Atmospheric Environment 45 (2011) 6085–6092

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

The Aerodyne Inverse Modeling System (AIMS): Source estimation applied to the FFT 07 experiment and to simulated mobile sensor data

Simón E. Albo, Oluwayemisi O. Oluwole*, Richard C. Miake-Lye

Aerodyne Research, Inc., 45 Manning Road, Billerica, MA 01821-3976, USA

ARTICLE INFO

Article history:
Received 31 January 2011
Received in revised form
15 July 2011
Accepted 18 July 2011

Keywords:
Source estimation
Inverse modeling
Source location
Air quality
Emissions modeling
Atmospheric pollution

ABSTRACT

The Aerodyne Inverse Modeling System was developed to enable location and characterization of hazardous atmospheric releases from dispersion and meteorological data. It combines an automatically-generated tangent-linear of SCIPUFF with a cost function tailored for practical applications and a minimization algorithm that can search for multiple instantaneous or continuous sources without requiring an initial guess. In this work AIMS was applied to estimate the sources in 84 FFT 07 cases that included instantaneous and continuous releases for up to four source locations. FFT 07 was a controlled short-range (~500 m) dispersion test using 100 digiPIDs evenly distributed over an area of 0.5 × 0.5 km. AIMS estimated sources were in average within 90–150 m of the real sources, with the distances from estimated to real source ranging from 0 to 510 m. AIMS performed better estimating the location of instantaneous sources than of continuous ones. It also performed better for single-source situations than for multiple source scenarios and when 16 sensors were used instead of 4. In addition to using stationary sensors, AIMS also has the capability of processing data from mobile sensors. This was applied using model-generated data in an example of a release in a setting similar to an industrial facility.

1. Introduction

The threat of atmospheric contamination by hazardous materials remains a high national security concern. There is a strong need for the development of emerging technologies, including sensors and algorithms, which can significantly advance risk assessment and response capabilities. Often, nothing is known about the event besides dispersed concentration levels detected by sensors. Moreover, sensors can vary in type and location and more information can be obtained by integrating the different observation types. The present paper describes the development of a new source estimation algorithm that utilizes the tangent linear of SCIPUFF (Second-order Closure Integrated Puff) (Sykes and Gabruk, 1997; Sykes et al., 1985) for characterizing the source of a chemical release using multiple types of observational data, and requiring no prior knowledge of the source.

The standard modeling approach for tracking atmospheric plumes falls in the category of forward modeling: given an initial state (atmospheric state and tracer concentration) and boundary values (space and time dependent tracer emissions and the time-evolving meteorological state), a transport model is stepped forward in time to produce a field of tracer concentrations at subsequent times. In source estimation, the goal is to determine the initial and/or boundary values that lead to a model trajectory that is most consistent with the observations. Source estimation is an active topic in research and development and a variety of solution approaches have been proposed. Rao (2007) presents a good discussion of source estimation methods. The methods discussed include Bayesian and Monte Carlo techniques, which consist of many forward runs from sources to receptors together with a suitable algorithm to adjust the source conditions until an acceptable level of agreement is reached. Other methods include Kalman filtering, adjoint and tangent-linear models and variational data assimilation. Adjoint models typically use only one run in reverse direction from the receptors to determine the upwind concentrations.

Some recent inverse model applications include those by Allen et al. (2007) that coupled a genetic algorithm with SCIPUFF for emission source characterization and validated it using synthetic and pollutant dispersion field data. The same group (Long et al., 2010) also used a genetic algorithm to assess the sensitivity of source term estimation to data quantity and quality and to determine the minimum data requirements to accurately estimate the source term and to obtain the relevant wind information. They have also investigated the treatment of binary or discrete variables (e.g. atmosphere stability or the presence of rain) using a mixed integer genetic algorithm (Haupt et al., 2011). Using a different

* Corresponding author. Tel.: +1 9786639500.
E-mail address: oluwoleo@aerodyne.com (O.O. Oluwole).

1352-2310/$ – see front matter
doi:10.1016/j.atmosenv.2011.07.033

approach, Lushi and Stockie (2009) combined linear least squares with a Gaussian plume solution for the advection–diffusion equation. Their method was validated using measurements from a lead-zinc smelting facility. Other works have focused on larger scale applications, such as constraining Asian sources of carbon monoxide using an adjoint method and a Bayesian solution (Kopacz et al., 2009); and using a variational data assimilation method to study emissions and their time profile over central and western Europe (Resler et al., 2010).

In this work we present the Aerodyne Inverse Modeling System (AIMS), which combines the tangent-linear of SCIPUFF with a minimization algorithm for estimating the number of sources and their location, mass, release times and durations that are most consistent with the concentration field provided. AIMS has the advantage of not requiring an initial estimate for the source(s) parameters from the user, thus making it suitable for situations of malicious or accidental releases where there is no *a priori* information regarding the source(s). A novel feature of AIMS is the ability to integrate data from stationary sensors and also from mobile sensors; it also incorporates many heuristics into the cost function definition and in the source search algorithm to improve the convergence of the source search algorithm. In the next section the development of the model is discussed, including cost function details, implementation of automatic differentiation to obtain the tangent-linear code and details of the search algorithm. Section 3 presents results and discussion of cases using FFT 07 field data from controlled releases and also SCIPUFF-generated data for a case of mobile sensor.

2. Model development

The source of an atmospheric release can be described using the following parameters: source location, source height, mass released, time of release and duration of release. In estimating a source, all or some of these parameters are determined. AIMS has been developed for source estimation, it takes as input all available measurement data, mostly concentration values and meteorological information. The algorithm can generate an initial guess, or it can take an estimate from the user. AIMS uses SCIPUFF as the atmospheric dispersion model. A gradient approach is employed to estimate the sources: a cost function is defined to quantify the mismatch between observed and modeled concentrations; then an optimal source estimate is obtained by applying a gradient-based minimization algorithm to find the model inputs that minimize the cost function. The output is the best set of source parameters for reproducing the measurement data using the forward model. Model outputs include number of sources, emission rates, locations, and release start and end times.

2.1. Cost function

The cost function quantifies the difference between the predicted concentration field and the observed concentration field. Typically, it is defined to be zero in the limit that model predictions match the data perfectly. The cost function used in AIMS is as follows:

$$\text{Cost} = \sum_{m=1}^{M} w_{\text{m}} \left\{ \sum_{t=1}^{T} \frac{\sum_{s=1}^{N} \left[Y_{s,t}^{\text{obs}} - Y_{s,t}^{\text{mod}} \right]^2}{\max\left\{ \sum_{s=1}^{N} \left(Y_{s,t}^{\text{obs}} \right)^2, \ 0.05*\max_{\forall t} \left(\sum_{s=1}^{N} \left(Y_{s,t}^{\text{obs}} \right)^2 \right)_t \right\}} \right\}_m \tag{1}$$

where Y^{obs} refers to the observed concentrations and Y^{mod} refers to the model-predicted concentrations at time t and sensor s. The cost function is a sum of M terms corresponding to the different types of measurements m (i.e. stationary and mobile point concentration measurements) provided to the algorithm. The contributions of each measurement type are added using a weight w_{m} that is provided by the user (by default all measurement types are weighted equally). The term of each measurement type, consists of a summation over T observation times. The numerator on each time term calculates the sum of squares of the differences between the observed and modeled concentration over all N sensors at the given time. The denominator on each time term is the maximum between the sum of squares of the observations at the given t time and 5% of the maximum denominator for all times. This form is introduced as a scaling mechanism because the concentration values may span many orders of magnitude. Applying this scaling avoids artificial inflation of the cost function by low-magnitude measurements at time t: notice that as $\sum_{s=1}^{N} (Y_{s,t}^{\text{obs}})^2 \to 0$, the total cost is dominated by the model-data discrepancies at time t, even for relatively low values of the numerator. On the other hand, omitting the denominator in Eq. (1) causes the cost function to be dominated by the absolute model-data discrepancies in the larger-magnitude measurements, while the lower-magnitude data are essentially ignored. In that case, one would often obtain source estimates with large model-data discrepancies at the lower-magnitude data points. The scaling mechanism described above was found to be highly effective for simultaneously maximizing the information content of the wide range of data values typically found in practical observational datasets for source estimation.

Additionally, the cost function is artificially enlarged when a source is unphysical (negative mass or negative release duration) or it is in a location that is undetectable by the sensors (this situation presents itself only when solving for multiple sources). These penalizations are introduced to force the algorithm to find only physically viable sources that contribute to reducing the cost; essentially creating a constrained minimization algorithm while retaining the unconstrained quasi-Newton solver described below. The penalizations are introduced for each source that falls within any of the cases using the following expressions:

$$\text{Negative release duration}: \ \text{Cost} = \text{Cost} - 100 * t_{\text{duration}}; \quad \text{where } t_{\text{duration}} < 0$$

$$\text{Negative source mass}: \ \text{Cost} = \text{Cost} - 10^{10} * \text{Mass}; \quad \text{where Mass} < 0$$

$$\text{Noncontributing release}: \ \text{Cost} = \text{Cost} + T * \left(NC - \max_{\forall t} \left(\sum_{s=1}^{N} Y_{s,t}^{\text{mod}} \right)_t \right);$$

a release is considered noncontributing if its downwind concentration is always below a specified threshold NC at all sensors and at all measurement times.