Identifying Sources of Fugitive Emissions in Industrial Facilities using Trajectory Statistical Methods

Carol A. Brereton and Matthew R. Johnson*

Energy & Emissions Research Lab.
Mechanical & Aerospace Engineering
Carleton University
1125 Colonel By Drive,
Ottawa, ON, Canada K1S 5B6

*Corresponding Author:
Email: Matthew_Johnson@carleton.ca
Phone: 1 613 520 2600 ext. 4039
Fax: 1 613 520 5715

NOTICE: this is the authors’ version of a work that has been published in Atmospheric Environment. Changes resulting from the publishing process, such as editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Please click the digital object identifier (doi: 10.1016/j.atmosenv.2012.01.057) to access the definitive version of the article as published, which can be cited as: Brereton, C.A. and Johnson, M.R. (2012), Atmospheric Environment, 51: 46-55. © Elsevier, 2012

Abstract

Fugitive pollutant sources from the oil and gas industry are typically quite difficult to find within industrial plants and refineries, yet they are a significant contributor of global greenhouse gas emissions. A novel approach for locating fugitive emission sources using computationally efficient trajectory statistical methods (TSM) has been investigated in detailed proof-of-concept simulations. Four TSMs were examined in a variety of source emissions scenarios developed using transient CFD simulations on the simplified geometry of an actual gas plant: potential source contribution function (PSCF), concentration weighted trajectory (CWT), residence time weighted concentration (RTWC), and quantitative transport bias analysis (QTBA). Quantitative comparisons were made using a correlation measure based on search area from the source(s). PSCF, CWT and RTWC could all distinguish areas near major sources from the surroundings. QTBA successfully located sources in only some cases, even when provided with a large data set. RTWC, given sufficient domain trajectory coverage, distinguished source areas best, but otherwise could produce false source predictions. Using RTWC in conjunction with CWT could overcome this issue as well as reduce sensitivity to noise in the data. The results demonstrate that TSMs are a promising approach for identifying fugitive emissions sources within complex facility geometries.

Highlights

- Trajectory statistical methods can identify fugitive emissions sources
- Test data from CFD model of wind and gas release over simplified gas plant geometry
- Among several TSMs, RTWC was best, but could give false sources with poor data
- Algorithms could be made relatively insensitive to concentration measurement noise
- RTWC used in conjunction with CWT is a promising approach
1. Introduction

Fugitive emissions can generally be described as unmonitored, unintended and/or uncontrolled releases of gas into the atmosphere. Potential sources of fugitive emissions include leaking valves, seals and fittings; evaporation losses; and process faults and failures (IPCC, 2006). In the oil and gas industry, fugitive emissions are an important source of greenhouse gases (GHG) and volatile organic compounds (VOC). In 2008, fugitive emissions (excluding venting and flaring) from this sector contributed an estimated 1-4% of total GHG emissions in the European Union, United States and Canada (UNFCCC, 2010). In the Russian Federation, this fraction may have been as high as 18% (UNFCCC, 2010).

Industry methods for estimating and locating sources of fugitive emissions can be time-consuming and challenging to implement. Rigorous surveys can be performed in accordance with the United States Environmental Protection Agency (USEPA) Method 21 (USEPA, 1995), a protocol for checking potential sources of fugitive emissions that requires the operator to use a probe (or bubble solution) to test each designated component. However, with Method 21, about 20% of potential leaks are not regularly checked because of a lack of resources, difficulty of access, and safety concerns (Chambers, 2006).

In response, alternative detection methods have been created. Ultrasonic leak detection can detect high-pressure jet leakage, though is sensitive to background noise and leak size (Gassonic, n.d.). The light absorption properties of methane and other hydrocarbons allow leaking gas to be viewed using equipment such as infrared cameras and various laser techniques (Chambers, 2004). Site surveys are generally performed sporadically due to their expense and their time consuming nature.

An alternative method is to combine downstream concentration measurements with wind data to locate major unknown fugitive emissions sources. An installation of wind measurement instruments and concentration sensors could determine likely pollutant source locations on a quasi-continuous basis. Maintenance could then be directed quickly to areas providing the most benefit in terms of GHG emission reduction and monetary value of lost gas, while enabling directed searches of a reduced search space that could be performed more often and to greater effect. Moreover, such a system could allow external verification of continued mitigation, which is critical for emissions crediting or trading.

Three principle challenges to overcome in such a system include:

1. Developing an appropriate sensor network,
2. Determining wind flow, and
3. Implementing effective source location algorithms based on collected data.

A novel tunable diode laser based sensor network, suitable for industrial deployment, is being designed in complementary research (Schoonbaert & Johnson, 2011). Wind flow simulations for the test-cases presented are accomplished using a commercially available computational fluid dynamics (CFD) code. This paper focuses on the development and testing of potential source location algorithms to solve the inverse dispersion problem that arises from combining wind flow information within and
around the plant and concentration data from a sensor network. Several different algorithms were considered and compared over a range of source scenarios in specified wind field conditions. Poor performance in this idealized situation would indicate that such algorithms are not worth further pursuit for this application.

1.1 Trajectory Statistical Methods
Several different approaches exist for solving inverse pollutant transport, most often by optimizing a solution or choosing a best solution among several candidates. Sohn et al. (2002) used Bayesian Monte Carlo to characterize releases in buildings. Hartley and Prinn (1993) determined surface fluxes of CFC$_3$ using a Kalman filter, transport models, and industry data. Allen et al. (2007) found the location of a release in an empty field using genetic algorithms. Thomson et al. (2007) identified sources of ethane dispersion in the desert with simulated annealing. Robertson and Persson (1993) applied variational data assimilation to a fictional release event. Unfortunately these methods tend to be very computationally intensive so are not ideally suited to the target fugitive emissions detection application. However, other methods such as trajectory statistical models (TSMs) developed for continental scale problems are potentially well-suited.

Trajectory statistical methods (TSMs) use backward trajectories, and in some cases residence times along these trajectories, to determine likely source regions. To date, TSMs have been applied to large, continental-scale pollutant source identification, although as discussed in this paper, the same theory is extensible to building-scale phenomena. The critical advantages of these methods are computational speed and implementation flexibility.

Four TSMs were investigated to test their ability to locate sources of fugitive emissions within a complex geometry representative of an actual petroleum industry gas plant: Potential Source Contribution Function (PSCF), Concentration Weighted Trajectory (CWT), Residence Time Weighted Concentration (RTWC), and Quantitative Transport Bias Analysis (QTBA). PSCF (Ashbaugh et al., 1985) links upwind residence time from measurement locations to a conditional probability field and has been used extensively to identify sources and transport pathways for trace substances and particulate over continental scales (Hsu et al., 2003; Ashbaugh et al., 1985; Lupu et al., 2002; Cheng & Lin, 2001). From a modified version of PSCF, Seibert et al. (1994) developed the CWT method, which was further refined by Stohl (1996) into the RTWC method. RTWC recognizes that hot spots can exist along a trajectory and reweights the relative importance of the field accordingly (Hsu et al., 2003; Stohl, 1996). Finally, QTBA (Keeler and Samson, 1989) attempts to account for the trajectory uncertainty inherent in TSMs.

To permit controlled investigation, synthetic test cases were created to analyze the performance of the algorithms under a common set of wind field and pollutant release conditions. Using a simplified geometry of an operating gas plant in Alberta, Canada, transient wind flow and pollutant releases were simulated using CFD. Concentration data could then be extracted from varying numbers of virtual sensors and, along with wind field data, used as inputs for TSM calculations. The predicted source locations for
the various algorithms could then be directly compared with the known, simulated leak(s). This work focuses on demonstrating applicability and determining best parameters of algorithms to locate pollutant sources within a complex plant geometry considering computational effort, sensitivity to poor quality input data, and ability to provide useful accuracy.

2. Data sources and model descriptions

2.1 Trajectory Statistical Method Theory

TSMs are receptor-based inverse models that determine where gas has been by calculating reverse trajectories starting at a sensor location and stepping backwards in time. Linked with the measurement data corresponding to sensor arrival time, the ensemble of trajectories can provide knowledge about the rest of the domain and determine pollutant sources and pathways.

2.1.1 Potential source contribution function (PSCF)

PSCF (Ashbaugh et al., 1985) is based on the idea that regions frequently occupied by high-concentration trajectories are more likely to be important source regions or pathways. Trajectories are identified as "polluted" or "unpolluted" by comparison of corresponding sensor readings with a chosen threshold. The domain is divided into cells, and points along a trajectory separated by a specified time interval, $\Delta t$, are used to calculate residence time statistics for each cell. More points within a cell correspond to longer residence times. Within cell $ij$, $a_{ij}$ is the total number of trajectory-points and $b_{ij}$ is the total number of polluted trajectory-points. Given $N$ total points within the domain, the probability that an air parcel passed through cell $ij$ to reach the sensor can be estimated as $P[A_{ij}]=a_{ij}/N$. Similarly, the probability that a polluted air parcel passed through $ij$ to reach the sensor is estimated as $P[B_{ij}]=b_{ij}/N$. The conditional probability that an air parcel passing through cell $ij$ arrives polluted at a sensor defines the potential source contribution as given by Equation (1), where high PSCF values correspond to likely important pollutant sources or pollutant pathways.

$$P_{SCF_{ij}} = \frac{P(B_{ij}|A_{ij})}{a_{ij}/N} = \frac{b_{ij}}{a_{ij}}$$

2.1.2 Concentration weighted trajectory (CWT) and residence time weighted concentration (RTWC)

For the CWT algorithm, each trajectory is weighted by the measured concentration. The current study used the variation on CWT of Hsu et al. (2003) based on the method developed by Seibert et al. (1994). Given $M$ total trajectories, $C_m$ is the measured concentration at the receptor for trajectory $m$, $\tau_{ijm}$ is the time spent by trajectory $m$ within cell $ij$, and the CWT field is calculated as:

$$C_{ij}^- = \frac{1}{\sum_{m=1}^{M} \tau_{ijm}} \sum_{m=1}^{M} C_m \tau_{ijm}$$

The RTWC refinement method of Stohl (1996) redistributes the CWT results field recognizing that cells intersected by both high and low concentration trajectories are unlikely to be major sources (which must be elsewhere along the high concentration trajectories). Again using the variation of Hsu et al. (2003), each trajectory $m$ is divided into $N_m$ segments, where $X_{nm}$ is the mean concentration of segment $n$ of trajectory $m$, and $\overline{X_m}$ is the average of concentrations along that trajectory. Segment $n$ contributes to trajectory $m$ via:
\[ C_{nm} = C_m \frac{X_{nm}N_m}{\sum_{j=1}^{N_m}X_{jm}} = C_m \frac{X_{nm}}{X_m} \]  

(3)

A new pseudo-concentration field is then calculated as:

\[ \overline{C}_{ij} = \frac{1}{\sum_{n=1}^{M} \sum_{m=1}^{N_m} t_{jnm}} \sum_{n=1}^{M} \sum_{m=1}^{N_m} C_{nm} t_{jnm} \]  

(4)

This redistribution repeats until a chosen convergence level for the final RTWC field, specified here as a deviation from the previous iteration that is less than 1% of the max value observed in the initial guess field.

2.1.3 Quantitative transport bias analysis (QTBA)

QTBA (Keeler and Samson, 1989) attempts to account for the uncertainties in the upwind trajectories by representing trajectory locations with Gaussian profiles. Zhou et al. (2004) simplified QTBA by neglecting deposition and chemical reactions and Zhao et al. (2007) replaced the complex empirical weighting filter with a simple weighting function. Following Zhao et al. (2007), the final QTBA field used is:

\[ QTBA(x, y|x', y') = \frac{\sum_{m=1}^{M} T_m(x, y|x', y') C_m}{\sum_{m=1}^{M} T_m(x, y|x', y')} w(x, y) \]  

(5)

where \( x \) and \( y \) define the grid cell coordinates, \( x' \) and \( y' \) are trajectory points upstream at time \( t' \), \( M \) is the total number of trajectories, \( C_m \) is the concentration corresponding to trajectory \( m \), \( T \) is the total time upstream, \( w \) is the filter applied to the results field, and the natural transport probability field is:

\[ T_m(x, y|x', y') = \frac{1}{\int_{t-T}^{t} \int_{t} \exp \left[ -\frac{1}{2} \left( \frac{x-x'(t')}{\sigma_x} \right)^2 + \frac{1}{2} \left( \frac{y-y'(t')}{\sigma_y} \right)^2 \right] dt'} \]  

(6)

\( \sigma_x \) and \( \sigma_y \) are the standard deviations about the trajectory and are approximated by \( \sigma(t') = at' \) where \( a \) is a constant. High regions within this field are related to pollutant pathways and sources.

2.2 Idealized Gas Plant Geometry and Test Case Creation

To evaluate the source location algorithms, simulations were performed on an idealized geometry of an actual gas plant located in Alberta, Canada (see Figure 1). The plant consists of approximately 70 buildings with an approximate average height of 10 m, and two 122 m tall stacks. A total domain size of 1 km x 1 km x 0.2 km was considered using the k-epsilon RNG turbulence model within ANSYS CFX 12.1. This maintained a blockage ratio of less than 3% and a minimum distance of 250 m from the outflow boundary in all directions, thus minimizing outlet boundary influence within the plant region. The mesh had 0.5 m maximum size and 1.3 expansion ratio at the buildings, 5 m maximum size and 1.4 expansion ratio at the ground away from the buildings, and 20 m maximum size overall. Three primary wind directions (varying within three of four quadrants) were simulated (Table 1), each for approximately twenty minutes of physical time at a timestep of 0.2 s, with data stored every second. The fourth wind quadrant was ignored recognizing that wind coverage in typical locations is not directionally complete. Power law velocity inlet profiles of

\[ u(z) = u_{10m} \left( \frac{z}{10 m} \right)^{1/7} \]  

with velocity \( (u) \) specified at a height \((z)\) of 10 m (Table 1) were used. Turbulence intensity \((I)\) profiles between 1 m and 160 m heights were approximated by the empirical relation

\[ I_u = \ln^{-1} \left( \frac{z}{z_0} \right) \]  

(Holmes, 2001) using a roughness height \((z_0)\) of 0.04 m and held constant outside this range. Full details of the simulations are described in Brereton (2010).
Neutrally buoyant gas releases were simulated using the transport equation. Industry investigation suggests that the majority of fugitive emissions in a typical oil and gas facility are due to a small number of leaking components (CAPP, 2007). Thus, four source locations at 1 m heights (Figure 1) with two source strengths representative of typical industry values (Table 2) were used to evaluate performance of the various TSMs in locating important sources under different scenarios. Since fugitive methane is of critical importance in the oil and gas industry, its kinematic diffusivity was specified for the simulated release gas.

Table 1: Inlet conditions at 10 m height using linear variation over duration

<table>
<thead>
<tr>
<th>Inlet Condition</th>
<th>Speed variation [km/h]</th>
<th>Directional variation [deg]</th>
<th>Duration [min]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17.0 to 14.9</td>
<td>190 to 215</td>
<td>21.5</td>
</tr>
<tr>
<td>2</td>
<td>6.0 to 6.8</td>
<td>100 to 126</td>
<td>22.0</td>
</tr>
<tr>
<td>3</td>
<td>5.0 to 7.5</td>
<td>10 to 35</td>
<td>21.5</td>
</tr>
</tbody>
</table>

Table 2: Source strengths for simulated fugitive emissions

<table>
<thead>
<tr>
<th>Number</th>
<th>Type</th>
<th>Mass Flow [g/s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Large</td>
<td>4.50</td>
</tr>
<tr>
<td>2</td>
<td>Medium</td>
<td>0.45</td>
</tr>
<tr>
<td>3</td>
<td>Large</td>
<td>4.50</td>
</tr>
<tr>
<td>4</td>
<td>Medium</td>
<td>0.45</td>
</tr>
</tbody>
</table>

2.3 Data and Trajectory Calculations

Thirty-one virtual sensors were placed throughout the domain (Figure 1) at 1 m elevation, and used in all tests except when effects of reduced sensor numbers were specifically being considered. In contrast to continental studies in which single sensors or close clusters of sensors are often used due to data availability, in an industrial plant environment placement of sensors is expected to be primarily limited by cost. For a fibre-optically coupled sensor network, systems with up to 32 or 64 sensors should be readily achieved (e.g. Whitenett et al., 2003).

Concentrations were interpolated to sensor locations and extracted every 30 s. As there is a physical limit to any detection system, concentrations below $10^{-6}$ kg/m$^3$ (~1.5 ppmv for methane, corresponding to the nominal measurement sensitivity of a Tuneable Diode Laser Absorption Spectroscopy (TDLAS) system, the anticipated technology for use in a plant environment) were reduced to zero. This detection threshold is equivalent to 1.5 ppmv above nominal background methane concentrations.

Backward trajectories corresponding to each concentration measurement were calculated and propagated upstream of each sensor for a maximum duration of 180 s in physical time. The trajectories were calculated with a 5th order explicit Runge-Kutta method (Butcher, 1964)
using velocity information extracted at all positions within the domain using linear interpolation. Velocity components between known timesteps were interpolated linearly to a resolution of 0.1 s.

From this total dataset of 4030 measurements and trajectories, several smaller datasets were created. Separate tests were performed using trajectories from each primary wind quadrant, using simulated trajectories from all three quadrants (i.e. a complete data set), using every third trajectory from all wind quadrants (i.e. increasing the time interval between calculated trajectories from every 30 s to every 90 s), and using data from every third sensor for all wind quadrants (i.e. using fewer sensors). Data set and parameter settings were fed into a C++ implementation of each algorithm. Post-processing including filtering was performed using Mathematica.

3. TSM Parameters
Several parameters that can influence the performance of the TSMs needed to be optimized prior to algorithm comparison to ensure an objective evaluation as discussed below.

3.1 Cell size
TSM cell size is a balance between the number of passing trajectories and spatial resolution. Square cells of four sizes (lengths of 5, 10, 20, or 40 m) on the horizontal plane were considered for each test case.

3.2 Filters
Zeng and Hopke (1989) noted that the PSCF value could be 1.0 (or 100%) even if a cell contains only a single polluted trajectory point. CWT and RTWC can similarly produce unreasonably high resulting values in areas with few trajectories. Several filtering approaches were explored to reduce the influence of regions in the domain with unreliable statistics.

Cells with low trajectory counts (i.e. number of different trajectories) or point counts (i.e. total residence time of all trajectories present) can simply be ignored (e.g. Zhou et al., 2004). In the present case, filters with minimum counts of four-trajectories or four-points per cell were used. Zeng and Hopke (1989) proposed a weighting function for PSCF, based on point count multiplied by the local PSCF value as follows:

$$\text{weight} = \begin{cases} 
0.5 & n = 1 \\
0.68 & n = 2 \\
0.85 & n = 3 \\
1.0 & n \geq 4 
\end{cases}$$

A modified version of this was tested in the present work in which trajectory counts per cell rather than point counts within a cell were used. This approach better recognizes the importance of trajectory statistics in ensuring coverage from multiple independent trajectories. This trajectory weighting function is multiplied by the local PSCF, CWT, or RTWC value.

Zhou et al. (2004) introduced a similar idea of an empirical weighting function to reduce noise in the QTBA algorithm. Zhao et al. (2007) found that a simpler weighting function performed just as well:

$$\text{weight} = \begin{cases} 
0.2 & p < 0.5\bar{p} \\
0.5 & 0.5\bar{p} \leq p < \bar{p} \\
0.75 & \bar{p} \leq p < 2\bar{p} \\
1.0 & p \geq 2\bar{p} 
\end{cases}$$

where $p$ is the QTBA value of the cell and $\bar{p}$ is the average of all cells. This form was considered in the current work.
3.3 PSCF threshold concentration

The PSCF polluted concentration threshold is a subjective parameter that must be set in accordance with the data being acquired. In many cases, percentiles and average cut-offs have been used (Lin et al., 2001; Cheng & Lin, 2001; Polissar & Hopke, 2001; Hsu et al. 2003). In the proposed fugitive emissions detection scheme, it is expected that some receptors would always measure near-zero concentration values. Since a threshold based on individual sensor statistics would lead to some of these low readings being misinterpreted as polluted, in the present case it was more appropriate to define a PSCF threshold value based on the pooled data set from all sensors. In accordance with Lin et al. (2001), who suggested that the 90th percentile performed better than the 50th and 75th percentiles used in previous studies, cut-off values corresponding to the 75th, 90th and 97.5th percentile concentration measurements of the pooled data set were specified.

3.4 RTWC segment length

Since RTWC is characterized by the redistribution of concentrations along trajectories, the trajectories must somehow be divided into segments for reweighting. Due to the difference in scale between the current study and previous continental-scale studies, it was not possible to derive guidelines in advance. Instead, a range of trajectory segment sizes were considered, calculated either in terms of propagation time (intervals of 15, 30 and 60 s) or physical length along the trajectory (as multiples of 2, 4, or 6 times cell size used for the redistributed concentration field).

3.5 QTBA standard deviation coefficient

In QTBA, the standard deviation for the trajectory position uncertainty, \( a(t')=at' \), is assumed to increase linearly with time upwind. The coefficient \( a \) defines the spread of the trajectory. On a continental-scale, \( a \) is typically assumed to be 5.4 km/h or 1.5 m/s (Zhou et al., 2004). For building-scale problems, where the largest scales of turbulence are reduced, it is expected that this value should be lower. Thus, coefficient values of 0.05, 0.1, 0.25, 0.5 and 1.5 m/s were investigated.

4. Results and Discussion

4.1 Results evaluation

Selected example results field contour plots, for each of the four algorithms, are shown in Figure 2, for test conditions detailed in Table 3. In each plot, the true source location is indicated as the center of the superimposed white crosshairs. From visual observation alone, it is easy to see that the predicted source location aligns well with the true source location in most cases, and that adjacent regions of high pseudo-concentration values are relatively contained.
Figure 2: Select algorithm field plots corresponding to Table 3
To permit more quantitative comparison of the various algorithms, an objective measure of the accuracy of any given source prediction was required. The goal of any fugitive emissions detection approach is to consistently detect the presence of major sources while minimizing the required search area to locate a specific emitter. Since search area scales with the radius squared, a reference field was created for comparison that was centered about the actual source location with a decay proportional to $1/radius^2$. The Pearson correlation was calculated between each TSM algorithm result and this reference field. For cases with two sources, the reference fields from each source were scaled by their relative source strengths and summed. A minimum radius of 1 m was specified in the reference field to avoid singularities. Calculated in this way, the Pearson correlation rewards less-diffuse predicted sources that are close to the correct location, while simultaneously penalizing any false sources located farther away.

To reduce computational overhead, the correlation was confined to a 600 m x 600 m region surrounding the buildings. Results for coarse grid cases were interpolated onto the fine grid to take into account spreading due to grid resolution. Regions within this domain without known values due to poor trajectory coverage, used the average value of known cells to maintain a constant correlation vector length for comparison between parameter settings and algorithms. Table 3 lists correlation values for each of the twelve plots shown in Figure 2.

### 4.1.1 PSCF

For single sources, such as that in Figure 2a, the PSCF algorithm was able to correctly identify the source location, but regions surrounding the source were also flagged as potential sources. Trails of higher probability upstream and downstream of the true source location were especially apparent in regions where trajectory coverage was from a single primary direction. For the two source cases, regions of similar sized sources were both distinguished (Figure 2b). When source strengths were mixed, larger (more important) sources were easily identified, but they tended to obscure the smaller sources. In regions with poor trajectory coverage, false source regions could also occur. This was especially obvious for Source 3 (Figure 2b), where in addition to the actual source area, the region North of the source was falsely identified. This error was directly attributable to unidirectional coverage with trajectories also traversing the actual source farther downstream.

### Table 3: Correlation values for select algorithm results

<table>
<thead>
<tr>
<th>Label</th>
<th>Algorithm</th>
<th>Source(s)</th>
<th>Data set</th>
<th>Filter</th>
<th>Algorithm Settings</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>PSCF</td>
<td>1</td>
<td>full</td>
<td>None</td>
<td>97.5th percentile cutoff</td>
<td>0.19</td>
</tr>
<tr>
<td>b</td>
<td>PSCF</td>
<td>1 &amp; 3</td>
<td>full</td>
<td>None</td>
<td>97.5th percentile cutoff</td>
<td>0.10</td>
</tr>
<tr>
<td>c</td>
<td>PSCF</td>
<td>1 &amp; 3</td>
<td>full</td>
<td>Four Traj.</td>
<td>97.5th percentile cutoff</td>
<td>0.12</td>
</tr>
<tr>
<td>d</td>
<td>CWT</td>
<td>1</td>
<td>full</td>
<td>None</td>
<td>-</td>
<td>0.26</td>
</tr>
<tr>
<td>e</td>
<td>CWT</td>
<td>1 &amp; 3</td>
<td>90 s intervals</td>
<td>None</td>
<td>-</td>
<td>0.10</td>
</tr>
<tr>
<td>f</td>
<td>CWT</td>
<td>1 &amp; 3</td>
<td>90 s intervals</td>
<td>Four Traj.</td>
<td>-</td>
<td>0.06</td>
</tr>
<tr>
<td>g</td>
<td>RTWC</td>
<td>1</td>
<td>full</td>
<td>None</td>
<td>20 m segment length</td>
<td>0.43</td>
</tr>
<tr>
<td>h</td>
<td>RTWC</td>
<td>2</td>
<td>full</td>
<td>None</td>
<td>10 m segment length</td>
<td>0.25</td>
</tr>
<tr>
<td>i</td>
<td>QTBA</td>
<td>1</td>
<td>full</td>
<td>None</td>
<td>$a = 0.25$ m/s</td>
<td>0.18</td>
</tr>
<tr>
<td>j</td>
<td>QTBA</td>
<td>2</td>
<td>full</td>
<td>None</td>
<td>$a = 0.25$ m/s</td>
<td>0.06</td>
</tr>
<tr>
<td>k</td>
<td>QTBA</td>
<td>3</td>
<td>full</td>
<td>None</td>
<td>$a = 1.5$ m/s</td>
<td>0.03</td>
</tr>
<tr>
<td>l</td>
<td>QTBA</td>
<td>3</td>
<td>full</td>
<td>None</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Results were not sensitive to the specified PSCF threshold (which was fixed at the 97.5th percentile for data shown in the paper), largely because the prescribed detection limit of the sensors functioned as a practical cut-off. In general, smaller cell sizes performed better, provided there were adequate numbers of trajectories passing through the cells, leading to improved spatial resolution of the prediction.

Multiple-directional coverage was found to provide more consistent results for all considered sources (see Figure 3). While some directions were particularly advantageous for a particular source (e.g. if there were multiple sensors up and downwind of a source as for Direction 2 for Source 1), poor results were also noted for insufficient coverage (e.g. Direction 1 for Sources 2 and 3).

![Figure 3: PSCF correlation for various data sets using 97.5th cut-off percentile, 5 m cell size and no filter. Coverage from multiple directions gives results that are more consistent.](image)

The use of filters can improve correlations by reducing the influence of cells with few trajectories. Four different filters were investigated: four-point, four-trajectory, point weighting and trajectory weighting. With the full data set, the filters had little effect for Sources 1, 2 and 4, and modest improvement for Source 3. With a reduced trajectory data set, the benefits and potential drawbacks of using filters were clearer, as shown in Figure 4. In this case, the correlation for Source 3 could actually be worsened with the four trajectory filter, even though this filter showed the most improvement for several of the other source cases. While each filter reduced the influence of the false source regions at the edge of the domain due to poor trajectory coverage (e.g. North of Source 3 noted previously in Figure 2b), if overall trajectory coverage was insufficient actual source regions within the domain could also be incorrectly removed (as shown for CWT in Figure 2f).

Weighting filters were much less sensitive to this problem. Thus, filters are useful in removing spurious sources at the edges of the domain, not improving a bad solution. If it is not feasible to improve the overall trajectory statistics, a trajectory-weighting filter is recommended as it improved or had negligible effect in all test cases (for all algorithms).
4.1.2 CWT

Figure 2d–f show three example CWT plots. Like PSCF, CWT can identify individual source regions (e.g. Figure 2d) and multiple sources (e.g. Figure 2e), though large differences in source strength still lead to small sources being overshadowed by larger sources. CWT behaved similarly to PSCF for the various data sets and cell sizes. The four-trajectory filter was tested on CWT with the same behavior demonstrated for PSCF (compare Figure 2c and Figure 2f). However, while the parameter trends were the same as PSCF, CWT had higher correlations overall.

4.1.3 RTWC

Figure 2g–i show sample RTWC fields for three cases. Compared with CWT, RTWC was capable of better spatial resolution, as apparent comparing Figure 2d with Figure 2g. The RTWC fields distinguished the larger sources when two sources were released. Smaller reweighting lengths (15-30s and 2 times cell size) were better for the complete data set in most cases (Figure 5). However, false sources could arise in regions of poor trajectory coverage that obscured the actual source locations (e.g. Source 2 in Figure 2h and Figure 5). If this spurious source is removed (e.g. with the trajectory filter as in Figure 2i), the actual source region becomes apparent.

Increasing the reweighting length removed this spurious source, as did increasing cell size at the expense of reducing spatial resolution. Smaller data sets emphasized spurious sources, as shown by the drop in correlation value with many of the smallest segments lengths noted in Figure 6.

Since false sources can be predicted in areas with few trajectories, RTWC is best used in conjunction with CWT to ensure RTWC reliability by comparing it with the non-reweighted CWT algorithm results (as noted by Zhou et al., 2004 in a continental scale application). Additionally, prior to running CWT and RTWC, multiple trajectory coverage in regions of interest should be checked and, if lacking, sources predicted in these regions should be excluded until favorable wind conditions produce desired coverage.
Figure 5: RTWC correlation vs. segment length for 5 m cell size and no filter. Poor trajectory statistics cause spurious source predictions for small segment lengths.

Figure 6: RTWC correlation vs. segment length for 5 m cell size and no filter using every third measurement time. Poor trajectory statistics cause spurious source predictions for small segment lengths are more obvious.

4.1.4 QTBA

Figure 2j–l show example QTBA plots. Some predictions matched expected source locations (Figure 2j), but this was not always the case (Figure 2k). The standard deviation coefficient “a” significantly influenced predictions. Figure 2k compared with Figure 2l shows how changing the coefficient value shifts the source prediction location. Overall, 0.25 m/s was best of the coefficients tested (Figure 7), which is an order of magnitude smaller than the 1.5 m/s value used on a continental scale (Zhou et al., 2004). Filtering and cell sizes had only modest effects, where the latter was expected given that QTBA is calculated at cell centers. Data sets from multiple directions and a large number of sensors provided more consistently high correlations for Sources 1, 2 and 4. However, Source 3 always had low correlation values.

Figure 7: QTBA correlation coefficient vs. various deviation coefficient values on 5 m cell size and no filter. A deviation coefficient of 0.25 m/s performs best for the full data set.

4.2 TSM Performance Comparison

Based on the results presented in the previous section, optimal parameters were selected for each algorithm as specified in Table 4, prior to overall performance comparison. Comparison results using the full data set are shown in Figure 8 and using reduced numbers of trajectories in Figure 9.
Table 4: Algorithm parameter settings

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Cell size [m]</th>
<th>Filter</th>
<th>Algorithm specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSCF</td>
<td>5</td>
<td>trajectory weighted</td>
<td>cut-off percentile: 97.5th</td>
</tr>
<tr>
<td>CWT</td>
<td>5</td>
<td>none</td>
<td>N/A</td>
</tr>
<tr>
<td>RTWC</td>
<td>5</td>
<td>none</td>
<td>redistribution length: 20 m</td>
</tr>
<tr>
<td>QTBA</td>
<td>5</td>
<td>none</td>
<td>deviation coefficient: 0.25 m/s</td>
</tr>
</tbody>
</table>

In every case for the full data set (Figure 8), RTWC provides a higher correlation value than the other algorithms. RTWC narrowed source regions and improved results of CWT correlations for most cases. However, for correlations that had regions with poor trajectory statistics (Figure 9), RTWC could falsely identify spurious source regions, being unable to correct poor input CWT predictions.

CWT provides the second best results. As CWT must be calculated before running RTWC, it could be used in addition to RTWC with no additional computations for cases where RTWC does not perform well due to poor trajectory statistics. QTBA performed the worst for every source location tested.

4.3 Computational time

All algorithms ran within a reasonable amount of time considering the target application (i.e. in under a day). Table 5 presents the run time on a single core of a standard desktop computer for each algorithm for the cases in Table 4 of Source 1 using the entire data set as a representative comparison. Both CWT and PSCF were very fast, significantly faster than the RTWC and QTBA calculations, although these were still accomplished within a few hours.

Table 5: Algorithm run times for the full Source 1 data set

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Run Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSCF</td>
<td>1 s</td>
</tr>
<tr>
<td>CWT</td>
<td>1 s</td>
</tr>
<tr>
<td>RTWC</td>
<td>44 min</td>
</tr>
<tr>
<td>QTBA</td>
<td>3.5 h</td>
</tr>
</tbody>
</table>

4.4 Effect of noise

Previous results considered the simulated concentration measurements as exact values. In reality, measurement noise is always present. To examine algorithm sensitivity to measurement
error, CWT and RTWC cases were re-evaluated with gaussian noise added to the input concentration values, with standard deviations of 0.5, 1.0, 1.5, and 2.0 ppm, corresponding to a range readily achievable by modern detectors.

Figure 10 shows the correlation value of the CWT algorithm on various levels of noise. For the large sources, only the 2 ppm case had a significant effect on the correlation value and, even then, the predicted source location was merely spread out and could still be identified. The effect was much greater on the smaller sources as the noise levels became the same order of magnitude as the measurements.

Figure 11 plots the RTWC results of the same cases. The large sources can still be found even at the highest noise level, however RTWC was in general more sensitive to noise levels than CWT as the reweighting emphasized the false sources that appeared in the CWT field.

5. Conclusions

TSMs can provide useful information about the location of pollutant sources within a building-scale domain. For areas with sufficient trajectory coverage, RTWC, CWT and PSCF can all find large leak sources. QTBA only worked well in some cases, even when trajectory coverage was good. In general, the determining factor in algorithm performance was trajectory statistics.

Fortunately, these can be evaluated prior to running the TSM algorithms to determine whether more measurements (i.e. trajectories) must be acquired to provide sufficient coverage of pertinent areas of the domain. While multidirectional information is desirable in all areas to reduce upstream and downstream smearing of source predictions, even smeared predictions can significantly reduce search space.

Of the algorithms tested, RTWC produced the best correlation results given good trajectory coverage. In cases of poor trajectory coverage, however, it could result in false source regions.
This could be overcome when used in conjunction with CWT and trajectory statistics plots. This same strategy proved effective at reducing susceptibility to measurement noise, since CWT proved more robust than RTWC.

Given the importance of trajectory coverage, sensor placement is key to locating sources successfully. Typical wind directions within a plant and regions likely to contain leaks should be considered when implementing the proposed system. In addition, weather may necessitate waiting a several days to acquire sufficient multidirectional wind data.

Overall, the proof of concept simulations presented demonstrate that TSMs have significant potential for advancing the complex problem of fugitive emissions detection within industrial facilities.

6. Acknowledgements
This project was supported by Natural Resources Canada (Project Manager Michael Layer), the Canadian Association of Petroleum Producers (CAPP), and the Natural Sciences and Engineering Research Council (NSERC) of Canada. We are grateful for the assistance of Claire Serdula and Brian Ross of Nexen Inc. in providing representative gas plant geometry data for the test simulations and for the assistance of Ian Joynes and Michael Shives in setting up the simulations.

7. References


