

**Investigation of Trajectory Statistical Methods for Locating Fugitive Emissions  
Sources on a Building Scale**

by

**Carol Anne Brereton**

A thesis submitted to The Faculty of Graduate Studies and Research

in partial fulfilment of the degree requirements of

**Master of Applied Science in Mechanical Engineering**

Ottawa-Carleton Institute for

Mechanical and Aerospace Engineering

Department of Mechanical and Aerospace Engineering

Carleton University

Ottawa, Ontario, Canada

September 2010

Copyright © 2010 – Carol Anne Brereton



Library and Archives  
Canada

Published Heritage  
Branch

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

Bibliothèque et  
Archives Canada

Direction du  
Patrimoine de l'édition

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file* *Votre référence*  
ISBN: 978-0-494-71528-4  
*Our file* *Notre référence*  
ISBN: 978-0-494-71528-4

**NOTICE:**

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

**AVIS:**

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

---

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

  
**Canada**

## **Abstract**

Fugitive emissions in the oil and gas industry are significant greenhouse gas contributors but their location is challenging. A proof-of-concept investigation was performed, combining wind and concentration data to determine source locations. Four simple trajectory statistical methods were investigated on an ideal test case from a CFD simulation of a simplified gas plant geometry: potential source contribution function (PSCF), concentration weighted trajectory (CWT), residence time weighted concentration (RTWC), and quantitative transport bias analysis (QTBA). Results were quantitatively compared using a correlation measure that took into account search area from the source(s). In the present application, PSCF, CWT and RTWC were all capable of distinguishing regions near major sources from the surroundings. QTBA was not. RTWC had the best source discrimination when provided sufficient domain trajectory coverage, but otherwise could produce false source regions, and should be used in conjunction with CWT and trajectory maps.

## **Acknowledgements**

I wish to offer my most sincere thanks to my supervisor, Dr. Matthew Johnson, who pushed me when I needed it, offering support and guidance with much patience and made this thesis much better than it started out. I am grateful for the hard work done by Michael Shives and Ian Joynes in aiding me with the CFD setup and the quick trajectory calculation coding that they accomplished. Funding from Carleton University and NSERC was also much appreciated. Finally, to my family and friends, without your support I would not be where I am today. Thank you.

## **Table of Contents**

Abstract .....	iii
Acknowledgements .....	iv
List of Tables .....	x
List of Figures .....	xii
Nomenclature .....	xxii
1 Introduction .....	1
1.1 Motivation .....	1
1.2 Current Industry Fugitive Emissions Location and Estimation .....	3
1.2.1 Emission Factor Estimates .....	3
1.2.2 Manual leak detection and measurement .....	4
1.2.3 Typical Leak Detection Methods .....	6
1.2.4 Leak Quantification Measurements .....	9
1.2.5 Remote and Downstream Detection and Measurements .....	11
1.3 Proposed Inverse Pollutant Locating Methodology .....	13
1.3.1 Thesis Overview .....	14
2 Background and Inverse Solution Methods .....	16
2.1 Forward Transport Modelling .....	16
2.1.1 Full Navier-Stokes Equations .....	17

2.1.2	Advection-Diffusion Equation.....	17
2.1.3	Box Models.....	18
2.1.4	Gaussian Plume Dispersion Model.....	20
2.1.5	Gaussian Puff Models.....	21
2.1.6	Particle Tracking.....	22
2.1.7	Chemistry Transport Models.....	24
2.1.8	Summary of Forward Transport Models.....	25
2.2	Inverse Transport Models.....	25
2.2.1	What is an adjoint?.....	26
2.2.2	Automatic Calculation.....	27
2.2.3	Advection-diffusion equation adjoint.....	28
2.2.4	Lagrangian Backtracking.....	30
2.2.5	Summary of Inverse Pollutant Transport Models.....	31
2.3	Inverse Problems.....	31
2.4	Solution Optimization Methods.....	33
2.4.1	Bayesian Inference.....	33
2.4.2	Kalman Filtering.....	36
2.4.3	Genetic Algorithms.....	38
2.4.4	Simulated Annealing.....	40
2.4.5	Variational Data Assimilation.....	41

2.5	Trajectory Statistical Methods .....	44
2.5.1	Potential Source Contribution Function.....	44
2.5.2	Concentration Weighted Trajectory and Residence Time Weighted Concentration.....	47
2.5.3	Quantitative Transport Bias Analysis .....	50
2.6	Discussion and Method Selection .....	52
3	Review of Trajectory Statistical Methods.....	55
3.1	Review of TSM development and validation studies.....	55
3.2	Summary and General Observations from Previous TSM Studies.....	65
4	CFD Simulation .....	70
4.1	General Approach .....	70
4.2	Simulation Geometry .....	71
4.3	CFX Model Implementation .....	72
4.3.1	Simulation Model Type .....	72
4.3.2	Turbulence Model Selection.....	77
4.3.3	Initial Simplifications.....	80
4.3.4	Forward Pollutant Release .....	85
4.3.5	Domain Size.....	91
4.3.6	Mesh.....	92
4.3.7	Boundary Conditions .....	98

4.3.8	Simulation Type.....	100
4.3.9	Initial Condition Generation .....	100
4.3.10	Mesh Discretization and Timestep Selection.....	101
4.3.11	CFD Setup Summary .....	111
5	Inverse Pollutant Tracking and Results .....	113
5.1	Receptors.....	113
5.2	Trajectory Calculation.....	113
5.2.1	Turbulent Dispersion .....	116
5.2.2	Concentration Limits .....	118
5.3	Parameters affecting source location algorithms .....	119
5.3.1	Parameters universally affecting source location algorithms .....	119
5.3.2	Algorithm Specific Parameters .....	123
5.4	Result Field Interpretation.....	126
5.5	Results Evaluation Methods.....	128
5.6	Analysis of Results.....	131
5.6.1	Potential Source Contribution Function.....	131
5.6.2	Concentration Weighted Trajectories .....	152
5.6.3	Residence Weighted Trajectory Concentration .....	157
5.6.4	Quantitative Transport Bias Analysis .....	165
5.7	Overall algorithm comparison.....	171

5.7.1	Correlation .....	171
5.7.2	Run Times .....	173
5.7.3	Measurement noise .....	174
5.7.4	Overall comparison .....	178
6	Conclusions and Recommendations .....	180
6.1	Conclusions .....	180
6.2	Field set-up recommendations .....	183
6.2.1	Placement of Sensors .....	183
6.2.2	Interpretation of Results.....	184
6.3	Recommendations for future work.....	185
	References.....	187
	Appendix A: Richardson Extrapolation and Convergence.....	204
	Appendix B: Correlation Measures.....	210

## List of Tables

Table 3.1: Some select Trajectory Statistical Method evaluation studies .....	68
Table 4.1: Source mass flow rates .....	86
Table 4.2: Mesh Setting Descriptions.....	95
Table 4.3: Mesh Expansion.....	95
Table 4.4: Balzac Mesh Settings.....	97
Table 4.5: Inlet conditions at 10 m height using linear variation over duration.....	99
Table 4.6: Richardson extrapolation error estimates for 1 min of simulated time at 5 km/h from 10° direction normalized by the value range.....	103
Table 4.7: Richardson extrapolation error estimates for 1 min of simulated time at 6 km/h from 100° direction normalized by the value range.....	104
Table 4.8: Difference between all three grids for 1 min of simulated time at 5 km/h from 10° direction.....	109
Table 4.9: Difference between all three grids for 1 min of simulated time at 6 km/h from 100° direction.....	110
Table 4.10: Difference estimates for 1 min of simulated time at 5 km/h from 10° direction normalized by the value range .....	111
Table 4.11: CFD summary.....	112
Table 5.1: Source combinations.....	122

Table 5.2: Trajectory filters considered in the present work .....	126
Table 5.3: Algorithm parameters .....	171
Table 5.4: Algorithm runtime .....	173

## List of Figures

Figure 1.1: Bagging Techniques (Picard et al., 1998) .....	10
Figure 2.1: Box Model (adapted from Heinsohn & Kabel, 1999) .....	18
Figure 2.2: A plume from stack with virtual reflection where the effect of the ground is considered using a second virtual plume. This requires an additional term in the standard equation. (Wark & Warner, 1981) .....	20
Figure 2.3: Movement of puffs A, B and C (Jung et al., 2003) .....	22
Figure 2.4: Interaction considerations for a CTM (Peters et al., 1995) .....	24
Figure 2.5: Tracer occupying domain $S$ at time $t_S$ moves partially to domain $D$ at time $t_D$ (dark grey). Looking backward in time, the air in $D$ at $t_D$ partially comes from domain $S$ at time $t_S$ . (Hourdin & Talagrand, 2006) .....	29
Figure 2.6: Sample trajectories of high (red) and low (blue) concentrations arrive at a single sensor .....	46
Figure 2.7: Sample initial field generated by CWT .....	49
Figure 2.8: Sample reweighting of initial CWT field by RTWC .....	49
Figure 4.1: Simulation boundaries containing the simplified gas plant geometry .....	72
Figure 4.2: Methods of numerical simulation and turbulence models (Murakami et al., 1994) .....	74
Figure 4.3: Source locations .....	86

Figure 4.4: Predicted buoyant plume rise for 4.5 g/s methane into air at 20°C and 1 atm for various wind speeds .....	90
Figure 4.5: Coarse Simulation Mesh .....	96
Figure 4.6: Points where Richardson estimation of speed available (blue) and unavailable (red) at 2 m height corresponding to 5 km/h wind from 10° .....	105
Figure 4.7: Points where Richardson estimation of speed available (blue) and unavailable (red) at 10 m height corresponding to 5 km/h wind from 10° .....	105
Figure 4.8: Random Sample Points .....	105
Figure 4.9: Speed at three points versus number of grid elements .....	106
Figure 4.10: Absolute difference in speed between coarse and fine mesh at 2 m height corresponding to 5 km/h wind from 10° .....	107
Figure 4.11: Absolute difference in speed between coarse and fine mesh at 10 m height corresponding to 5 km/h wind from 10° .....	107
Figure 4.12: Isosurface 0.25 m/s difference between fine and coarse grids coloured by fine grid average speed .....	108
Figure 5.1: Receptor locations .....	113
Figure 5.2: Trajectories from receptor 20 for wind quadrant 1 .....	114
Figure 5.3: Maximum difference of 0.1 s timestep trajectories compared with timestep of 0.001 s .....	116
Figure 5.4: Side view of domain with trajectories for full case.....	120
Figure 5.5: Trajectory locations for all three wind quadrants (full case).....	121

Figure 5.6: Number of trajectories passing through each cell, 5 m cell size ..... 121

Figure 5.7: RTWC results for Source 1 calculated with 10 m segment lengths (2x cell size) and 5 m cell size. The actual source location is indicated by the white circle. The source is identified within 5-10 m. (Pearson correlation = 0.43, explained in detail in Section 5.5) ..... 127

Figure 5.8: Comparative correlation map for Source 1 based on the actual location of the released gas ..... 130

Figure 5.9: PSCF Source 2, 97.5th percentile, 5 m cell size (Pearson correlation = 0.27)  
The region surrounding the source is clearly visible, though it is smeared into the surrounding area..... 133

Figure 5.10: PSCF Source 2, 97.5th percentile, 10 m cell size (Pearson correlation = 0.24)  
Highest value is to the northwest of the source at a greater distance than in the 5 m prediction. The second highest cell corresponds to the actual location. .... 133

Figure 5.11: PSCF Source 2, 97.5th percentile, 20 m cell size (Pearson correlation = 0.25)  
The prediction at the actual source location corresponds to the second highest predicted cell. The distance from the highest cell to the source is less than that found with the 10 m cell size..... 134

Figure 5.12: PSCF correlation vs. cell size for a single source using 97.5th cut-off percentile and no filter ..... 135

Figure 5.13: PSCF Source 1 and Source 3, 97.5th percentile, 5 m cell size (Pearson correlation = 0.10) Both "large" sources are visible, though the highest source prediction region is to the North of Source 3..... 136

Figure 5.14: PSCF Source 1 and Source 2, 97.5th percentile, 5 m cell size (Pearson correlation = 0.20) "Large" Source 1 is within the highest predicted region, but "medium" Source 2 can no longer be distinguished..... 137

Figure 5.15: PSCF Source 1 and Source 4 , 97.5th percentile, 5 m cell size (Pearson correlation = 0.19) "Large" Source 1 is within the highest predicted region, but "medium" Source 4 can no longer be distinguished..... 137

Figure 5.16: PSCF Source 2 and Source 3, 97.5th percentile, 5 m cell size (Pearson correlation = 0.09) "Large" Source 3 is within the secondary high of the prediction smear. "Medium" Source 2 cannot be distinguished from the smear's tail. .... 138

Figure 5.17: PSCF Source 3 and Source 4, 97.5th percentile, 5 m cell size (Pearson correlation = 0.09) "Large" Source 3 is within the secondary high of the prediction smear. "Medium" Source 2 cannot be distinguished from the smear's tail. .... 138

Figure 5.18: PSCF correlation vs. cell size for two sources using 97.5th cut-off percentile and no filter. Correlation values are dominated by the algorithm prediction of the larger source. .... 139

Figure 5.19: PSCF zeroed data set correlation vs. cut-off percentile for single source using 5 m cell size and no filter. Due to few trajectories above the detection threshold, there is no or little difference between the correlations at the 75<sup>th</sup> and 90<sup>th</sup> percentiles. 142

Figure 5.20: PSCF zeroed data set correlation vs. cut-off percentile for two sources using 5 m cell size and no filter. Due to few trajectories above the detection threshold, there is no or little difference between the correlations at the 75<sup>th</sup> and 90<sup>th</sup> percentiles..... 142

Figure 5.21: PSCF non-zeroed data set correlation vs. cut-off percentile for single source cases using 5 m cell size and no filter. With no detection threshold, increasing the cut-off percentile improves the source prediction. Thus, the optimum threshold for these cases lies somewhere above the 90<sup>th</sup> percentile. .... 143

Figure 5.22: PSCF non-zeroed data set correlation vs. cut-off percentile for cases with two sources using 5 m cell size and no filter. With no detection threshold, increasing the cut-off percentile improves the source prediction. Thus, the optimum threshold for these cases lies somewhere above the 90<sup>th</sup> percentile. .... 143

Figure 5.23: PSCF correlation with various filters for single source using 97.5th cut-off percentile and 5 m cell size. Benefits of the filters are negligible for Source 1 and Source 2. There is a small improvement for Source 4. Source 3 had significant differences between the results of various filters. .... 144

Figure 5.24: PSCF Source 3, 97.5th percentile, 5 m cell size with no filter. (Pearson correlation = 0.10) An additional high area exists to the North-East of the source region. .... 145

Figure 5.25: PSCF Source 3, 97.5th percentile, 5 m cell size with four-trajectory filter. (Pearson correlation = 0.13) The four-trajectory filter removes a large portion of the additional high region which existed to the North-East of the source. .... 145

Figure 5.26: PSCF correlation with various filters for using every third measurement interval and 97.5th cut-off percentile and 5 m cell size. Reducing the data set increases the effect of filters for all sources. The four-trajectory filter greatly reduces the correlation for all cases involving Source 3. .... 146

Figure 5.27: PSCF Source 3, 97.5th percentile, 5 m cell size from every 3rd measurement interval using four-trajectory filter. (Pearson correlation = 0.0) The four-trajectory filter has removed all cells in the vicinity of Source 3. The source can no longer be located. .... 147

Figure 5.28: PSCF correlation for various data sets using 97.5th cut-off percentile, 5 m cell size and no filter. Coverage from multiple directions gives more consistent results. .... 149

Figure 5.29: PSCF Source 3, 97.5th percentile, 5 m cell size using every third sensor. (Pearson correlation = 0.0) There are no trajectories passing over Source 3 so it cannot be found, though upstream and downstream smears are still visible. .... 151

Figure 5.30: CWT correlation vs. cell size for a single source using no filter ..... 153

Figure 5.31: CWT correlation vs. cell size for two sources using no filter ..... 153

Figure 5.32: CWT Source 1 and Source 3, 5 m cell size. (Pearson correlation = 0.11) The highest predicted region is located at Source 1. A high smear region exists around Source 3, but the source itself cannot be distinguished from the rest of the smear. .... 154

Figure 5.33: CWT Source 1 and Source 4, 5 m cell size. (Pearson correlation = 0.26) The highest predicted region is located at "large" Source 1. "Medium" Source 4 cannot be distinguished. .... 154

Figure 5.34: CWT correlation for various filters and sources using 5 m cell size. The four-trajectory filter has negligible effect for cases other than those containing Source 3. .... 155

Figure 5.35: CWT correlation for various filters and sources using 5 m cell size and every 3<sup>rd</sup> measurement interval. The four-trajectory filter removes all information in the region around Source 3 so the correlation values are low. .... 155

Figure 5.36: CWT correlation for various data sets using 5 m cell size and no filter. Coverage from multiple directions gives more consistent results. .... 156

Figure 5.37: RTWC correlation vs. segment length for cell size 5 m and no filter. Poor trajectory statistics cause a spurious source prediction for Source 2 when the segment length is small. .... 158

Figure 5.38: RTWC Source 2, 5 m cell size, 2x cell size segment length, iteration 1. High smear is visible at Source 2..... 159

Figure 5.39: RTWC Source 2, 5 m cell size, 2x cell size segment length, iteration 10. Highest point is in domain center (arrow). Source 2 region is now difficult to distinguish. .... 159

Figure 5.40: RTWC Source 2, 5 m cell size, 2x cell size segment length, iteration 10 using four-trajectory filter. Removing spurious source with the four-trajectory filter reveals a secondary high smear at the actual source. .... 160

Figure 5.41: RTWC Source 2, 5 m cell size, 4x cell size segment length, iteration 10. Increasing the weighting length removes the spurious high prediction found with the 2x cell size segment length. .... 161

Figure 5.42: RTWC correlation vs. segment length for cell size 10 m and no filter ..... 162

Figure 5.43: RTWC correlation vs. reweighting length for cell size 5 m, no filter and every 3<sup>rd</sup> measurement interval. Spurious sources appear in regions of poor trajectory statistics for Source 1, Source 2 and Source 4 using the smallest segment length. .... 163

Figure 5.44: RTWC correlation for various data sets for cell size 5 m and reweighting length of 2x cell size. Multiple wind directions provide more consistent results as does increased number of sensors. .... 165

Figure 5.45: QTBA correlation vs. cell size for a single source using 1.5 m/s deviation coefficient and no filter. Cell size has no significant effect..... 166

Figure 5.46: QTBA correlation vs. cell size for two sources using 1.5 m/s deviation coefficient and no filter. Cell size has no significant effect..... 166

Figure 5.47: QTBA Source 1, 1.5 m/s deviation coefficient, 5 m cell size (Pearson correlation = 0.09)..... 167

Figure 5.48: QTBA Source 1, 1.5 m/s deviation coefficient, 20 m cell size. (Pearson correlation = 0.10) Center of high prediction region does not change from 5 m cell size. .... 167

Figure 5.49: QTBA correlation vs. deviation coefficient for a single source using 5 m cell size ..... 168

Figure 5.50: QTBA correlation vs. deviation coefficient for two sources using 5 m cell size ..... 168

Figure 5.51: QTBA Source 3, 1.125 m/s deviation coefficient, 5 m cell size (Pearson correlation = 0.03)..... 169

Figure 5.52: QTBA Source 3, 1.875 m/s deviation coefficient, 5 m cell size. (Pearson correlation = 0.02) Increasing the deviation coefficient moves the predicted high region inward. .... 169

Figure 5.53: QTBA correlation for various filters for a single source using 5 m cell size and 1.5 m/s deviation coefficient. The weighting filter had no significant effect. .... 170

Figure 5.54: QTBA correlation for various data sets using a 1.5 m/s deviation coefficient, 5 m cell size and no filter. Multiple wind directions gave more consistent results. .... 171

Figure 5.55: Comparison of algorithms using optimum parameters for calculations on the complete data set. RTWC performs better in all cases when there are good trajectory statistics. QTBA did not perform well when scaled down to a building scale problem. 172

Figure 5.56: CWT correlation for various noise levels, 5 m cell size. Source regions can still be identified for noise levels up to 2 ppm for the large sources. For the medium sources, once the noise reached 1 ppm levels, the true values were of the same scale as the noise and false source regions began to appear. At a noise level of 2 ppm, these false sources became common and the actual source region became difficult to distinguish. 175

Figure 5.57: CWT Source 1, 5 m cell size, 2 ppm<sub>v</sub> noise level. (Pearson correlation = 0.24) Source 1 is still clearly visible. .... 176

Figure 5.58: CWT Source 4, 5 m cell size, 2 ppm<sub>v</sub> noise level. (Pearson correlation = 0.12) Noise related highs are the same intensity as the region at Source 4. .... 176

Figure 5.59: RTWC correlation for various noise levels, 5 m cell size, 4x cell size segment length. Source regions can still be identified for noise levels up to 2 ppm for the

large sources. For the medium sources, once the noise reached 1 ppm levels, the results became sensitive to the false sources which appeared in the CWT results. .... 177

Figure 5.60: RTWC Source 1, 5 m cell size, 10 m segment length, 2 ppm<sub>v</sub> noise level (Pearson correlation = 0.27)..... 178

Figure 5.61: RTWC Source 4, 5 m cell size, 10 m segment length, 2 ppm<sub>v</sub> noise level (Pearson correlation = 0.07)..... 178

## Nomenclature

<i>Symbol</i>	<i>Description</i>	<i>Defined in</i>	<i>First usage</i>
$A$	Gas source [kg/s]		Equation 2.17
$A$	Tracer arrival probability		Equation 2.35
$\mathbf{a}$	Input vector		Equation 2.8
$a$	Number of segment points in cell		Equation 2.30
$a$	Upwind standard deviation constant [m/s]		Equation 2.36
$\mathbf{a}'$	Perturbed vector		Equation 2.9
$a_k$	Input vector at position $k$		Equation 2.9
$a_k'$	Perturbed input vector at position $k$		Equation 2.9
$B$	Vector operator		Equation 2.8
$B$	Gas removal [kg/s]		Equation 2.18
$\mathbf{B}$	Covariance matrix for background error		Equation 2.24
$\mathbf{b}$	Output vector		Equation 2.8
$b$	Number of polluted segment points in cell		Equation 2.30
$b_j$	Output vector at position $j$		Equation 2.9
$b_j'$	Perturbed output vector at position $j$		Equation 2.9
$C$	Concentration [kg/m <sup>3</sup> ]		Equation 1.1

$C^*$	Concentration in Volume $D$ [ $\text{kg}/\text{m}^3$ ]		Equation 2.14
$C_a$	Concentration ahead of box [ $\text{kg}/\text{m}^3$ ]		Equation 2.2
$C_{background}$	Background concentration fraction		Equation 1.2
$C_{ij}$	Variable proximity measure	Equation B.3	Equation B.1
$C_m$	Concentration corresponding to trajectory $m$ [ $\text{kg}/\text{m}^3$ ]		Equation 2.32
$C_{nm}$	Concentration contribution of segment $n$ on trajectory $m$ [ $\text{kg}/\text{m}^3$ ]	Equation 2.33	
$Cp_a$	Specific heat of air [ $\text{kJ}/\text{kg K}$ ]		Equation 4.25
$Cp_p$	Specific heat of the plume [ $\text{kJ}/\text{kg K}$ ]		Equation 4.25
$C_{sample}$	Sample concentration fraction		Equation 1.2
$C_{ss}$	Steady state concentration [ $\text{kg}/\text{m}^3$ ]	Equation 2.3	
$C_{\varepsilon 1}, C_{\varepsilon 2}, C_{\mu},$	Closure coefficients		Equation 4.11-4.13
$\sigma_k, \sigma_{\varepsilon}$			
$C_{1RNG}, C_{\varepsilon 1},$	Turbulence model constant	Equation 4.15	Equation 4.14
$C_{\varepsilon 2}, C_{\mu}, \sigma_k,$			
$\sigma_{\varepsilon}$			
$\mathbf{c}^B$	Background concentration vector		Equation 2.24
$\mathbf{c}^k$	Model concentration vector at time $k$		Equation 2.24
$\mathbf{c}_{obs}^k$	Observed concentration vector at time $k$		Equation 2.24
$D$	Volume $D$		Equation 2.13

$D$	Dimension in space		Equation A.2
$D_f$	Turbulent diffusivity [ $\text{m}^2/\text{s}$ ]	Equation 5.1	
$\mathbf{D}_t$	Covariance matrix at time $t$	Equation 2.23	Equation 2.22
$D_\phi$	Kinematic diffusivity [ $\text{m}^2/\text{s}$ ]		Equation 4.24
$DE$	Discretization error		Equation A.8
$d_i$	Rank difference		Equation B.7
$d_p$	Diameter of plume exit [m]		Equation 4.25
$E'_1$	Relative error of fine grid	Equation A.12	
$E_1$	Approximate relative error of fine grid	Equation A.13	
$E_{1,f}$	Approximate error of fine grid relative to value range	Equation A.14	
$E_{1,\max}$	Maximum approximate error of fine grid relative to value range	Equation A.15	
$E_{1,rms}$	Root Mean Square approximate error of fine grid relative to value range	Equation A.16	
$E_R$	Autocorrelation statistic moments	Equation B.4- B.5	
$\mathbf{E}_t$	Vector of source strengths at time $t$		Equation 2.21
$ER$	Emission rate [kg/h]		Equation 1.1
$ER$	Emission rate [cfm]		Equation 1.2
$F_b$	Buoyancy flux [ $\text{m}^4/\text{s}^3$ ]	Equation 4.25	

$f$	Matrix map		Equation B.3
$f_k$	Numerical solution		Equation A.1
$f_{ex}$	Exact solution		Equation A.1
$G_{ij}$	Spatial proximity measure	Equation B.2	Equation B.1
$\mathbf{G}_t$	Gain matrix at time $t$	Equation 2.22	Equation 2.21
$g$	Gravitational acceleration [ $\text{m/s}^2$ ]		Equation 4.25
$g$	Coefficient		Equation A.1
$g$	Matrix map		Equation B.3
$H$	Effective release height [m]		Equation 2.5
$\mathbf{H}_k$	Observer operator at time $k$		Equation 2.24
$h$	Height [m]		Equation 2.2
$h$	Grid width measure		Equation A.1
$I$	Turbulence intensity	Equation 4.22	
$i, j, k$	Cell identifiers		Equation 2.30
$i, j$	Location in matrix		Equation B.1
$J$	Jacobian		Equation 2.11
$J$	Cost function		Equation 2.24
$K$	Eddy diffusion tensor [ $\text{m}^2/\text{s}$ ]		Equation 2.1
$K$	Number of simulations		Equation 2.19
$k$	Position in vector		Equation 2.9
$k$	Turbulence kinetic energy [ $\text{m}^2/\text{s}^2$ ]		Equation 4.10
$k$	Mesh		Equation A.1
$k_{vK}$	von Karman's constant		Equation 4.17

$L$	Likelihood function	Equation 2.20	Equation 2.19
$M$	Total number of trajectories		Equation 2.32
$M_a$	Molecular mass of air [u]		Equation 4.25
$M_p$	Molecular mass of plume [u]		Equation 4.25
$\dot{m}$	Mass added with time [kg/s]		Equation 2.2
$\dot{m}$	Pollutant emission rate [kg/s]		Equation 2.5
$m$	Mass [kg]		Equation 2.6
$m$	Current trajectory		Equation 2.32
$m_{ex}$	Mass exchanged [kg]		Equation 2.13
$m_D$	Mass of air in Volume $D$ [kg]		Equation 2.13
$m_S$	Mass of air in Volume $S$ [kg]		Equation 2.13
$m_{trace}$	Mass of tracer [kg]		Equation 2.16
$N$	Number of steps, iterations or model components		Equation 2.10
$N$	Number of segment points in domain		Equation 2.30
$N$	Number of nodes/cells in mesh		Equation A.2
$N_m$	Number of segments in trajectory $m$		Equation 2.33
$N_t$	Noise matrix at time $t$		Equation 2.22
$n$	Current step, iteration or model component		Equation 2.10
$n$	Total number of observations		Equation B.4
$O$	Vector of sensor measurements		Equation 2.19

$P$	Time averaged pressure [Pa]		Equation 4.6
$\mathbf{P}_t$	Matrix of partial derivative model values with respect to source strength at time $t$		Equation 2.22
$p$	Probability		Equation 2.19
$p$	Instantaneous pressure [Pa]		Equation 4.3
$P$	Order of accuracy	Equation A.6	Equation A.1
$p_k, q_m$	Number of ties		Equation B.8
$Q$	Volumetric sweep rate [ $\text{m}^3/\text{h}$ ]		Equation 1.1
$Q$	Sample flow rate [cfm]		Equation 1.2
$R$	Ratio of change		
$\mathbf{R}_k$	Covariance matrix for observation error		Equation 2.24
$r$	Grid refinement ratio	Equation A.2	
$r$	Pearson rank correlation	Equation B.10	Equation B.9
$r$	Spatial correlation index	Equation B.6	
$r_s$	Spearman rank correlation	Equation B.7	
$S$	Volume $S$		Equation 2.13
$S$	Correlation constant		Equation B.4
$Sc_t$	Turbulent Schmidt number		Equation 4.24
$S_{ij}$	Strain rate tensor		Equation 4.15
$S_\phi$	Volumetric source term [ $\text{kg}/\text{m}^3\text{s}$ ]		Equation 4.24
$s$	Side length [m]		Equation 2.2

$s$	Stability parameter	Equation 4.29	Equation 4.28
$T$	Time period of trajectory [s]		Equation 2.35
$\tilde{T}$	Weighted concentration field [kg/m <sup>3</sup> ]		Equation 2.38
$T$	Correlation constant		Equation B.4
$\bar{T}_m$	Mass transfer potential field		Equation 2.37
$T$	Temperature [K]		Equation 4.29
$T_L$	Lagrangian turbulent integral timescale [s]	Equation 5.2	Equation 5.1
$TP$	Transition probability density	Equation 2.36	Equation 2.35
$t$	Time [s]		
$t_S$	Time at Volume $S$		Equation 2.13
$t_D$	Time at Volume $D$		Equation 2.13
$\bar{U}$	Wind velocity field [m/s]		Equations 2.1
$U$	Mean velocity [m/s]		Equation 2.2
$U$	Velocity component in $x$ direction [m/s]		Equation 2.5
$U$	Mean velocity component [m/s]		Equation 4.22
$U_i$	Error band estimate for grid 1	Equation A.17 - Equation A.19	
$U_i$	Velocity component $i$ [m/s]		Equation 2.7
$U_i$	Mean velocity component $i$ [m/s]		Equation 4.1
$\langle U_i \rangle$	Time averaged velocity component $i$ [m/s]		Equation 4.9

$U_p$	Velocity at start of plume [m/s]	Equation 4.25
$u$	Velocity [m/s]	Equation 4.16
$u, v, w$	Velocity components [m/s]	
$u^*$	Frictional velocity [m/s]	Equation 4.17
$u_i$	Velocity component $i$ [m/s]	Equation 4.1
$u_i'$	Fluctuating velocity component $i$ [m/s]	Equation 4.1
$u_i''$	Modelled turbulence velocity component $i$ [m/s]	Equation 4.9
$u_i'''$	Resolved fluctuating velocity component $i$ [m/s]	Equation 4.9
$w$	Side length [m]	Equation 2.2
$X, Y$	Cell center positions [m]	Equation 2.36
$\bar{X}_m$	Average concentration of field along trajectory $m$ [kg/m <sup>3</sup> ]	Equation 2.33
$X_{nm}$	Average concentration of field of trajectory segment $n$ on trajectory $m$ [kg/m <sup>3</sup> ]	Equation 2.33
$x$	Distance along $x$ axis [m]	Equation 2.6
$x, y$	Coordinate locations	Equation B.2
$x, y$	Vectors	Equation B.10
$x^*$	Distance until final plume rise [m]	Equation 4.27, Equation 4.28

$x_i$	Distance in direction $i$ [m]	Equation 2.7
$Y_k$	Prediction of $k^{\text{th}}$ simulation	Equation 2.19
$y$	Horizontal plume distance [m]	Equation 2.5
$z$	Vertical distance [m]	Equation 2.5
$z$	Vertical height [m]	Equation 4.16
$z_0$	Roughness height [m]	Equation 4.18
$z_h$	Zero plane displacement [m]	Equation 4.19
$z_{ref}$	Reference height [m]	Equation 4.21

## Greek

### Symbols

$\alpha$	Power law constant	Equation 4.21	Equation 4.20
$\beta$	Turbulence model constant		Equation 4.15
$\Gamma$	Adiabatic lapse rate [K/m]		Equation 4.30
$\Gamma$	Autocorrelation statistic	Equation B.1	
$\Delta h_B$	Buoyant plume rise [m]	Equation 4.26	
$\delta_i$	Discretization error of grid $i$		Equation A.17
$\varepsilon$	Exchange ratio		Equation 2.13
$\varepsilon$	Turbulent dissipation [ $\text{m}^2/\text{s}^3$ ]		Equation 4.11
$\zeta$	Diffusion operator [ $\text{m}^2/\text{s}$ ]		Equation 2.17
$\eta, \eta_0$	Turbulence model constants		Equation 4.15
$\theta$	Potential temperature [K]		Equation 4.29

$\lambda^k$	Interim iteration value at step $k$	Equation 2.27
$\mu$	Viscosity [Pa·s]	Equation 4.3
$\mu_T$	Eddy viscosity [Pa·s]	Equation 4.11
$\mu_t$	Turbulent viscosity [Pa·s]	Equation 4.24
$\rho$	Density [kg/m <sup>3</sup> ]	Equation 2.14
$\rho_{air}$	Density of air [kg/m <sup>3</sup> ]	Equation 4.16
$\rho_a$	Density of air [kg/m <sup>3</sup> ]	Equation 4.25
$\rho_p$	Density of plume [kg/m <sup>3</sup> ]	Equation 4.25
$\sigma_e$	Measurement error standard deviation	Equation 2.20
$\sigma_u$	Standard deviation of velocity [m/s]	Equation 4.22
$\sigma_x, \sigma_y$	Standard deviation about the trajectory [m]	Equation 2.36
$\sigma_y$	Standard deviation in horizontal direction [m]	Equation 2.5
$\sigma_z$	Standard deviation in vertical direction [m]	Equation 2.5
$\tau$	Residence time in cell [s]	Equation 2.32
$\tau$	Reynolds-stress tensor	Equation 4.8
$\tau_0$	Surface shear stress [Pa]	Equation 4.16
$\phi$	Concentration [kg/m <sup>3</sup> ]	Equation 4.24
$\chi_t^{obs}$	Vector of observed mixing ratios at time $t$	Equation 2.21

$\chi_t^{model}$

Vector of modelled mixing ratios at  
time  $t$

Equation 2.21

## Acronyms

BMC	Bayesian Monte Carlo
CAPP	Canadian Association of Petroleum Producers
CFD	Computational fluid dynamics
CPPI	Canadian Petroleum Products Institute
CTM	Chemistry transport model
CWE	Computational wind engineering
CWT	Concentration weighted trajectory
DES	Detached eddy simulation
DI&M	Directed inspection and maintenance
DIAL	Differential absorption LIDAR
DNS	Direct numerical simulation
EPA	U. S. Environmental Protection Agency
GRI	Gas Research Institute

IMSS	Image multi-spectral sensing
INGAA	Interstate Gas Association of America
IPCC	Intergovernmental Panel on Climate Change
IR	Infrared
LDAR	Leak detection and repair
LES	Large eddy simulation
LIDAR	Light Detection And Ranging
MC	Monte Carlo
OVA	Organic vapour analyzer
ppm	Parts per million
PSCF	Potential source contribution function
QTBA	Quantitative transport bias analysis
RANS	Reynolds averaged Navier-Stokes
RNG	Re-Normalisation Group
RTWC	Residence time weighted concentration
TSM	Trajectory statistical method
VOC	Volatile organic compound

# **1 Introduction**

## **1.1 Motivation**

Fugitive emissions in the oil and gas industry are large, typically 'invisible' sources of greenhouse gas emissions. Though a small number of fugitive emissions could be considered planned, such as from pneumatics that use pressurized methane gas and vent it into the atmosphere, most are unintended. Although various definitions exist, in general, fugitive emissions can be described as unmonitored, unintended and/or uncontrolled gas releases into the atmosphere. These releases can come from leaking valves, seals and fittings; evaporation losses; and process faults and failures (Carras et al., 2006). Even for planned releases such as the venting of process gases, composition and quantity are uncertain values (Carras et al., 2006). This results in thousands of potential emission sources in a typical refinery or plant along miles of piping.

Fugitive emissions from such sources are a significant problem. The Canadian Association of Petroleum Producers (CAPP), in a report prepared by Clearstone Engineering based on data from the year 2000 (CAPP, 2004), estimated that annually in Canada, 477 kt of methane leaks from plant equipment alone. Based on the current IPCC global warming potential (GWP) value for methane of 25 [100-year GWP] (Forster et al., 2008), these methane emissions alone are equivalent to 12 Mt of CO<sub>2</sub>e. Fugitive equipment leaks in general were estimated to cause 12% of the emissions within the upstream oil and gas industry and 2% of the total greenhouse gas emissions in the country (CAPP, 2004). Based on emission factors (CAPP, 1999; CAPP 2004; CPPI, 2004; CAPP, 2006) and production levels, Environment Canada (2010) estimated that

1262 kt of fugitive methane emissions from the oil and natural gas industry (excluding venting and flaring) were generated in 2008. Using the latest IPCC GWP, this is equivalent to 31.6 Mt CO<sub>2</sub>e (or 4% of Canada's total 2008 GHG emissions).

At an individual plant, significant improvements in emission levels can be made by identifying and eliminating only a small number of sources. Investigation by industry has suggested that the majority of fugitive emissions are often the result of a small number of large leaks rather than a large number of small leaks (Ross & Picard, 1996; Picard et al., 1998). More recently, CAPP published best management practices stating that, for upstream oil and gas facilities, "only the top 5 to 10 percent of leaking components account for 80 to 90 percent of emissions" (CAPP, 2007). Resources are therefore best allocated to eliminate these locations as they make the largest difference. Finding these locations, however, is not always easy, especially in inaccessible or dangerous locations. Novel approaches to reduce the difficulty in identifying emissions sources at individual facilities would directly enable elimination and repair of leaks. Given the overall magnitudes of fugitive emissions and the benefits of fixing even a small number of the large sources, development of new detection strategies has the potential to make a significant difference in the fight against climate change. Furthermore, any technological approach that might permit continuous, *in-situ* monitoring at a site would enable identification of new leaks as they arise so they can be controlled in less time than with intermittent measurement approaches. Such a system could then be used to ensure that mitigation measures remain in place and are functioning and would simultaneously enable the types of continued verification of emissions reduction necessary for emissions trading and crediting strategies.

## **1.2 Current Industry Fugitive Emissions Location and Estimation**

Several different approaches are currently used to identify and quantify fugitive emission sources in industry. These are typically based on the use of emissions factor estimates, manual identification of individual leaking components, and direct measurement or remote downstream measurements. These various existing approaches are briefly reviewed below.

### **1.2.1 Emission Factor Estimates**

Estimates based on emission factors provide a simple way for a facility or site to determine approximate emission rates by using the factors in conjunction with activity data. The activity data can be general (such as lengths of pipe or number of compressor stations) or specific (such as the equipment fuel consumption or number of facility leaks), depending on the “tier” or level of detail (Innovative Environmental Solutions, Inc., 2005).

Not surprisingly, however, the use of volume-based emission factors can introduce large errors into overall estimates of fugitive emissions. This is also true of using system-wide default factors and activity data, because facility type, age, and maintenance practices have large effects. Alternatively, performing bottom-up estimates and measurement programs is difficult and costly. In addition, estimates of leakage based on lost and unaccounted gas are affected by meter accuracy and throughput because fugitives emissions are often on the same order of magnitude as meter uncertainty (Innovative Environmental Solutions, Inc., 2005).

The Interstate Gas Association of America (INGAA) recently published detailed guidelines for estimations at various levels of detail for natural gas storage and transmission facilities (Innovative Environmental Solutions, Inc., 2005) using factors typically derived from Gas Research Institute (GRI)/U. S. Environmental Protection Agency (EPA) field measurements and emission estimate protocols. The tier 1 approach to provide high-level qualitative values is based on the default emission factors with pipeline length and storage station counts (Innovative Environmental Solutions, Inc., 2005). Errors based on this approach can typically be up to several orders of magnitude from the amount predicted (Innovative Environmental Solutions, Inc., 2005). While higher tier approaches incorporate more detailed factors, the nature of fugitive emissions is still inherently uncertain, and default emission factors are a large contributor to this uncertainty (Innovative Environmental Solutions, Inc., 2005).

Higher tier approaches could be combined with screening methodologies such as EPA's Method 21, described below, to create specific emissions estimates based on the U.S. EPA Protocol for Equipment Leak Emission Estimates (EPA, 1995). This has been suggested as a method to improve emissions estimate quality (Innovative Environmental Solutions, Inc., 2005). Finally, although emissions estimation may provide an inventory for total emissions from a site, this type of approach does not help locate specific emission sources. To identify individual sources, some other method is required, such as direct measurement.

### **1.2.2 Manual leak detection and measurement**

Leak Detection and Repair (LDAR) and Directed Inspection and Maintenance (DI&M) programs that screen for individual leaks are the most common approaches to

determining and quantifying fugitive emissions. In a full leak detection program, a portable organic vapour analyzer (OVA) is used to measure gas concentration adjacent to potentially leaking components to estimate emission rates in accordance with EPA's Method 21 (Picard et al., 1998). Method 21 specifies a methodology for the determination of volatile organic compound (VOC) leaks from process equipment using portable hand-held sensors (EPA, 1995). It is intended to locate and classify leaks which exceed the gas concentration threshold at the leaking component's surface, not to directly measure emission rates (EPA, 1995). The instrument detector is not specified by Method 21, but must meet certain minimum performance criteria (EPA, 1995).

LDAR campaigns require that potential emission sources be monitored and screened individually, typically using Method 21, at regular intervals and leaking components replaced within a specified timeframe (EPA, 2007). This individual screening can be used in conjunction with emission factors to estimate emission quantities or with direct measurement to construct an emissions inventory for the facility (Picard et al., 1998). However, this individual screening is typically not practical or possible in large facilities since complete site surveys are extremely labour intensive and time consuming. Method 21 requires the operator to place the probe next to each valve or fitting resulting in approximately 20% of potential leak sources not being assessed routinely due to lack of resources, difficulty of access, and safety concerns (Chambers, 2006). The major LDAR campaign inefficiency is that the majority of areas surveyed have no leaks or have leaks with negligible overall effect on total emissions.

DI&M focuses on inspection and correction in the areas most likely to offer significant, cost-effective emission control with coarser screening in other areas (CAPP,

2007). CAPP recently released a Best Practice Guide for the use of DI&M campaigns for upstream oil and gas facilities (CAPP, 2007). DI&M also used Method 21 or equivalent alternative methods (CAPP, 2007). This methodology is anticipated to be more cost effective than LDAR, but suffers from similar limitations due to component accessibility and labour intensiveness.

### **1.2.3 Typical Leak Detection Methods**

#### ***1.2.3.1 Portable Gas Detectors***

Multiple types of hand held gas detectors exist which meet the specifications for Method 21 such as catalytic oxidation, flame ionization, infrared absorption and photoionization based detectors (EPA, 1995). Method 21 simply specifies that the instrument must be of standard dimensions and flow rate range and must be readable to within 2.5 percent of the specified leak concentration threshold which, for methane, is 10 000 ppm (EPA,1995).

The probe is used on the surface of or directly adjacent to a potentially leaking component, which may be difficult to access due to location within the facility and such things as insulation.

#### ***1.2.3.2 Soap Solution***

One low cost alternative to portable gas detectors suggested in Method 21 is the use of soap solution. Bubble formation can be used as a visual indicator of leaking components. While this approach can be effective, it does have multiple limitations. Soap solution cannot be used where there are continuously moving parts, where there is liquid leakage, or where there are large gaps open to the atmosphere that need to be bridged (as often occurs at pipe flanges). In addition, test locations must have a surface temperature

between freezing and boiling of the solution (EPA, 1995). Soap solution cannot be used to determine leak significance since it gives no quantitative measurement, and suffers from the same or worse accessibility limitations of portable gas detectors. In spite of these limitations, using a soap solution to tag leaking components is generally 3 to 4 times faster than applying Method 21 with an organic vapour analyzer (Picard et al., 1998).

#### ***1.2.3.3 Ultrasonic Leak Detection***

Ultrasonic (or acoustic) leak detection can be used to detect leakage of high pressure jets. The jet expands from high to low pressure as it moves into the atmosphere and, if the pressure difference is sufficiently high, the resulting pressure waves can be detected. In commercial systems, the required pressure difference is typically  $> 10$  bar (Gassonic, 2009). Portable and permanent installations are both currently available (Gassonic, 2009; Russelltech, 2009). Whereas the portable models are generally faster than portable gas detectors and permanent installations can be used for long term monitoring, acoustic detectors are insensitive to smaller leaks and are affected by background noise (Gassonic, 2009).

#### ***1.2.3.4 Infrared Camera Leak Detection***

Methane and other hydrocarbons absorb in the wavelength spectrum of modern infrared (IR) cameras. Using appropriate wavelength filters, it is possible to produce an image of the leaked emission plume. Whereas it is not possible to determine hydrocarbon species or emission rate with this approach alone, IR camera detection is a fast method of identifying leak locations (Chambers, 2004). In combination with other methods, such as spectrometers, infrared cameras can provide a method to determine concentration and

estimate emission rates. Leak Surveys Inc. is one example of a company which has developed such a commercial infrared camera for leak detection (the GasFindIR HAWK, [www.leaksurveysinc.com](http://www.leaksurveysinc.com)), which can be used for both valve to valve leak detection and in helicopter flybys for large areas, relaying leak information back to ground operators. However, complex backgrounds and lighting conditions may adversely affect the ability of the operator to identify emission sources.

#### ***1.2.3.5 Image Multi-Spectral Sensing***

Unlike the infrared cameras mentioned above, the Image Multi-Spectral Sensing (IMSS) system developed by Pacific Advanced Technology ([www.patinc.com](http://www.patinc.com)) is capable of detecting a range of species in real time. The system combines a diffractive imaging spectrometer and a tuneable filter. The IMSS system permits pictures to be taken at a range of different wavelengths. Using a processing algorithm combined with the particular infrared signatures of various species, only the desired species are retained (Hinnrichs, 2000). Like the infrared camera, the remote nature of this method allows imaging to be done in areas which might otherwise be considered inaccessible and speeds up leak identification compared with the portable gas detector (Envirotech Engineering, 2007). However, the ability to identify sources can be affected by lighting conditions and wind strength at sources (Envirotech Engineering, 2007).

#### ***1.2.3.6 Backscatter Absorption Gas Imaging***

For backscatter absorption gas imaging, the component being screened is illuminated by an infrared laser in front of a reflective target. The laser's wavelength must be tuned so as to be absorbed by the gas of interest, but not by the atmosphere. When the scene is illuminated, the gas plume attenuates part of the backscattered laser light forming a dark

cloud in the resulting recorded image (Kulp et al., 1998). This method is limited by the need for reflective surfaces behind the leak. Portable systems are also hindered by the weight and power requirements of the artificial lighting system (Chambers, 2004).

#### **1.2.4 Leak Quantification Measurements**

Once a leak source has been identified, it may still be necessary to quantify the size of the leak. Two of the most common techniques for quantifying individual sources are bagging or use of hi-flow gas samplers. Both methods are limited by accessibility of components and labour cost.

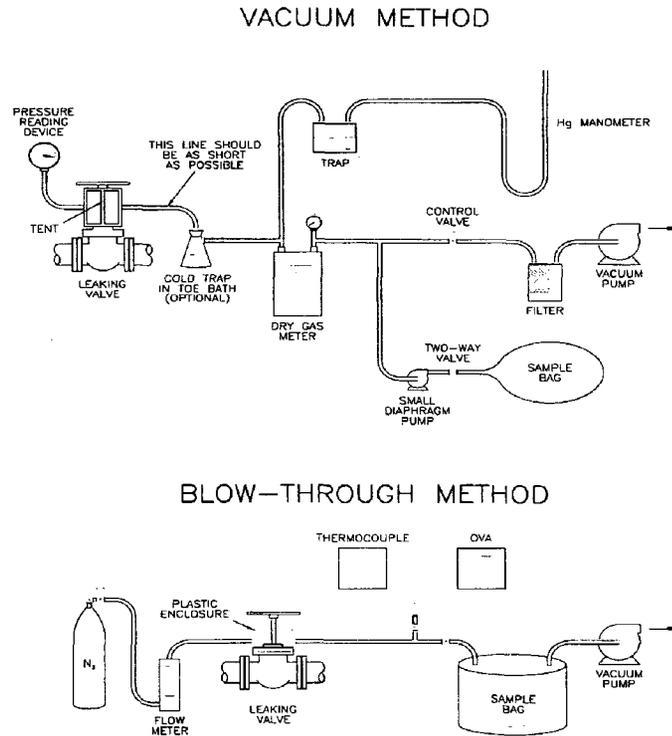
##### ***1.2.4.1 Bagging Techniques***

For conventional bagging (Picard et al., 1998), an enclosure is placed around the target source. A flow of clean gas (e.g. nitrogen) is passed through this enclosure, which sweeps the source gas to the sample bag to create a well-mixed environment for measurement. The emission rate can then be determined from the sweep rate (i.e. the measured flow rate of clean gas used to sweep the target source) and the steady-state concentration within the enclosure using:

$$ER = \frac{QC}{1 - C} \quad (1.1)$$

where  $Q$  is the measured volumetric sweep rate and  $C$  is the concentration of methane in the outlet air. Figure 1.1 shows two basic bagging setups as illustrated by Picard et al. (1998). In general, the enclosure is formed using non-permeable plastic sheeting and duct tape. Depending on the enclosure size, 15 to 20 minutes are required in general to

reach the necessary steady state conditions for the emission rate measurement (Picard et al., 1998).



**Figure 1.1: Bagging Techniques (Picard et al., 1998)**

#### *1.2.4.2 Hi-Flow Sampler*

The Hi-Flow Sampler is portable with the instrument packaged inside a back pack worn by the operator (Bacharach Inc., 2005). Air in the vicinity of a leaking component is drawn into the sampler at a sufficiently high flow rate that it can be assumed that all of the leaking gas is being captured. If the flow rate and concentration are measured accurately, the gas leak emission rate can be found as (Bacharach Inc., 2005):

$$ER = Q(C_{sample} - C_{background}) \quad (1.2)$$

where  $Q$  is the sample flow rate and  $C$  is the concentration fraction. The gas sample is drawn in through a 1.5" ID hose which has various attachments in order to capture all the leaked gas from the component. A venturi meter within the unit is used to calculate the sample flow rate. The concentrations of portions of the sample and the background level are directed to a combustible gas sensor which can measure in the range of 0.05% to 100% methane gas by volume. The sample is finally exhausted back into the atmosphere away from the sampling area (Bacharach Inc., 2005). The Hi-Flow Sampler performs with the same order of accuracy as bagging but improves measurement speed to purportedly make it as fast as typical leak screening techniques.

## **1.2.5 Remote and Downstream Detection and Measurements**

### ***1.2.5.1 Differential Absorption LIDAR***

Differential Absorption LIDAR (DIAL) is based on the use of lasers to determine the concentration along a line. Using multiple scans, it is possible to measure within a plane in the region downstream of the facility. Similarly, it is possible to use DIAL on a horizontal plane as permitted by geometry to develop concentration surface contour maps (EPA, 2007). Two laser wavelengths are used: one within the absorption band of the target gas and one that is not absorbed. The beams are directed to the measurement area and the back-scattered light from the atmosphere and/or the ground is collected. The difference in intensities of the backscattered signals is a measure of the absorption along the optical path, which can be used to determine the gas concentration (Demidov, 2002). From the relative time and strength of the returned signals, a profile of concentration along the light path can be calculated. Using a system of mirrors, it is possible to scan an area and develop a 2D or 3D map of gas concentration (Chambers, 2004). Unfortunately,

equipment complexity and power requirements prohibit the development of easily portable units. Spectrasyne Ltd. ([www.spectrasyneltd.uk](http://www.spectrasyneltd.uk)) is one group which has developed a mobile system mounted within a large “recreational vehicle” or RV that can be used to quantify “whole facility” fugitive emissions and identify significant leaks (Chambers, 2004). The DIAL method can be combined with aircraft to form an airborne system of gas pipeline leak detection in which the laser is transmitted downwards to illuminate buried pipe areas (Chambers, 2004) or over plant locations. Constructing a DIAL system such as the RV described above has been estimated to cost \$3-4 million (Chambers, 2003). Single site surveys using existing equipment can cost hundreds of thousands of dollars (Chambers, 2001).

#### ***1.2.5.2 Remote Sensing Leak Detection Approach***

Tracer gas techniques can also be used to estimate fugitive emission rates from a facility. In this approach, the tracer gas is released at a known rate within the facility. The ratio of concentrations of the tracer and the pollutant of interest measured downstream can then be used to estimate the pollutant emission rate (Picard et al., 1998).

A related approach uses a virtual tracer based on meteorological data, pollutant monitoring sensors, an accurate positioning system, and computational dispersion modelling to determine emission rates. The pollutant concentration is measured at a known downwind location and the source strength is found by back calculating from the site specific meteorological data and the chosen dispersion model (Picard et al., 1998). This method requires that the source locations be known in advance, as well as their relative strength and size in the case of multiple or non-point sources. This makes this method an emissions estimation technique rather than one useful for leak detection.

However, the combination of downstream concentration measurements and simulated meteorology and wind conditions has been used to determine emission fluxes and release locations within other fields, especially for large scale (e.g. continental) pollutant tracking.

### **1.3 Proposed Inverse Pollutant Locating Methodology**

An alternative method to those currently used in industry, is to use a combination of downstream concentration measurements and wind flow information to locate unknown, major fugitive emission sources. With permanent installation of onsite concentration sensors and wind measurement instruments, such a system could be used in conjunction with facility geometry data to determine likely pollutant source locations on a continuous basis. This type of approach would offer a significant advantage over current sporadic measurements necessitated by lack of equipment, trained personnel, and expense. A monitoring technique based on a permanent sensor network could also be used in combination with other techniques to enable directed searches into areas of interest with such tools as gas sniffers or IR cameras. This would enable intelligent deployment of resources to efficiently achieve the most significant emissions reductions.

There are multiple challenges that need to be met for such an approach. These include:

1. Sensor network design and development,
2. Wind flow determination, and
3. Development and implementation of effective and efficient source location algorithms.

The focus of the present work is on the source location algorithms themselves. A synthetic data set with known source locations and strengths was used to permit controlled investigation of a range of potential source location algorithms. This synthetic data set was generated using a CFD model for wind flow and pollutant dispersion over the simplified geometry of an actual operating gas plant in Alberta, Canada. In this approach, the wind field and the concentrations could be considered known at all locations. Virtual sensors extracted concentration values to provide, along with the wind information, necessary input data to the source location algorithms. This left an ill-posed, inverse dispersion problem to determine the source location(s). Solutions using the various test algorithms could then be compared with the known, simulated leak(s).

Key questions to be addressed in this work included:

1. Can the source location algorithms determine the pollutant source location regions to useful accuracy within a plant geometry consisting of a group of buildings?  
For algorithms developed for continental-scale applications, can they be applied at building scale?
2. If so, how well do the algorithms perform in these idealized circumstances?
3. How long is necessary to achieve a solution?
4. How can the parameter settings be used to improve the solution accuracy?

### **1.3.1 Thesis Overview**

The remainder of this thesis is organized as follows:

Chapter 2 provides a survey of background information on forward and inverse transport modelling. In addition, a summary of optimization algorithms and trajectory statistical

methods is provided which can be used to determine source locations. Finally, a selection of the methods considered in the current study is made.

Chapter 3 provides a detailed literature review of the source location algorithms under study: potential source contribution function, concentration weighted trajectory, residence time weighted concentration, and quantitative transport bias analysis.

Chapter 4 describes the computational fluid dynamic modelling of the wind and the pollutant dispersion around the virtual plant site in order to generate the synthetic data.

In Chapter 5, the results of parameter variation on the ability of the four chosen algorithms to determine source location are presented and discussed.

Chapter 6 gives conclusions derived from the results and discussion of Chapters 5 and outlines potential areas for further investigation.

## **2 Background and Inverse Solution Methods**

Whereas Chapter 1 gave an overview of the proposed methodology combining sensor data, CFD calculations and source location algorithms, this chapter expands on necessary background information and various existing methods for solving types of inverse problems relevant to the current project.

The modelling of gas transport provides an essential link between the measurement sensor locations and emission source distribution within the domain. Forward transport models are source location centric and model the movement of pollutant outward from sources and forward in time. Inverse transport models link a measurement location to points upstream by moving backward in time. As both approaches are central to any source location algorithm, this chapter provides an overview of these models. Brief descriptions of solution optimization methods and residence-time based statistical algorithms are subsequently presented. These algorithms can be used in combination with forward and/or backward transport simulations to determine pollutant origin. Finally, the rationale for the selection of the type of source location algorithms pursued in the present work is presented.

### **2.1 Forward Transport Modelling**

The ability to determine the downwind concentration of a released gas under various conditions is an important part of pollution control engineering. The desire to predict these concentration values in advance of such a pollution release event has led to various models being developed to deal with the complexity of the problem.

### **2.1.1 Full Navier-Stokes Equations**

One possible method for modelling a pollutant release is to use the fundamental equations for momentum in conjunction with conservation of mass and energy. The pollutant released is solved along with the flow field of the air, rather than as a separate step, which allows for the possibility that the pollutant may influence the flow. More details of methods to solve these equations are given in Section 4.3 with the discussion of wind modelling. This method requires the fewest approximations and simplifications, and can be combined with chemical reactions, but is also the most computationally intensive. It is not possible to take advantage of a pre-computed flow field with changing release parameters as the two must be solved simultaneously. This is a further disadvantage when multiple sets of release parameters need to be modelled for the same wind conditions. In practice, solutions based on solving full Navier-Stokes equations are not feasible on large, industrially relevant domains.

### **2.1.2 Advection-Diffusion Equation**

Rather than solve the full Navier-Stokes equations, in many cases the pollutant release is modelled separately from the wind field. The advection-diffusion equation is a popular method based on the conservation of species' mass, assuming a non-reacting, neutrally buoyant gas where transport is driven by advection with the wind field as well as diffusion. This still allows for the representation of temporal and spatial variations for the source strength and the wind properties within the domain. In general, molecular diffusion is considered insignificant in comparison to turbulent diffusion. This gives a representation of the atmospheric diffusion process of (Runca et al., 1979):

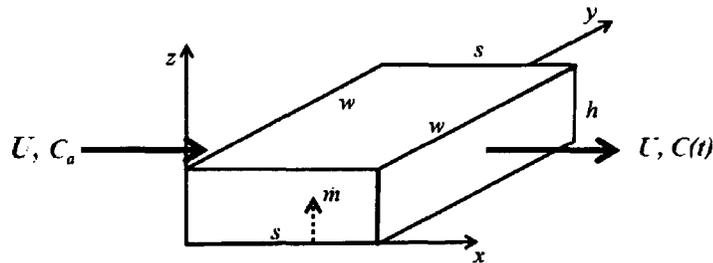
$$\frac{\partial C}{\partial t} + \vec{U} \cdot \nabla C = \nabla \cdot (K \nabla C) + \text{Source} \quad (2.1)$$

where  $C$  and  $\vec{U}$  are the pollutant concentration and wind field respectively and  $K$  is the eddy diffusion tensor. For many practical applications, this is still too complicated. Thus, even simpler methods for ideal conditions have been developed.

### 2.1.3 Box Models

Box models are the simplest form of pollutant dispersion models. Whereas pollutants mix at finite rates causing downwind concentrations to vary, box models assume infinitely fast mixing. This simplification greatly reduces the complexity of the calculations (Heinsohn & Kabel, 1999).

In a typical box model, a volume is defined in which this rapid mixing is assumed to occur. Air at velocity  $U$  is defined to enter and leave through area  $hw$  as defined in Figure 2.1.



**Figure 2.1: Box Model (adapted from Heinsohn & Kabel, 1999)**

This is reasonable in cases where the physical space is bounded by tall buildings, tunnels, indoor environments, or within a valley (Heinsohn & Kabel, 1999). The height of the box must be chosen so that it is of a sufficient altitude that the pollutants do not pass through the upper surface as the box model is based on the conservation of mass of

both the air and the pollutants and does not consider flux in that direction. The conservation of pollutant mass equation includes a term for the addition or subtraction of mass  $\dot{m}$ , which could occur during a pollutant release or disposition within the region (Heinsohn & Kabel, 1999):

$$whs \frac{dC}{dt} = \dot{m} + UC_a hw - UChw \quad (2.2)$$

For a steady state solution, this equation simplifies to (Heinsohn & Kabel, 1999):

$$C_{ss} = C_a + \frac{\dot{m}}{Uhw} \quad (2.3)$$

It is also possible to integrate Equation (2.2) directly. Assuming that the initial concentration is the same as the ambient conditions and that the variables are constant, the concentration is given by (Heinsohn & Kabel, 1999):

$$\frac{C(t) - C_a}{C_{ss} - C_a} = 1 - \exp\left[-\frac{Ut}{s}\right] \quad (2.4)$$

where  $s$  is the length of the box in the wind direction. In cases in which variables such as wind speed keep changing, the integration can be performed using less direct techniques (Heinsohn & Kabel, 1999).

The box model's simplicity makes it practical to use without great computational expense. However, the model assumes that the pollutant is distributed homogeneously and, although this can provide useful information, this approach does not provide much detail of the distribution of concentration throughout the plume itself or downstream locations.

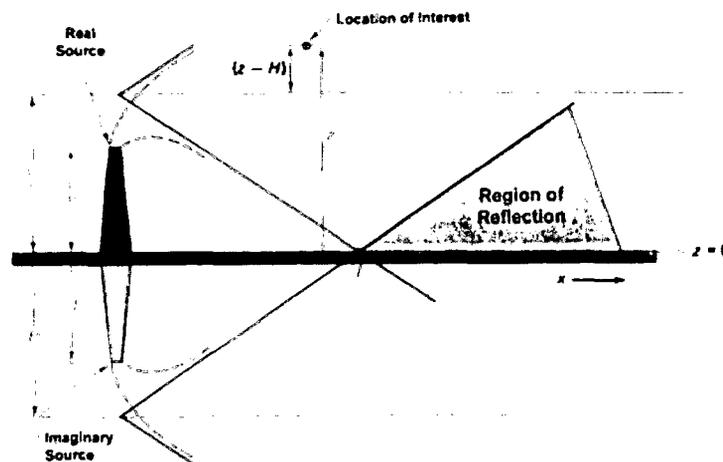
### 2.1.4 Gaussian Plume Dispersion Model

The Gaussian plume model is another simple and commonly used approach for predicting atmospheric dispersion. This is a steady state model for continuous releases, and advection by the wind is considered the primary transport mechanism along the wind direction rather than turbulent mixing (Reible, 1999).

Referring to Figure 2.2, the basic form of the standard Gaussian plume model is:

$$C(x, y, z, H) = \frac{\dot{m}}{2\pi\sigma_y\sigma_zU} \exp\left[-\frac{y^2}{2\sigma_y^2}\right] \exp\left[-\frac{(z-H)^2}{2\sigma_z^2}\right] \quad (2.5)$$

where  $C$  is the steady state concentration,  $\dot{m}$  is the emission rate of the pollutant,  $U$  is the wind speed (where the wind direction is the  $x$  direction),  $y$  is the horizontal distance from the center of the plume,  $z$  is the distance from the ground and  $H$  is the effective release height (Jung et al., 2003).



**Figure 2.2: A plume from stack with virtual reflection where the effect of the ground is considered using a second virtual plume. This requires an additional term in the standard equation. (Wark & Warner, 1981)**

This equation is equivalent to the Gaussian probability distribution and is referred to as the Gaussian plume model (Reible, 1999). The dispersion parameters  $\sigma_y$  and  $\sigma_z$  represent the standard deviations in the crosswind and vertical spread of the concentration distribution which are estimated by empirical correlations. These correlations are typically assumed to be proportional to time or distance traveled (Reible, 1999) and depend on the atmospheric stability class.

The model is simple and quick to solve even by hand, but assumes that the wind is constant and uniformly directional, which does not hold in complex terrain (Reible, 1999). The model also has no memory so that previous concentrations have no effect on the prediction.

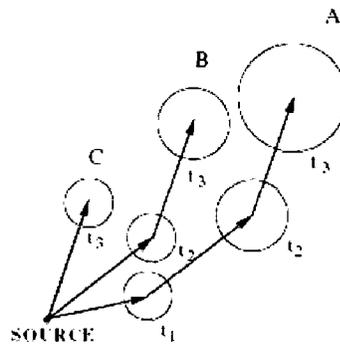
### 2.1.5 Gaussian Puff Models

Puff models approximate plumes with a series of equivalent puffs. The mass of each puff is the equivalent mass released for the prescribed time interval, and the dispersion parameters evolve as in the plume theory above. Thus, for a single puff (Jung et al., 2003):

$$C(x, y, z, H) = \frac{m}{(2\pi)^{3/2} \sigma_y^2 \sigma_z} \exp\left[-\frac{x^2}{2\sigma_y^2}\right] \exp\left[-\frac{y^2}{2\sigma_y^2}\right] \exp\left[-\frac{(z-H)^2}{2\sigma_z^2}\right] \quad (2.6)$$

where  $m$  is the mass of the puff (amount released during the time interval). The concentration at a location is found from the sum of all existing puffs in the domain. The puff's center is advected by the local wind vector which allows variation in both time and space. The dispersion parameters' growth can be calculated from virtual horizons and vertical distance. Thus, each puff can follow its own path through the domain as seen in

Figure 2.3 where puffs A, B and C are generated at times  $t_1$ ,  $t_2$  and  $t_3$  respectively (Jung et al., 2003).



**Figure 2.3: Movement of puffs A, B and C (Jung et al., 2003)**

Two types of puff models exist. The first is the plume segment type model, for which each puff represents the concentration distribution of over a period of time of the pollutant release, as described above. The second is the cluster dispersion type model, in which the puff is identified with clusters of particles. Turbulent eddies smaller than the puff size contribute to its growth while larger ones move the puff itself (de Haan et al., 2000).

The accuracy of the model depends on the dispersion coefficients and the advecting wind field. Unlike the simpler Gaussian plume model, the puff model no longer is restricted to steady state, uniform wind flows.

### 2.1.6 Particle Tracking

The particle tracking method is based on releasing large numbers of particles, rather than puffs, into the flow field. The velocity field can be pre-computed offline as is true with

the advection-diffusion equation. In general, the particle position in each of the coordinate directions after a release time interval of  $\Delta t$  is solved as (Kim et al., 2008)

$$x_i(t + \Delta t) = x_i(t) + U_i(t) \cdot \Delta t \quad (2.7)$$

where velocity  $U_i$  is made up of mean modelled wind and turbulent fluctuations in direction  $i$ .

One of the difficulties of such a method in comparison to the Gaussian plume or puff models is determining the concentration in a certain location (de Haan, 1999). For atmospheric dispersion modelling, common practice is to use a volume average over a grid cell in space, i.e. by summing the mass of all particles within the cell and dividing by the cell volume (de Haan, 1999). However, using large cell volumes can lose important details, while small volumes can suffer from random fluctuations so it is difficult to select the best cell size for determining concentration (de Haan, 1999). Alternatively, the density kernel approach can be used in which density distributions of different shapes (e.g. Gaussian) are used to distribute each particle in space (de Haan, 1999).

The particle tracking method allows for a more complete view of dispersion over a complex domain than the simple Gaussian plume model (Kim et al., 2008). This method can also be used to avoid solving large numbers of simultaneous differential equations (Kim et al., 2008). Whereas this reduces the computational resources required relative to a full Eulerian solution, the particle tracking method cannot model nonlinear processes and chemistry (Peters et al., 1995).



### **2.1.8 Summary of Forward Transport Models**

The selection of a forward transport model is a trade-off between computational time and complexity and must consider the applicability of the simplifications used in the various models. Box models can provide useful information and multiple boxes can be combined for more complex geometrical regions, but are not good at providing useful information about the concentration profiles downstream. Thus, for applications where this is of larger concern than computational simplicity, box models would not be the first choice. Gaussian plume models are also fast and quick to compute and give, for applicable flow fields, a concentration distribution downstream. Unfortunately, in the case of the flow around buildings, the unsteadiness associated with building wakes and other flow field complexities are at odds with the assumption of uniform wind direction for pollutant releases at the scales of interest. Puff and particle tracking models, as well as solving the complete Navier-Stokes or Advection-Diffusion equations allow for these types of spatially and time varying flows. Chemical transport models with additional chemistry are not required for the time and lengths scales of transport anticipated in modelling fugitive emissions within the plant environments (on the order of a few minutes over a maximum of several hundred meters).

## **2.2 Inverse Transport Models**

Forward transport modelling for gas dispersion starts from the gas release point and marches the solution forward in time. There is an alternative paradigm to investigating gas motion. Inverse transport models determine not where gas will go, but rather where it has been before reaching a point in the domain.

### 2.2.1 What is an adjoint?

Adjoint models are a tool to estimate the sensitivity and evolution of sensitivity through time of a few model output variables with respect to multiple input variables (Errico, 1997; Rao, 2007). These sensitivity fields can be used in conjunction with observations to determine which inputs best explain the outputs or to analyze how uncertainties in the model inputs affect the results (Errico, 1997). These model inputs can include pollutant release locations and strength.

Tangent linear models are models which compute the output perturbation from an input perturbation (Rao, 2007). Consider a model operator  $B$  (Errico, 1997):

$$\mathbf{b} = B(\mathbf{a}) \quad (2.8)$$

With a perturbation of the input  $\mathbf{a}'$ , the change in output can be approximated by a first order Taylor series approximation.

$$\Delta b_j \approx b_j' = \sum_k \frac{\partial b_j}{\partial a_k} a_k' \quad (2.9)$$

For small perturbations, this is valid as long as the derivatives exist (Errico, 1997). A sequence of operations to determine the net result can be considered instead of looking at each of the individual values. Using the chain rule, for  $1 \leq n \leq N$  this net result can be determined as (Errico, 1997):

$$\begin{aligned} b_j'^{(0)} &= \sum_k \frac{\partial b_j^{(0)}}{\partial a_k} a_k' \\ b_j'^{(n)} &= \sum_k \frac{\partial b_j^{(n)}}{\partial b_k^{(n-1)}} b_k'^{(n-1)} \\ b_j' &= b_j'^{(N)} \end{aligned} \quad (2.10)$$

where  $b_j^{(n)}$  is the  $j^{\text{th}}$  component after step  $n$  (Errico, 1997).

The Jacobian (or propagator) is the matrix describing the derivatives from the equations above, determined with respect to the model input (Errico, 1997). Now consider (Errico, 1997):

$$J = J(\mathbf{b}) = J[B(\mathbf{a})] \quad (2.11)$$

and

$$\frac{\partial J}{\partial a_j} = \sum_k \frac{\partial b_k}{\partial a_j} \frac{\partial J}{\partial b_k} \quad (2.12)$$

Equation (2.9) relates perturbations while Equation (2.12) relates sensitivity gradients and the input and outputs are reversed. The Jacobian of Equation (2.12) is transposed and is, mathematically, the adjoint corresponding to Equation (2.9). This adjoint operates backward in time and determines a gradient with respect to the input from a gradient with respect to the output (Errico, 1997).

This linear model is useful to the extent that the first order Taylor approximation gives good results for the perturbation size in the forward direction. In general, this is true. The accuracy of the approximations used in creating the Jacobian and the adjoint determine how well the adjoint will perform (Errico, 1997). The question is then how to determine these adjoint models.

### 2.2.2 Automatic Calculation

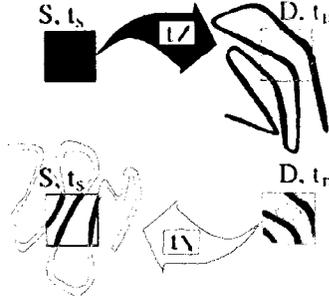
In many complex models, it is too difficult to determine the adjoint in an analytic fashion. The adjoint can be determined, however, using computational differentiation (Rao, 2007).

An example of this is the Tangent Linear Adjoint Model Compiler (TAMC), which can be used to perform automatic differentiation. Kaminski et al. (1999) used this to generate the adjoint of 3D CO<sub>2</sub> transport on a global scale. More details of the method used can be found in Giering and Kaminski (1998).

### 2.2.3 Advection-diffusion equation adjoint

While the multiple reactions and chemical reactions of the more complex chemical transport models may require automatic differentiation, in the case of the advection-diffusion equation, the adjoint equation can be determined analytically. An important concept in source location literature is the "retroplume", which designates the past history of a tracer sample. The "retroplume" concept acknowledges that air sampled at a measurement location arrives from somewhere upstream. This is combined with the idea that concentration measurement is influenced by sources, and is the basis for determining a source-receptor relationship (Issartel & Baverel, 2003).

Following the discussion presented by Issartel and Baverel (2003), consider the release of a tracer passively transported by air motion, with two volumes  $S$  at time  $t_s$  and  $D$  at time  $t_D$  where  $t_s < t_D$  as seen in Figure 2.5. The top part of the figure indicates forward transport of the tracer and the bottom part of the image indicates the adjoint movement backward in time.



**Figure 2.5: Tracer occupying domain  $S$  at time  $t_S$  moves partially to domain  $D$  at time  $t_D$  (dark grey). Looking backward in time, the air in  $D$  at  $t_D$  partially comes from domain  $S$  at time  $t_S$ . (Hourdin & Talagrand, 2006)**

The masses of air in the two volumes are denoted  $m_S$  and  $m_D$  respectively and the mass of particles exchanged between them at the prescribed times is  $m_{ex}$ . Thus, an exchange ratio between the two volumes can be defined as (Issartel & Baverel, 2003):

$$\varepsilon(S, t_S, D, t_D) = \frac{m_{ex}}{m_S m_D} \quad (2.13)$$

This ratio describes both the dispersion of air from  $S$  and the origin of air in  $D$ . At some time  $t$  following  $t_S$ , the local concentration of tracer emitted from volume  $S$  per unit mass of air can be denoted as  $C(x, t)$  and the total air released as  $m_S$ . In the inverse, let  $C^*(x, t)$  be the concentration of the plume sampled at  $D$  at some time  $t$  prior to  $t_D$  and the amount of air acting as an inverse tracer as  $m_D$ . The mass exchanged and the exchange ratio previously defined can now be evaluated as (Issartel & Baverel, 2003):

$$m_{ex} = \int_D \rho C(x, t_D) dx = \int_S \rho C^*(x, t_S) dx \quad (2.14)$$

$$\varepsilon = \frac{1}{m_D} \int_D \rho \frac{C(x, t_D) dx}{m_S} = \frac{1}{m_S} \int_S \rho \frac{C^*(x, t_S) dx}{m_D} \quad (2.15)$$

The normalized plume and retroplume concentrations are  $\hat{C} = \frac{C}{m_S}$  and  $\hat{C}^* = \frac{C^*}{m_D}$ .

The reciprocity relation is given by Equation (2.16) (Issartel & Baverel, 2003), where the overbars stand for averages in the volume at the denoted time:

$$\varepsilon(S, t_S, D, t_D) = \overline{\hat{C}(D, t_D)} = \overline{\hat{C}^*(S, t_S)} \quad (2.16)$$

Based on the above, an amount of tracer  $m_{trace}$  released in  $S$  at time  $t_S$  gives a plume with an average concentration per unit mass of air of  $\varepsilon(S, t_S, D, t_D)m_{trace}$  in  $D$  at  $t_D$ . Releasing the same amount  $m_{trace}$  in  $D$  at  $t_D$  and transporting backwards in time will lead to the same average concentration in  $S$  at  $t_S$ . Thus, forward and backward transport is equivalent.

Looking at the advection-diffusion equation with a wind field of  $\vec{U}$  and a diffusion operator of  $\zeta$ , the backward transport equation may be obtained. For forward transport, the equation can be expressed as (Issartel & Baverel, 2003):

$$\frac{\partial \hat{C}}{\partial t} + \vec{U} \cdot \nabla \hat{C} + \zeta(\hat{C}) = A \quad (2.17)$$

which is simply Equation (2.1) rewritten. As the two are equivalent, the backwards normalized retroplume may then be described as (Issartel & Baverel, 2003):

$$-\frac{\partial \hat{C}^*}{\partial t} - \vec{U} \cdot \nabla \hat{C} + \zeta(\hat{C}) = \hat{B} \quad (2.18)$$

#### 2.2.4 Lagrangian Backtracking

As forward and inverse plumes are equivalent with inverted flow fields, other methods of transport modelling besides the advection-diffusion equation can be used. Utilizing a

large number of Lagrangian backtracks (Lagrangian particle tracks with an inverted velocity field) from the measurement point can be considered the equivalent to the Eulerian approach to inverse transport described above (Issartel & Baverel, 2003). Moving backward in time, a single Lagrangian backtrack on the mean flow field representing the likely path of air transport can be considered a very simple adjoint of forward transport showing in which regions of the domain a pollutant release would affect the measurement point concentration.

### **2.2.5 Summary of Inverse Pollutant Transport Models**

Adjoint and inverse transport models allow determination of the sensitivity of measurement points to changes within other regions in the domain. Whereas some forward models have adjoints that are difficult to compute analytically, the advection-diffusion equation and particle backtracking adjoints are more easily found.

Unfortunately, whereas both forward and backward models of pollutant transport give pertinent information about the domain, transport models alone are not enough to determine where a pollutant release has taken place. Various inference techniques are required in order to determine locations (Rao, 2007). Thus it is also necessary to investigate methods for solving inverse problems.

## **2.3 Inverse Problems**

The use of downstream data to estimate upstream source locations is not as simple as first appearances may suggest. There exist a near infinite number of possible source and release strength combinations in the upstream field which must be narrowed down into a solution. The problem itself is an inverse, ill-posed problem which cannot be solved

directly. Modelling in a forward sense for which the location and strength are known is very simple in comparison. Working in the inverse direction requires different techniques.

The study of a physical system can be considered in three parts (Tarantola, 1987):

1. Parameterization of the system: determining the model parameters which completely define the system;
2. Forward problem: physical laws which, for given model parameters, can be used to make predictions of observable parameters; and
3. Inverse problem: the use of observable measurements to infer model parameters.

Unfortunately, the typical difficulty of an inverse problem is that it is ill-posed. A well-posed problem follows three conditions (Lavrentev et al., 1986):

1. the solution exists;
2. the solution is unique; and
3. the solution depends continuously on the problem data.

In the case of locating pollutant sources from point measurements downstream, there is not enough information to directly solve for the source locations and strengths. A single solution resulting in the observed measurements is unlikely in a mathematical sense. So, a solution must exist assuming that the measurements are correct, but the problem is more complicated. Fortunately, various techniques have been developed to determine most-likely solutions.

## 2.4 Solution Optimization Methods

One way of solving an inverse problem is to use optimization methods to either choose the best possible solution from a series of potential solutions, or to adapt and improve an existing solution.

### 2.4.1 Bayesian Inference

Bayesian inference uses observations (evidence) to determine the likelihood of a hypothesis being true. For example, the difference between modelled concentrations of a released gas and the actual measurements can be used to determine the probability of various combinations of source locations and strengths existing in reality. For the basic Bayesian Monte Carlo (BMC) method, results from a Monte Carlo (MC) analysis are assumed to correspond to the *prior* distributions of the model parameters and output (Rao, 2007). This *prior* distribution “represents the information about an uncertain parameter” (Gelman, 2002). The *prior* distribution in the case of locating emission sources is generally a vector containing information on source location and strength. With large sample sizes and well-defined parameters, reasonable choices of prior distributions have only minor effects on the posterior inference. This dependence can be checked by sensitivity analysis and comparing the posterior inferences under several different reasonable choices of prior distribution. The choice of prior distribution becomes more important for small sample sizes or when the data available provides only indirect information about the parameters of interest (Gelman, 2002).

Of the forward simulations performed during the Monte Carlo analysis, only those simulations that are consistent with observations within acceptable limits are kept (Rao, 2007). A continuous likelihood function that defines the probability of obtaining a

specified consistency refines the uncertainty estimates by weighing the results of the individual MC simulations (Rao, 2007). This allows the transport and dispersion modelling to be decoupled from the interpretation of measurements (Rao, 2007; Sohn et al., 2002).

One example of the use of BMC is the work of Sohn et al. (2002) who used the method to locate and characterize pollutant releases in buildings. The approach was divided into two stages. First, a transport model was selected, a computer model was then built of the buildings, uncertainties of the model inputs were characterized, and finally simulations of many airflow and pollutant transport scenarios were performed using a MC approach. These were done in advance of an actual pollutant release. During an actual release event, Bayesian updating was used to evaluate the agreement between the models and the sensor data.

The model inputs were taken from tracer gas flow experiments and modelling, literature and professional judgement (Sohn et al., 2002). Uncertain input parameters or variable building characteristics as well as source location, strength and emission duration were assigned an uncertainty distribution describing the range of possible values. From this, a library of model simulation was generated using a Monte Carlo technique. Sufficient sampling size could be tested by continuing calculations until summary statistic changes (e. g. mean, variance) would become negligible (Sohn et al., 2002).

During the actual release, data from the sensors was compared with the library of simulations using Bayesian updating. The data were compared with each simulation in

the library to assess the likelihood that it described the event. A high likelihood estimate meant that there was a high probability that the model inputs described the event being considered. By comparing these relative fits with Bayesian statistics, the best fitting model inputs and their uncertainties were estimated. In this comparison, the difference between data and predictions due to such factors as measurement error and model imperfections was also considered. The prior and posterior probabilities refer to the probability of each model before and after assessing their agreement (Sohn et al., 2002).

The posterior probability,  $p(Y_k|O)$ , of the  $k^{\text{th}}$  MC simulation making prediction  $Y_k$  given sensor measurements  $O$  can be calculated using Bayes' rule (Sohn et al., 2002):

$$p(Y_k | O) = \frac{L(O | Y_k) p(Y_k)}{\sum_{i=1}^K L(O | Y_i) p(Y_i)} \quad (2.19)$$

where  $L(O|Y_k)$  is the likelihood of observing  $O$  given prediction  $Y_k$ ,  $p(Y_k)$  is the prior probability of the  $k^{\text{th}}$  MC simulation, and  $K$  is the total number of MC simulations.

Before beginning the comparison, each is equally likely so that  $p(Y_k)$  is assumed to be  $1/K$ . The likelihood function,  $L(O|Y_k)$ , quantifies the error structure of the data, which are the result of such factors as “measurement error, spatial and temporal averaging or correlations and imperfect model representation” (Sohn et al., 2002).  $L$  is Gaussian in cases with uncorrelated normally distributed measurement errors (Sohn et al., 2002):

$$L(O | Y_k) = (\sqrt{2\pi}\sigma_e)^{-1} \exp \left[ -0.5 \left( \frac{O - Y_k}{\sigma_e} \right)^2 \right] \quad (2.20)$$

where  $\sigma_e$  is the standard deviation of measurement error. This may not be applicable in cases where there are correlated errors such as calibration bias (Sohn et al., 2002). The prior uncertainty of each of the model input and output parameter is then updated based on how well predictions in the prior uncertainty distribution agree with the measured data (Sohn et al., 2002).

The Bayesian inference method is a forward transport based method that requires a large number of transport simulations. Depending on the geometry, wind field variation and source distribution, this may be non-trivial task as some of the simpler and faster methods of forward transport are not applicable. For example, the assumptions made for Gaussian plume models do not generally hold true in the presence of buildings and other structures that impose unsteady wind conditions and create impassable objects. Thus, performing multiple forward runs of more computationally costly transport models may be necessary.

#### **2.4.2 Kalman Filtering**

In general, the Kalman filter is a set of equations to estimate a process state efficiently and recursively while minimizing the mean of the squared error (Welch & Bishop, 2006). Kalman filtering is another method of data assimilation which has been applied to determining pollution sources and improving their estimations. The standard filter can, in many cases, be considered a prototype algorithm on which other estimation procedures are based for specific purposes so Kalman filtering “is not an end unto itself” (Cohn, 1997). While the exact details may differ depending on the exact problem and the research group, the Kalman filter and its variants have been used to find optimized solutions to time-continuous problems. The algorithms used to estimate a state vector

(e.g. emissions) which is then combined with a forward model and the model's output is then used to update the state vector so that the model results better match the observations (Rao, 2007).

Hartley and Prinn (1993) used a linear Kalman filter to determine constant surface fluxes of  $\text{CFCl}_3$  in different regions in conjunction with a CTM and industry data to estimate an initial guess of sources and sinks. This filter was “derived on a linear relationship between the change in mixing ratio with time and the emissions over that time period” (Hartley & Prinn, 1993). This assumption holds true for long lived species as the nonlinear chemistry effects are on a much shorter timescale (Hartley & Prinn, 1993). The update equation (Kalman filter) was applied as (Hartley & Prinn, 1993):

$$\mathbf{E}_{t+1} = \mathbf{E}_t + \mathbf{G}_t (\chi_t^{\text{obs}} - \chi_t^{\text{model}}) \quad (2.21)$$

The vector of source strengths in each region at time  $t$ ,  $\mathbf{E}_t$ , is updated to  $\mathbf{E}_{t+1}$  based on a comparison of the mixing ratios  $\chi_t$  (volume of pollutant/volume of air) at each station to the model values. This is done through the “gain matrix”  $\mathbf{G}$ , which is also updated at each timestep (Hartley & Prinn, 1993). This matrix gives the sensitivity of the adjustment to differences between the model and observations (Rao, 2007).

$$\mathbf{G}_t = \mathbf{D}_t \mathbf{P}_t^T [\mathbf{P}_t \mathbf{D}_t \mathbf{P}_t^T + \mathbf{N}_t]^{-1} \quad (2.22)$$

$\mathbf{D}_t$  is the covariance matrix representing errors in the estimated emissions in each region;

$\mathbf{P}_t$  is the partial derivative matrix of the model values at each station with respect to the

regional source strengths; and  $\mathbf{N}_t$  is the covariance matrix of measurement errors at each

site (“noise matrix”) (Hartley & Prinn, 1993). The covariance matrix for emission errors is also updated as (Hartley & Prinn, 1993):

$$\mathbf{D}_{t+1} = \mathbf{D}_t - \mathbf{G}_t \mathbf{P}_t \mathbf{D}_t \quad (2.23)$$

The covariance matrix  $\mathbf{D}_t$  is necessary to calculate the gain and can be considered the difficult portion of the Kalman filter method (Rao, 2007). To initialize this matrix, it is necessary to define the goodness of the a priori estimate of the emissions from each region (Hartley & Prinn, 1993). After initializing the first guess, the model is run forward and the filter routine is used to obtain a new estimate for the initial conditions and this continues until the solution is considered sufficiently converged (Hartley & Prinn, 1993). Haas-Laursen et al. (1996) extended Hartley and Prinn's (1993) work to deal with time varying sources and sinks as many processes vary by season when dealing with emission location problems on a global scale.

Kalman filters work best when the relationship between the internal variables and the output is strong and the covariance matrix is well conditioned. In the cases of turbulent flow and transport, these mechanisms are not linear (Rao, 2007), making it difficult to determine the covariance matrix.

### 2.4.3 Genetic Algorithms

Genetic algorithms were introduced and popularized by Holland (1975) and are based on the biological idea of the chromosome. The model inputs for a potential solution are represented as a numerical string, which can be thought of as a chromosome by analogy. An initial population of these strings is chosen where each string defines a randomly chosen solution. The match between the potential solution and the observations is used

to determine its fitness and, using the fitness of the solution to determine the probability, a new population is created from the old members. This includes the use of “mating” which shuffles segments of the strings to create new ones and random mutations to prevent the search from settling on suboptimal answers (Cartwright & Harris, 1993). This method is basically an intelligent search where poor solutions are discarded and the population fitness increases with each iteration (Cartwright & Harris, 1993).

The fitness of a solution is in general evaluated by modelling the pollutant release forward using one of the methods described in the preceding section and then a cost function must be used to evaluate the resulting concentration field in comparison with the measured data. An example of this can be found in the work done at Pennsylvania State University using simple Gaussian plume models for forward dispersion in an empty field and root mean square difference for the cost function between the predicted and receptor data values for identical twin experiments (Allen et al., 2007). Work continued using puff models rather than plume models, different cost functions and continuing to blend sensor and model data (Haupt et al., 2007).

In the simple case of a single source, a genetic algorithm can be implemented with simple strings representing locations and source strength. In the case of multiple receptors and sources, it is inappropriate to use a single dimensional string and becomes necessary to use a matrix of dimensions dictated by the product of *receptors*  $\times$  *sources* (Cartwright & Harris, 1993). This makes the process more complex and computationally expensive, since each possible solution must be simulated for all forward iterations. In cases where the number of sources is unknown, determining how many potential sources to include within the algorithm is difficult.

#### **2.4.4 Simulated Annealing**

Simulated annealing is another method for solving global optimization problems (Spall, 2003). In essence, a neighbouring state to the current state is simulated and the decision of whether to move to this new state or remain at the current one is determined in a probabilistic fashion. This allows occasional moves to states that increase the cost function in order to avoid local minima (Bertsimas & Tsitsiklis, 1993).

The annealing principle, inspired by the physical process, introduces random perturbations which are reduced (annealed) in a controlled fashion to find a global minimum of a function without being caught in local minima (Spall, 2003). This willingness to allow a temporary increase in the cost function is the distinction of this type of optimization algorithm (Spall, 2003).

Molecules at high temperature have more mobility than at lower temperatures, potentially allowing them to align themselves into crystalline structures that minimize the system's energy state. Temperature alone does not govern this process. If the cooling takes place too quickly the resulting state may not be that of minimum energy (Spall, 2003). This minimum energy state represents the cost function and the aim is the lowest value in the case of an optimization problem. Temporary higher values are allowed as the alignment process continues. This is based on the Boltzmann-Gibbs probability distribution describing the probability of a particular energy state where, even at low temperatures, the probability of reaching a higher energy state is non-zero though decreasing with temperature (Spall, 2003). Metropolis et al. (1953) were responsible for introducing the idea of simulated annealing to numerical analysis.

Like the other forward based source location algorithms, simulated annealing is used in conjunction with a forward plume dispersion model and a cost function is used to assess the resulting values in comparison with measurements. Thomson et al. (2007) used such a method in conjunction with Gaussian plume models and random search algorithms to identify the source of ethane dispersion in a desert environment. This method tends to share the same strengths and weaknesses of the other forward algorithms, though is less likely to stay in a local minimum.

#### **2.4.5 Variational Data Assimilation**

Data assimilation is the use of measurements to constrain model predictions (Zhang, 2007). The measurement information is used to obtain better initial conditions, boundary conditions, etc. (Zhang, 2007). Sasaki (1958) developed variational data assimilation as a meteorological tool. Penenko and Obraztsov (1976) were the first to use variational methods in conjunction with adjoint equations. Time dependent problems, such as concentration measurements, use four-dimensional variational data assimilation to account for the distributed observations in time (Robertson & Langner, 1998).

Variational data assimilation uses a cost function measuring the misfit between a model and corresponding observations (Robertson & Langner, 1998). This cost function is often the sum of squares of the differences between the model and the observations (Rao, 2007; Robertson & Langner, 1998), though this may include estimations of background values and background error (Zhang, 2007; Elbern et al., 1997). Finding the minimum cost function is equivalent to determining the conditions that best explain the observations (Rao, 2007). The gradient of the cost function shows its sensitivity with

respect to changes of these conditions and indicates the directions that the initial guess should be changed in order to model the output better (Rao, 2007).

For example, consider the cost function,  $J$ , used by Zhang (2007) in conjunction with a chemical transport model concerned with ozone levels defined as

$$J(\mathbf{c}^0) = \frac{1}{2}(\mathbf{c}^0 - \mathbf{c}^B)^T \mathbf{B}^{-1}(\mathbf{c}^0 - \mathbf{c}^B) + \frac{1}{2} \sum_{k=0}^N (\mathbf{H}_k \mathbf{c}^k - \mathbf{c}_{obs}^k)^T \mathbf{R}_k^{-1} (\mathbf{H}_k \mathbf{c}^k - \mathbf{c}_{obs}^k) \quad (2.24)$$

where  $\mathbf{c}^B$  represents background (or a priori estimate) values,  $\mathbf{B}$  is the covariance matrix for estimating the background error,  $\mathbf{H}_k$  is the observer operator which ensures that the model values and the measured values are equivalent and have comparable units, and  $\mathbf{c}_{obs}^k$  are the observations measured at time  $k$ .  $\mathbf{R}_k$  is the covariance matrix, which accounts for observation and representativeness errors. The gradient with respect to the initial values is (Zhang, 2007):

$$\nabla_{\mathbf{c}^0} J(\mathbf{c}^0) = \mathbf{B}^{-1}(\mathbf{c}^0 - \mathbf{c}^B) + \sum_{k=0}^N \left( \frac{\partial \mathbf{c}^k}{\partial \mathbf{c}^0} \right)^T \mathbf{H}_k^T \mathbf{R}_k^{-1} (\mathbf{H}_k \mathbf{c}^k - \mathbf{c}_{obs}^k) \quad (2.25)$$

This gradient can be determined using a direct sensitivity method, where each of the derivatives is found with respect to all components of the initial state (Zhang, 2007).

$$\frac{\partial \mathbf{c}^k}{\partial \mathbf{c}^0} = \left[ \frac{\partial \mathbf{c}^k}{\partial c_1^0}, \frac{\partial \mathbf{c}^k}{\partial c_2^0}, \dots, \frac{\partial \mathbf{c}^k}{\partial c_n^0} \right] \quad (2.26)$$

The direct sensitivity method computes each derivative in the above matrix by using integrations of the tangent linear models with different initial conditions.

Depending on the forward model, this can be a huge computational expense. Zhang (2007) estimated that for a realistic CTM, this approach would take approximately

$10^7$  runs, making the task impossible in a practical sense. For more basic models that do not contain a large number of chemical reactions, this number would be much lower, but still is a distinct disadvantage. However, using the adjoint to compute this gradient is possible. Backward integration from the last timestep with condition (Zhang, 2007)

$$\lambda^N = \mathbf{H}_N^T \mathbf{R}_N^{-1} (\mathbf{H}_N \mathbf{c}^N - \mathbf{c}_{obs}^N) \quad (2.27)$$

then looping backward in time to the initial timestep with

$$\lambda^k = \left( \frac{\partial \mathbf{c}^{k+1}}{\partial \mathbf{c}^k} \right)^T \lambda^{k+1} + \mathbf{H}_k^T \mathbf{R}_k^{-1} (\mathbf{H}_k \mathbf{c}^k - \mathbf{c}_{obs}^k) \quad (2.28)$$

results in the gradient defined as

$$\nabla_{\mathbf{c}^0} J(\mathbf{c}^0) = \mathbf{B}^{-1} (\mathbf{c}^0 - \mathbf{c}^B) + \lambda^0 \quad (2.29)$$

This gradient can be used with a decent method (e.g. conjugate gradient method) to adjust the initial state guess toward the optimal solution as an iterative process (Robertson & Persson, 1993). The iterative process begins with a guess of the initial parameters. The forward model is run and then the adjoint in order to determine the gradient. The gradient is then used to adjust the initial state and so forth (Robertson & Persson, 1993).

Similar to the adjoint technique, variational continuous assimilation alters the model time derivatives instead of the initial conditions (Derber, 1989). Robertson and Langner (1998) used such a method to reconstruct source terms using data from the European Tracer Experiment.

Variational data assimilation takes advantage of the knowledge provided by the adjoint model as well as the forward transport simulation, reducing the number of required transport simulations relative to optimization methods using the forward transport model alone.

## **2.5 Trajectory Statistical Methods**

In addition to solution optimization methods, there exist other statistical source location algorithms which are specific to the problem of pollutant source identification. The trajectory statistical methods (TSMs) use physical ideas such as residence time and backward trajectories directly within each algorithm in order to determine probability fields or pseudo-concentration fields, which indicate likely pollutant source regions. An overview of each of the main TSM methods is provided below and these algorithms are discussed in further depth in Chapter 3.

### **2.5.1 Potential Source Contribution Function**

Potential source contribution function analysis (PSCF), also known as residence time probability analysis, is the simplest of the methods discussed. PSCF links residence time upwind of measurement locations and a set pollution criterion with a conditional probability field. Originally developed to analyze sulphur concentrations at Grand Canyon National Park (Ashbaugh et al., 1985), PSCF has been used extensively to identify sources and transport pathways for trace substances and particulate (Hsu et al., 2003; Ashbaugh et al., 1985; Lupu et al., 2002; Cheng et al., 2001).

In an implementation of PSCF, the backward air trajectory starting from the receptor location is determined for each measurement sample. Each trajectory is denoted

as polluted if the arrival concentration is above a chosen threshold. Trajectories are then divided by a set time interval  $\Delta t$  to create segment points. More segment points existing within a region corresponds to longer the residence time within this region. For the  $ijk^{th}$  cell within the domain,  $a_{ijk}$  is the total number of segment points present during the experimental time interval and  $b_{ijk}$  is the number of polluted segment points. If there are a total of  $N$  segment points within the domain from back trajectories originating at the receptor location, then the cumulative probability that an air parcel passed through the  $ijk^{th}$  cell before reaching the receptor site can be estimated as  $P[A_{ijk}] \approx a_{ijk}/N$ . Similarly, the probability that a polluted air parcel passed through the  $ijk^{th}$  cell before reaching the receptor is  $P[B_{ijk}] \approx b_{ijk}/N$ .

The conditional probability that an air parcel in a cell arrives at the receptor location with a concentration over the chosen criterion is then

$$PSCF_{ijk} = P[B_{ijk} | A_{ijk}] = \frac{b_{ijk} / N}{a_{ijk} / N} = \frac{b_{ijk}}{a_{ijk}} \quad (2.30)$$

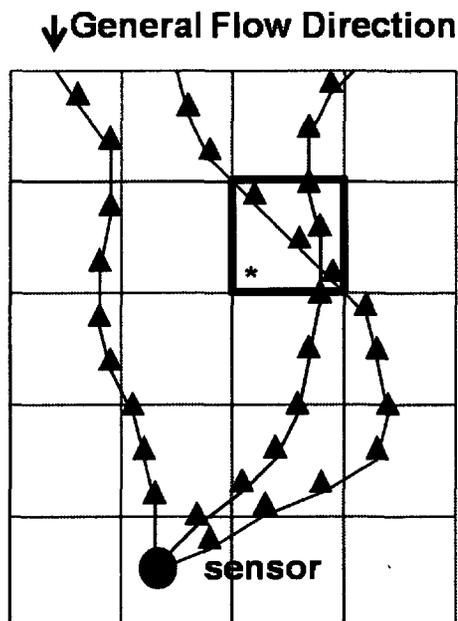
Thus, air parcels passing over high valued regions are more likely to be polluted.

To more clearly illustrate the concept of PSCF, Figure 2.6 shows three different trajectories from three different times that arrive at a sensor. Since the flow varies in time, the three trajectories are not identical. Each triangle along a trajectory marks divisions in time so that the number of triangles within a cell is representative of the residence time. If the concentration measured at that sensor at the time corresponding to the arrival of the "red" trajectory was high compared with that of the "blue" trajectories, intuitively the source is more likely to be along the region passed by the "red" trajectory. In the cell denoted \*, a low concentration trajectory passes over the high concentration

trajectory. Thus, it is less likely that this is the part of the red trajectory that passes over the source since the low concentration trajectory also crossed this region. However, the source is still more likely to be here than if two low concentration trajectories passed through this cell. Looking again at cell\*, the PSCF field value for that cell can be calculated as

$$PSCF_* = \frac{b_*}{a_*} = \frac{2}{5} \quad (2.31)$$

whereas the cells above and below have PSCF field values of 1.0. Cells transected by blue trajectories only (below the specified concentration threshold) have PSCF field values of zero.



**Figure 2.6: Sample trajectories of high (red) and low (blue) concentrations arrive at a single sensor**

The primary advantage of the PSCF method is its computational speed and simplicity. PSCF requires only one simple calculation, rather than an iterative approach, and does not require prior information about the sources within the domain. Determining the back-trajectories makes up the vast majority of the computational cost. Due to its presumption of a cut-off pollutant threshold, PSCF does not tend to pick up on "hot-spot" zones as well as some other methods and is known to have good angular resolution from the receptors, but, while some studies have shown good radial resolution, others have not (Vasconcelos et al., 1996). The determination of the concentration pollutant cut-off value and the data set size are important to the resulting source estimation, and higher values of both tend to improve estimates (Cheng & Lin, 2001; Poirot et al., 2001).

## **2.5.2 Concentration Weighted Trajectory and Residence Time Weighted**

### **Concentration**

Based on a modified version of PSCF, Seibert et al. (1994) developed the concentration weighted trajectory method, which was further refined by Stohl (1996) into the residence time weighted concentration (RTWC) method. RTWC considers the idea of hot spots along a trajectory (Hsu et al., 2003; Stohl, 1996).

Initially, a first guess concentration field, known as the concentration weighted trajectory (CWT) field, is calculated using the variation of Hsu et al. (2003) based on the method developed by Seibert et al. (1994) as follows.  $M$  is the total number of trajectories considered,  $C_m$  is the measured concentration at the receptor for trajectory  $m$ , and  $\tau_{ijkm}$  is the time spent by trajectory  $m$  within cell  $ijk$ . The initial mean concentration is weighted by residence time and is calculated as

$$\overline{C}_{ijk} = \frac{1}{\sum_{m=1}^M \tau_{ijkm}} \sum_{m=1}^M C_m \tau_{ijkm} \quad (2.32)$$

The refined method of Stohl (1996) redistributes the weighting of concentrations based on the idea that a cell along a 'clean' trajectory is unlikely to be a major pollutant source. Thus, each trajectory  $m$  is divided into  $N_m$  segments, where  $X_{nm}$  is the mean concentration of segment  $n$  of trajectory  $m$ , and is the average of concentrations along that trajectory. The contribution of segment  $n$  along trajectory  $m$  is then weighted according to:

$$C_{nm} = C_m \frac{X_{nm} N_m}{\sum_{j=1}^{N_m} X_{jm}} = C_m \frac{X_{nm}}{X_m} \quad (2.33)$$

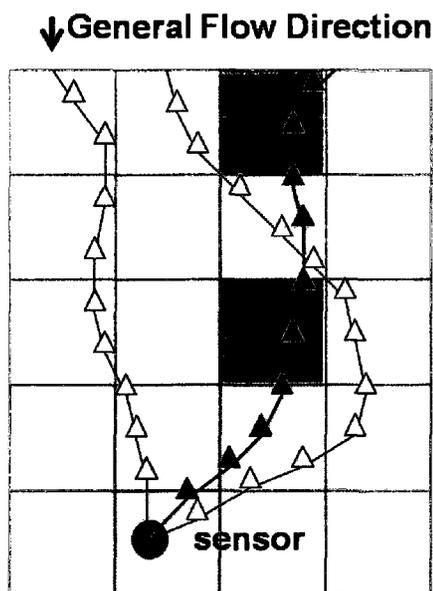
Finally, a new concentration field is calculated as

$$\overline{C}_{ijk} = \frac{1}{\sum_{m=1}^M \sum_{n=1}^{N_m} \tau_{ijknm}} \sum_{m=1}^M \sum_{n=1}^{N_m} C_{nm} \tau_{ijknm} \quad (2.34)$$

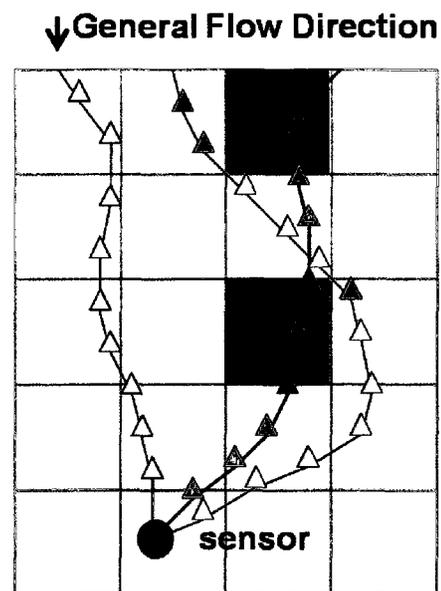
The calculation process Equations (2.33) and (2.34) are repeated until a chosen convergence level.

Figure 2.7 and Figure 2.8 show simple example fields for CWT and RTWC respectively to clarify the concepts. CWT uses the residence time within each cell similar to PSCF (illustrated above in Figure 2.6), but it does not incorporate a pollution cut-off threshold. Rather, CWT weighs the results field by the measured concentration. Thus, a cell where high concentration trajectories have long residence times has a higher CWT value than those with long residence times of low concentration trajectories.

RTWC starts from the initial CWT field and reweights it to produce the RTWC field as shown in Figure 2.8 for the simplified example case. Note that the highest concentration cells are darker in Figure 2.8 relative to Figure 2.7 due to the reweighting and that cells with both relatively clean and polluted trajectories are suppressed. RTWC incorporates this reweighting by exploiting the concept that the source location is more likely to be in areas where only high concentration trajectories pass, and is less likely to be in cells also containing low concentration trajectories. To perform this reweighting and define the overlap, each trajectory is divided into segments of a specified length or specified time period. Thus, in Figure 2.8, the region where the high and low concentration trajectories overlap are reweighted to have lower intensity in the results field. However, since the source is still most likely along the high concentration trajectory, the RTWC field values along the other regions of the trajectory are increased.



**Figure 2.7: Sample initial field generated by CWT**



**Figure 2.8: Sample reweighting of initial CWT field by RTWC**

PSCF, CWT and RTWC are all based on simple back-trajectories, giving a potential computational cost and speed advantage over solution optimization methods that require iteration with large numbers of increasingly complex transport models. These methods need to be run only once to generate a solution, even with the redistribution of the data in the case of RTWC, and do not require an initial guess or the number of sources to be defined. While not as straightforward as PSCF, CWT and RTWC do not use pollutant cut-offs. This allows the ability to distinguish between major and minor source regions. RTWC has good spatial resolution and is able to reproduce the main features emission features when compared with inventories (Lupu & Maenhaut 2002; Zhou et al., 2004). The redistribution, however, can create false positives so it is recommended to use RTWC in conjunction with a non-redistributed method such as PSCF or CWT (Zhou et al., 2004).

### **2.5.3 Quantitative Transport Bias Analysis**

Unlike PSCF and RTWC, quantitative transport bias analysis attempts to account for the trajectory uncertainties that dominate these methods (Zhou et al., 2004). In QTBA, back trajectories are assumed to have Gaussian profiles rather than a single definite direction. This, in turn, results in a similar idea to an inverse puff model combined with concentration measurements, though the Gaussian profile is based on trajectory uncertainty rather than diffusion.

Developed by Keeler et al. (1989), the original QTBA attempted to take into account the various uncertainties of trajectory. Simplified QTBA (Zhou et al., 2004) ignores the uncertainties caused by deposition and chemical reaction. For the sake of

readability, the method will be described for the two-dimensional case, but can be extended into the third dimension.

The probability of a tracer emission arriving at a given point in space at time  $t$  is (Zhou et al., 2004):

$$A(x, y, t) = \int_{t-T}^t \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} TP(x, y, t | x', y', t') dx' dy' dt' \quad (2.35)$$

where  $TP$  is the transition probability density function of the air parcel found at position  $(x', y', t')$  arriving at  $(x, y, t)$ . This probability density function is assumed to have a normal distribution about the trajectory, which increases linearly with time upwind. If  $X, Y$  are the locations of the grid cell centers,  $x'(t'), y'(t')$  are the positions along the trajectory, and  $\sigma_x, \sigma_y$  are the standard deviations about the trajectory (approximated as  $\sigma(t')=at'$  where  $a$  is a constant), then  $TP$  is (Zhou et al., 2004):

$$TP(x, y, t | x', y', t') = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left[ -\frac{1}{2} \left( \left( \frac{X - x'(t')}{\sigma_x} \right)^2 + \left( \frac{Y - y'(t')}{\sigma_y} \right)^2 \right) \right] \quad (2.36)$$

Integrating over an upwind period for trajectory  $m$  over time period  $T$ , gives the potential mass-transfer field (Zhou et al., 2004):

$$\bar{T}_m(x, y | x', y') = \frac{\int_{t-T}^t TP(x, y, t | x', y', t') dt'}{\int_{t-T}^t dt'} \quad (2.37)$$

Weighting by the sample concentration, gives the weighted concentration field

$$\tilde{T}(x, y | x', y') = \sum_{m=1}^M \bar{T}_m(x, y | x', y') C_m(x, y) \quad (2.38)$$

Dividing the concentration weighted mass transfer potential by the unweighted value gives the QTBA concentration field (Han et al., 2007).

$$QTBA(x, y | x', y') = \frac{\tilde{T}(x, y | x', y')}{\sum_{m=1}^M \bar{T}_m(x, y | x', y')} \quad (2.39)$$

The higher this concentration field is, the more likely it is that a given location is important to pollutants reaching the receptor.

Again, QTBA requires a single iteration and no prior knowledge of the domain. QTBA has been found to give consistent results in estimating locations for large and clearly defined sources. However, its performance is less consistent for smaller and multiple sources (Zhou et al. 2004), which make QTBA a candidate method to use in conjunction with RTWC to identify spurious source locations. The spatial resolution is not as good as RTWC, perhaps due to the assumption of constant dispersion (Han et al. 2007). In addition, the standard deviation coefficient for the trajectory error must be defined. While recommendations exist in the literature for this coefficient when applied to continental scale problems, there are no readily obvious values to apply on the building scale since the flows differ greatly between the two scales.

## 2.6 Discussion and Method Selection

Though investigating and comparing each method for the current study would be desirable, this was far beyond the possible scope of this work. Thus, it was necessary to select a limited number of potential approaches for further investigation.

The central question in this project is not whether it is possible to locate sources from downstream concentration measurements, but whether it is possible to do so in a

reasonable amount of time for the scale and geometry of the problem with a limited sensor grid. Unfortunately, the type of obstacles and flow regimes within the domain increase the cost of performing transport model calculations, as the assumptions around which the faster models (such as Gaussian plumes) do not necessarily apply. Increasing the number of required transport calculations would therefore have a significant effect on the cost and speed of the algorithm. The more complex source optimization models with large number of these forward simulations may work, but their use may not be justified in relation to the difference in costs compared with simpler models provided that the simpler models are sufficient.

Forward based methods require a large number of forward simulations, though this may be reduced as knowledge of the domain is increased. As stated in Chapter 1, there are thousands of potential sources within a plant. BMC needs simulations of all the potential sources. Kalman filters and simulated annealing are time intensive due to the computational cost as well. Genetic algorithms, too, require a fairly large search space in the case of multiple receptors and sources.

To reduce the number of transport simulations, receptor oriented approaches were considered instead of forward transport based approaches. This left variational data assimilation and the trajectory statistic approaches. Variational data assimilation utilizes a combination of adjoint and forward transport models, which should reduce the number of transport simulations required. However, this approach is still more complex and costly than the trajectory statistical methods. The primary advantages of the trajectory statistical methods are their calculation simplicity, low computational cost, and their lack of reliance on advance knowledge of the sources.

Based on the above premise of "as simple as possible, but no simpler", the trajectory statistical methods were selected for evaluation and comparison as they require no domain knowledge and, if adequate location estimates are achieved, provide a significant computational advantage over other models. Adequate location estimates would also demonstrate their applicability to small scale problems.

### **3 Review of Trajectory Statistical Methods**

Over the past several decades, multiple methods have been developed to use trajectories and pollutant sensor measurements to determine the origin of pollutants. Many of these have been widely used on regional and continental scales, though there is no theoretical reason that they could not be used on a much smaller domain size.

#### **3.1 Review of TSM development and validation studies**

Much of the early work on relating downstream measurements with upstream sources was focussed on source apportionment based on assessing relative contributions of various known sources to measured pollutant levels via elemental analysis on collected samples (e.g. Blifford & Meaker, 1967; Miller et al., 1972; Hopke et al., 1976). Though this type of approach allowed logic to be used to hypothesize likely sources, pollutant source locations were not being solved for explicitly. Meteorological information, however, is an important part of interpreting changes in trace gases, aerosols, etc. at measurement sites, as noted by Miller (1981).

Miller (1981) classified a large number of trajectories from the Mauna Loa Observatory in Hawaii into directional sectors and speed and suggested that these trajectories could be used to perform comparisons between measurements of gases and aerosols and climatology variations. Multiple authors subsequently used trajectories to link precipitation chemistry and the air mass movement. Henderson and Weingartner (1982) classified trajectories in quadrants and time periods, and related these to precipitation acidity. Colin et al. (1989) used directional sectors separated by industrial activity to analyze acidity of snow and rain precipitation. Miller et al. (1993) used

directional sectors to relate direction with radon measurements and precipitation amounts. Moody and Samson (1989) and Harris and Kahl (1990) replaced the “somewhat arbitrary trajectory classification by compass sector” by cluster analysis, which is an automated multivariate statistical method to split the trajectories into groups by a specified distance measure, such as the spatial distance between trajectory points at given times (Moody & Samson, 1989). Unfortunately, it is difficult to determine source areas simply from the relations between trajectory clusters and concentration measurements, especially if there are multiple sources (Stohl, 1996).

During the same time period as Miller (1981), Pitchford et al. (1981) analyzed data from Project VIEW (Visibility Investigative Experiment in the West), which measured visibility parameters and particle concentrations for multiple locations in the Midwestern United States. Beginning with principle component analysis and data comparison for measurements taken at the majority of network sites in the fall of 1979, Pitchford et al. (1981) performed additional analysis on the best and worst visibility days at the Grand Canyon. The causes of each of these days were hypothesised by computing average trajectories back in time, based on the idea that clean and dirty air may originate along specific corridors. Their findings revealed that during the majority of the high visibility days, arriving trajectories of air originated from the North and Northwest over remote and unpopulated areas, while during most of the worst days, trajectories came from the West, Southwest and Southeast. In the exceptions, trajectory velocity was suggested as a factor since the highly visible trajectories had higher speeds (and thus spent less time over source regions) than the lower visibility trajectories.

This idea of time spent over source regions was used by Ashbaugh et al. (1985) in the development of a statistical method that used back trajectories to identify source regions. Residence time analysis, based on the Ph. D. dissertation of Ashbaugh (1982), is the estimation of the probability that a randomly selected air parcel was in a location during a specified time duration. This may also be expressed as the spatial probability distribution of previous positions for air arriving at a location during a certain time period. The analysis can be performed for all trajectories and measurements, or special case trajectories that correspond to concentrations over a selected cut-off value.

However, just using the residence time of trajectories associated with high concentration samples gives a peak at receptor locations, since all receptor based trajectories must end at the receptor. Thus, without some sort of correction or normalization, the receptors are always predicted to be sources.

There are two main approaches to prevent the algorithm from predicting source locations at receptors. One method is to normalize the computed probability field with a hypothetical probability field, such as a hypothetical symmetric peak at the receptors (Ashbaugh et al., 1985). The other method invokes a conditional probability function, also known as the potential source contribution function (PSCF) in literature (see 2.5.1), which normalizes the high pollutant residence time function by the overall residence time function. Ashbaugh et al. (1985) used PSCF to determine contributing regions for high sulphur episodes in 1980 at Grand Canyon National Park and found agreement with directional trends noticed by Pitchford et al. (1981).

Zeng and Hopke (1989) analyzed acid rain contribution regions using factor analysis including back trajectory information and the PSCF method. They noted that if

a cell contains a single endpoint, but it happens to be polluted, the PSCF value will be 1.0 even though the confidence in that value is low. Rather than simply discard the values in these cells, a weighting function was used where an endpoint count of  $n_{ijk} = 1$  has a weight of 0.5,  $n_{ijk} = 2$  has a weight of 0.68,  $n_{ijk} = 3$  has a weight of 0.85 and  $n_{ijk} \geq 4$  a weight of 1.0. Zeng and Hopke (1989) combined PSCF runs based on data from different sensors by multiplying the resulting cell values together to show the source areas common to both runs. Regions appearing on multiple PSCF plots for various single receptors were emphasized. As these high regions existed in multiple measurement sets, they were considered to be important source regions. However, this requires PSCF results to exist in these areas for each receptor measurement set. Thus, this is similar, but distinct, from joint probabilities calculated by pooling data sets to improve the trajectory coverage (and hence improve the trajectory statistics) of the domain from various upwind directions (e.g. Hsu et al. (2003) and Peng et al. (2007)). This latter approach allows PSCF to be calculated in a larger area and with increased confidence.

Cheng et al. (1993) used a similar weighting function to the one by Zeng and Hopke (1989). They also introduced a method of integrating multiple trajectory height releases for the same measurement based on the assumption of good mixing in the planetary boundary layer and the idea of total probability. Cheng et al. (1993) found that there was no systematic method of choosing an optimal pollutant criterion in advance, but comparing maps with multiple values was useful for observing the source spatial patterns.

Vasconcelos et al. (1996a, 1996b) investigated the statistics of the PSCF method and the influence of the trajectory variability on the results. For example, if air that passes over a source region always passes over a neighbouring cell and vice versa, it is impossible to tell which of the cells contains the source. Using a single receptor and calculated back trajectories, imaginary sources were selected and trajectories passing over them were considered polluted. With a single sensor, the predicted source locations were smeared out in the radial direction, and in general varying angular and radial source locations showed good angular, but poor radial resolution. This is likely due to the convergence of trajectories about the single starting point. Vasconcelos et al. (1996b) also noted the applicability of bootstrap resampling methods, in addition to binomial tests based on endpoint numbers, to determine the statistical significance of the PSCF results. With bootstrapping, a large number of PSCF fields are calculated from random sample subsets of the original data set in order to estimate the standard deviation of each cell, which is used to determine the confidence in the resulting field estimate. This does not require any assumptions to be made about the sample distribution, unlike the binomial tests which ignore the interdependence of points along a single trajectory.

Polissar et al. (2001) empirically determined a PSCF weighting function based on the ideas of Zeng and Hopke (1989) and Cheng et al. (1993). They found that using an upper cut-off for the weighting function at approximately 2.5-3 times the average number of end points within a cell worked well.

A study by Lin et al. (2001) concluded that increasing temporal resolution (and thus increasing the number of measurement and trajectory pairs) improved PSCF resolution and continued to identify major source regions. Later that year, Cheng and Lin

(2001) looked at verifying PSCF using the 1998 forest fires in South America as a well known source. They found that using a higher pollutant criterion value (90<sup>th</sup> percentile as compared with the 50<sup>th</sup> and 75<sup>th</sup> percentiles often used in previous studies) reduced the risk of source misidentification and that PSCF plots for lower values should be compared with trajectory plots before conclusions are drawn. They concluded that source locations could be found successfully with carefully chosen inputs and were improved with higher criterion values and longer data sets, but the model results could not be taken for granted and should be inspected in conjunction with information provided by cluster analysis or back trajectory plots.

Lee and Ashbaugh (2007) compared single receptor PSCF with a variation that used trajectory hits from multiple receptors on a cell rather than segment points. In essence, instead of basing everything on the residence time in the cell, this modified approach is based on the residence time of the trajectory at a receptor. This new method was able to identify more sources than trajectory point single receptor PSCF, but the increased number of receptors was found to be the most important factor in this improvement.

Keeler et al. (1989) investigated if back trajectories could be used to determine systematic regional differences for unique tracer elements in the North Eastern United States from measurements at 6 sites in August 1983. Rather than using existing methods, they created the quantitative transport bias analysis (QTBA) algorithm (see 2.5.3) based on the idea that the back trajectories are uncertain. In this approach, each backward trajectory point was treated as a normally distributed “puff” and the resulting mass transfer probability weighed by the measured concentration accompanying the trajectory.

It was also noted that using a single receptor location would show preferred pathways, but would not address distance of the source from the receptor, which requires the use of multiple receptors. The study was able to show the systematic differences for each of the tracer elements (primary differences in arrival directions), validating the QTBA approach.

PSCF has difficulty determining the relative importance of sources. Seibert et al. (1994) developed the concentration weighted trajectory method (CWT), which would allow for the extraction of additional information (see 2.5.2). The original CWT uses the logarithmic mean time-weighted concentration in each grid cell, with higher concentrations on the grid indicating regions more likely to be sources. In addition, Seibert et al. (1994) used a point filter to smooth the field while restricting CWT field values to within a 95% confidence interval to remove small scale variations that had limited statistical significance, though simply discarding cells with insufficient trajectories is typical in later studies (e.g. Lupu & Maenhaut (2002) and Hsu et al. (2003)). Variations that weight the field using the mean time-weighted concentration rather than the logarithmic mean also exist (Hsu et al, 2003).

The CWT method does neglect some information available in the data set. Stohl (1996) noted that air pollutants are usually concentrated in 'hot-spots' while CWT equally attributes concentrations along entire trajectories. Imagine a single grid cell in the domain through which multiple trajectories pass, only one of which has a high accompanying concentration measurement while the rest correspond to low concentrations. The clean trajectories probably don't have a source along their paths, including in the cell they share with the polluted trajectory. Thus, the polluted trajectory

must have acquired its pollutant load somewhere else along its length. The residence time weighted concentration (RTWC) (sometimes also called redistributed time weighted concentration) redistributes the concentration field based on this idea (for details, see 2.5.2). This is an iterative process until a chosen convergence criteria (for example, average difference below 0.5% between iterations) is reached.

Wotawa and Kroger (1999) investigated CWT and RTWC by using a box model to create a synthetic data case and isolate the trajectory statistic methods from measurement and trajectory errors, as well as a data set with artificial errors introduced. General features were reproduced by both methods. It was noted that accuracy improved with increased numbers of trajectories and measurements, particularly for shorter length data sets in which CWT and RTWC initially performed poorly. Thus, using methods over longer periods with more data to create stable results was recommended. Also, improving cell counting statistics improved results and including additional receptors near areas of interest was recommended when the domain was increased in size. In areas with low trajectory statistics and significance levels, unrealistic structures tended to occur. In addition, representativeness of the trajectories to the actual transport and the measurement accuracy were found to be critical to estimating spatial emission distribution to both methods.

Lupu and Maenhaut (2002) looked at PSCF and CWT methods together and implemented the bootstrapping suggested by Vasconcelos et al. (1996b) to determine the PSCF significance. They also considered the ability to test the "smearing" of PSCF with ideal sources, as discussed in Vasconcelos et al. (1996b), an advantage which made it complimentary to other methods when determining if it was possible to distinguish the

source location from up and downstream locations. However, bootstrapping could not overcome this smearing of results up and downstream from a source. While unimplemented due to the receptor sites available, it was also noted that increasing the sensor locations to improve trajectory statistics from multiple directions was the most straightforward method of improving the accuracy of the predictions.

Hsu et al. (2003) compared PSCF, CWT and RTWC. Rather than use a single receptor for PSCF, Hsu et al. (2003) used a joint PSCF method and multiple receptors for CWT and RTWC which pooled data sets from multiple receptors to improve trajectory statistics. All three methods indicated similar source locations, but had different advantages and disadvantages suggesting that all three should be calculated to provide more information about the potential source locations. Setting the cut-off parameter for PSCF lower allowed moderate, as well as large, source regions to be identified. CWT was found to be useful in distinguishing between moderate and large sources. RTWC reduced trailing effects seen in the PSCF and CWT results, as did pooling data sets. While consistent results were found for some sources, others could not be found without further up and downstream measurements to narrow the source location.

Zhou et al. (2004) compared QTBA and RTWC. In addition, empirical down-weighting was used for the QTBA method as an equivalent to the filters on the RTWC field that reduce trailing. They found that for large, clearly defined sources, both methods gave consistent results. QTBA required less spatial resolution for reliable results, but had a trailing effect. RTWC allowed for better spatial resolution and was able to find minor source regions, but was sensitive to such things as varying emission rate

and the authors recommended using it in conjunction with non-redistributing methods such as PSCF, QTBA or CWT.

For Zhao et al. (2007) , the complicated QTBA empirical filter of Zhou et al. (2004) was replaced by a simple weighting function similar to that used widely for PSCF fields (e.g. Zeng and Hopke (1989)). This weighting function was found to be an effective alternative to the more complicated empirical filter. An RTWC cut-off for minimum trajectories of half of the mean trajectory crossings was suggested. RTWC was found to reduce the predicted high regions to a number of hot spots which in QTBA were represented as a single high region.

Han et al. (2007) performed a quantitative comparison using the Spearman correlation index and a spatial correlation index against emissions inventories for multisite PSCF, RTWC and QTBA methods. Each showed positive statistical correlations, but multisite PSCF produced the strongest correlation. It was suggested, however, that this may have been due to uncertainties in the inventory itself. RTWC was able to distinguish contributions from very high and low source concentration regions, while PSCF and QTBA could not.

Scheifinger and Kaiser (2007) noted that the majority of works to validate trajectory statistical methods investigated PSCF, CWT and RTWC methods used subjective comparison with known sources rather than more quantitative approaches (see Table 3.1 below). Since the volume source information consisted of only regions inside and outside the source area, rather than use a spatial correlation index, a 'hit ratio' which compared the mean weighted residence time inside and outside the volume sources was

used by Scheifinger and Kaiser (2007) as a comparison parameter. This 'hit ratio' relies on an area source, and cannot be implemented for a point emission source. TSMs were found to be able to reconstruct isolated source positions for virtual source areas, especially RTWC. However, comparison with real world data showed decreased performance compared with the virtual release, and the difference was attributed to a maximum useful transport time, above which gas removal in the real-world setting became important. For example, SO<sub>2</sub> experienced a useful range of 2.5 days in the Scheifinger and Kaiser (2007) experiments of continental scale emission reconstruction. This scale of trajectory is far beyond the building scale problem under current consideration.

### **3.2 Summary and General Observations from Previous TSM Studies**

Overall, each of the algorithms (PSCF, CWT, RTWC, and QTBA) was found to be useful in determining important emission source locations for continental scale problems. None, however, have been used for smaller scale source location estimation. The studies can still provide valuable guidance for their use for building scale pollutant source location as well as reveal opportunities for further improvement of algorithms.

Although RTWC was able to reconstruct more complex scenarios, it also tended to identify false source regions so none of the algorithms can be immediately eliminated. In fact, previous work suggests that the combined use of multiple methods can be beneficial. Looking at the initial parameters which cause such false identifications may also be useful in determining which algorithm to trust depending on circumstances.

As noted multiple times, multiple receptors can be used to reduce the trailing effect that appears in PSCF, CWT and RTWC and improve results. Extending amount of time used in data gathering to produce larger data sets can also provide more consistent results. Neither effect should change when moved to a smaller scale problem. Thus, a larger network of sensors at various locations around the plant should provide better information than using only one or a few sensors. The importance of sufficient trajectories was iterated several times with several different filters used in an attempt to limit false predictions regions with poor trajectory statistics. Further research is required to determine whether filters can be used to eliminate these regions and whether their use will change the overall predictions when dealing with a large multi-sensor network.

For PSCF specifically, the choice of the pollutant criterion has been shown to be important across multiple studies. High cut-off percentiles seem to provide better ability to locate the more important sources. However, as PSCF has most often been run using single sensors, in a multi-sensor network, where most sensors would not be expected to register high concentrations, it is unclear how high should this criterion be and is another issue to be investigated. Since multiplying multiple PSCF fields requires PSCF information throughout the domain, a pooled data set method is more applicable to the small scale problem under consideration in the present work to reduce time required to obtain measurements and trajectories.

While most previous algorithm comparisons (see Table 3.1) were largely qualitative, a few quantitative comparisons have been completed. Hit ratio is not applicable to point sources, but the use of correlation coefficients is a useful approach that can be applied in the present work. Comparison based on quantitative correlation

statistics provides a method of determining relative performance of each algorithm under the influence of different input parameters without inserting observer bias.

**Table 3.1: Some select Trajectory Statistical Method evaluation studies**

<b>Trajectory Statistical Method</b>	<b>Evaluation Method</b>	<b>Region</b>	<b>Species measured</b>	<b>Result type</b>	<b>Authors</b>
PSCF	Comparison with known sources	North America	Particulate sulphate	Qualitative	Cheng et al. (1993)
Modified PSCF	Comparison between imaginary sources and single receptor	South-western United States	Virtual pollutant	Qualitative	Vasconcelos et al. (1996b)
CWT, RTWC	Comparison with EMEP emission inventory	Central Europe	NO <sub>x</sub>	Spearman rank order correlation	Wotawa and Kroger (1999)
CWT	Comparison with EMEP inventory	Europe	SO <sub>2</sub> , SO <sub>4</sub> <sup>-2</sup> , NO <sub>x</sub> , NO <sub>x</sub> <sup>-</sup> , NH <sub>3</sub>	Spearman rank order correlation	Charron et al. (2000)
PSCF	Comparison with known sources	Northern hemisphere	Condensation nuclei counts, black carbon, aerosol light scattering	Qualitative	Polissar et al. (2001)
PSCF, RTWC	Comparison with known sources	Europe, Middle East, Northern Africa	Sulphur, arsenic, vanadium, PM, BC, Zn, NH <sub>4</sub> <sup>+</sup> , Pb, NO <sub>3</sub> <sup>-</sup>	Qualitative	Lupu and Maenhaut (2001)
PSCF	Comparison with known sources	North America	Optical aerosol properties and aerosol counts	Qualitative	Cheng and Lin (2001)

<b>Trajectory Statistical Method</b>	<b>Evaluation Method</b>	<b>Region</b>	<b>Species measured</b>	<b>Result type</b>	<b>Authors</b>
RTWC	Comparison with virtual sources & sinks	Europe	CO <sub>2</sub>	Qualitative	Apadula et al. (2003)
RTWC, QTBA	Comparison with known sources	North-eastern United States, Ontario, Quebec	Particles	Qualitative	Zhou et al. (2004)
PSCF, CWT, RTWC	Comparison with virtual source; Comparison with EMEP emission inventory	Europe	Virtual pollutant SO <sub>2</sub>	Hit ratio; Spearman rank order correlation; Normalized difference field between results and inventory	Scheifinger and Kaiser (2007)
RTWC, QTBA	Comparison with known sources	Eastern United States	Nitrate, sulphate	Qualitative	Zhao et al. (2007)
PSCF, modified PSCF	Comparison with virtual sources	United States, Mexico	Virtual SO <sub>2</sub>	Number of sources correctly identified	Lee and Ashbaugh (2007)
PSCF, RTWC, QTBA	Comparison with emissions inventory	United States and Southern Canada	Atmospheric mercury	Spearman rank order correlation; Spatial correlation index	Han et al. (2007)

## **4 CFD Simulation**

### **4.1 General Approach**

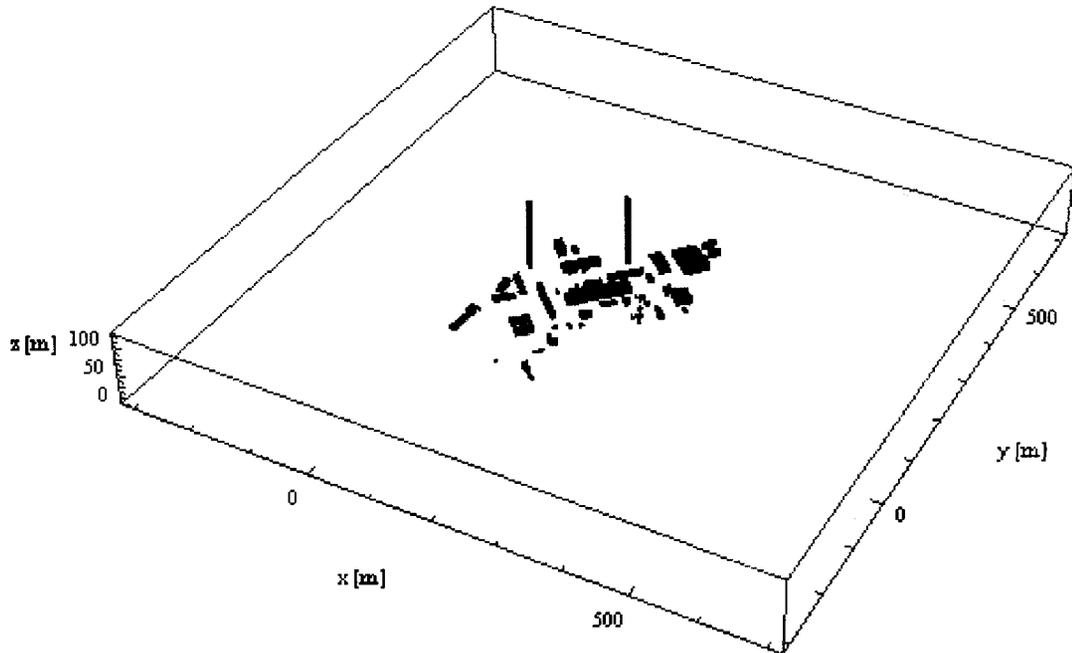
To evaluate a range of source location algorithms, time-resolved velocity and concentration data were required at various points throughout the domain for a range of test conditions. Test data were generated using computational fluid dynamics (CFD) simulations of wind flow and gas releases over a simplified geometry of an actual operational gas plant in Alberta, Canada. CFD makes it possible to capture the influence of plant geometry on the wind flow and gas dispersion. The use synthetic data in an initial simplified simulation allowed for a controlled evaluation of potential algorithms, albeit in an idealized scenario. In this approach, the potential influence of errors related to the accuracy of predicted wind flow and subsequently determined upwind trajectories was not explicitly considered.

The first step in performing a CFD simulation is to define the domain within which the flow is modelled. A geometry file is used to describe this region and the obstructions within it. This domain is divided into discrete elements (cells), where increasing domain resolution increases the knowledge of the domain at the expense of increased computational resource requirements. The problem physics are then defined such as the properties of the fluid and turbulence conditions, as well as the conditions on the boundaries of the domain. In the current case, ANSYS CFX version 12.1 was used as the flow solver. CFX uses an iterative solution process on the variables of interest at each cell within the domain and can compute steady-state or transient solutions, for which the choice of timestep determines the temporal resolution. Since flows around

buildings and in urban areas tend to be unsteady, transient simulations are generally required even with non-varying wind conditions. A detailed description of the chosen domain, wind, source conditions and the implementation of the simulation is described below.

## **4.2 Simulation Geometry**

Rather than use an arbitrary test geometry, the simplified geometry of an actual gas plant was chosen for initial testing. This simplified geometry was based on an operational gas plant located in Balzac, Alberta and can be seen in Figure 4.1. The plant consists of two 122 m tall stacks and approximately seventy smaller buildings with an average height of 10.5 m. The geometry was provided in the form of building vertex coordinates and heights, limiting the details in the geometry generated and neglecting the presence of pipes and other smaller objects. However, use of simplified geometry is not unusual in wind simulations which often use only the building geometry rather than further simulation detail (Mochida & Lun, 2008).



**Figure 4.1: Simulation boundaries containing the simplified gas plant geometry**

### **4.3 CFX Model Implementation**

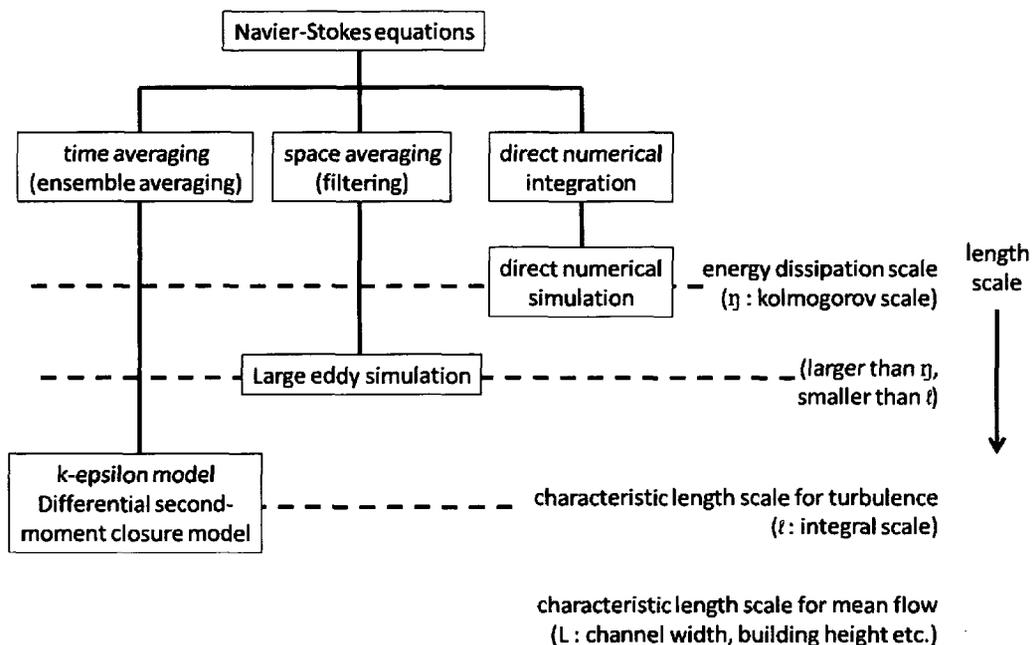
CFX is a general purpose finite-volume CFD code that allows both structured and unstructured grids, multiple turbulence models, steady and unsteady flow cases, and various numerical schemes. Further details of the solution methodology are available in the CFX User Manual (ANSYS, 2008).

#### **4.3.1 Simulation Model Type**

Wind modelling is a growing field in computational fluid dynamics with its own unique challenges. Determining the best model type is an important first step in simulating such an exterior flow. A schematic of the major model types is shown in Figure 4.2 below. In principle, direct numerical simulation (DNS) would be the ideal method. However, the computational resources necessary are prohibitive in practical applications such as wind

engineering simulations of flow over a group of buildings (Ferziger, 1993; Speziale, 1998).

More practical wind models exist for the purpose of computational wind engineering (CWE). Linear models based on Jackson and Hunt (1975) can be useful when estimating locations for wind plants, but do not match the detail and accuracy of more complex models and are not recommended for other applications. Although large eddy simulation (LES) is often a preferred approach in simpler problems where sufficient computational resources are available, for most purposes Reynolds Averaged Navier-Stokes (RANS) models remain the only practical method (Rodi, 1997; Tominaga et al., 2008). Detached eddy simulation (DES) models are a possible compromise, but are more difficult to implement and increase computational time over RANS models (Hasama et al., 2008). RANS models remain an industry standard for wind engineering due to their simplicity, robustness and computational economy (Hanjalic & Kenjeres, 2008). Due to its place as an industry standard, its computational advantages and its availability, RANS modelling was chosen to determine the synthetic wind flow. A description of the equations solved in the RANS model is given below.



**Figure 4.2: Methods of numerical simulation and turbulence models (Murakami et al., 1994)**

#### 4.3.1.1 RANS Equations

As first shown by Reynolds (1895), the instantaneous velocity,  $u_i$ , can be represented statistically as the sum of mean,  $U_i$ , and fluctuating,  $u_i'$ , components:

$$u_i = U_i + u_i' \quad (4.1)$$

Following the summary given by Wilcox (1994) and Narasimhamurthy (2004), this statistical model can be applied to the continuity and momentum equations for an incompressible flow. The basic form of these equation is show in (4.1) and (4.2).

Continuity:

$$\frac{\partial u_i}{\partial x_i} = 0 \quad (4.2)$$

Momentum:

$$\rho \frac{\partial u_i}{\partial t} + \rho u_j \frac{\partial u_i}{\partial x_j} = -\frac{\partial p}{\partial x_i} + \frac{\partial}{\partial x_j} \left( \mu \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \right) \quad (4.3)$$

Combining 4.2 and 4.3, the conservation form of the Navier-Stokes equations is

$$\rho \frac{\partial u_i}{\partial t} + \rho \frac{\partial}{\partial x_j} (u_j u_i) = -\frac{\partial p}{\partial x_i} + \frac{\partial}{\partial x_j} \left( \mu \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \right) \quad (4.4)$$

The ensemble averaged continuity equation is:

$$\frac{\partial U_i}{\partial x_i} = 0 \quad (4.5)$$

Subtracting Equation (4.5) from the continuity equation shows that the fluctuating velocity term must have zero divergence. Taking the ensemble average of the momentum equation gives

$$\rho \frac{\partial U_i}{\partial t} + \rho \frac{\partial}{\partial x_j} (U_j U_i + \overline{u_j' u_i'}) = -\frac{\partial P}{\partial x_i} + \frac{\partial}{\partial x_j} \left( \mu \left( \frac{\partial U_i}{\partial x_j} + \frac{\partial U_j}{\partial x_i} \right) \right) \quad (4.6)$$

Other than the replacement of the instantaneous variables with their mean values, the only difference between the averaged equation and the original equation is the term  $\overline{u_j' u_i'}$  (Wilcox, 1994). Equation (4.6) can be rearranged into the familiar form of the

Reynolds-averaged Navier-Stokes equation:

$$\rho \frac{\partial U_i}{\partial t} + \rho U_j \frac{\partial U_i}{\partial x_j} = -\frac{\partial P}{\partial x_i} + \frac{\partial}{\partial x_j} \left( \mu \left( \frac{\partial U_i}{\partial x_j} + \frac{\partial U_j}{\partial x_i} \right) - \overline{\rho u_j' u_i'} \right) \quad (4.7)$$

The last term is known as the Reynolds-stress tensor:

$$\tau_{ij} = -\overline{\rho u_j' u_i'} \quad (4.8)$$

In the case of the unsteady RANS, the typical RANS equations are maintained along with the time dependant  $\frac{\partial U_i}{\partial t}$  term (Narasimhamurthy, 2004). It is important to note that defining the averages as ensemble means retains dependency on time (Narasimhamurthy, 2004). The velocity described in Equation (4.1) can be decomposed into

$$u_i = U_i + u_i' = \langle U_i \rangle + u_i'' + u_i' \quad (4.9)$$

where  $\langle U_i \rangle$  is the time-averaged velocity,  $u_i''$  is the resolved fluctuation and  $u_i'$  is the modelled turbulence (Narasimhamurthy, 2004).

The Reynolds-stress tensor has six independent components which are new unknown quantities for the system of equations. In a three-dimensional flow, there are four unknown mean properties (pressure and three velocity components) in addition to these Reynolds-stress terms for a total of ten unknowns. Mass conservation and the various components of Equation (4.7) give a total of four equations only, so the system is not closed and cannot be solved without additional information (Wilcox, 1994). Thus, a turbulence model is required.

### 4.3.2 Turbulence Model Selection

The atmospheric boundary layer of the Earth is always turbulent and the bluff bodies typical in urban flow situations require consideration when determining the turbulence model. For the RANS modelling of wind, the k-epsilon turbulence model has been tested often enough to reveal an over-estimation of turbulence energy in stagnation regions (Franke, 2006; Tominaga et al., 2008). This has led to the development of other models to improve simulation results. One model that decreases the production of turbulence energy in this region is the k-epsilon RNG model.

#### 4.3.2.1 k-epsilon and k-epsilon RNG models

Both k-epsilon and k-epsilon RNG models are considered two-equation models since they characterize not only the turbulence kinetic energy in the form of a transport equation, but also the size of the large, energy-containing eddies (Mankbadi, 1994). The most widely used two-equation model is the k-epsilon model. The central work on the model was performed by Jones and Launder (1972).

The turbulence kinetic energy (per unit mass),  $k$ , can be defined as (Wilcox, 1994)

$$k = \frac{1}{2} \overline{u_i' u_i'} = \frac{1}{2} (\overline{u'^2} + \overline{v'^2} + \overline{w'^2}) \quad (4.10)$$

The velocity scale is defined by  $\sqrt{k}$  (Mankbadi, 1994) and the length scale by

$\ell = C_\mu k^{3/2} / \varepsilon$  (Wilcox, 1994). The eddy viscosity is thus (Wilcox, 1994)

$$\mu_T = \rho C_\mu k^2 / \varepsilon \quad (4.11)$$

and the turbulence kinetic energy and dissipation rate,  $\varepsilon$ , correspond respectively to

$$\rho \frac{\partial k}{\partial t} + \rho U_j \frac{\partial k}{\partial x_j} = \tau_{ij} \frac{\partial U_i}{\partial x_j} - \rho \varepsilon + \frac{\partial}{\partial x_j} \left( (\mu + \mu_T / \sigma_k) \frac{\partial k}{\partial x_j} \right) \quad (4.12)$$

$$\rho \frac{\partial \varepsilon}{\partial t} + \rho U_j \frac{\partial \varepsilon}{\partial x_j} = C_{\varepsilon 1} \frac{\varepsilon}{k} \tau_{ij} \frac{\partial U_i}{\partial x_j} - C_{\varepsilon 2} \rho \frac{\varepsilon^2}{k} + \frac{\partial}{\partial x_j} \left( (\mu + \mu_T / \sigma_\varepsilon) \frac{\partial \varepsilon}{\partial x_j} \right) \quad (4.13)$$

with typical closure coefficients of (Versteeg & Malalsekera, 1995):

$$C_{\varepsilon 1} = 1.44 \quad C_{\varepsilon 2} = 1.92 \quad C_\mu = 0.09 \quad \sigma_k = 1.0 \quad \sigma_\varepsilon = 1.3$$

The k-epsilon model is probably the most popular RANS turbulence model and has been widely used and verified for many varieties of flows. Compared with the more complex stress models, the k-epsilon model is computationally inexpensive and tends to prove successful in applications where normal Reynolds stresses are relatively unimportant (Easom, 2000). However, there are weaknesses that become apparent when the k-epsilon model is applied to wind engineering.

For flows around a cube, the k-epsilon model always predicts a positive turbulence energy production term, even in regions where the value should be negative (Murakami, 1990). In the recirculation region and wake, the turbulence is anisotropic, but this cannot be reproduced (Murakami, 1990) due to the definition of  $k$  (see Equation (4.10)). The stream-wise normal stress dominates on the roof of a cube and the lateral stress dominates in the wake and recirculation zone (Murakami, 1990). The error is itself rooted in the isotropic eddy viscosity concept (Easom, 2000). These issues are well known and have led to the development of other models for their use in wind engineering.

Yakhot and Orzag (1986), performed renormalisation group analysis of the Navier-Stokes equations, thus the name RNG model. The transport equations for turbulence generation and dissipation match the standard k-epsilon model, but the constants differ (Stangroom, 2004).

$$\rho \frac{\partial \varepsilon}{\partial t} + \rho U_j \frac{\partial \varepsilon}{\partial x_j} = (C_{\varepsilon 1} - C_{s1RNG}) \frac{\varepsilon}{k} \tau_{ij} \frac{\partial U_i}{\partial x_j} - C_{\varepsilon 2} \rho \frac{\varepsilon^2}{k} + \frac{\partial}{\partial x_j} \left( (\mu + \mu_T / \sigma_\varepsilon) \frac{\partial \varepsilon}{\partial x_j} \right) \quad (4.14)$$

where

$$\begin{aligned} C_{s1RNG} &= \frac{\eta(1-\eta/\eta_0)}{1+\beta\eta^3} \\ \eta &= \left( 2S_{ij} \cdot S_{ij} \right)^{1/2} \frac{k}{\varepsilon} \\ \eta_0 &= 4.377 \\ \beta &= 0.012 \end{aligned} \quad (4.15)$$

with typical k-epsilon RNG closer coefficients of (Versteeg & Malalasekera, 1995):

$$C_{\varepsilon 1} = 1.42 \quad C_{\varepsilon 2} = 1.68 \quad C_\mu = 0.0845 \quad \sigma_k = 0.7179 \quad \sigma_\varepsilon = 0.7179$$

While this does not eliminate the isotropic assumption, the additional term in the dissipation equation enhances dissipation of turbulence kinetic energy, which is typically overproduced in the standard model near stagnation points (Jeong et al., 2002).

Kim and Patel (2000) found that the RNG model gave improved results with respect to separation length and flow profile for terrain modelling containing regions of separation and recirculation. They concluded that the model was a good candidate for modelling under neutral conditions (where the temperature lapse rate and dry adiabatic rate are equal and vertical motion is not accelerated/decelerated by temperature induced

buoyancy effects in the atmosphere). Jeong et al. (2002) found similar improvements for simulations around a bluff body.

Many commercial CFD codes, such as CFX, include the k-epsilon RNG model. While there are more complicated models with non-linear equations and non-isotropic eddy viscosity assumptions, these add additional computational cost to the solution and few examples of their accuracy for wind engineering problems are available (Tominaga et al., 2008). Thus, the RNG turbulence model was chosen to model the turbulence for the synthetic data set.

### **4.3.3 Initial Simplifications**

Some simplifications were necessary for model implementation. These include both domain and boundary profile shape simplifications, which are described in more detail below.

#### ***4.3.3.1 Vegetation and geometry simplification***

As no large vegetation was present at the gas plant location, any potential effects were neglected. Also, due to the simplified nature of the geometry data provided, other stationary objects such as piping outside of buildings were also ignored due to lack of information available. This assumption is used in most conventional wind assessments of urban areas, where only topographic features such as buildings are considered (Mochida & Lun, 2008). This is sufficient for the purposes of generating synthetic test data, but for the purposes of simulating wind flow in order to locate unknown leaks in an actual plant, stationary objects such as pipes outside of buildings, large vegetation or other such objects that affect wind flow may become important considerations. As this increases the

model complexity, the choice to include finer details of a simulation geometry is clearly dependant on the domain location and the predicted effects of these smaller, non-building objects. The addition of such objects is a current trend in urban flow modelling (Mochida & Lun, 2008).

#### ***4.3.3.2 Thermal Convective Currents***

If thermal emissivity varies greatly in the ground features, thermal convective mixing can have a significant influence on the flow field. Such mixing is one of the main causes of wind flow in general. On the smaller scale, an example would be the difference between the emissivity of a lake and the surrounding terrain. At night, air over a lake cools faster than air above a ground surface, causing a convective current. This can affect the turbulence generation, boundary layer profiles or even cause significant air flow in extreme cases. This is sometimes included in more complex models, such as canopy modelling of urban wind flows (Mochida & Lun, 2008), but it is in many cases neglected. To further simplify the wind flow and gas releases, the domain was assumed to be isothermal and wind currents due to heat differences were neglected.

#### ***4.3.3.3 Velocity Profile***

##### *Logarithmic Law Wind Profile*

The “logarithmic law” derived originally from the turbulent boundary layer on a flat plate is generally considered the most accurate representation of wind speed variation with height above the ground up to a height of approximately 100-200 m (Holmes, 2001).

Assuming that the change of wind speed with height ( $u(z)$ ) is a function of height ( $z$ ), surface shear stress ( $\tau_0$ ) and the density of air ( $\rho_{air}$ ) only, and that because the wind

is turbulent the effects of molecular viscosity are negligible, a non-dimensional version of the wind shear can be expressed as (Holmes, 2001):

$$\frac{du(z)}{dz} z \sqrt{\frac{\rho_{air}}{\tau_0}} \quad (4.16)$$

The inverse of the last term,  $\sqrt{\frac{\tau_0}{\rho_{air}}}$ , is known as the frictional velocity ( $u_*$ ). Setting this

equal to a constant ( $1/k_{vK}$ ) gives (Holmes, 2001):

$$\frac{du(z)}{dz} \frac{z}{u_*} = \frac{1}{k_{vK}} \quad (4.17)$$

Integrating, with the roughness length  $z_0$  as the integration constant, results in the expression (Holmes, 2001):

$$u(z) = \frac{u_*}{k_{vK}} \ln \left[ \frac{z}{z_0} \right] \quad (4.18)$$

In the case of very rough areas such as forests, there is often an effective height displacement. Thus, the mean wind velocity as a function of height can be expressed as (Holmes, 2001):

$$u(z) = \frac{u_*}{k_{vK}} \ln \left[ \frac{z - z_h}{z_0} \right] \quad (4.19)$$

where

$u_*$  is the friction velocity,

$k_{vK}$  is the von Karman's constant (experimentally has a value of approximately 0.4),

$z_h$  is the zero-plane displacement, or the effective height in cases of very rough terrain such as forests and urban areas, and

$z_0$  is the roughness length of the ground surface (between 0.01-0.05 m for open terrain such as the current domain).

While the logarithmic law is based on theory, it has mathematical characteristics which make it difficult to use. Thus, for practical applications, the alternative 'power law' approach is a common alternative.

#### *Power Law Wind Profile*

Although the power law wind profile is not based on theory, it is used because it produces a profile shape similar to that of the logarithmic model, while avoiding the mathematical difficulties arising from logarithms at zero and negative values. Relating the mean wind speed with that measured at 10 m, the power law is written as (Holmes, 2001):

$$u(z) = u_{10m} \left( \frac{z}{10m} \right)^\alpha \quad (4.20)$$

The exponent is an empirical function of the terrain roughness and the height at which the logarithmic and power law profiles are matched ( $z_{ref}$ ) (Holmes, 2001):

$$\alpha = \frac{1}{\ln(z_{ref} / z_0)} \quad (4.21)$$

For example, in a rural area with a roughness height of 0.04 m and matching the profiles at a 50 m height, the exponent becomes 0.14 or, approximately, 1/7. Due to ease of

implementation within the present simulation, the power law profile with the exponent of 1/7 was used to simulate a wind profile entering the domain.

#### ***4.3.3.4 Turbulence Intensity Profile***

The turbulence, or gustiness, of the wind is also a consideration. The turbulence intensity is the ratio of the standard deviation of each velocity component about its mean value.

Thus, the turbulence intensity may be defined as (Holmes, 2001):

$$I = \frac{\sigma_u}{U} \quad (4.22)$$

where  $\sigma_u$  is the standard deviation of the velocity and  $U$  is the mean velocity component.

Average turbulence intensities gathered from field measurements typically lie in the range of 15 to 20 percent (Jones, 1968). For neutral stability conditions, 15 percent is common, though this varies with roughness, and unstable conditions can have intensities up to 5 times that of stable conditions (Jones, 1968).

Turbulence intensity also varies with height. Since the purpose of the simulation is to generate synthetic data, an empirical relation of  $\sigma_u$  equal to  $2.5u_*$  was used in order to create an inlet profile (Holmes, 2001). Using the logarithmic law profile, the intensity can be given by (Holmes, 2001):

$$I_u = \frac{2.5u_*}{(u_* / 0.4) \ln(z / z_0)} = \frac{1}{\ln(z / z_0)} \quad (4.23)$$

While the empirical relation used was found initially for high wind conditions (Beaufort number<sup>1</sup> 7-10), the final equation has also been suggested in general for the reverse purpose of determining roughness height from turbulence intensity measurements for neutrally stable conditions (Wieringa, 1996). The resulting turbulence intensities fall within the range of expected values of Jones (1968) and were therefore used as a reasonable approximation of turbulence intensities for the synthetic data generation.

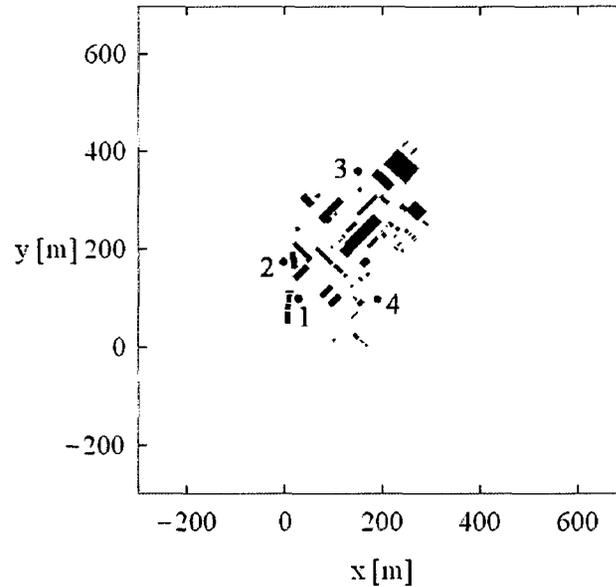
#### **4.3.4 Forward Pollutant Release**

##### ***4.3.4.1 Simulated Leak Sources***

Rather than separately solving the complete Navier-Stokes equations for all of the simulated releases, the virtual gas leaks were created using additional variables introduced to the flow domain at a given source rate and location. These variables acted as passively transported scalars which did not impact the wind flow field as the amount of leaked gas is very small in comparison with the bulk flow. For evaluation purposes, source locations were chosen as shown in Figure 4.3.

---

<sup>1</sup> Beaufort numbers refer to the measure on the international Beaufort scale of wind force ranging from 0 (calm) to 12 (hurricane).



**Figure 4.3: Source locations**

Two source strengths were used based on ranges of values which are typical in an oil and gas facility (based on a private communication with David Picard, Clearstone Engineering Ltd.). Locations 1 and 3 were selected as 'large' sources, while locations 2 and 4 were 'medium' sized sources (Table 4.1).

**Table 4.1: Source mass flow rates**

<b>Number</b>	<b>Type</b>	<b>Mass Flow [g/s]</b>
1	Large	4.5
2	Medium	0.45
3	Large	4.5
4	Medium	0.45

The sources were emitted as unique variables so they could be considered individually or in combination.

#### 4.3.4.2 Forward Simulation of Leaked Gas

The transport equation used to determine the motion of this scalar quantity (ANSYS, 2008) is the version of the advection-diffusion equation given in Equation (2.1).

$$\left[ \frac{\partial \phi}{\partial t} + \nabla \cdot (U\phi) \right] = \nabla \cdot \left[ \left( \rho D_{\phi} + \frac{\mu_t}{Sc_t} \right) \nabla \left( \frac{\phi}{\rho} \right) \right] + S_{\phi} \quad (4.24)$$

where

$\phi$  is the scalar (i.e. concentration),

$\rho$  is the mixture density,

$S_{\phi}$  is a volumetric source term,

$D_{\phi}$  is the kinematic diffusivity for the scalar,

$\mu_t$  is the turbulent viscosity, and

$Sc_t$  is the turbulent Schmidt number.

The kinematic diffusivity term must be specified to use of the transport equation.

As methane gas is a primary fugitive emission of interest in the oil and gas industry, its kinematic diffusivity was used. This was calculated to be  $2.2 \times 10^{-5} \text{ m}^2/\text{s}$  using the method of Turns (2000).

The use of the transport equation does not allow for the inclusion of buoyancy, so the gas was assumed to be neutrally buoyant. For methane, this is not the case.

However, it is possible to construct a simple estimate for the implications of this

assumption as detailed below. Depending on the gas being detected, the potential importance of buoyancy should be re-considered on a case-by-case basis.

#### 4.3.4.3 An Estimate for the Effect of Buoyancy on the Transport of Methane

Simple models of plume rise exist and can be calculated in advance to determine the significance of the buoyancy on the plume rise. First, the buoyancy flux of the plume must be determined. In cases where the plume gas has a very different specific heat and molecular mass than air, the buoyancy flux is defined as (Schnelle & Dey, 2000):

$$\begin{aligned}
 F_b &= \frac{\rho_p}{\rho_a} g U_p \frac{d_p^2}{4} \left[ \left( 1 - \frac{M_p}{M_a} \right) + \left( \frac{\rho_a}{\rho_p} - 1 \right) \left( \frac{Cp_p}{Cp_a} \right) \right] \\
 &= \frac{\dot{m}g}{\pi\rho_a} \left[ \left( 1 - \frac{M_p}{M_a} \right) + \left( \frac{\rho_a}{\rho_p} - 1 \right) \left( \frac{Cp_p}{Cp_a} \right) \right]
 \end{aligned}
 \tag{4.25}$$

where

$M_p$  is the molecular weight of the plume gases

$M_a$  is the molecular weight of air

$Cp_p$  is the specific heat of the plume gases

$Cp_a$  is the specific heat of the air

$\rho_p$  is the density of the plume gas

$\rho_a$  is the density of the air

$U_p$  is the initial velocity of the plume

$d_p$  is the diameter of the plume gas exit

$g$  is gravitational acceleration

$\dot{m}$  is the mass flow rate of the plume

Once the buoyancy flux is known, the gradual plume rise Briggs equation (Briggs, 1972) can be used to determine the plume rise at a specified distance downstream.

Taking  $U$  as the wind velocity at the plume exit and  $x$  as the distance downstream, the plume rise is (Turner, 1994):

$$\Delta h_B = 1.6 \frac{F_b^{1/3} x^{2/3}}{U} \quad (4.26)$$

There is, however, a distance at which the plume reaches a maximum height ( $x_*$ ). In cases where the distance downstream is greater than the final distance, the final distance value is used rather than the actual distance downstream. The distance at which the plume reaches the final height depends on atmospheric stability class. For neutral or unstable conditions (Pasquill stability classes A-D), the following can be used (Turner, 1994):

$$\begin{aligned} \text{If } F_b &\geq 55 \text{ m}^4 / \text{s}^3, \\ x_* &= 119 F_b^{2/5} [m] \\ \text{If } F_b &< 55 \text{ m}^4 / \text{s}^3, \\ x_* &= 49 F_b^{5/8} [m] \end{aligned} \quad (4.27)$$

In stable conditions (Pasquill stability classes E-G), the final distance requires some additional information about stability and the change of temperature potential with height. The final distance is (Turner, 1994):

$$x_* = 2.07 U s^{-1/2} [m] \quad (4.28)$$

The stability parameter is

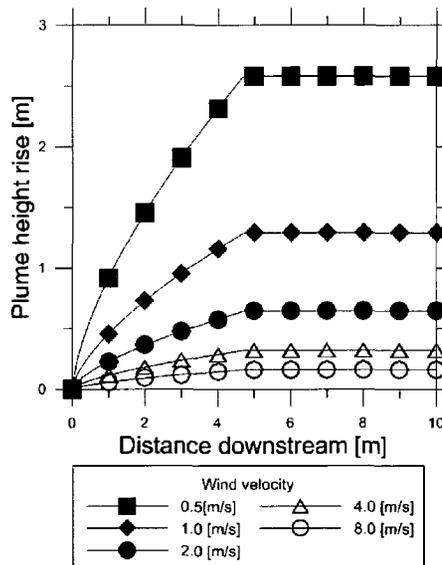
$$s = \left( g \frac{d\theta}{dz} \right) T \quad (4.29)$$

and the change of potential temperature with height is

$$\frac{d\theta}{dz} = \frac{dT}{dz} + \Gamma \quad (4.30)$$

where  $T$  is the ambient temperature and  $\Gamma$  is the adiabatic lapse rate (0.0098 K/m) for the standard atmosphere (Turner, 1994).

The predicted plume rise in an example case of 4.5 g/s release of methane into air at 20°C and 1 atm is given below in Figure 4.4 for an unstable or a neutrally stable atmosphere.



**Figure 4.4: Predicted buoyant plume rise for 4.5 g/s methane into air at 20°C and 1 atm for various wind speeds**

From Figure 4.4 it is apparent that in all cases the final rise of the large methane leak is not significant relative to the size of the domain and the resolution of the source location

algorithms (which is discussed in greater detail in Chapter 5). Thus, for the present case, it is quite reasonable to neglect the effects of buoyancy in the simulations.

#### 4.3.5 Domain Size

Determining the size of the domain is another important step in preparing the simulation. While a full domain size convergence study can be done for optimization purposes, recommendations exist to determine a starting size from the domain without one. In the case of the current domain, several recommendations were considered from various sources.

Blockage ratio area was kept under the minimum 3%. Taking  $H$  as the building height, a recommendation of  $5-8H$  in front of the buildings and  $15H$  behind (though this can be reduced in the case of multiple buildings) is recommended (Franke et al., 2007). As the wind acting on the domain changes direction, the mesh must have the minimum distance for outflow in each direction. In the case of the domain, the maximum height is the two stacks at 122 m. Following these recommendations for the stack heights creates a domain of several kilometres in length, which is unattainable for the simulation. However, the majority of the buildings and the area of interest are much smaller. The average building, excluding the stacks, is approximately 10 m high and of relatively small cross-section. This average height was taken as the height of interest. The domain was created with a length of 1 km, giving a minimum distance of 250 m in any direction from a building to the domain edge, which is an approximate ratio of at least  $25H_{\text{avg}}$ . A height of 200 m ( $20H_{\text{avg}}$ ) was used in the vertical direction.

Using a constant pressure outlet, simulation results from the North-East direction showed that even tens of meters upstream from the outlet, producing pressure differences in the order of only  $1 \times 10^{-2}$  Pa across cuts of the domain. For simulations of real geometries, it may be worthwhile to attempt a full domain size convergence study, but this was not considered necessary for the generation of purely synthetic data.

#### **4.3.6 Mesh**

The major determining issue of mesh generation for the computational domain was the compromise between sufficient resolution and the computational expense of the resulting simulation. In regions of high gradients, especially near the ground and buildings, or within wakes of building, smaller elements provided better resolution of the resulting flow field. This was also the region where gas leaks would be expected. In addition to the element size, the mesh quality needed to be considered. For the Balzac geometry, three meshes were created for the purposes of a grid convergence study, as discussed in further detail below.

##### ***4.3.6.1 Mesh Type***

Meshes may be either structured or unstructured. A structured mesh is characterized by all interior nodes having an equal number of adjacent elements. The elements are typically quad or hexahedral in shape. Unstructured meshes allow any number of elements to meet at a single node. Elements are commonly triangle or tetrahedral in shape, though quadrilateral and hexahedral unstructured meshes are possible (Owen, 1998).

While structured orthogonal hexahedral meshes are considered to provide more accurate solutions than unstructured tetrahedral meshes of similar resolution (Hefny & Ooka, 2009), the difficulty of constructing an appropriate mesh for complicated geometries is prohibitive. For example, when constructing structured hexahedral based grids, in general, there must be continuous grid lines on all faces (Hefny & Ooka, 2009). This requires the domain to be subdivided into blocks where the grid lines must be continuous through the geometry making them difficult to construct. (Hefny & Ooka, 2009). Conversely, grid generation can be done automatically in unstructured grids, since continuous gridlines are not necessary. Unstructured grids are formed by connecting vertices, which are arbitrarily distributed within the domain by non-intersecting lines that produce the cell elements of the domain (Hefny & Ooka, 2009). This makes them attractive for mapping complex domains (Hefny & Ooka, 2009) of which a building cluster at different sizes, shapes and angles is one example. Thus, for the Balzac geometry, an unstructured mesh generation method was chosen.

#### ***4.3.6.2 Mesh Generation***

The domain mesh was created using ANSYS ICEM CFD software. For unstructured tetrahedral meshes, three generation algorithms were available in ICEM CFD: Octree, Delaunay and Advancing Front. Of these three methods, only the Octree algorithm produced meshes of consistently good quality for the Balzac geometry files.

##### *Octree Method*

With the Octree method, cubes containing the geometry are subdivided in a recursive fashion until the desired resolution is achieved. From this decomposition, irregular cells

are created where these cubes intersect surfaces. Tetrahedral elements are then created from these irregular boundary cells and the regular internal cells. Better grids can be obtained by limiting differences in Octree subdivision levels for adjacent cubes to prevent large size changes and smoothing to improve the resulting element shapes (Owen, 1998).

#### ***4.3.6.3 Mesh Quality***

Tetrahedral mesh quality refers to the aspect ratio of the cells. In general, cells should have ratios greater than 0.3 (where 1.0 refers to the best quality aspect ratio and 0.0 the worst), though 0.15 is generally sufficient for complicated models (ANSYS, 2009). Poor quality meshes can cause difficulties in reaching a converged solution. Thus, a minimum ratio of 0.3 was maintained for the base mesh and the coarser meshes used in the convergence study.

#### ***4.3.6.4 Mesh Settings***

To control the mesh generation, it is necessary to have some understanding of the influencing parameters. Mesh options can be defined globally or in a part specific fashion, which can be used to override global mesh settings. For example, it is possible to define the ground and buildings as different parts and require a finer mesh near the buildings than farther away on the ground. The parameters used to control mesh settings are listed in Table 4.2, and the values chosen to used to define the Balzac grids are listed at the end of this section in Table 4.4.

**Table 4.2: Mesh Setting Descriptions**

<b>Parameter</b>	<b>Description</b>
Global Max Size	Maximum element size in the mesh
Part Max Size	Maximum element size on specific part
Part Min Size	Minimum element size on specific part
Tetra Size Ratio/Expansion Factor	Expansion rate moving away from the surface

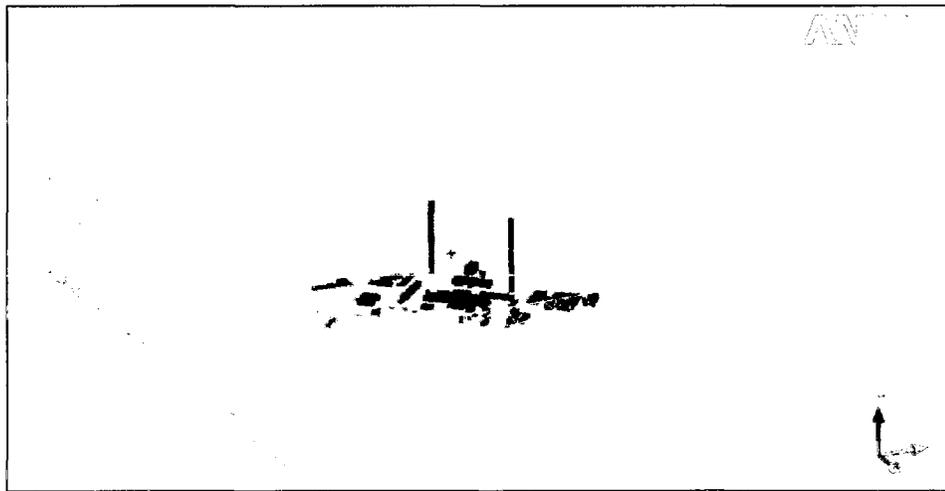
It is desirable to have small mesh sizes near the regions of high gradients, but this is unnecessary for parts of the domain far away from the buildings such as near the top of the domain. However, this change in size should be gradual. In order to move from the small element sizes to the larger elements, it is necessary to define an expansion ratio. This tetra size ratio defines the size of the elements in the layer as the previous layer size, multiplied by the ratio. In the case of the Octree meshing algorithm, the sizes are restricted to powers of 2 multiplied by the smallest size in the model. Thus, the expansion cannot be smooth. An example is given in Table 4.3 below which shows an initial size of 0.5 for an arbitrary length unit and the expansion of layers outwards with a ratio of 1.4. This can come into effect when defining the grid spacing where the closest power of two had to be used (second row in Table 4.3) to match the ideal expansion ratio (first row).

**Table 4.3: Mesh Expansion**

<b>Expansion Method</b>	<b>Tetra Sizes</b>									
Smooth Expansion	0.5	0.7	0.98	1.37	1.92	3.84	5.37	7.52	10.58	14.81
Actual Size (Octree)	0.5	0.5	0.5	1.0	1.0	2.0	4.0	4.0	8.0	8.0

#### 4.3.6.5 Balzac Mesh

For the sake of a grid refinement study (see Section 4.3.10), three meshes were created at varying levels of refinement. Each mesh consists of an identical single domain which extends 200 m above the ground, as shown in Figure 4.5. On the horizontal plane, the domain consists of a square with 1 km long sides. A more detailed discussion of domain size is given in Section 4.3.5.



**Figure 4.5: Coarse Simulation Mesh**

Due to computational restrictions, the finest computational mesh generated consisted of 20.8 million elements. Urban scale simulations have been documented using mesh spacing of between 1 m and 5 m at ground level in the area of interest and up to 40 m at the top of the simulation domain (Hanna, 2006). Franke et al. (2007) recommend sizes at the buildings of at least 10 cells per cube root of the building volume. As the height is typically the smallest dimension, a more conservative estimate is 1/10 of the height, giving an approximate value of 1 m for the current domain based on the average

building height. A maximum size ratio of 1.3 in regions of high gradients (such as near the buildings) was recommended (Franke et al., 2007).

To meet or exceed these recommendations, for the fine mesh, a maximum size of 0.5 m and a minimum size of 0.05 m were set at the buildings, with an expansion ratio of 1.3. The maximum size at the ground at a distance away from the buildings was set to 5 m with a minimum of 0.5 m and an expansion ratio of 1.4. The overall maximum element size at the extremes of the domain was set to 20 m.

To perform grid convergence tests, two coarser meshes were created. The expansion ratios were maintained while the maximum and minimum grid sizes were increased to have a relatively constant change in spacing throughout the domain. A minimum mesh quality of 0.3 was maintained for each grid. A summary of these settings is provided in Table 4.4.

**Table 4.4: Balzac Mesh Settings**

<b>Mesh Name</b>	<b>Coarse</b>	<b>Medium</b>	<b>Fine</b>
<b>Statistics</b>			
# elements	6019226	9921034	20803627
# nodes	1074734	1772426	3716883
min Quality	0.3	0.3	0.3
<b>Settings</b>			
Max element [m]	40	30	20
<i>Building Surfaces</i>			
max size [m]	1	0.75	0.5
min size [m]	0.1	0.075	0.05
expansion factor	1.3	1.3	1.3
<i>Ground Surface</i>			
max size [m]	10	7.5	5
expansion factor	1.4	1.4	1.4

### **4.3.7 Boundary Conditions**

To generate the synthetic data, boundary conditions on the outside of the domain must be specified. The selections for each of these boundaries are discussed below.

#### ***4.3.7.1 Inlet***

The inflow boundary condition was prescribed based on the power law profile. The plant was located in a rural location and, since the rural area contained no large obstacles,  $z_h$  was assumed to be zero. The roughness length was set to a rural terrain value of 0.04 m (Holmes, 2001). For the purpose of generating a synthetic data set, the inlet flow was specified as a smooth, linear sweep in velocity. This varied in both direction and speed. The speed magnitudes were based on hourly measurements from an arbitrary day at the Ottawa International Airport. To aid in the definition of inflow and outlet boundary conditions to the computational domain, wind directions were segregated into three general quadrants as shown in Table 4.5 where each quadrant represents a small variation around a prevailing wind direction. Winds from the fourth and final quadrant were purposefully omitted, since in a typical field situation, one could not expect full 360° wind coverage on a regular basis. The wind direction was assumed to be constant with height. It was also necessary to have some measure of the turbulence so the turbulence intensities from Section 4.3.3.4 were used, which were considered to be reasonable for the slowly varying speed conditions.

**Table 4.5: Inlet conditions at 10 m height using linear variation over duration**

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

#### ***4.3.7.2 Wall and Ground Boundary***

A no-slip (zero velocity) boundary condition was enforced at the ground and building walls. For the purposes of the simulation, resolution of the full boundary layer was not a priority and, on the scale of the element sizes of the simulation, CFX implemented logarithmic wall functions.

It was also necessary to specify the wall properties. Tominaga et al. (2007) followed the recommendation of Franke et al. (2004) to use smooth wall conditions for the building area. Smooth walls were similarly chosen for the present simulation.

#### ***4.3.7.3 Top***

The upper domain surface employed an opening boundary condition, which allowed both inflow and outflow. However, due to the need to maintain a boundary condition consistent with the inlet boundary (Franke et al., 2007), the velocity and turbulence intensity was specified to match the inlet condition and improve convergence and stability.

#### ***4.3.7.4 Outflow***

Typically, outflow or constant static pressure boundaries are specified behind obstacles (Franke et al., 2007). Rather than specify all the outlet velocity values, a zero relative

static pressure was specified at the outlet, which assumed that beyond the domain boundaries the pressure remained constant.

#### **4.3.8 Simulation Type**

Simulations can be either steady-state or transient. A steady-state simulation attempts to find the statistically unchanging flow values for given boundary conditions. A timestep is chosen for the simulation and the solution is marched forward in “time” until a solution is reached that does not change more than a specified convergence level. However, if transient behaviour is caused, it may no longer be possible to reach a converged solution and a transient simulation is required. For the Balzac geometry, even with steady inlet wind profiles, transient behaviour is observed. Thus, it becomes important to model the transient behaviour of the flow for this application.

#### **4.3.9 Initial Condition Generation**

For transient simulations, it is necessary to have an initial flow field from which the solution marches forward in time. Typically, without other available data, this flow field is created from a steady state simulation. This method was also used to initialize the Balzac meshes. A steady state simulation was attempted for the fine mesh for 100 timesteps, since convergence greater than the order of  $10^{-3}$  could not be reached due to the transient nature of the flow. The results of this simulation was used to initialize the transient simulations. For the grid convergence study, these steady results were interpolated onto the coarser meshes for the sake of consistency.

#### **4.3.10 Mesh Discretization and Timestep Selection**

Both spatial and temporal discretization can have a large effect on the results of a CFD simulation. The theoretical accuracy of the simulations is between first and second order in space using the CFX high resolution numerical discretization. Thus, if the grid spacing is halved, the error caused by spatial discretization should decrease 2-4 times. Both spatial and temporal convergence studies were carried out for the Balzac simulation as described below.

Franke et al. (2007) and Franke & Frank (2008) have recommended attempting Richardson extrapolation for wind predictions, which is one method for determining the discretization error. It should be noted that other methods exist for determining error in CFD simulations (Celik & Li, 2005) and research into the applicability of all methods is ongoing.

It is especially important in wind engineering to look at the error as grid converged solutions are usually beyond current computational capabilities. This discretization error arises from modelling with a discrete system with a finite grid and time step (Gokaltun el al., 2005). While reaching solutions in the asymptotic range is challenging, and even impossible in many cases, it is recommended to attempt Richardson extrapolation for wind engineering problems to gain a sense of the error due to the problem discretization (Franke et al., 2007). Details of Richardson extrapolation can be found in Appendix A. An increased safety factor was used to account for oscillatory convergence (Franke and Frank, 2008).

Three grids were used to perform Richardson extrapolation on the results. The baseline grid was the maximum node count possible on the current hardware (i.e. 20.8 million elements), with the other two grids of smaller node count. Rather than perform the entire simulation of 20 minutes on three grids, transient simulations were performed on a steady inlet condition for two directions for a period of one minute. A timestep of 0.2 s was selected as will be further discussed in Section 4.3.10.2 below. Simulations were run to a convergence level of  $10^{-5}$ . The forward dispersion calculations also resulted in a 4 to 5 order of magnitude residual drop for this convergence condition, although the starting ( $10^{-3}$  to  $10^{-4}$ ) and ending ( $10^{-7}$  to  $10^{-9}$ ) order of magnitudes were lower.

#### ***4.3.10.1 Spatial Discretization Error Study***

Spatial convergence was performed using Richardson extrapolation based on the methodology used by Franke and Frank (2008) for an urban street intersection. Twenty circular planes of 300 m radius separated by 1 m vertical height were centered on the building complex. This allowed for importance to be placed on the area of interest for the pathline generation. One thousand points were selected using approximately equal spacing on each plane to provide locations to extract velocity data from the simulations. At each of these points the mean, maximum, minimum, and RMS values for speed and each velocity component were extracted.

Richardson extrapolation was used to estimate the actual value of each at the selected positions. Due to computational limits, it was not possible to attain the asymptotic range for the grid triplet at all the locations, something which is difficult even for simple flows. Franke and Frank (2008) found that in the case of urban flows larger

error bars reflected the greater uncertainty even when grids outside the asymptotic range were used as part of the triplet, so the procedure was performed to determine the order of magnitude of the discretization error. The safety factor was increased from 1.25 to a conservative value of 3 for oscillatory convergence regions and in regions where the order of convergence deviated from the expected range based on Franke and Frank (2008) and further described in Appendix A.

To determine the relative error, the resulting extrapolation values were normalized. To prevent normalization by zero, the error was normalized by the total range of the value within the domain. Table 4.6 and Table 4.7 show the Richardson extrapolation values for two directions using steady inlet conditions for one minute of physical time simulation at 0.2 s timesteps. These values were calculated for the four flow variables (i.e. total and component magnitudes of the velocity vector) extracted at each point. The mean and RMS normalized error values for the entire domain were then determined for points where an estimate could be found.

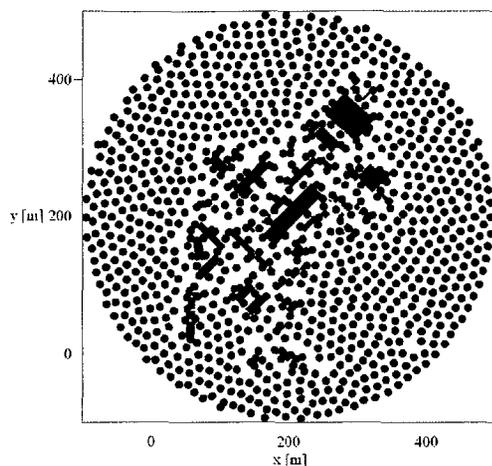
**Table 4.6: Richardson extrapolation error estimates for 1 min of simulated time at 5 km/h from 10° direction normalized by the value range**

Extracted Data at Point	Mean		Max		Min		RMS	
	Mean	RMS	Mean	RMS	Mean	RMS	Mean	RMS
Speed	0.102	0.312	0.155	0.433	0.093	0.357	0.100	0.302
U	0.029	0.122	0.034	0.153	0.035	0.189	0.054	0.206
V	0.076	0.238	0.098	0.266	0.074	0.287	0.097	0.301
W	0.020	0.149	0.034	0.259	0.028	0.214	0.041	0.301

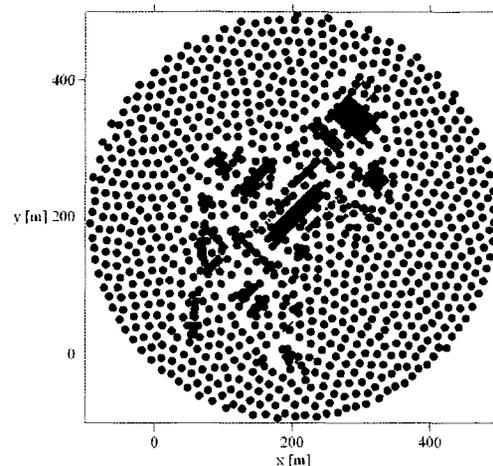
**Table 4.7: Richardson extrapolation error estimates for 1 min of simulated time at 6 km/h from 100° direction normalized by the value range**

Extracted Data at Point	Mean		Max		Min		RMS	
	Mean	RMS	Mean	RMS	Mean	RMS	Mean	RMS
Speed	0.142	0.465	0.179	0.492	0.117	0.444	0.144	0.442
U	0.104	0.314	0.132	0.369	0.088	0.330	0.141	0.431
V	0.036	0.190	0.038	0.164	0.041	0.202	0.053	0.201
W	0.027	0.143	0.036	0.215	0.029	0.178	0.049	0.244

While these results indicate that the velocity values have errors less than the total velocity range, as shown in simulations of Franke and Frank (2008), in this case it was not possible to determine values for a large number (45-55%) of the selected grid triplets even with the use of safety factors for oscillatory convergence. In the simulations of Franke and Frank (2008), approximately 30% of points selected estimations could not be found. With the more complex geometry under current study and point selection in more diverse regions, there were even more points where the extrapolation could not be calculated (between 45-55% of points depending on the variable). Figure 4.6 and Figure 4.7 show points where Richardson extrapolation was attempted for two different heights. In both cases, there are regions upstream, as well as downstream, of the buildings where the estimation cannot be calculated, though this is more common at the lower height.



**Figure 4.6: Points where Richardson estimation of speed available (blue) and unavailable (red) at 2 m height corresponding to 5 km/h wind from 10°**



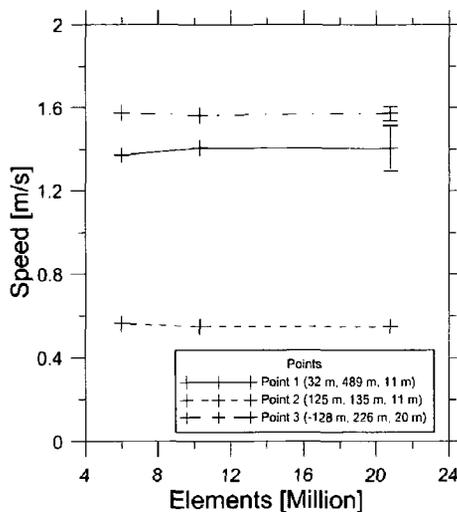
**Figure 4.7: Points where Richardson estimation of speed available (blue) and unavailable (red) at 10 m height corresponding to 5 km/h wind from 10°**

Three random points were chosen in the domain for the purpose of illustration of the Richardson extrapolation values for individual points. These are shown in Figure 4.8 and the coordinates are defined in Figure 4.9.



**Figure 4.8: Random Sample Points**

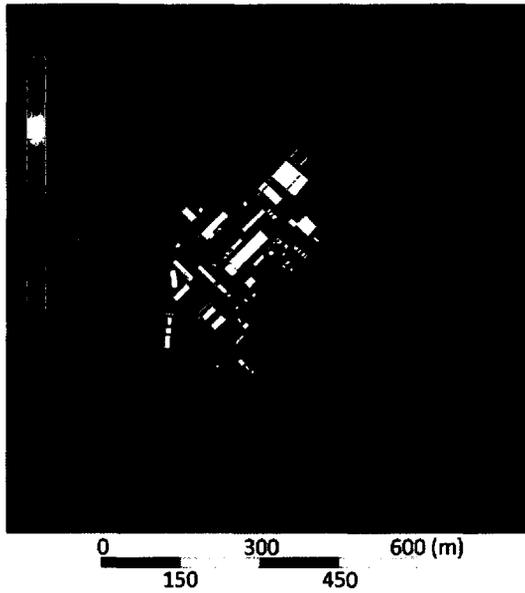
Figure 4.9 gives the speed at each point location for the three grids and the range of the Richardson extrapolation error on the finest grid point corresponding to 5 km/h wind from 10°.



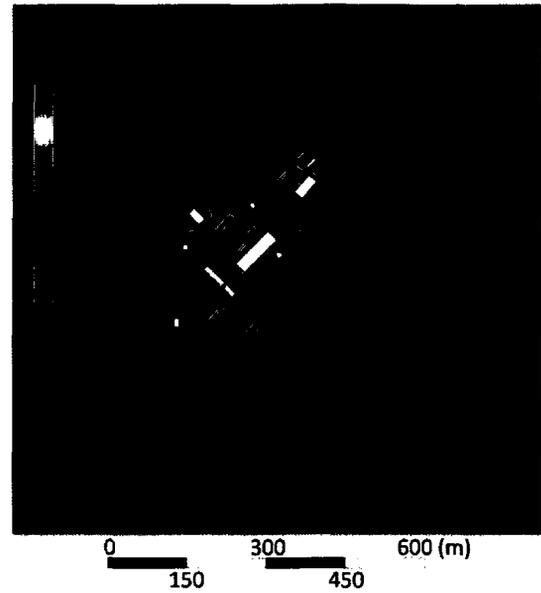
**Figure 4.9: Speed at three points versus number of grid elements**

Note that Point 2 does not have a error range displayed as it was one of the points at which a value could not be calculated. The difference between successive grid values must decrease for it to be possible to estimate error. While the computed speeds for Point 2 are within cm/s of each other for each grid, the ratio of difference was not decreasing and an error estimate could not be calculated. Thus, Richardson extrapolation was considered insufficient to characterize the error and the maximum difference between the three grids at each point was also considered.

Figure 4.10 and Figure 4.11 show the absolute difference in speed between the finest and coarsest meshes at two different heights. At 10 m in height, the largest difference occurs at the upwind building edge (up to 1.7 m/s).

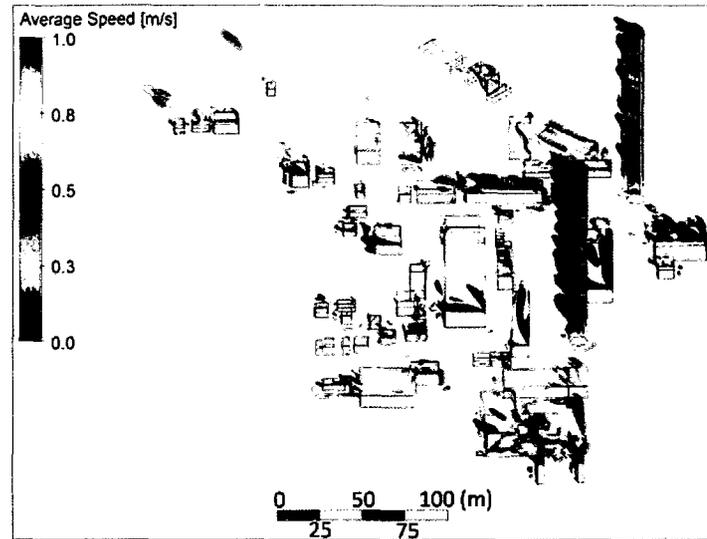


**Figure 4.10: Absolute difference in speed between coarse and fine mesh at 2 m height corresponding to 5 km/h wind from  $10^\circ$**



**Figure 4.11: Absolute difference in speed between coarse and fine mesh at 10 m height corresponding to 5 km/h wind from  $10^\circ$**

To get a clearer picture, a three dimensional plot was constructed. An isosurface at an absolute speed difference of 0.25 m/s was coloured by the speed values on the fine grid (Figure 4.12). As shown here, the largest regions of difference between the grids occur at the upwind corner of the domain and directly downstream of the two stacks where the fine grid is able to improve the resolution of the flow field.



**Figure 4.12: Isosurface 0.25 m/s difference between fine and coarse grids coloured by fine grid average speed**

The errors will have an effect on trajectories passing over these buildings, but it is not essential that the synthetic data set for testing purposes match a physical case as these trajectories correspond to the releases performed on the same flow field. Table 4.8 and Table 4.9 summarize maximum differences between all three test grids, giving the mean and 95<sup>th</sup> percentile values for each of the extraction points.

**Table 4.8: Difference between all three grids for 1 min of simulated time at 5 km/h from 10° direction**

Extracted Data at Point	Mean		Max		Min		RMS	
	Mean	95 <sup>th</sup> percentile						
<b>Absolute [m/s]</b>								
Speed	0.033	0.081	0.048	0.116	0.030	0.122	0.032	0.082
U	0.012	0.044	0.018	0.074	0.015	0.065	0.011	0.040
V	0.032	0.080	0.048	0.116	0.028	0.116	0.032	0.080
W	0.007	0.034	0.012	0.061	0.010	0.050	0.007	0.033
<b>Normalized by fine grid solution</b>								
Speed	0.038	0.118	0.053	0.147	0.035	0.169	0.034	0.116
U	0.389	0.309	0.250	0.590	0.128	0.254	0.054	0.190
V	0.136	0.127	0.186	0.249	0.077	0.179	0.041	0.123
W	19.3	5.33	2.61	6.12	3.20	6.17	0.481	2.04

The mean absolute difference between velocity values for the coarsest and finest grids is less than 0.1 m/s in all cases. The 95<sup>th</sup> percentile difference is less than 0.13 m/s among the grids. This variation is probably less than what could be realistically measured in the field. The third velocity component has, overall, values very close to zero so small differences cause large percentage errors, which are reflected in the table above. The same trend is reflected with the other inlet direction, with mean differences between the time-averaged velocities on the order of cm/s as shown in Table 4.9.

**Table 4.9: Difference between all three grids for 1 min of simulated time at 6 km/h from 100° direction**

Extracted Data at Point	Mean		Max		Min		RMS	
	Mean	95 <sup>th</sup> percentile						
<b>Absolute [m/s]</b>								
Speed	0.041	0.103	0.057	0.142	0.039	0.144	0.040	0.102
U	0.040	0.100	0.060	0.150	0.037	0.140	0.039	0.099
V	0.017	0.065	0.022	0.098	0.026	0.104	0.016	0.057
W	0.011	0.048	0.019	0.100	0.016	0.075	0.011	0.049
<b>Normalized by fine grid solution</b>								
Speed	0.042	0.127	0.055	0.166	0.036	0.158	0.036	0.12
U	0.103	0.145	0.309	0.351	0.061	0.172	0.041	0.127
V	0.243	0.456	0.146	0.325	0.404	0.917	0.062	0.227
W	1.07	2.47	2.03	2.52	27.2	3.42	0.283	0.922

While a completely grid independent solution was not identified due to computational limits, Richardson extrapolation and evaluation of the absolute and percentage difference between the grids suggest that the velocity change between grids is sufficiently small when compared with the velocity range of the fine grid solution. In addition, the trajectories and concentrations ultimately used both correspond to calculations made on the fine grid solution so errors in wind flow calculation used in determining the trajectories will not affect the algorithm testing.

#### ***4.3.10.2 Timestep Selection***

As with the spatial convergence, temporal convergence was also performed using three timesteps (0.5 s, 0.2 s and 0.05 s). Initially, Richardson extrapolation was attempted in order to determine the mean flow characteristics for the exact solution. However, the

differences between the solutions were so small that this resulted in a large number of points with zero difference between them. This made it difficult to apply the Richardson extrapolation equations which rely on dividing by the difference between two grids. Instead, Equation (A.19) as described in Appendix A on Richardson extrapolation was applied uniformly and the results were normalized by the value range.

As can be in Table 4.10, the order of magnitude of the temporal discretization error is several times less than that of the spatial discretization. Thus, timesteps less than 0.5 s were considered sufficient for capturing the important flow characteristics on the scale of the trajectories. A timestep of 0.2 s was chosen since this allowed the convergence criteria to be reached within 3-5 iterations for each timestep as recommended by the ANSYS CFX manual, allowing for a balance between computational time and accuracy.

**Table 4.10: Difference estimates for 1 min of simulated time at 5 km/h from 10° direction normalized by the value range**

Statistical Component for Duration	Mean		Max		Min		RMS	
	Mean (10 <sup>-4</sup> )	RMS (10 <sup>-4</sup> )						
Max normalized diff. between solutions								
Speed	3.4	9.8	7.8	30	7.6	28	3.3	9.2
U	2.3	7.8	5.9	25	5.8	28	4.1	32
V	2.5	7.4	4.8	20	5.4	20	3.2	9.0
W	2.4	8.2	7.4	39	6.2	32	4.2	15

#### 4.3.11 CFD Setup Summary

Following the detailed discussion presented in this chapter, parameters for the final simulations of the flow within the simplified plant geometry for use in trajectory

statistical model evaluation are summarized in Table 4.11. Due to the size of the result files, the results for every second, rather than every 0.2 s, were saved for the purposes of data extraction.

**Table 4.11: CFD summary**

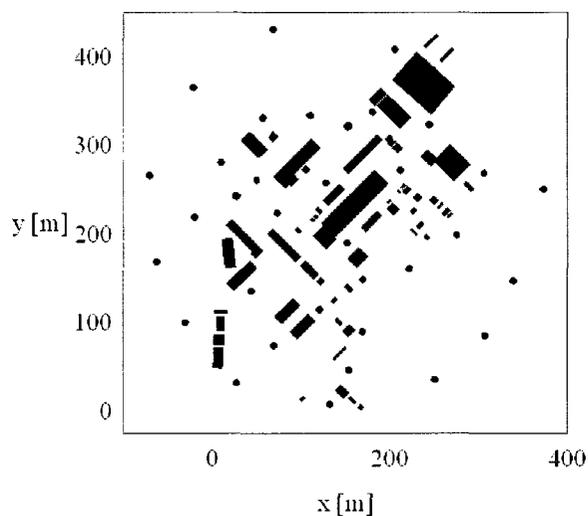
<b>Parameter</b>	<b>Value</b>
Number of mesh elements	20.8 million
Mesh type	Tetrahedral unstructured
Max element size at buildings	0.5 m
Max element size at ground	5 m
Max element size in domain	20 m
Domain size	1000 m x 1000 m x 200 m
Inlet velocity profile	Power law profile (exponent 1/7)
Inlet turbulence profile	See 4.3.3.4
Inlet velocity direction and magnitude	See Table 4.5
Wall boundary	Smooth, no slip
Top boundary	Opening
Outflow boundary	Zero relative static pressure
Turbulence model	k-epsilon RNG
Timestep	0.2 s
Convergence criteria	$10^{-5}$

The effects of gustiness, transient wind variation with height (as opposed to the simple, mean profile) and other such complexities to wind inlet conditions were not considered in this simplified simulation. Finally, the advection-diffusion equation was used to simulate the release of a neutrally buoyant gas from four locations within the domain to be used in algorithm testing.

## 5 Inverse Pollutant Tracking and Results

### 5.1 Receptors

Thirty-one synthetic receptor locations (virtual sensors) were chosen as represented by the filled circles shown in Figure 5.1. Locations were first specified to follow equal spacing within a 250 m radius as calculated by CFX, and any locations beside walls were then manually repositioned outward. All receptors were placed at a height of 1 m above the ground.

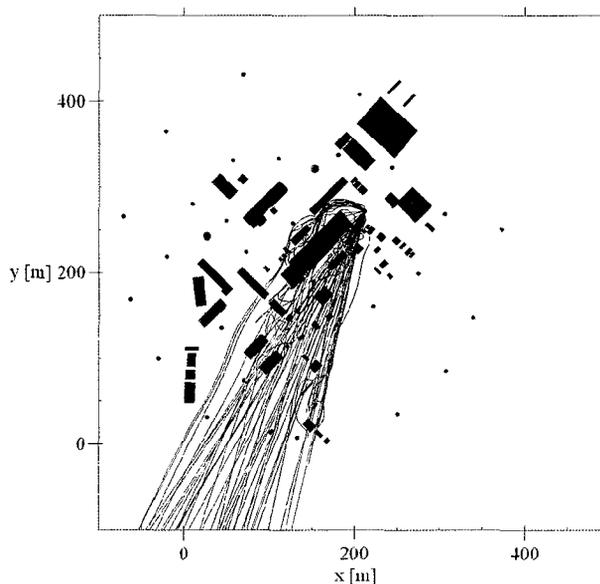


**Figure 5.1: Receptor locations**

### 5.2 Trajectory Calculation

Prior to performing any of the TSMs, algorithms needed to be defined to calculate upstream trajectories of air measured at the receptor locations for each concentration measurement. From the CFD results, air velocity data were available at each node in the mesh for every second of physical time. Trajectories were calculated on this wind data,

which are discretized both in time and space. These trajectories marched backward in time from when the concentration measurement data were extracted. Figure 5.2 shows example trajectories arriving at a single sensor, calculated for the first wind quadrant.



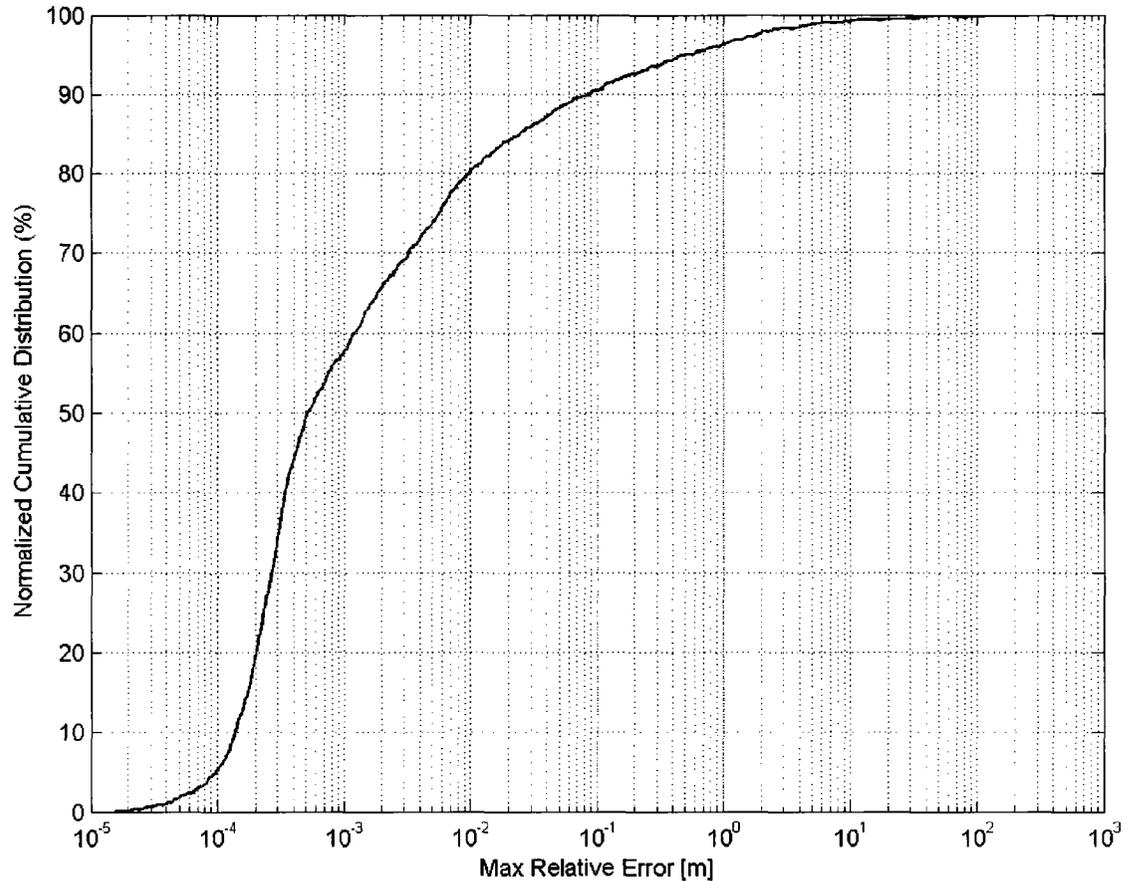
**Figure 5.2: Trajectories from receptor 20 for wind quadrant 1**

To march backward in time, it is first necessary to extract the velocity information at points in the domain at specific times. For a specific timestep, the velocity coordinates are stored at each node of an element. Within an element, the velocity is interpolated linearly using Barycentric coordinates. To calculate trajectories, a 5<sup>th</sup> order explicit Runge-Kutta method was used (Butcher, 1964).

While velocity field data were available at one second time resolution, this was considered too coarse for the trajectory calculations when marching backward in time. Velocity components between timesteps were thus interpolated linearly to a timestep resolution of 0.1 s. The influence of this trajectory timestep on calculated trajectories

was separately investigated and it was found that reducing this timestep resulted in little change in the calculated trajectories as discussed below.

For every 30 s interval of simulated flow time, a new trajectory calculation was initiated at each receptor, and marched backward in time for a maximum of 180 s. To determine the error due to the trajectory timestep selection, multiple trajectory calculation timesteps were used. The potential significance of round-off error was also investigated by selectively using single and double precision variables in the calculations. During test calculations using velocity data representing 20 minutes of physical time (i.e. flow time), the maximum distances between the corresponding trajectories within the data sets for trajectory timesteps of 0.1 s and 0.001 s were compared. Figure 5.3 plots the cumulative distribution of the maximum differences between these trajectories and shows that there was a maximum difference of less than 1 m for 95% of the trajectories used. Since the finest resolution source grid used to estimate source locations had a 5 m cell size, this difference between trajectories calculated using different timesteps was deemed negligible. In rare cases where the differences were larger, examination of the velocity field and calculated trajectories showed that this occurred when a trajectory passed along the edge of a building induced recirculation zone, such that the smallest perturbation could dictate whether the trajectory got drawn into the recirculation, or continued past. Differences of this nature were understood to be part of the flow physics rather than an indication that the resolution of the trajectory calculations was insufficient.



**Figure 5.3: Maximum difference of 0.1 s timestep trajectories compared with timestep of 0.001 s**

### 5.2.1 Turbulent Dispersion

When calculating the trajectories, the relative effect of modelled turbulence versus the large scale resolved structures has implications on the scalability of the algorithms used. On a continental scale, the turbulence structures and scales are very different than those found in a modelled gas plant. To quantify contribution of modelled turbulence to trajectory motion, the turbulent diffusivity was estimated.

The average diffusivity in homogeneous and isotropic turbulence can be defined as (Mazzitelli & Lohse, 2004)

$$D_f = (u')^2 T_L \quad (5.1)$$

where  $u'$  are the root-mean-square of the velocity fluctuations and  $T_L$  is the Lagrangian integral timescale. The Lagrangian timescale has been estimated based on turbulent kinetic energy ( $k$ ) and turbulent eddy dissipation ( $\varepsilon$ ) as (Picart et al., 1986)

$$T_L \sim 0.135 \frac{k}{\varepsilon} \quad (5.2)$$

Assuming isotropic turbulence, then the velocity fluctuations have the same magnitude for each coordinate direction and can be related to the turbulent kinetic energy via

$$k = \frac{1}{2} \overline{u_i' u_i'} = \frac{3}{2} \overline{u'^2} \quad (5.3)$$

Thus, the diffusivity can be estimated as

$$\begin{aligned} D_f &= (u')^2 T_L \\ &\sim \left( \frac{2}{3} k \right) \left( 0.135 \frac{k}{\varepsilon} \right) \\ &\sim 0.09 \frac{k^2}{\varepsilon} \end{aligned} \quad (5.4)$$

In the current simulations, the diffusivity was estimated at twenty thousand points in the area relevant to trajectory calculations (see Section 4.3.10.1) for three primary wind directions. This resulted in an average diffusivity of  $5.4 \times 10^{-2} \text{ m}^2/\text{s}$ . Based on a time interval of 180 s (corresponding to the length of time over which each trajectory was calculated), the diffusion of gas due to the modelled turbulent diffusivity is approximately  $10 \text{ m}^2$ . This is equivalent to a circle with a radius of approximately 2 m. This is less than 1% of the average trajectory length (217 m) and smaller than the 5 m cell size of the finest resolution used. Even at an order of magnitude larger, this distance

is still within a useful range for determining source locations. This suggests that large scale resolved structures tend to dominate the trajectory motion.

### 5.2.2 Concentration Limits

Concentration data at each receptor were also extracted and linked to any trajectories arriving at the receptor in a given timestep. However, at some receptors, the simulation predicted negligibly small concentration values, much smaller than what would be practically measureable in a field experiment. In general, these low concentrations were observed in regions where the plume had passed previously, leaving local concentrations that were no longer exactly zero, even though they would be far below measureable limits. While this would have little effect on CWT, RTWC and QTBA end results which use concentration weightings, these residual concentrations would affect PSCF since it uses binary pollution cut-off limits. For example, by using the raw simulated concentration data in a PSCF calculation, concentrations of  $10^{-20}$   $\text{kg/m}^3$  ( $1.5 \times 10^{-14}$   $\text{ppm}_v$  for methane) could be considered polluted whereas concentrations of  $10^{-21}$   $\text{kg/m}^3$  ( $1.5 \times 10^{-15}$   $\text{ppm}_v$  for methane) could be considered unpolluted, even though both are effectively zero. To account for this in the analysis presented below, all methane concentrations below  $10^{-6}$   $\text{kg/m}^3$  (approximately 1.5  $\text{ppm}_v$  for methane) were levelled to zero. The nominal 1.5  $\text{ppm}_v$  detection limit corresponds to the anticipated sensitivity limit of a measurement via Tuneable Diode Laser Absorption Spectroscopy (TDLAS), which is an anticipated technology for use in a plant environment. The 1.5  $\text{ppm}_v$  value is slightly less than typical ambient concentrations of methane in the atmosphere ( $\sim 1.8$   $\text{ppm}_v$ ), since the synthetic data set only included gas added at the various sources and the methane concentration in the inflow to the domain was set to zero.

### **5.3 Parameters affecting source location algorithms**

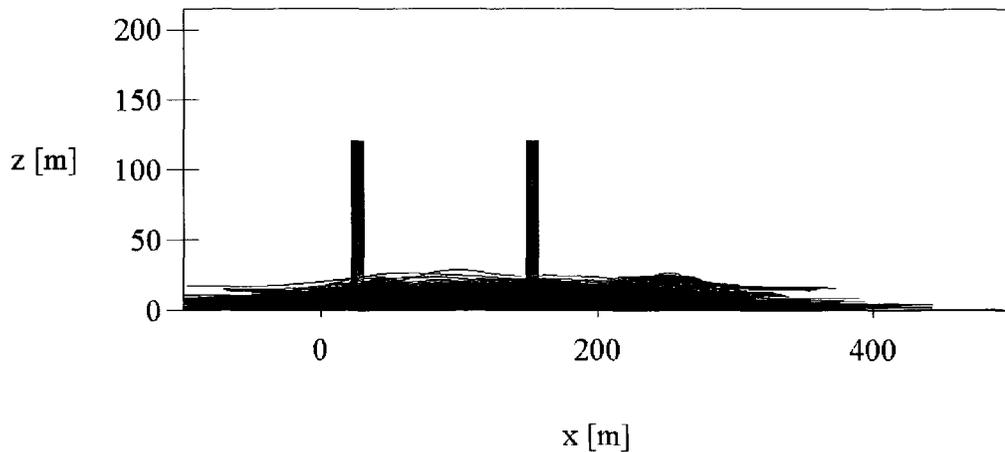
There are multiple possible factors that have the potential to influence the performance of the various source location algorithms. Comparing and evaluating algorithms without some attempt to optimize these influences could result in the rejection of an algorithm simply due to user inexperience in parameter selection, rather than inherent weaknesses.

#### **5.3.1 Parameters universally affecting source location algorithms**

While some input parameters are algorithm specific, others can potentially alter the results for all of the TSMs. Parameters in this latter category are outlined first below, along with the inputs tested for this study.

##### ***5.3.1.1 Cell Size***

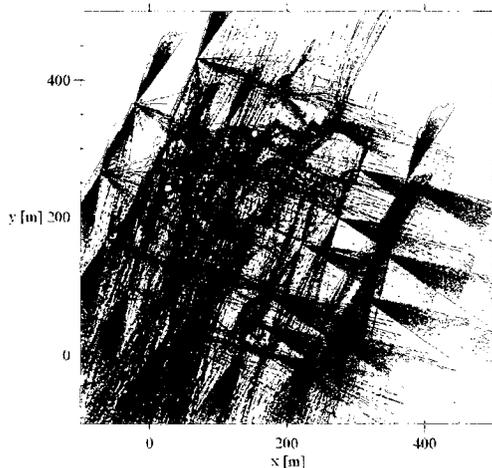
To locate sources, the domain must be divided into potential source regions. Trajectory statistical methods are a balance between coverage of the domain and spatial resolution. As such, the size of the cells considered for pollutant sources can have a very large influence on results. Four cell sizes on the horizontal plane were chosen. These cells were square, with lengths of 5, 10, 20, or 40 m. Although trajectories were calculated in three dimensions (which allowed them to freely pass over and around buildings), it was found that they had very little height variation in comparison with the horizontal distances travelled. This is apparent in Figure 5.4, which shows a side view of all calculated trajectories in the domain. Thus, the source location algorithms were analyzed in two dimensions by considering the horizontal coordinates only (i.e. differences in height between overlapping trajectories in the vertical plane were assumed to be negligible). The trajectory data shown in Figure 5.4 suggests this was a reasonable assumption.



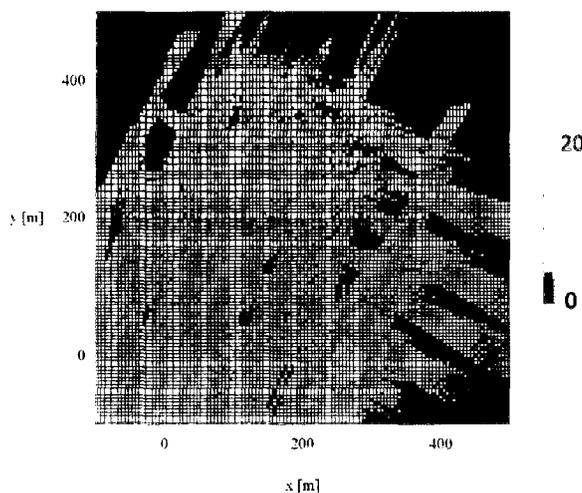
**Figure 5.4: Side view of domain with trajectories for full case**

### ***5.3.1.2 Number of Receptor Measurements***

Increasing the number of receptor measurements (and the number of trajectories) also affects domain coverage. This can be affected by changing both the number of receptors and the measurement intervals. The trajectories (and corresponding measured concentrations at receptor locations) were combined to form several data sets. Separate tests were performed for four different cases using trajectories from each wind quadrant, trajectories from wind all quadrants, 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 every third sensor for all wind quadrants (i.e. reducing the number of sensors used). Figure 5.5 and Figure 5.6 show the locations of trajectories for the combined data set of all trajectories.



**Figure 5.5: Trajectory locations for all three wind quadrants (full case)**



**Figure 5.6: Number of trajectories passing through each cell, 5 m cell size**

### ***5.3.1.3 Number of Sources***

Increasing the number of pollutant emitting sources increases the difficulty of correctly detecting and locating them using backtracking algorithms. For this reason, different configurations of single and dual sources of equal or different magnitudes were considered during testing, as indicated in the Table 5.1 below and Figure 4.3.

**Table 5.1: Source combinations**

Case	Source numbers (locations [m])	Source magnitudes
Single source	1 (28, 99) 2 (-2, 174) 3 (150, 360) 4 (189, 98)	large (4.5 g/s) medium (0.45 g/s) large medium
Dual large sources	1 and 3	large and large
Dual medium sources	2 and 4	medium and medium
Large and medium source	1 and 2 1 and 4 2 and 3 3 and 4	large and medium large and medium medium and large medium and large

#### *5.3.1.4 Timestep*

Each trajectory is defined by points that were calculated at a regular time intervals (0.1 s in the present case). However, the trajectories can be resampled to consider the effect of trajectory resolution on the analysis. In the case of PSCF, CWT and RTWC, the time within a cell is based on the number of trajectory points (i.e. points along the trajectory separated by a fixed time interval) that are present in the cell.. Smaller timesteps increase the computational time but improve the residence time measurement, though this is limited by the resolution of the trajectories themselves. For QTBA, the potential influence of a trajectory is calculated as a field throughout the domain at each point along the trajectory which can improve the results, but greatly increases the computational time. For the purposes of this study, the resampling timestep was maintained at 1 s to correspond with the available velocity field resolution from the CFD simulations.

### **5.3.2 Algorithm Specific Parameters**

Several other parameters that could affect the source location solution are specific to the individual algorithms. These are described below along with the ranges of values considered in the present work.

#### ***5.3.2.1 PSCF Pollution Cut-off***

One of the defining characteristics of PSCF is the use of a concentration cut-off to determine whether a trajectory is considered "clean" or "polluted". Having low cut-offs makes it less likely to miss the source region, but increases the likelihood of finding false positives. Increasing the cut-off value can improve the spatial resolution, but risks missing the source. Thus, finding a "rule of thumb" for this cut-off in this type of environment is desirable. In the present work, three cut-off percentiles were tested. Lin et al. (2001) suggested that the 90<sup>th</sup> percentile performed better than the 50<sup>th</sup> and 75<sup>th</sup> used in previous studies. As the current application uses a pooled data set and should have a larger number of sensors with zero and low pollutant readings somewhat higher cut-off values were chosen corresponding to the 75<sup>th</sup>, 90<sup>th</sup> and 97.5<sup>th</sup> percentile concentration measurement

#### ***5.3.2.2 RTWC segment length***

RTWC is characterized by the redistribution of concentrations along trajectories. Each trajectory is divided into segments and the concentrations are reweighted to vary among these segments. The reweighted trajectories are then used to calculate a redistributed results field. Although reducing the trajectory segment length does tend to improve the resolution of the final results, the minimum segment length is limited by the cell size used

to calculate the redistributed results field. Thus, RTWC segment lengths of 2, 4, and 6 times the cell size of the redistributed concentration field were tested.

### ***5.3.2.3 RTWC convergence criteria***

Since RTWC is an iterative process, it is necessary to specify some criterion at which to stop the iterations of field redistribution. This criterion should be sufficiently tight that there are no visible differences between the subsequent fields. While the average difference between fields has been used (e.g. Han et al., 2007), the maximum difference was considered more conservative as a convergence criterion. Rather than continually changing the convergence criterion, the maximum difference between fields was set at the beginning of the run to be no greater than 1% of the concentration range of the initial guess field, as calculated from CWT run used to initialize the RTWC calculation. Since the range of the reweighted concentration field is expected to increase with successive redistributions, using the initial guess field to set the 1% convergence criterion was more conservative than revising the criterion with every iteration. This also allowed the convergence criterion to be known in advance and run time predicted.

### ***5.3.2.4 QTBA trajectory standard deviation coefficient***

In QTBA, the standard deviation for the trajectory position uncertainty is assumed to be a linear function of time upwind,  $\sigma(t')=at'$ . Since the coefficient  $a$  is not generally known, it is typically assumed to be 5.4 km/h or 1.5 m/s (Zhou et al., 2004). Two additional values of 0.75 and 1.25 times the standard value were also investigated.

### 5.3.2.5 Filters

Zeng and Hopke (1989) noted that for the PSCF approach, if a cell contains a single polluted trajectory point, the value of the PSCF statistic is 1.0, even though the confidence in that value should be low. This is also true for the CWT and RTWC algorithms. Many authors ignore cells with trajectory counts (i.e. number of different trajectories) or point counts (i.e. total number of points from all trajectories present) under a certain value (e.g. Zhou et al., 2004).

An alternative to simply ignoring cells with trajectory counts below a fixed threshold is to use a weighting function. Zeng and Hopke (1989) proposed a weighting function for PSCF, based on point count that could be multiplied by the local PSCF value as follows:

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

Given the importance of trajectory statistics, an alternative weighting filter is proposed. Rather than use point counts,  $n$  in the above filter is replaced by the trajectory counts per cell. This trajectory weighting function is, again, multiplied by the resulting local PSCF value.

Weighting functions are not restricted to PSCF. Zhou et al. (2004) introduced the idea of an empirical weighting function to QTBA to reduce noise. Zhao et al. (2007) implemented a simpler weighting function for the same purpose, which was found to

perform just as well. The weighting function of Zhao et al. (2007) had the following form where  $p$  is the QTBA value of the cell:

$$weight = \begin{cases} 1.0 & p \geq 2Average(p) \\ 0.75 & 2Average(p) > p \geq Average(p) \\ 0.5 & Average(p) > p \geq 0.5Average(p) \\ 0.2 & p < 0.5Average(p) \end{cases}$$

In the present work, for each algorithm, various filters were implemented to see if these could remove false predictions and improve the solution. The final filters tested are noted below in Table 5.2.

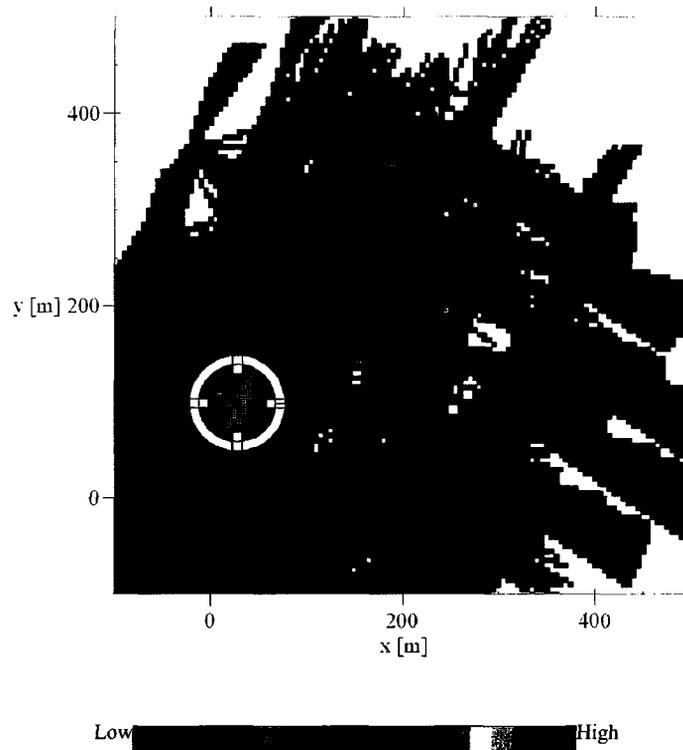
**Table 5.2: Trajectory filters considered in the present work**

Filter tested	PSCF	CWT	RTWC	QTBA
No filter	X	X	X	X
Minimum point count of 4	X			
Minimum trajectory count of 4	X	X	X	
Weighting function (Zeng and Hopke, 1989)	X			
Trajectory weighting function (present work)	X			
Weighting function (Zhao et al., 2007)				X

#### 5.4 Result Field Interpretation

Running each algorithm for the various tested parameters resulted in fields of values which required interpretation. As different units were used for different algorithms, each result field was normalized to a range of 0 to 1 for unbiased comparison. Regions which are highly likely to be sources are denoted in red and regions where sources are unlikely are denoted in blue.

An example plot is shown in Figure 5.7 for Source 1 using the RTWC algorithm on the fine 5 m mesh, with a reweighting length along the trajectories of 10 m. In regions where trajectory or endpoint filter requirements were not met, no result is plotted since the values were unknown. The actual location of Source 1 is denoted by the white circle. In this case, Source 1 is predicted to be located within a few metres of the actual source location. Trajectory coverage of the domain is also visible by comparing the coloured regions with the uncoloured regions that had no trajectories passing through them.



**Figure 5.7: RTWC results for Source 1 calculated with 10 m segment lengths (2x cell size) and 5 m cell size. The actual source location is indicated by the white circle. The source is identified within 5-10 m. (Pearson correlation = 0.43, explained in detail in Section 5.5)**

## 5.5 Results Evaluation Methods

Summarizing the large results matrix and performing parameter optimization required some method of quantifying the results, rather than simply using visual observation alone. So, while qualitative comparison was performed by observing plots of the resulting fields by eye, statistical correlation was used as well. Ideally, high correlations would be given if the source alone was correctly and precisely located, and correlation should decrease as the precision of the location was reduced or if other regions were incorrectly identified as possible secondary sources. Thus, results indicating high probability source location regions surrounding the actual source cell should be considered relatively good compared with those at a distance away and it is essential that the method consider the distance from the source region.

The two most prevalent comparison methods found in the literature review, aside from qualitative comparison, were the spatial correlation index and the Spearman rank order correlation between the reconstructed fields and estimated emissions. The spatial correlation index uses both concentration and distance between regions to determine the match between the extremes of the domain. While this method was initially implemented, it proved to be more computationally intensive and time consuming than the source location algorithms themselves and it was necessary to investigate alternative correlation methods with more reasonable run times.

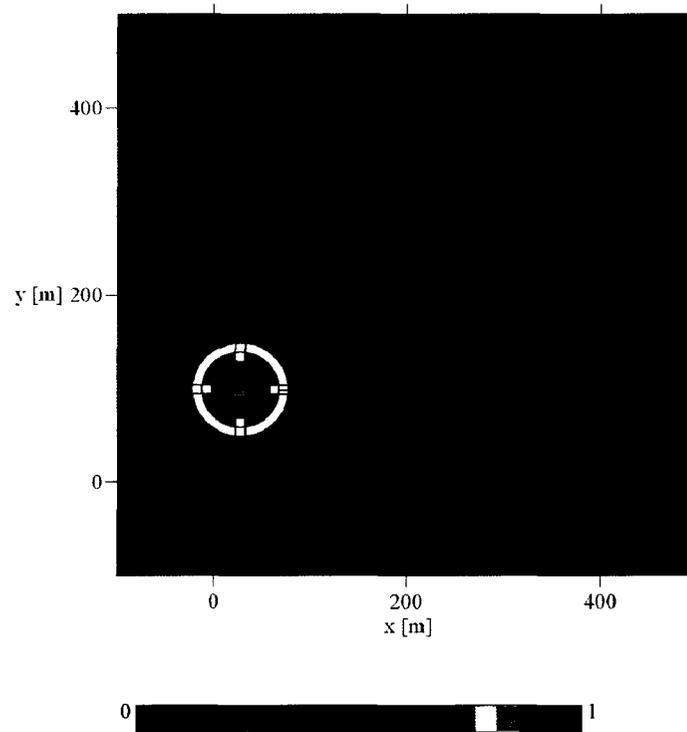
The Spearman rank correlation finds a correlation between two vectors based on their relative rank order from largest to smallest. In the present case, a result vector can be simply defined as the array of all intensity values in the result field image. This vector can then be compared with a vector created from a reference field containing the actual

location of the source as explained below. Due to the nature of the constructed fields, there were many points which were of similar value (such as near zero regions) which lead to different ranking values. To obtain rank correlation values that agreed with qualitative observations, it was necessary to divide the field into bins which were considered the same rank. However, this made the correlation very dependent on the relative bin size. Thus, a method based on a continuous field was preferred.

By contrast, the Pearson correlation does not require the use of ranks and can be used on continuous fields. Tests using the Pearson correlation proved to be consistent with visual observations of the reconstructed plots. Thus, in light of the difficulties identified with measures based on the spatial correlation index and the Spearman rank correlation, the Pearson correlation was chosen as a quantitative measure for algorithm comparison in the present work. Further details of each correlation method may be found in Appendix B.

To implement the Pearson correlation method, a relevant field must be generated for comparison. This field must have high values at the source(s) and low values in other regions. Distance from the source should also be considered, where high likelihood values far from the actual source are considered worse than moderate likelihood values near the source. Similarly, as PSCF, CWT and RTWC cannot give estimates in regions where there are no trajectories, in the event that sufficient trajectories do not cross the actual source, then successful identification of a potential source adjacent to the actual source would be beneficial, though not as good as locating the actual source exactly. Locating high or moderate regions far from the source should reduce the correlation value.

Considering the practical application of locating sources of fugitive emissions for repair, the search area to cover (and hence the required time and effort for the search) scales with  $radius^2$ . Thus, for the purpose of evaluating the Pearson correlation, a comparative field centered about the actual source with a decay proportional to  $\frac{1}{radius^2}$  was specified. A minimum radius of 1 m was specified to avoid singularities at zero. The comparative map for Source 1 is plotted in Figure 5.8. In cases of multiple sources, the two reference fields were added together after being scaled by the relative source strengths.



**Figure 5.8: Comparative correlation map for Source 1 based on the actual location of the released gas**

The correlation was performed for the normalized TSM algorithm output values at the center of the cells of the finest source location grid. These centers defined where the points for the vectors would be extracted from both the results grid and the comparative field. The coarser grid results were interpolated onto the finest grid for correlation calculation. Thus, the spreading of the source regions due to large cell sizes can be taken into account. The domain considered for comparison was confined to a 500 m by 500 m region surrounding the buildings since no sources were expected at greater distances and the observer could discount them automatically. Regions in this domain without known values (due to a lack of passing trajectories, for example) were interpolated. This interpolation was necessary so that the correlation vectors would be of constant length and to enable quantitative comparisons among different cases based on correlation magnitude. This interpolation also makes sense from an application standpoint, since an observer would similarly be forced to interpolate unknown parts of the domain during visual examination of results.

## **5.6 Analysis of Results**

The various input parameters described previously and their influence on each of the algorithms were investigated. Once the optimum parameter settings were identified for each algorithm, the algorithms were subsequently compared to evaluate their relative performance against each other.

### **5.6.1 Potential Source Contribution Function**

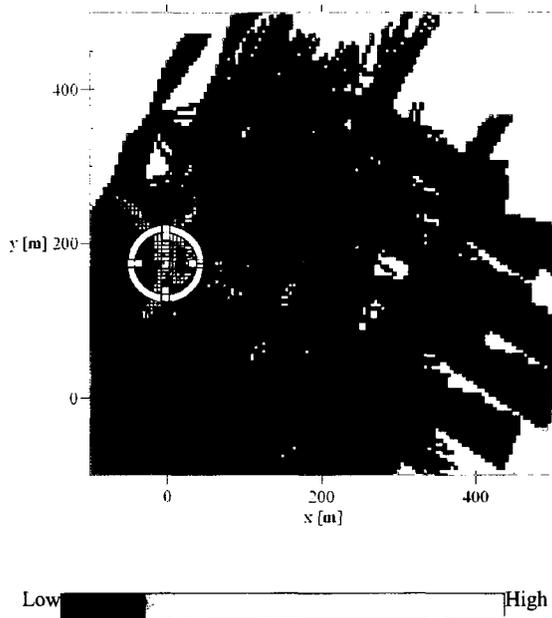
Potential source contribution function (PSCF) is the simplest algorithm under investigation. As it was expected that there would invariably be several receptors measuring low to zero concentration values and some which always measure high due to

their position relative to the source, the cut-off level was defined based on all the measurements in the pooled data set rather than on an individual receptor basis as is sometimes done (for example, Hsu et al., 2003). A cut-off percentile based on the pooled data set allows receptors that always measure high values due to their proximity to a source to always have polluted trajectories. Using sensor specific cut-offs based on individual sensor data would require receptors, even if they are always directly downstream of a source, to reject a portion of its trajectories as non-polluted in spite of their high concentration readings.

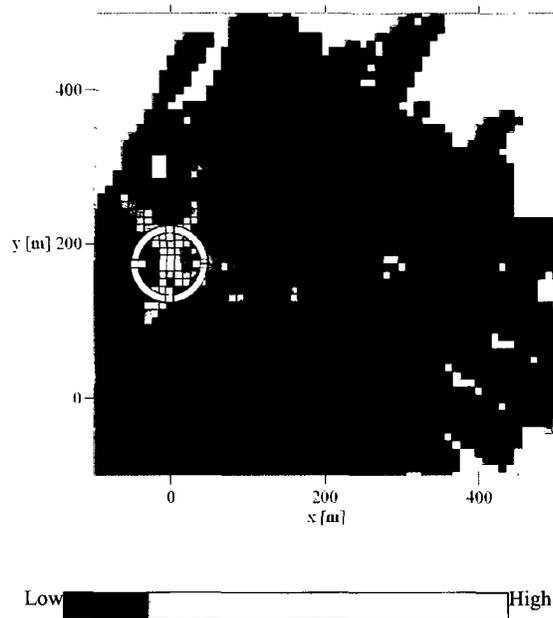
#### ***5.6.1.1 Grid cell size***

Variation in grid cell size introduces a trade-off between the number of trajectories which pass through the cell (which increases with cell size and improves the reliability of the source prediction) and the spatial resolution with which a source can be identified (which improves with decreasing cell size). For each source, there is an optimum size which performs best. As an example, Figure 5.9 to Figure 5.11 show plots of the results for Source 2 calculated for three different grid sizes.

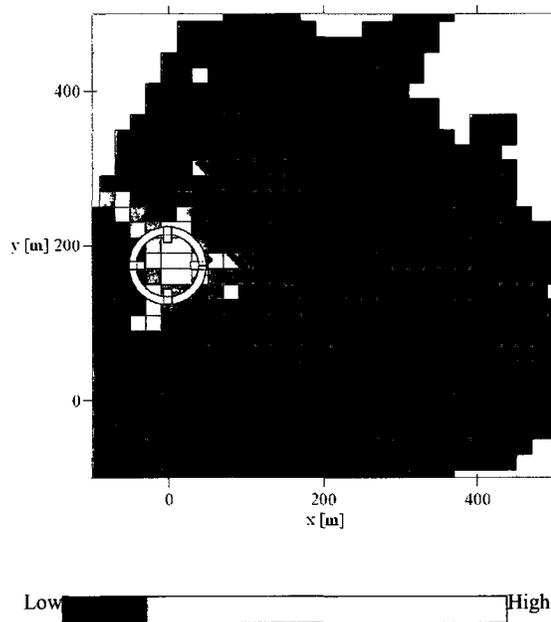
For Source 2, the smallest cell size provides a high valued smear (see Figure 5.9) in the source region. For the 10 m cells (Figure 5.10), the highest point in this smear ends up Northwest of the source, rather than the centered on it, though there are secondary high points at the source. This is filtered out when the cell size is increased to 20 m (Figure 5.11). Thus, the correlation coefficient would be expected to experience a dip at the 10 m cell size for Source 2 as shown in the correlation plots presented in Figure 5.12.



**Figure 5.9: PSCF Source 2, 97.5th percentile, 5 m cell size (Pearson correlation = 0.27) The region surrounding the source is clearly visible, though it is smeared into the surrounding area.**

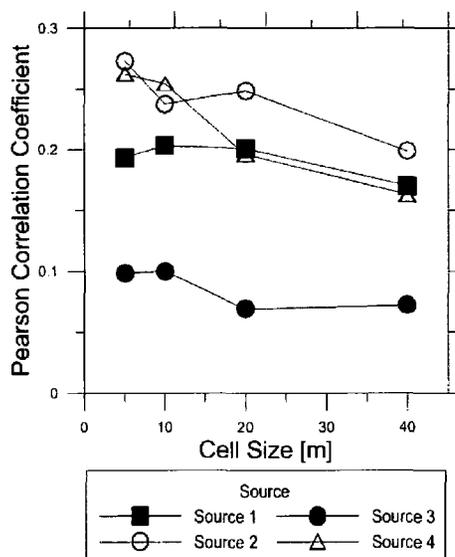


**Figure 5.10: PSCF Source 2, 97.5th percentile, 10 m cell size (Pearson correlation = 0.24) Highest value is to the northwest of the source at a greater distance than in the 5 m prediction. The second highest cell corresponds to the actual location.**



**Figure 5.11: PSCF Source 2, 97.5th percentile, 20 m cell size (Pearson correlation = 0.25) The prediction at the actual source location corresponds to the second highest predicted cell. The distance from the highest cell to the source is less than that found with the 10 m cell size.**

Figure 5.12 shows the variation in correlation coefficient with cell size for all single source cases. In the majority of cases the largest cell size (40 m spacing) is the worst case due to the decreased spatial resolution, although for Source 3 there is negligible difference between the 20 and 40 spacing cases as shown in Figure 5.12. For large cell sizes, the position of the cell relative to the location of the actual source affected the correlation. A source centered in a cell provides a better correlation than one on the edge of a cell. This can be seen in the slight upward variation in the correlation for Source 1 alone in Figure 5.12. However, the change in correlation value between the 5 m and 10 m case are not significant.

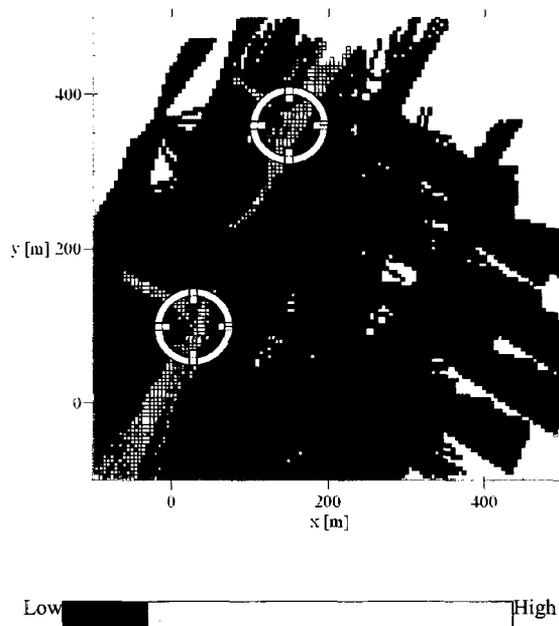


**Figure 5.12: PSCF correlation vs. cell size for a single source using 97.5th cut-off percentile and no filter**

In general, a cell size of 5-10 m provided better correlation than the larger cell sizes. As shown in Figure 5.12, Source 1 and Source 3 were identified with slightly higher correlation values with the 10 m cell size than the 5 m cell size; however the opposite is true for Sources 2 and 4. Source 2 shows a slight dip at the 10 m spacing as discussed previously. An averaging filter may be one way to filter out such variations, but this could have a detrimental effect on the ability to pinpoint the source region.

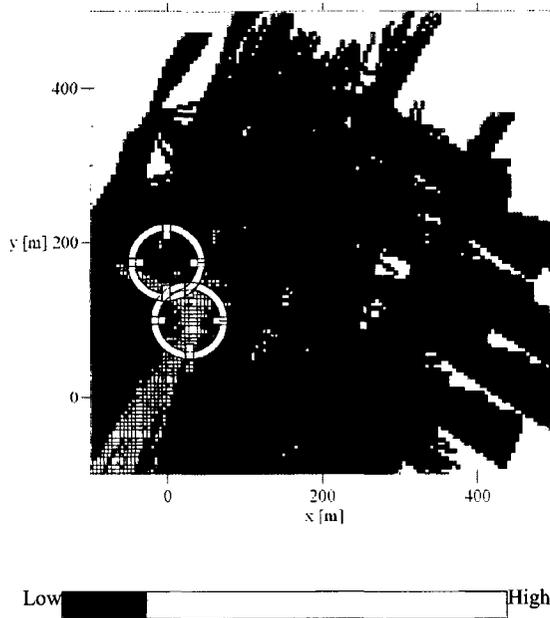
When multiple sources were considered, identifying the larger source was considered more important than identifying the smaller source since the larger source contributes more emissions into the atmosphere. Looking at the result field plots for two large sources (Figure 5.13), both sources are visible. The highest PSCF results field values (Figure 5.13) in the vicinity of Source 3 occur at the edge of the trajectory coverage where there are few trajectories passing through that region and all must pass by

the source. For the smallest cell case (5 m), there is no coverage from the other directions at all to reduce the values in this region.

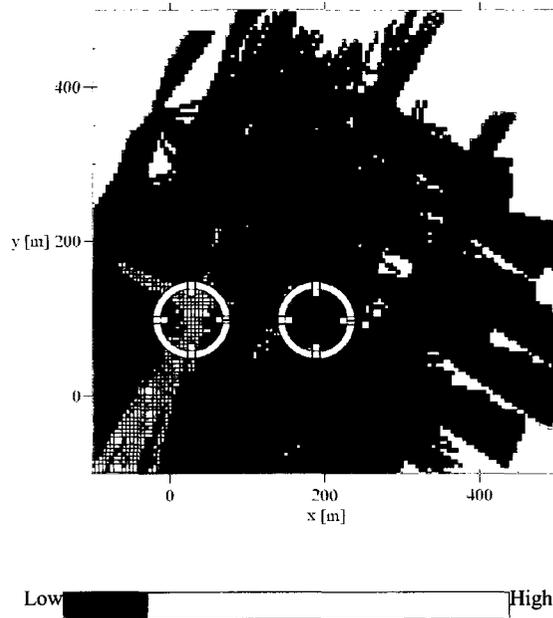


**Figure 5.13: PSCF Source 1 and Source 3, 97.5th percentile, 5 m cell size (Pearson correlation = 0.10) Both "large" sources are visible, though the highest source prediction region is to the North of Source 3.**

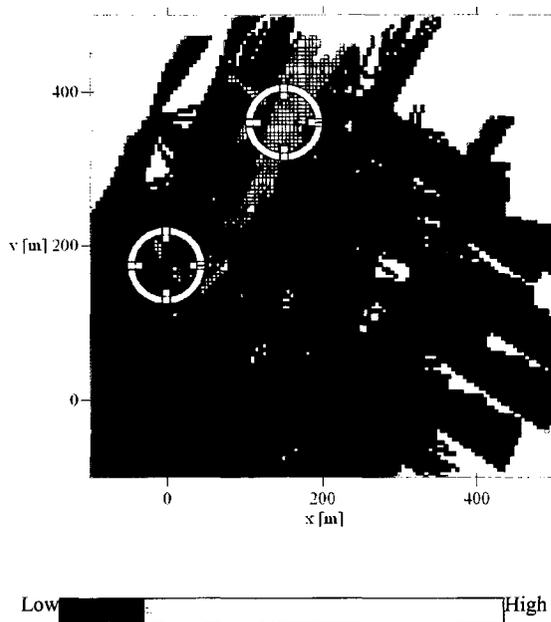
Figure 5.14 to Figure 5.17 show result fields for a mix of large and small sources. As expected, the larger source generally obscures the smaller sources even though the smaller sources are easily found in a single source setting. The larger source, however, can still be identified.



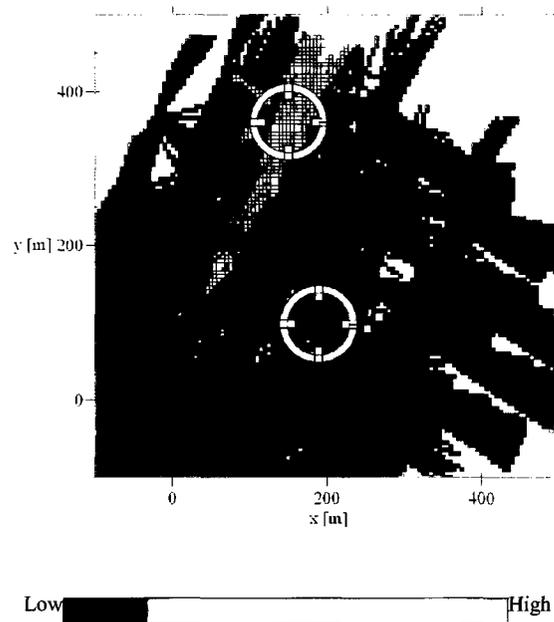
**Figure 5.14: PSCF Source 1 and Source 2, 97.5th percentile, 5 m cell size (Pearson correlation = 0.20) "Large" Source 1 is within the highest predicted region, but "medium" Source 2 can no longer be distinguished.**



**Figure 5.15: PSCF Source 1 and Source 4, 97.5th percentile, 5 m cell size (Pearson correlation = 0.19) "Large" Source 1 is within the highest predicted region, but "medium" Source 4 can no longer be distinguished.**



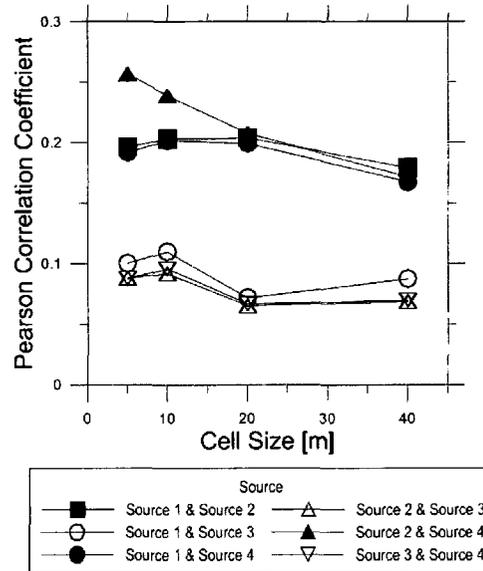
**Figure 5.16: PSCF Source 2 and Source 3, 97.5th percentile, 5 m cell size (Pearson correlation = 0.09) "Large" Source 3 is within the secondary high of the prediction smear. "Medium" Source 2 cannot be distinguished from the smear's tail.**



**Figure 5.17: PSCF Source 3 and Source 4, 97.5th percentile, 5 m cell size (Pearson correlation = 0.09) "Large" Source 3 is within the secondary high of the prediction smear. "Medium" Source 2 cannot be distinguished from the smear's tail.**

In cases of two sources of differing strength, the larger source prediction dominates the correlation value (Figure 5.18) as the emission rate of the large source is an order to magnitude higher than the small source. Thus, the correlation coefficient trends are similar to that of a single large source alone. As discussed previously, the results field plots show that the larger source tends to obscure the smaller source (refer Figure 5.14 to Figure 5.17). However, the large source is still visible and this is the location that is of the most interest. Sources 1 and 3 together (large sources) do not have as high a correlation as Source 1 individually, though both are visible in Figure 5.13. This is due to the high region caused by a low trajectory count North of Source 2. For the two small

sources (Source 2 and 4), smaller cells provide the best correlation, as was the case with both the individual sources alone.



**Figure 5.18: PSCF correlation vs. cell size for two sources using 97.5th cut-off percentile and no filter. Correlation values are dominated by the algorithm prediction of the larger source.**

### 5.6.1.2 Correlation Magnitudes

Reviewing Figure 5.7 and Figure 5.9 to Figure 5.18, some general statements can be made about the magnitude of the correlation coefficient and the prediction plots. While there is some smearing around the source, Figure 5.7 shows a prediction with the highest region within a few meters of the actual source and the surrounding smear is lower than that at the actual source location. This has a correlation value of 0.43. Thus, an excellent prediction should be in the range of a correlation value of 0.4.

Figure 5.9 to Figure 5.11 all have correlation values of approximately 0.25. In each case, the smear surrounds the source, though the highest prediction location is removed from the actual source location. The high regions of the smear are also larger and more spread out than that seen in Figure 5.7. This is still extremely useful information that could be used to significantly reduce the search area.

Figure 5.13 shows the PSCF results for two large sources and has a correlation value of approximately 0.10. The highest prediction region is not at the source, but is upstream of Source 3. Though the sources are contained within the medium probability regions, these regions are quite large and extend well upstream and downstream of each source. However, the search area that would be required to find these sources is still much less than the area of the entire plant, making even this prediction useful.

Based on these observations, in general, correlation coefficients on the order of 0.1 are moderately useful predictions, on the order of 0.2 are very useful predictions and on the order of 0.4 extremely useful predictions. In addition, based on the number of points sampled, correlation differences of less than 0.014 should not be considered significant because they could arise by chance.

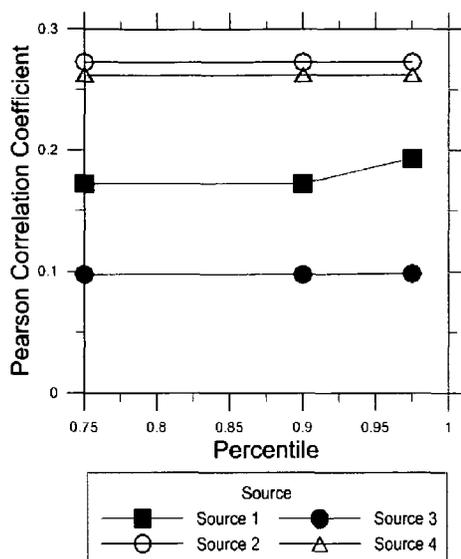
#### ***5.6.1.3 Effect of PSCF Pollutant Cut-off***

In general, high PSCF cut-off percentile values would be expected to improve the correlation by isolating strongly polluted trajectories from those that may have low concentration readings but not be along the direct path of the plume. However, as the cut-off point is increased significantly, there is a trade-off between initial correlation improvements and reducing the number of valid polluted trajectories. An arbitrarily high

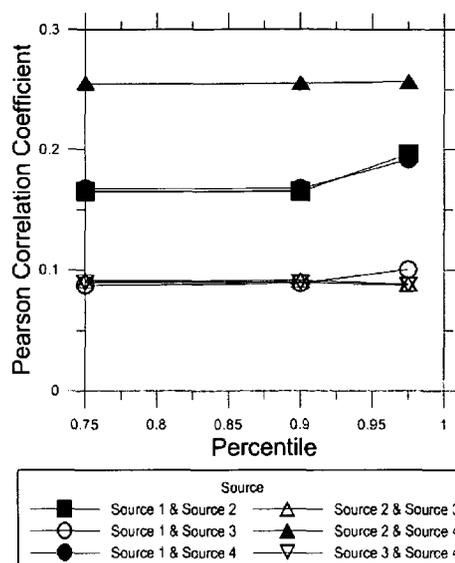
cut-off reduces the PSCF prediction in areas where these high-concentration but "unpolluted" trajectories pass along with the remaining polluted trajectories, even if they move over the source. In general, with large numbers of sensors and directions, and in the case of a small number of leaks, it is expected that a large number of the measurements should be non-polluted, and higher cut-off values will give better correlation results.

Figure 5.19 and Figure 5.20 shows the strength of the Pearson correlation versus the percentile cut-offs for the various source configurations. Note the flat trend and lack of change of correlation values between the 75<sup>th</sup> and 90<sup>th</sup> percentile cut-offs in all cases. As there are few sources and a large number of sensors, there are few trajectories which have concentration measurements above the detection threshold. Thus, the effective concentration value of the cut-off is the same for both the 75<sup>th</sup> and 90<sup>th</sup> percentile. Plots of the correlation values for data without the detection threshold (Figure 5.21 and Figure 5.22) illustrate this point.

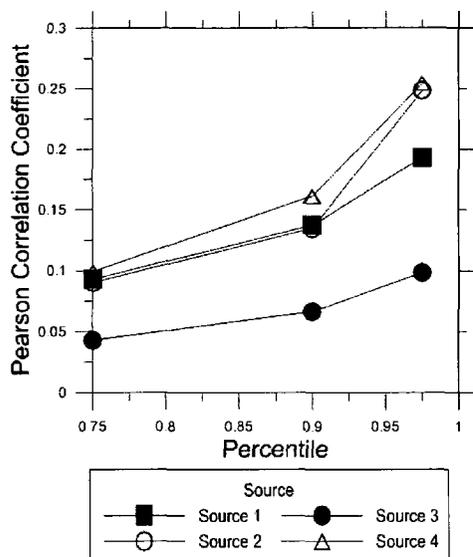
As a detection threshold is a limit on any practical measurement device, for a large number of sensors with few sources, high effective percentile cut-off values are to be expected. Defining a universal optimum, however, was not possible and this limitation is a principle shortcoming of the PSCF algorithm. Thus, PSCF's main use would be to support other algorithms and investigate the influence of trajectories. PSCF should always be used in conjunction with the testing of multiple cut-off values, in agreement with the findings of Cheng and Lin (2001).



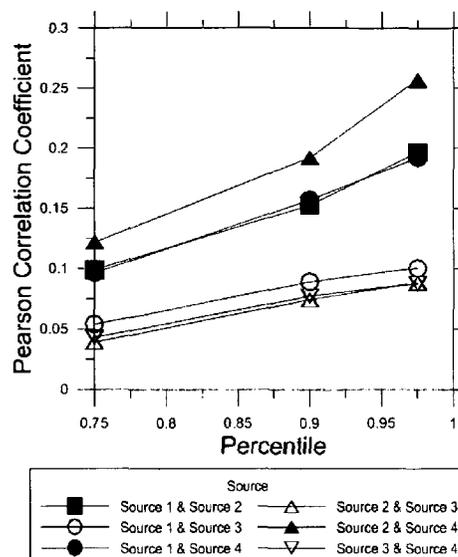
**Figure 5.19: PSCF zeroed data set correlation vs. cut-off percentile for single source using 5 m cell size and no filter. Due to few trajectories above the detection threshold, there is no or little difference between the correlations at the 75<sup>th</sup> and 90<sup>th</sup> percentiles.**



**Figure 5.20: PSCF zeroed data set correlation vs. cut-off percentile for two sources using 5 m cell size and no filter. Due to few trajectories above the detection threshold, there is no or little difference between the correlations at the 75<sup>th</sup> and 90<sup>th</sup> percentiles.**



**Figure 5.21: PSCF non-zeroed data set correlation vs. cut-off percentile for single source cases using 5 m cell size and no filter. With no detection threshold, increasing the cut-off percentile improves the source prediction. Thus, the optimum threshold for these cases lies somewhere above the 90<sup>th</sup> percentile.**

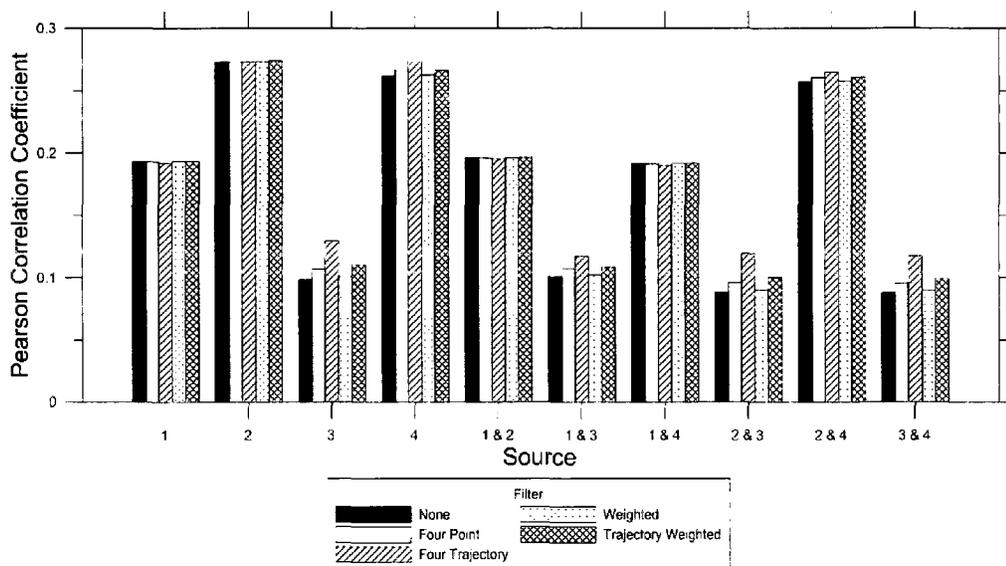


**Figure 5.22: PSCF non-zeroed data set correlation vs. cut-off percentile for cases with two sources using 5 m cell size and no filter. With no detection threshold, increasing the cut-off percentile improves the source prediction. Thus, the optimum threshold for these cases lies somewhere above the 90<sup>th</sup> percentile.**

#### 5.6.1.4 Filter effects

Figure 5.23 shows the effect of filters on PSCF solutions using the entire trajectory data set. The inclusion of any of the four filters (four-point, four-trajectory, weighting and trajectory weighting) had little effect on the single source correlations for Source 1 and Source 2. For Source 3 (large source), the correlation increased visibly by 0.025 when using the four-trajectory filter as discussed further below. Source 4 (small source) also had increased correlation with the use of filters, though to a lesser extent than Source 3. The filters had similarly small effects for the multiple source cases. The best performing filter was the four-trajectory filter, which in general showed a measurable improvement

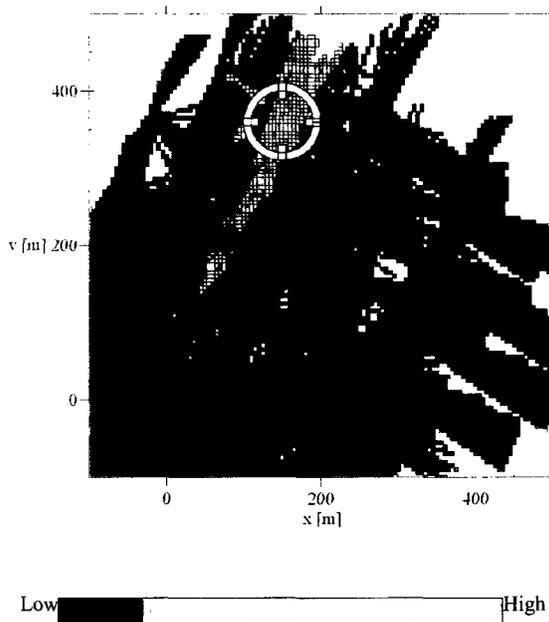
or negligible benefit in all cases, and in no case caused a decrease in measured correlation. Source 3 again stood as a bit of an outlier, since most of the filters did provide some improvement in the correlation for cases involving this source, with the four-trajectory filter providing the largest benefit.



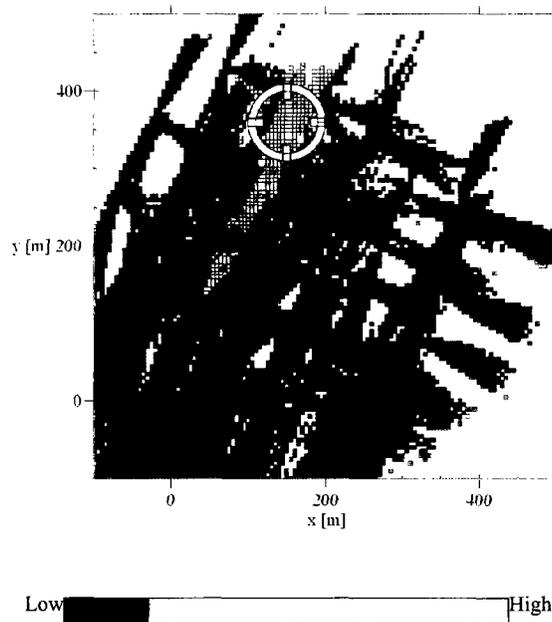
**Figure 5.23: PSCF correlation with various filters for single source using 97.5th cut-off percentile and 5 m cell size. Benefits of the filters are negligible for Source 1 and Source 2. There is a small improvement for Source 4. Source 3 had significant differences between the results of various filters.**

Considering Source 3 in greater detail, Figure 5.24 and Figure 5.25 show the associated four-trajectory filtered and unfiltered field plots for this case. Results calculated using the filter show better correlation as the filter effectively removes the false source locations at the edge of the domain, where there are only a few trajectories providing coverage, all of which presumably must have passed near the source on their way to the receptor. Thus, this region is less trustworthy due to poor coverage, but still has an artificially high PSCF number. The four-point filter is less effective than the four-

trajectory filter at eliminating these cells, since multiple points are still available even if from a small number of trajectories.



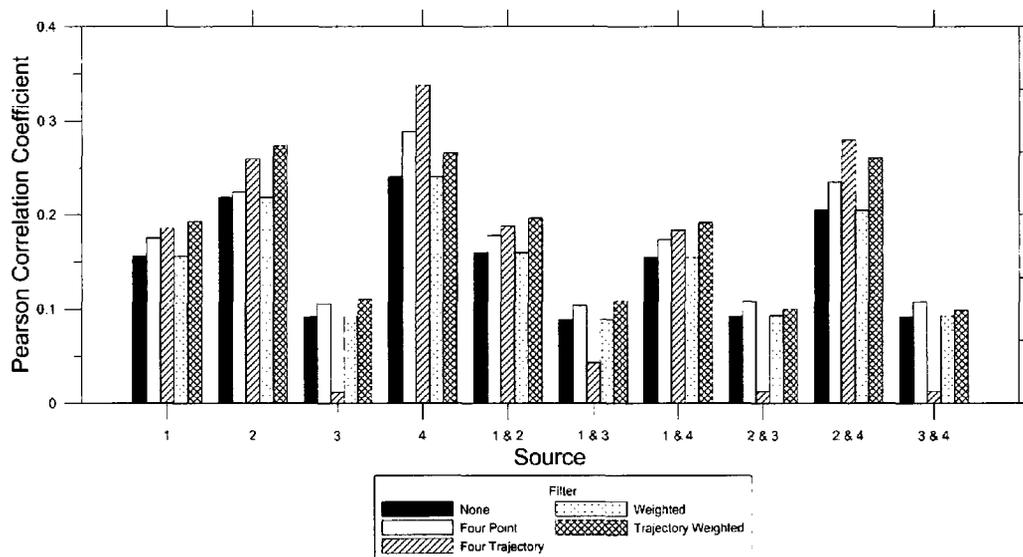
**Figure 5.24: PSCF Source 3, 97.5th percentile, 5 m cell size with no filter. (Pearson correlation = 0.10) An additional high area exists to the North-East of the source region.**



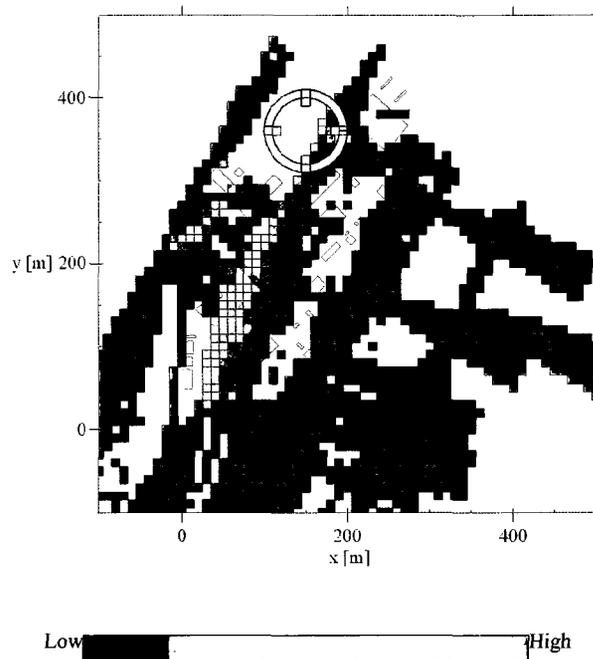
**Figure 5.25: PSCF Source 3, 97.5th percentile, 5 m cell size with four-trajectory filter. (Pearson correlation = 0.13) The four-trajectory filter removes a large portion of the additional high region which existed to the North-East of the source.**

Figure 5.26 shows the same plot using an identical number of sensors, but with trajectories calculated only at every 3rd measurement interval, which has the effect of increasing the time between measurements. With this reduced data set, the benefit of the filters is more clear. Removing areas with few trajectories leads to improvement using the point and trajectory filters for all cases except Source 3. Although the filters gave improved correlations for cases involving Source 3 when the full trajectory data set was considered, the opposite is true when using a reduced trajectory data set. This reason for

this effect is apparent in Figure 5.27, which shows results for Source 3 using the four-trajectory filter, and reveals that no results are now available in the region of the source, so that Source 3 cannot be located. Thus, in this case, the filter has removed the only information available to allow detection of Source 3.



**Figure 5.26: PSCF correlation with various filters for using every third measurement interval and 97.5th cut-off percentile and 5 m cell size. Reducing the data set increases the effect of filters for all sources. The four-trajectory filter greatly reduces the correlation for all cases involving Source 3**



**Figure 5.27: PSCF Source 3, 97.5th percentile, 5 m cell size from every 3rd measurement interval using four-trajectory filter. (Pearson correlation = 0.0) The four-trajectory filter has removed all cells in the vicinity of Source 3. The source can no longer be located.**

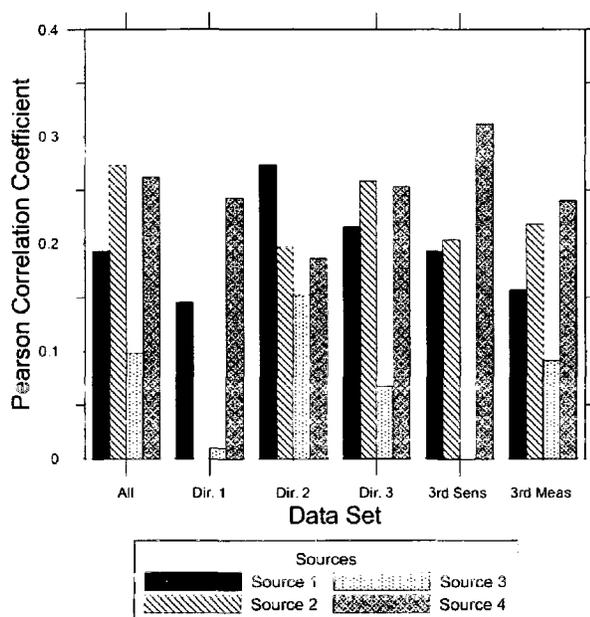
Using a trajectory weighting filter, like the four-point filter, improved the correlation for all sources from the unfiltered case and did not filter out Source 3 when using the reduced data set (Figure 5.26). However, as the trajectory weighting merely reduced the influence of regions containing poor trajectory statistics (rather than removing them completely), the apparent benefit was not as significant as the four-trajectory filter in regions with reasonable trajectory coverage (e.g. Source 4). Due to the position of the trajectories within the domain, using the four-trajectory filter removes the up and downstream smear from Source 4 while the actual source region is left alone, resulting in a higher improvement than with the trajectory weighting filter. It is noted that the use of the four-point filter netted similar correlation improvements as the

trajectory-weighting filter, although the trajectory-weighting filter arguably had better overall performance.

While using filters can remove spurious sources resulting from poor trajectory statistics, some work better than others in specific cases. Rather than blindly applying filters, it is apparent that filters should be chosen carefully in consultation with plots of the trajectory statistics to see where sources can still be located. Filters are primarily meant to remove spurious sources at the edges of the domain, not in the center, and if central regions are filtered out, more data collection is likely necessary. While the four-trajectory filter had the most dramatic improvements, it could also remove the source region entirely. If it is not feasible to improve the trajectory statistics for the use of the four-trajectory filter, the trajectory-weighting or the four-point filter should be used as both either improved the result or had negligible effect.

#### ***5.6.1.5 Data set effect***

Figure 5.28 shows the correlation differences when using data based on different numbers of trajectories and using trajectories originating from more than one dominant location. Set “All” uses trajectories and measurements corresponding to every sensor with data extraction every 30 seconds, evaluated for all considered wind directions. The directional data sets correspond to single wind direction quadrants (see Chapter 4), giving approximately 1/3 of the measurements as set “All”. The final two data sets also approximately use 1/3 of the original data set by only considering every 3<sup>rd</sup> sensor location (i.e. reducing the number of sensors) or only considering every 3<sup>rd</sup> measurement at each sensor (i.e. reducing the frequency of measurement).



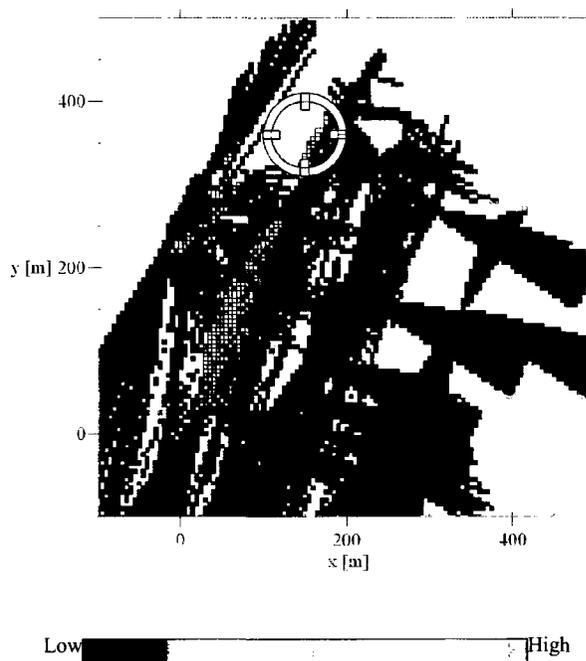
**Figure 5.28: PSCF correlation for various data sets using 97.5th cut-off percentile, 5 m cell size and no filter. Coverage from multiple directions gives more consistent results.**

The directional measurements give an indication of which directional data set is most important in determining a source location for the complete data set and how sensor location can affect results. Figure 5.28 emphasizes that while a single direction can give good results in a direction advantageous to identifying a particular source (typically when there are multiple sensors downwind of a source as in the case of Direction 2 with Source 1), poor results are achieved if the available data do not contain the sufficient coverage (e.g. Direction 1 for Sources 2 and 3). Thus, multiple direction coverage improves results in general. In a practical setting, the prevailing wind direction and likely source areas can be used to place sensors so that some of the sensors will always be downwind of these regions of interest for common weather conditions. Plotting the trajectory positions themselves will also indicate the areas where sources can be

identified and if it will be worthwhile to wait for different wind directions to improve trajectory coverage of the important potential source regions.

Reducing the number of sensors can have the same effect as reducing the directional information since domain coverage is also reduced. The results plotted in Figure 5.29 for Source 3 using every third sensor has an extremely low correlation value (see also Figure 5.28), while the other sources have correlation values over 0.175. In this case, Source 3 no longer has any polluted trajectory information available over the source region in order to make a good identification. The only sensor giving polluted readings is far downstream with few crossing trajectories leading to a false source between this sensor and the source region. Thus, lack of information in source regions not only prevents sources from being identified there, but can also lead to false predictions. If a single sensor provides all the high readings, the region it covers without other trajectories cannot be narrowed down to the actual source. Increasing the amount of directional information to get more domain coverage can be considered an alternative to increasing the number of sensors, although this can only be controlled in a practical setting with increased data collection times (i.e. waiting for a change in wind conditions). This effect is apparent in Figure 5.28 when comparing results for directional data sets with the complete data set for a fixed number of sensors where some sources cannot be found at all in some cases (e.g. Direction 1). When the number of sensors is reduced (such as the 3<sup>rd</sup> Sens. data set in Figure 5.28), sensor placement to obtain good domain coverage should be considered carefully. As fewer sensors are used, more care is required in their placement to obtain coverage with maximum effect. This should be done with

consideration of typical wind conditions and potential source locations (such as piping, buildings etc.).



**Figure 5.29: PSCF Source 3, 97.5th percentile, 5 m cell size using every third sensor. (Pearson correlation = 0.0) There are no trajectories passing over Source 3 so it cannot be found, though upstream and downstream smears are still visible.**

Comparing the full data set with data using only every third measurement interval, shows that reducing the number of measurements by increasing the time between them has a small negative effect on the correlation. The increased time between measurements results in a slight reduction in the coverage of the domain.

The preceding discussion shows that thorough domain coverage by trajectories is critical to making good source predictions. To achieve good spatial resolution, it is also critical to have information from multiple directions so that more than one sensor covers

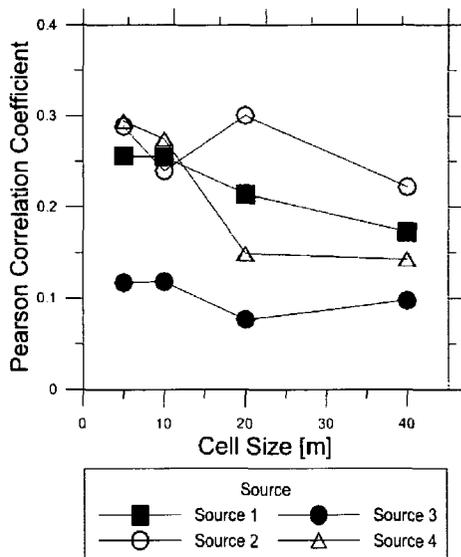
each part of the domain. This is necessary to allow distinction to be made between a source region and an upstream or downstream region along a single trajectory.

## **5.6.2 Concentration Weighted Trajectories**

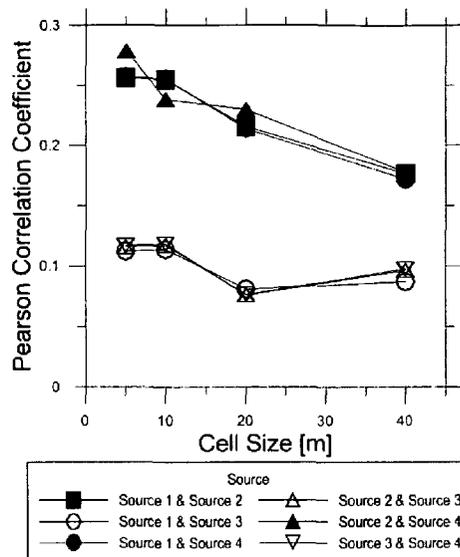
### ***5.6.2.1 Cell size effects***

The concentration weighted trajectories (CWT) algorithm is defined in a similar fashion to PSCF. However, each trajectory in CWT is weighted by the measured concentration while PSCF is weighted in a binary fashion using the pollutant cut-off.

Figure 5.30 shows the correlation values for single source releases using CWT. Comparing with Figure 5.12 and the PSCF correlation for various cell sizes, both graphs show similar trends. As with PSCF, there is a trade-off between small cell sizes (which improve the ability to pinpoint locations) and the number of trajectories within a cell (which give weight to the prediction in that area). For multiple sources (see Figure 5.31), the strength of the correlation is dominated by the accuracy with which the larger source is predicted, since the larger source had a emission rate ten times greater than the smaller source. Thus, the trends in Figure 5.31 involving large and small sources together (e.g. 1&2, 1&4, 3&2, 3&4) follow the trends for the large sources alone (e.g. source 1 and source 3) shown in Figure 5.30.

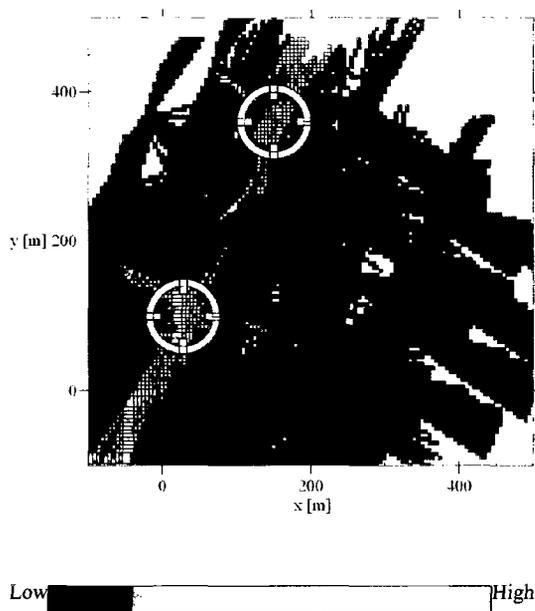


**Figure 5.30: CWT correlation vs. cell size for a single source using no filter**

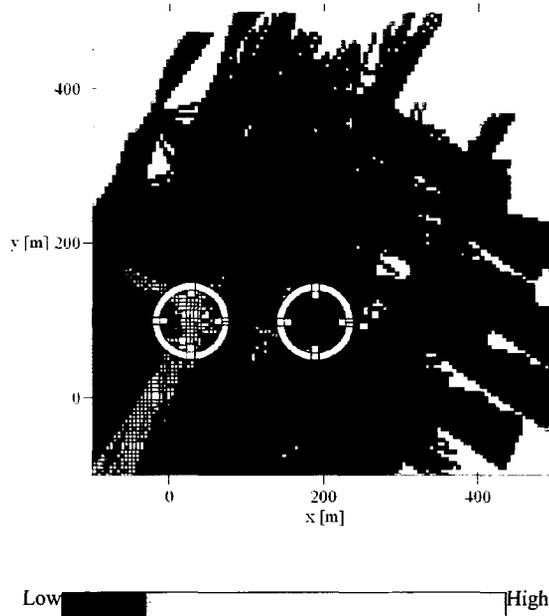


**Figure 5.31: CWT correlation vs. cell size for two sources using no filter**

With sources of the same magnitude (Figure 5.32), the high regions around the source are both distinct. Looking at Figure 5.33, the larger source does again have a tendency to obscure the smaller source (as was the case for PSCF as shown in Figure 5.15). As the large source is ten times the magnitude of the small source, the reconstructed field should have lower values in this region, but they are difficult to tell apart from the trails around the larger source.



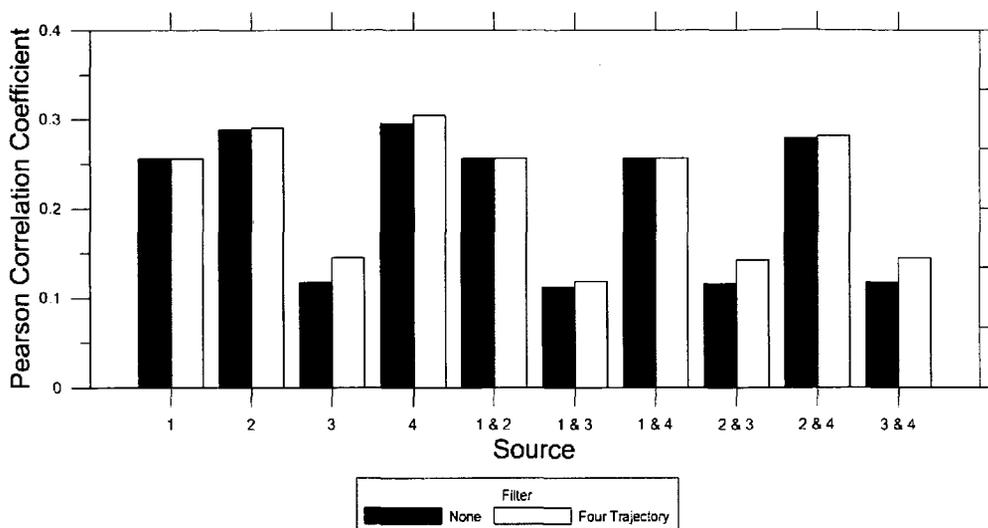
**Figure 5.32: CWT Source 1 and Source 3, 5 m cell size. (Pearson correlation = 0.11) The highest predicted region is located at Source 1. A high smear region exists around Source 3, but the source itself cannot be distinguished from the rest of the smear.**



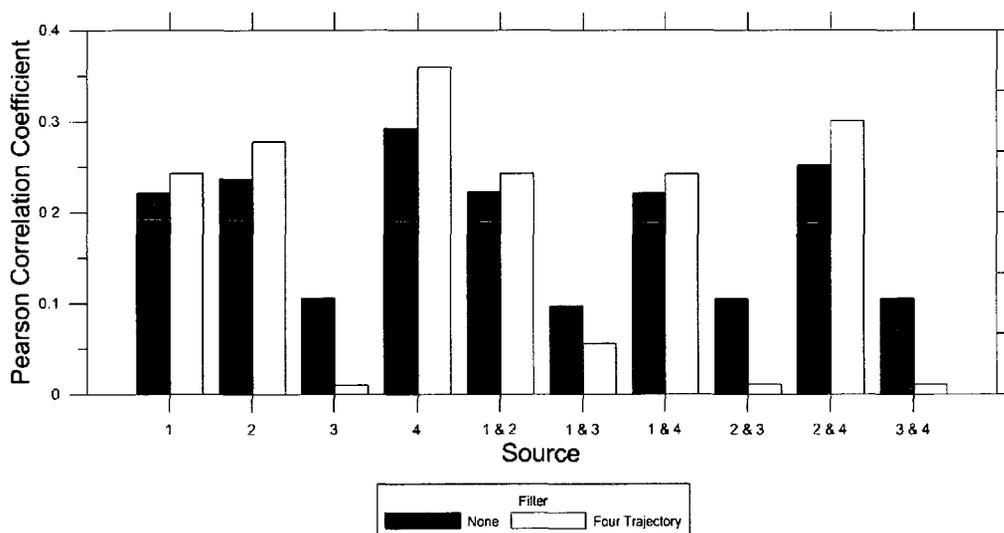
**Figure 5.33: CWT Source 1 and Source 4, 5 m cell size. (Pearson correlation = 0.26) The highest predicted region is located at "large" Source 1. "Medium" Source 4 cannot be distinguished.**

### 5.6.2.2 Filter effects

Figure 5.34 and Figure 5.35 show the effects of filters when using the larger and smaller data sets. When all data are considered, the four-trajectory filter improves results by eliminating cells with low trajectory count, as was the case with the PSCF (Figure 5.23). With the smaller data set, there are more cells eliminated with the four-trajectory filter and Source 3 can no longer be found, since less than 4 trajectories pass through the cells in this region. This is similar to what was found in PSCF calculations (Figure 5.27).



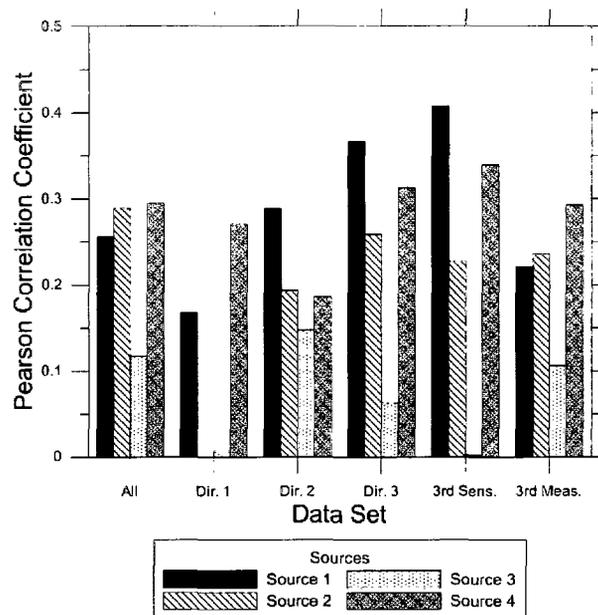
**Figure 5.34: CWT correlation for various filters and sources using 5 m cell size. The four-trajectory filter has negligible effect for cases other than those containing Source 3.**



**Figure 5.35: CWT correlation for various filters and sources using 5 m cell size and every 3<sup>rd</sup> measurement interval. The four-trajectory filter removes all information in the region around Source 3 so the correlation values are low.**

### 5.6.2.3 Data set effects

Figure 5.36 shows the correlation for runs using different data sets. As was the case with PSCF, using a larger number of trajectories from a variety of directions provided more consistent and significant results, though individual directions could show improvement for individual sources depending on their position. Overall, increasing the number of wind directions would provide better correlation significance for more source locations within the domain. Alternatively, resolution is improved with increased numbers of sensors. Again, it is not possible to find sources in regions where no trajectories are present, as was evident for Source 3 when only every 3<sup>rd</sup> sensor was used.



**Figure 5.36: CWT correlation for various data sets using 5 m cell size and no filter. Coverage from multiple directions gives more consistent results.**

### **5.6.3 Residence Weighted Trajectory Concentration**

RTWC is a reweighted version of CWT, so filter effects are similar and will not be discussed again in detail. It was found that cell size alone could not be separated from the reweighting length, so discussion of the two parameters has been combined.

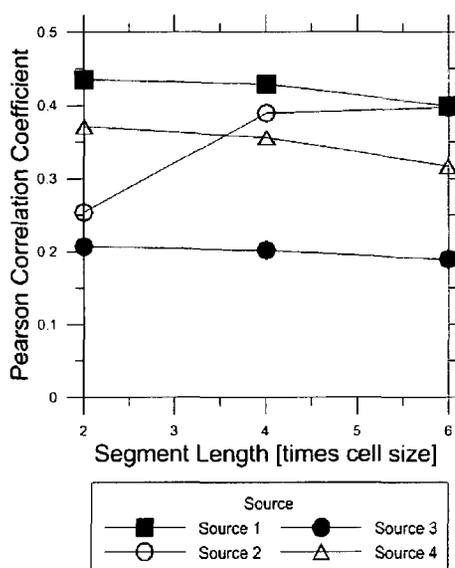
#### ***5.6.3.1 Reweighting segment length and cell size***

RTWC is based on the idea that if a high and low concentration trajectory pass over the same region, the source is unlikely to be at this location and instead is more likely to be on another part of the high concentration trajectory. Thus, the results field is reweighted with this in mind. To define the regions where trajectories pass over each other, the trajectories are divided into segments for reweighting. However, the reweighting segment length cannot be entirely separated from other parameters, since it relates to the size of the cell used to calculate the concentration field within the source location domain, as explained in Section 5.3.2.2 above.

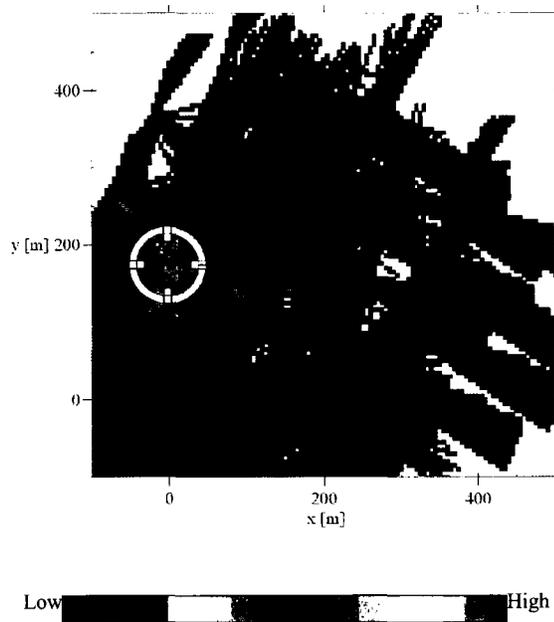
Figure 5.37 shows the change in correlation for various length trajectory segments when using the smallest sized cell for the source location domain. In general, the correlation tends to get worse as the size of the reweighting segment length is increased. While this was not significant in the case of Source 3, the same cannot be said universally. This decrease with segment length is not universally true as seen in Figure 5.37 for Source 2 alone. Figure 5.38 and Figure 5.39 give the field plots for two different iterations of the redistribution for Source 2. When using the smallest cell size, the initial plot field (Figure 5.38) shows that there are high probability source location regions near the source but also in a lone orange cell near the center of the domain. Each iteration increases the concentration value (i.e. the probability of a source) in this lone cell, by redistributing

concentrations from the nearby cells (Figure 5.39). By iteration 10, the actual source region is now no longer as easily distinguished.

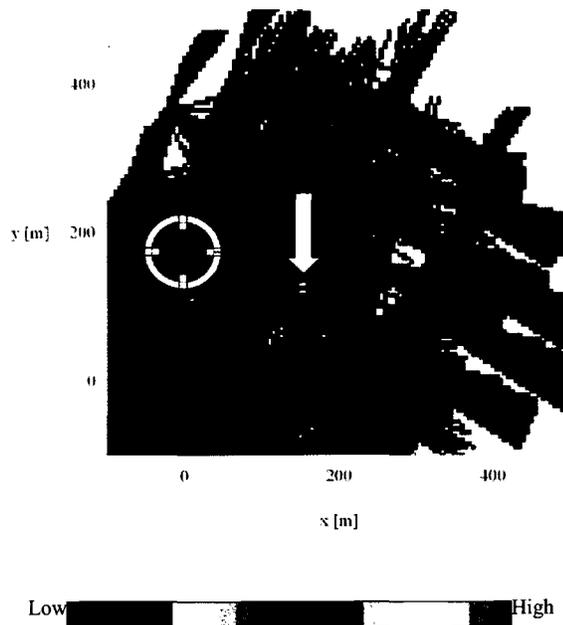
However, when filters are used which remove this spurious source as shown in Figure 5.40, the actual source region is again quite visible compared with the surroundings. This improvement comes at the expense of reduced trajectory coverage in the domain as previously discussed for PSCF and CWT.



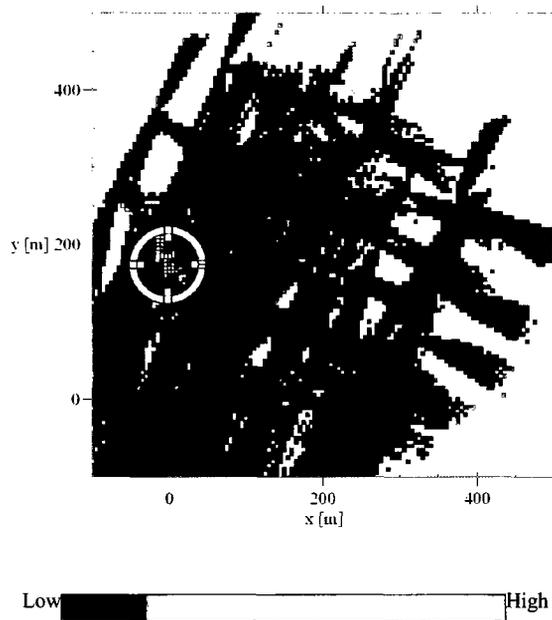
**Figure 5.37: RTWC correlation vs. segment length for cell size 5 m and no filter. Poor trajectory statistics cause a spurious source prediction for Source 2 when the segment length is small.**



**Figure 5.38: RTWC Source 2, 5 m cell size, 2x cell size segment length, iteration 1. High smear is visible at Source 2.**

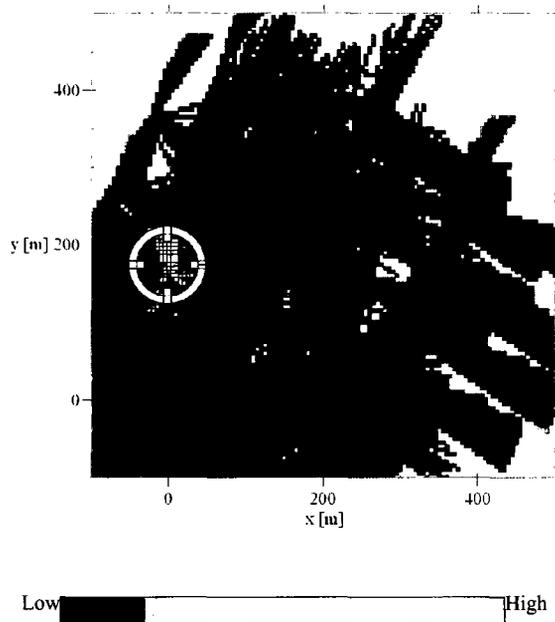


**Figure 5.39: RTWC Source 2, 5 m cell size, 2x cell size segment length, iteration 10. Highest point is in domain center (arrow). Source 2 region is now difficult to distinguish.**



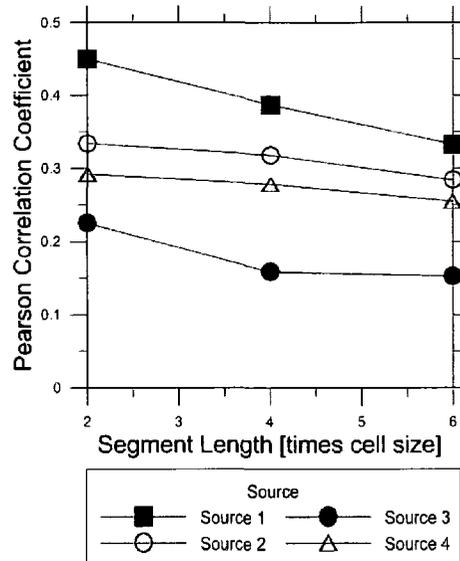
**Figure 5.40: RTWC Source 2, 5 m cell size, 2x cell size segment length, iteration 10 using four-trajectory filter. Removing spurious source with the four-trajectory filter reveals a secondary high smear at the actual source.**

When the reweighting segment length is increased (Figure 5.41), the spurious source area is smeared out since the surrounding field (which is counted as part of the same trajectory segment) has low concentration values.



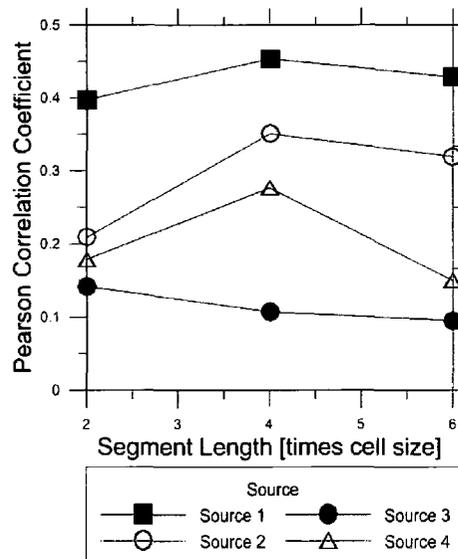
**Figure 5.41: RTWC Source 2, 5 m cell size, 4x cell size segment length, iteration 10. Increasing the weighting length removes the spurious high prediction found with the 2x cell size segment length.**

This issue was not observed when calculations were performed using a larger cell size for the concentration field, as is evident when comparing Figure 5.42 (10 m cell size) with Figure 5.37 (5 m cell size). With a larger number of trajectories per cell, the effect of the spurious false source is eliminated. Under these conditions shorter trajectory reweighting lengths universally improve the correlation.



**Figure 5.42: RTWC correlation vs. segment length for cell size 10 m and no filter**

Figure 5.43 shows a plot of correlation versus segment length for the data set using every third measurement interval. The reduced correlation coefficient for the 2x reweighting segment length compared with the 4x segment length now can be seen for three of the four source predictions (as compared with just Source 2 in Figure 5.37). Thus, if trajectory statistics are poor, spurious sources are more common but their effect can be reduced by increasing the trajectory reweighting length.



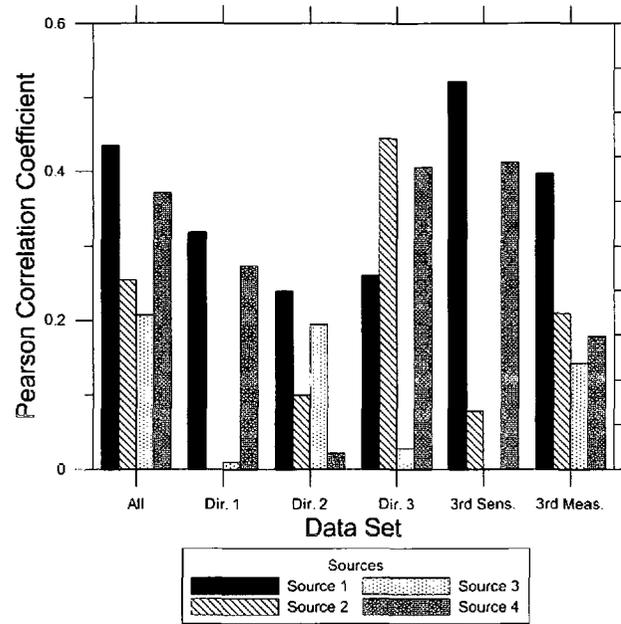
**Figure 5.43: RTWC correlation vs. reweighting length for cell size 5 m, no filter and every 3<sup>rd</sup> measurement interval. Spurious sources appear in regions of poor trajectory statistics for Source 1, Source 2 and Source 4 using the smallest segment length.**

While smaller reweighting lengths work well in cases where there are multiple trajectories for the cells within the domain, when there are regions with poor trajectory statistics, spurious sources can result. Thus, before determining whether to use a smaller or larger trajectory length, a plot of the trajectory coverage should be investigated. For the current domain, in cases where there were regions of the domain with poor coverage, a reweighting length of four times the cell size was a good compromise between the improvement of shorter reweighting lengths and the effect of poor trajectory coverage. Though the increased reweighting length performs slightly worse when there is a great deal of information known about the section of interest of the domain, it improves correlation when there are areas where less information is available. Increasing the cell size also gives a similar effect of reducing spurious sources by increasing the number of trajectories per cell, but this reduces the spatial resolution. This may be necessary if there

are no or single trajectories in many cells of the domain. If there are multiple trajectories passing through all cells in the domain of interest where sources can be present, the reweighting segment length should be reduced from 4x to 2x the cell size. While 1x cell size was not tested, lower values may continue this improvement.

### ***5.6.3.2 Data set effects***

Comparing Figure 5.44 for the RTWC algorithm with Figure 5.36 for the CWT algorithm, reweighting sometimes narrows the region around the desired source but, in other cases, reduces the strength of the correlation. If the CWT correlation was already above 0.25, the RTWC reweighting improves it further. However, for lower correlations caused by poor trajectory statistics, RTWC may cause spurious source regions and RTWC cannot correct for regions where there is no information. As spurious sources may arise in regions of poor trajectory coverage, the present results suggest that RTWC should not be used alone, but in conjunction with the CWT plot. In addition, trajectory coverage plots should be used in advance of both CWT and RTWC to determine if there are multiple trajectories in the cells of the domain of interest and, if not, to wait for wind conditions which will allow for trajectory coverage of these areas (or to potentially exclude these regions from consideration).

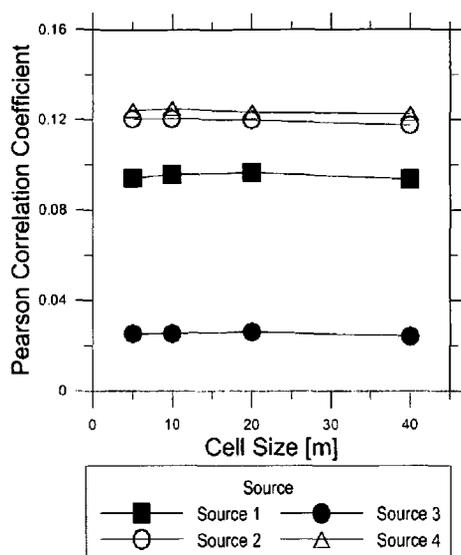


**Figure 5.44: RTWC correlation for various data sets for cell size 5 m and reweighting length of 2x cell size. Multiple wind directions provide more consistent results as does increased number of sensors.**

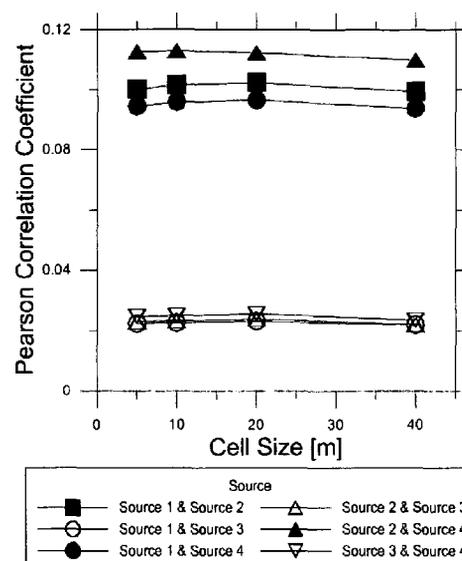
## 5.6.4 Quantitative Transport Bias Analysis

### 5.6.4.1 Cell size effects

Figure 5.45 and Figure 5.46 show the effect of cell size on the QTBA source prediction. There is no significant effect observed with the changing cell size, and slight variations are likely due to the positioning of the cell centers in comparison with the source point location. QTBA is calculated at the center points of the cells, rather than considering the cell size so this lack of variation is not unexpected since the larger grid cells share center points with the smaller cell cases.

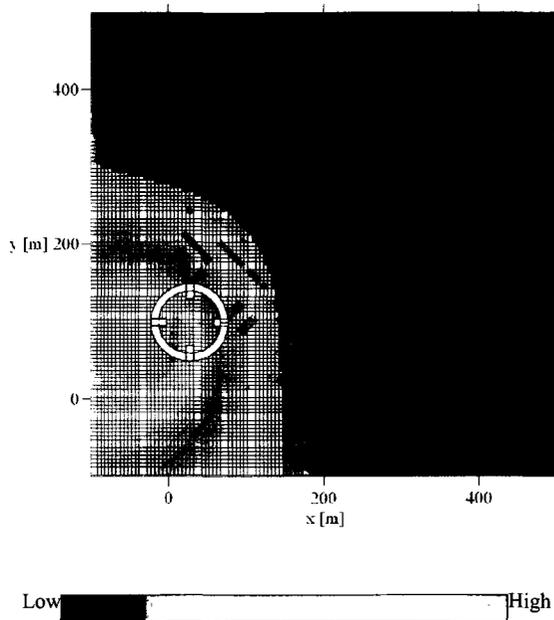


**Figure 5.45: QTBA correlation vs. cell size for a single source using 1.5 m/s deviation coefficient and no filter. Cell size has no significant effect.**

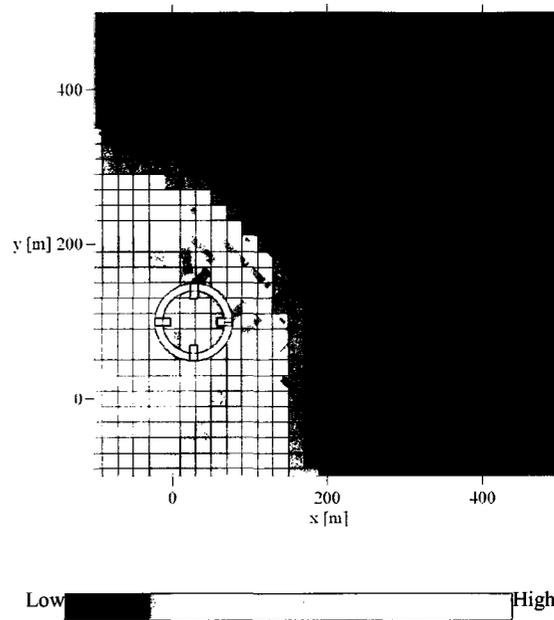


**Figure 5.46: QTBA correlation vs. cell size for two sources using 1.5 m/s deviation coefficient and no filter. Cell size has no significant effect.**

Figure 5.47 and Figure 5.48 show the field plot variation with changing cell size for a prediction of Source 1. Increasing the size makes the contours of the various regions jagged in comparison, but does not change the source prediction location. Similar results were apparent for the other test cases as well. Again, since the QTBA prediction is based at the cell centers and increasing the number of cells can only improve resolution rather than change the source prediction locations.



**Figure 5.47: QTBA Source 1, 1.5 m/s deviation coefficient, 5 m cell size (Pearson correlation = 0.09)**

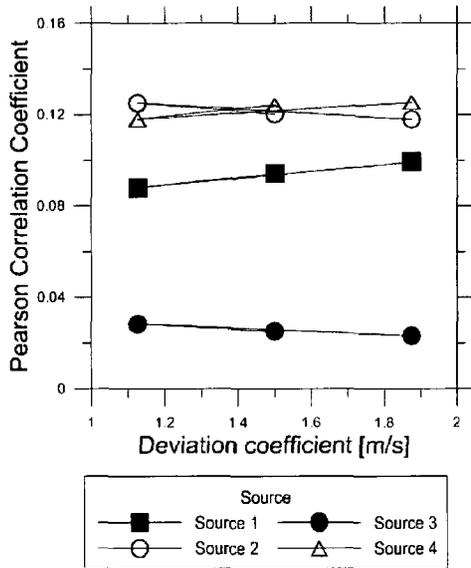


**Figure 5.48: QTBA Source 1, 1.5 m/s deviation coefficient, 20 m cell size. (Pearson correlation = 0.10) Center of high prediction region does not change from 5 m cell size.**

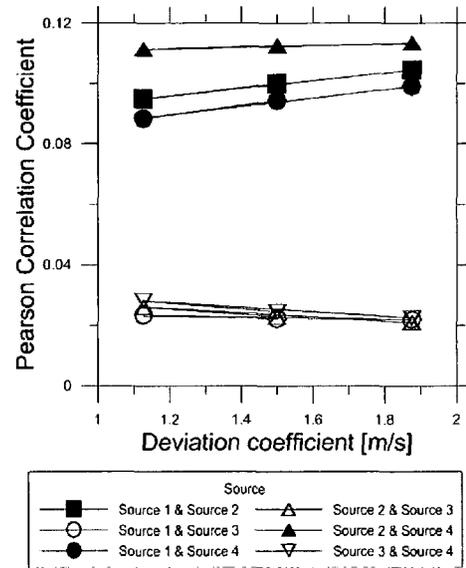
#### **5.6.4.2 Trajectory standard deviation coefficient**

Figure 5.49 and Figure 5.50 show the change in correlation with changing the standard deviation coefficient used along each trajectory. As this coefficient increases, the spread of the trajectory's influence increases faster with time upstream. Figure 5.51 and Figure 5.52 show plots for Source 3 for different coefficients using the finest grid. As the deviation coefficient is increased, the prediction moves closer to the receptors as its influence decreases faster along its length. Thus, the changes in correlation seen in Figure 5.49 and Figure 5.50 are determined by whether moving the prediction inward toward the receptors takes it closer or farther away from the actual source. Thus, if the

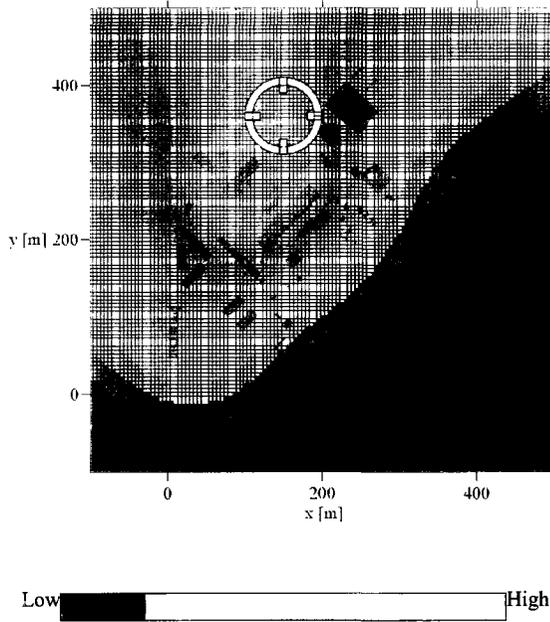
source is far from the center of the receptors, increasing the coefficient decreases the correlation.



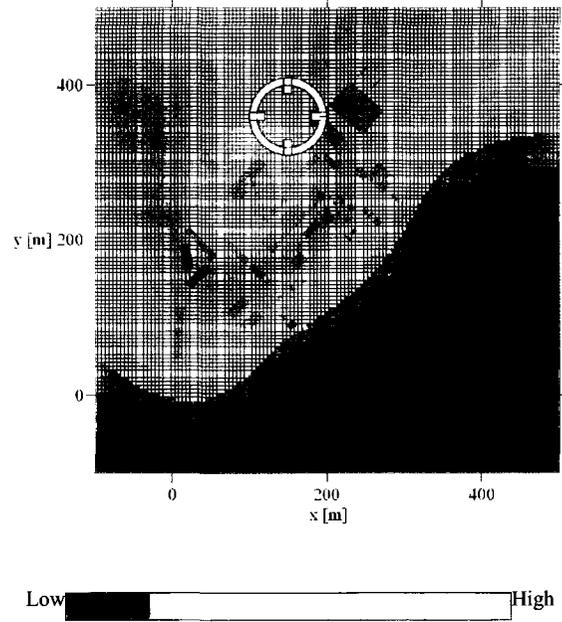
**Figure 5.49: QTBA correlation vs. deviation coefficient for a single source using 5 m cell size**



**Figure 5.50: QTBA correlation vs. deviation coefficient for two sources using 5 m cell size**



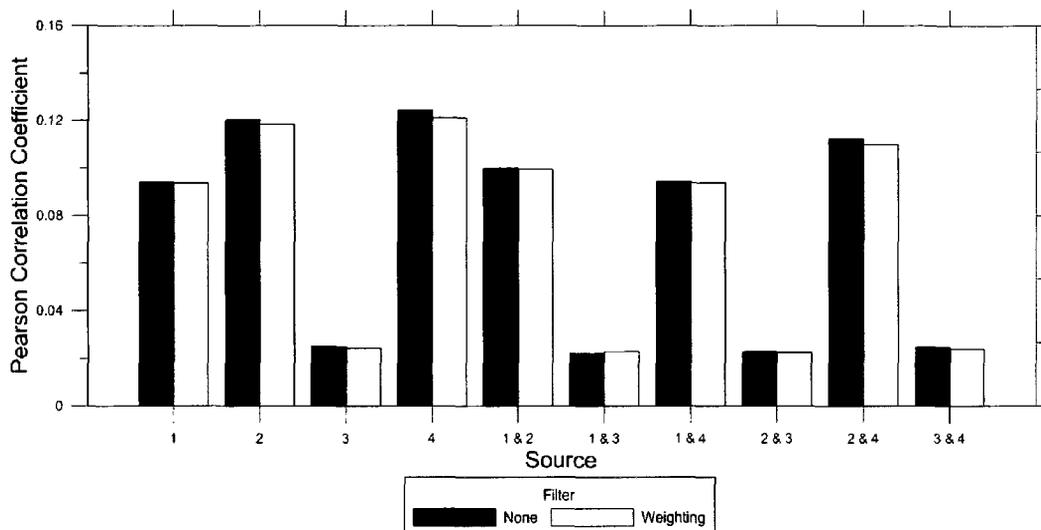
**Figure 5.51: QTBA Source 3, 1.125 m/s deviation coefficient, 5 m cell size (Pearson correlation = 0.03)**



**Figure 5.52: QTBA Source 3, 1.875 m/s deviation coefficient, 5 m cell size. (Pearson correlation = 0.02) Increasing the deviation coefficient moves the predicted high region inward.**

#### **5.6.4.3 Filter effects**

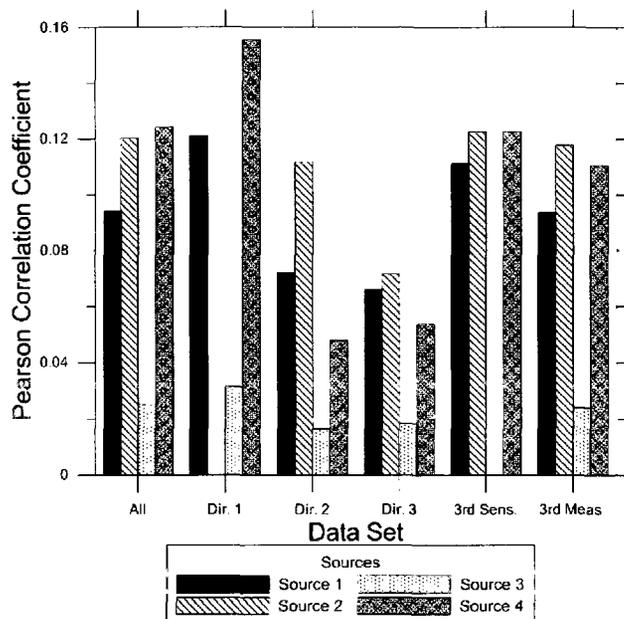
Figure 5.53 shows the effect of the weighting filter on the correlation between the prediction and the expected source location. As Figure 5.51 and Figure 5.52 illustrated, the shape of the QTBA prediction radiates outward from a single source point prediction. The weighting filter spreads this pattern even more, although the effect is not significant (correlation coefficient changes on the order of 0.01). The effect is small since the predicted location does not change and, as there was no noise to smooth out, the QTBA weighting filter only reduces the correlation as the predicted source region becomes larger.



**Figure 5.53: QTBA correlation for various filters for a single source using 5 m cell size and 1.5 m/s deviation coefficient. The weighting filter had no significant effect.**

#### *5.6.4.4 Data set effects*

Figure 5.54 reveals a similar trend to that observed in Figure 5.28 for PSCF, in that increasing the amount of data from multiple directions gives more consistent results for all sources.



**Figure 5.54: QTBA correlation for various data sets using a 1.5 m/s deviation coefficient, 5 m cell size and no filter. Multiple wind directions gave more consistent results.**

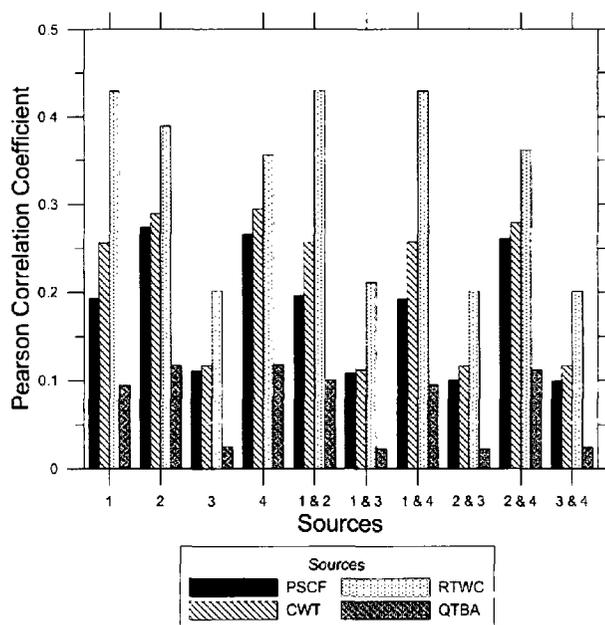
## 5.7 Overall algorithm comparison

### 5.7.1 Correlation

Figure 5.55 compares the correlation values achieved by each algorithm for combinations of sources using the complete data set. Each algorithm was set using the optimum parameters / compromises discussed in the previous section, as summarized in Table 5.3.

**Table 5.3: Algorithm parameters**

Algorithm	Cell size[m]	Filter	Algorithm specific
PSCF	5	trajectory weighted	cut-off percentile: 90 <sup>th</sup>
CWT	5	none	N/A
RTWC	5	none	redistribution length: 4x cell size
QTBA	5	none	deviation coefficient: 1.5 m/s



**Figure 5.55: Comparison of algorithms using optimum parameters for calculations on the complete data set. RTWC performs better in all cases when there are good trajectory statistics. QTBA did not perform well when scaled down to a building scale problem.**

In every case for the full data set, RTWC provides a higher correlation value than the other three algorithms. If the segment length is reduced to two times the cell size, this remains true except for Source 2 alone, where the RTWC correlation is slightly less than CWT and PSCF. CWT provides the next best results after RTWC. As CWT must be calculated before running the RTWC reweighting, it could be used as a check on the RTWC results with no additional computations for cases where RTWC does not perform as well due to poor trajectory statistics. The PSCF correlation is always lower than that of CWT. This is due to the PSCF pollutant percentile cut-off, which is difficult to optimize and does not need to be included for CWT to run. QTBA always correlates

poorly as compared with the other three algorithms for every source combination considered.

### 5.7.2 Run Times

One of the preliminary questions asked in this document was whether the algorithms could be performed in a reasonable amount of time (i.e. in under a day). This proved to be true for all cases. There was still, however, a significant difference between some algorithms in terms of run time. Table 5.4 presents run time data for each algorithm, calculated for the finest grid case for Source 1 which was representative of the entire data set. In calculating these data, for each algorithm the middle algorithm specific variable value was chosen (i.e. pollutant cut-off, segment length, and deviation coefficient). QTBA was run only on the section of the domain used for the correlation comparison. Each run was performed on the same computer and implemented using the C++ programming language.

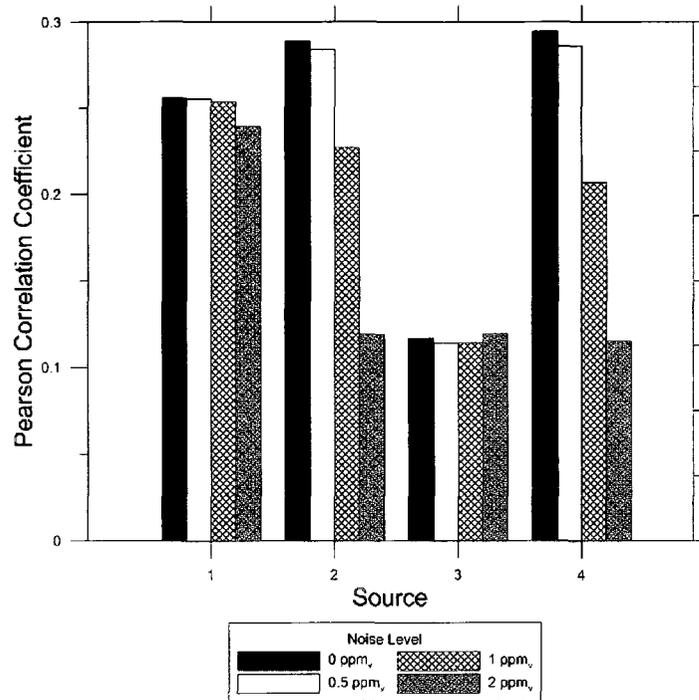
**Table 5.4: Algorithm runtime**

<b>Algorithm</b>	<b>Run Time (s)</b>
PSCF	1
CWT	1
RTWC	2653
QTBA	6847

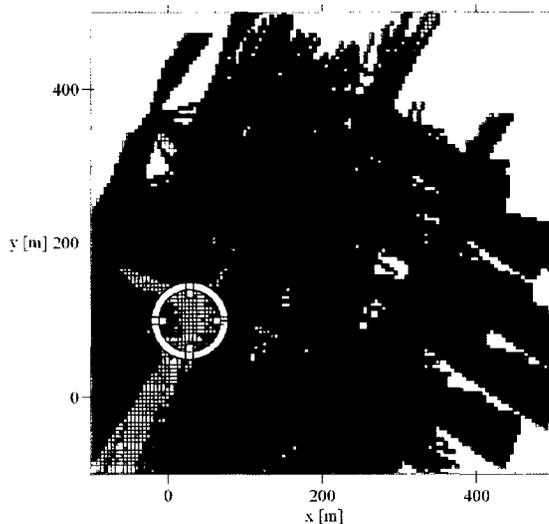
Both CWT and PSCF run very fast with more time taken to read in the data set than to run the algorithm itself. This is significantly faster than the RTWC reweighting which itself is significantly faster than the QTBA calculation.

### 5.7.3 Measurement noise

Previous results considered the concentration measurements as exact values. In reality, measurement noise will be present in the concentration values. To test the sensitivity of the CWT and RTWC algorithms to measurement noise, a Gaussian distribution about zero with various standard deviations (noise levels) was applied to the concentration values used by the two algorithms. Noise levels of 0.5, 1.0, 1.5, and 2.0 ppm<sub>v</sub> were considered, corresponding to a range thought to be readily achievable with modern detectors. Figure 5.56 shows the correlation value for the CWT algorithm for various levels of noise. For the large sources, only the 2 ppm<sub>v</sub> noise case had a significant effect on the algorithm's correlation for Source 1. Even in this case, the predicted source location was merely spread out and could still be identified (Figure 5.57). The effect was much greater on the medium sources as the noise levels became the same order of magnitude as the original measurements (Figure 5.58).

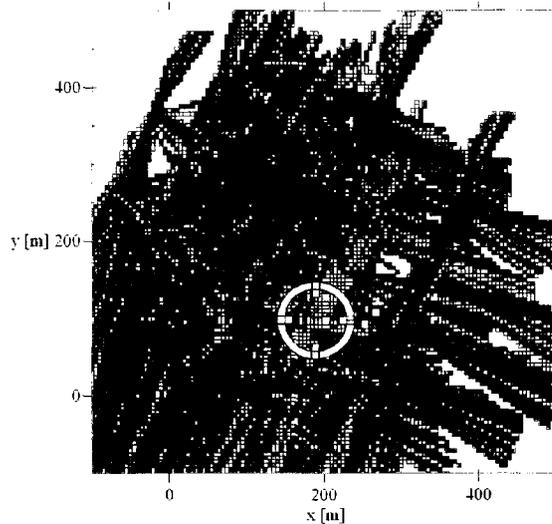


**Figure 5.56: CWT correlation for various noise levels, 5 m cell size. Source regions can still be identified for noise levels up to 2 ppm for the large sources. For the medium sources, once the noise reached 1 ppm levels, the true values were of the same scale as the noise and false source regions began to appear. At a noise level of 2 ppm, these false sources became common and the actual source region became difficult to distinguish.**



Low  High

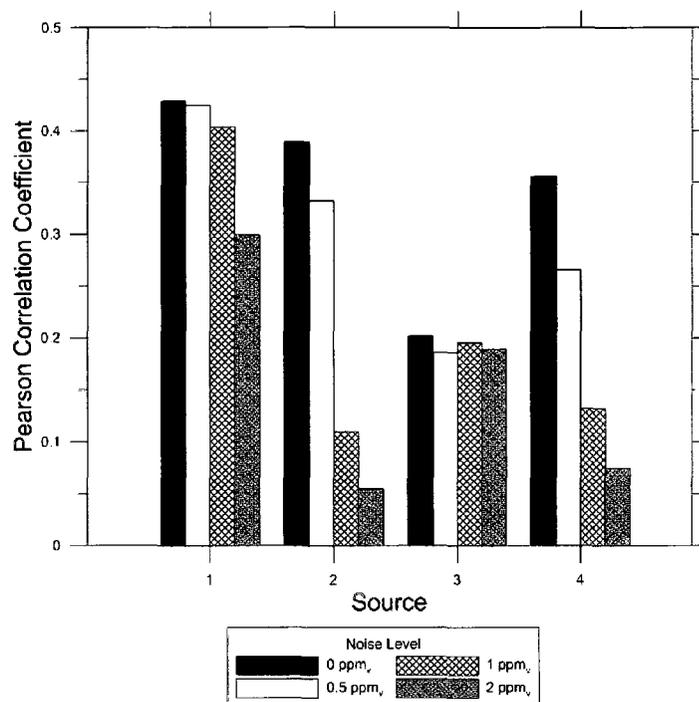
**Figure 5.57: CWT Source 1, 5 m cell size, 2 ppm<sub>v</sub> noise level. (Pearson correlation = 0.24) Source 1 is still clearly visible.**



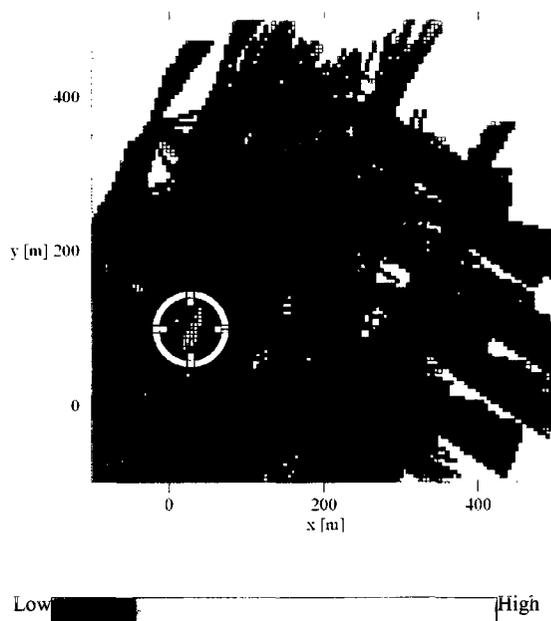
Low  High

**Figure 5.58: CWT Source 4, 5 m cell size, 2 ppm<sub>v</sub> noise level. (Pearson correlation = 0.12) Noise related highs are the same intensity as the region at Source 4.**

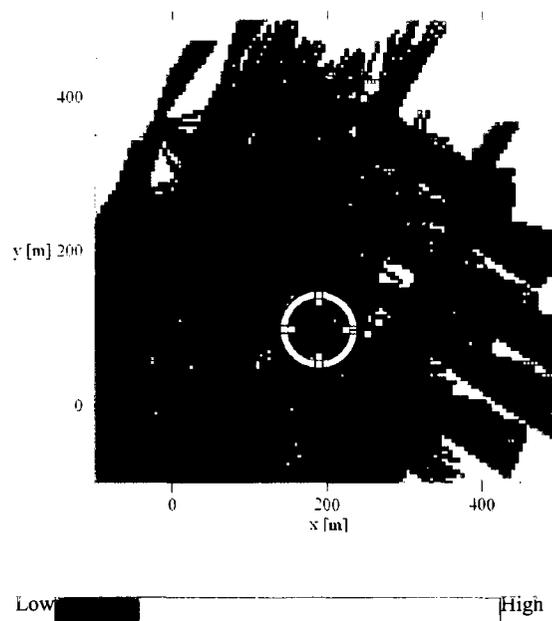
Figure 5.59 plots the RTWC results for the same cases. The large source can still be found even at the highest considered noise levels of 2 ppm<sub>v</sub> (Figure 5.60). However, RTWC was in general more sensitive to noise levels than CWT, as the reweighting could emphasize the false sources found in the CWT field (Figure 5.61).



**Figure 5.59: RTWC correlation for various noise levels, 5 m cell size, 4x cell size segment length. Source regions can still be identified for noise levels up to 2 ppm for the large sources. For the medium sources, once the noise reached 1 ppm levels, the results became sensitive to the false sources which appeared in the CWT results.**



**Figure 5.60: RTWC Source 1, 5 m cell size, 10 m segment length, 2 ppm<sub>v</sub> noise level (Pearson correlation = 0.27)**



**Figure 5.61: RTWC Source 4, 5 m cell size, 10 m segment length, 2 ppm<sub>v</sub> noise level (Pearson correlation = 0.07)**

As the 0.5 ppm<sub>v</sub> level is a conservative noise estimate for the measurement system under consideration, the addition of measurement noise should not prevent the sources of the sizes under consideration from being identified. Even at 2 ppm<sub>v</sub> levels, the large sources can still be found.

#### 5.7.4 Overall comparison

Of the tested algorithms, RTWC allows for the most precise source discrimination as long as there is good trajectory coverage of the domain. The run times for RTWC would be achievable in the envisioned practical implementation to semi-continuously monitor fugitive emissions in an industrial setting. RTWC should however be used in conjunction with CWT and trajectory statistics plots as a secondary verification to ensure that trajectory coverage is sufficient to avoid identification of spurious sources. Above all,

the results show that location of sensors and sufficient wind variation are critical to achieving good results. Advance determination of a map of possible source locations, and locating sensors to attempt to achieve a minimum number of trajectories in the cells of those regions (i.e. making sure that trajectory coverage is maintained even when using a four-trajectory filter, for example) before running the source location algorithms is highly recommended. Encouragingly, the results show that with further development it should be possible to detect fugitive emissions in an industrial setting with sufficient precision. Speed of the algorithms themselves appears to be advantageous over existing approaches mostly based on manual searches with hand-held detectors. If wind simulation time is taken into consideration, developing a database of pre-computed flows may take several months. However, these only need to be performed once on a given site during initial setup.

## **6 Conclusions and Recommendations**

### **6.1 Conclusions**

The original question asked at the beginning of this project was whether it was possible to use trajectory statistical methods to locate fugitive emission leak locations for an idealized case with complex geometry as proof of concept. Results have shown that PSCF, CWT and RTWC can find leak locations on a building scale when provided with sufficient trajectory coverage of the domain corresponding to concentration measurements. QTBA does not work well in identifying individual source locations within a gas plant geometry even for this idealized case. It does not appear that this method can be directly downsized to a building scale application.

Each of the algorithms ran in a reasonable time. PSCF and CWT were near instantaneous, with more time spent setting up the problem and reading in concentration, geometry and trajectory data than running the algorithms themselves. RTWC run times depended on the number of iterations required to reach convergence which could range from a few minutes to several hours for the data sets used. QTBA also ran on the scale of a few hours for the data sets used.

PSCF, CWT and RTWC were able to locate sources provided that there was sufficient trajectory coverage of the regions of interest within the domain. This coverage depended on the wind direction, the measurement intervals, the number of sensors, and the positions of the sensors. Sensor positioning increases in importance as the number of sensors decreases and typical wind directions should be considered so that trajectory coverage of the domain is maintained. In the range of test cases considered, the original

31 sensors were able to locate the source regions for three of the four sources alone, while the fourth had trailing effects upstream where there was poor trajectory coverage. Decreasing the sensor count to 1/3 of the original, three sources were still located fairly well, but there was no information available to locate the fourth source. Thus, adding one or two additional sensors to cover this region would provide sufficient coverage in the case of this set of leaks. Coverage of a region from multiple wind directions provides better leak location abilities than from a single direction so that sources can be ruled out in upstream and downstream regions. For multiple leak sources, large source tend to smear out smaller sources. However, these large leaks are the most important emission contributors as the majority of fugitive emissions are often the result of a small number of large leaks (Ross and Picard, 1996; Picard et al., 1998).

In general, trajectory statistics can be used to determine whether sufficient information has been gathered to locate sources within the domain or if more measurements are required. While multiple directional information is desirable for all areas to reduce the upstream and downstream smearing of source predictions, even the smeared predictions are useful in reducing the search space.

Of the four algorithms, RTWC in general produced the best correlation results when locating sources, though CWT performed better in the case of Source 2 where trajectory coverage was poor when a short segment length was used. When noise is added to the calculation, CWT was less affected than RTWC. As CWT is computed as part of the RTWC calculation, it could be used to provide a useful check for the resulting prediction fields.

While the data set used was the most important parameter for the identification of sources in all cases, the other parameters could be optimized for better correlation values. For PSCF, the use of the four-point filter and the trajectory weighting filter eliminated some false positives in the smaller data sets and in regions with few trajectories where little information was known, without eliminating large regions of the domain where sources were located. This indicates that it may be useful to extend these filters to use with CWT and RTWC. It was not possible to determine an optimal pollutant cut-off range for all cases, and this is a general practical limitation of the PSCF algorithm. It was also noted that the concentration detection threshold of the measurement system influences the choice of percentile value in the PSCF calculation. Although percentile values are commonly used in the literature, it seems more appropriate to specify PSCF cut-offs based on actual concentration values.

CWT performed better with a 5 m cell spacing and no filter as these reduced the domain coverage too much for the smaller data sets. The reweighting length for RTWC should be at least 4 times the cell size for smaller cells, though larger cells can be used to compensate for the number of trajectories passing through cells, in order to smooth out spurious regions away from the source. Using small reweighting lengths can lead to local highs in single cells that have few passing trajectories.

Overall, RTWC combined with CWT and trajectory statistics plots show promise in greatly reducing the search space for large fugitive emissions sources in an industrial setting. In the ideal circumstances considered, sources could be distinguished within 5-10 m if good trajectory coverage was available. Run times are feasible ranging from a

few minutes to a few hours. Even with increased data set size, this is still significantly faster than searching using manual methods alone.

## **6.2 Field set-up recommendations**

Based on the performance of the trajectory statistical algorithms, some recommendations can be made for attempts to use them in the field.

### **6.2.1 Placement of Sensors**

Since trajectory coverage was clearly identified as having a strong influence on results, it is worthwhile to plan the sensor placement carefully. As of yet, the best method for determining the wind field for the trajectory calculations is an ongoing challenge for application in the field. However, based on the current research, the following steps are recommended prior to attempting field measurements:

1. Determine regions onsite which could contain sources. These are the regions where trajectory coverage is key. Spending money on sensors to cover regions which cannot have leaks is a waste of resources which would be better spent increasing the sensor coverage of the rest of the domain.
2. Investigate typical wind conditions onsite. Generate a wind rose with the percentage of time the wind blows in each direction to help optimize sensor placement.
3. Based on the potential source regions and the wind rose, trajectory coverage for various sensor positions can be estimated. A simple, preliminary method would be to use straight lines in the prevailing directions corresponding to the percentage of time it is expected for the wind to be in each direction.

4. Based on these tentative sensor positions and predicted coverage, attempts could also be made to optimize sensor position choice to provide coverage for the important regions within the domain. While having coverage of the important domain regions is desirable from multiple directions, it is more important to have coverage of these regions from at least a single direction. Even if the source cannot be pinpointed and is smeared up and downstream, this still narrows the search region to find the source.

### **6.2.2 Interpretation of Results**

As measurements are made and corresponding trajectory calculations performed, plotting the trajectories passing through each cell can provide insight into whether sufficient data has been obtained to make a source estimation. In areas with insufficient coverage, it will be necessary either to wait for the winds to shift until this region or to reposition sensors. Data for regions with a large number of trajectories are more trustworthy. Sources in regions with no trajectories may still lead to source identification in regions upstream and downstream from the actual source. Thus, interpreting a results plot for a small number of trajectories can be quite challenging and becomes easier as the amount of information about the domain increases. The cost of investigating potential source regions which could be false must be weighed against waiting for more measurements and trajectories. If a sufficient number of trajectories are available to maintain coverage of the important regions, using a trajectory filter can eliminate false sources at the edges of the domain, but if few trajectories are available in the coverage areas desired, these filtered results should only be used in conjunction with the unfiltered field rather than eliminating information of the region entirely.

### 6.3 Recommendations for future work

This study was essentially a proof of concept for the use of source location methodologies on a building scale, which allowed for the elimination of some algorithms from consideration. However, given the importance of domain coverage as revealed in the results, it would be worthwhile to look in greater depth at source smearing and filter optimization for a larger number of data and source sets. The question of how much data coverage is required to pinpoint a location within a few meters, rather than tens of meters requires further study to answer.

While PSCF did not perform as well as CWT or RTWC, the binary nature of the PSCF method may also be useful for investigation of upstream and downstream smearing due to trajectory coverage. If all trajectories passing over a cell are considered polluted, the trailing pattern can be matched with that predicted with CWT to see how the trajectory coverage is influencing the source prediction.

One of the greatest challenges remaining to practical application of this approach using any algorithm including those discussed in the literature review, is the wind field determination. Going from point wind measurements at a few sensors within the domain to a complete wind field is not simple. Computational wind engineering as performed for the synthetic data generation is time consuming and computationally costly. While vorticity conservation methods may offer potential speed up of computational time and are being investigated in related work, these approaches are unlikely to be sufficient to enable wind computations on the fly. It is instead likely that a pre-computed database of wind fields corresponding to different wind conditions (direction & velocity) would be used to characterize the flow field of the domain. Trajectories could then be computed in

advance based on these data and would require only a small amount of data storage. In the present work, this was essentially performed using three, distinct, short wind-field computations from three different dominant directions. However, the performance of various source location algorithms with varying quality wind field data thus needs to be investigated.

The use of the adjoint of the advection-diffusion equation may also provide a way to narrow results further. Using variational data assimilation with the adjoint equation, in combination with the current algorithms to generate an initial guess, may allow for the estimation of source strengths as well as location.

Finally, after having shown that it is possible to locate source regions for the ideal case, algorithm testing on a real case would be desired. The Joint Urban 2003 study in Oklahoma City may provide an interesting data set for dispersion among buildings as a preliminary test case based in the physical world where errors in the wind field would matter, before moving on to a full physical experiment testing the methodology. In cases of wind field and trajectory uncertainty, multiple trajectories with a random element for each concentration measurement (Lagrangian inverse plume) are another possible area of investigation.

## References

- ANSYS. (2008). *ANSYS CFX, Release 11.0 user manual*. Canonsburg: ANSYS Inc.
- ANSYS. (2009). *ANSYS ICEM CFD, Release 12.1 user manual*. Canonsburg: ANSYS Inc.
- Allen, C. T., Young, G. S., & Haupt, S. E. (2007). Improving pollutant source characterization by better estimating wind direction with a genetic algorithm. *Atmospheric Environment*, *41*, 2283-2289.
- Apadula, F., Gott, A., Pignini, A., Longhetto, A., Rochett, F., Cassardo, C., et al. (2003). Localization of source and sink regions of carbon dioxide through the method of the synoptic air trajectory statistics. *Atmospheric Environment*, *37*, 3757-3770.
- Ashbaugh, L.L. (1982). *Large scale transport of fine sulfur particles in the Intermontane United States*. (Ph.D. dissertation), University of California, Davis, CA .
- Ashbaugh, L., Malm, W., & Sadeh, W. (1985). A residence time probability analysis of sulphur concentrations at Grand Canyon National Park. *Atmospheric Environment*, *19*, 1263-1270.
- Bacharach Inc. (2005). HI FLOW sampler natural gas leak rate measurement instruction 55-9017 operation & maintenance. *Bacharach*. Retrieved March 22, 2009, from [www.bacharach-inc.com/PDF/Instructions/55-9017.pdf](http://www.bacharach-inc.com/PDF/Instructions/55-9017.pdf)
- Bertsimas, D., & Tsitsiklis, J. (1993). Simulated annealing. *Statistical Science*, *8*(1), 10-15.
- Billford, H. J., & Meaker, G. O. (1967). A factor analysis model of large scale pollution. *Atmospheric Environment*, *1*, 147-157.

Briggs, G. A. (1972). Discussion on chimney plumes in neutral and stable surroundings.

*Atmospheric Environment*, 6, 507-510.

Butcher, J. C. (1964). On Runge-Kutta processes of higher order. *Journal of the*

*Australian Mathematical Society*, 4, 179-194.

Canadian Association of Petroleum Producers (CAPP) (1999). *CH<sub>4</sub> and VOC emissions*

*from the Canadian upstream oil and gas industry, Vols. 1 and 2*. Calgary, AB:

Clearstone Engineering Ltd.

Canadian Association of Petroleum Producers (CAPP) (2004). *A national inventory of*

*greenhouse gas (GHG), criteria air contaminant (CAC) and hydrogen sulphide*

*(H<sub>2</sub>S) emissions by the upstream oil and gas industry - volume 1, overview of the*

*GHG emissions inventory*. Calgary, AB: CAPP.

Canadian Association of Petroleum Producers (CAPP) (2006). *An inventory of GHGs,*

*CACs, and H<sub>2</sub>S emissions by the Canadian bitumen industry: 1990 to 2003*.

Calgary, AB: Clearstone Engineering Ltd.

Canadian Association of Petroleum Producers (CAPP) (2007). *Best management*

*practice: management of fugitive emissions at upstream oil and gas facilities*.

Calgary, AB: CAPP.

Canadian Petroleum Products Institute (CPPI) (2004). *Economic and environmental*

*impacts of removing sulphur from Canadian gasoline and distillate production*.

Calgary, AB: Levelton Consultants Ltd.

Carras, J. N., Franklin, P. M., Hu, Y., Singh, A. K., Tailakov, O., Picard, D., et al (2006).

Chapter 4: Fugitive emissions. *2006 IPCC guidelines for national greenhouse*

*gas inventories*. Hayama, Japan: IPCC.

- Cartwright, H. M., & Harris, S. P. (1993). Analysis of the distribution of airborne pollution using genetic algorithms. *Atmospheric Environment*, 27A(12), 1783-1791.
- Celik, I. B., & Li, J. (2005). Assessment of numerical uncertainty for the calculations of turbulent flow over a backward-facing step. *International Journal for Numerical Methods in Fluids*, 49, 1015-1031.
- Chambers, A.K. (2001). *Well test flare plume monitoring - literature review*. Edmonton, AB: Carbon and Energy Management, Alberta Research Council Inc. Retrieved July 15, 2010, <http://www.ptac.org/env/dl/envp0201afr.pdf>
- Chambers, A.K. (2003) *Well test flare plume monitoring phase II: DIAL testing in Alberta* (CEM 7454-2003). Edmonton, AB: Carbon and Energy Management, Alberta Research Council Inc. Retrieved July 15, 2010, <http://www.ptac.org/env/dl/envp0402fr.pdf>
- Chambers, A.K. (2004). *Optical measurement technology for fugitive emissions from upstream oil and gas facilities* (ARC Contract Report No. CEM – P004.03), Calgary, AB: Canadian Association of Petroleum Producers.
- Chambers, A.K. (2006) *Refinery demonstration of optical technologies for measurement of fugitive emissions and for leak detection* (ARC Contract Report No. CEM 9643). Edmonton, AB: Carbon and Energy Management, Alberta Research Council Inc. Retrieved March 22, 2009, from <http://www.ptac.org/env/dl/envp0403.pdf>
- Charron, A., Plaisance, H., Sauvage, S., Coddevile, P., Gallo, J. C., & Guillermo, R. (2000). A study of the source-receptor relationships influencing the acidity of

- precipitation collected at a rural site in France. *Atmospheric Environment*, 34, 3665-3674.
- Cheng, M., & Lin, C. (2001). Receptor modeling for smoke of 1998 biomass burning in Central America. *Journal of Geophysical Research*, 106, 22871-22886.
- Cheng, M., Hopke, P. K., & Zeng, Y. (1993). A receptor-oriented methodology for determining source regions of particulate sulfate observed at Dorset, Ontario. *Journal of Geophysical Research*, 98(D9), 16839-16849.
- Colin, J. L., Renard, D., Lescoat, V., Jaffrezo, J. L., Gros, J. M., & Strauss, B. (1989). Relationship between rain and snow acidity and air mass trajectory in eastern France. *Atmospheric Environment*, 23(7), 1487-1498.
- de Haan, P. (1999). On the use of density kernels for concentration estimations within particle and puff dispersion models. *Atmospheric Environment*, 33, 2007-2021.
- de Haan, P., Scire, J. S., Strimaitis, D. G., & Rotach, M. W. (1999). Introduction of a puff-particle approach for near-source dispersion into the CALPUFF model. *Air pollution modeling and its application XIII (v. 13)* (pp. 147-154). New York: Plenum Press.
- Demidov, A. A. (2002). Laser remote sensing. In D. L. Andrews & A. A. Demidov (Eds.), *An Introduction to Laser Spectroscopy* (2nd ed.): (pp. 338-368). New York: Kluwer Academic Press.
- Derber, J. C. (1989). A variational continuous technique. *Monthly Weather Review*, 117, 2437-2446.
- Easom, G. (2000). *Improved turbulence models for computational wind engineering*. (PhD. thesis). University of Nottingham, Nottingham, UK.

- Eca, L. & Hoekstra, M. (2004). A verification exercise for two 2-D steady incompressible turbulent flows. In: Neittaanmaki, P., Rossi, T., Majava, K. & Pironneau, O. (Eds.), *Proceedings of the ECCOMAS 2004, Jyväskylä, Finland, 24-28 July 2004*. Jyväskylä, Finland: ECCOMAS.
- Elbern, H., Schmidt, H., & Ebel, A. (1997). Variational data assimilation for tropospheric chemistry modeling. *Journal of Geophysical Research*, 102(D13), 15967-15985.
- Environment Canada. (2010). *National inventory report 1990–2008: Greenhouse gas sources and sinks in Canada*. Gatineau, QC: Environment Canada. Retrieved July 10, 2010, from <http://www.ec.gc.ca/ges-ghg>
- Environmental Protection Agency. (2007). *Measurement of total site mercury emissions for a chlor-alkali plant using open-path UV-DOAS* (EPA Publication No. 600-R-07-077). Triangle Park, NC: U. S. Environmental Protection Agency. Retrieved March 22, 2009, from <http://www.epa.gov/nrmrl/pubs/600r07077/600r07077.pdf>
- Environmental Protection Agency (1995). *Protocol for equipment leak emission estimates* (EPA Publication No. 453-R-95-017). Triangle Park, NC: U. S. Environmental Protection Agency.
- Envirotech Engineering (2007). *Review and update of methods used for air emissions leak detection and quantification*. Calgary, AB: Petroleum Technology Alliance Canada. Retrieved March 22, 2009, from <http://www.ptac.org/eet/dl/eetp0701.pdf>
- Errico, R. M. (1997). What is an adjoint model? *Bulletin of the American Meteorological Society*, 78(11), 2577-2591.
- Ferziger, J. (1993). Simulation of complex turbulent flows: recent advances and prospects

in wind engineering. *Journal of Wind Engineering and Industrial Aerodynamics*, 46-47, 195-212.

Forster, P., Ramaswamy, V., Artaxo, P., Bernsten, T., Betts, R., Fahey, D.W., et al.

(2007). Changes in atmospheric constituents and in radiative forcing. In S.

Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt et al. (Eds.),

*Climate change 2007:the physical science basis. Contribution of working group I*

*to the fourth assessment report of the Intergovernmental Panel on Climate*

*Change*. New York: Cambridge University Press.

Franke, J., & Frank, W. (2008). Application of generalized Richardson extrapolation to

the computation of the flow across an asymmetric street intersection. *Journal of*

*Wind Engineering and Industrial Aerodynamics*, 96, 1616-1628.

Franke, J., Hellsten, A., Schlunzen, H. & Carissimo, B. (Eds.). (2007). *Best practice*

*guideline for the CFD simulation of flows in the urban environment – COST*

*action 732*, Brussels, Belgium: COST Office.

Franke, J., Hirsch, C., Jensen, A.G., Krus, H.W., Schatzmann, M., Westbury, P.S., et al.

(2004). Recommendations on the use of CFD in wind engineering. In van Beeck,

J.P.A.J. (Ed.), *COST Action C14, Impact of wind and storm on city life built*

*environment. Proceedings of the International Conference on Urban Wind*

*Engineering and Building Aerodynamics, 5–7 May 2004*. Sint-Genesius-Rode,

Belgium: von Karman Institute.

Gassonic.(n.d.). Ultrasonic Leak Detection Technology. *Gassonic: Ultrasonic gas leak*

*detector*. Retrieved March 22, 2009, from

[http://www.gassonic.com/technology/ultrasonic\\_gas\\_leak\\_detection/](http://www.gassonic.com/technology/ultrasonic_gas_leak_detection/)

- Gelman, A. (2002). Prior distribution. In V Barnett & K. F. Turkman (Eds.).  
*Encyclopaedia of Envirometrics, Volume 3* (pp. 1634-1637). Chichester: John Wiley & Sons, Ltd.
- Giering, R., & Kaminski, T. (1998). Recipes for adjoint code construction. *ACM Transactions On Mathematical Software*, 24(4), 437-474.
- Gokaltun, S., Skudarnov, P. V. & Lin, C.-X. (2005). *Verification and validation of CFD simulation of pulsating laminar flow in a straight pipe*. (AIAA 2005-4863).  
Symposium conducted at the 17th AIAA Computational Fluid Dynamics Conference. Toronto, Canada.
- Gravetter, F. J., & Wallnau, L. B. (1994). *Essentials of statistics for the behavioral sciences* (2nd ed.). United States of America: West Publishing Company.
- Haas-Laursen, D. E., Hartley, D. E., & Prinn, R. G. (1996). Optimizing an inverse method to deduce time-varying emissions of trace gases. *Journal of Geophysical Research*, 101(D17), 22823-22831.
- Han, Y., Holsen, T., & Hopke, P. (2007). Estimation of source locations of total gaseous mercury measured in New York State using trajectory-based models. *Atmospheric Environment*, 41(28), 6033-6047.
- Hanjalic, K., & Kenjeres, S. (2008). Some developments in turbulence modeling for wind and environmental engineering. *Journal of Wind Engineering and Industrial Aerodynamics*, 96, 1537-1570.
- Hartley, D., & Prinn, R. (1993). Feasibility of determining surface emissions of trace gases using an inverse method in a three-dimensional chemical transport model. *Journal of Geophysical Research*, 98(D3), 5183-5197.

- Hasama, T., Kato, S., & Ooka, R. (2008). Analysis of wind-induced inflow and outflow through a single opening using LES and DES. *Journal of Wind Engineering and Industrial Aerodynamics*, 96, 1678-1691.
- Haupt, S. E., Young, G. S., & Allen, C. T. (2007). A genetic algorithm method to assimilate sensor data for a toxic contaminant release. *Journal of Computers*, 2(6), 85-93.
- Hefny, M. M., & Ooka, R. (2009). CFD analysis of pollutant dispersion around buildings: effect of cell geometry. *Building and Environment*, 44, 1699-1706.
- Heinsohn, R. J., & Kabel, R. L. (1999). Chapter 9: Atmospheric dispersion. *Sources and control of air pollution* (pp. 386-337). New York: Prentice-Hall.
- Henderson, R. G., & Weingartner, K. (1982). Trajectory analysis of MAP3S precipitation chemistry data at Ithaca, NY. *Atmospheric Environment*, 16(7), 1657-1665.
- Hinnrichs, M. (2000). Infrared hyperspectral imaging sensor for gas detection. In M. R. Descour & S. S. Shen, (Eds.), *Imaging spectrometry VI: 31 July-2 August, 2000, San Diego, USA (Proceedings of Spie--the International Society for Optical Engineering, V. 4132.)* (illustrated edition ed., pp. 344-355). Bellingham: Spie-International Society For Optical Engineering.
- Holland, J. H. (1975). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. Ann Arbor, MI: The MIT Press.
- Holmes, J. (2001). *Wind loading of structures* (1 ed.). New York: Taylor & Francis.
- Hopke, P., Gladney, E., Gordon, G., Zoller, W., & Jones, A. (1976). The use of multivariate analysis to identify sources of selected elements in the Boston urban

- aerosol. *Atmospheric Environment*, 10, 1015-1025.
- Hsu, Y., Holsen, T. M., & Hopke, P. K. (2003). Comparison of hybrid receptor models to locate PCB sources in Chicago. *Atmospheric Environment*, 37, 545-562.
- Innovative Environmental Solutions, Inc. (2005). *Greenhouse gas emission estimation guidelines for natural gas transmission and storage Volume 1 - GHG emission estimation methodologies and procedures*. Washington, DC: Interstate Natural Gas Association of America. Retrieved March 22, 2009, from <http://www.ingaa.org/File.aspx?id=5485>
- Issartel, J., & Baverel, J. (2003). Inverse transport for the verification of the Comprehensive Nuclear Test Ban Treaty. *Atmospheric Chemistry and Physics*, 3, 475-486.
- Jackson, P. S., & Hunt, J. C. (1975). Turbulent wind flow over a low hill. *Quarterly Journal of the Royal Meteorological Society*, 101, 929-955.
- Jeong, U. Y., Koh, H., & Lee, H. S. (2002). Finite element formation for the analysis of turbulent wind flow passing bluff structures using the RNG k-epsilon model. *Journal of Wind Engineering and Industrial Aerodynamics*, 90, 151-169.
- Jones, M. E. (1968). A review of the literature on the structure of wind turbulence, with special regard to its effect on buildings. *Building Science*, 3, 41-60.
- Jones, W. P., & Launder, B. E. (1972). Prediction of relaminarization with a two equation model of turbulence. *International Journal of Heat and Mass Transfer*, 15, 301-314.
- Jung, Y., Park, W., & Park, O. (2003). Pollution dispersion analysis using the puff model with numerical flow field data. *Mechanics Research Communications*, 30, 277-

286.

- Kaminski, T., Heimann, M., & Giering, R. (1999). A coarse grid three-dimensional global inverse model of the atmospheric transport. 1. Adjoint model and Jacobian matrix. *Journal of Geophysical Research*, *104*(D15), 18535-18553.
- Keeler, G. J., & Samson, J. (1989). Spatial representativeness of trace element ratios. *Environmental Science & Technology*, *23*, 1358-1364.
- Kim, C., Song, C., Lee, S., & Song, S. (2008). Simulating mesoscale transport and diffusion of radioactive noble gases using the Lagrangian dispersion model. *Journal of Environmental Radioactivity*, *99*, 1644-1952.
- Kim, H. G., & Patel, V. C. (2000). Test of Turbulence models for wind flow over terrain with separation and recirculation. *Boundary-Layer Meteorology*, *94*, 5-21.
- Kulp, T., Powers, P., Kennedy, R., & Goers, U. (1998). Development of a pulsed backscatter-absorption gas-imaging system and its application to the visualization of natural gas leaks. *Applied Optics*, *37*(18), 3912-3922.
- Lavrentev, M. M., Romanov, V. G., & Shishatskii, S. P. (1986). *Ill-posed problems of mathematical physics and analysis [translated by J.R. Schulenberg. Translations of mathematical monographs, vol. 64]*. Providence, RI: American Mathematical Society.
- Lee, S., & Ashbaugh, L. (2007). Comparison of multi-receptor and single-receptor trajectory source apportionment (TSA) methods using artificial sources. *Atmospheric Environment*, *41*, 1119-1127.
- Lin, C., Cheng, M., & Schroedar, W. H. (2001). Transport patterns and potential sources of total gaseous mercury measured in Canadian high Arctic in 1995. *Atmospheric*

- Environment*, 35, 1141-1154.
- Lupu, A., & Maenhaut, W. (2002). Application and comparison of two statistical trajectory techniques for identification of source regions of atmospheric aerosol species. *Atmospheric Environment*, 36, 5607-.
- Malalasekera, W., & Versteeg, H. (1995). *An introduction to computational fluid dynamics: The finite volume method*. (2nd ed.). Alexandria, VA: Prentice Hall.
- Mankbadi, R. R. (1994). *Transition, turbulence, and noise: theory and applications for scientists and engineers (The Springer international series in engineering and computer science)* (1 ed.). New York: Springer.
- Mazzitelli, I. M., & Lohse, D. (2004). Lagrangian statistics for fluid particles and bubbles in turbulence. *New Journal of Physics*, 6, 203.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21, 1087-1092.
- Miller, J. M. (1981). A five-year climatology of back trajectories from the Mauna Loa Observatory, Hawaii. *Atmospheric Environment*, 15(9), 1553-1558.
- Miller, J. M., Moody, J. L., Harris, J. M., & Gaudry, A. (1993). A 10-year trajectory flow climatology for Amsterdam Island. *Atmospheric Environment*, 27A(12), 1909-1916.
- Miller, M. S., Friedlander, S. K., & Hidy, G. M. (1972). A chemical element balance for the Pasadena aerosol. *Journal of Colloid and Interface Science*, 39, 165-176.
- Mochida, A., & Lun, I. Y. (2008). Prediction of wind environment and thermal comfort at pedestrian level in urban area. *Journal of Wind Engineering and Industrial*

*Aerodynamics*, 96, 1498-1527.

Moody, J. L., & Samson, P. J. (1989). The influence of atmospheric transport on precipitation chemistry at two sites in the Midwestern United States.

*Atmospheric Environment*, 23(10), 2117-2132.

Murakami, S., Mochida, A., & Hayashi, Y. (1990). Examining the k-epsilon model by means of a wind tunnel test and large-eddy simulation of the turbulence structure around a cube. *Journal of Wind Engineering and Industrial Aerodynamics*, 35, 87-100.

Murakami, S., Mochida, A., & Tominaga, Y. (1994). Numerical simulation of turbulent diffusion in cities. In J. E. Cermak, A. G. Davenport, E. J. Plate & D. X. Viegas (Eds.). *Wind climate in cities* (pp. 681-701). Dordrecht: Kluwer Academic Publishers.

Narasimhamurthy, V. (2004). *Unsteady-RANS simulation of turbulent trailing-edge flow*. (Masters Thesis), Chalmers University of Technology, Goteborg, Sweden.

NDE Systems Inc. (n.d.). VPAC 5131 digital AE leak detector for valves. *Russell NDE Systems Inc.* Retrieved March 22, 2009, from [www.russelltech.com/UT/VPAC%205131.html](http://www.russelltech.com/UT/VPAC%205131.html)

Owen, S. J. (1998). *A survey of unstructured mesh generation technology*. Symposium conducted at the 7th International Meshing Roundtable, Dearborn, Michigan.

Penenko, V. V., & Obraztsov, N. N. (1976). A variational initialization method for the fields of meteorological elements (English translation). *Soviet Meteorological Hydrology*, 1, 1-11.

Peng, X. L., Choi, M. P., & Wong, M. H. (2007). Receptor modeling for analyzing

- PCDD/F and dioxin-like PCB sources in Hong Kong. *Environmental Modeling & Assessment*, 12, 229-237.
- Peters, L. K., Berkowitz, C. W., Saylor, R. D., Tsang, T. T., Carmichaels, G. R., Easter, R. C., et al. (1995). The current state and future direction of Eulerian models in simulating the tropospheric chemistry and transport of trace species: a review. *Atmospheric Environment*, 29(2), 189-222.
- Picard, D., Stribrny, M. & Harrison, M. R. (1998). *Handbook for estimating methane emissions from Canadian natural gas systems*. Don Mills, ON: Clearstone Engineering Ltd., Enerco Engineering Ltd. and Radian International.
- Picart, A., Berlemont, A., & Gouesbet, G. (1986). Modelling and predicting turbulence fields and the dispersion of discrete particles transported by turbulent flows. *International Journal of Multiphase Flow*, 12(2), 237-261.
- Pitchford, A., Pitchford, M., Malm, W., Flocchini, R., & Cahill, T. (1981). Regional analysis of factors affecting visual air quality. *Atmospheric Environment*, 15(10-11), 2043-2054.
- Poirot, R. L., Wishinski, P. R., Hopke, P. K., & Polissar, A. V. (2001). Comparative application of multiple receptor methods to identify aerosol sources in northern Vermont. *Environmental Science & Technology*, 35, 4622-4636.
- Polissar, A. V., Hopke, P. K., & Harris, J. M. (2001). Source regions for atmospheric aerosol measured at Barrow, Alaska. *Environmental Science & Technology*, 35, 4214-4226.
- Rao, K. S. (2007). Source estimation methods for atmospheric dispersion. *Atmospheric Environment*, 41, 6964-6973.

- Reible, D. D. (1999). *Fundamentals of environmental engineering - 1999 publication*. Boca Raton, FL: Lewis Publishers.
- Reynolds, O. (1895). On the dynamical theory of incompressible viscous fluids and the determination of the criterion. *Philosophical Transactions of the Royal Society of London, 186*, 123-164.
- Roache, P. J. (1994). Perspective: a method for uniform reporting of grid refinement studies. *Journal of Fluids Engineering, 116*, 405-413.
- Robertson, L., & Persson, C. (1993). Attempts to apply four dimensional data assimilation of radiological data using the adjoint technique. *Radiation Protection Dosimetry, 50(2)*, 333-337.
- Robertson, L., & Langner, J. (1998). Source function estimate by means of variational data assimilation applied to the ETEX-I tracer experiment. *Atmospheric Environment, 32(24)*, 4219-4225.
- Rodi, W. (1997). Comparison of LES and RANS calculations of the flow around bluff bodies. *Journal of Wind Engineering and Industrial Aerodynamics, 69-71*, 55-75.
- Ross, B. D., & Picard, D. J. (1996). *Measurement of methane emissions from western Canadian natural gas facilities*. Don Mills, ON: Clearstone Engineering Ltd.
- Roy, C. J. (2005). Review of code and solution verification procedures for computational simulation. *Journal of Computational Physics, 205*, 131-156.
- Rummel, R. J. (1976). The correlation coefficient. *Understanding correlation*. Retrieved June 23, 2010, from [www.hawaii.edu/powerkills/UC.HTM](http://www.hawaii.edu/powerkills/UC.HTM)
- Runca, E., Melli, P., & Spirito, A. (1979). Real-time forecast of air pollution episodes in the Venetian region. Part 1: the advection-diffusion model. *Applied*

- Mathematical Modelling*, 3, 402-408.
- Sasaki, Y. (1968). An objective analysis based on the variational method. *Journal of the Meteorological Society of Japan*, 36, 1-12.
- Scheifinger, H., & Kaiser, A. (2007). Validation of trajectory statistical methods. *Atmospheric Environment*, 41, 8846-8856.
- Schnelle, K. B., & Dey, P. R. (2000). *Atmospheric dispersion modeling compliance guide*. New York: McGraw-Hill Professional.
- Seibert, P., Kromp-Kolb, H., Baltensperger, U., Jost, D. T., Schwikowski, M., Kasper, A., et al. (1994). Trajectory analysis of aerosol measurements at high alpine sites. *Transport and Transformation of Pollutants in the Toposphere* (pp. 689-693). Den Haag: Academic Publishing.
- Sohn, M. D., Reynolds, P., Singh, N., & Gadgil, A. J. (2002). Rapidly locating and characterizing pollutant releases in buildings. *Journal of the Air and Waste Management Association*, 52, 1422-1432.
- Spall, J. C. (2003). *Introduction to stochastic search and optimization: Estimation, simulation, and control*. Hoboken, NJ: John Wiley and Sons.
- Speziale, C. G. (1998). Turbulence modeling for time-dependent RANS and VLES: a review. *AIAA Journal*, 36(2), 173-184.
- Stohl, A. (1996). Trajectory statistics - a new method to establish source-receptor relationships of air pollutants and its application to the transport of particulate sulphate in Europe. *Atmospheric Environment*, 30, 579-587.
- Tarantola, A. (1987). *Inverse problem theory*. Amsterdam: Elsevier Science & Technology.

- Thomson, L. C., Hirst, B., Gibson, G., Gillespie, S., Jonathan, P., Skeldon, K. D., et al. (2007). An improved algorithms for locating a gas source using inverse methods. *Atmospheric Environment*, *41*, 1128-1134.
- Tominaga, Y., Mochida, A., Yoshie, R., Kataoka, H., Nozu, T., Yoshikawa, M., et al. (2008). AIJ guidelines for practical applications of CFD to pedestrian wind environment around buildings. *Journal of Wind Engineering and Industrial Aerodynamics*, *96*, 1749-1761.
- Turner, D. B. (1994). *Workbook of Atmospheric Dispersion Estimates: Second Edition* (2nd ed.). USA: CRC-Press.
- Turns, S. R. (2000). Appendix D. *An introduction to combustion: Concepts and applications w/software* (2 ed., pp. 656-658). New York: McGraw-Hill Science/Engineering/Math.
- Vasconcelos, L. A., Kahl, J. D., Liu, D., Macias, E. S., & White, W. H. (1996). Spatial resolution of a transport inversion technique. *Journal of Geophysical Research*, *101*, 19337-19342.
- Vasconcelos, L. A., Kahl, J. D., & Liu, D. (1996). A tracer calibration of back trajectory analysis at the Grand Canyon. *Journal of Geophysical Research*, *101*(D14), 19329-19335.
- Wark, K., & Warner, C. P. (1981). *Air pollution: its origin and control* (2nd ed.). New York: Harper & Row.
- Welch, G., & Bishop G. (2006). An introduction to the Kalman filter. Chapel Hill, NC: University of North Carolina. Retrieved Sept. 9, 2010, [http://www.cs.unc.edu/~welch/media/pdf/kalman\\_intro.pdf](http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf)

- Wieringa, J. (1996). Does representative wind information exist? *Journal of Wind Engineering and Industrial Aerodynamics*, 65, 1-12.
- Wilcox, D. C. (1994). *Turbulence modelling for CFD*. La Canada, California: DCW Industries Inc.
- Wotawa, G., & Kroger, H. (1999). Testing the ability of trajectory statistics to reproduce emission inventories of air pollutants in cases of negligible measurement and transport errors. *Atmospheric Environment*, 33, 3037-3043.
- Wright, N. G., & Easom, G. J. (2003). Non-linear k-epsilon turbulence model results for flow over a building at full-scale. *Applied Mathematical Modeling*, 27, 1013-1033.
- Yakhot, V., & Orszag, S. A. (1986). Renormalized group analysis of turbulence. I. basic theory. *Journal of Scientific Computing*, 1, 3-51.
- Zeng, Y., & Hopke, P. K. (1989). A study of the sources of acid precipitation in Ontario, Canada. *Atmospheric Environment*, 23(7), 1499-1509.
- Zhang, L. (2007). *Large-scale simulations using first and second order adjoints with applications in data assimilation*. (Master's thesis). Virginia Polytechnic Institute and State University, Blacksburg, VA.
- Zhao, W., Hopke, P. K., & Zhou, L. (2007). Spatial distribution of source locations for particulate nitrate and sulfate in the upper-midwestern United States. *Atmospheric Environment*, 41, 1931-1847.
- Zhou, L., Hopke, P., & Liu, W. (2004). Comparison of two trajectory based models for locating particle sources for two rural New York sites. *Atmospheric Environment*, 38, 1955-1963.

## Appendix A: Richardson Extrapolation and Convergence

Verification and validation in computational wind engineering are critical for creating confidence in CFD solutions. Validation is used to assess model errors by comparison with experimental data. This is not possible when, as in the current case, the modelled situation is purely hypothetical. Verification, on the other hand, deals with numerical errors in the CFD results. Richardson extrapolation is one method for determining the discretization error (Franke & Frank, 2008). This is especially important in wind engineering as grid converged solutions are usually beyond current computational capabilities. This discretization error arises from modelling with a discrete system with a finite grid and time step (Gokalktun el al., 2005). While reaching solutions in the asymptotic range is challenging, or impossible, in many cases, it is recommended to attempt Richardson extrapolation for wind engineering problems to gain a sense of the error due to the problem discretization (Franke et al., 2007).

### Generalized Richardson Extrapolation

Richardson extrapolation is an a posteriori error estimator that can be used independent of the numerical scheme (Franke et al., 2007). First, consider spatial discretization with the exact solution of  $f_{ex}$  and a numerical solution of  $f_k$  on mesh  $k$ . The numerical and exact solutions can be related as (Franke et al., 2007):

$$f_k = f_{ex} + g_p h_k^p + g_{p+1} h_k^{p+1} + g_{p+1} h_k^{p+1} + \dots \quad (\text{A.1})$$

where  $h$  is a measure of the grid width,  $p$  is the order of accuracy, and  $g$  are coefficients. If the solution is within the asymptotic range, higher order terms than  $p$

can be neglected and  $g$  and  $p$  do not depend on  $h$ . This leaves three unknowns since  $h$  for the grid and  $f_k$  are already known. Thus, three equations (and three different meshes) are required to estimate the exact solution. If  $k=1,2,3$  corresponds to a fine, medium and coarse grid respectively, the grid refinement ratios between them are  $r_{21} = h_2 / h_1$  and  $r_{32} = h_3 / h_2$  (Franke et al., 2007).

In the case of unstructured meshes, the refinement factor can be calculated as

$$r_{21} = (N_1 / N_2)^{1/D} \quad (\text{A.2})$$

where  $D$  is the dimension in space and  $N$  the number of nodes or cells within the mesh. (Franke et al., 2007)

The series expansions are thus (Franke et al., 2007):

$$f_1 = f_{ex} + g_p h_1^p \quad (\text{A.3})$$

$$f_2 = f_{ex} + g_p h_2^p = f_{ex} + g_p (r_{21} h_1)^p \quad (\text{A.4})$$

$$f_3 = f_{ex} + g_p h_3^p = f_{ex} + g_p (r_{21} r_{32} h_1)^p \quad (\text{A.5})$$

In order to use Richardson extrapolation, the meshes must display monotonic convergence. The ratio of change is defined as  $R = (f_2 - f_1) / (f_3 - f_2)$ , the three possible behaviours are (Franke et al., 2007):

- $0 < R < 1$  monotonic convergence
- $R < 0$  oscillatory convergence
- $|R| > 1$  divergence

It is not possible to obtain error estimates from diverging grids. If the difference between the medium and fine grid approaches zero, the maxima and minima of the local solution should be analyzed (Franke et al., 2007) .

The order of accuracy can be calculated from (Franke et al., 2007)

$$p = \frac{\ln \left[ \frac{(f_2 - f_1)/(f_3 - f_2)}{\ln r_{21}} \right]}{\ln r_{21}} - \frac{1}{\ln r_{21}} \left[ \ln(r_{32}^p - 1) - \ln(r_{21}^p - 1) \right] \quad (\text{A.6})$$

using an iterative method.

An estimate of the exact value is then given by (Franke et al., 2007)

$$f_{ex} = f_1 + \frac{f_1 - f_2}{r_{21}^p - 1} \quad (\text{A.7})$$

and, thus, the spatial discretization errors for each grid are

$$DE_1 = \frac{f_1 - f_2}{r_{21}^p - 1} \quad (\text{A.8})$$

$$DE_2 = r_{21}^p DE_1 \quad (\text{A.9})$$

$$DE_3 = (r_{21} r_{32})^p DE_1 \quad (\text{A.10})$$

Rearranged, these equations can be used to check if the solutions are within the asymptotic range if the following identity holds (Franke et al., 2007):

$$DE_1 = \frac{DE_2}{r_{21}^p} = \frac{DE_3}{(r_{21} r_{32})^p} \quad (\text{A.11})$$

## Grid Convergence

Though the above approaches provide an estimate of a corrected value ( $f_{ex}$ ) at the points selected for comparison (generally the nodes on the coarsest grid), this corrected value cannot be used as the actual solution since it no longer fulfills the general equations (Franke et al., 2007). Thus, the corrected value is typically used to find the relative error or an error band (Franke et al., 2007).

The relative error of the fine grid is (Franke et al., 2007)

$$E_1' = \frac{f_1 - f_{ex}}{f_{ex}} \quad (\text{A.12})$$

This can be approximated by (Franke et al., 2007)

$$E_1 = \frac{1}{r_{21}^p - 1} \frac{f_2 - f_1}{f_1} = E_1' + O(h^{p+1}) \quad (\text{A.13})$$

Some more practical estimators are (Franke et al., 2007)

$$E_{1,f} = \frac{1}{r_{21}^p - 1} \frac{|f_2 - f_1|}{\text{range}(f_1)} \quad (\text{A.14})$$

$$E_{1,\max} = \frac{1}{r_{21}^p - 1} \frac{\max |f_2 - f_1|}{\text{range}(f_1)} \quad (\text{A.15})$$

$$E_{1,rms} = \frac{1}{r_{21}^p - 1} \frac{\text{rms}(f_2 - f_1)}{\text{range}(f_1)} \quad (\text{A.16})$$

The normalization is performed over the difference between the maximum and minimum values on the fine grid to prevent problems in areas where the value approaches zero (Franke et al., 2007).

The magnitude of the relative error defines an air band around the solution, but only provides a confidence level of 50%. Thus, the error is multiplied by a factor of safety in order to increase the confidence level. The exact value of the safety factor to reach 95% confidence is still being debated (Roy, 2005). Roache (1994) suggested a safety factor of 1.25 when using three meshes and the observed and formal order match and a factor of 3 for two meshes or cases where the formal order and observed order are different (Franke et al., 2007).

In the case of urban flows, Franke and Frank (2008), in their study of the use of Richardson extrapolation to produce error bars for flow through street canyons used a variation on Eca and Hoestra's (2004) extension of Roache's grid convergence index. For an expected order of accuracy between one and two, with monotonic convergence, the error band was calculated based on the observed order of accuracy  $p$ .

This is done using  $\delta_i = DE_i$  as described in the section on Richardson extrapolation and solving the grid triplet for  $\delta'_i = f_i - f_{ex} \approx g_1 h_i + g_2 h_i^2$  (Eca & Hoekstra, 2004).

- For  $0.5 < p \leq 2$

$$U_1 = 1.25 |\delta_1| \quad (\text{A.17})$$

- For  $2 < p \leq 3$

$$U_1 = 1.25 \max\left(|\delta_1|, |\delta'_1|\right) \quad (\text{A.18})$$

- For  $p \leq 0.5$  or  $p > 3$

$$U_1 = 3 \max\left(|f_3 - f_2|, |f_2 - f_1|\right) \quad (\text{A.19})$$

where the last equation is also used for oscillatory convergence and no estimation is made in cases of divergence or when the observed order of accuracy is outside the range of 0 to 10. Franke and Frank (2008) examined the velocity components using this method. The results were normalized by the maximum component value to obtain a percent error.

### **Temporal Convergence**

Roache (1994) suggested separate convergence studies for spatial and temporal convergence. Thus, a similar approach can be used to determine the effect of the timestep size on the results. Gokalktun et al. (2005) used this method for an unsteady, pulsating flow.

## Appendix B: Correlation Measures

### B.1 Spatial Correlation Index

A spatial correlation index,  $r$ , can be used as a measure of the association between two maps (Haining, 1990). A general autocorrelation ( $\Gamma$ ) statistic between two spatial matrices is represented as

$$\Gamma = \sum_i \sum_j G_{ij} C_{ij} \quad (\text{B.1})$$

where  $G_{ij}$  is a measure of the spatial proximity of locations  $i$  and  $j$  and  $C_{ij}$  of the variable value proximity at the same locations. One possible set of measures for spatial and variable proximity (in 2D, though this is easy extendable) is (Hopke et al., 2005):

$$G_{ij} = \frac{1}{(x_i - x_j)^2 + (y_i - y_j)^2 + 1} \quad (\text{B.2})$$

$$C_{ij} = \frac{(g_i - \bar{g})(f_j - \bar{f}) + (g_j - \bar{g})(f_i - \bar{f})}{2} \quad (\text{B.3})$$

where  $g$  and  $f$  are matrices representing the two maps,  $x_i$  and  $y_i$  are the coordinate locations of location  $i$ ,  $g_i$  and  $f_i$  are the values of the two maps at location  $i$  and  $\bar{g}$  and  $\bar{f}$  are the average map values. Similar high or low variable values have the most influence on  $C$  (Hopke et al., 2005). To determine how meaningful the correlation is, the null hypothesis ( $H_0$ ) that there is no correlation between the two maps must be considered. Assuming a total number of observations  $n$ , using the randomization assumption, the first two moments of the autocorrelation statistic are (Haining, 1990; Hopke et al., 2005)

$$E_R(\Gamma) = \frac{S_0 T_0}{n(n-1)} \quad (\text{B.4})$$

$$E_R(\Gamma^2) = \frac{S_1 T_1}{2n^{(2)}} + \frac{(S_2 - 2S_1)(T_2 - 2T_1)}{4n^{(3)}} + \frac{(S_0^2 + S_1 + S_2)(T_0^2 + T_1 + T_2)}{n^{(4)}} \quad (\text{B.5})$$

where

$$S_0 = \sum_{ij} G_{ij}$$

$$S_1 = \frac{1}{2} \sum_{ij} (G_{ij} + G_{ji})^2$$

$$S_2 = \sum_i \left( \sum_j G_{ij} + \sum_j G_{ji} \right)^2$$

$$n^{(b)} = n(n-1)(n-2)\dots(n-b+1)$$

and  $T_k$  is the same as  $S_k$  with  $G_{ij}$  replaced by  $C_{ij}$ .

The correlation index  $r$  is (Haining, 1990; Hopke et al., 2005):

$$r = \frac{\Gamma - E(\Gamma)}{\left( E(\Gamma^2) - (E(\Gamma))^2 \right)^{\frac{1}{2}}} \quad (\text{B.6})$$

This can be treated as a normal distribution of mean 0 and standard deviation 1 (Haining, 1990). It is necessary to determine the significance of the correlation index. Using a one tailed test, the value which results in the rejection of the null hypothesis can be found for a certain significance level. Thus, for 95% confidence, correlations under 1.6 should be rejected.

## B.2 Spearman Rank Order Correlation Coefficients

Spearman rank correlation coefficient,  $r_s$ , is another possible measure of association which evaluates the “degree of linear association between two independent variables” (Gauthier, 2001). It is given as (Haining, 1990):

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^2 - n} \quad (\text{B.7})$$

where  $d_i$  is the difference in rank of the  $i^{\text{th}}$  observation for the two maps. This becomes more complicated in cases where there are ties in rank. In cases where there are multiple identical values, the mean of ranks that would have been used had the values differed slightly is assigned instead so that the sum of ranks is the same total sum as that of integers from 1 to  $n$  (Press et al., 1992). Let  $p_k$  be the number of ties in the  $k^{\text{th}}$  group of ties in the  $f_i$ 's and  $q_m$  be number of ties in the  $m^{\text{th}}$  group of ties among the  $g_i$ 's. Then,

$$r_s = \frac{1 - \frac{6}{n^3 - n} \left[ \sum_i d_i^2 + \frac{1}{12} \sum_k (p_k^3 - p_k) + \frac{1}{12} \sum_m (q_m^3 - q_m) \right]}{\left[ 1 - \frac{\sum_k (p_k^3 - p_k)}{n^3 - n} \right]^{\frac{1}{2}} \left[ 1 - \frac{\sum_m (q_m^3 - q_m)}{n^3 - n} \right]^{\frac{1}{2}}} \quad (\text{B.8})$$

If there are no ties, this reduces the previous equation.

## B.3 Pearson Correlation Coefficient

Pearson correlation is a measure of the linear relationship between a pair of variables. In a conceptual sense, this can be expressed as (Gravetter & Wallnau, 2008):

$$r = \frac{\text{degree X and Y vary together}}{\text{degree X and Y vary separately}} \quad (\text{B.9})$$

Thus, if both variables increase together, the relationship is positive. If one variable increases while the other decreases, the relationship is negative. If they do not vary together at all and there is no correlation, the value is zero. Specifically (Rummel, 1976),

$$r = \frac{\sum(x_i - \bar{x})(y_j - \bar{y})}{\sqrt{(\sum(x_i - \bar{x})^2)(\sum(y_j - \bar{y})^2)}} \quad (\text{B.10})$$

Thus, if value of both vectors is greater than the mean, the correlation is positive, if they are both less than the mean, the value is also positive and if they are moving in opposite directions from their mean, there is a negative contribution.