
**Pacific Northwest
National Laboratory**

Operated by Battelle for the
U.S. Department of Energy

Comparison of Two Gas Selection Methodologies: An Application of Bayesian Model Averaging

AS Renholds
SE Thompson
KK Anderson
LK Chilton

February 2006



Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DISCLAIMER

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

PACIFIC NORTHWEST NATIONAL LABORATORY

operated by

BATTELLE

for the

UNITED STATES DEPARTMENT OF ENERGY

under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service,
U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161
ph: (800) 553-6847
fax: (703) 605-6900
email: orders@ntis.fedworld.gov
online ordering: <http://www.ntis.gov/ordering.htm>



This document was printed on recycled paper.

(9/2003)

Comparison of Two Gas Selection Methodologies: An Application of Bayesian Model Averaging

AS Renholds
SE Thompson
KK Anderson
LK Chilton

February 2006

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99352

Executive Summary

One goal of hyperspectral imagery analysis is the detection and characterization of plumes. Characterization includes identifying the gases in the plumes, which is a model selection problem. Two gas selection methods compared in this report are Bayesian model averaging (BMA) and minimum Akaike information criterion (AIC) stepwise regression (SR).

Simulated spectral data from a three-layer radiance transfer model were used to compare the two methods. Test gases were chosen to span the types of spectra observed, which exhibit peaks ranging from broad to sharp. The size and complexity of the search libraries were varied. Background materials were chosen to either replicate a remote area of eastern Washington or feature many common background materials.

For many cases, BMA and SR performed the detection task comparably in terms of the receiver operating characteristic curves. For some gases, BMA performed better than SR when the size and complexity of the search library increased. This is encouraging because we expect improved BMA performance upon incorporation of prior information on background materials and gases.

Acknowledgements

We would like to acknowledge and thank Sheila Bennett for her editing services, and Alejandro Heredia-Langner for his advice.

Abbreviations and Acronyms

SYMBOL	DEFINITION
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
BMA	Bayesian Model Averaging
BMS	Bayesian Model Selection
GTK	Gas Tool Kit
HIP	Hyperspectral Image Processing
IR-SAGE	Infrared Systems Analysis in General Environments
KLI	Kullback-Liebler Information
LANL	Los Alamos National Laboratory
LLNL	Lawrence Livermore National Laboratory
MCMC	Markov Chain Monte Carlo
MC ³	Markov Chain Monte Carlo model composition
NEFDS	Nonconventional Exploitation Factors Data System
N/FAPs	Nuisance/False Alarm Probabilities
PDs	Probabilities of Detection
PMP	Posterior Model Probability
PNNL	Pacific Northwest National Laboratory
ROC	Receiver Operating Characteristic
SR	Stepwise Regression
SSVS	Stochastic Search Variable Selection
WMF	Whitened Matched Filter

Contents

Executive Summary	iii
Acknowledgements	v
Abbreviations and Acronyms	vii
1.0 Introduction	1.1
2.0 Model Setup	2.1
2.1 Background	2.2
2.2 Radiance Transfer Model	2.2
2.2.1 Development of the Whitening	2.3
2.3 Chemicals of Interest	2.4
3.0 Model Selection	3.1
3.1 Select Single Model	3.1
3.2 Incorporate Multiple Models	3.2
3.2.1 History of Model Averaging	3.2
3.2.2 Current Methods for Model Averaging	3.3
3.2.3 How to Address Large Model Space	3.3
3.3 Bayesian Theoretical Development	3.4
3.3.1 Bayesian Model	3.4
3.3.1.1 Priors	3.5
3.3.1.2 Priors on the Models	3.6
3.3.2 Derivation of Posterior Model Probability	3.7
3.3.2.1 Approximations	3.8
3.3.3 Derivation of Quantity of Interest Distribution	3.8
4.0 Simulation	4.1
4.1 Description	4.1

4.1.1	Wavelengths and Instruments	4.1
4.1.2	Background Materials Available	4.1
4.1.3	Gases Available	4.2
4.1.4	Whitening	4.4
4.1.5	Gas Concentrations	4.4
4.2	Comparison Metrics	4.5
4.2.1	Receiver Operating Characteristic Curve	4.5
4.2.2	Other Metrics	4.5
4.3	Simulation Study Results	4.6
4.3.1	Study 1	4.6
4.3.2	Study 2	4.7
5.0	Conclusions	5.1
6.0	Future Work	6.1
6.1	Statement of Work - FY 2006	6.1
6.1.1	Investigate Bayesian Model Averaging	6.1
6.1.2	Explore Prior Distributions of Nuisance Parameters	6.1
6.1.3	Chemical Universe	6.1
6.1.4	Algorithm Development - Image Clustering	6.1
6.1.5	Temporal Change Analysis for Persistent Monitoring	6.4
6.2	Long-Term Planning	6.4
6.2.1	Model Validation/Testing	6.4
6.2.2	Prototype Development	6.4
7.0	References	7.1
	Appendix A – Simulation Variables	A.1
	Appendix B – Study Figures	B.1

Appendix C – Simulation Code C.1

Figures

2.1	Simple physical radiance model.	2.2
2.2	Raw and whitened ethene absorbance spectra.	2.5
2.3	Raw and whitened background spectra.	2.5
4.1	Relative emissivities of four background materials for simulated scene.	4.2
4.2	A generic scene from Study 1.	4.3
4.3	Chemical spectra for five gas plumes at instrument resolution.	4.4
4.4	Receiver operating characteristic curve.	4.6
4.5	2-sigma uncertainties for ROC curves.	4.7
6.1	1D image with four spectral bands, and weak plume at intersection of regions.	6.3
6.2	ROC curves for clustered (solid line) and unclustered (dashed line) matched filter detection.	6.4
B.1	ROC ammonia study 1	B.1
B.2	ROC ethene study 1	B.2
B.3	ROC F-113 low conc study 1	B.3
B.4	ROC F-113 very low conc study 1	B.4
B.5	ROC furan study 1	B.5
B.6	ROC ammonia correlated study 2	B.6
B.7	ROC ammonia uncorrelated study 2	B.7
B.8	ROC ethene correlated study 2	B.8
B.9	ROC ethene uncorrelated study 2	B.9
B.10	ROC F-113 correlated study 2	B.10
B.11	ROC F-113 uncorrelated study 2	B.11
B.12	ROC F-114 correlated study 2	B.12
B.13	ROC F-114 uncorrelated study 2	B.13
B.14	ROC furan correlated study 2	B.14

B.15 ROC furan uncorrelated study 2 B.15

Tables

A.1	Full Gas Library	A.1
A.2	Simulated Sensor Error Model Values	A.4
A.3	Backgrounds for Study 2	A.5

1.0 Introduction

One goal of hyperspectral imagery analysis is the detection and characterization of plumes. Examples include detection of sulfur dioxide plumes (Marino 1999) and general effluent detection (O'Donnell et al. 2004). Characterization includes identifying and quantifying the gases in the plumes. In data analysis terms, identification of gases is a model selection problem; what, if any, gases in a given chemical spectral library are present in the pixel of interest? Techniques for solving the model selection problem include using banks for whitened matched filters and stepwise regression, both of which involve linear modeling of the spectral data.

A valuable model selection methodology is Bayesian model averaging (BMA), (Hoeting et al. 1999a). The most attractive feature of BMA is that it provides explicit estimates of the probabilities and uncertainties that the gases in the given chemical spectral library are present in the pixel of interest. This is in contrast to the more dichotomous results of stepwise regression techniques; in which a list of gases is included in the model, explicitly excluding the remainder of the library.

The goal of this report is to compare BMA with minimum Akaike information criterion (AIC) stepwise regression (Venables and Ripley 2002) for detecting optically thin plumes with small spatial coverage (plume pixels are few in comparison to the whole image). We analyzed simulated hyperspectral datasets based on the physical radiance transfer modeling and whitened matched filtering of Villeneuve and Stocker (2000) and Stocker (2000). Simulated datasets were used so that ground-truth was available, with known gases present, gas concentration, and background materials.

In Section 2, we present the physical radiance transfer model used in our study. The two model selection methodologies are described in Section 3, with great detail on the development of the BMA algorithm used. Section 4 details the results of two simulation studies comparing BMA and minimum AIC stepwise regression across various background materials, chemical gases and concentrations, and chemical spectral library size and complexity. Conclusions are given in Section 5, and possible directions for future work are given in Section 6. References are in Section 6.2.2, and supplemental material is provided in the appendices.

2.0 Model Setup

The plume analysis problem can be divided into three stages: detection, identification, and quantification. Each stage in the process is conditional, based on the prior stages and may make use of spectral, spatial, and temporal information. The spectral information is used to select a small number of pixels that differ from the general spectral behavior of other pixels in the image or from other pixels with similar background characteristics. The spatial information is used to further identify small sets of neighboring pixels that appear to be plume shaped. In many cases, a sequence of images through time can help in selecting organic shapes that move through time as the wind moves the plume.

The traditional technique for plume detection consists of applying a set of whitened matched filters to an image or subset of an image (Young 2002). A common method for chemical identification is to compare the spectral information in a whitened plume pixel (or small set of neighboring pixels) to a chemical library using a χ^2 test. Finally, the chemical concentration is quantified. One technique for quantification is to compare model coefficients from the whitened matched filter results for the selected gas when there are multiple gases in the plume.

There may be significant interaction between and within stages, modifying the information required for detection to improve results. In each stage, the methods used in identification provide a measure of the strength of one individual conclusion using standard mathematical and statistical methods. One major challenge in this process is that the strength of each decision does not account for any of the other decisions that have been made during the process or the iterative nature of plume detection. Therefore, decision strength may be overestimated, leading to incorrect decisions.

Estimation, model selection, and the assessment of errors are often the primary goals in many areas of statistics. The modeling paradigm typically involves the following four steps. First, a class of candidate models must be selected. In general, the modeler will select a set of explanatory variables and a model class such as regression models, linear models, or nonlinear models. After the set of potential explanatory variables has been chosen, the analysis proceeds without considering other sets of explanatory variables that may fit the data equally well. For example, in a general linear model framework, this would include the choice of explanatory variables and decisions about how the explanatory variables will be included in the model, including transformations, interactions, and model size. The second stage involves selecting the best model using one or more of a suite of measures of model fit, including Mallows' C_p , R^2 , adjusted R^2 , BIC, and AIC to name a few. The third stage is estimation of the parameters in the best model. Finally, the model is used to provide an assessment of model error.

Bayesian methods have the potential to provide a more "honest" assessment of uncertainty in the decision-making process by incorporating additional information into the decision-making process in a quantifiable way and reducing the amount of iteration. In this paper, we begin to compare Bayesian procedures with traditional, quantifiable methods of plume detection, primarily whitened matched filtering. The intended outcome of this analysis is defensible scientific decisions that better assess the confidence in the decision made. This paper discusses the first phases of Bayesian methods, examining the results of BMA applied to plume detection, and looking into Bayesian clustering as a method for improving on and off-plume pixel identification.

2.1 Background

This section discusses the construction of the modeling framework used in detection of chemical plumes. Using first principles of atmospheric physics, the framework builds a three-layer atmospheric model that includes aspects of the sensor used to measure the data, the platform the sensor sits on (generally airborne or spaceborne), atmospheric layers, the plume, and the surface of the ground in terms of both temperature and emissivity. This section introduces the model one layer at a time.

2.2 Radiance Transfer Model

This section describes the physical radiance transfer model we use for remote sensing. A general scene containing a plume is displayed in Figure 2.1. L_{obs} , the radiance or luminance observed at the sensor, is a linear combination of the atmospheric upwelling radiance, downwelling radiance, ground radiance, reflected downwelling, and plume radiance. All the radiance measurements depend on wavelength (or wavenumber). Our model consists of three layers: ground, plume, and atmosphere. The ground radiance is defined as

$$L_g(\lambda) = \varepsilon_g(\lambda)B(\lambda, T_g), \quad (2.1)$$

where $\varepsilon_g(\lambda)$ is the emissivity of the ground, and $B(\lambda, T_g)$ is Planck's Blackbody function evaluated at the temperature of the ground. Reflected downwelling radiance is assumed to be negligible for this paper. The off-plume pixels have the following radiance:

$$L^{off}(\lambda) = L_g(\lambda)\tau_a(\lambda) + L_a(\lambda), \quad (2.2)$$

where $\tau_a(\lambda)$ is the upward transmissivity of the atmosphere and $L_a(\lambda)$ is the atmospheric upwelling radiance. On-plume radiance is modeled by having the ground radiance pass through

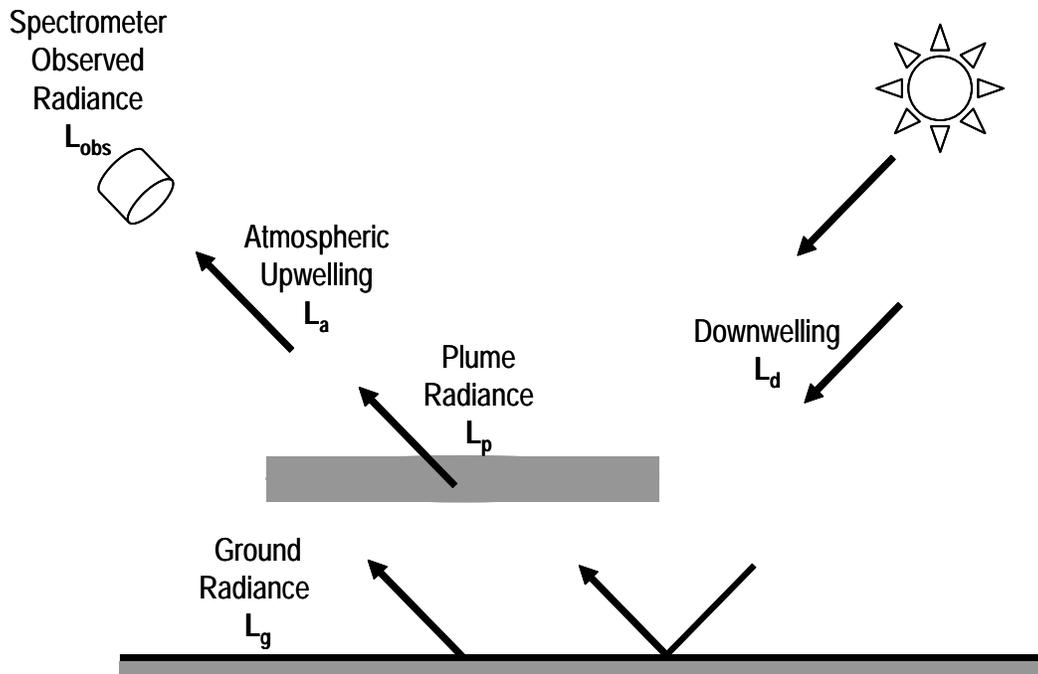


Figure 2.1. Simple physical radiance model.

the plume, which then passes through the atmosphere before reaching the sensor.

$$L^{on}(\lambda) = (L_g(\lambda)\tau_p(\lambda) + L_p(\lambda))\tau_a(\lambda) + L_a(\lambda) \quad (2.3)$$

$$= (L_g(\lambda)\tau_p(\lambda) + (1 - \tau_p(\lambda))B(\lambda, T_p))\tau_a(\lambda) + L_a(\lambda), \quad (2.4)$$

where $L_p(\lambda)$ is the plume radiance term and $\tau_p(\lambda)$ is the transmissivity of the plume. By Beer-Bouguer-Lambert Law (Liou 2002),

$$\tau_p(\lambda) = \exp\left(-\sum_{j=1}^J A_j(\lambda)C_j\right), \quad (2.5)$$

where $A_j(\lambda)$ is the absorbance for gas j , and C_j is the burden (concentration times path length) for the same gas, with J total gases to choose from.

When the plume is a thin layer of gas in the atmosphere, i.e., the path length is short and of low concentration, the τ_p term can be expanded via a Taylor series expansion, keeping just the first order approximation, (Salas and Hille 1995). This linearization results in Equation (2.5) becoming

$$\hat{\tau}_p(\lambda) = 1 - \sum_{j=1}^J A_j(\lambda)C_j \quad (2.6)$$

2.2.1 Development of the Whitening

For this report, we will be applying the spectral matched filter techniques in conjunction with the model to detect and estimate the gas burdens. Spectral matched filtering is a specialized application of unconstrained multivariate regression. In order to apply this technique, we must first rearrange our model into a linear form, similar to Stocker (2000).

The pixel contrast will be defined as the difference between the on-plume and off-plume radiance, given as

$$L^{on}(\lambda) - L^{off}(\lambda) = \tau_a(\lambda)(1 - \tau_p(\lambda))(B(\lambda, T_p) - L_g(\lambda)). \quad (2.7)$$

The above equation is exact, whereas in the real world we cannot know the on and off-plume radiance of each pixel at the exact same time.

We can only measure one radiance per pixel, and in practice the pixel contrast will include $\delta(\lambda)$ as the residual error. If we then substitute $\hat{\tau}_p$, the estimate of τ_p , for τ_p and bring the off-plume radiance back to the right side of the equation, (2.7) becomes

$$L^{on}(\lambda) = \left(\sum_{j=1}^J A_j(\lambda)C_j\right)\tau_a(\lambda)(B(\lambda, T_p) - L_g(\lambda)) + L^{off}(\lambda) + \delta(\lambda). \quad (2.8)$$

In the case of spectral matched filtering, there is only one basis vector for the off-plume radiance, $\overline{L^{off}}(\lambda)$, the spectral mean of the off-plume pixels. Off-plume pixels are found either by having an a priori mask where the plume pixels are (as in simulated datasets), or by applying some method (like principal component analysis) to the whole scene to make a first estimate at a plume mask. One can use the whole image to find the mean and covariance structure of the off-plume wavelengths, provided the image is dominated by the off-plume pixels. Let β_j be a rearrangement of some terms in the above equation defined and approximated as

$$\beta_j = \tau_a(\lambda)C_j(B(\lambda, T_p) - L_g(\lambda)) \approx C_j(B(\bar{\lambda}, T_p) - L_g(\bar{\lambda})), \quad (2.9)$$

where $\bar{\lambda}$ is the average wavelength. If $\tau_a(\lambda)$ is approximately one, given the height above the ground that the data were taken at, then the final linear model is put in standard regression form. Let \mathbf{Q} denote the cube of hyperspectral data that is an $N \times M \times \Lambda$ dimensional array, where N and M are spatial dimensions and Λ is the spectral dimension. Given that $\mathbf{Q}_{nm}(\lambda)$ is one pixel's measured value of $L^{on}(\lambda)$, then

$$L^{on}(\lambda) = \left(\sum_{j=1}^J A_j(\lambda) \beta_j \right) + \overline{L^{off}}(\lambda) + \delta(\lambda) \text{ or,} \quad (2.10)$$

$$\mathbf{Q}_{nm}(\lambda) - \overline{L^{off}}(\lambda) = \left(\sum_{j=1}^J A_j(\lambda) \beta_j \right) + \delta(\lambda). \quad (2.11)$$

Putting the same equation into matrix notation with λ as the row dimension, we have for a pixel column of data,

$$(\mathbf{Q}_{nm} - \overline{\mathbf{L}^{off}}) = \mathbf{A}\boldsymbol{\beta} + \boldsymbol{\delta}. \quad (2.12)$$

For the linear regression to have a variance equal to a diagonal matrix $\sigma^2 I$, the final step of whitening, we must “divide by the square root” of the spectral covariance, $\boldsymbol{\Sigma}(\lambda)$, of either the whole image or the off-plume pixels.

$$\boldsymbol{\Sigma}^{-1/2}(\mathbf{Q}_{nm} - \overline{\mathbf{L}^{off}}) = (\boldsymbol{\Sigma}^{-1/2}\mathbf{A})\boldsymbol{\beta} + \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\delta}. \quad (2.13)$$

The final whitened linear regression model in standard statistical notation takes the form

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi}, \quad (2.14)$$

where $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{Q}_{nm} - \overline{\mathbf{L}^{off}})$, $\mathbf{X} = (\boldsymbol{\Sigma}^{-1/2}\mathbf{A})$, and the error is $\boldsymbol{\xi} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\delta}$.

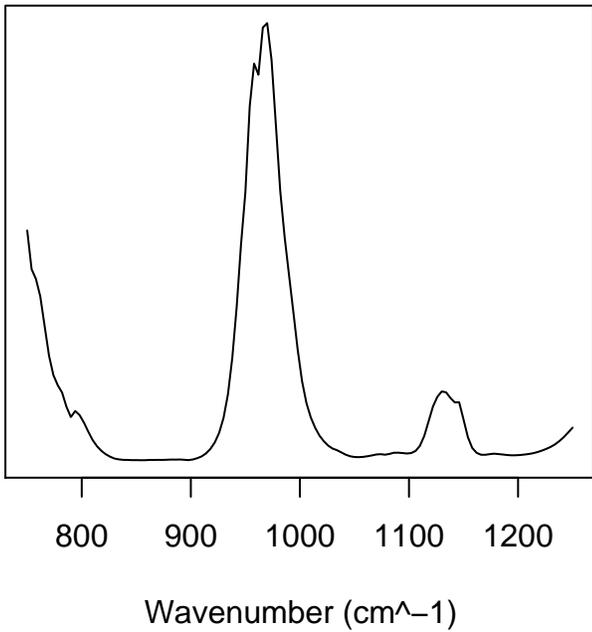
Examples of the whitening of a gas absorbance spectrum and a background pixel are given in Figures 2.2 and 2.3. Figure 2.2 shows the differences in the gas ethene between raw spectra and whitened spectra. On the left is the gas ethene at airborne sensor resolution and range, and on the right is the column of the \mathbf{X} matrix that corresponds to the whitened ethene spectra. For the background, a sample pixel going through the whitening process is displayed in Figure 2.3.

2.3 Chemicals of Interest

A wide variety of chemicals can be detected in plumes or in the atmosphere. Much of the current research in detecting chemicals in the atmosphere focuses on estimating the major constituents such as nitrogen, oxygen, argon, water, and carbon dioxide (Aires et al. 2002) or pollutants such as sulfur dioxide (Marino 1999). As discussed earlier, we are most interested in detecting low-concentration gases present in optically thin plumes.

The full gas library that we consider in this effort for Study 2 is the intersection of the standard sets in Gas Toolkit (GTK) from Lawrence Livermore National Laboratory (LLNL), Hyperspectral Image Processing (HIP) from Los Alamos National Laboratory (LANL), and Infrared Spectral Library from Pacific Northwest National Laboratory (PNNL). The library for Study 1 was taken entirely from the PNNL spectral library. Table A.1 of Appendix A contains a complete listing of chemicals.

Raw Ethene Spectra



Whitened Ethene Spectra

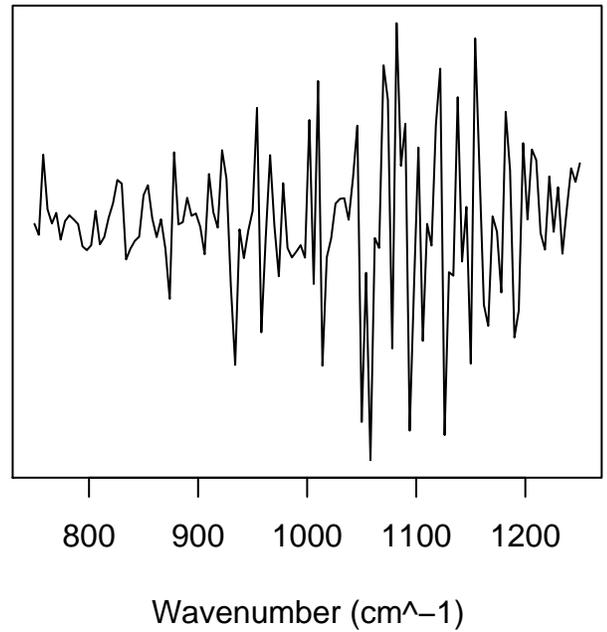
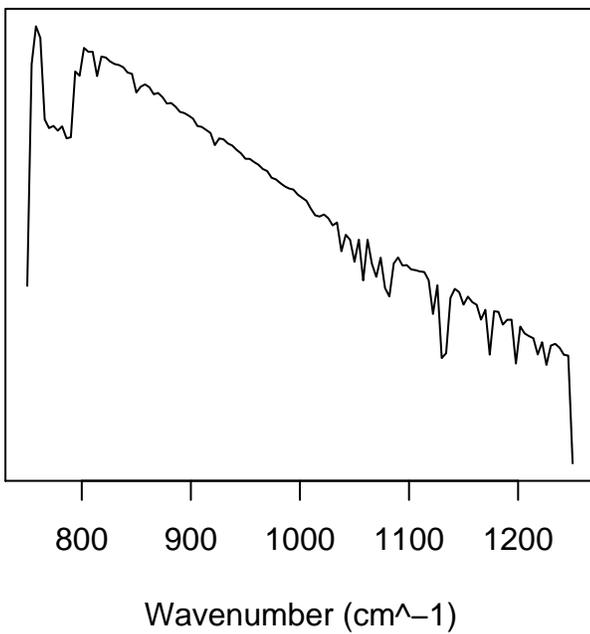


Figure 2.2. Raw and whitened ethene absorbance spectra.

Raw Background Pixel



Whitened Background Pixel

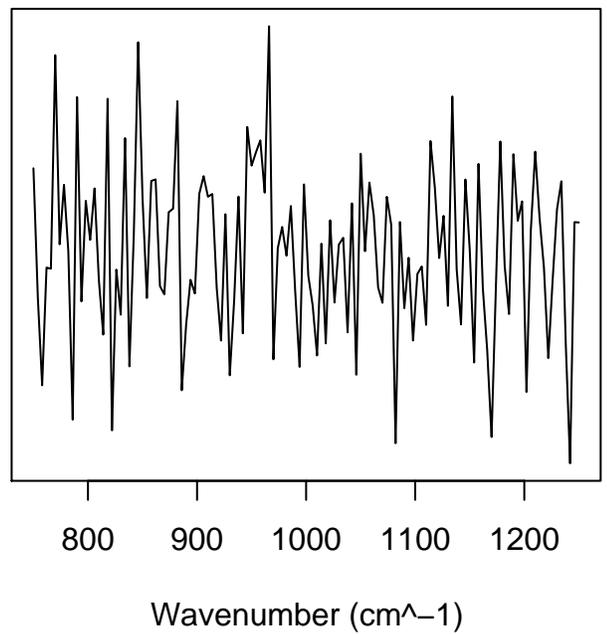


Figure 2.3. Raw and whitened background spectra.

3.0 Model Selection

The radiance modeling described in Section 2.2 produced a linear regression model for a whitened spectrum from a cube of hyperspectral data. Fitting the model, that is, estimating β , is a simple least squares optimization. However, determining which J gases to include in the \mathbf{X} matrix is not so simple. In the typical application, the goal of model selection is to find the best-fitting model that can be used to explain the behavior of the data, or occurrences of future observations. For our remote sensing application, the goal is to detect the plume and determine the gases present; are any gases present, and, if so, which ones?

The general statistical problem of selecting the “best” linear regression equation has a long history. Many algorithms and criteria have been proposed for the selection of the variables to include in the \mathbf{X} matrix. These include all possible regressions, best subset regressions, backward elimination, and stepwise regression, with various criteria: correlation coefficient (R^2), mean squared error, and Mallows’s C_p , to name a few (Draper and Smith 1981). All methods have advantages and disadvantages due to their measurement of different aspects of model misfit. For example, the correlation coefficient measures the linear relationship of the model errors with the response vector. If there is a strong nonlinear response, it may not be reflected in R^2 . One recent method that has become widely accepted is minimum AIC stepwise regression (Akaike 1983, Venables and Ripley 2002).

Another approach to model selection is Bayesian model averaging (BMA). Instead of selecting the best model, the idea is to average over a set of models. The goal of this paper is to compare BMA to minimum AIC stepwise regression (SR). In this section we describe these two methods and their advantages and disadvantages.

3.1 Select Single Model

The concept behind minimum AIC SR is to penalize a proposed model (a particular set of variables) for its complexity (the number of variables used). So while a model with a large number of variables will fit better (in terms of increased multiple correlation coefficient and reduced sum of squared error values), we are searching for the simplest model that can adequately describe available data. Thus, an information criterion is used that combines the fit and the penalty. The SR algorithm finds the set of variables that minimizes the information criterion in a stepwise approach, adding and removing variables from the regression equation. Information criteria take the form

$$\text{IC}(c) = -2\log(L(M_J)) + cJ, \quad (3.1)$$

where M_J represents a model with a particular set of J gases, $\log(L(M_J))$ is the maximized log likelihood for that model, and c is the penalizing parameter. For the Akaike Information Criterion (AIC), $c = 2$ (Akaike 1983). While other penalizing parameters have been proposed, model comparisons based on AIC have a number of optimal properties related to prediction, including asymptotic equivalence to cross-validation methods and Bayesian methods (based on Bayes factors).

The SR algorithm used in this research was written by Ripley (2002) in the R language and environment for statistical computing and graphics. R is available as Free Software under the terms of the Free Software Foundation’s GNU General Public License (see www.r-project.org). The biggest advantage of SR is its computational efficiency compared with other regression model selection methods.

Another single model method is Bayesian model selection (BMS). BMS selects models based on the posterior distribution of the data. This metric differs from stepwise regression in that it allows for the inclusion of prior distributions on the model parameters. This method does not incorporate uncertainty in model selection, but serves as a technique for selecting a set of explanatory variables.

Bayesian model selection is similar to the methodology of Bayes Factors (Kass and Raftery 1995, Raftery 1995) where models are chosen based on the ratio of the posterior distributions. A Bayes factor comparing model M_1 and model M_0 is defined to be

$$B_{1,0} = \frac{pr(Z|M_1)}{pr(Z|M_0)}, \quad (3.2)$$

where $pr(Z|M_k)$ is the integrated likelihood for model M_k . Bayes factors provide a method for model comparison. Kass and Raftery (1995) provide a framework for evaluation of the evidence for model M_1 fitting the data better than model M_0 . These criteria help to quantify decisions made about explanatory variables. Bayes factors have been used for model selection in a large variety of statistical models, including linear models, stochastic processes, survival analysis, and multivariate analysis.

For BMS, the model with the largest posterior model probability (PMP) is selected for further examination. BMS forms a middle ground between stepwise model selection (no prior distributions) and BMA (prior distributions on both model parameters and the models themselves) that incorporates parameter uncertainty into the selection of the model and parameter estimation, but does not take model uncertainty into account.

BMS is a building block in performing BMA. When uninformative prior distributions are used, BMS reduces to the BIC model criteria. For chemical plume detection, we will incorporate informative prior distributions on the model parameters in future model construction.

3.2 Incorporate Multiple Models

BMA provides a framework where the results from many models can be combined and the uncertainty in the selection of explanatory variables can be included in the prediction intervals. BMA has been applied to general linear models (Hoeting et al. 1999b) and to survival analysis models (Volinsky 1997). This report applies to the plume detection, identification, and estimation problem BMA methodology to provide more accurate assessments of the probability of a plume, the particular gases in the plume, and their concentrations.

3.2.1 History of Model Averaging

Volinsky (1997) is a good source on the history of combining models. The origins of combining models can be found in “Laplace’s Deuxieme Supplement a la Theorie Analytique des Probabilities” (1818). Stigler (1973) provides the translation of Laplace (1818), showing how “two estimators could be combined to provide a new estimator which would be better than either.” This is not model averaging but formed a basis for future work. Cochran (1937) established the foundation for incorporating estimators from multiple samples. In the quality control literature, Bates and Granger (1969) were the first to combine results from different models fit to the same data. Articles on combining predictions from different models soon followed. Bates and Granger (1969) demonstrated that the weighted average of two-model forecasts is better than forecasts from the individual models as long as each model contains some “independent information.”

The philosophy of BMA and model uncertainty can be traced to two articles from the statistical community. Roberts (1965) proposed a distribution that combines the results of two models, foreshadowing the development of BMA for multiple models. Volinsky (1997) credits Leamer (1978) with addressing the idea that “averaged estimators incorporate the ambiguity about the correct model into the posterior distribution,” verifying the philosophy that averaging over models can account for model uncertainty.

3.2.2 Current Methods for Model Averaging

A variety of articles have reviewed methodology for BMA and have addressed some of the problems with selecting one “true model.” Draper (1995) focuses on Bayesian hierarchical models, building an expanded modeling structure where the highest level of structure corresponds to the probability of a specific model. Chatfield (1995) challenges the statistical community to account for model uncertainty, stating that failure to do so may be “more serious than other sources of uncertainty.” Kass and Raftery (1995) focus on Bayes factors, the ratio of posterior odds of a model to its prior odds, as a tool for comparing and combining models. Hoeting et al. (1999b) summarize BMA methodology.

Burnham and Anderson (1998) detail a frequentist model averaging framework that uses the AIC statistic to estimate the Kullback-Liebler Information (KLI) to provide model weights. This framework does not average over all models but focuses on “good” models and incorporating expert opinion.

In the last decade, a variety of methods, both Bayesian and frequentist, have been proposed that average the results from multiple models. Implementation of BMA for plume detection will be described in the following sections.

3.2.3 How to Address Large Model Space

Although this report focuses on averaging over all possible models, there are situations where the number of models in the model space \mathcal{M} is so large that averaging over all models is not viable. Methods for either reducing the model space or approximating variable posterior distributions have been developed. Some of the methods include Occam’s window (Madigan and Raftery 1994, Hoeting 1994), leaps and bounds (Volinsky et al. 1997, Volinsky 1997), Markov Chain Monte Carlo model composition (MC³) (Madigan and York 1995, Hoeting 1994) and stochastic search variable selection (SSVS) (George and McCulloch 1993). The following discussion focuses on the model space defined as the set of all possible explanatory variables.

Occam’s window is built on two ideas. First, a typical situation is that most of the models in the model space \mathcal{M} do not predict as well as a smaller subset of models in \mathcal{M} . These poor-fitting models can be removed from the model space. Therefore, the posterior predictive distribution does not average over any models whose fit is inadequate. The second idea is Occam’s Razor: if the probabilities of two models given the data are equal, then the simpler model is better and is therefore used. These two principles can reduce the model space significantly depending on the user’s definition of poorly fitting and equivalent models.

Leaps and bounds (Furnival and Wilson 1974) is an algorithm for linear regression that reduces the model space by “trimming” sections of model space that do not fit well. Leaps and bounds is used by Volinsky et al. (1997) to quickly identify a subset of models to be used in the posterior.

On the other hand, some methods approximate the posterior distribution of interest using a

stochastic search algorithm. For example, MC³ builds a stochastic process that travels through model space. In this approach, the proportion of time spent at each model should approximate the posterior model probability (PMP). Furthermore, an approximation of the average of the posterior predictive distribution is the average posterior prediction for each step in the stochastic process.

Another method that uses a stochastic search methodology is SSVS. SSVS embeds the regression setup into a hierarchical Bayes normal mixture model. In this framework, a set of latent variables specifies the inclusion of explanatory variables. As the chain progresses, the frequency of specific sets of latent variables is useful in identifying promising models. Sets of explanatory variables that appear often fit the data better.

A more complete description of methods to use when the model space is large can be found in Hoeting et al. (1999b). In summary, both the Occam's window and leaps and bounds approaches average over a smaller subset of the models that are upheld from the data. When the number of explanatory variables is large, i.e., $K > 30$, these methods "are too expensive computationally or do not explore a large enough region of the model space" (Clyde 1999).

The last two methods require Markov chain Monte Carlo (MCMC) or importance sampling techniques to travel through the model space. Clyde (1999) points out that these methods "can be viewed as special cases of reversible jump MCMC algorithms." In these algorithms, the chain travels through model space as well as parameter space of potentially different dimensions.

To address the problem of large model space, this report fits all possible models up to a fixed number of parameters. The leaps and bounds function "leaps" coded in R is also incorporated into the code to allow for easy expansion. Fitting all possible models up to a fixed size will miss models that include a larger number of chemicals than the maximum allowed in our code.

3.3 Bayesian Theoretical Development

In this section, we describe constructing a Bayesian model around the model described in Section 2. This will include selection of priors and the derivation of terms needed in development of model selection and model averaging techniques.

3.3.1 Bayesian Model

A first step in accounting for model uncertainty is to define a set of models, \mathcal{M} , for prediction. Let \mathcal{M} be the "model space," the set of models under consideration, where $\mathcal{M} = \{M_1, M_2, \dots, M_K\}$. The models M_i in this set \mathcal{M} may differ by including different sets of explanatory variables, auto-correlation functions, or transformations on the response Z . The set of models is determined by a variety of methods, including all possible models given the explanatory variables and models deemed important by previous studies or other scientific considerations.

The second step is to compute the BMA posterior predictive distribution. Define Δ to be a quantity of interest, such as a quantity proportional to plume gas concentration. Note that Δ must have the same meaning for all models.

In performing BMA for remote sensing, the formulation of the problem follows the standard regression setup described in Hoeting et al. (1999a).

$$Z(s) = X(s)\beta + \epsilon(s),$$

where $\varepsilon(s)$ is assumed to have mean zero and variance σ^2 . Standard Bayesian methods incorporate uncertainty of the parameters of the model into the analysis. The parameters are seen as random variables. To accomplish this, prior distributions will be placed on the parameters. If $f(Z|\beta, \sigma^2, \theta)$ is the likelihood equation and $f(\beta, \sigma^2, \theta)$ is the prior distribution on the parameters, then using Bayes rule the posterior distribution of the parameters given the data Z is

$$f(\beta, \sigma^2, \theta|Z) = \frac{f(Z|\beta, \sigma^2, \theta) f(\beta, \sigma^2, \theta)}{\int \dots \int f(Z|\beta, \sigma^2, \theta) f(\beta, \sigma^2, \theta) d\theta d\sigma^2 d\beta}. \quad (3.3)$$

The prior distribution represents knowledge and uncertainty about the parameters, while the posterior distribution represents the updated distribution of the parameters after observing the data. Equation 3.3 is referred to as the posterior model probability (PMP).

Prior distributions are typically classified according to several characteristics. The term ‘‘proper’’ prior refers to a prior distribution that integrates to 1. An improper prior distribution integrates to infinity.

Prior distributions can be further classified as informative or noninformative. A noninformative prior contains no information about the parameters and therefore does not favor any value over any other. For example, $f(\beta) = c > 0$, for $-\infty < \beta < \infty$, is both improper and noninformative because $\int_{-\infty}^{\infty} f(\beta) d\beta = c \int_{-\infty}^{\infty} d\beta = \infty$, and $f(b_i) = f(b_j)$ for all $b_i, b_j \in \mathbb{R}$. A frequently used noninformative prior for scale parameter σ is $f(\sigma) = \sigma^{-1}$. In general, noninformative priors can be chosen via Jeffreys method (Berger 1985, p. 87-88) where

$$f(\beta, \sigma^2, \theta) = |I(\beta, \sigma^2, \theta)|^{\frac{1}{2}}, \quad (3.4)$$

where $I(\beta, \sigma^2, \theta)$ is the expected Fisher information matrix. Noninformative and improper prior distributions are acceptable prior distributions as long as the resulting posterior distribution is proper.

Calculation of the posterior distribution can be simplified by the use of conjugate priors. Certain likelihood functions and prior distributions result in closed-form posterior distributions. When using conjugate priors, it is not necessary to calculate the denominator in Equation (3.3) because the posterior distribution can be identified from the form of the numerator. For example, if the likelihood function of a linear model is Gaussian, then the conjugate prior distribution on (β, σ^2) is Gaussian and inverse gamma distributions, respectively, with β and σ^2 independent. The resulting posterior distribution is also Gaussian.

One of the challenges in performing BMA is determining both prior structure and prior parameterization. The results of analysis can be sensitive to the prior distributions.

3.3.1.1 Priors

For determining prior structure, in many cases the conjugate family structure can be used to adequately describe the underlying behavior. The conjugate family of priors for standard linear regression uses a multivariate normal distribution on the regression coefficients β , and the prior distribution on σ^{-2} is gamma distributed. β is normally distributed with mean μ and variance $\sigma^2 V$, while $\frac{v\lambda}{\sigma^2}$ is distributed χ^2 with v degrees of freedom.

There are a wide variety of methods for determining hyperparameters for the k^{th} model. Determination of hyperparameters used in our code for the two distributions follows the recommendations of Fernandez et al. (2001), where they formulate a benchmark set of priors for linear

regression models. This corresponds to setting μ_k equal to the sample mean of the data and letting V_k , the correlation matrix of the parameters, equal $g_{0k}X_k'X_k$ with $g_{0k} > 0$. Determination of hyperparameters of σ can be difficult and have significant impact on the results of the analysis. For this reason, we use a non-informative prior on σ such that $p(\sigma) \propto \sigma^{-1}$. The last hyperparameter, $g_{0k} = \frac{1}{n_k^2}$ if $n \leq n_k^2$, otherwise $g_{0k} = \frac{1}{n}$, where n is the number of observations and n_k is the total number of gases.

The other commonly used prior distribution assumes the same distributions but assumes that the regression coefficients are uncorrelated, so V_k is a diagonal matrix as indicated above. The correlations between variables better reflect the reality of gas identification, where many of the gases of interest are highly correlated.

The resulting Bayes factor, in closed form, comparing model i to model j , is

$$B_{i,j} = \left(\frac{g_{0i}}{g_{0i} + 1} \right)^{\binom{n_i+1}{2}} \left(\frac{g_{0j}}{g_{0j} + 1} \right)^{\binom{-n_j+1}{2}} \left(\frac{\frac{1}{g_{0j+1}}y'M_{X_j}y + \frac{g_{0j}}{g_{0j+1}}(y - \mu_j)'(y - \mu_j)}{\frac{1}{g_{0i+1}}y'M_{X_i}y + \frac{g_{0i}}{g_{0i+1}}(y - \mu_i)'(y - \mu_i)} \right)^{(n-1)/2}, \quad (3.5)$$

where $y'M_{X_j}y = y'y - y'X_j(X_j'X_j)^{-1}X_j'y$.

3.3.1.2 Priors on the Models

The specification of a prior distribution on the model space is the last step in BMA. Like the prior distributions described in Section 3.3.1 both noninformative and informative priors can be used. The inclusion of informative priors on the model space has been successful in improving predictive performance (Spiegelhalter et al. 1993, Lauritzen et al. 1994). When the models include different numbers of explanatory variables, it is possible to place a prior on the model space that will assign different probabilities on models with specific variables. For example, let

$$f(M_i) = \prod_{j=1}^p \pi_j^{\delta_{ij}} (1 - \pi_j)^{1 - \delta_{ij}}, \quad (3.6)$$

where p is the total number of explanatory variables, $\pi_j \in [0, 1]$ is the prior probability that the coefficient for predictor j , β_j , does not equal 0, and δ_{ij} equals 1 if explanatory variable j is in model M_i and zero otherwise (Hoeting et al. 1999b). If $\pi_j = 0.5$ for all j , this prior corresponds to the uniform prior described above. If $\pi_j > 0$ for all j , models with a large number of explanatory variables have less weight. Expert opinion can also be included by using different values of π_j for different j , such as if $\pi_j = 1$, explanatory variable j is forced to be included in all models. This prior does assume that the inclusion of each explanatory variable is independent of other explanatory variables, although a more complex prior distribution could be created if necessary.

The prior distribution on the model space used in this paper is the assumption that all models are equally likely, so that $p(M_i) = \frac{1}{K}$, where K is the number of models. Draper (1999) suggested a potential problem with this uniform prior for model space. If two explanatory variables are highly correlated, there may be some duplication in the models; a model is essentially counted twice by the prior. There are two potential outcomes from including models that have correlated variables (Hoeting et al. 1999b). First, the two explanatory variables could have significantly different interpretations and mechanisms and therefore describe the process from different points of view. In this instance, both models are useful and should be included. Second, if the two

explanatory variables measure the same mechanism, more thought is required in the determination of \mathcal{M} . In this instance, either one of the models should be removed to reduce the double counting or the prior distribution on the model space should take into account the correlation structure.

In the context of detecting chemical plumes, many of the chemicals in the library are highly correlated. The presence of the two or more highly correlated gases may have dramatically different interpretation and implications. Furthermore, knowing that two gases fit the process near equally well, and knowing the uncertainty in the decision can be vitally important.

3.3.2 Derivation of Posterior Model Probability

The computation of the PMPs (see Equation 3.3) plays a key role in calculating the posterior predictive distribution. The PMPs are found by first integrating out the parameters, and then using Bayes' rule to find the probability of a model given the data. For simplicity, the following derivations are assumed to be for a model M_k and notation specifying the model M_k has been dropped.

Estimating the PMP requires computing the integrated likelihood. Under a given model, M_k , the marginal likelihood is given by

$$f(Z) = \int \int \int f(Z|\beta, \sigma^2, \theta) f(\beta, \sigma^2, \theta) d\beta d\sigma^2 d\theta, \quad (3.7)$$

where, for notational simplicity, $f(Z) = f(Z|M_k)$. This computation will be done in several steps, evaluating one integral at a time.

First, the calculation of

$$f(Z|\sigma^2, \theta) = \int f(Z|\beta, \sigma^2, \theta) f(\beta | \sigma^2, \theta) d\beta$$

is considered. Write $Z = X\beta + \varepsilon$, where ε and β are independent with $\varepsilon \sim N(0, \sigma^2\Psi)$, $(\beta | \sigma^2, \theta) \sim N(\mu, \sigma^2V)$, and μ , and V are as defined in Section 3.3.1.1. The covariance of the noise ε is $\sigma^2\Psi$ and Ψ is assumed to be diagonal. Then $E[Z] = X\mu$ and $Var[Z] = \sigma^2(XVX' + \Psi)$. Because β and ε are both normally distributed, $(Z|\sigma^2, \theta) \sim N(X\mu, \sigma^2(XVX' + \Psi))$.

Next, integrating out σ^2 gives

$$f(Z|\theta) = \int f(Z|\sigma^2, \theta) f(\sigma^2|\theta) d\sigma^2 \quad (3.8)$$

$$= \int (2\pi)^{-n/2} u^{n/2} |XVX' + \Psi|^{-1/2} \exp\left\{-\frac{u}{2} \left\{ (Z - X\mu)' (XVX' + \Psi)^{-1} (Z - X\mu) \right\}\right\} \frac{\left(\frac{v\lambda}{2}\right)^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right)} (u)^{\frac{v}{2}-1} \exp\left\{-\frac{vu}{2}\right\} du. \quad (3.9)$$

Therefore,

$$\begin{aligned}
f(Z|\theta) &= \left(\frac{1}{2\lambda} \left\{ (Z - X\mu)' (XVX' + \Psi)^{-1} (Z - X\mu) + v\lambda \right\} \right)^{-\frac{n+v+1}{2}} \\
&\quad (2\pi)^{-n/2} (\lambda)^{v/2} |XVX' + \Psi|^{-1/2} \frac{\left(\frac{v}{2}\right)^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right)} \int \exp\{-s\} s^{\left(\frac{n+v}{2}-1\right)} ds \\
&= (\pi)^{-n/2} (v\lambda)^{v/2} |XVX' + \Psi|^{-1/2} \Gamma\left(\frac{v}{2}\right)^{-1} \Gamma\left(\frac{n+v}{2}\right) \\
&\quad \left(\left\{ (Z - X\mu)' (XVX' + \Psi)^{-1} (Z - X\mu) + v\lambda \right\} \right)^{-(n+v)/2}.
\end{aligned} \tag{3.10}$$

This is a multivariate noncentral Student's t distribution with mean $X\mu$, variance $\frac{v\lambda}{v-2} (XVX' + \Psi)$, and v degrees of freedom.

The PMP is determined using Bayes' rule from the product of the prior model probability $f(M_k)$ and the integrated likelihood,

$$f(M_k|Z) = \frac{f(Z|M_k) f(M_k)}{\sum_{l=1}^K f(Z|M_l) f(M_l)}. \tag{3.11}$$

3.3.2.1 Approximations

In many cases, the integration of Equation (3.7) is difficult to compute in closed form. Two approximations are commonly used to evaluate the marginal distribution for Z , a Laplace approximation and the "Bayesian information criterion" (BIC) approximation. The Laplace approximation is defined by

$$\int \exp\{L(\theta)\} d\theta \approx (2\pi)^{b/2} |\Sigma|^{1/2} \exp\{L(\hat{\theta})\} \tag{3.12}$$

where $L(\theta) = \log f(Z|\theta, M_k) + \log f(\theta|M_k)$ is the product of Equation (3.10) and the prior distribution for θ given the model, b is the dimension of θ , $\hat{\theta}$ maximizes $L(\theta)$, and Σ is the negative inverse Hessian of $L(\theta)$ evaluated at $\hat{\theta}$. This approach was used for generalized linear models by Raftery (1996), and for censored survival models by Volinsky (1997).

The Bayesian information criterion (Raftery 1996) is an approximation to the Laplace approximation, incorporating $\log |\Sigma| \approx b \log n$. Both the BIC approximation and the Laplace approximation improve as the distribution of the vector θ approaches normality, and the approximations are exact when θ is normally distributed.

In this application, we are using the priors developed in Section 3.3.1.1 for two primary reasons. First, one of the features of this set of priors is that we can calculate the posterior model probability in closed form. Second, current model development does not include prior information about the chemicals under investigation. Future work will investigate exactly what sort of prior chemical information should be incorporated into the model.

3.3.3 Derivation of Quantity of Interest Distribution

Let us consider the distribution of a general variable, Δ . In many examples, Δ can represent a quantity proportional to the gas concentration. Δ can also represent the probability that the concentration is not zero.

The posterior distribution of Δ given data Z is

$$f(\Delta|Z) = \sum_{k=1}^K f(\Delta|M_k, Z) f(M_k|Z), \quad (3.13)$$

where $f(M_k|Z)$ is the PMP for model M_k and $f(\Delta|M_k, Z)$ is the predictive distribution of the quantity of interest given the data under a specific model M_k . This BMA posterior distribution is a weighted average of the posterior predictive distributions for each model, where the weights are specified by the PMP.

4.0 Simulation

To compare the two methodologies, we need a test where the truth is known. Therefore, datasets were simulated using research software that combines an instrument model, physical radiance model, and standard atmospheric model to simulate a hyperspectral image. Noise is then added to the image before outputting the simulated signal. The first study set of images was created with four distinct stripes of background materials. The second study set was created with random assignment of 100 different background materials. Both sets were overlaid with the plume gases in distinct bands of constant concentrations. A description of how the simulated datasets were created follows.

4.1 Description

Infrared Systems Analysis in General Environments (IR-SAGE), which was authored at PNNL, is the research software that was used to create the simulated datasets. IR-SAGE has a built-in sensor noise model because it was developed to test different sensor configurations before building new sensor systems. IR-SAGE uses Monte-Carlo techniques to combine multiple sources of variability from the instrument, ground, atmosphere, and plume into a simulated instrument response. It can create atmospheric terms using FASCODE or input them from a measured response. IR-SAGE incorporates a physical radiance model similar to but distinct from the model described in Section 2.2. All sources of variability are incorporated into the received pixel value at the same time. This is done by adding noise, $N(0, nesr^2)$, where $nesr$ is noise equivalent concentration path length (please see Appendix A of Sheen et al. (2001)).

4.1.1 Wavelengths and Instruments

We used IR-SAGE to simulate images from a generic infrared spectrometer with 126 equally spaced bands covering the range of 750 to 1250 inverse centimeters. IR-SAGE automatically convolves the library spectra (both gas and background material) to the resolution of the supposed instrument using a linear interpolation. The generic spectrometer is a passive Fourier Transform Infrared (FTIR) instrument. For this simulation, we used a standard US76 atmosphere from file instead of generating one using the FASCODE bundled with IR-SAGE. The complete list of sensor error model parameters and their levels for this simulation is available in Table A.2 of Appendix A.

4.1.2 Background Materials Available

Background materials were all chosen from the Nonconventional Exploitation Factors Data System (NEFDS), which is a database of reflectance spectra, as well as surface reflection parameters used for computer rendering of objects (NIST 2001).

For Study 1, backgrounds were chosen that might resemble a scene from an rural highway in eastern Washington. The four backgrounds used were:

1. Grey paint on thin metal,
2. Matte side of black construction tar,
3. Light red/brown clay soil,
4. Reddish brown fine sandy loam soil.

Figure 4.1 shows their relative emissivities. These backgrounds were spatially distributed in

equal width stripes to form a scene. As can be seen in Figure 4.2, pixels vary due to the noise term within a concentration by material type.

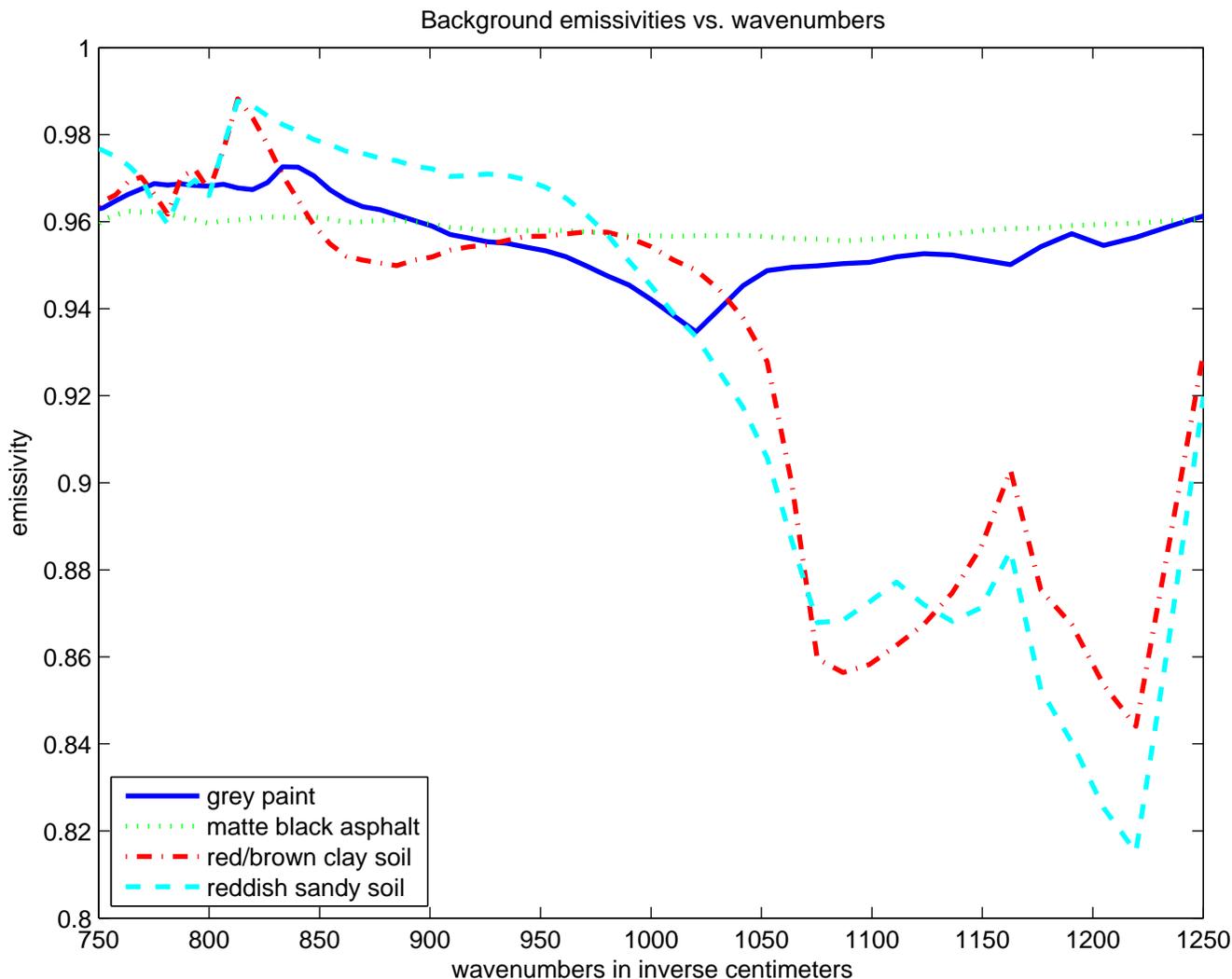


Figure 4.1. Relative emissivities of four background materials for simulated scene.

For the second study, 100 backgrounds chosen from NEFDS were randomly distributed to form a scene. Because the analysis does not use spatial information, there is no information lost by randomly assigning background materials to pixels. Table A.3 in Appendix A is a list of the backgrounds for Study 2.

4.1.3 Gases Available

The initial set of gases was chosen to represent the extremes in spectral shapes. Ethene and ammonia both have very sharp peaks that can make them easy to identify. However, if the instrument resolution is too coarse, those sharp peaks turn into broad peaks. Freon-113 and furan have much broader peaks, and could be harder to detect. Those four gases were used in Study 1. Freon-114 was included in Study 2. Figure 4.3 displays their relative shapes at the resolution of the simulated instrument.

Study 1 Scene, Full Color Contrast

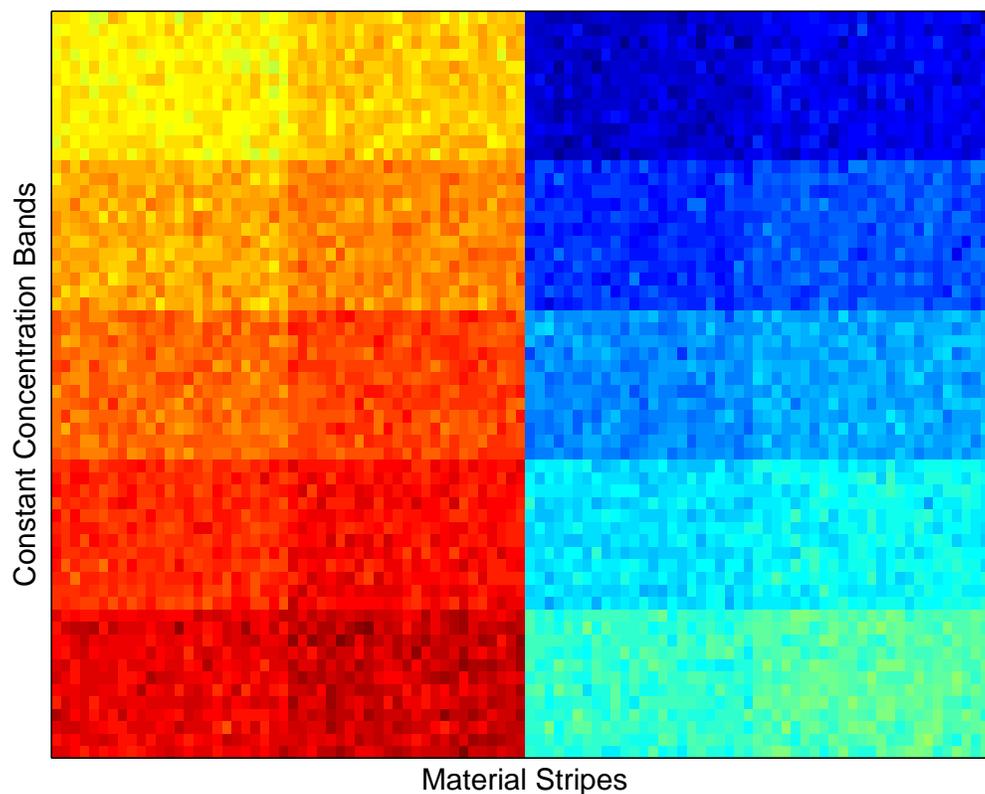


Figure 4.2. A generic scene from Study 1.

Constant concentrations of gases were implemented over a gradient addition because of ease of calculation. The metric and display that were used for detection of the gases must have repeated measures at the same concentration level. Constant concentrations allowed us to control how many pixels were available at a given level.

In Study 1, a search library was used that consisted of 14 gases in addition to the four test gases. The “confounding” gases in the search library have been used by others within the community. For Study 2, we wanted to add more complexity to the search library. Specifically, we asked whether increasing the number of gases in the library would affect which gases are detected. To assess this, we increased the number of gases to include search libraries of size 5, 10, 25, and 55 gases in addition to the five plume gases.

To further confuse the searching algorithms, we introduced two categories of search libraries: correlated and uncorrelated. The superset of 93 gases available for the search library was taken from the intersection of the LLNL, LANL, and PNNL databases. From this superset, the correlation between all gases was obtained based on their instrument resolution. The highest correlation between the plume gases and the superset was 0.59. The correlated and uncorrelated libraries are nested in that five additional gases in the C 5 library (as labeled in Table A.1 of Appendix A) are contained in the C 10 library, which is contained within the C 25 library, and those are contained in the C 55 library. The same holds true for the uncorrelated libraries. The superset is not large enough to have mutually exclusive correlated and uncorrelated full libraries, so there are some gases that are present in both the C 55 and UnC 55 libraries. Table A.1 of Appendix A has a full listing of the “confounding” gases and which libraries they belong to.

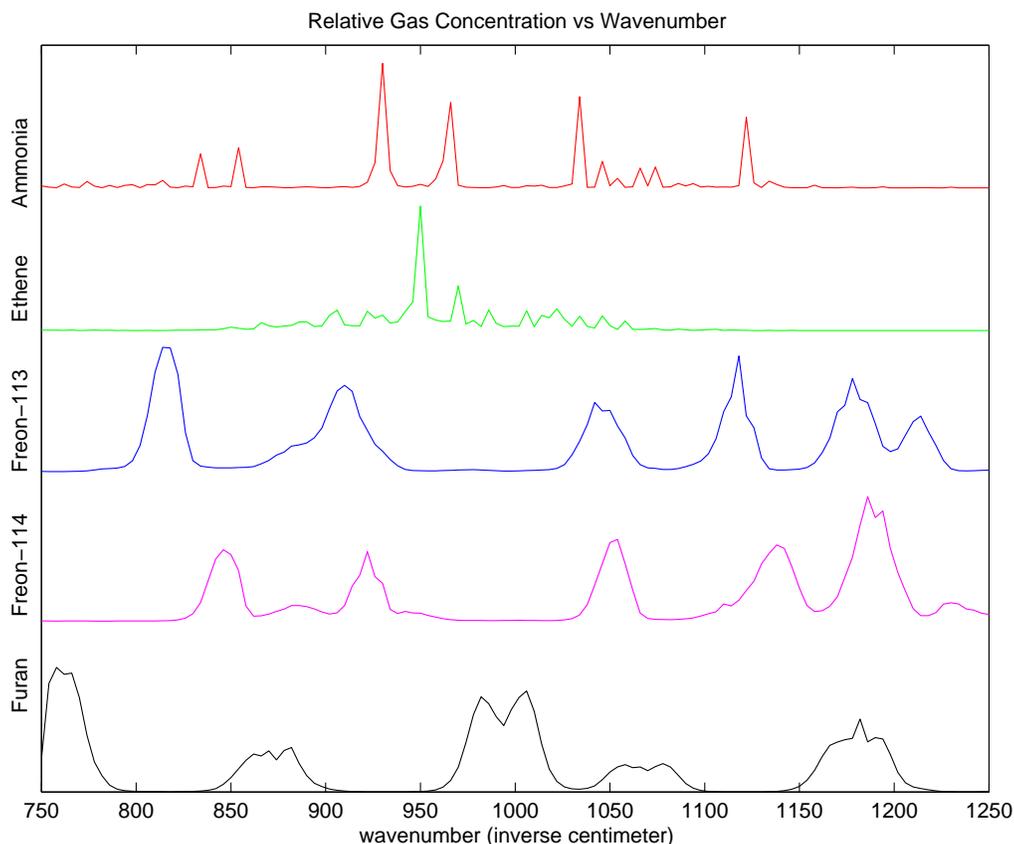


Figure 4.3. Chemical spectra for five gas plumes at instrument resolution.

4.1.4 Whitening

There are several methods for whitening an image. They all differ on which pixels are used to find the spectral covariance matrix. One method is to use all the pixels in an image. This assumes that the plume pixels are small in comparison to the non-plume pixels, and therefore contribute very little to the covariance. Another method uses a mask of some sort to pick only the non-plume pixels to calculate the covariance. This mask can be known a priori, as in simulated datasets, or can be found through an iterative process with an algorithm.

For this simulation, we used the knowledge of where the plume pixels were located within an image so that we used only the non-plume pixels in the calculation of the spectral covariance matrix. This is the best possible case because in practice we will not have a known mask to apply to an image.

4.1.5 Gas Concentrations

Through this simulation study we found that the concentrations of plume gases that can be detected are much lower than our initial guess. We first used a set of parts per million (ppm) (1, 26, 51, 75, 100). These levels were too high, dwarfing the background and confusing both algorithms, and most likely were very unrealistic. We lowered the levels to (1, 7, 13, 19, 25), which encompasses the lower limit of detection capabilities with our simulated instrument and wavenumbers. Indeed, for Study 1, F-113 required even lower concentrations (1, 3, 6, 8, 10) to not reach saturation on probability of detection. See Figures B.3 and B.4 in Appendix B for

visual display.

4.2 Comparison Metrics

Comparing the two gas selection methodologies, Bayesian model averaging and minimum AIC stepwise regression, is the main goal of this report. When analyzing a hyperspectral image, both methodologies determine the subset of gases present in a pixel-by-pixel approach. Such determination of a particular gas of interest is a decision process. The BMA and SR decision rules may be compared through their receiver operating characteristic curves, which use nuisance/false alarm probabilities and probabilities of detection as the metrics, as discussed in the following section.

4.2.1 Receiver Operating Characteristic Curve

Receiver operating characteristic (ROC) curves are used to graphically display the performance characteristics of a statistical decision rule. Such curves plot the nuisance/false alarm probabilities (N/FAPs) versus the probabilities of detection (PDs) associated with the decision rule. In the simplest decision applications, the decision is based on whether a scalar test statistic exceeds a given scalar threshold. ROC curves are calculated by varying the threshold, each setting of which determines an associated N/FAP and PD point on the ROC curve. They aid the decision maker in setting the threshold for a decision rule by deciding a tolerable N/FAP and a necessary PD, or vice versa.

Decision rules may be compared through their ROC curves. A rule whose PDs are always greater than those of another rule over the domain of acceptable N/FAPs is preferred. The best possible ROC curve would be a straight horizontal line from $(PD, N/FAP) = (1, 0)$ to $(PD, N/FAP) = (1, 1)$, corresponding to perfect chemical detection.

Unlike other SR procedures, the minimum AIC SR procedure does not involve a threshold that determines when a variable (a gas in our application) enters or leaves the model. This means that only one point in a ROC curve is produced: that point for the resulting (N/FAP, PD) pair for whether the gas of interest is included in the final model or not. Because the procedure provides a t-statistic that may be used to assess the significance of the gas of interest's coefficient, we may apply a second-stage decision on whether to remove the gas of interest from the model (if the t-statistic is too small). In this way we may set a threshold on that t-test and produce part of a ROC curve for the minimum AIC SR. It is just part of a ROC curve because we may only consider N/FAPs less than the single (N/FAP, PD) point discussed above. This second-stage ROC curve does allow us to compare the minimum AIC SR procedure with BMA, which was our research goal. As an example, Figure 4.4 plots the ROC curve for an ethene plume over a background of gray paint on thin metal.

4.2.2 Other Metrics

Once a model is selected, the fitting of the WMF produces regression coefficients that provide quantifications of the selected gases. BMA and SR could be compared through these quantifications in terms of bias (accuracy) and uncertainty (precision). However, because any analysis of quantification must always be conditional on detection, we chose to consider and compare only the detection capabilities of BMA and SR in this report. A comparison of their quantification capabilities can be addressed in a future study.

ROC for Ethene and bkgrnd 1

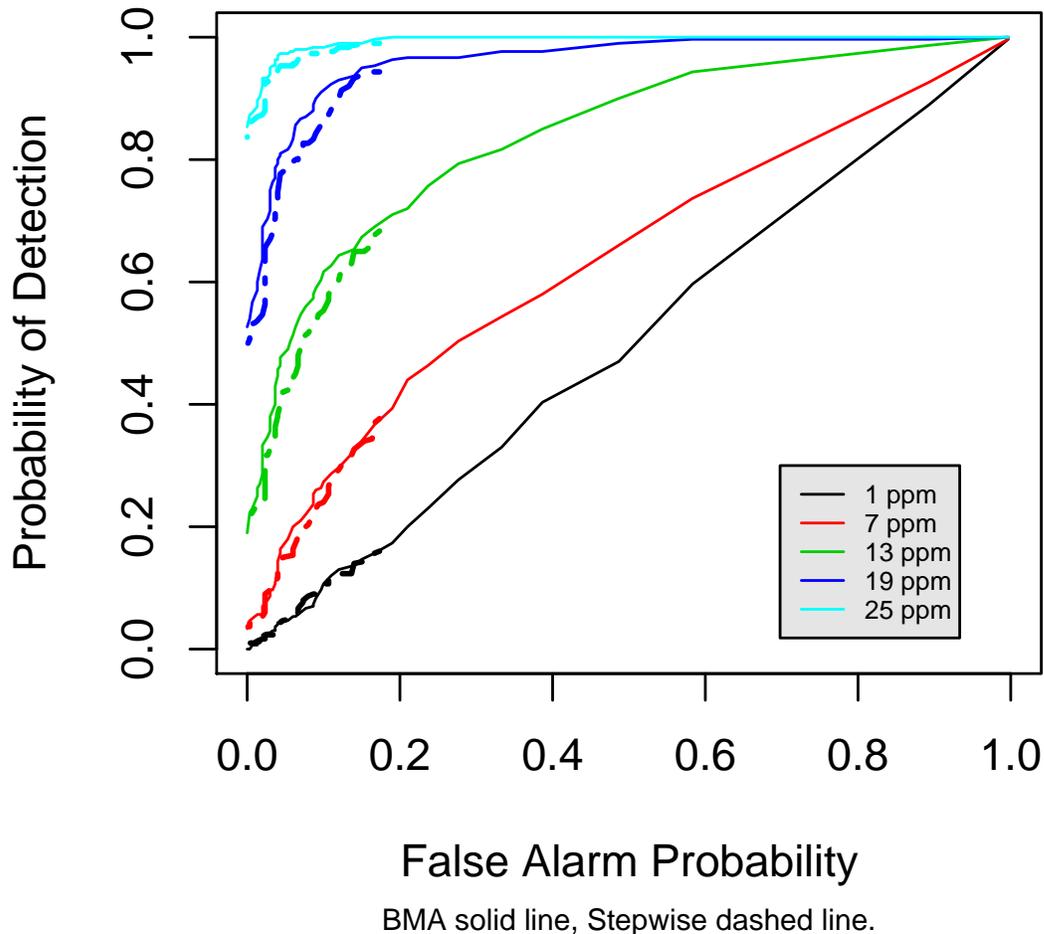


Figure 4.4. Receiver operating characteristic curve.

4.3 Simulation Study Results

The ROC curves from the two studies are presented in Appendix B. These ROC curves are graphical representations of the detection capabilities of the two gas selection methodologies, BMA and SR. All results and conclusions derive from these ROC curves.

4.3.1 Study 1

The study consisted of injections of four gases over four different background materials at six known gas concentrations (including no gas) that simulated ground truth. All 16 ROC curves suggest that BMA does slightly better than SR in detecting the four test gases, ammonia, ethene, Freon-113, and furan. For example, Figure 4.4 shows the ROC curves for ethene against the first background, grey paint on thin metal. The uncertainties in the empirical probability of detections is approximately ± 0.04 , so only the 19-ppm curve has BMA statistically significantly better than SR. In general, there is no significant difference in the BMA and SR ROC curves, but the BMA curves usually plot just above the SR curves.

Of the four test gases, Freon-113 is the most easily detected and furan is the hardest to detect in terms of concentrations required to achieve the same PDs. The four backgrounds show similar ROCs within each gas, indicating no effect in detection capability for the four test gases. Figure 4.5 gives approximate 2-sigma uncertainties as a function of the probability of detection in the ROC curves.

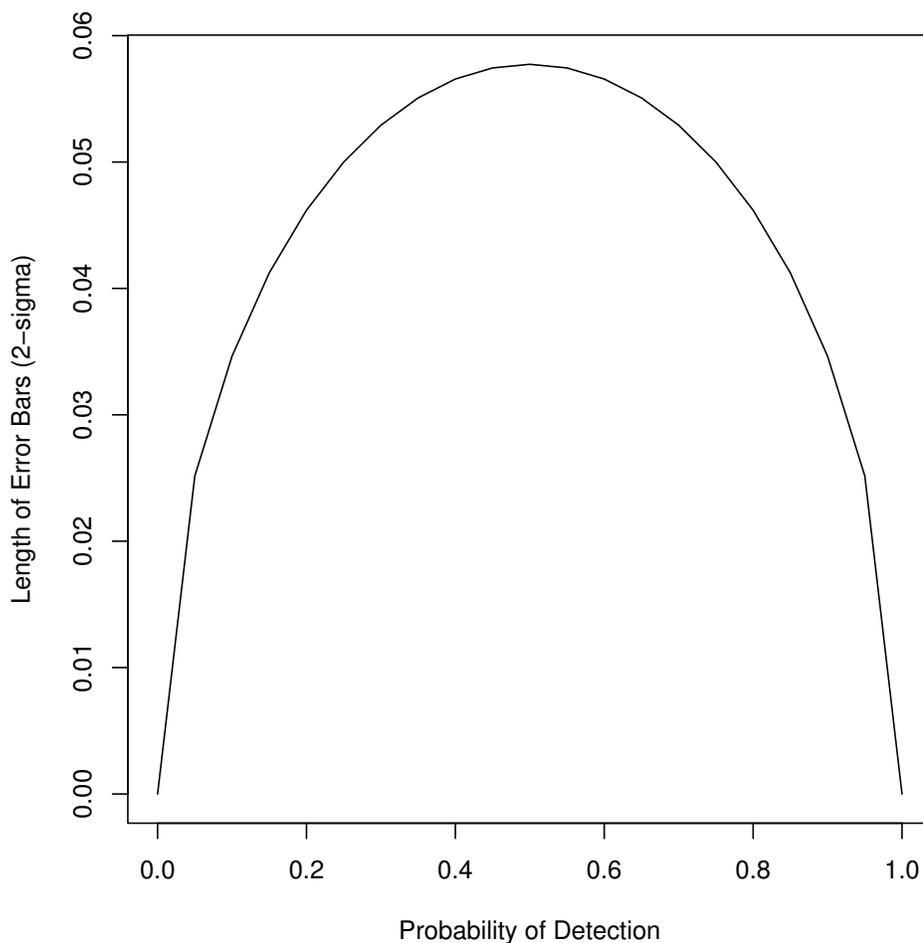


Figure 4.5. 2-sigma uncertainties for ROC curves.

4.3.2 Study 2

Forty ROC curves are presented in Figures B.6 - B.15 in Appendix B. These curves give the detection capabilities averaged across the 100 background materials considered. For ethene and furan, the BMA and SR ROC curves are similar across library size and complexity. For ammonia, Freon-113, and Freon-114, the BMA and SR ROC curves are similar for the smaller libraries. However, for the larger and more complex libraries, BMA does increasingly better than SR. This is most evident with Freon-113 and Freon-114.

5.0 Conclusions

Bayesian model averaging and minimum AIC stepwise regression were compared using simulated datasets. Bayesian model averaging did not incorporate gas information (informative priors) in the analysis. Test gases were chosen to span the types of spectra available from the PNNL library, broad peaks to sharp peaks. The size and complexity of the search library were varied. Background materials were chosen to either replicate a remote area of eastern Washington or feature many common background materials. We selected the off-plume whitening information from the same background materials as the corresponding plumes to replicate the iterative nature of off-plume pixel selection. Typically, an image analyst iteratively identifies on-plume and relevant off-plume pixels to obtain the best detection.

For many cases, BMA and SR performed the detection task comparably in terms of the ROC metric. Background materials did not affect the performance because the off-plume whitening information was selected from the same materials. For some gases, BMA did perform better than SR when the size and complexity of the search library increased.

One advantage of BMA over SR is the ability to incorporate other sources of information into the prior distributions and therefore into the decision process. For example, by using more informative priors developed for background materials collected through persistent surveillance, we expect improved BMA performance. Incorporation of prior information on gases developed through clustering the gas libraries should also lead to improved performance. Both of these sources of prior information, as well as others, have the potential to increase automation of plume detection.

6.0 Future Work

6.1 Statement of Work - FY 2006

In FY 2006 we will work in the following areas: 1) BMA, 2) prior distributions of nuisance parameters, 3) chemical universe, 4) algorithm development, 5) temporal change analysis for persistent monitoring.

6.1.1 Investigate Bayesian Model Averaging

BMA has been studied in the context of plume gas identification/quantification in hyperspectral image analysis, and significant progress has been made. The next step is to write a journal article documenting this research. Specifically, this article will address how BMA can be used to identify and quantify the composition of gases in a plume. It will also compare the performance of BMA with stepwise regression.

6.1.2 Explore Prior Distributions of Nuisance Parameters

In the hyperspectral image context, nuisance parameters refer to the image background and the other parameters in the Radiance Transfer Model. This is the part of the image that is not due to plume gases. To produce good estimates of nuisance parameters, several issues must be considered.

- How to integrate a time series of images of the same scene.
- How to combine images from different sources/sensors.
- How to perform image segmentation and what is the advantage of segmenting the image into nearly homogeneous regions.

The outcome of this research will have significant impact on algorithm development.

6.1.3 Chemical Universe

This task will continue development of methods to identify groups of gases with similar spectra. These results will be combined with sensor characteristics to construct a chemical triage method that produces a hierarchy of chemical groupings based on measured hyperspectral data.

6.1.4 Algorithm Development - Image Clustering

A critical initial step in this project is to identify the pixels where the plume is present. Typically a matched filter is used to detect a weak target signal in hyperspectral images. This requires an estimate of the spectral covariance of the image. In most cases the image will not be homogeneous, meaning that it will consist of regions that have distinct spectral statistics. If the image can be segmented or clustered into statistically homogeneous regions, a matched filter optimized for the individual regions can be constructed. These filters are based on the covariance matrix approximated only by the pixels in the region. In general, the local matched filter is no worse than the matched filter based on the whole image and may produce considerably better results.

Clustering results in a set of matched filters for an image, one for each region or cluster. We then apply the cluster-specific matched filters to detect the target signal. This approach has been shown to improve detection of weak target signals significantly compared with using a matched

filter based on the covariance of the whole image (Funk et al. 2001).

The image clustering problem is nontrivial. Common approaches include k-means, fuzzy c-means, hierarchical, and mixture of Gaussian. Each approach has strengths and weaknesses. We propose developing Bayesian image segmentation techniques similar to the work by Neher (2004). He used spatial statistics including image gradients in a Markov random field setting to label or classify image regions.

To emphasize the advantage of accurately segmenting the image, we provide the following discussion and example. We begin by outlining the match filter signal detection approach for a nonsegmented image. We model the image as a linear combination of a target signal s multiplied by strength α and a background that consists of a constant u (the mean radiance over the entire image) and a zero-mean noise term ε . Then the radiance of an image pixel can be written as the sum:

$$r = \alpha s + u + \varepsilon .$$

The random background noise term ε represents all sources of noise in the image, including sensor noise, background variation, inhomogeneous atmospheric effects, etc. We want to determine the linear filter q that will maximize the signal-to-noise ratio. The linear filter q is applied to each pixel r in the image to produce a scalar filter image by taking the inner product of q with each pixel r :

$$q^T r = \alpha q^T s + q^T u + q^T \varepsilon .$$

The first term is proportional to target signal strength α , the second term can be subtracted out given an estimate for u , and the variance of the third term is $q^T E[\varepsilon \varepsilon^T] q$. So the signal-to-noise ratio for the filter q is

$$\frac{S}{N} = \frac{\alpha q^T s}{\sqrt{q^T E[\varepsilon \varepsilon^T] q}} .$$

Now let $R = E[\varepsilon \varepsilon^T]$. To find q we solve the following optimization problem:

$$\max_q \frac{S}{N} = \max_q \frac{\alpha q^T s}{\sqrt{q^T R q}} . \quad (6.1)$$

We assume R is symmetric and that it can be factored as $R = R^{\frac{1}{2}} R^{\frac{1}{2}}$. Letting $y = R^{\frac{1}{2}} q$, the objective function in Equation (6.1) becomes

$$\begin{aligned} \frac{\alpha q^T s}{\sqrt{q^T R q}} &= \frac{\alpha y^T R^{-\frac{1}{2}} s}{\sqrt{y^T y}} \\ &= \alpha \hat{y}^T R^{-\frac{1}{2}} s , \end{aligned}$$

where \hat{y} is a unit vector. So the optimization problem (6.1) becomes

$$\max_{\|y\|=1} \alpha y^T R^{-\frac{1}{2}} s .$$

Thus, y is the unit vector in the direction of $R^{-\frac{1}{2}} s$. Or

$$y = \frac{R^{-\frac{1}{2}} s}{\sqrt{s^T R^{-1} s}} ,$$

and

$$q = R^{-\frac{1}{2}} y = \frac{R^{-1} s}{\sqrt{s^T R^{-1} s}}. \quad (6.2)$$

The usual estimate for R is

$$R \approx \frac{1}{N} \sum_{i=1}^N (r_i - u)(r_i - u)^T.$$

If the signal is absent, this is exactly what is needed because R is the covariance of the noise. On the other hand, if the signal is present, it will degrade the approximation. Because we are interested in detecting weak signals, we will assume the effect of the signal is not significant. The q given in Equation 6.2 is then used to construct the matched filter image. Due to numerical instabilities, care must be taken when constructing $R^{-1} s$.

If the image is clustered or segmented into k nonintersecting regions, a matched filter q_i is constructed for each region:

$$q_i = \frac{R_i^{-1} s}{\sqrt{s^T R_i^{-1} s}} \quad i = 1, \dots, k. \quad (6.3)$$

Then the matched filter image is constructed by taking the inner product $q_i^T r$ over each region.

To illustrate the advantage of this approach, we perform a simple experiment. The test image is one spatial dimension with a spectral vector of length 4 at each of 256 pixels. The image consists of two clusters with a weak plume present in the region where the two clusters meet, which is illustrated in Figure 6.1. In Figure 6.2 the ROC curves are given for matched filter detection on the unclustered versus the clustered image. We see the significant advantage of clustering the image before applying the matched filter. This emphasizes the importance of a robust image-clustering algorithm.

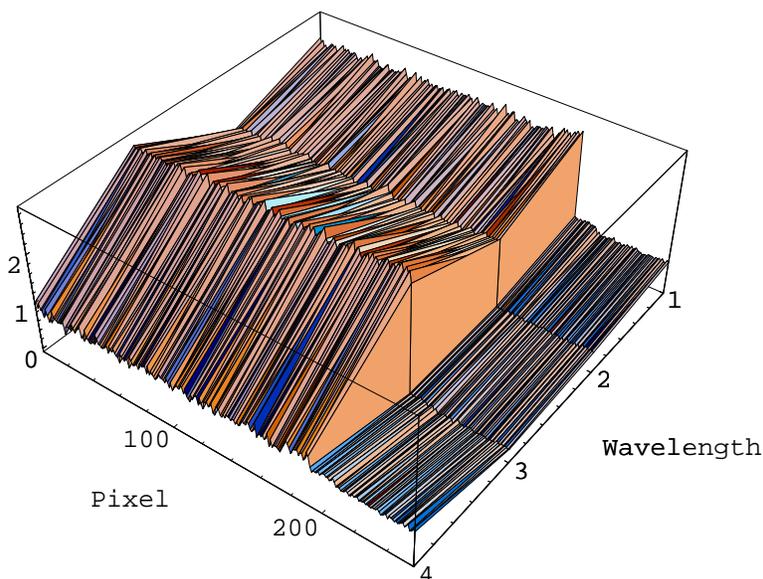


Figure 6.1. 1D image with four spectral bands, and weak plume at intersection of regions.

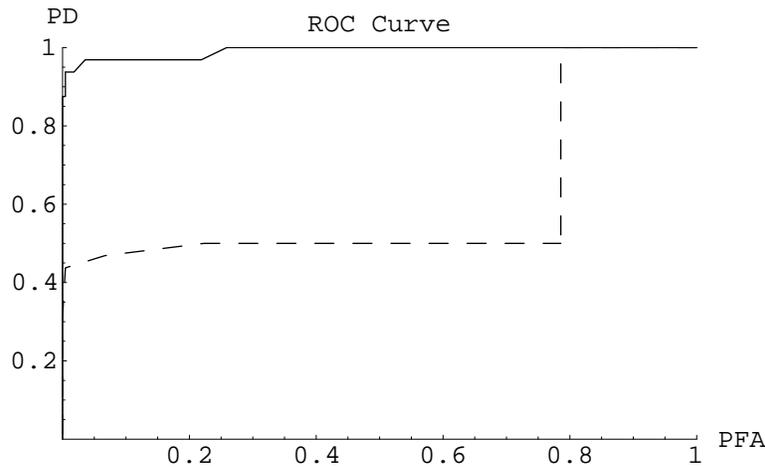


Figure 6.2. ROC curves for clustered (solid line) and unclustered (dashed line) matched filter detection.

6.1.5 Temporal Change Analysis for Persistent Monitoring

The goal of this task is to use a sequence of images of the same scene taken over time to improve the ability to detect, track, and quantify gas plumes. Issues such as image registration, approximating prior distribution of the image background, image segmentation, and change detection will all be considered.

6.2 Long-Term Planning

Even though the project SOW for FY 2006 does not address these issues, we propose they be given preliminary attention and planning. These include validation and testing as well as code prototype development.

6.2.1 Model Validation/Testing

Obtaining a high-fidelity ground truth dataset for plume detection is probably not possible. As a result, our current approach is to use synthetic data. Most journal articles in this area also rely on synthetic data. We would like to go beyond synthetic images if possible and propose a two-pronged approach. First, to find out what validation data are available. Efforts have been made to produce actual test images, and we need to take advantage of this progress.

6.2.2 Prototype Development

Now is a good time to initiate prototype development so that algorithm development and testing can be done in a way that supports the final product as much as possible. Several issues must be addressed to guide this process. These include development language/environment, computer platform, intended end user, throughput requirement, etc. We would like to clarify these issues and initiate prototype development in FY 2006.

pnnl]

7.0 References

- Aires F, W Rossow, N Scott, and A Chedin. 2002. . “Remote sensing from the infrared atmospheric sounding interferometer instrument 2. Simultaneous retrieval of temperature, water vapor, and ozone atmospheric profiles.” *Journal of Geophysical Research* 107(D22).
- Akaike H. 1983. . “Information measures and model selection.” *Bulletin of the International Statistical Institute* 50:227–290.
- Bates JM and CWJ Granger. 1969. . “The combination of forecasts.” *Operational Research Quarterly* 20:451–468.
- Berger JO. 1985. . *Statistical Decision Theory and Bayesian Analysis*, second edition. Springer-Verlag: New York.
- Berger JO, VD Oliveira, and B Sanso. 2000. . *Objective Bayesian Analysis of Spatially Correlated Data*. 00-12, Institute of Statistics and decision Sciences, Duke University.
- Burnham KP and DR Anderson. 1998. . *Model Selection and Inference: A Practical Information Theoretic Approach*. Springer: New York.
- Chatfield C. 1995. . “Model Uncertainty, Data Mining and Statistical Inference.” *Journal of the Royal Statistical Society, Series A* 158:419–466.
- Clyde M. 1999. . “Comment on ‘Bayesian Model Averaging: A Tutorial’.” *Statistical Science* 14:401–404.
- Cochran WG. 1937. . “Problems in the analysis of a series of similar experiments.” *Journal of the Royal Statistical Society (supplement)* 4:102–118.
- Draper D. 1995. . “Assessment and Propagation of model uncertainty.” *Journal of the Royal Statistical Society, Series B* 57:45–97.
- Draper D. 1999. . “Comment on ‘Bayesian Model Averaging: A Tutorial’.” *Statistical Science* 14:405–409.
- Draper N and H Smith. 1981. . *Applied Regression Analysis*, second edition. John Wiley & Sons, New York.
- Fernandez C, E Ley, and M Steel. 2001. . “Benchmark priors for Bayesian model averaging.” *Journal of Econometrics* (100):381–427.
- Funk C, J Theiler, D Roberts, and C Borel. 2001. . “Clustering to improve matched filter detection of weak gas plumes in hyperspectral thermal imagery.” *IEEE Transactions on Geoscience and Remote Sensing* 39(7).
- Furnival GM and RW Wilson. 1974. . “Regression by leaps and bounds.” *Technometrics* 16:499–511.
- George EI and R McCulloch. 1993. . “Variable selection via Gibbs sampling.” *Journal of the American Statistical Association* 88:881–889.
- Hoeting J, D Madigan, A Raftery, and C Volinsky. 1999a. . “Bayesian Model Averaging: A Tutorial.” *Statistical Science* 14(4):382–417.
- Hoeting JA. 1994. . *Accounting for Model Uncertainty in Linear Regression*. PhD thesis, University of Washington.

- Hoeting JA, D Madigan, AE Raftery, and CT Volinsky. 1999b. . “Bayesian Model Averaging: A Tutorial with Discussion.” *Statistical Science* 14:382–417.
- Kass RE and AE Raftery. 1995. . “Bayes Factors.” *Journal of the American Statistical Association* 90:773–795.
- Laplace PS. 1818. . *Deuxiem Supplement a la Theorie analytique des Probabilities*. Gauthier-Villars: Paris.
- Lauritzen SL, B Thiesson, and DJ Spiegelhalter. 1994. . “Diagnostic systems created by model selection methods: a case study.” *Uncertainty in Artificial Intelligence*, Cheeseman P and W Oldford, eds., pp. 143–152. Springer: Berlin.
- Leamer EE. 1978. . *Specification Searches*. Wiley: New York.
- Liou K. 2002. . *An Introduction to Atmospheric Radiation*, second edition. Academic Press, California.
- Madigan D and AE Raftery. 1994. . “Model Selection and accounting for model uncertainty in graphical models using Occam’s Window.” *Journal of the American Statistical Association* 89:1535–1546.
- Madigan D and J York. 1995. . “Bayesian graphical models for discrete data.” *International Statistical Review* 63:215–232.
- Marino SA. 1999. . “Operation and Calibration of the NPS Ultraviolet Imaging Spectrometer (NUVIS) in the Detection of Sulfur Dioxide Plumes.” Master’s thesis, Naval Postgraduate School.
- Neher RE Jr.. 2004. . *A Bayesian mrf framework for labeling terrain using hyperspectral imaging*. PhD thesis, Florida State University.
- NIST. 2001. . *Nonconventional Exploitation Factors Data System*. National Institute of Standards and Technology, <http://math.nist.gov/FHunt/appearance/nefds.html>.
- O’Donnell E, D Messinger, C Salvaggio, and J Schott. 2004. . “Identification and detection of gaseous effluents from hyperspectral imagery using invariant algorithms.” In *Proceedings of the SPIE, Sensor Data Exploitation and Target Recognition, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery X, Vol. 5425*.
- Raftery AE. 1995. . “Bayesian Model Selection in Social Research (with Discussion).” *Sociological Methodology* 25:111–196.
- Raftery AE. 1996. . “Approximate Bayes factors and accounting for model uncertainty in generalized linear models.” *Biometrika* 83:251–266.
- Roberts HV. 1965. . “Probabilistic prediction.” *Journal of the American Statistical Association* 60:50–62.
- Salas S and E Hille. 1995. . *Calculus: one and several variables*, seventh edition. John Wiley & Sons, New York.
- Sheen D, B Wise, N Gallagher, S Sharpe, P Heasler, K Anderson, and J Schultz. 2001. . *Infrared Chemical Detection Systems Modeling and Advanced Chemometric Analysis*. PNNL-13737, Pacific Northwest National Laboratory, Richland, WA.

- Spiegelhalter DJ, Dawid A, Lauritzen S, and Cowell R. 1993. . “Bayesian analysis in expert systems.” *Statistical Science* 8:219–283.
- Stigler SM. 1973. . “Studies in the History of Probability and Statistics. XXXII: Laplace, Fisher and the Discovery of the Concept of Sufficiency.” *Biometrika* 60:439–445.
- Stocker A. 2000. . *HIRIS Baseline Algorithms for Level 3 Analysis, Version 2.0*. UCRL-CR-140736, Lawrence Livermore National Laboratory, Livermore, CA.
- Venables W and Ripley B. 2002. . *Modern Applied Statistics with S*, fourth edition. Springer Verlag, New York.
- Villeneuve P and Stocker A. 2000. . *HIRIS Algorithm Physics Model Description, Version 2.3*. UCRL-CR-141865-B346001, Lawrence Livermore National Laboratory, Livermore, CA.
- Volinsky CT. 1997. . *Bayesian Model Averaging for Censored Survival Models*. PhD thesis, University of Washington.
- Volinsky CT, Madigan D, Raftery AE, and Kronmal RA. 1997. . “Bayesian model averaging in proportional hazard models: assessing the risk of a stroke.” *Journal of the Royal Statistical Society, Series C* 46:433–448.
- Young S. 2002. . *Detection and Quantification of Gases in Industrial-Stack Plumes Using Thermal-Infrared Hyperspectral Imaging*. ATF-2002(8407)-1, The Aerospace Corporation, El Segundo, California.

Appendix A

Simulation Variables

Appendix A – Simulation Variables

Full Gas Library

C refers to correlated and UnC refers to uncorrelated for Study 2.

Table A.1. Full gas library for Studies 1 and 2.

Gas	Study1	UnC 5	C 5	UnC 10	C 10	UnC 25	C 25	UnC 55	C 55
ACETOIN							X		X
Acetone	X								
ACROL							X		X
ALLYLAM						X		X	X
ASH3						X		X	
BENZENE		X		X		X		X	
BPINENE								X	
BUTANE			X		X		X		X
BUTENE1							X		X
BZALDEH								X	X
C13DCLP						X		X	X
C2H2								X	X
C3H8								X	
C4F10								X	X
C6H5BR								X	X
C6H5F									
CCLF3									X
CH3CL	X								
CH4									
CH4									
CHBR3									
CHCL2F								X	X
CHF3								X	
CLACET		X		X		X		X	X
CLETHOH									
CLPICRIN								X	
CLTOL2						X		X	
CLTOL3									X
CLTOL4						X		X	
CUMENE					X	X	X	X	X
CYCHEXE				X		X		X	
CYCLOHX									
CYCLOPR									
DCBZ12								X	
DCBZ13								X	X
DCBZ14				X		X		X	
DCE11								X	
DCLM			X		X		X		X
DCLP12								X	X
DCLP13								X	X
DEA							X		X

Continued on next page...

Table A.1 – Continued

Gas	Study1	UnC 5	C 5	UnC 10	C 10	UnC 25	C 25	UnC 55	C 55
DIETHANIL				X		X	X	X	X
DIMAMINE								X	X
DMHYDRZ		X	X	X	X	X	X	X	X
DMMP	X								
DODEC	X					X		X	X
EDA				X		X		X	X
EDB			X		X				X
EDC12								X	
EPX12BUT								X	X
ETHENE	X	X	X	X	X	X	X	X	X
ETNO2									
F113	X	X	X	X	X	X	X	X	X
F114	X	X	X	X	X	X	X	X	X
F12						X		X	X
F125				X		X		X	X
F218						X		X	X
FURAN	X	X	X	X	X	X	X	X	X
HCFC124									X
HCFC142B					X				X
HCHO						X		X	
HEPTENE		X		X		X		X	X
HFC134									X
HNO3	X								
HYDRAZINE					X				X
IBUTANE					X		X	X	X
ISBTENE						X		X	
ISOCT								X	
ISOPREN						X		X	X
ISPROAM								X	
MBUTENE									
MCDS									
ME2PEN2									
ME3BUT1								X	X
ME3PEN		X		X		X	X	X	X
MENO2						X		X	
MEOH	X		X		X	X	X	X	X
MMAM								X	X
N2O	X								
NBUTALD									X
nButanol	X							X	X
NCCN								X	X
NH3	X	X	X	X	X	X	X	X	X
NITROPROP							X		X
NO2	X								
PERC	X								
PH3							X		X
PHOSGENE								X	
PROPYNE									
PRTHIOL1									X

Continued on next page...

Table A.1 – Continued

Gas	Study1	UnC 5	C 5	UnC 10	C 10	UnC 25	C 25	UnC 55	C 55
PRTHIOL2									
SO2	X								
T12DCE							X		X
T13DCLP					X		X	X	X
TBPscaledtoTIPP	X								
TCE	X					X		X	X
TCE1112							X		X
TCE1122									
TEA								X	X
TFAA								X	X
THF								X	
TMA							X		X

IR-SAGE Information

Table A.2. Simulated sensor error model values

Parameter	Value
spectrometer type dispersive of fts	fts
modulation efficiency	1.0
jitter parameter	0.0
detector quantum efficiency	0.35
efficiency of the telescope	0.9
efficiency of the cold filter	0.9
efficiency of the spectrograph	0.8
pixel area in cm ²	$\frac{\pi}{4}0.005^2$
F number	3.5
emissivity of the spectrograph	0.05
emissivity of the cold filter	0.05
emissivity of the cold shield	0.05
emissivity of the fore-optics	0.1
lowest wavenumber for detector response	750.0
low-frequency cutoff cm ⁻¹	750.0
high-frequency cutoff cm ⁻¹	1250.0
cutoff for infinite black-body integration	10000.0
temperature of spectrograph body	10.0
temperature of coldshield	10.0
temperature of fore-optics	195.0
integration time (seconds)	0.004*128
resolution cm ⁻¹	1.0
instrument line shape	triangular
electrons/sec due to 1/f noise	1.5
electrons/sec due to readout	5
electrons/sec due to ADC	5
daylight (1) or night(0)	1
solar zenith angle	$\frac{\pi}{4}$
atoms tras from alt to grnd	0.0 con
atmos. transmissivity	us76-nadir-01km-tr.spc
atmos. radiance	us76-nadir-01km-ra.spc
ground reflectances	refl-4
telescope diameter (m)	0.15
src types are ambient, heated, sender	ambient
heated src diam or sender tel diam (m)	0.15
sender src elem diam (m)	1e-3
src or sender tel area cm ²	0.8888
heated src brightness temp (K)	425.0
heated src with sender tel brightness temp (K)	1300.0
range to src or background	100.0
plume length (m)	1.0
temperature of the plume (K)	310.0
temperature of the scene (K)	300.0

Study Background Information

Table A.3. Backgrounds for Study 2.

Filename	Description
0002UUUPLS	black plastic sheet
0005UUUFAB	light-brown ripstop nylon fabric cloth, translucent
0011UUUFAB	green fabric cloth
0014UUUALM	gray oxidized metal aluminum plate
0015UUUCAM	green/green plastic camouflage net
0020UUUCAM	white/white coated fabric camouflage net, translucent
0028UUUCNT	brown felt camouflage net, translucent
0034UUUCNT	green cloth camouflage net, multiple layers
0039UUUCNT	black plastic wire-like camouflage net, translucent
0050UUUPRI	Yellow paint on metal
0059UUUPNP	dark-green paint on thin metal
0069UUUPNT	gray paint on thin metal
0075UUUPNT	brown paint on thin metal
0081UUUPNT	black paint on thin metal
0083UUUMTL	gray metal aluminum
0085UUUPNT	brown paint on metal
0086UUUABS	absorber (tcmm) used for scattering of radar impulses
0101UUUCIN	cinders from railroad bed
0103UUUSND	[very dark brown] black? sand soil
0108UUUCLO	brown metallized woven fabric cloth (clean)
0113UUUCNT	green/brown rubberized fabric camouflage net (brown side)
0121UUUCNT	camouflage netting of synthetic garnish
0123UUUCLO	brown metallized woven fabric cloth (weathered)
0128UUUSOL	brown fine sandy loam soil
0133UUUSOL	reddish-brown fine sandy loam soil
0136UUUSOL	brown to dark-brown loamy sand soil
0139UUUSOL	brown silty loam soil
0145UUUSOL	brown loamy sand soil
0148UUUSOL	reddish-tan silty loam soil
0151UUUSOL	light-tan fine silty loam soil
0205UUUCLO	brown metallized woven fabric cloth (weathered)
0209UUUSOL	very dark grayish-brown silty loam soil
0212UUUSOL	brown loam soil
0215UUUSOL	dark-brown fine sandy loam soil
0218UUUSOL	very dark grayish-brown loam soil
0221UUUSOL	reddish-brown fine sandy loam soil
0224UUUSOL	dark grayish-brown & very dark brown stony coarse sandy loam soil
0227UUUSOL	brown sandy loam soil
0332UUUMTL	gray titanium, thick metal block, textured surface
0339UUUCBN	dark-gray variegated carbon-teflon composite tile, grooved
0342UUURBR	dark-red foam rubber tile, vesicular, bubbles
0343UUUCBN	dark-gray carbon-phenolic composite tile, grooved
0345UUUEXY	orange carbon-epoxy composite tile, grooved, translucent
0350UUUSIL	white very translucent fused silica glass composite tile
0355UUUCBN	dark-gray carbon-carbon composite tile, alternating dark/light bands
0357UUUGLS	white fiberglass composite tile, slightly translucent, distinct pattern

Continued on next page...

Table A.3 – Continued

Filename	Description
0360UUUCBN	dark-gray carbon-phenolic composite tile, wavy/linear pattern, threads
0366UUUCBN	dark-gray carbon-carbon composite tile, small alternating dark/light bands
0374UUUEXY	orange fiberglass-epoxy composite tile, pattern, slightly translucent
0377UUUSIL	white rubber coating on yellow silica-silicone composite tile
0380UUUSIL	orange rubber coating on orange-brown silica-phenolic composite tile, glossy surface
0385UUUALM	Green paint (1 coat) on aluminum
0388UUUALM	gray uncoated weathered metal aluminum plate, grooved surface
0391UUUEXY	very light brown translucent epoxy-fiberglass composite plate, grooved surface
0394UUUEXY	clear paint (1 coat) on translucent epoxy-fiberglass composite, uniform coating
0398UUUCNC	variegated light/dark gray construction concrete, smooth cut exposed interior aggregate
0406UUUWOD	black paint (2 coats) on construction wood, grain visible
0409UUUWOD	light-brown uncoated clean wood board (ponderosa pine), distinct grain visible
0413UUUBRK	brick
0415UUUASP	black construction asphalt core, exterior surface, weathered, soiled and worn down
0417UUUASPi	variegated black/dark gray construction asphalt core, exposed interior aggregate
0419UUUIRN	green paint (2 coats) on iron [steel] metal, translucent
0422UUUIRN	black paint (1 coat) on iron [steel] metal
0424UUUCMT	variegated light-brown/gray construction concrete, embedded rock aggregate
0426UUUASP	asphalt
0428UUUFIGI	light-gray paint on fiberglass composite, 2-panels of large 4-panel checker-board
0430UUUMTL	light-gray paint on metal
0435UUUFIG	light-green paint on fiberglass, weathered, originally bright green
0437UUUFAB	light-gray, nearly white, metallized [plastic-coated?] fabric cloth, 60%, translucent
0440UUUFAB	dark-gray metallized fabric cloth, 12%, translucent
0443UUUCAMb	green-brown rubberized camouflage net (brown side)
0444UUUFAB	black plasticized [metallized] fabric cloth, old [calibration?] panel, orientation arrow
0449UUUPNT	dark-brown paint on metal, uniform surface, front/back indistinguishable
0451UUUPNT	dark-green paint on metal, uniform surface, front/back indistinguishable
0453UUUFAB	black plasticized [metallized] fabric cloth, new [calibration?] panel, orientation arrow
0457UUUSOL	light-brown, sandy, pebbly gravel soil, dirt from parking lot, Elephant Butte Lake, NM
0459UUUPNT	thin cardboard with green paint
0491UUUASP	weathered asphalt from helipad
0495UUUSTD	Plasmagold, part # ACT-02B
0498UUUSTD	Plasmagold, part # ECF-035
0506UUUPNT	aluminum painted with flat black paint
0521UUUPNT	weathered, painted galvanized steel from rooftop

Continued on next page...

Table A.3 – Continued

Filename	Description
0522UUUTAR	heavily weathered, tarred roofing paper from rooftop
0525UUUSTLb	galvanized steel from rooftop, one side weathered
0534UUUIGN	igneous rock from lava flow
0537UUUPNT	aluminum painted low-emissivity green
0539UUUALM	polished aluminum
0561UUUGRV	weathered rock ground cover
0588UUUCNC	weathered concrete outer wall
0591UUUCNC	weathered concrete outer wall
0596UUUSTO	weathered rock (fill material)
0598UUUCNC	weathered concrete apron over aggregate base
0599UUUASP	weathered asphalt over aggregate base
0627UUUCAMb	weathered camouflage net with brown and green sides
0629UUUPNT	weathered, green-painted plastic chip
0667UUUPNT	unweathered steel painted two coats black
0674UUUASP	weathered asphalt runway
0705UUUASP	weathered asphalt shingle
0706UUUPNT	unweathered black paint on plywood
0709UUUPNT	dark-gray paint on aluminum - aircraft

Appendix B

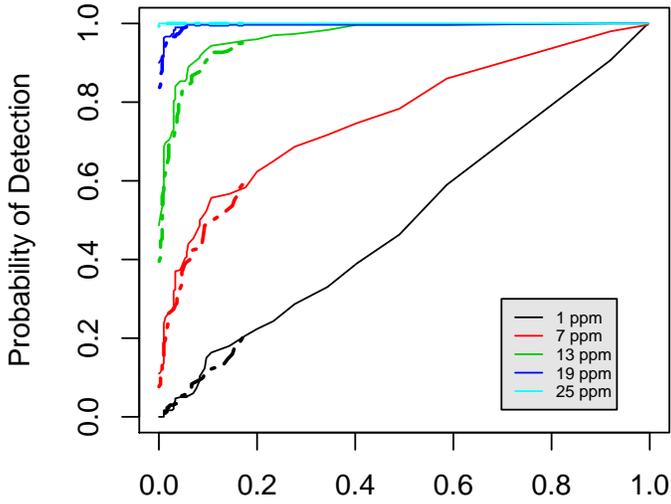
Study Figures

Appendix B – Study Figures

Study 1

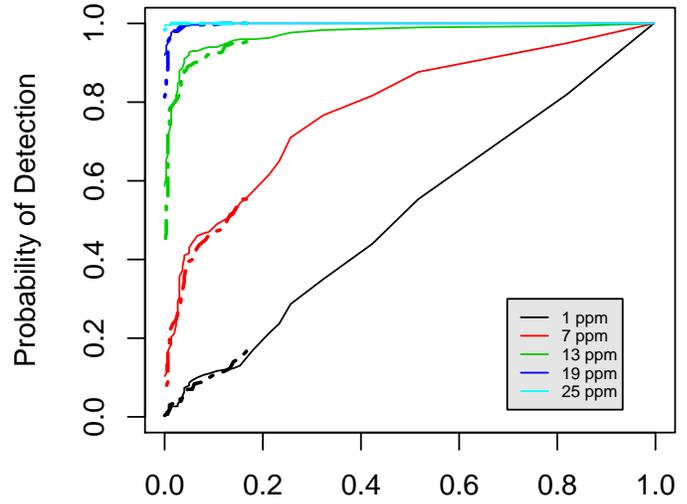
Complete set of ROC curves from Study 1.

ROC for Ammonia and bkgrnd 1



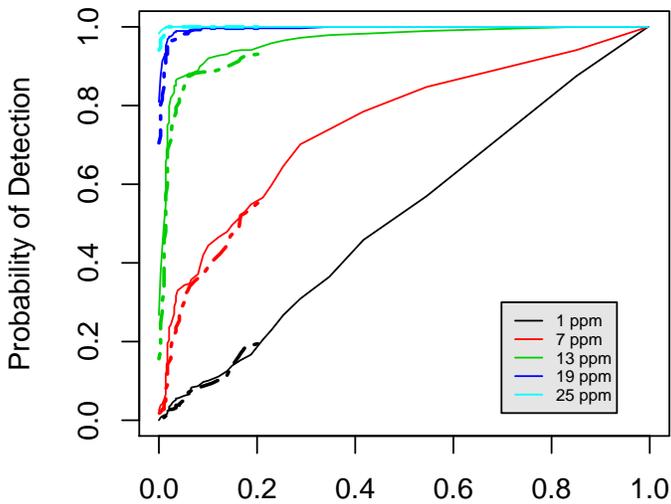
False Alarm Probability
BMA solid line, Stepwise dashed line.

ROC for Ammonia and bkgrnd 2



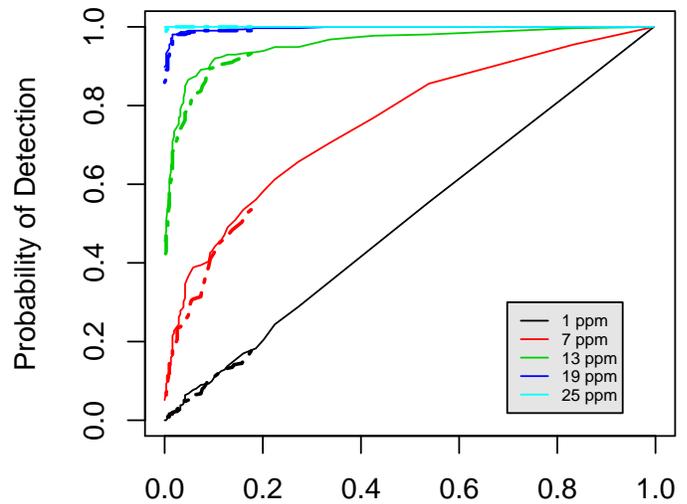
False Alarm Probability
BMA solid line, Stepwise dashed line.

ROC for Ammonia and bkgrnd 3



False Alarm Probability
BMA solid line, Stepwise dashed line.

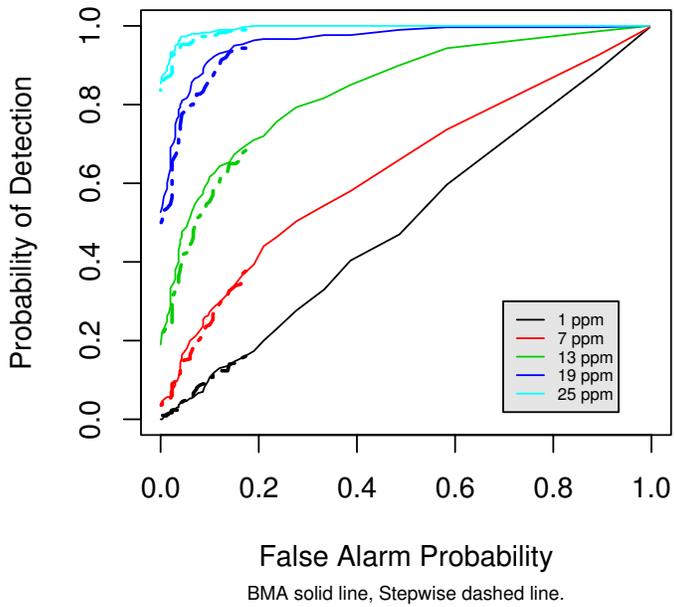
ROC for Ammonia and bkgrnd 4



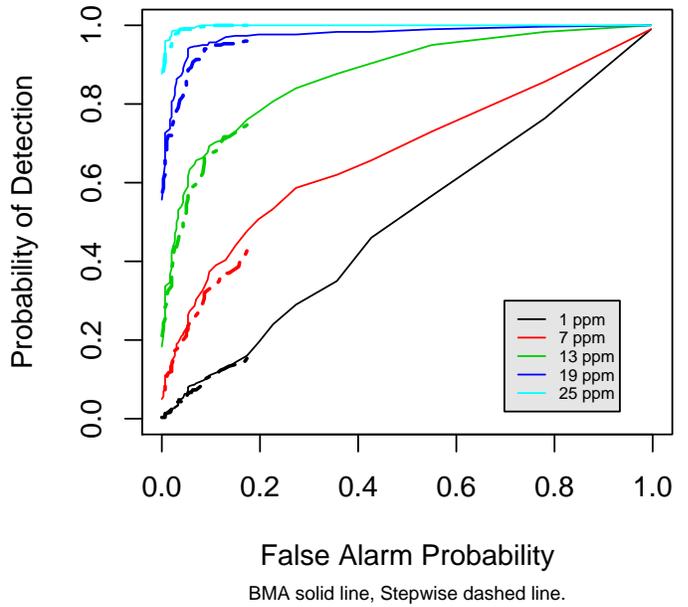
False Alarm Probability
BMA solid line, Stepwise dashed line.

Figure B.1. Receiver operating characteristic curve for ammonia.

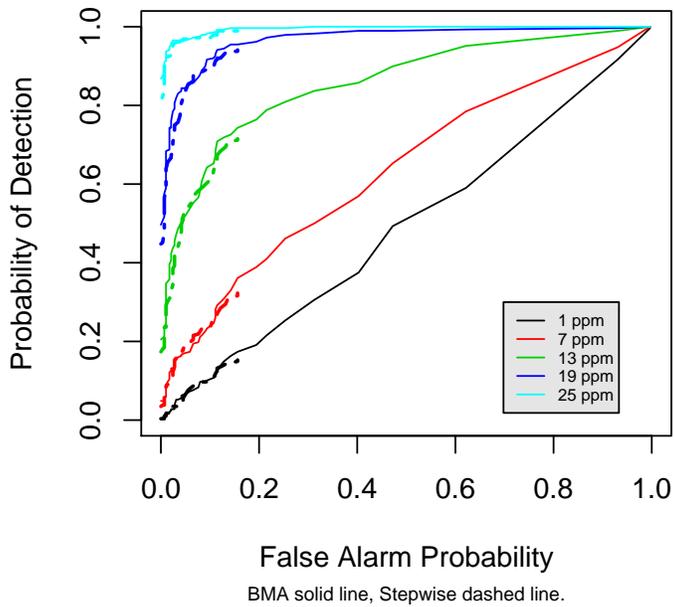
ROC for Ethene and bkgrnd 1



ROC for Ethene and bkgrnd 2



ROC for Ethene and bkgrnd 3



ROC for Ethene and bkgrnd 4

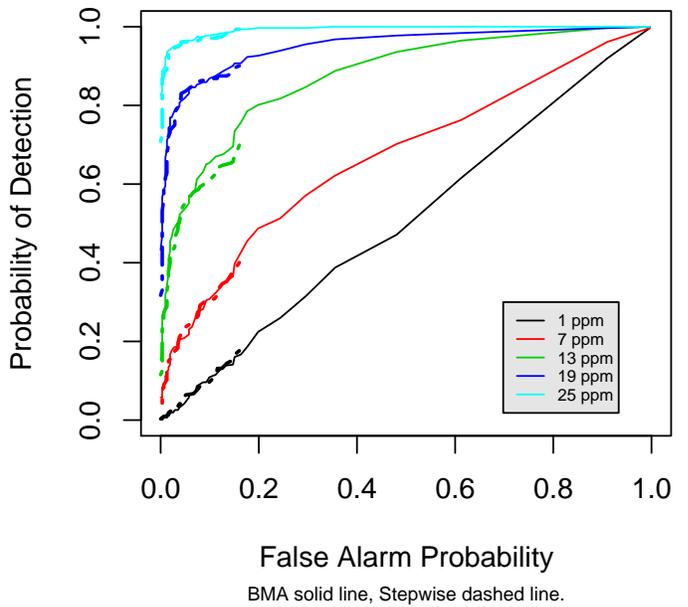
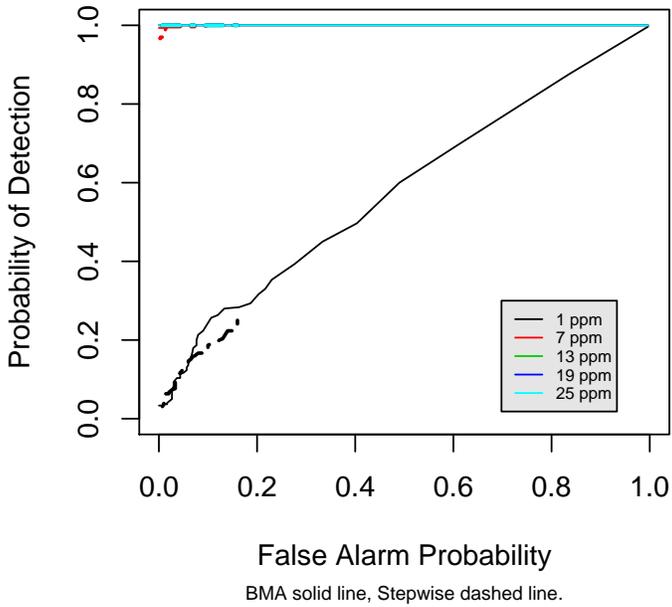
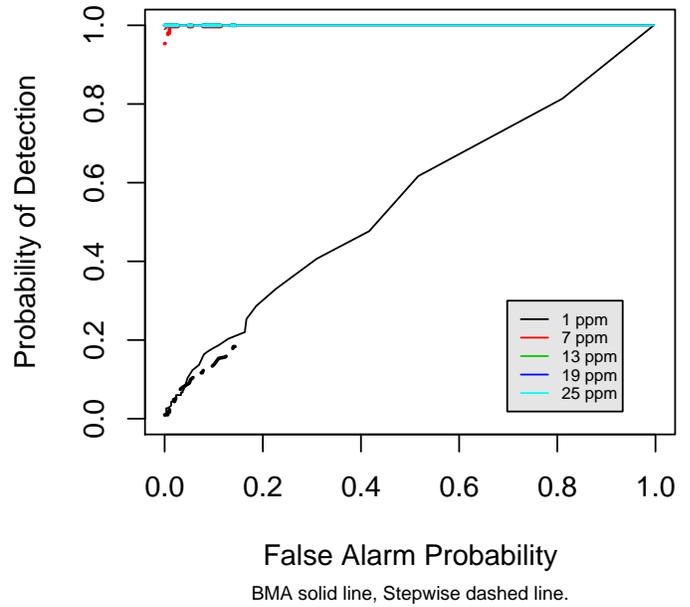


Figure B.2. Receiver operating characteristic curve for ethene.

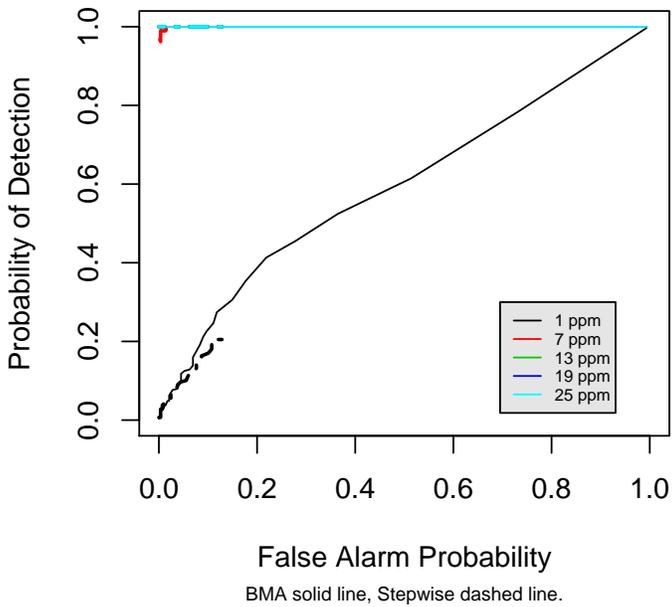
ROC for F113 and bkgrnd 1



ROC for F113 and bkgrnd 2



ROC for F113 and bkgrnd 3



ROC for F113 and bkgrnd 4

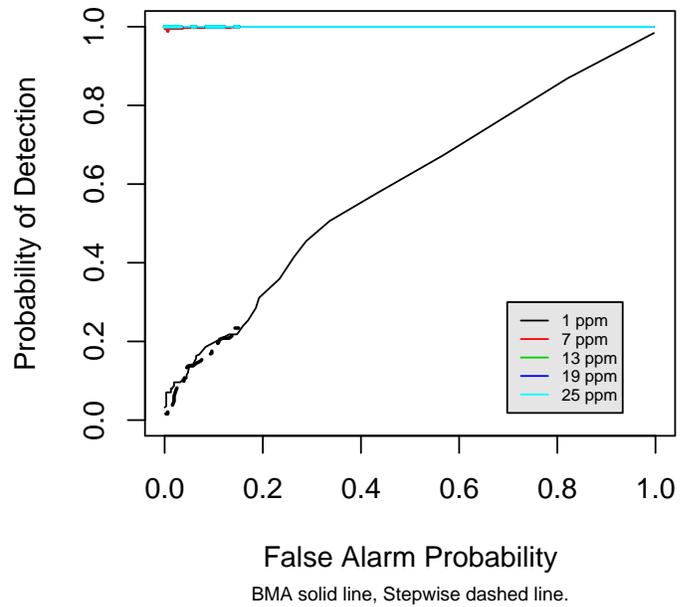
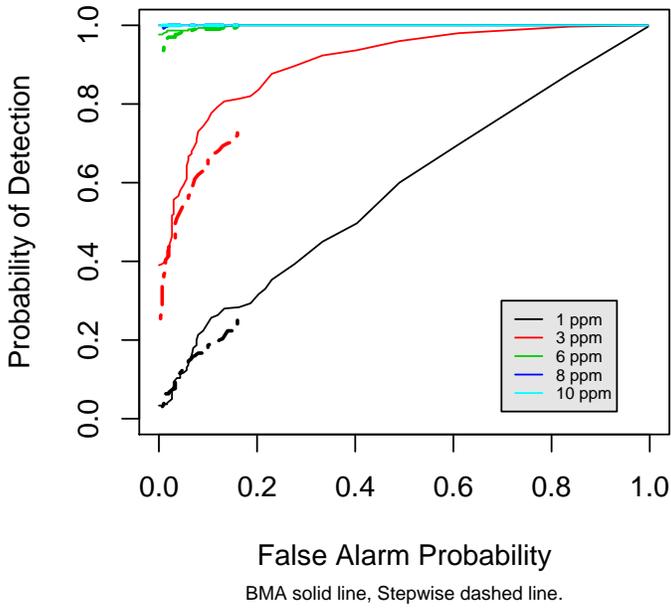
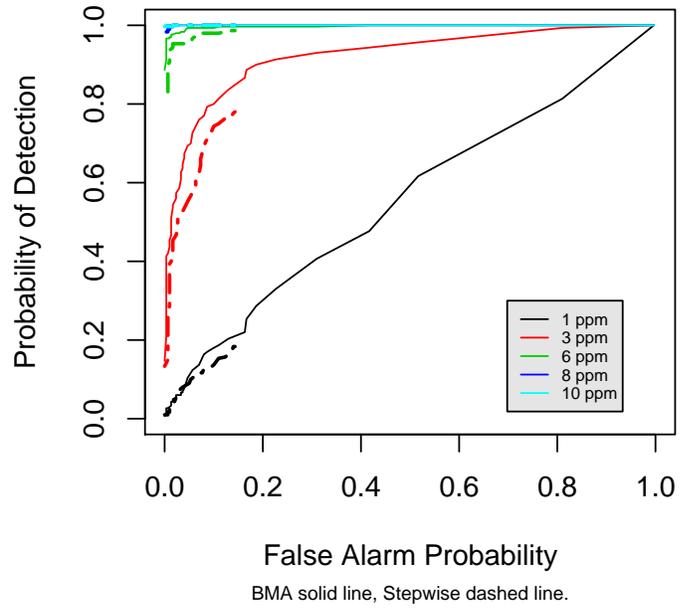


Figure B.3. Receiver operating characteristic curves for F-113 low concentration.

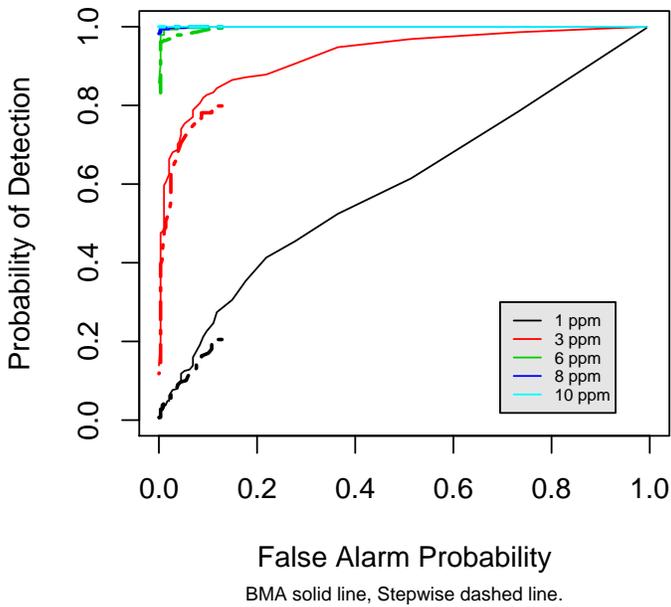
ROC for F113 and bkgrnd 1



ROC for F113 and bkgrnd 2



ROC for F113 and bkgrnd 3



ROC for F113 and bkgrnd 4

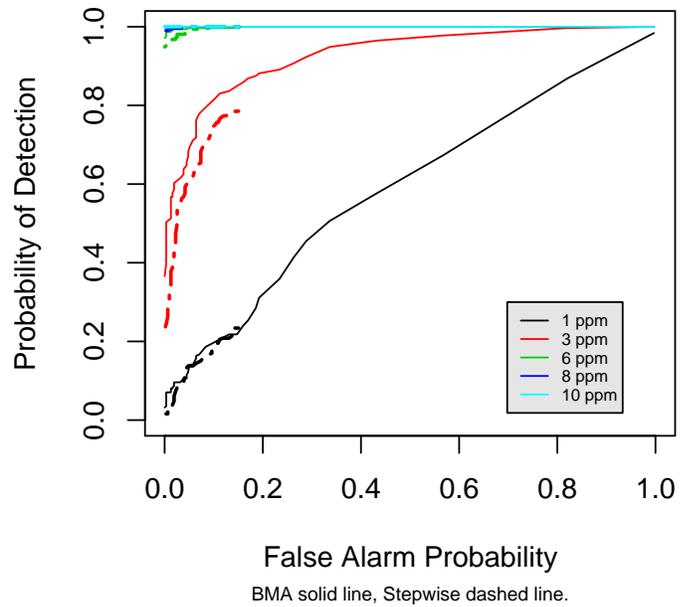
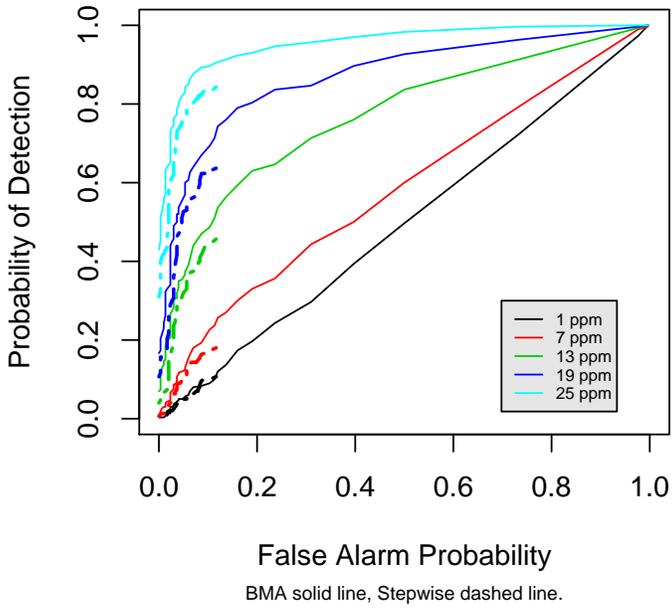
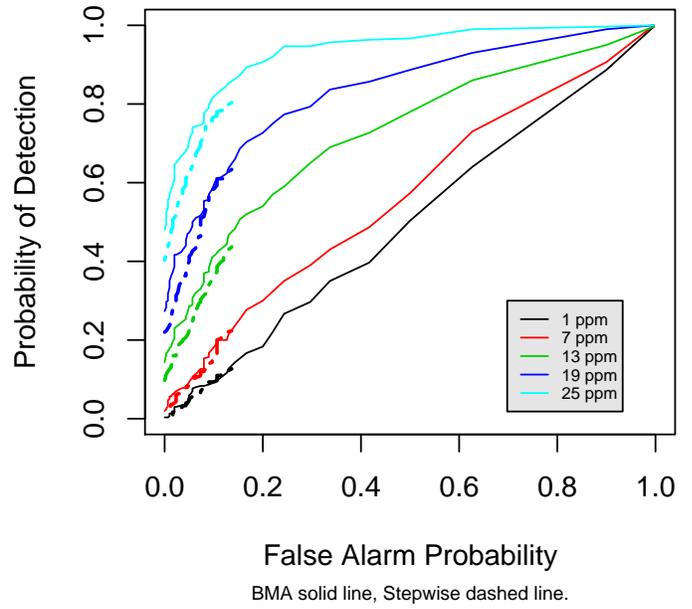


Figure B.4. Receiver operating characteristic curves for F-113 very low concentration.

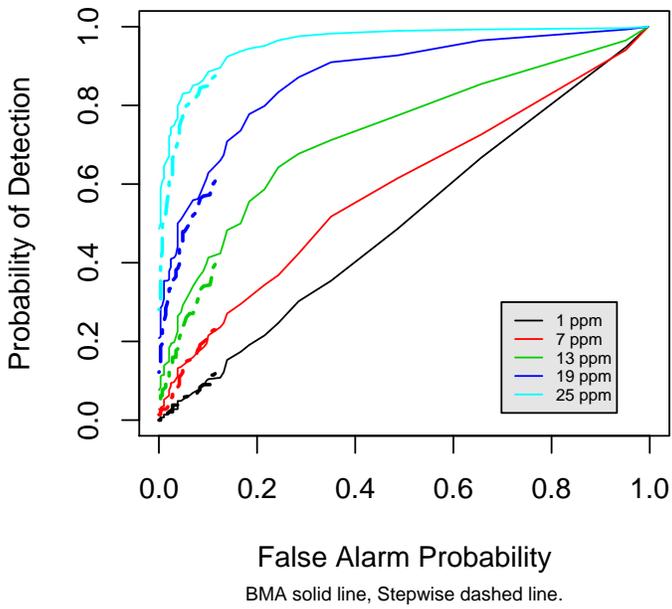
ROC for Furan and bkgrnd 1



ROC for Furan and bkgrnd 2



ROC for Furan and bkgrnd 3



ROC for Furan and bkgrnd 4

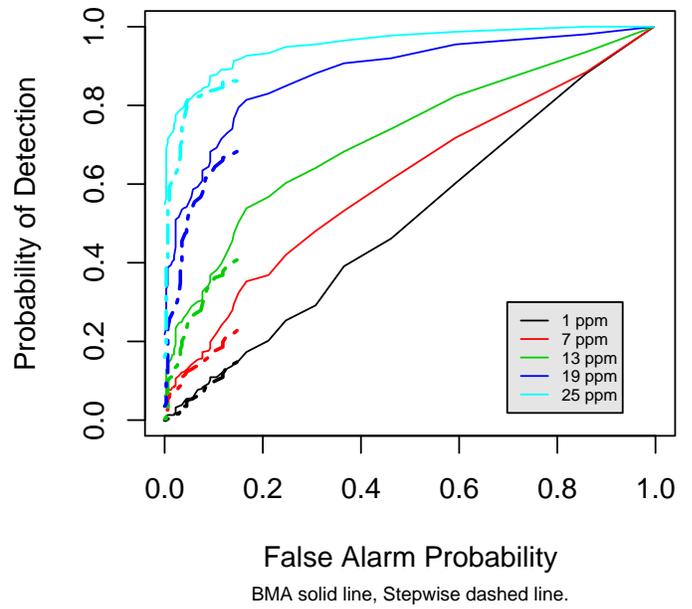


Figure B.5. Receiver operating characteristic curve for furan.

Study 2

Complete set of ROC curves from Study 2.

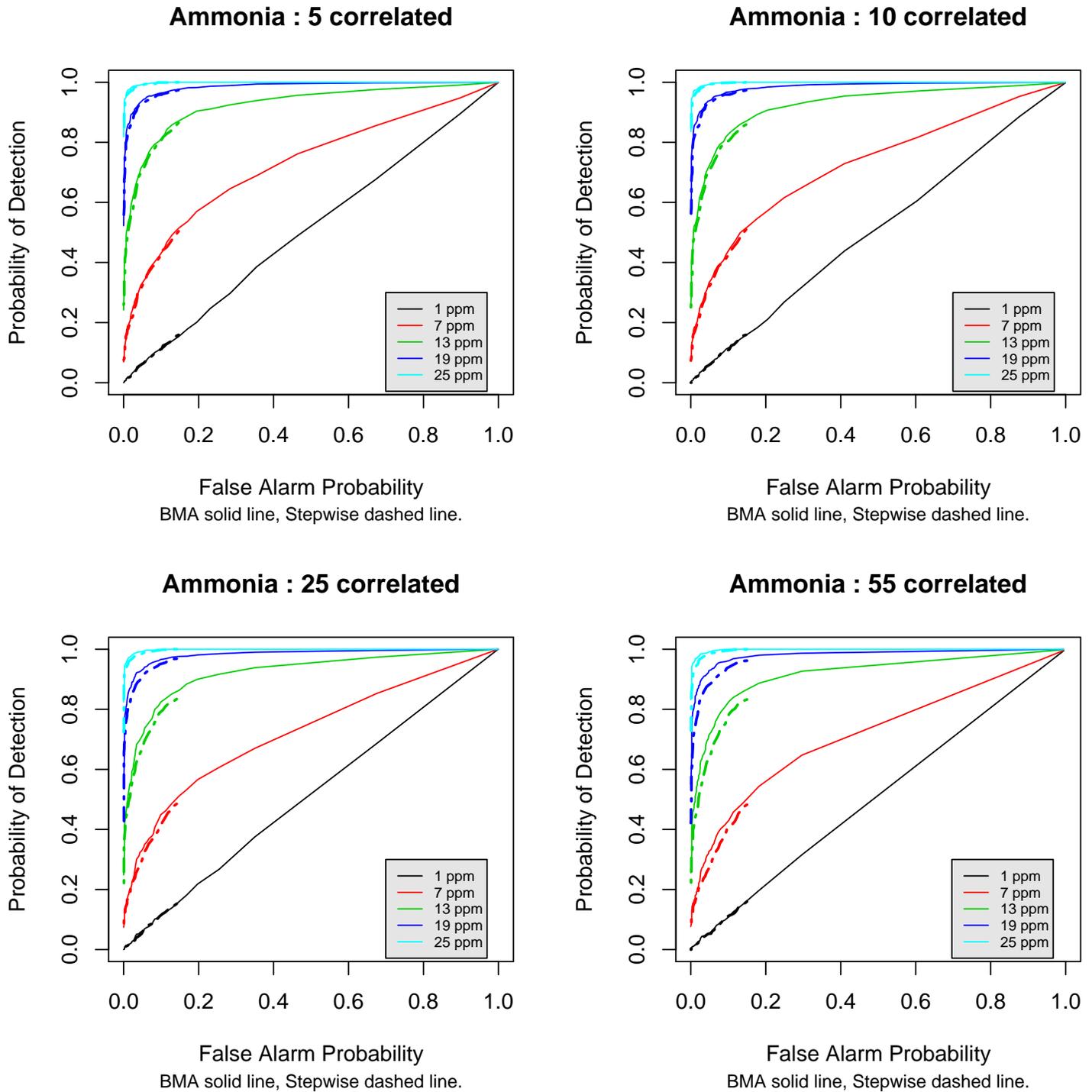
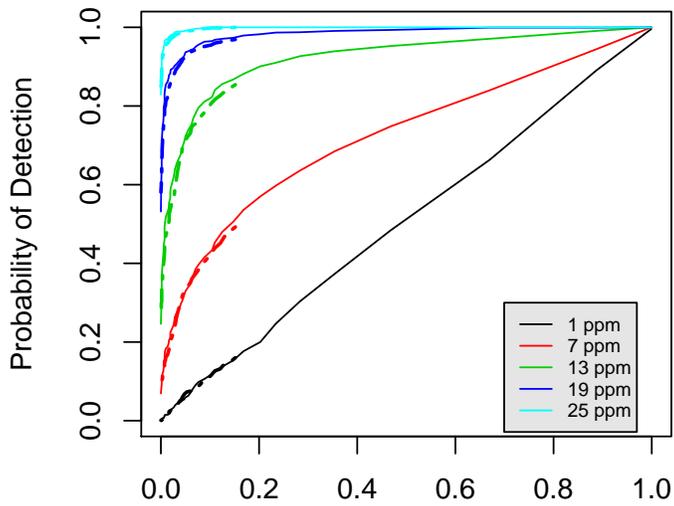


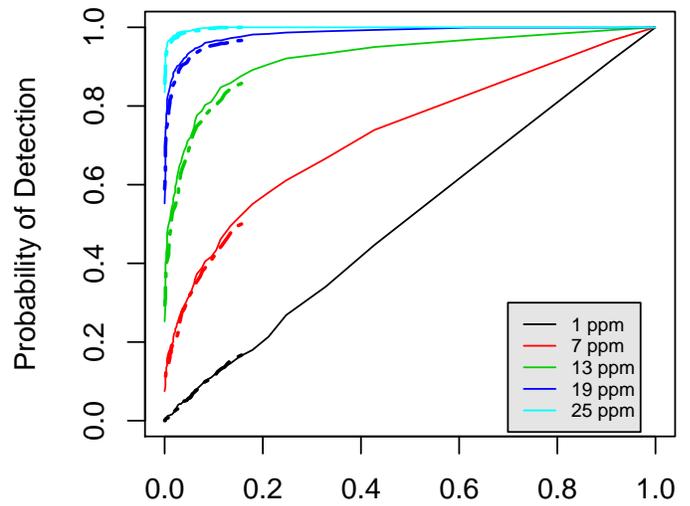
Figure B.6. Receiver operating characteristic curve for correlated ammonia.

Ammonia : 5 uncorrelated



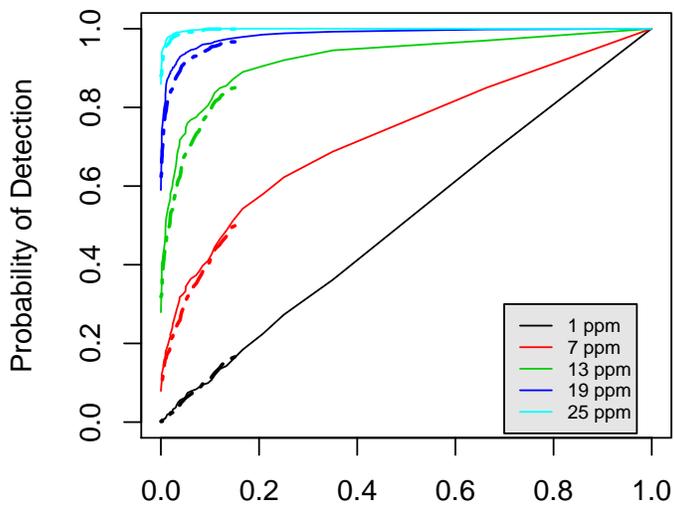
BMA solid line, Stepwise dashed line.

Ammonia : 10 uncorrelated



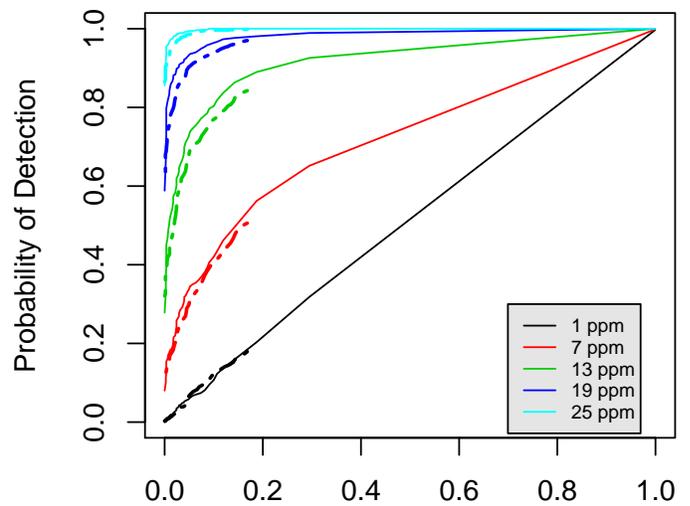
BMA solid line, Stepwise dashed line.

Ammonia : 25 uncorrelated



BMA solid line, Stepwise dashed line.

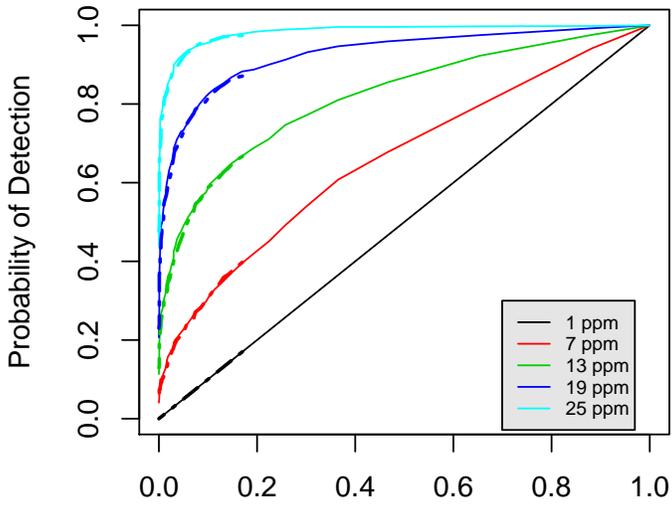
Ammonia : 55 uncorrelated



BMA solid line, Stepwise dashed line.

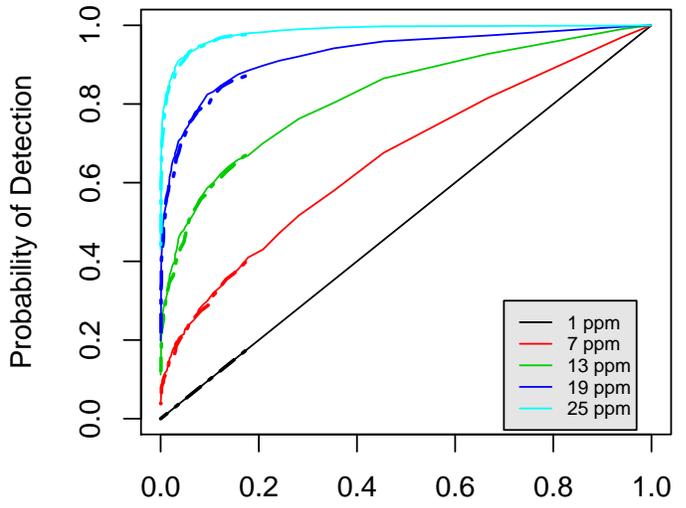
Figure B.7. Receiver operating characteristic curve for uncorrelated ammonia.

Ethene : 5 correlated



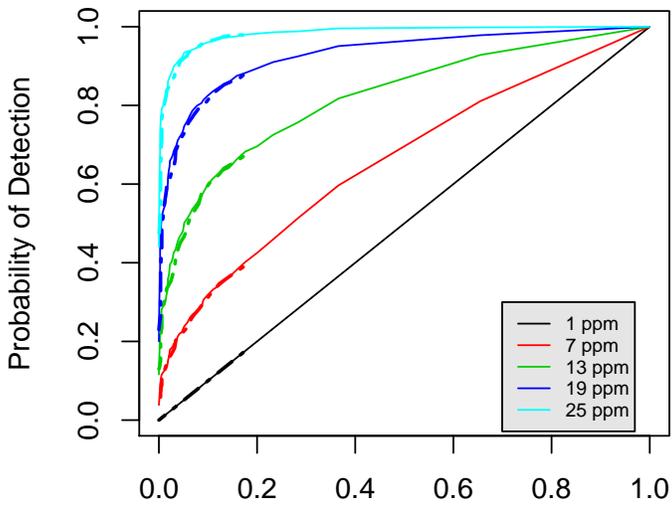
False Alarm Probability
BMA solid line, Stepwise dashed line.

Ethene : 10 correlated



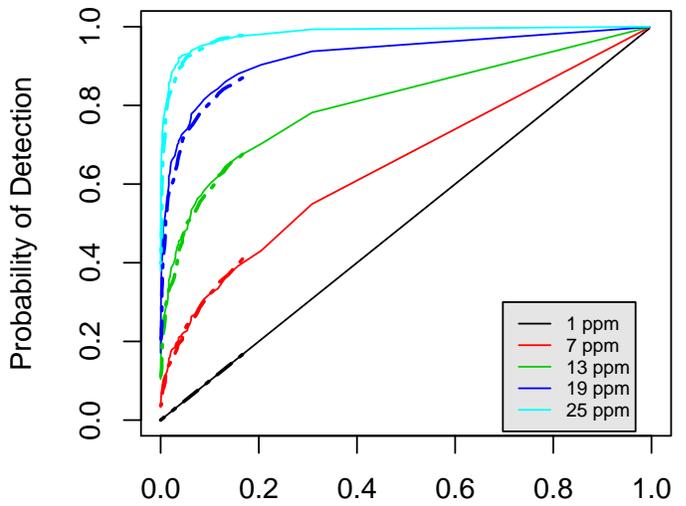
False Alarm Probability
BMA solid line, Stepwise dashed line.

Ethene : 25 correlated



False Alarm Probability
BMA solid line, Stepwise dashed line.

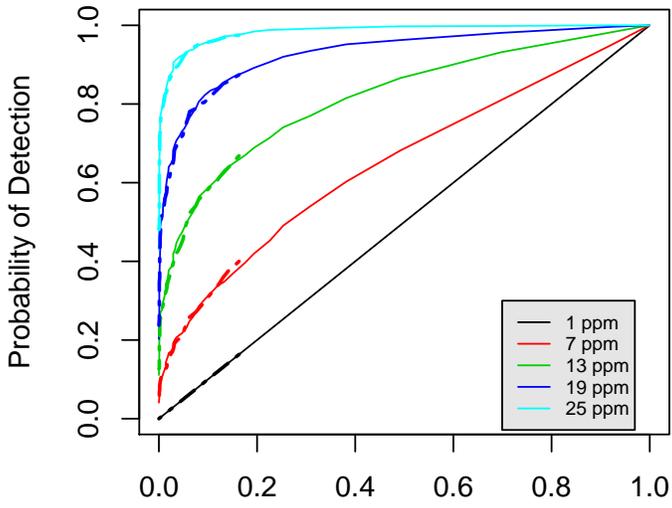
Ethene : 55 correlated



False Alarm Probability
BMA solid line, Stepwise dashed line.

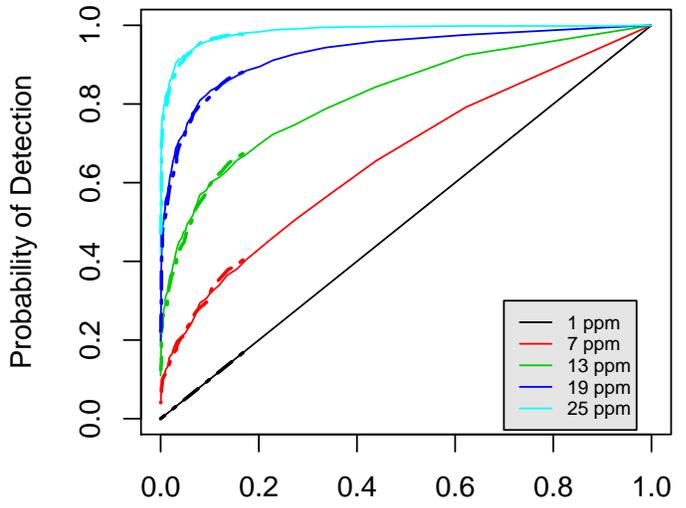
Figure B.8. Receiver operating characteristic curve for correlated ethene.

Ethene : 5 uncorrelated



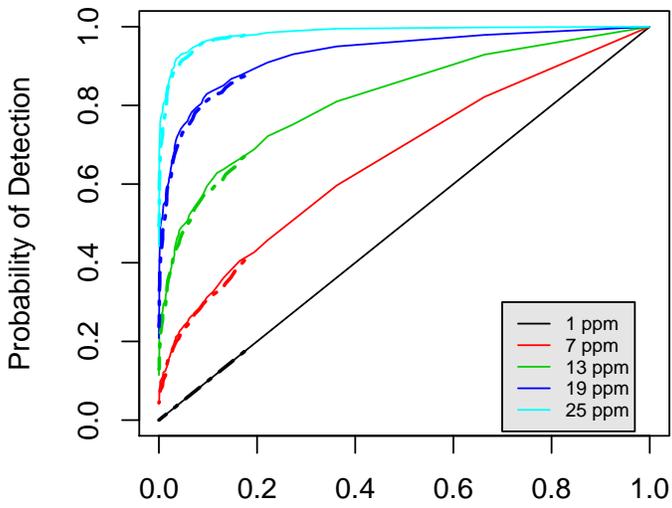
False Alarm Probability
BMA solid line, Stepwise dashed line.

Ethene : 10 uncorrelated



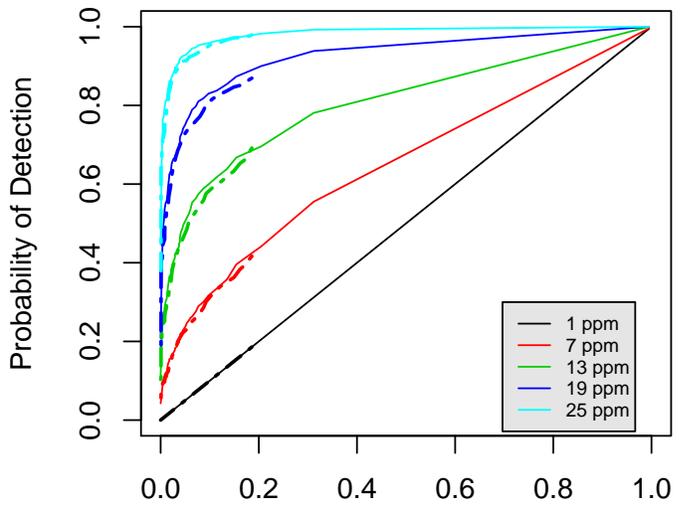
False Alarm Probability
BMA solid line, Stepwise dashed line.

Ethene : 25 uncorrelated



False Alarm Probability
BMA solid line, Stepwise dashed line.

Ethene : 55 uncorrelated



False Alarm Probability
BMA solid line, Stepwise dashed line.

Figure B.9. Receiver operating characteristic curve for uncorrelated ethene.

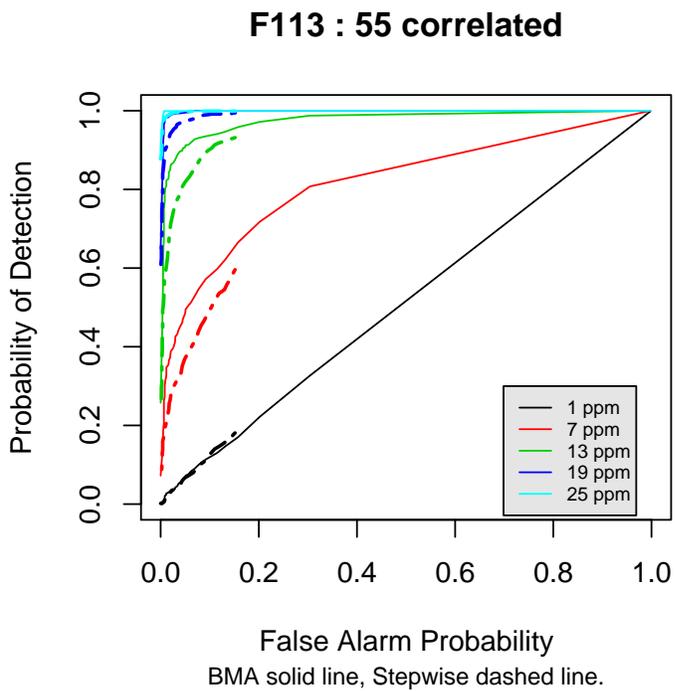
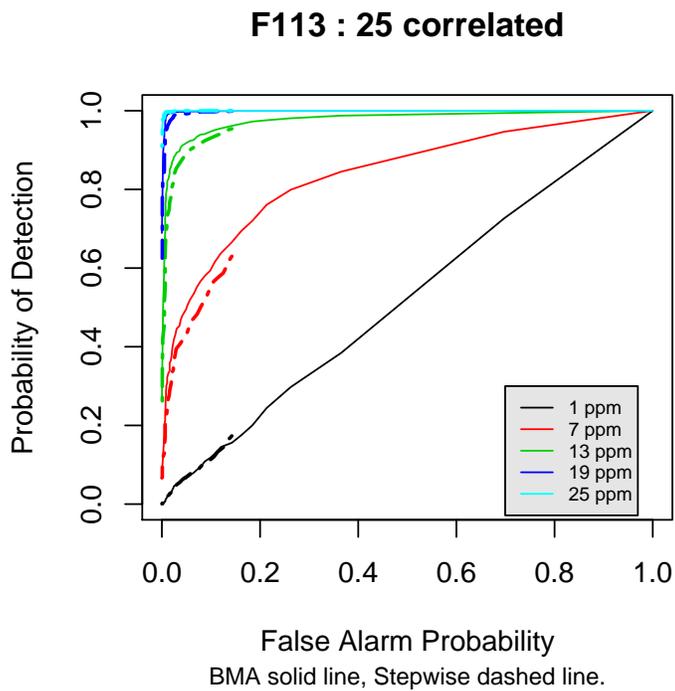
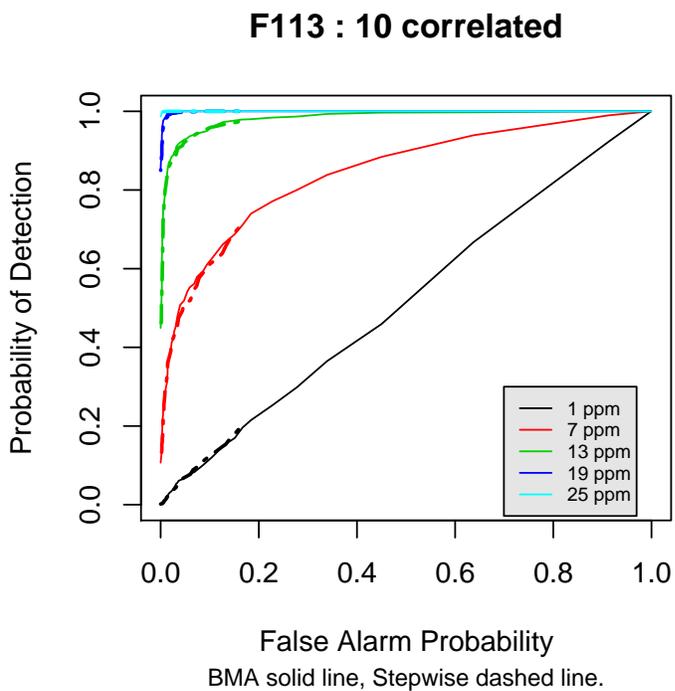
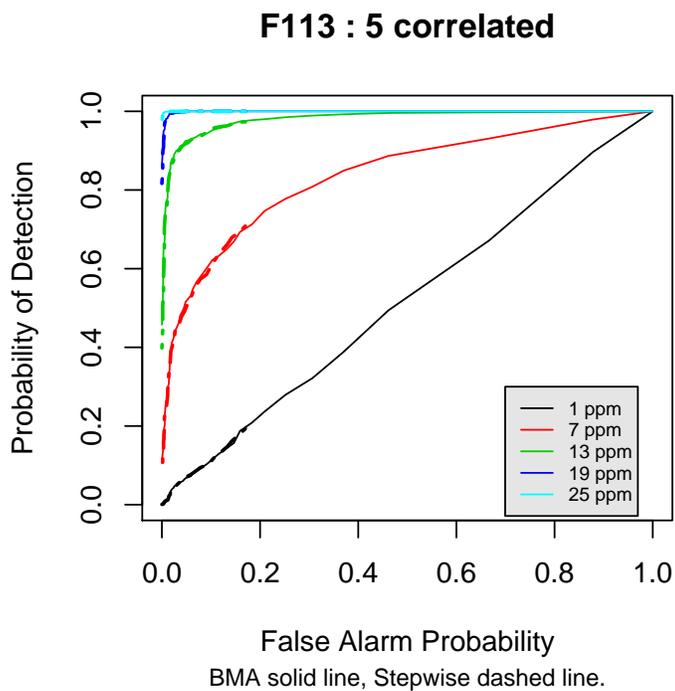


Figure B.10. Receiver operating characteristic curve for correlated F-113.

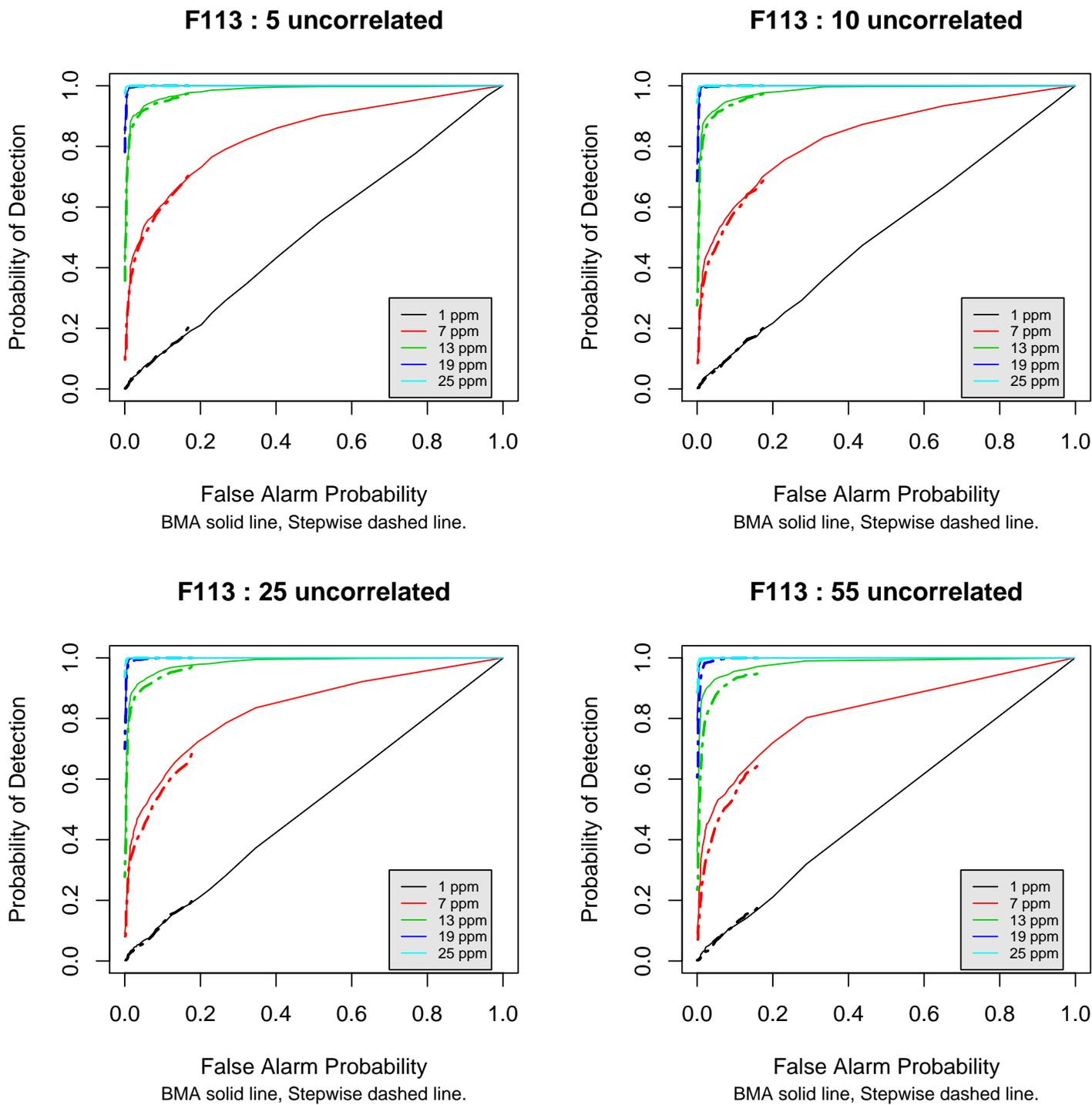
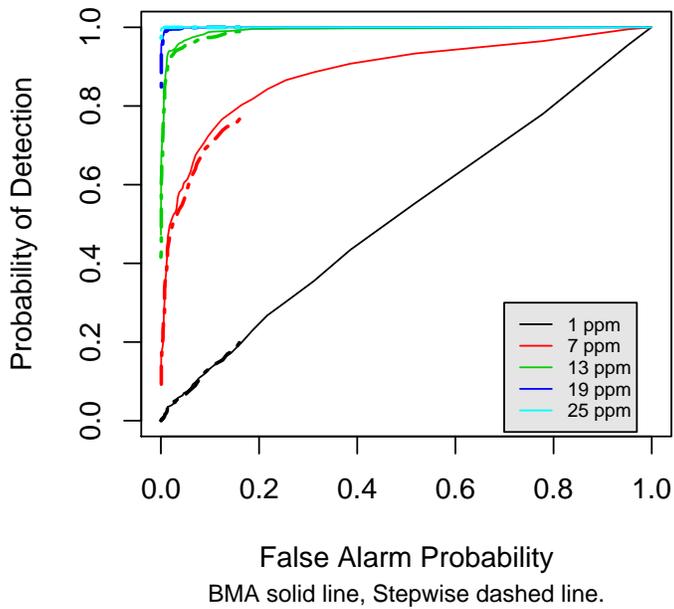
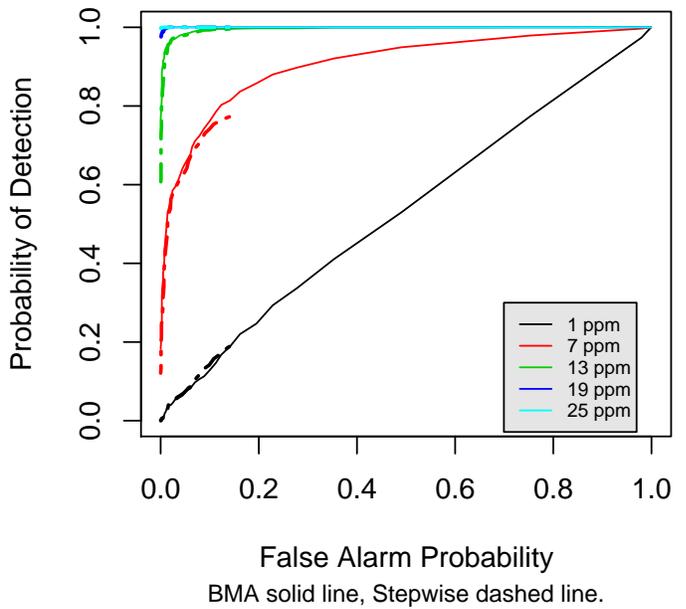


Figure B.11. Receiver operating characteristic curve for uncorrelated F-113.

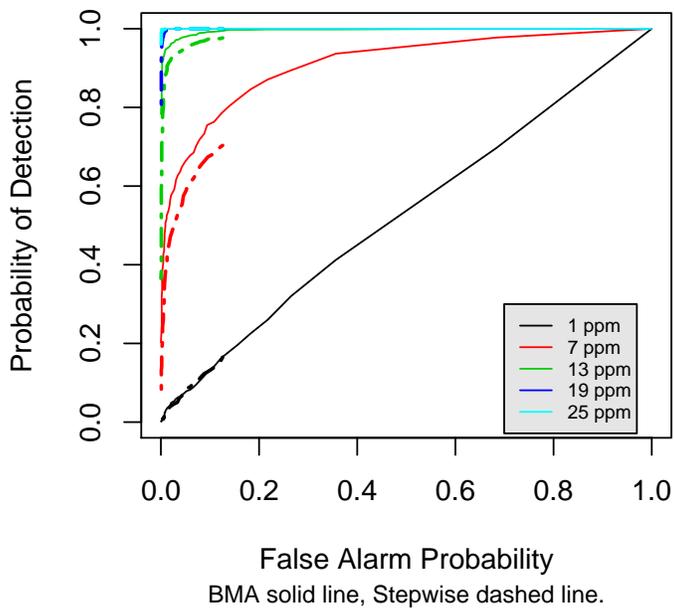
F114 : 5 correlated



F114 : 10 correlated



F114 : 25 correlated



F114 : 55 correlated

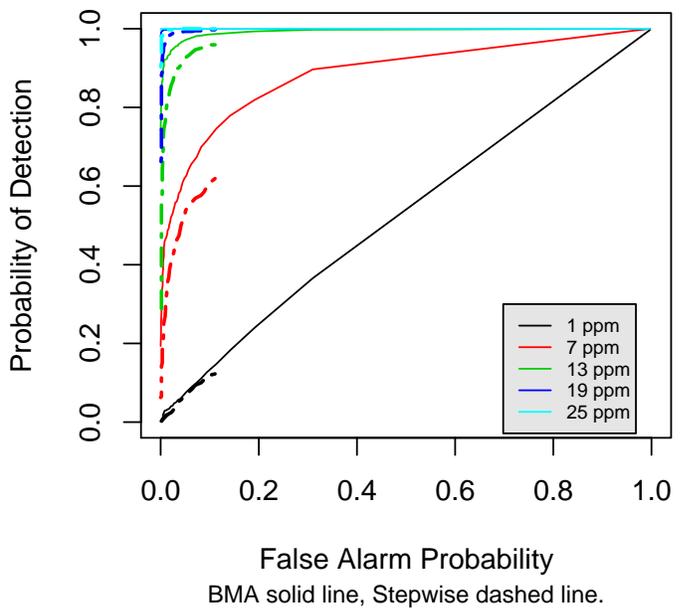
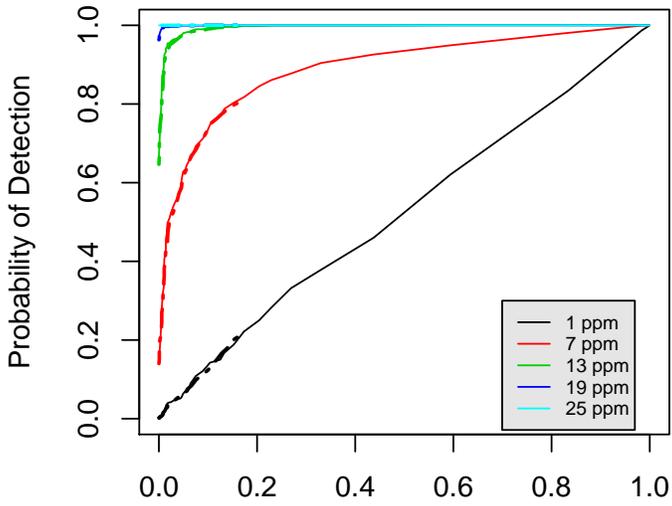
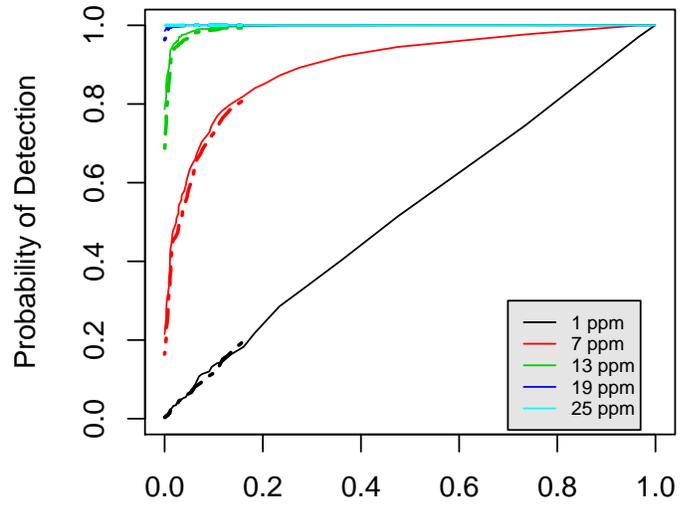


Figure B.12. Receiver operating characteristic curve for correlated F-114.

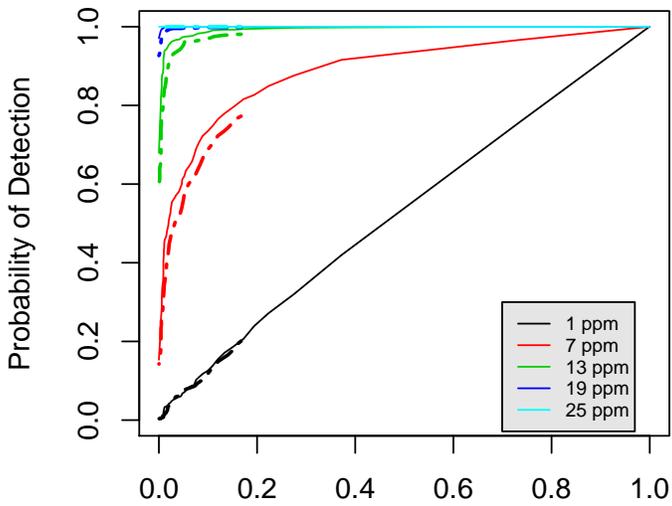
F114 : 5 uncorrelated



F114 : 10 uncorrelated



F114 : 25 uncorrelated



F114 : 55 uncorrelated

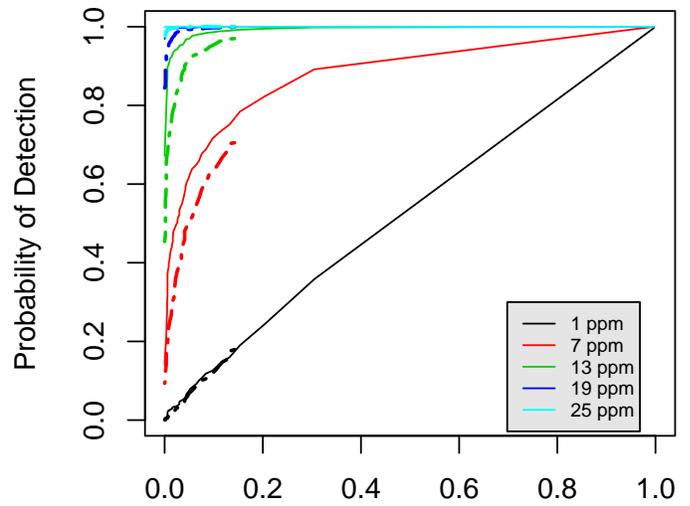
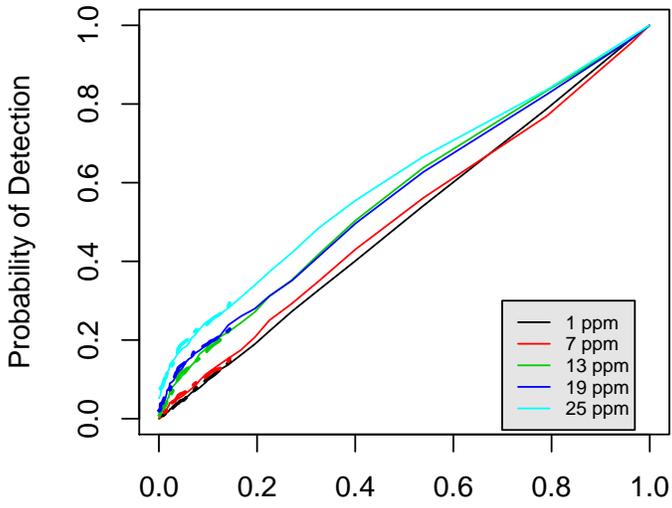


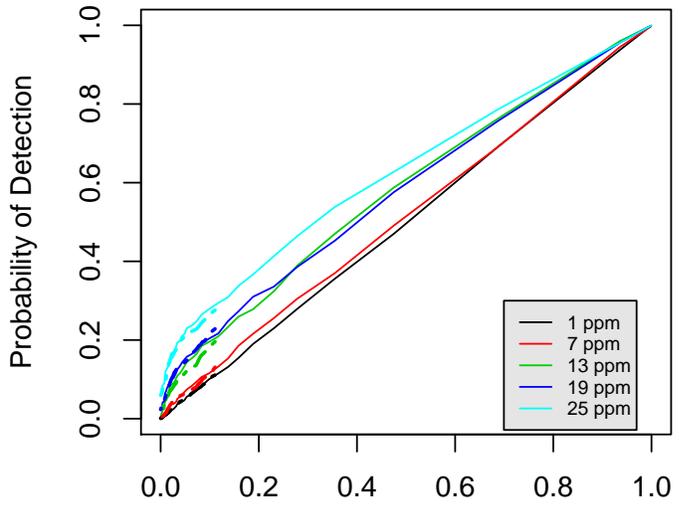
Figure B.13. Receiver operating characteristic curve for uncorrelated F-114.

Furan : 5 correlated



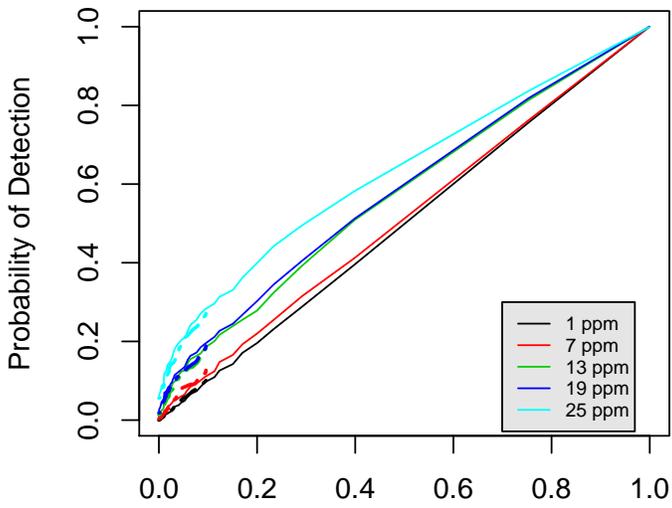
False Alarm Probability
BMA solid line, Stepwise dashed line.

Furan : 10 correlated



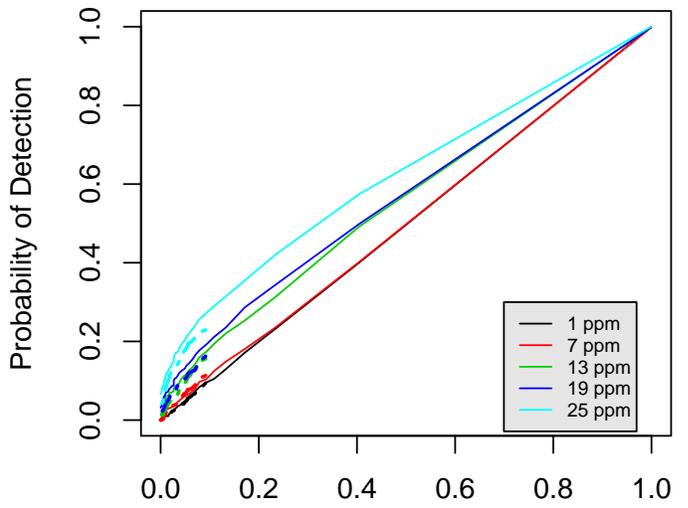
False Alarm Probability
BMA solid line, Stepwise dashed line.

Furan : 25 correlated



False Alarm Probability
BMA solid line, Stepwise dashed line.

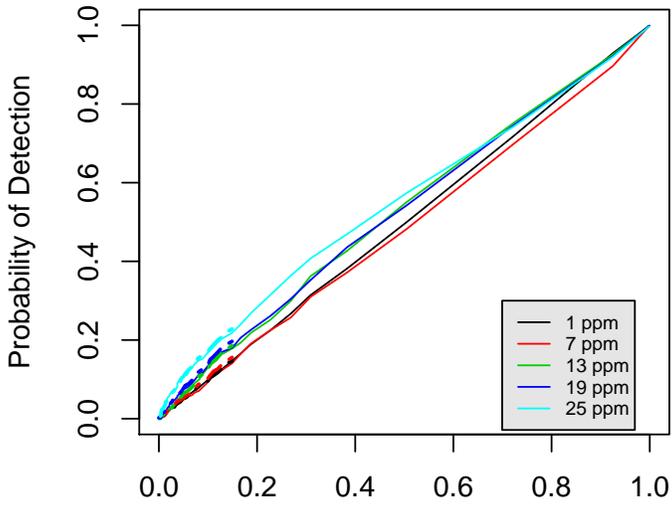
Furan : 55 correlated



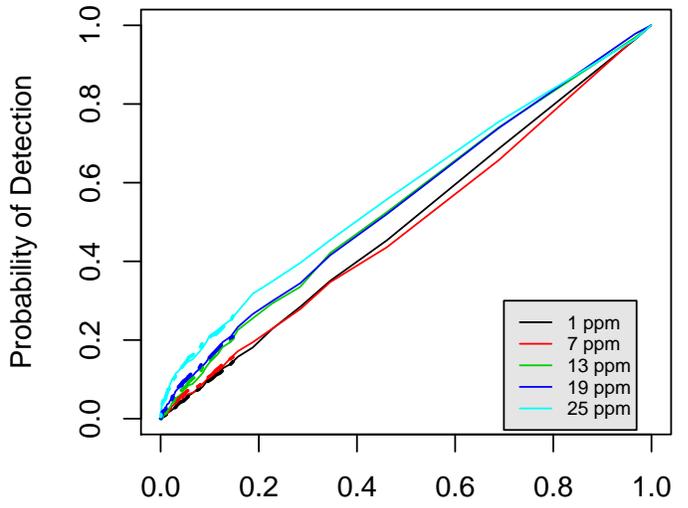
False Alarm Probability
BMA solid line, Stepwise dashed line.

Figure B.14. Receiver operating characteristic curve for correlated furan.

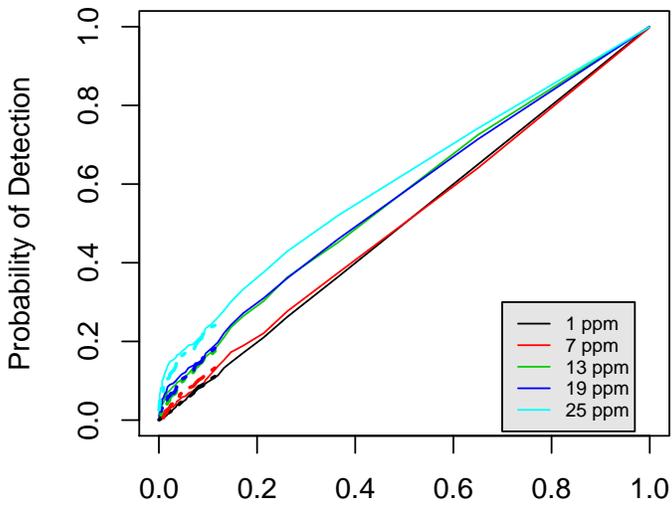
Furan : 5 uncorrelated



Furan : 10 uncorrelated



Furan : 25 uncorrelated



Furan : 55 uncorrelated

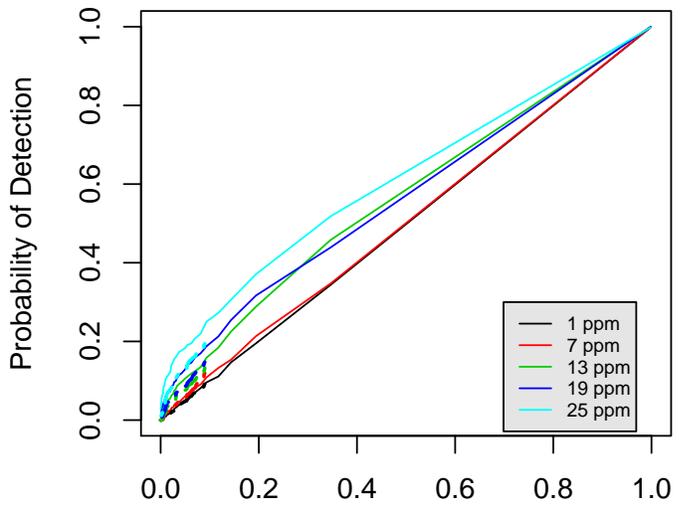


Figure B.15. Receiver operating characteristic curve for uncorrelated furan.

Appendix C

Simulation Code

Appendix C – Simulation Code

BMA functions.R

```
#This file contains functions necessary for running the examples in
  BMA.examples
#Functions include:
# BMAcalc - estimates posterior model probability for one model
# pmp.apply - estimates pmp for a list of models
# bic.parm.apply - estimates bic, estimates, and sd for a list of
  models
# BMAfitting - does the whole shebang

BMAcalc <- function(Xmat, y, g="default"){
#This work uses the prior distributions described in Fernandez
  et. al. 2001
#ie non-informative on sigma and non-informative on intercept
  (ie sample mean)
k = ncol(Xmat) #number of regressors
n = nrow(Xmat) #number of samples
if (g=="default") {g=1/max(n,k^2)}

part1 = .5*k*log(g/(g+1))
Mx = Xmat%%solve(t(Xmat)%*%Xmat)%*%t(Xmat)
yMxy = t(y)%*%y - t(y)%*%Mx%%*%y
yybar = y-mean(y)
part2 = (1/(g+1))*yMxy
part3 = (g/(g+1))*t(yybar)%*%yybar
part4 = -.5*(n-1)*log(part2 + part3)
part5 = part1+part4
#output is log(bayes factor numerator)
return(part5)}

pmp.apply = function(variables, resp, Xmat){
pred=Xmat[,variables]
pmp=BMAcalc(pred, resp)
return(pmp)}

bic.parm.apply = function(variables, resp, Xmat){
pred=Xmat[,variables]
model.fits=ls.print(lsfit(pred, resp, int=F), print.it=F)$
  coef.table
return(model.fits)}

BMAfitting <- function(resp, pred, nvmax){
#use leaps to calculate bic and all models
var.names= c("Int", colnames(pred))
```

```

#constants
nvar =ncol(pred)
fake.models<-regsubsets(resp~.,data=pred,intercept=T,method=
  "exhaustive",
nbest=100,nvmax=nvmax,really.big=T)
#grab the matrix identifying variables
which.var=summary(fake.models)$which
#calculate number of models that were fit
nmodels=nrow(summary(fake.models)$which)
#divide the models into a list
var.nums = t(t(which.var)*(1:(ncol(which.var))))
modellist=split(var.nums,1:nmodels)
varlist= lapply(modellist,unique)
#add the mean vector to the predictor matrix
fullpred = data.frame(rep(1,nrow(pred)),pred)
#calculate the pmp
pmp.all=lapply(varlist,pmp.apply,resp=resp,Xmat=
  as.matrix(fullpred))
logpmp=(unlist(pmp.all))
fullpostprob = exp(logpmp)/sum(exp(logpmp))

#calculate the bic approximation
bic = summary(fake.models)$bic-min(summary(fake.models)$
  bic)
bicpostprob = exp(-.5*bic)/sum(exp(-.5*bic))

#calculate E(bi) and V(bi) for the standard(no prior) model
#goes with the bic approximation.
est.all=lapply(varlist,bic.parm.apply,resp=resp,Xmat=
  as.matrix(fullpred))
Ebi <- matrix(NA,nrow=nmodels,ncol=(nvar+1))
SDbi <- matrix(NA,nrow=nmodels,ncol=(nvar+1))
for (i in 1:nmodels){
Ebi[i,varlist[[i]]] = est.all[[i]][[1]][,1]
SDbi[i,varlist[[i]]] = est.all[[i]][[1]][,2]
}

#calculate E(bi) and V(bi)
#goes with the Fernandez et. al. 2001
#needs to be calculated

#gather results
var.data=matrix(ncol=4,nrow=ncol(which.var))
dimnames(var.data)=list(var.names,c("BMApmp","bicpmp","E(b_i)",
  "SD(b_i)"))
#note, we may want to include the other two approximations
  mentioned by the LANL folk
#also need to calculate the estimates of Ebi_prior, V(bi_prior)
#cols are pmp, bic,E(bi),v(bi),E(bi_prior),v(bi_prior)
var.data[,1]=round(100*t(which.var)%*%fullpostprob,2)
var.data[,2]=round(100*t(which.var)%*%bicpostprob,2)

```

```

var.data[,3]=apply(Ebi*bicpostprob,2,sum,na.rm=T)
SDvar = (SDbi*SDbi+ Ebi^2)*bicpostprob
var.data[,4]=sqrt(apply(SDvar,2,sum,na.rm=T)-var.data[,3]^2)

#also return best bic model
best.bic=seq(1:nmodels)[bic==0]
best.var=unique(unlist(var.nums[best.bic,]))
best.var.names=var.names[best.var]
best.est=est.all[[best.bic]][[1]]
rownames(best.est)=best.var.names
return(list(bma.table=var.data,
best.est=best.est,best.var=best.var.names,best.bic=
  min(summary(fake.models)$bic)))}

StepFitting <- function(resp,pred,trace=0) {
#use stepAIC to find best stepwise model via AIC
  varNames = names(pred)
  LM1 = lm(resp~1)
  attach(pred)
  upperEq = as.formula(paste("upper=~",paste(varNames,collapse="+")))
  stepwise = step(LM1,scope=list(upper=upperEq,
                                lower=~1),trace=trace,direction="both")
  coefMat = summary(stepwise)$coefficients
  variables = rownames(coefMat)
  detach(2)
  return(list(variables=variables,coefMatrix=coefMat))
}

StepTvalue <- function(StepResult,VarName) {
  varcheck <- StepResult$variables==VarName
  tvalue <- ifelse(any(varcheck),
    StepResult$coefMatrix[varcheck,3],0)
  tvalue
}

sum.get <- function(x,a) sum(x>a)
sum.abs.get <- function(x,a) sum(abs(x)>a)

BMAPvalue <- function(BMAPResult,VarNumber) {
  BMAPResult$bma.table[VarNumber,1]
}

# bkg.mean returns the Spectral-mean from an HS-image

bkg.mean <- function(img.data) apply(img.data,3,mean)

# bkg.vihalf returns the sqrt-inverse of the variance of
# the spectrum from an HS-image, used to whiten

bkg.vihalf <- function(img.data) {

```

```

    dims <- dim(img.data)
    sigma2 <- var(matrix(as.vector(img.data),dims[1]*dims[2],dims[3]))
    ttt <- La.svd(sigma2)
    vihalf <- t(ttt$u)/sqrt(ttt$d)
    vihalf
}

# whiten.v  whitens a vector using input mean and vihalf

whiten.v <- function(data.vec,meanvec,vihalf) {
  as.vector(vihalf %*% (data.vec - meanvec))
}

# whiten.m  whitens an HS-image using input S-mean and S-vihalf

whiten.m <- function(img.data,meanspec,specvhalf) {
  aperm(apply(img.data,1:2,whiten.v,meanspec,specvhalf),c(2,3,1))
}

```

ExampleScript 2study.R

```

## Bring in the wavelengths and convert to correct units
dataNu <- as.vector(unlist(read.table("D:\\brs\\datasets2\\
  NoGas126Bands100Bkgrnds\\nu_126bands.txt",sep="\t")))
## want things in lambda (nm), so will convert
dataWavelengths <- 1e4/dataNu
numWavelengths <- length(dataWavelengths)

#chemSpectra <- matrix(0,nrow = numWavelengths, ncol = numChems)
#for (k in 1:numChems)
#{
# ## bring in a chemical, along with it's wavelength
# thisFile <- paste(chemDir,chemFiles[k,1],sep="\")
# thisChem <- read.table(thisFile,sep="\t")
# thisLambda <- 1e4/thisChem[,1]
# chemSpectra[,k] <- LibConv(thisLambda,thisChem[,2],dataWavelengths)
#}
#dimnames(chemSpectra) <- list(NULL,chemFilesSplit)
#write.table(chemSpectra,"D:\\brs\\datasets2\\chem92Lib126.csv",
  sep=",")
### Can also read them in.  It takes 45 mins to do this right now,
  so might want to do that ###
chemSpectra <- (read.table("D:\\brs\\datasets2\\chem92Lib126.csv",
  sep=",",header=T))
chemSpectra <- array(unlist(chemSpectra),dim = dim(chemSpectra),
  dimnames = NULL)
dimnames(chemSpectra) <- list(NULL,chemFilesSplit)
## Having trouble with the right shape, wrong sign
chemSpectra <- abs(chemSpectra)
#####

```

```
#####
## Start with study2startHere.Rwork and continue below
#####
#####

## want R to open the leaps package without using the mouse ('cause I
  like it that way)
library("leaps",character.only = TRUE)

## seperate out the chemicals for each run into a sub setChemSpectra
chemsRun <- read.table("D:\\IrSage\\IRSageSpec\\chem\\Andrea\\
  Library2\\top5corr.txt",sep="\t")
setchemCol <- charmatch(as.character(chemsRun[,1]),
  as.character(chemFiles[,1]))
setChemSpectra <- chemSpectra[,setchemCol]

## Gases to use for our example, then backgrounds, then concentrations.
gasesFileNames <- c("Ammonia","Ethene","F113","F114","Furan")
gases <- toupper(c("NH3","Ethene","F113","F114","Furan"))
gasesToMatch <- unlist(dimnames(setChemSpectra)[2])
##gasColumn <- c(76,47,49,50,54) ## which number gas the ones in the
  images are
gasColumn <- charmatch(gases,gasesToMatch)
gasConcLevels <- read.table("D:\\BRS\\Datasets2\\top5corr\\
  AmmoniaGas126Bands100Bkgrnds\\Ammonia_gas_conc_ultraLow.txt",sep=",")
numGases <- length(gases)
numGasConc <- length(gasConcLevels)

baseDir <- "D:\\brs\\datasets2\\top5corr\\"
numMaxPred <- 5
numROCpts <- 50

#####

## Off plume cube, used to make the mean and covaraince
dataCubeNO <- read.table(paste(baseDir,"NoGas126Bands100Bkgrnds\\
  No_gas_25_126bands.txt",sep=""),sep=",")
dataCubeNO <- array(unlist(dataCubeNO), dim = c(dim(dataCubeNO)[1],
  (dim(dataCubeNO)[2])/numWavelengths,numWavelengths),dimnames = NULL)

# Note, Nogas data has 5 times as many rows, so compute the mean and
## covariance with the most pixels, then subset.
meanOffPlume <- bkg.mean(dataCubeNO)
covHalfOffPlume <- bkg.vihalf(dataCubeNO)
# Now, "whiten" or "standardize" the chemical Library
## could either call LibStd, or LibWhite.
chemSpectraWhite <- as.data.frame(LibStd(setChemSpectra,
  covHalfOffPlume))

## Now, for each gas cycle through the concentrations
```

```

## Now, get rid of some rows of Nogas
dataCubeNO <- dataCubeNO[1:12,,]
dimsData <- dim(dataCubeNO)
numPixels <- (dimsData[1])*dimsData[2]##Nogas has all 5
concentrations
## We want to have zero concentration for the bmaRes and stepRes to
work like before.
numPixels <- dim(dataCubeNO)[1]*dim(dataCubeNO)[2]##Nogas has all 5
concentrations

for (k in 1:numGases)
{
  thisGasColumn <- 1 + gasColumn[k] ## number of column in
predictor matrix of the gas in image
  ## create the bmaRes matrix for plotting the ROC curves, load it
with the zero concentration
  bmaRes <- matrix(NA,nrow = numPixels,ncol = numGasConc+1,
dimnames = list(NULL,c(0,gasConcLevels)))
  dataCubeWhite <- whiten.m(dataCubeNO,meanOffPlume,covHalfOffPlume)
  dataCubeWhiteFlatten <- matrix(dataCubeWhite,nrow = numPixels ,
ncol = numWavelengths)
  dataList <- split(dataCubeWhiteFlatten,1:numPixels)
  dataListBMA <- lapply(dataList,BMAfitting,
as.data.frame(chemSpectraWhite),numMaxPred)
  bmaRes[,1] <- unlist(lapply(dataListBMA,BMAPvalue,thisGasColumn))
  ## needed to sequence for probability curves
  pPc <- seq(min(bmaRes[,1]),max(bmaRes[,1]),length = numROCpts)
  ## Now for stepwise regression
  stepRes <- matrix(NA,nrow = numPixels,ncol = numGasConc + 1,
dimnames = list(NULL,c(0,gasConcLevels)))
  dataListStep = lapply(dataList,StepFitting,(chemSpectraWhite))
  stepRes[,1] = unlist(lapply(dataListStep,StepTvalue,gases[k]))
  tVc <- seq(min(abs(stepRes[stepRes[,1]!=0,1])),
max(abs(stepRes[stepRes[,1]!=0,1])),length = numROCpts)
  #if (any(is.nan(tVc)))
  #{
  # tVc <- seq(1,100,length=numROCpts)
  #}

  for (l in 1:numGasConc)
  {
    ## get the correct directory and file to read in
    thisDir <- paste(baseDir,gasesFileNames[k],
"Gas126Bands100Bkgrnds\\",sep="")
    thisFile <- paste(thisDir,gasesFileNames[k],"_126Bands_",
gasConcLevels[l],"conc_100bkgrnd_ultraLow.csv",sep="")
    ## read in the file, then get into correct array format
    dataCube <- read.table(thisFile,sep=",")
    dataCube <- array(unlist(dataCube), dim = c(dim(dataCube)[1],
(dim(dataCube)[2])/numWavelengths,numWavelengths),dimnames = NULL)
    ## whiten the cube, then flatten for using lapply for BMA and

```

```

stepwise.
  dataCubeWhite <- whiten.m(dataCube,meanOffPlume,covHalfOffPlume)
  dataCubeWhiteFlatten <- matrix(dataCubeWhite,nrow = numPixels ,
ncol = numWavelengths)
  dataList = split(dataCubeWhiteFlatten,1:numPixels)
  dataListBMA=lapply(dataList,BMAfitting,
as.data.frame(chemSpectraWhite),5)
  ## get the BMA results into a matrix for plotting
  bmaRes[,l+1] <- unlist(lapply(dataListBMA, BMAPvalue,
thisGasColumn))
  ## Now, for the stepwise
  dataListStep = lapply(dataList,StepFitting,
as.data.frame(chemSpectraWhite))
  stepRes[,l+1] = unlist(lapply(dataListStep,StepTvalue,gases[k]))

  } ## end l loop
  ## write to disk for later
  write.table(bmaRes,paste(thisDir,gasesFileNames[k],
"_126Bands_bma_100bkgrnd_ultraLowConc.csv",sep=""),row.names = FALSE)
  write.table(stepRes,paste(thisDir,gasesFileNames[k],
"_126Bands_step_100bkgrnd_ultraLowConc.csv",sep=""),row.names =
FALSE)
  ## Now, get the bmaROCmat together and the stepROCmat together
  bmaROCmat <- matrix(NA,nrow = numROCpts,ncol = numGasConc+1,
dimnames = list(NULL,c(0,gasConcLevels)))
  stepROCmat <- matrix(NA,nrow = numROCpts,ncol = numGasConc+1,
dimnames = list(NULL,c(0,gasConcLevels)))
  for (i in 1:numROCpts)
  {
    bmaROCmat[i,] <- apply(bmaRes,2,sum.get,pPc[i])/numPixels
    stepROCmat[i,] <- apply(stepRes,2,sum.abs.get,tVc[i])/numPixels
  }

  ## Now, plot both up and save them
  matplot(bmaROCmat[,1],bmaROCmat[,-1],type='l',lty=1,xlab=
"False Alarm Probability",ylab = "Probability of Detection", main =
paste("BMA ROC for ", gasesFileNames[k],sep=""),col=1:5,xlim=
c(0,round(max(bmaROCmat[,1]),digits = 3)), ylim=c(0,1))
  legend(.7*round(max(bmaROCmat[,1]),digits=3),0.3,legend =
paste(unlist(c(dimnames(bmaROCmat[,-1])[2])), " ppm",sep=""),lty=1,
col=1:5,cex = 0.6,bg="grey90")
  savePlot(filename = paste(thisDir,"bmaROCplot",gasesFileNames[k],
"_100bkgrnd_ultraLowConc",sep=""),type="pdf")

  matplot(stepROCmat[,1],stepROCmat[,-1],type = 'l',lty=1,xlab=
"False Alarm Probability",ylab="Probability of Detection",main=
paste("STEP ROC for ", gasesFileNames[k],sep=""),col=1:5,xlim=
c(0,round(max(stepROCmat[,1]),digits = 3)),ylim=c(0,1))
  legend(0.7*round(max(stepROCmat[,1]),digits=3),0.7,legend =
paste(unlist(c(dimnames(bmaROCmat[,-1])[2])), " ppm",sep=""),lty=1,
col=1:5,cex = 0.6,bg="grey90")

```

```

        savePlot(filename = paste(thisDir,"stepROCplot",
gasesFileNames[k],"_100bkgrnd_ultraLowConc",sep=""),type="pdf")

## Now, plot both up on the same axes and save them
#postscript(file = paste(thisDir,"ROCplot",gasesFileNames[k],
"_ultraLowConc.ps",sep=""),paper="letter")

        matplot(bmaROCmat[,1],bmaROCmat[,-1],type='l',lty=0,lwd=1,xlab=
"False Alarm Probability",ylab = "Probability of Detection", main =
paste("ROC for ", gasesFileNames[k],sep=""),sub = "BMA solid line,
Stepwise dashed line.",cex.sub = 0.7,col=1:5,xlim=c(0,1), ylim=
c(0,1))

matlines(bmaROCmat[,1],bmaROCmat[,-1],type='l',lty=1,lwd=1,
xlim=c(0,1),ylim=c(0,1))
        matlines(stepROCmat[,1],stepROCmat[,-1],type='l',lty=6,lwd=2,
col=1:5,xlim=c(0,1),ylim=c(0,1))
legend(.7*round(max(bmaROCmat[,1]),digits=3),0.3,legend =
paste(unlist(c(dimnames(bmaROCmat[,-1])[2])), " ppm",sep=""),lty=1,
col=1:5,cex = 0.6,bg="grey90")
        #dev.off()

        savePlot(filename = paste(thisDir,"ROCplot",gasesFileNames[k],
"_100bkgrnd_ultraLowConc",sep=""),type="pdf")

} ## end k loop

#####
#####
### Create updated figures for the paper
## Load in the start here stuff.
library("leaps",character.only = TRUE)

libSize <- 25
libType <- "uncorrelated"
thisType <- paste("top",libSize,substr(libType,1,1+regexpr("r",
libType)),sep="") # looks for first r in libType

## separate out the chemicals for each run into a sub setChemSpectra
chemsRun <- read.table(paste("D:\\IrSage\\IRSageSpec\\chem\\Andrea\\
Library2\\",thisType,".txt",sep=""),sep="\t")
setchemCol <- charmatch(as.character(chemsRun[,1]),
as.character(chemFiles[,1]))
setChemSpectra <- chemSpectra[,setchemCol]

gasesFileNames <- c("Ammonia","Ethene","F113","F114","Furan")
gases <- toupper(c("NH3","Ethene","F113","F114","Furan"))
gasesToMatch <-unlist(dimnames(setChemSpectra)[2])
##gasColumn <- c(76,47,49,50,54) ## which number gas the ones in the

```

```

images are
gasColumn <- charmatch(gases,gasesToMatch)
gasConcLevels <- read.table("D:\\BRS\\Datasets2\\top5corr\\
  AmmoniaGas126Bands100Bkgrnds\\Ammonia_gas_conc_ultraLow.txt",sep=",")
numGases <- length(gases)
numGasConc <- length(gasConcLevels)
numMaxPred <- 5
numROCpts <- 50

for (k in 1:numGases){
  thisDir <- paste("D:\\brs\\datasets2\\",thisType,"\\",
    gasesFileNames[k],"Gas126Bands100Bkgrnds\\",sep="")

  bmaRes <- read.table(paste(thisDir,gasesFileNames[k],
    "_126Bands_bma_100bkgrnd_ultraLowConc.csv",sep=""),header = TRUE)
  stepRes <- read.table(paste(thisDir,gasesFileNames[k],
    "_126Bands_step_100bkgrnd_ultraLowConc.csv",sep=""),sep=" ",header =
    TRUE)
    ## needed to sequence for probability curves
    pPc <- seq(min(bmaRes[,1]),max(bmaRes[,1]),length = numROCpts)
    tVc <- seq(min(abs(stepRes[stepRes[,1]!=0,1])),
    max(abs(stepRes[stepRes[,1]!=0,1])),length = numROCpts)
  ## Now, get the bmaROCmat together and the stepROCmat together
    bmaROCmat <- matrix(NA,nrow = numROCpts,ncol = numGasConc+1,
    dimnames = list(NULL,c(0,gasConcLevels)))
    stepROCmat <- matrix(NA,nrow = numROCpts,ncol = numGasConc+1,
    dimnames = list(NULL,c(0,gasConcLevels)))
    for (i in 1:numROCpts)
      {
        bmaROCmat[i,] <- apply(bmaRes,2,sum.get,pPc[i])/
        dim(bmaRes)[[1]]#numPixels
        stepROCmat[i,] <- apply(stepRes,2,sum.abs.get,tVc[i])/
        dim(stepRes)[[1]]#numPixels
      }

  ## Plot up, with appropriate labels this time...

matplot(bmaROCmat[,1],bmaROCmat[,-1],type='l',lty=0,lwd=1,xlab=
  "False Alarm Probability",ylab = "Probability of Detection", main =
  paste(gasesFileNames[k],":",libSize,libType,sep=" "),sub =
  "BMA solid line, Stepwise dashed line.",cex.sub = 0.8,col=1:5,xlim=
  c(0,1), ylim=c(0,1))

matlines(bmaROCmat[,1],bmaROCmat[,-1],type='l',lty=1,lwd=1,
  xlim=c(0,1),ylim=c(0,1))
  matlines(stepROCmat[,1],stepROCmat[,-1],type='l',lty=6,lwd=2,
  col=1:5,xlim=c(0,1),ylim=c(0,1))
legend(.7*round(max(bmaROCmat[,1]),digits=3),0.3,legend =
  paste(unlist(c(dimnames(bmaROCmat[,-1])[2])), " ppm",sep=""),lty=1,

```

```
col=1:5,cex = 0.7,bg="grey90")

    savePlot(filename = paste(thisDir,"ROCplot",gasesFileNames[k],
"_100bkgrnd_",substr(libType,1,1+regexpr("r",libType)),libSize,
sep=""),type="pdf")
    ## write to two places
savePlot(filename = paste("D:\\brs\\clientreports\\sandy\\ROCplot",
gasesFileNames[k],"_100bkgrnd_",substr(libType,1,1+regexpr("r",
libType)),libSize,sep=""),type="pdf")
}
```

Distribution List

No. of Copies

OFFSITE

- 1 Rhys Williams * * *
Physical Scientist
NA-22/Forrestal Building
U.S. Department of Energy
1000 Independence Ave., S.W.
Washington, DC 20585

ONSITE

- 4 Pacific Northwest National Laboratory

Gordon Dudder *
Steve Sharpe *
Dale Anderson *
Brent Pulsipher *

Distribution method:

- * Email notification of availability at
ERICA
- ** CD
- * * * Hardcopy