Survey of Evolutionary and Probabilistic Approaches for Source Term Estimation

Branko Kosović

George Young, Kerrie J. Schmehl, Dustin Truesdell (PSU), Sue Ellen Haupt, Andrew Annunzio, Luna Rodriguez (NCAR),

Julie K. Lundquist (Univ. Colorado), Luca Delle Monache (NCAR)
Outline

- Common components of source term estimation (STE)
- Probabilistic approach using Bayesian inference
  - Example: Algeciras accidental release
- Optimization using Genetic Algorithms
  - Example: Redoubt volcano release
- Summary
Source Term Estimation Problem

Requirements for STE methodology:

- Effective (quantitative & accurate)
- Efficient (within time constraints)
- Flexible (adaptable, multiple data types)
- Robust (operational use)
- Quantifies uncertainty (probabilistic)
Source Term Estimation Process

Input Data

Transport and Dispersion

Metric

Convergence

Updated Parameter Estimate
Source Term Estimation Process

Input Data
- Sensor Data
- Met Data (Measurements, NWP)
- Initial Source Parameter Estimates (Guesses/Samples from Prior)

Transport and Dispersion

Metric

Updated Parameter Estimate

Convergence
Source Term Estimation Process

**Input Data**
- Sensor Data
- Met Data (Measurements, NWP)
- Initial Source Parameter Estimates (Guesses/Samples form Prior)

**Transport and Dispersion**
- Analytic, Eulerian, Lagrangian
- Forward
- Backtrajectory
- Adjoint

**Metric**

**Updated Parameter Estimate**

**Convergence**
Source Term Estimation Process

**Input Data**
- Sensor Data
- Met Data (Measurements, NWP)
- Initial Source Parameter Estimates (Guesses/Samples from Prior)

**Transport and Dispersion**
- Analytic, Eulerian, Lagrangian
- Forward
- Backtrajectory
- Adjoint

**Metric**
- Cost Function
- Likelihood Function

**Updated Parameter Estimate**

**Convergence**
Convergence

Source Term Estimation Process

Input Data:
- Sensor Data
- Met Data (Measurements, NWP)
- Initial Source Parameter Estimates (Guesses/Samples from Prior)

Transport and Dispersion
- Analytic, Eulerian, Lagrangian
- Forward
- Backtrajectory
- Adjoint

Metric
- Cost Function
- Likelihood Function

Updated Parameter Estimate
- Variational Approach
- Optimization/Minimization (with or without gradients)
- Stochastic Sampling

Convergence
Source Term Estimation Process

Input Data:
- Sensor Data
- Met Data (Measurements, NWP)
- Initial Source Parameter Estimates (Guesses/Samples from Prior)

Transport and Dispersion
- Analytic, Eulerian, Lagrangian
- Forward
- Backtrajectory
- Adjoint

Metric
- Cost Function
- Likelihood Function

Updated Parameter Estimate
- Variational Approach
- Optimization/Minimization (with or without gradients)
- Stochastic Sampling

Convergence
- Single Set of Parameters
- Probability Distribution
Source Term Estimation

Input Data:
- Sensor Data
- Met Data (Measurements, NWP)
- Initial Source Parameter Estimates (Guesses/Samples from Prior)

Transport and Dispersion
- Analytic, Eulerian, Lagrangian
- Forward
- Backtrajectory
- Adjoint

Metric
- Cost Function
- Likelihood Function

Updated Parameter Estimate
- Variational Approach
- Optimization/Minimization (with or without gradients)
- Stochastic Sampling

Convergence
- Single Set of Parameters
- Probability Distribution

PROBABILISTIC APPROACH USING BAYESIAN INFERENCE WITH STOCHASTIC SAMPLING
Source Term Estimation

Input Data:
- Sensor Data
- Met Data (Measurements, NWP)
- Initial Source, Met Data Estimates (Guesses/Samples from Prior)

Transport and Dispersion
- Analytic, Eulerian, Lagrangian
- Forward
- Backtrajectory
- Adjoint

Metric
- Cost Function
- Likelihood Function

Optimization Approach using Genetic Algorithm

Updated Parameter Estimate
- Variational Approach
- Optimization/Minimization (with or without gradients)
- Stochastic Sampling

Convergence
- Single Set of Parameters
- Probability Distribution
Outline

- Common components of source term estimation (STE)
- Probabilistic approach using Bayesian inference
  - Example: Algeciras accidental release
- Optimization using Genetic Algorithms
  - Redoubt volcano release
- Summary
Models and Observations are Coupled Through Bayesian Inference

**STOCHASTIC SAMPLING OF UNKNOWN PARAMETERS**
- Informed prior and improved proposal distribution
- Markov Chain Monte Carlo
- Sequential Monte Carlo
- Hybrid and multi-resolution methods

**METEOROLOGY**

**DISPERSION MODELS**
- Urban models: (empirical puff, CFD)
- Global, regional models: (2D, 3D, puff, particle)

**OBSERVED DATA**

**ERROR QUANTIFICATION**

**BAYESIAN COMPARISON**
(Bayes Theorem)

$$P(\theta | d) = P(d | \theta) P(\theta) / P(d)$$

*Update likelihood until convergence to a posterior distribution*

Dynamic Data-Driven Event Reconstruction for Atmospheric Releases, UCRL-TR-229417
Algeciras Accidental Release

What: CS-137
When: 0100-0300 UTC, 30 May 1998
How much: 8-80 Ci

Stations: Ispra (2-5/06), Cadarache (1-2/6), Capomele (1-2/6), Ivrea (2/6), Marcuole (1/6), Milano (1-2/6), Montfaucon (2/6), Montpelier (1-2/6), Nice (1-2/6), Palermo (6-7/6), Tórno (1-2/6), Vercelli (4-6/6)
Simulation Set-up and Assumptions

- Surface point source
- Time and duration of the release
  (0130-0200 UTC, May 30, 1998)
- Sampling box (next slide)
- 11 stations for 17 observations
  (9 on 06/02/1998 + 8 on 06/03/1998)
- Zero concentrations not used
Prediction of Source Location Using Three Markov Chains

~3600 km

~1800 km
Location and Release Rate Probability Densities
Plume Prediction
After Several Hours (@ 1200 UTC)

May 30, 1998:
24-h Average Concentration (μBq/m³)
Prediction plume after 1 day
Prediction plume after 2 days
Prediction plume after 3 day
Prediction plume after 5 day
Prediction plume after 6 day

June 5, 1998:
24-h Average Concentration ($\mu$Bq/m$^3$)
Outline

- Common components of source term estimation (STE)
- Probabilistic approach using Bayesian inference
  - Example: Algeciras accidental release
- Optimization using Genetic Algorithms
  - Redoubt volcano release
- Summary
Genetic Algorithm

- Initial Random Population of Chromosomes $Q, u, \theta$

Dispersion Model

Evaluate Cost Function

New Variables

- Low Cost
  - Mate
  - Mutate
  - Iterate
  - No

- High Cost
  - Discard

Convergence?

Yes

Final Variables

No

cost function =

$$\frac{\sqrt{\sum_{s=1}^{TS} [O_s - C_s]^2}}{\sqrt{\sum_{s=1}^{TS} [O_s] \sqrt{\sum_{s=1}^{TS} [C_s]}}}$$
Redoubt Volcano Eruption

March 23, 2009 Eruption

The Alaska Volcano Observatory (AVO) observed eleven major explosive events during the first week, and a total of 19 events over the 14 day explosive eruptive period in March and early April.

Coordinates: 60°29′ N 152°44′ W
Elevation: 3108 meters (10,197 ft.)
Volcano Type: Stratovolcano
Eruption Timeline

Sunday
March 22

First Eruption Begins

First Sounding
@ 4 PM

10PM 11PM 12AM 1AM 2AM 3AM 4AM 5AM 6AM 7AM

Monday
March 23

Fifth Eruption Ends

Second Sounding

Satellite Pass

!!!!!!!!!!10PM!!!!!!!11PM!!!!!!!12AM!!!!!!!1AM!!!!!!!!!!2AM!!!!!!!!!3AM!!!!!!!!!4AM!!!!!!!!!5AM!!!!!!!!!6AM!!!!!!!!!7AM!
GOES-11 satellite provided data for use with the GA
Single Uniform Release Results

Ten Runs

<table>
<thead>
<tr>
<th>Run</th>
<th>Wind direction (°)</th>
<th>Wind speed (m s⁻¹)</th>
<th>Emission Rate (kg s⁻¹)</th>
<th>Cost Function Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>215.7</td>
<td>16.4</td>
<td>6.8x10⁴</td>
<td>0.1631</td>
</tr>
<tr>
<td>2</td>
<td>213.0</td>
<td>23.8</td>
<td>1.1x10⁵</td>
<td>0.1589</td>
</tr>
<tr>
<td>3</td>
<td>214.0</td>
<td>18.0</td>
<td>6.4x10⁴</td>
<td>0.1592</td>
</tr>
<tr>
<td>4</td>
<td>213.2</td>
<td>23.7</td>
<td>9.7x10⁴</td>
<td>0.1593</td>
</tr>
<tr>
<td>5</td>
<td>213.3</td>
<td>19.3</td>
<td>8.0x10⁴</td>
<td>0.1555</td>
</tr>
<tr>
<td>6</td>
<td>212.4</td>
<td>21.2</td>
<td>1.0x10⁵</td>
<td>0.1594</td>
</tr>
<tr>
<td>7</td>
<td>213.5</td>
<td>21.8</td>
<td>9.1x10⁴</td>
<td>0.1567</td>
</tr>
<tr>
<td>8</td>
<td>213.0</td>
<td>29.3</td>
<td>1.5x10⁵</td>
<td>0.1701</td>
</tr>
<tr>
<td>9</td>
<td>211.8</td>
<td>34.7</td>
<td>2.3x10⁵</td>
<td>0.1903</td>
</tr>
<tr>
<td>10</td>
<td>213.7</td>
<td>24.0</td>
<td>9.9x10⁴</td>
<td>0.1587</td>
</tr>
<tr>
<td>Mean</td>
<td>213.4</td>
<td>23.2</td>
<td>1.1x10⁵</td>
<td>0.1631</td>
</tr>
<tr>
<td>STD</td>
<td>1.0</td>
<td>5.4</td>
<td>4.8x10⁴</td>
<td>0.0104</td>
</tr>
</tbody>
</table>

Population 64
Generations 200
Mutation Rate 20%
Selection 50%
Single Uniform Release: GA Initialized Best Solution

<table>
<thead>
<tr>
<th>Emission Rate</th>
<th>Total Mass Emitted</th>
<th>Wind Direction</th>
<th>Wind Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.0x10^4 kg/s</td>
<td>2.6x10^9 kg</td>
<td>213.3°</td>
<td>19.3 m/s</td>
</tr>
</tbody>
</table>

Satellite data

Single Release - Initialized with GA Best Solution
Horizontal Slice at z = 6000.0m
Total ASH at 23-Mar-09 15:00Z (8.00 hrs)
How to achieve effectiveness and efficiency while dealing with complexity and uncertainty?

Both algorithms are effective but require significant computational expense.

Bayesian approach:
• Provides probabilistic solution
• Flexible framework


Genetic Algorithm approach:
• Efficient exploration of parameter space
• Single best solution, but also probabilistic