Detection and Characterization of Chemical Vapor Fugitive Emissions from Hyperspectral Infrared Imagery by Nonlinear Optimal Estimation

Christopher M. Gittins

gittins@psicorp.com

978-689-0003

SPIE Defense, Security, and Sensing
Orlando, FL
April 5-9, 2010

All slides previously approved for public release or derived from unlimited-distribution material:
DTRA PA Control 08-556
**Detection and Characterization of Chemical Vapor Fugitive Emissions from Hyperspectral Infrared Imagery by Nonlinear Optimal Estimation**

**Physical Sciences Inc, 20 New England Business Center, Andover, MA, 01810-1077**

**Approved for public release; distribution unlimited**

**U.S. Government or Federal Rights License**

**Security Classification of:**
- a. Report: unclassified
- b. Abstract: unclassified
- c. This Page: unclassified

**Limitation of Abstraction:**
- Same as Report (SAR)
- Number of Pages: 35

**Form Approved**
OMB No. 0704-0188
Agenda

- **Introduction**

- **Nonlinear estimation**
  - Algorithm formulation
  - Test data
  - Results

- **Conclusions**
Algorithm Development: Overview

- **Objectives**
  - Improve pixel-level detection: Reduce probability of false alarm for given \( P_d \)
  - Address optically-thick plumes: Improve accuracy of estimated path integrated concentration (column density, CL)
  - Compatible with real-time processing

- **Limitations of current practice**
  - Matched-filter-based detection presumes optically-thin plume
  - Other approaches require prior measurements of background – not compatible with on-the-move detection

- **Payoff:** Improve detection immediately following large-scale release, low-lying plumes; improve mass estimate
Problem Formulation

- Ensemble of measured spectra
- Measured spectra are nonlinear functions of atmospheric temperature, constituent profiles, background characteristics, etc.
- Desire inverse solution to radiative transfer equation (RTE)
- Inverse solution is mathematically ill-posed – no unique solution for $R_n$
Relation to Atmospheric Profile Retrieval

- **Stratified atmosphere model**
- **Profile retrieval**
  - Many stratifications
  - Simple background
  - Apply constraints to layer-to-layer variation
- **Plume detection**
  - Simple atmosphere
  - Complicated background
  - Apply constraints to background characterization
Simplified Radiative Transfer Model

- **Simplifying assumptions:**
  - Homogeneous atmosphere between sensor and vapor cloud
  - Cloud is at air temperature

- **Compare performance of non-linear (exact) RT model with linearized approximation**

\[
x_p = x_0 + (1 - \tau_p) \cdot \left[ (L_a - x_0) + \tau_a \cdot (L_p - L_a) \right]
\]

<table>
<thead>
<tr>
<th></th>
<th>Linear (approx.)</th>
<th>Non-Linear (exact)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plume transmission (\tau_p)</td>
<td>(1 - \alpha_s)</td>
<td>(\exp(-\alpha_s))</td>
</tr>
<tr>
<td>Radiance contrast (L_a - x_0)</td>
<td>(\propto \Delta T_{\text{eff}})</td>
<td>(\text{any})</td>
</tr>
<tr>
<td>Plume temperature (T_p)</td>
<td>(T_p = T_a)</td>
<td>(T_p = T_a)</td>
</tr>
<tr>
<td>Atmospheric scattering</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
Adaptive InfraRed Imaging Spectroradiometer (AIRIS)

- **Imaging Fabry-Perot spectrometer**
  - Mirror spacing $\sim \lambda$
  - Staring IR FPA
  - Band sequential data acquisition
  - Co-registration of narrowband images
  - Tune time $\sim 2$ ms

- **Selective sampling of wavelengths**
  - Acquire imagery only at wavelengths which facilitate target ID
  - Minimize data volume

- **Wide field-of-view, wide spectral coverage**
TEP Detection:
Shortcoming of Thin Plume Approximation

- Triethyl phosphate (TEP) release
- Post-processing:
  - Non-linear estimator in IDL
  - False alarm mitigation: 4 of 8 spatial filter
  - Bad pixels substituted
- Detection key:
  - TEP only
  - Yellow: OD ~ 0
  - Red: OD ≥ 1
Agenda

- Introduction

- **Nonlinear estimation**
  - Algorithm formulation
  - Test data
  - Results

- Conclusions
Optimal Estimation: Bayesian Approach

- **Bayesian posterior pdf for model parameter values:**
  \[
  p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}
  \]

- **Maximum likelihood parameter values maximize posterior:**
  \[
  \hat{\theta} = \arg \max \{p(\theta|x)\} = \arg \min \{-\ln p(\theta|x)\}
  \]

- **Multi-variate normal pdf for deviation between model and measurement:**
  \[
  -\ln p(x|\theta) = \frac{1}{2} [\bar{x} - f(\theta)]^T D^{-1} [\bar{x} - f(\theta)] + c_{x|\theta}
  \]
  \[
  D = \text{diag}\{\sigma_1^2, \sigma_2^2, ..., \sigma_k^2\}
  \]

- **Prior pdf for model parameter values**
  \[
  -\ln p(\theta) = \frac{1}{2} [\theta - \theta_a]^T R[\theta - \theta_a] + c_{\theta}
  \]
Optimal Estimation: Signal Model

- **Signal model:**
  \[ f(\theta) = \tau_e \circ x_0 + \left[1 - \tau_e \right] \circ L_a \]

- **Plume transmission:**
  \[ \tau_e = \exp[-\alpha s] \]

- **Infrared background:**
  - Linear mixing model
  - Probabilistic Principal Components
  - Robust estimate of sample covariance (Huber-type M-estimator)

- **Model parameters:**
  - \( \alpha \): Plume OD
  - \( T_a \): Plume/air temperature
  - \( \beta \): Parameters which account for bkgd. radiance given bkgd. model
Minimize Cost Function

- Maximum likelihood parameter values minimize cost function
- Multivariate normal pdfs result in "quadratic" cost function

\[ C = \left[ x - f(\theta) \right]^T D^{-1} \left[ x - f(\theta) \right] + \left[ \theta - \theta_a \right]^T R_\theta \left[ \theta - \theta_a \right] \]

- Quadratic formulation: \( C = r^T r \)
- Prior applied to background coefficients only: \( \left[ \theta - \theta_a \right]^T R_\theta \left[ \theta - \theta_a \right] = \beta^T \beta \)
- Residuals vector: \( r \equiv \left[ D^{-1/2} \left[ x - f(\theta) \right], \beta \right] \)

- Determine maximum likelihood parameter values by nonlinear estimation
  - Approach not limited to quadratic cost function
  - Quadratic cost function amenable to computationally-efficient solution
Nonlinear Optimization Algorithms

- Iterative determination of parameters, e.g., Newton's Method:

  \[ \theta_{i+1} = \theta_i - H_i^{-1} (\nabla C)_i \]

- Gauss-Newton algorithm
  - Approximate Hessian matrix: \( H \approx 2J^T J \)
  - Parameter update equation:
    \[ \theta_{i+1} = \theta_i - (J_i^T J_i)^{-1} J_i^T r_i \]
  - Initial guess at \( \theta \) from linear model

- Levenberg-Marquardt algorithm also applicable
Agenda

- Introduction

- Nonlinear estimation
  - Algorithm formulation
  - Test data
  - Results

- Conclusions

- Next generation algorithm(s)
Test Regions

- **Plume-free data augmented with synthetic plumes:**
  - 64 x 5 pixels
  - Max OD from 0 to 3.0 (base e)
  - $T(\text{plume}) = T(\text{air}) = 25.0 \text{ deg C}$

- **Thermal contrast**
  - $\sim 0 \text{ K along horizon}$
  - Monotonic increase with elev. angle

- **Test both favorable and unfavorable regions**
Simulation: Synthetic R-134a Plumes

- **Effective plume transmission:**
  - Reference spectrum from PNNL library
  
  \[ \tau(\lambda) = \exp[-CL \cdot \sigma(\lambda)] \]
  
  - Specify column density
  - Beer’s Law + instrument resolution function

- **Data augmentation:**
  - Partition measurement into estimated signal, noise
  - Modify signal w/plume signature
  - Add back estimated noise

\[ x_p = \hat{x}_0 + [1 - \tau_p] \odot [L_a - \hat{x}_0] + \hat{\epsilon} \]

\[ \tau_p(\lambda_s) = \int \tau(\lambda) \cdot g(\lambda, \lambda_s) \cdot d\lambda \]
Performance Metric: ROC Curves

- Binary decision hypotheses
  - $H_0$ ("plume absent") and $H_1$ ("plume present")
  - pdfs for detection statistic:
    $$ p(F \mid H_1) $$

- ROC curve is $P_d(F_{th})$ vs $P_{fa}(F_{th})$
  - $P_d$ from plume-augmented region
  - $P_{fa}$ from rest of scene

- ROC "surface": $P_d(\alpha; F_{th})$

$$ P_{fa}(F_{th}) = \int_{F_{th}}^{\infty} p(F \mid H_0) d\eta $$

$$ P_d(F_{th}) = \int_{F_{th}}^{\infty} p(F \mid H_1) d\eta $$
Agenda

• Introduction

• Nonlinear estimation
  – Algorithm formulation
  – Test data
  – Results

• Conclusions
Performance Comparison with Matched Filter

- **Objective:** Compare nonlinear estimation with matched filter estimation
  - Detection statistics
  - Column density/optical density
- **Detection with nonlinear estimator:** F test

\[
F(\tilde{x}) = (k - 1) \cdot \left[ \frac{C(\tilde{x}, \hat{\theta}_0)}{C(\tilde{x}, \hat{\theta})} - 1 \right]
\]

- Analogous metric for clutter-matched filter: Adaptive Cosine Estimator (ACE)

\[
D_{MF}(\tilde{x}) = \frac{\left(s^T \hat{\Sigma}^{-1} [\tilde{x} - \mu]\right)^2}{\left(s^T \hat{\Sigma}^{-1} s\right) \left(\tilde{x} - \mu\right)^T \hat{\Sigma}^{-1} [\tilde{x} - \mu]}
\]

\[
F_{MF} = (k - 1) \frac{D_{MF}}{1 - D_{MF}}
\]

- Matched-filter optical density estimate:

\[
\hat{\alpha}_{MF} = \frac{s^T \hat{\Sigma}^{-1} (\tilde{x} - \mu)}{s^T \hat{\Sigma}^{-1} s} \cdot \frac{\Delta T_0}{\Delta T_{eff}}
\]

- **Expect near identical results for optically-thin plumes**
R-134a Detection: Optically-Thin Plume, OD=0.1

- Plume column density = 82 mg/m² (20 ppmv-m)
- Detection statistics not favorable in either Region
- ACE and Gauss-Newton ROC curves are nearly identical
  - 20 bands in test datacube
  - OD=0 reference spectrum
R-134a Detection: Optically-Thin Plume, OD=0.3

- Plume column density = 246 mg/m² (59 ppmv-m)
- Detection statistics not favorable in Region 1, marginal in Region 2
  - Lower thermal contrast
  - ~2 orders of magnitude reduction in $P_{fa}$ from Region 1 to Region 2
- ACE and Gauss-Newton ROC curves are nearly identical
R-134a Detection: Optically-Thick Plume, OD=1.0

- **Plume column density** = 822 mg/m² (197 ppmv-m)
- **Detection statistics favorable in Region 2, marginal in Region 1**
  - >2 orders of magnitude reduction in $P_{fa}$ from Region 1 to Region 2
- **Gauss-Newton produces significantly more favorable ROC curves than ACE**
  - Factor of ~2 improvement in Region 1 ($P_{fa}$ for fixed $P_d$)
  - Multiple orders of magnitude improvement in Region 2
R-134a Detection: Optically-Thick Plume, OD=2.0

- Plume column density = 1643 mg/m² (394 ppmv-m)
- Detection statistics favorable in both Regions
- Gauss-Newton produces significantly more favorable ROC curves than ACE
  - >1 order of magnitude improvement in Region 1
  - Multiple orders of magnitude improvement in Region 2
Column Density Estimation

- Increased thermal contrast reduces uncertainty, no effect overall accuracy
- Nonlinear estimation
  - Accurately recovers embedded OD (CL)
  - Systematic deviation at OD>1 is instrument resolution effect
- Matched Filter systematically underestimates CL
- **Nonlinear estimator always as good or better than MF**
Algorithm Execution

- Gauss-Newton algorithm is iterative
- Termination criterion:
  \[ 0 < \left[ 1 - \frac{C_{i+1}}{C_i} \right] < \delta_{\text{max}} \]
- Initial guess is Iteration 0
- Typical results:
  - 1-2 iterations for no plume (plume OD=0)
  - 3 iterations to converge for OD~2-3 TEP plume
- Decreasing $\delta$ to 0.0001 increase no. iteration but no statistically-significant effect on CL

\[ \delta_{\text{max}} = 0.01 \]
Summary and Conclusions

- Developed nonlinear estimator for plume detection and characterization based on RTE
  - Bayesian formulation
  - Statistical model for IR background
  - Gauss-Newton algorithm to estimate maximum a posteriori (MAP) values

- Signal model developed for non-scattering atmosphere, single layer plume
  - Easily modified to address more complicated atmospheres

- Nonlinear estimation significantly outperforms matched-filter-based with optically-thick plumes
  - "Orders of magnitude" improvement
  - NL estimator and matched filter produce equivalent results for optically-thin plumes

- This work was performed under Contracts from the Defense Threat Reduction Agency (HDTRA01-07-C-0067) and US Army ECBC Aberdeen Proving Ground, MD (W911SR-06-C-0022). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of HDRA or the Army.
Additional Material
Focus is pixel-level target detection

New background characterization approach facilitates improved pixel-level detection

"A chain is only as strong as its weakest link."
  - Provide higher quality input to False Alarm Mitigation block
  - False Alarm Mitigation is separate issue
Technical Approach

- Adapt methodology used for atmospheric profile retrieval from space-based sensor data (e.g. AIRS, IASI, MODIS, TES)
  - Parameterize Radiative Transfer Equation (RTE)
  - Apply Estimation Theory to determine max. likelihood parameter values
  - Exploit large data set: utilize ensemble statistics

- Rationale:
  - Physics-based model for observations
  - Statistically-justified constraints
  - Strong theoretical foundation (see, e.g., C.D.Rodgers, *Inverse Methods for Atmospheric Sounding*)

- Benefits
  - Adaptable framework
  - Immediate application to non-scattering atmosphere
  - Can modify RTE to address more complicated atmospheres
Linear Models

\[ x_p = x_0 + \alpha' s' \]

\[ x_0 = \mu + B \beta \quad \text{"Structured background"} \]

\[ x_0 = \nu_b \sim N(\mu, \Sigma) \quad \text{"Unstructured background"} \]

- **"Structured Background"**
  - Values of \( \beta \) are unconstrained
  - Generalized Likelihood Ratio Test:
    \[ D_{GLRT}(x) = \frac{x^T P_B^\perp x}{x^T P_{SB}^\perp x} \]
    \[ P_B^\perp = I - B(B^T B)^{-1} B^T \]
  - Typical implementation: \( B = \) eigenvectors of sample covariance matrix

- **"Unstructured Background"**
  - \( \nu_b \) is a random vector
  - Adaptive Cosine Estimator:
    \[ D_{ACE}(x) = \frac{\left[ s^T \Sigma^{-1} x \right]^2}{\left[ s^T \Sigma^{-1} s \right] \left[ x^T \Sigma^{-1} x \right]} \]

Pros and Cons of Linear Approximation

- **Pro:** Matrix multiplication results in fast computation
  - All spectra in ensemble may be processed in parallel
  - Major computational expense is diagonalization of sample covariance matrix
  - \textit{AIRIS-WAD:} \(<150\) ms to process 65536 twenty element spectra for four target signatures (using 2005 vintage technology)

- **Pro:** Detection statistics well-understood for Gaussian noise

- **Con:** Underlying physical assumptions not valid for detection scenarios of interest
  - Mathematical model not matched to physics
  \[
  \tau_p = \exp(-\alpha s) \approx 1 - \alpha s
  \]
  - Linear approximation to Beer's Law can introduce significant error
Why Gauss-Newton Yields Better Results

- Model is matched to the data
- Fit residuals are systematically larger with linear model
  - Result of least-squares minimization
  - Location of largest residuals highly correlated with strongest R-134a absorption features
Adaptive Infrared Imaging Spectroradiometer – Wide Area Detector (AIRIS-WAD)

- **Optical:**
  - 256 x 256 pixels
  - 30 deg x 30 deg FOV
  - spectral coverage: 7.9 to 11.2 µm at ~0.1 µm resolution (~1% of λ)

- **Datacubes:**
  - 20 wavelengths
  - user selectable λ’s, specified prior to mission

- **Real-time datacube processing:** up to 3 Hz

- **Detection algorithm history:**
  - ACE: since Spring 2006
Hyperspectral Background Model

- **Probabilistic Principal Components-based**

- **Linear mixing model**
  
  \[ x = \mu + B\beta \]

- **Eigenvalue-based covariance regularization**

  \[
  \Sigma \approx \hat{\Sigma} = BB^T + \epsilon D \\
  \Sigma = D^{1/2}(U\Lambda U^T)D^{1/2} \\
  B = D^{1/2}U_m(\Lambda_m - \epsilon I_m)^{1/2}
  \]

- \( \Sigma = \text{robust estimate of sample covariance: Huber-type M-estimator} \)
Gauss-Newton Algorithm

- Follows from Newton's method – simplifying approximations
- Good for solving weakly nonlinear equations
- Hessian matrix:

\[
H_{jk} = 2 \sum_{q=1}^{m} \left[ \frac{\partial r_q}{\partial \theta_j} \frac{\partial r_q}{\partial \theta_k} + r_q \frac{\partial^2 r_q}{\partial \theta_j \partial \theta_k} \right]
\]

\[
\approx 2 \sum_{q=1}^{m} \left[ \frac{\partial r_q}{\partial \theta_j} \frac{\partial r_q}{\partial \theta_k} \right] = 2J^T J
\]

- Gradient:

\[
[\nabla_{\theta} C]_j = \frac{\partial C}{\partial \theta_j} = 2 \sum_{q=1}^{m} \left[ r_q \frac{\partial r_q}{\partial \theta_j} \right]
\]

\[
\nabla_{\theta} C = 2J^T r
\]

- Parameter update equation:

\[
\theta_{i+1} = \theta_i - (J_i^T J_i)^{-1} J_i^T r_i
\]

- Initial guess at \( \theta \) from linear model