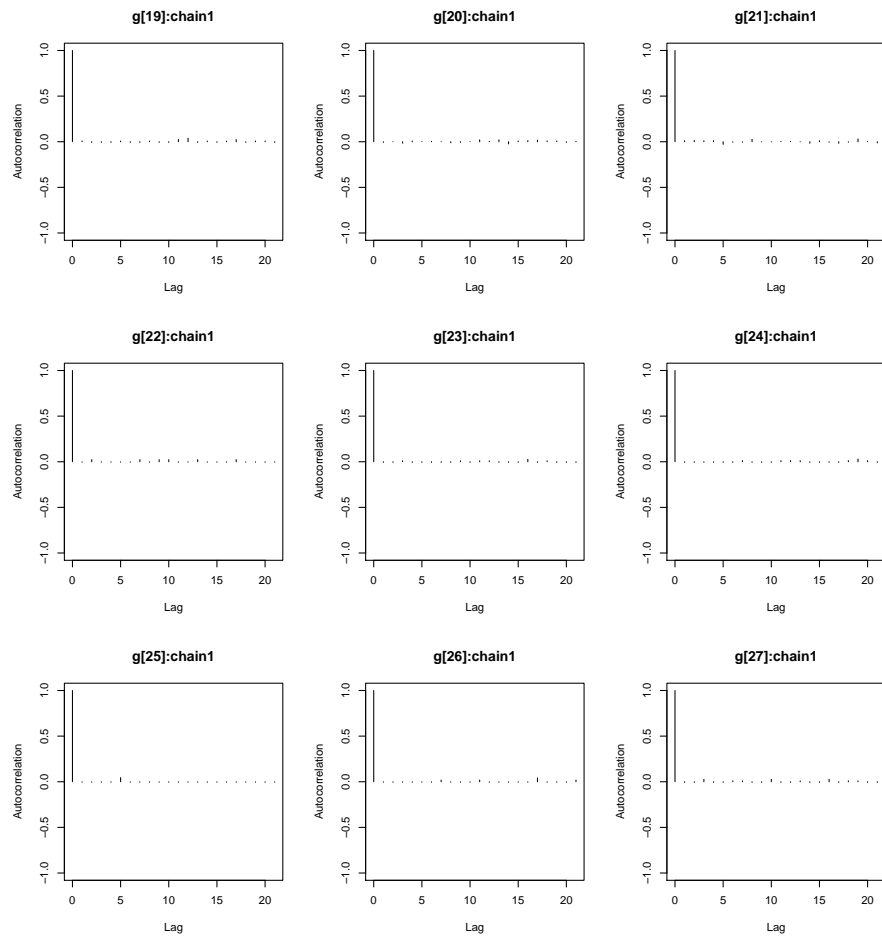
Figure 7.17: ACF plots from GVS
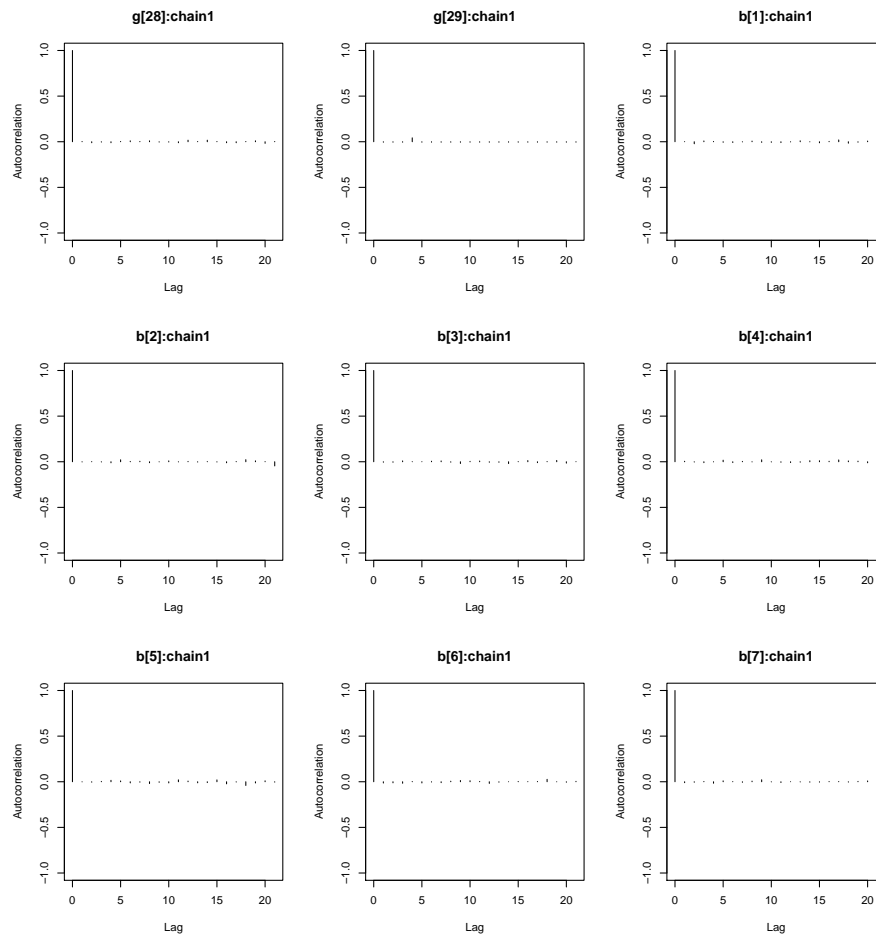
Figure 7.18: ACF plots from GVS
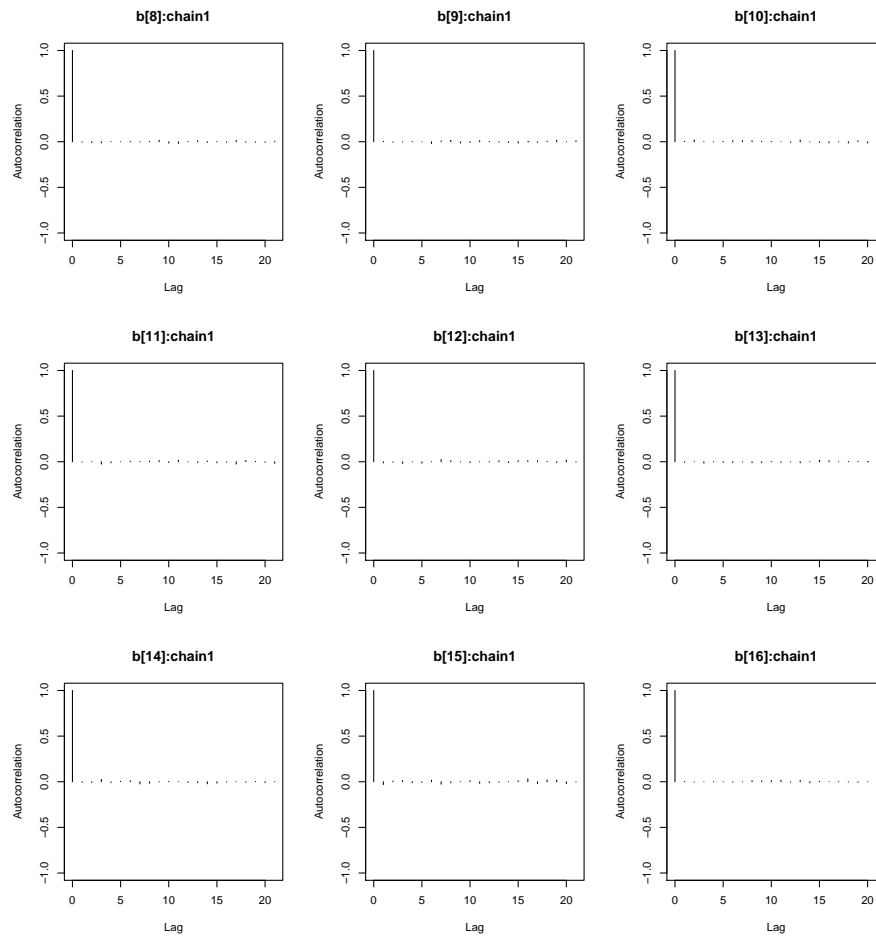
Figure 7.19: ACF plots from GVS
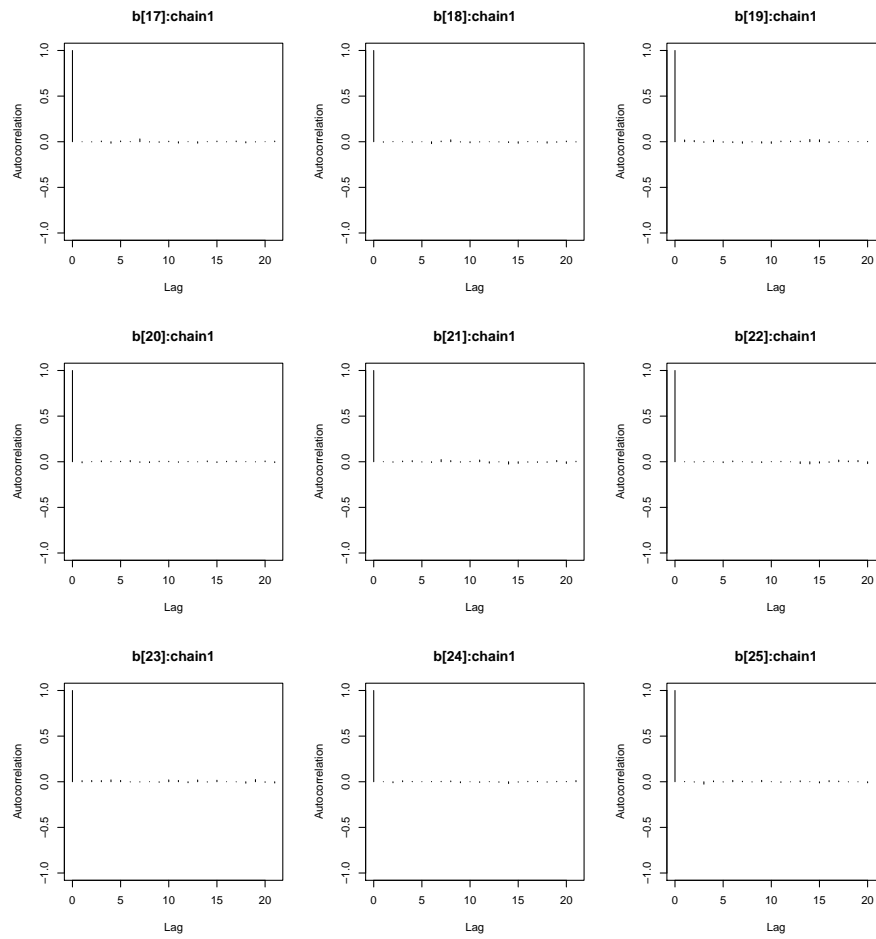
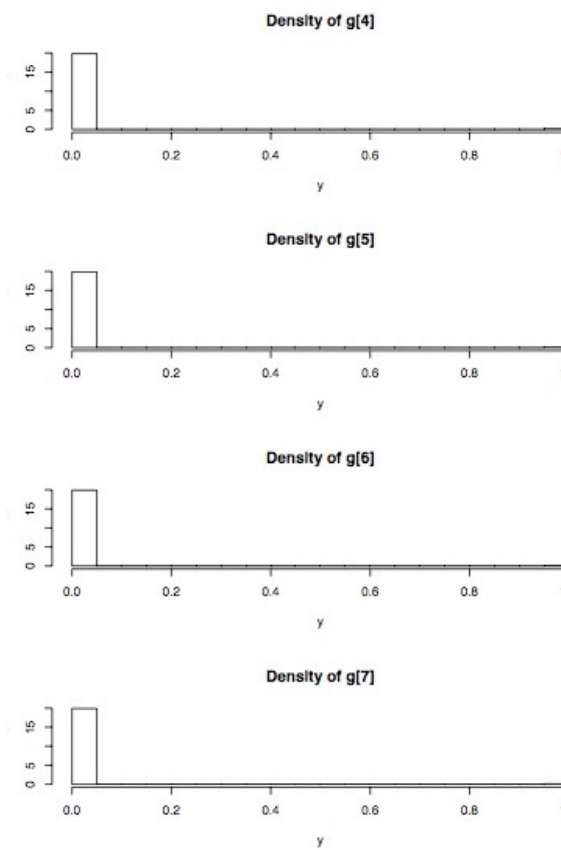Figure 7.20: ACF plots from GVS
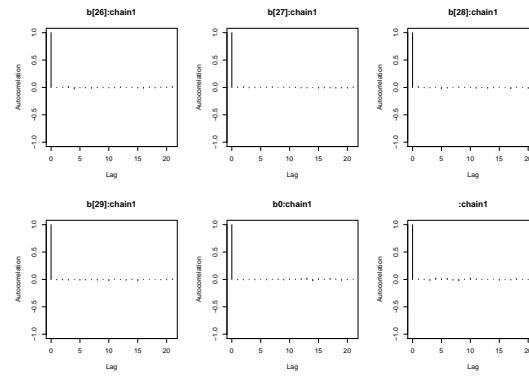
Figure 7.21: ACF plots from GVS
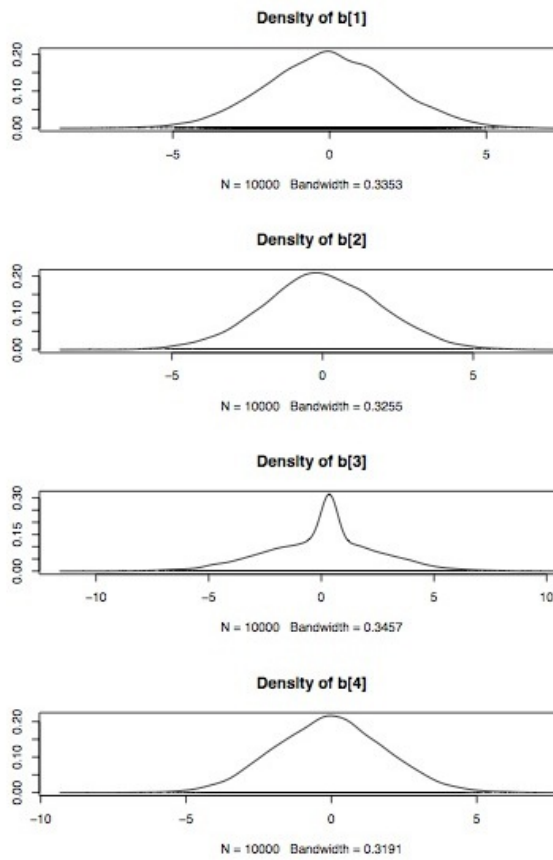
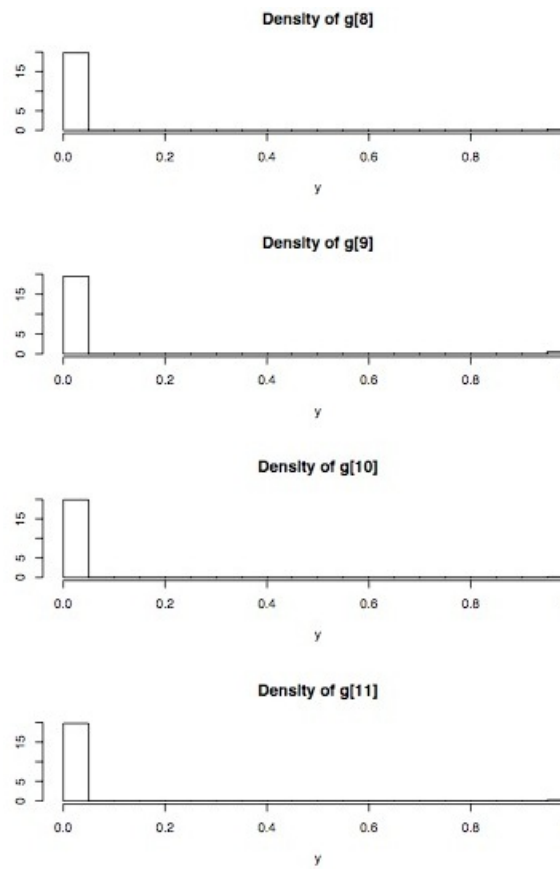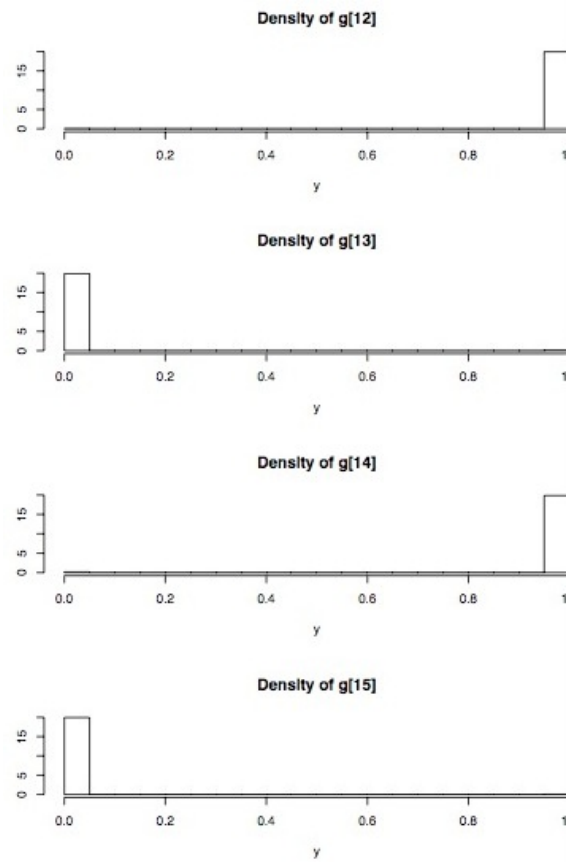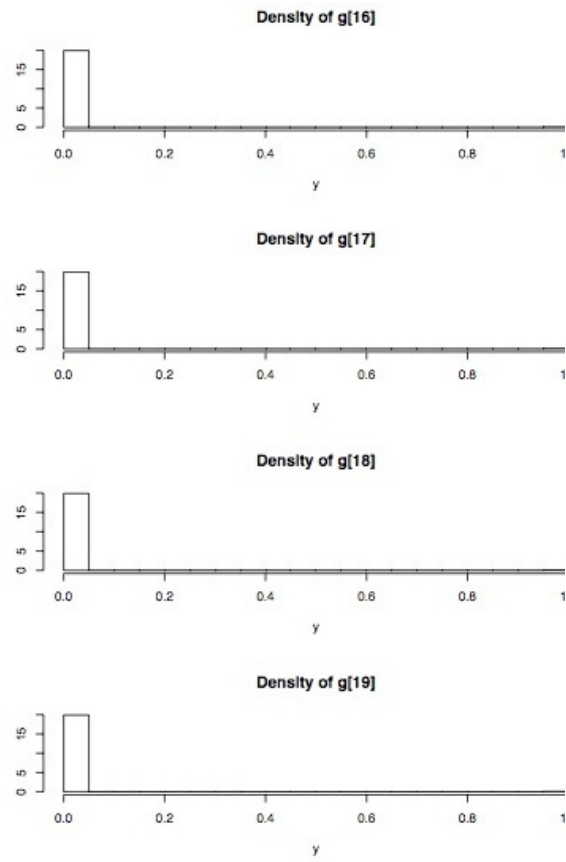Figure 7.22: Posteriors estimated from GVS
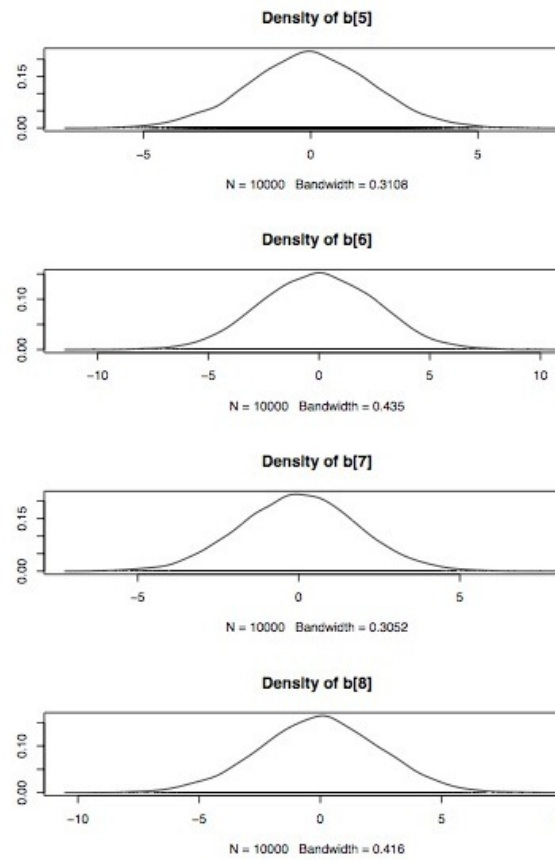
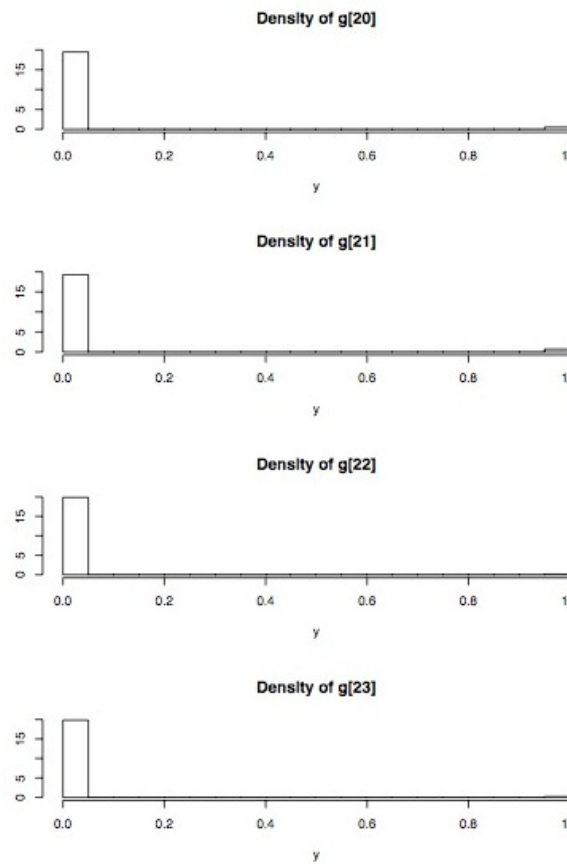Figure 7.23: Posteriors estimated from GVS
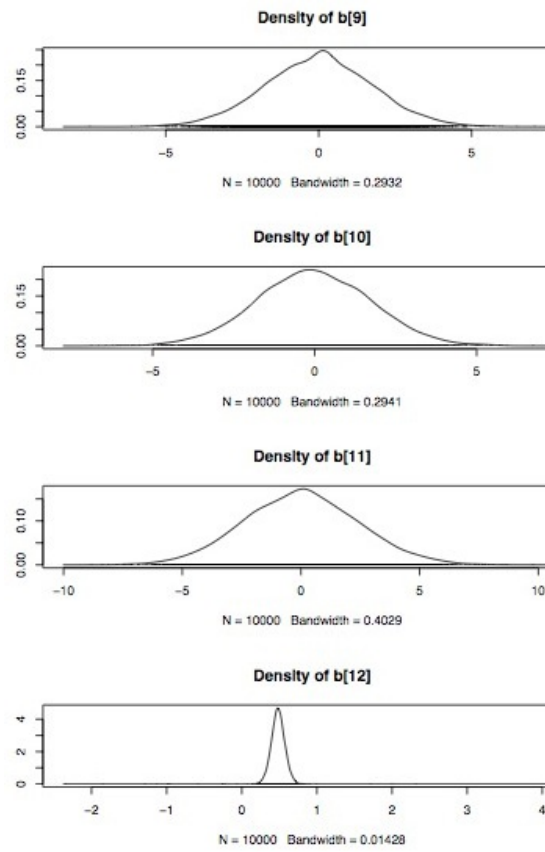
Figure 7.24: Posteriors estimated from GVS
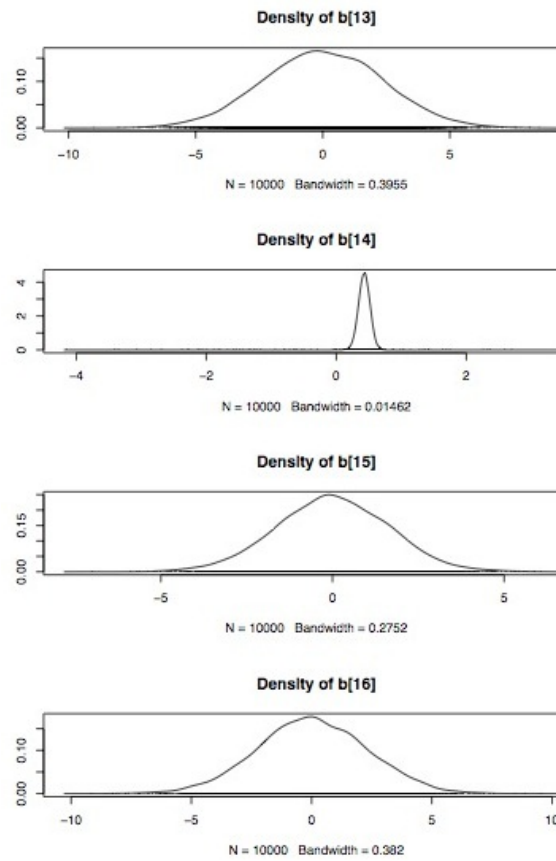
Figure 7.25: Posteriors estimated from GVS
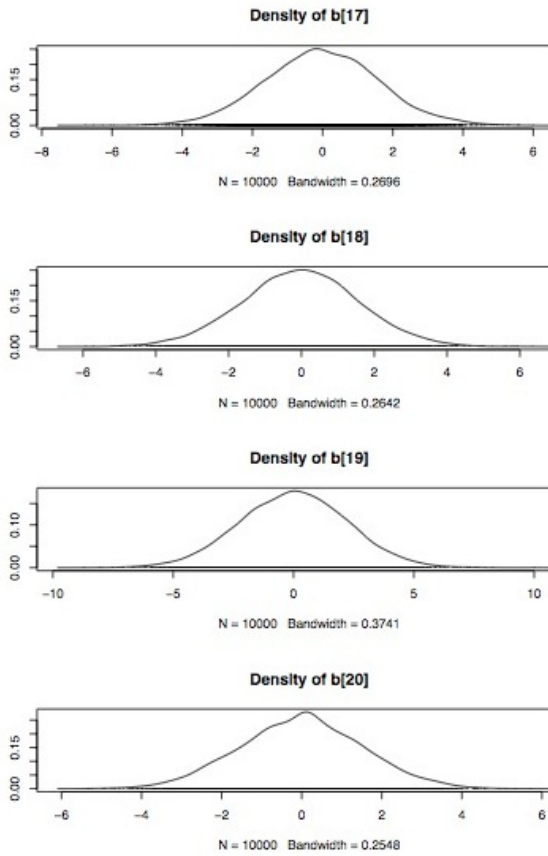
Figure 7.26: Posteriors estimated from GVS
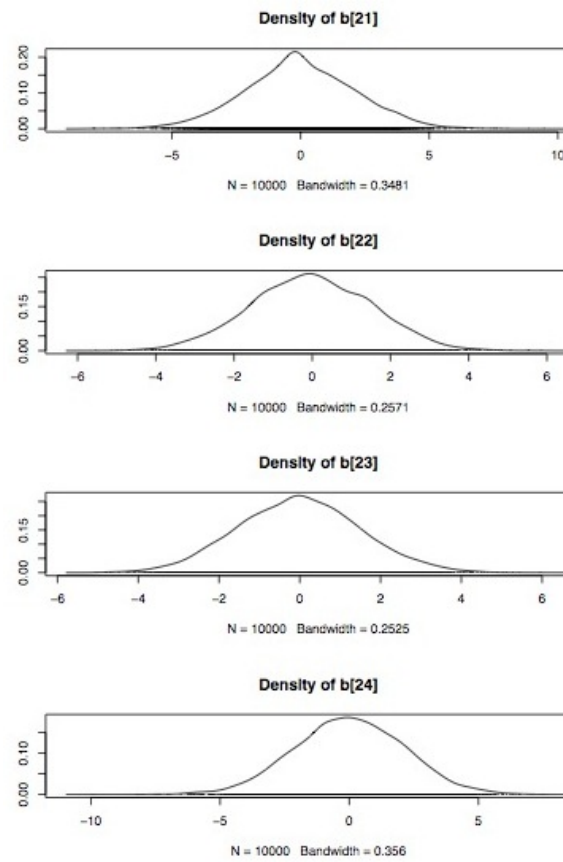
Figure 7.27: Posteriors estimated from GVS

Density of b[25]

N = 10000   Bandwidth = 0.2511

Density of b[26]

N = 10000   Bandwidth = 0.344

Density of b[27]

N = 10000   Bandwidth = 0.245

Density of b[28]

N = 10000   Bandwidth = 0.2396
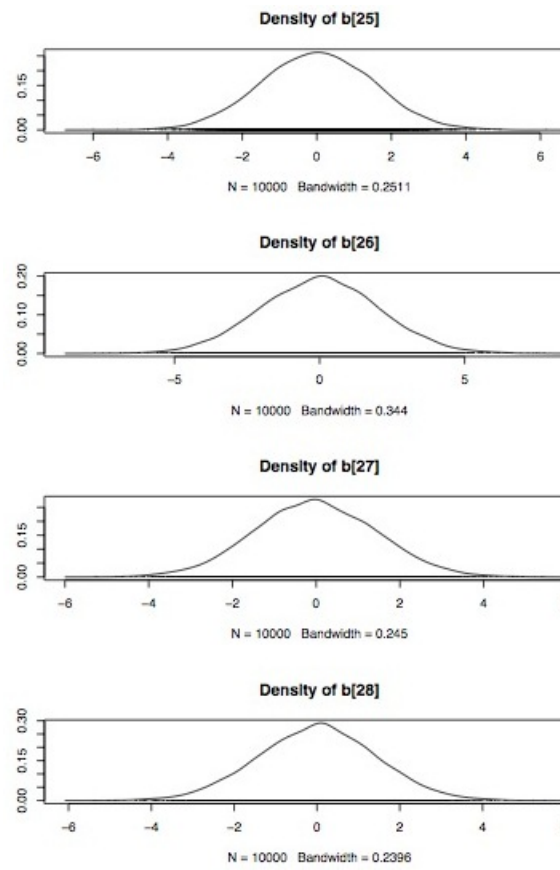
Figure 7.28: Posteriors estimated from GVS
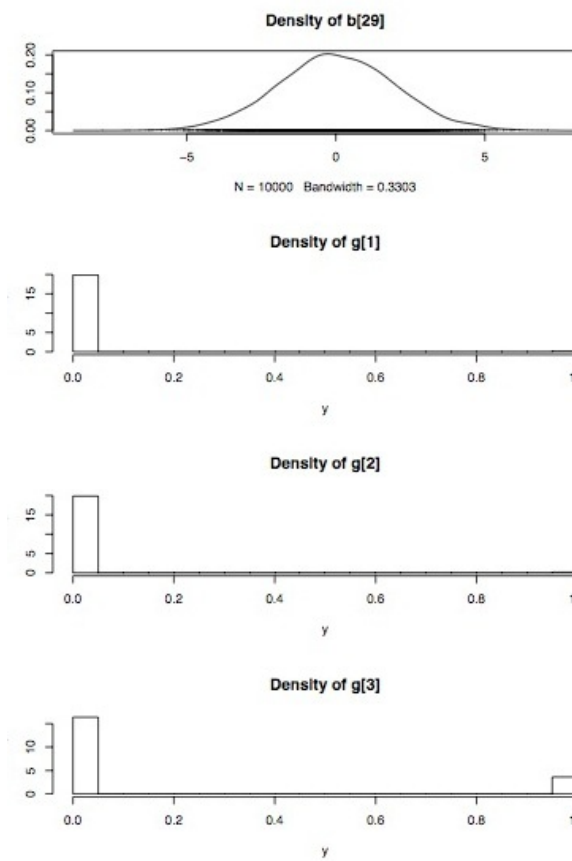
Figure 7.29: Posteriors estimated from GVS

Density of g[4]

Density of g[5]

Density of g[6]

Density of g[7]

Figure 7.30: Posteriors estimated from GVS

Figure 7.31: Posteriors estimated from GVS

Density of g[12]

Density of g[13]

Density of g[14]

Density of g[15]

Figure 7.32: Posteriors estimated from GVS

Figure 7.33: Posteriors estimated from GVS

Density of g[20]

Density of g[21]

Density of g[22]

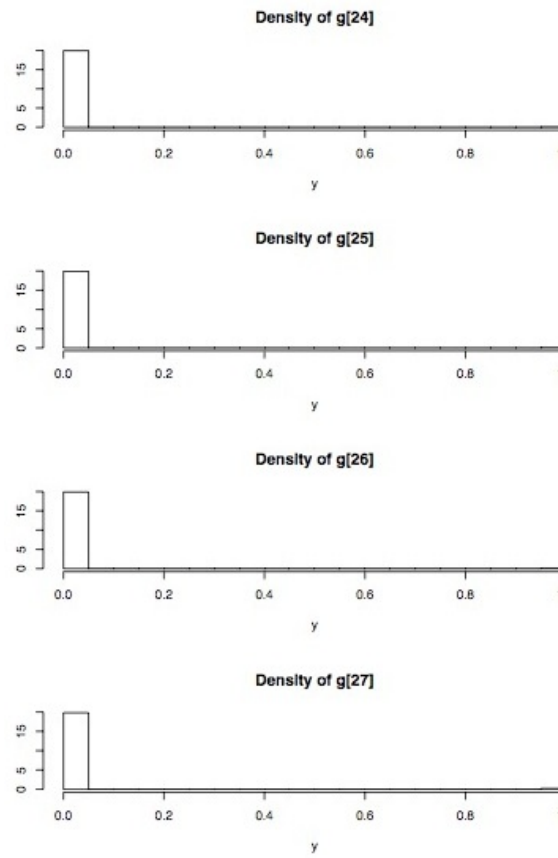**Density of g[23]**

Figure 7.34: Posteriors estimated from GVS
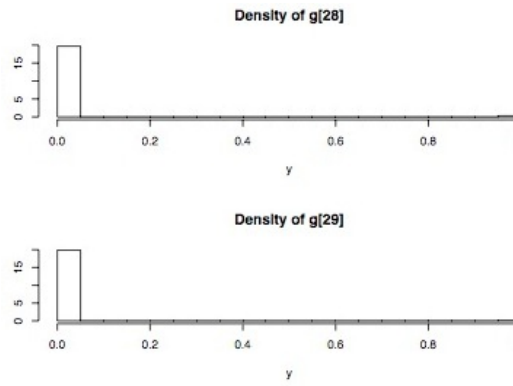
Figure 7.35: Posteriors estimated from GVS

Figure 7.36: Posteriors estimated from GVS
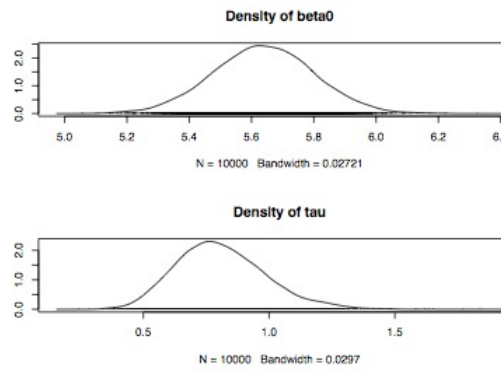

Figure 7.37: Posteriors estimated from GVS

Bibliography

[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] J. M. Brown, *Molecular Spectroscopy*, Oxford Science Publications, 1998.

[3] J. Schott, *Remote Sensing: The Image Chain Approach*, Oxford University Press, Oxford, NY, 2007.

[4] 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.

[5] 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 ingrared hyperspectral imager," in *SPIE Conference on Airborne Reconnaissance XXII*, 1998.

[6] 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.

[7] 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.

[8] N. Gallagher, B. Wise, and D. Sheen, "Error analysis for estimation of trace vapor concentration pathlength in stack plumes," *Applied Spectroscopy* , 2003.

[9] L. Chilton, "Model and method for gas identification." Personal communication, 2006.

[10] A. E. Raftery, D. Madigan, and J. A. Hoeting, "Bayesian model averaging for linear regression models," *Journal of the American Statistical Association* **92**, pp. 179–191, 1997.

[11] C. Fernandez, E. Ley, and M. Steel, "Benchmark priors for bayesian model averaging," *Journal of Econometrics* , 2001.

[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] A. Gelman, J. Carlin, H. Stern, and D. Rubin, *Bayesian Data Analysis*, Chapman and Hall, 2 ed., 2004.

[14] 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.

[15] S. T. Kacenjar, D. F. Gill, J. A. Lelii, J. Foreman, and C. B. Batrone, "Spectral effluent detection sensitivity study," in *SPIE Conference on Algorithms for Multispc?4fal and Hyperspectral Imagery IV*, 1998.

[16] 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.

[17] 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.

[18] M. Z. Salvador, R. G. Resmini, and R. B. Gomez, "Hyperspectral trace gas detection using the wavelet packet transform," *SPIE* **6966**, 2008.

[19] J. Neter, M. H. Kutner, W. Wasserman, and C. J. Nachtsheim, *Applied Linear Statistical Models*, McGraw Hill, 1996.

[20] P. Dellaportas, J. Forester, and I. Ntzoufras, "On bayesian model and variable selection using mcmc," *Statistics and Computing* **12**, 2002.

[21] I. Ntzoufras, "Gibbs variable selection using bugs," *Journal of Statistical Software,* **7**(7), 2002.

[22] M. Plummer, N. Best, K. Cowles, and K. Vines, *coda: Output analysis and diagnostics for MCMC*, r package version 0.10-7 ed., 2006.

[23] M. K. Cowles and B. P. Carlin, "Markov chain monte carlo convergence diagnostics: A comparative review," *Journal of the American Statistical Association* , 1996.

[24] J. Eidsvik, S. Martino, and H. Rue, "Approximate bayesian inference in spatial generalized linear mixed models," 2006. Unpublised.

[25] N. Cressie, *Statistics for Spatial Data*, Wiley, 1991.

[26] A. Blackadar, *Turbulence and Diffusion in the Atmosphere: Lectures in Environmental Sciences*, Springer, 1998.

[27] H. Ishida, T. Nakamoto, and T. Moriizumi, "Remote sensing of gas/odor source location and concentration distribution using mobile system," *Senors and Actuators B* , 1998.

[28] R. Deming, S. Higbee, D. Dwyer, M. Welser, and L. Perlovsky, "Adaptive estimation for spectral-temporal characterization of energetic transient events," in *World Congress on Computational Intelligence (WCCI), Vancouver, Canada*, 2006.

[29] R. Hogg and A. Craig, *Introduction to Mathematical Statistics*, Prentice Hall, 5th ed., 1995.

[30] B. P. Carlin and T. A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman and Hall/CRC, 1996.

[31] P. J. Green, "Reversible jump markov chain monte carlo computation and bayesian model determination," *Biometrika* **82**(4), 1995.

[32] S. B. Vardeman, "Stat 544 course outline." http://www.public.iastate.edu/ vardeman/stat544/544outline-07.pdf.