18th International Conference on
Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes
9-12 October 2017, Bologna, Italy

USING SENSOR DATA AND INVERSION TECHNIQUES TO SYSTEMATICALLY REDUCE DISPERSION MODEL ERROR

D. J. Carruthers[1], A. L. Stidworthy[1], D. Clarke[2], K.J. Dicks[3], R. L. Jones[4], I. Leslie[5], O. A. M. Popoola[4], A. Billingsley[6] and M. Seaton[1]

[1]Cambridge Environmental Research Consultants, Cambridge, UK
[2]Cambridgeshire County Council, Cambridge, UK
[3]Cambridge City Council, Cambridge, UK
[4]Department of Chemistry, University of Cambridge, UK
[5]Computer Laboratory, University of Cambridge, UK
[6]Environmental Instruments, Stratford-upon-Avon, UK

Abstract: An optimisation scheme has been developed that uses inversion techniques to modify pollution emission rates based on sensor data to improve dispersion model accuracy. The scheme minimises a cost function using a non-negative least squares solver. Error covariance is defined in relatively simple terms for both emissions and monitored concentrations. The scheme has been tested in an initial case study in Cambridge using monitored data from four reference monitors and twenty AQMesh sensor pods for the period, 30 June 2016 – 30 September 2016. Hourly NO_x concentrations from road sources modelled using ADMS-Urban and observed concentrations were processed using the optimisation scheme and the adjusted emissions were re-modelled. The optimisation scheme improved model accuracy and reduced average road emissions on average by 6.5% compared to the original estimates. Future work will focus on developing more complex representations of error covariance and on extending the scheme to multiple source types and pollutants.

Key words: *inversion, optimisation, emissions, ADMS-Urban*

INTRODUCTION

Compiling an accurate emissions inventory for an urban area is a challenging and time consuming task. Even where comprehensive and detailed emissions inventories exist, errors in rates of emissions account for a significant proportion of dispersion model error; for example there is high uncertainty in published NO_X emission factors for light-duty diesel vehicles (Anenberg *et al.*, 2017) or for PM from residential burning (Denier van der Gon *et al.*, 2015). Traditionally, dispersion models used in urban areas are validated by comparing measured and modelled concentrations at well-established monitoring sites (Stocker *et al.*, 2014); at best, modellers manually refine the dispersion modelling to minimise error at these locations; at worst, modellers calculate 'adjustment factors' and apply these to modelled concentrations. Meanwhile, the increasing availability of relatively low cost air pollution sensors that are easy to install and to maintain is allowing networks of such sensors to be installed across urban areas (Kumar *et al.*, 2015). Although these sensors have reduced reliability and accuracy compared with traditional monitors they allow much greater spatial coverage. This trend requires the dispersion modelling community to examine how data from these networks can be used most effectively to assess and improve dispersion models because the traditional model validation methods may not be appropriate. A systematic method that integrates data from these low cost sensors with models could deliver real benefits in terms of understanding and improving the quantification of emissions and improving model calculations of concentrations of pollutants. It also offers the opportunity to examine important questions such as: what spatial separation or number of sensors is sufficient to optimise emissions through inverse modelling; and what is the relative effectiveness of a small number of reference monitors and a larger number of sensors. This paper presents the implementation of an inversion technique (e.g. Webster *et al.*, 2016) in the street scale resolution urban dispersion model ADMS-Urban (Owen *et al.*, 2000). The methodology has been tested using data from four reference monitors and twenty AQMesh sensor pods (Carruthers *et al.*, 2016) in Cambridge.

METHODOLOGY

The inversion method (e.g. Webster *et al.*, 2016) requires minimisation of the cost function $J(\boldsymbol{x})$ defined in equation (1); the equation parameters together with their dimensions are defined in Table 1.

$$J(\boldsymbol{x}) = (\boldsymbol{Mx} - \boldsymbol{y})^T\boldsymbol{R}^{-1}(\boldsymbol{Mx} - \boldsymbol{y}) + (\boldsymbol{x} - \boldsymbol{e})^T\boldsymbol{B}^{-1}(\boldsymbol{x} - \boldsymbol{e}) \qquad (1)$$

The first term represents model error taking into account observation uncertainty; the second term represents emissions error taking into account emissions uncertainty. Given an initial set of emissions data, this cost function is minimised using a non-negative least squares solver to find a revised set of emissions data that reduces model error.

Table 1. Definition of cost function equation parameters

Quantity	Definition	Dimensions
$\boldsymbol{x}$	Vector of emissions (result)	n
$\boldsymbol{M}$	Transport matrix relating the source term to the observations	n by k
$\boldsymbol{y}$	Vector of observations	k
$\boldsymbol{R}$	Error covariance matrix for the observations	k by k
$\boldsymbol{e}$	Vector of first guess emissions	n
$\boldsymbol{B}$	Error covariance matrix for the first guess emissions	n by n

A key challenge of the implementation of the inversion technique is to quantify the covariance of emissions error between sources of the same type, between sources of different types and between pollutants; similarly, to quantify the covariance of observation error between monitoring sites and between pollutants. The diagonal values in the error covariance matrices represent the variance σ^2. To obtain the error covariance matrix values it is assumed that the standard deviation σ is equal to the uncertainty in the measurement or emission and that this is proportional to the measurement or emission. U_E is a fraction that represents the uncertainty in the emissions; U_{OR} and U_{OS} are fractions that represent the uncertainties in the observations at the reference monitors and sensors respectively. The error covariance matrix values off the diagonal relate to the proportion of the error that is due to systematic error between different sources and different sensors. For example, in the case of road sources, one source of systematic error would be error in the emissions factors (e.g. road traffic emissions factors, which are used by all sources). Unsystematic error might be, for example, an error in the traffic count on a particular road. It is assumed that a fraction U_{EF} of error is due to systematic errors. For the observations, we assume that there is zero error covariance between monitors of different types (e.g. between reference monitors and sensors) and we simply assume that a fraction U_{ORF} of reference monitor error U_{OR} is systematic and a fraction U_{OSF} of sensor error U_{OS} is systematic.

CASE STUDY

During 2016 twenty AQMesh sensor pods (Carruthers *et al.*, 2016) were deployed across Cambridge in addition to four reference monitors already in situ (see Figure 1). The sensors measure NO, NO_2, NO_x, O_3, CO, SO_2, PM1, $PM_{2.5}$, PM_{10} and TPC at intervals of fifteen minutes. For the optimisation scheme case study, only hourly averaged NO_x concentrations were considered and it was assumed that local emissions were dominated by road traffic emissions. The period analysed was 30 June 2016 – 30 September 2016. The aims of this initial case study were two-fold: firstly, to test that the optimisation scheme behaves as expected and to find and correct any errors in the scheme; secondly, to examine whether, even with the relatively

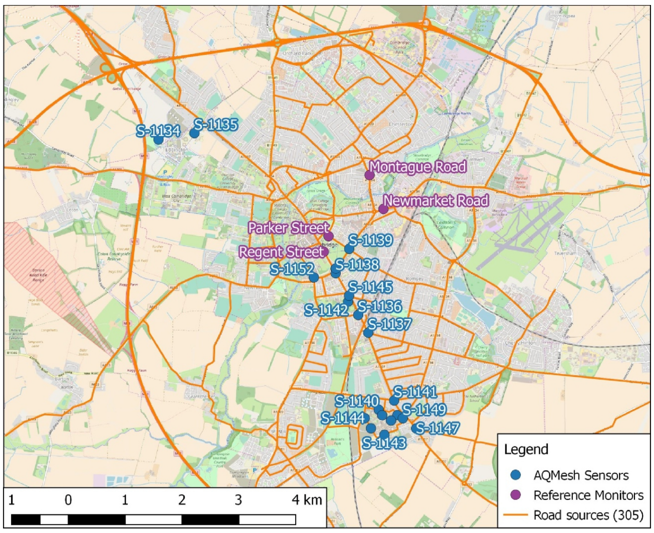

Figure 1. Map of Cambridge showing the locations of the AQMesh sensors, the reference monitors and the 305 road sources modelled.

simple representation of error co-variance described above, the adjusted emissions calculated by the optimisation scheme deliver any improvements in model predictions. The stages in the analysis were as follows: road emissions were modelled using ADMS-Urban for each hour of the three month period using a road traffic emission inventory for Cambridge and meteorological data from the Andrewsfield weather station 40 km to the south-east of Cambridge; the transport matrix, emissions vector and monitored data vectors were formed; the optimisation scheme was executed for each hour independently to determine the adjusted emissions for each hour; and finally the adjusted road emissions were re-modelled with ADMS-Urban keeping all other inputs unchanged. The optimisation was performed twice: firstly including both the reference monitor data and AQMesh sensor data in the optimisation scheme; secondly including only the AQMesh sensor data in the optimisation scheme. The uncertainty and covariance factor values used are shown in Table 2. These values represent plausible estimates but would need refinement in any further study. It was assumed that the error covariance factors for both the sensors and reference instruments were small, but are more significant for emissions since the latter depend on road traffic emission factors common to all sources.

Table 2. Variance and co-variance uncertainty factors

Parameter name	Description	Value
U_{OR}	Observation uncertainty (reference monitors)	0.1
U_{OS}	Observation uncertainty (sensors)	0.3
U_{ORF}	Observation uncertainty covariance factor (reference monitors)	0.05
U_{OSF}	Observation uncertainty covariance factor (sensors)	0.1
U_E	Emissions uncertainty	0.5
U_{EF}	Emissions uncertainty covariance factor	0.4

RESULTS

The analysis of the case study results focuses on two aspects: firstly on how the modelled concentrations using adjusted emissions compare with observed concentrations and the original modelled values in the two cases; and secondly how the adjusted emissions compare with the original emissions in the two cases.

Effect of the optimisation on modelled concentrations

The modelled concentrations using adjusted emissions when both the reference monitor data and the AQMesh sensor data are included in the optimisation demonstrates the expected behaviour. Figure 2 shows an example for one hour, comparing observed NO_x concentrations with the original modelled concentrations and with the modelled concentrations using adjusted emissions. Since the uncertainty for the AQMesh sensors is higher than that assumed for the reference monitors, the adjustments at the reference sites are larger. To examine the effect of the optimisation in the two cases on model performance, observed concentrations at the reference monitors were compared with modelled concentrations using: unadjusted emissions (‘Original’); emissions adjusted using reference monitor data and sensor data (‘Adjusted, all sensors’); and emissions adjusted using sensor data only (‘Adjusted, AQMesh sensors only’).

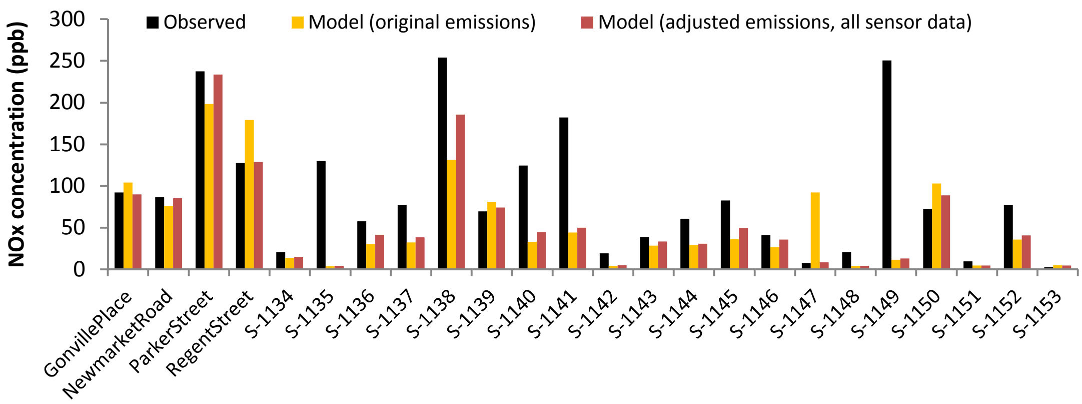

Figure 2. Change in mean emission rate per road source

Figure 3 shows scatter plots of the comparison; Table 3 shows model evaluation statistics for the same cases. The results show that as expected, if the reference data are included in the optimisation then the model performance at the reference sites using the adjusted emissions improves dramatically. Of course, any monitor data being used for model validation should be excluded from the optimisation, but this result demonstrates that the scheme is behaving in the expected way. The few points above the y=2x line in this case represent data points that were omitted from the optimisation process because the monitored value was lower than the background concentration. The more important result is that if only the AQMesh sensor data are included in the optimisation the model performance at the reference sites improves noticeably: the scatter is reduced, the bias is reduced and the correlation is increased.

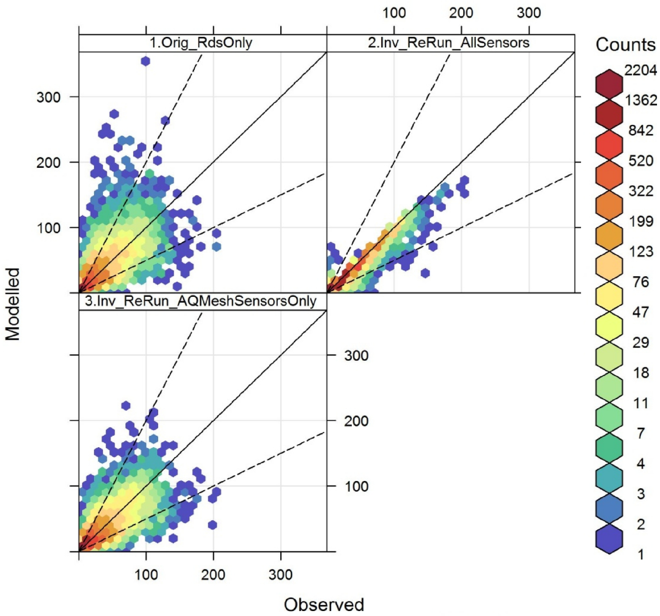

Figure 3. Frequency scatter plots of modelled versus observed hourly NO_x concentrations (ppb) at the reference monitors for three emissions cases: original emissions; adjusted emissions using all reference monitor and AQMesh sensor data; and adjusted emissions using AQMesh sensor data only

Table 3. Model performance statistics at the reference sites for three emissions cases: original emissions; adjusted emissions using all reference monitor and AQMesh sensor data; and adjusted emissions using AQMesh sensor data only

Statistics		Original	Adjusted, all sensors	Adjusted, AQMesh sensors only
Mean	**Obs**	31.2	31.2	31.2
	Mod	34.5	29.3	31.3
StDev	**Obs**	27.9	27.9	27.9
	Mod	31.0	26.0	27.0
MB		3.30	-1.91	0.10
NMSE		0.51	0.05	0.39
R		0.70	0.97	0.75
Fac2		0.71	0.94	0.73

Effect of the optimisation on emissions

The average diurnal emission factors and average emission rates were calculated for the original emissions and for the two sets of adjusted emissions. Figure 4 compares the diurnal emission factors, Figure 5 compares the average emission rate per road source and Table 4 gives the average difference in emission rate over all sources.

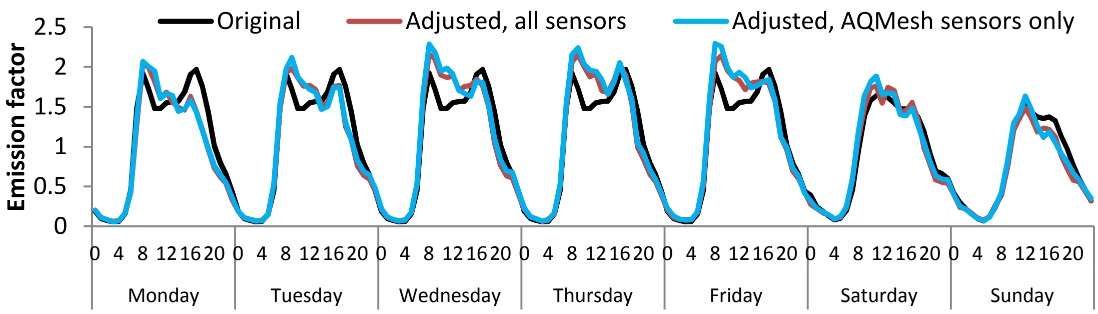

Figure 4. Comparison of diurnal emission factor profiles calculated from the original and adjusted emissions

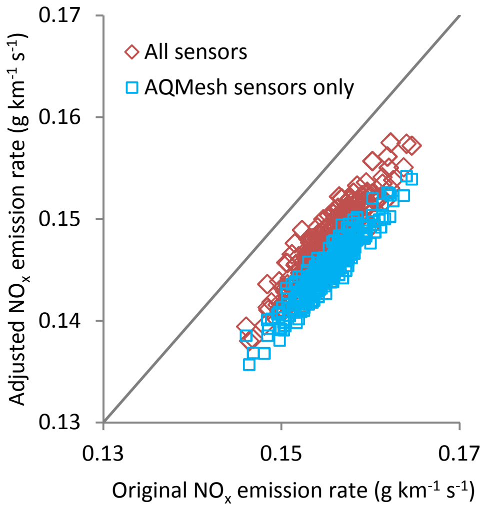

Figure 5. Scatter plot of adjusted versus original NO_x emission rates. The y=x line is shown in grey.

Table 4. Average change in NO_x emission rate (g km^{-1} s^{-1})

	Original	Adjusted	Change
All sensors	0.1552	0.1478	-4.8%
AQMesh only		0.1452	-6.5%

The diurnal emission factor profiles show that on weekdays the optimisation reduces the evening rush-hour emissions peak and increases the signal of the morning rush-hour peak. The magnitude of the change is slightly greater when the reference monitor data are not included in the optimisation, but the patterns of the changes are very similar irrespective of whether the reference monitors are used. The effect of the optimisation scheme on average emission rates is to reduce the emission from every source. When the reference monitors are included in the optimisation the calculated reduction in emissions is lower because on average the reference monitors' measured concentrations are slightly more consistent with the initial estimate of emissions.

CONCLUSION

The optimisation scheme presented here, using inversion techniques to modify pollution emission rates based on sensor data, has been shown to improve the accuracy of modelled concentrations. The current version of the scheme uses a relatively simple representation of error covariance. Indicators of emissions error covariance that are not yet accounted for include: distance between sources and meteorological factors such as temperature. Multiple pollutants and different source types also need to be accounted for; in the study presented here only road source emissions of NO_x have been considered, but a comprehensive ADMS-Urban modelling study of an urban area will also include pollutants such as PM_{10} and $PM_{2.5}$, and point, line, area and volume sources in addition to road sources. Defining the covariance in error between different pollutants and between difference source types presents a challenge, but the encouraging initial results presented here suggest that this approach could make practical use of large networks of low-cost sensors to improve dispersion model results.

REFERENCES

Anenberg, S., J. Miller, R. Minjares, L. Du, D. Henze, F. Lacey, C. Malley, L. Emberson, V. Franco, Z. Klimont and C. Heves, 2017: Impacts and mitigation of excess diesel-related NO_x emissions in 11 major vehicle markets. *Nature*, **545**, 467–471.

Carruthers, D., D. Clarke, K. J. Dicks, R. A. Freshwater, M. Jackson, R.L. Jones, C. Lad, I. Leslie, A. J. Lewis, H. Lloyd, O. A. M. Popoola, A. Randle and S. Ulrich, 2016: Using a commercial low cost sensor network (AQMesh) to quantify urban air quality: comparing measured and modelling (ADMS-Urban) pollutant concentrations, Conference on 'Air Quality Monitoring: Evolving Issues and New Technologies', London, UK, December 2016.

Denier van der Gon, H. , R. Bergström, C. Fountoukis, C. Johansson, S. Pandis, D. Simpson and A. Visschedijk, 2015: Particulate emissions from residential wood combustion in Europe – revised estimates and an evaluation. *Atmos. Chem. Phys.*, **15**, 6503–6519.

Kumar, P., L. Morawska, C. Martani, G. Biskos, M. Neophytou, S. Di Sabatino, M. Bell, L. Norford, R. Britter, 2015: Rise of low-cost sensing for air pollution in cities. *Environment Intl.*, **75**, 199-205.

Owen, B., Edmunds, H.A., Carruthers, D.J. and Singles, R.J., 2000: Prediction of total oxides of nitrogen and nitrogen dioxide concentrations in a large urban area using a new generation urban scale dispersion model with integral chemistry model. *Atmospheric Environment*, **34(3)**, 397-406.

Stocker, J., C. Hood, D. Carruthers, M. Seaton, K. Johnson and J. Fung, 2014: The development and evaluation of an automated system for nesting ADMS-Urban in regional photochemical models. *13th Annual CMAS Conference, Chapel Hill, NC, USA, October 2014.*

Webster, H. N., D. J. Thomson, M. C. Cooke and R. E. Pelley, 2016: Improvements to an operational inversion method for estimating volcanic ask source parameters using satellite retrievals, 17th Intl. Conference on Harmonisation for Regulatory Purposes, Budapest, Hungary, May 2016.