Presented at the 16th Annual CMAS Conference, Chapel Hill, NC, October 23-25, 2017

Development and validation of a rapid urban scale dispersion modelling platform

Scott L. Hamilton*
Ricardo Energy and Environment, UK

Nicola Masey, Iain Beverland
University of Strathclyde, UK

1. INTRODUCTION

Dispersion models simulate atmospheric transport and transformation of air pollutants emitted from sources to allow estimation of concentrations at receptors. Gaussian dispersion models can suffer from lengthy run-times for large urban areas which can quickly become prohibitive due the computational demands of calculating concentrations at a suitable number of locations. RapidAir offers an elegant solution, particularly when annual average concentrations are required.

Here we describe the development and evaluation of a new dispersion model (RapidAir®, Ricardo-AEA Ltd) designed as a decision support platform, and describe a recent validation exercise in London, UK which was carried out by Strathclyde University, UK. Ricardo's RapidAir model comprises several libraries written in the python programming language with functionality specific to air quality analysis (e.g. handling time series observation data, array based processing of road emissions).

The modelling system is written using open source scientific computing libraries. RapidAir makes extensive use of the numpy, scipy and pandas python libraries which enable very efficient scientific computation though their use is relatively rare in the air quality community in the UK. The model is built inside an Anaconda environment with the above libraries in python 2.7.

RapidAir has the following functions:

- Convenient control of AERMET and AERMOD to create convolution kernels
- Calculation of gridded road emissions with COPERT5 coefficients (>240 vehicle types) and Geographical Abstraction Library (GDAL)
- Period mean road source dispersion model based on convolution of emission grids with AERMOD derived dispersion kernels
- Automated background concentrations assimilation (UK only)
- Two street canyon models (USEPA STREET and AEOLIUS)
- Statistical evaluation/plotting module based on python's statsmodels library.

RapidAir is set up using USEPA methods provided in their 'Hotspot Conformity' guidance (USEPA, 2015, Appendix J). Appendix J sets out methods specific to dispersion modelling of road traffic emissions in AERMOD- such as setting release height, initial plume depth and other factors which have been adopted in RapidAir. A useful feature of the model is the fact that an entire workflow (from gathering met observations through to model validation) can be set up and run from a single text file/python IDE, though we have also developed a menu driven UI in the open-source Jupyter Notebook (http://jupyter.org/) format to make it easier for users to learn the model.

A key motivation for the development of RapidAir was our experience of a lack of a cost-effective (to us) operational city-scale dispersion model with convenient run times, which does not require large amounts of manpower to operate.

In this paper, we present some results from of a validation exercise based on annual mean concentrations of NO_2 in London in 2008, closey following methods set out in a study by the UK Government.

2. Methods

2.1 Study area and receptor location

We modelled concentrations of NOx and NO_2 in Greater London. This was the study area used in a previous Department for Environment, Food and Rural Affairs (DEFRA) Urban Model Evaluation exercise, which evaluated several existing models (Carslaw, 2011). We modelled annual average NOx and NO_2 concentrations for 2008; the same year used by the DEFRA study to enable statistical comparison between RapidAir and the models assessed in the DEFRA comparison.

**Corresponding author:* Scott L. Hamilton, Ricardo Energy and Environment, 18 Blythswood Square, Glasgow, UK; e-mail: scott.hamilton@ricardo.com

We evaluated the model at 86 continuous monitoring locations from the London Air Quality Network (LAQN) monitoring network. For model evaluation purposes, the monitoring sites are designated as kerbside, roadside, suburban and urban background according to their proximity to road traffic sources. As in the DEFRA Urban Model Evaluation, we excluded any sites which collected less than 75 % data during 2008.

2.2 Model description

RapidAir uses a dispersion kernel based convolution procedure which operates in a similar way to algorithms used in image processing software (e.g. Gaussian blur, edge detection). Convolution is also widely used in the astronomy community. An idealized dispersion model plume (the kernel) calculated for a single small area source in AERMOD is convolved with a road traffic emission grid at the same resolution. The convolution procedure (which uses a scipy library in python, followed by a GDAL conversion procedure to GIS ready data) yields a continuous concentration surface comprising millions of overlapping plumes derived from the road source emissions. A useful feature is that model run time is linearly dependent only on the output number of cells (typically about 500 million) and is unaffected by the number of emissions sources in the domain (recent runs which are not reported here comprise several million discrete road sources). This is a key benefit compared with some models whose run time is linearly dependent both on resolution/number of receptors and number of sources. The RapidAir convolution model computes concentrations at hundreds of millions of discrete receptors in less than 3 minutes on a standard office laptop with 16GB RAM and an Intel i7 processor. For cases requiring a bigger grid the model includes a splitting algorithm which iterates over “tiles” describing the emissions- early tests have recently produced continuous concentration fields of annual mean NO_2 at 3m resolution for the largest 6 cities in the UK with a total run time of less than 20 minutes.

During the London study, regional background concentrations calculated by the DEFRA Pollution Climate Mapping (PCM) model were added to the pollution surface generated by RapidAir. We removed hot exhaust road sources of NOx prior to adding the background to the modelled pollution concentrations above to prevent double-counting of traffic related air pollutants. The last step was the empirically derived conversion of NOx to NO_2.

2.3 Street canyons

In built-up urban areas air pollution can become trapped in street canyons surrounded by tall buildings. Exposure estimates may be improved by combining additional models that account for urban topography. Street canyon models range from complex computational fluid dynamics models to simpler empirical (e.g. USEPA STREET box-model) and semi-empirical models (e.g. Danish Operational Street Pollution Model (OSPM).

Geospatial surrogates have been investigated to model the effect of street canyons on air quality in urban locations. Such metrics are commonly used in studies of urban climate where temperature and wind patterns are affected by building density and height. For example, sky view factor (SVF) can be used to indicate the presence of tall buildings and has been incorporated into a LUR models to estimate the presence of street canyons. Building height and/or volume information has also been observed to improve the accuracy of LUR model estimates (Gillespie et al., 2016). Geospatial surrogates can be readily applied across entire cities in automated processes which reduce geometrical errors and aid efficiency. The use of geospatial surrogates also has potential to improve the reproducibility of dispersion model pollution estimates as the number of model design choices is reduced substantially (with corresponding substantial reduction in manpower costs).

Of the 86 receptor locations we identified 19 sites that were located within urban street canyons. Concentrations of NOx within these street canyons were estimated using two street canyon models: the STREET model (Dabberdt et al., 1973; Johnson et al., 1973) and the AEOLIUS Model (Buckland and Middleton, 1999). Both canyon algorithms are included in RapidAir.

Building height data for London was used to create a 5 m raster created of the maximum building height within each cell.

Sky view factor (SVF) was used to calculate this value using the building height raster as the input and a search radius of 200 m.

Hill shading (HS) is commonly used to identify areas in shade as a result of surrounding topographical features. We calculated this using an elevation angle of 45 degrees (suggested to be most appropriate for steep terrain as is encountered in an urban environment. We calculated HS values for 8 sectors (i.e. every 45 degrees) and averaged the HS values calculated to produce estimated HS value over the study area.

Wind Effect (WE) was estimated using a module in SAGA-GIS which predicts if an area is wind shadowed or exposed, where values below 1 are shadowed and above 1 are exposed. As above for HS, WE values were calculated for 8 sectors and the average if these values were used. A search radius of 200 m was used.

A recent update to RapidAir has added an automatic street canyon allocator in a custom python module. Our most recent version of the London model has around 3000 individually treated street canyons which is populated with emissions and concentrations in the model in a few minutes.

3. Results and discussion

3.1 Model with no urban effects

The baseline RapidAir kernel model (i.e. no urban morphology treatment) highlighted contributions from major roads in London, and Heathrow Airport in the west of the study area, however the model underestimated observed NOx concentrations at the receptor locations. The underestimation of the concentrations by the base model may be the result of uncertainties in background concentrations, road traffic emissions or monitoring locations. However, as we used publicly available open source data to generate the model we did not investigate these uncertainties further to ensure reproducibility of the model results and comparability with other groups who used the same data sets.

It is likely that road traffic NOx emissions data are underestimated in the inventory we used. The European Environment Agency's COPERT road traffic emissions model has been observed to under-predict historical NOx emissions from diesel vehicles in the UK fleet (Carslaw et al., 2011). Given the date of the data used in the model, the fact that evidence for COPERT under predicting traffic NOx only came to light around 2011, and the fact that COPERT was used by the GLA to make the emissions estimates in 2008, it is likely that reported under-prediction of emissions in the diesel fleet biased the road traffic NOx inventory towards under-prediction.

We corrected the NOx kernel model for systematic underestimation bias using the regression equation derived between the modelled and measured concentrations following UK statutory guidance provided by DEFRA (2016). The receptor locations were split randomly into training ($n = 57$) and test ($n = 29$) data sets, with the latter used as an independent verification data set.

The linear regression (using the training data) for the model adjustment of the raw model is shown in Equation 1:

$$NOx_{obs} = 1.98 * NOx_{mod} \quad (1)$$

Where NOx_{obs} and NOx_{mod} are concentrations in μg/m^3.

Legislative limit values specified by the European Union and UK government are for NO_2, and not NOx, therefore we converted RapidAir NOx concentrations to NO_2 concentrations using the DEFRA NOx to NO_2 model (version 3.2) which is recommended for use in UK air quality assessment for statutory purposes. The model was set to use the built-in fleet composition for London (which automatically sets the fraction of NOx emissions as NO_2 (f-NO_2)) and average regional oxidant concentration over the study area from the PCM model. Estimated NO_2 concentrations were plotted against NOx concentrations and fitted with a polynomial regression equation (Equation 1):

$$NO_2 = -0.0001 * NOx^2 + 0.2737 * NOx + 18.6 \quad (2)$$

Where NOx and NO_2 concentrations are in μg/m^3. This equation was used to convert annual mean NOx to NO_2 in the model.

NO_2 concentrations predicted by RapidAir were similar to measured NO_2 concentrations at most monitoring stations; however, the model underestimated concentrations at some very high concentration kerbside measurement sites. Underestimation could be attributed to urban morphologies (including street canyon effects) or underestimation in the emissions rates used to predict the NOx concentrations (Beevers et al., 2012).

The correlation between modelled and observed NO_2 concentrations was high ($r = 0.77$) and of similar magnitude to previous evaluations of dispersion models (e.g. $r = 0.74$ during an evaluation of NOx dispersion models carried out during the ESCAPE study (de Hoogh et al., 2014)).

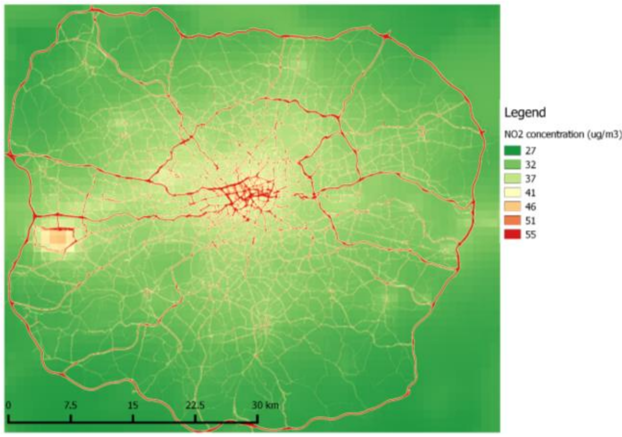

Figure 1: Annual mean NO_2 concentrations over the Greater London conurbation for the RapidAir kernel model

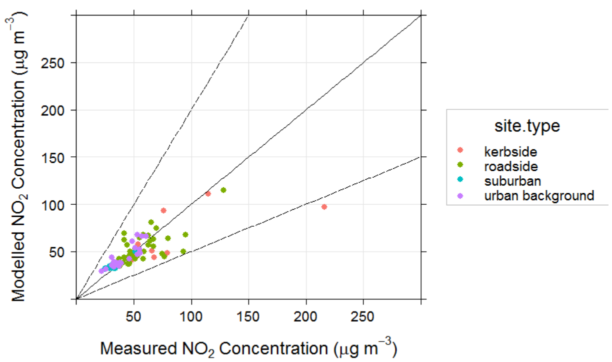

Figure 2: Scatter plot of NO2 estimated by RapidAir kernel model vs. observed concentrations at 86 stations. Receptors are colour coded to represent the different site types. Solid line represents 1:1 dashed lines represent FAC2 values.

Station Type	n	FAC2	NMB	RMSE ($\mu g/m^3$)	r
All	86	0.99	-0.05	17.1	0.8
Kerbside	8	0.88	-0.25	45.1	0.7
Roadside	40	1.00	-0.07	13.9	0.7
Suburban	13	1.00	0.08	4.0	0.9
Urban background	25	1.00	0.06	6.0	0.9

Table 1: NO_2 kernel model evaluation statistics (after adjustment for systematic bias) for LAQN receptor locations categorized by site type.

DEFRA recommend that an air quality model is acceptable for use if more than half of its observations fall within a factor of 2 of the observations (Williams et al., 2011).

The NO_2 RapidAir model meets the FAC2 criterion for all site types, with the lowest FAC2 value calculated for kerbside sites (FAC2 = 0.88). Similar findings were reported in the DEFRA urban model evaluation exercise for NO_2 which found that FAC2 values were lower for the kerbside sites than the three other site types tested (Carslaw, 2011).

DEFRA also suggest that NMB values should lie between -0.2 and 0.2 (Williams et al., 2011). NMB values for RapidAir meet this criterion when all sites were considered together; and for the individual site types, with the exception of the kerbside sites. None of the models tested during the DEFRA model evaluation exercise met the NMB 'acceptance values' proposed by DEFRA at the kerbside sites.

3.2 Model with canyon treatments

A calibration equation was derived between *Unadjusted modelled NOx* vs. both *modelled NOx* and *Surrogate* for each of the three surrogate values investigated (Table 2a). Applying the calibration equations to the test NOx data resulted in similar coefficients of determination and regression equations to the RapidAir estimates (Table 2b).

(a)	Surrogate	Measured NO_x	R^2
	RapidAir	1.98 * $RapidAir\_NO_x$	0.88
	SVF	1.87 * $RapidAir\_NO_x$ – 70.61 * SVF + 55.90	0.71
	WE	2.00 * $RapidAir\_NO_x$ – 90.99 * WE + 85.43	0.70
	HS	2.01 * $RapidAir\_NO_x$ – 54.04 * HS + 49.57	0.69
(b)	Model	Measured NO_x	R^2
	RapidAir	0.79*$RapidAir\_NO_x$	0.94
	SVF	0.79*$RapidAir\_NO_x$	0.94
	WE	0.78*$RapidAir\_NO_x$	0.94
	HS	0.78*$RapidAir\_NO_x$	0.94

Table 2: (a) Linear regression equations between receptor NOx (Measured NOx), and kernel model NOx concentrations and the surrogate variables for the training data set (n = 59); (b) Ordinary least squares regression equations between the measured and kernel model NOx concentrations (baseline and after surrogate correction) for the test data set (n = 29).

Model	$Model\_NO_x$ ($\mu g/m^3$)
Kernel	1.98 * Meas_NOx, R^2 = 0.88
STREET	1.04 * $Meas\_NO_x$ + 34.45, R^2 = 0.75
AEOLIUS	1.41 * $Meas\_NO_x$ + 18.87, R^2 = 0.73

Table 3: Linear adjustment equations to account for systematic bias in kernel model performance. Equations are shown for the kernel model; the kernel model corrected and including surrogates; and kernel model including street canyon models.

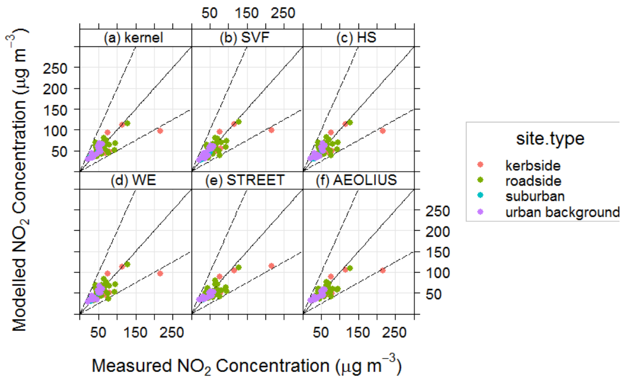

Figure 3: Scatter plot of NO_2 estimated by RapidAir kernel model *vs*. observed concentrations (NO_2) (n = 86): (a) uncorrected concentrations from the base-kernel model; and the kernel model after correction using the surrogates for street canyons: (b) sky view factor (SVF), (c) hill shading (HS), (d) wind effect (WE), (e) STREET canyon model and (f) AEOLIUS canyon model.

The difference in modelled concentrations between the STREET and AEOLIUS models was very small which is similar to previously published findings (Zhu et al., 2015).

Despite the small change in model evaluation statistics the canyon models require less adjustment for systematic bias than the kernel only model (Table 3). Therefore, when this model is applied to areas of the city which do not have any measurements the model is less likely to be subject to over or under estimation than the standard model which does not attempt to address urban morphology.

Model	n	FAC2	NMB	RMSE ($\mu g/m^3$)	r
Kernel	86	0.99	-0.05	17.1	0.8
SVF	86	0.99	-0.06	16.3	0.8
WE	86	0.98	-0.05	17.0	0.8
HS	86	0.99	-0.06	17.0	0.8
STREET	86	1.00	-0.09	15.9	0.9
AEOLIUS	86	0.99	-0.08	16.4	0.8

Table 4: Summary model evaluation statistics for annual mean NO2 at receptor locations. Statistics are given for the bias corrected Kernel only model, the kernel model after correction using the surrogates for street canyons and then bias corrected, and using the street canyon models with bias correction.

4. Advantages and disadvantages

The central focus of this work is to evaluate an air quality modelling platform aimed at the operational setting. The RapidAir model succeeds as an operational air quality model in the context of very large urban areas and as a decision support tool.

A significant benefit with RapidAir is reduced computational burden. Run times of 10 minutes or less present a benefit for the operational modeler and decision makers who require fast but robust analyses. The RapidAir platform allows extremely efficient policy testing and other "what if" model runs for new emission scenarios.

The performance metrics are very similar to those computed for other dispersion modelling systems in the DEFRA inter comparison exercise. For example, the suite of RapidAir outputs for kerbside locations in London have NO_2 RMSE values of 38.91 – 45.26 µg/m3 (r = 0.65 - 0.84, n = 8) where the comparable models in the inter comparison have RMSE values ranging from 29.39 to 67.09 µg/m3 (r = 0.15 - 0.93, n = 7). At roadside locations, the RapidAir outputs have NO_2 RMSE values of 12.78 – 14.28 µg/m3 (r = 0.70 - 0.76, n = 40) where the models in the inter comparison have RMSE values ranging from 9.94 to 19.69 µg/m3. (r = 0.38 - 0.89, n = 30). The key model metrics for the 2008 model run in London are very similar to those for standard modelling suites used in the UK and which are used and accepted by DEFRA for use in compliance assessments at the highest level of statutory European air quality reporting.

The performance statistics for the surrogates for urban morphology are reasonably close to those from the models which treat canyons discretely. Again, our focus is on operational modelling where reproducible and efficient workflows are important. We would suggest that for compliance assessment RapidAir is used with either the STREET or AEOLIUS model options included as the run times are not significantly impacted by including these models. The model results should be compared with measured concentrations and the modeler may choose the best performing street canyon model for their case. The surrogate canyon models could be used as screening tools and perhaps to spatially delineate locations where the street canyon models should be invoked.

5. Educational value and engagement

The RapidAir project would not have been possible without important contributions from groups like the USEPA, NOAA, UK Met Office, DEFRA, and the authors of python itself and the open source libraries we use in the model. Hence Ricardo take seriously the notion that we should return some value to the community as we have benefited so much from the efforts of others. Mainly we do this via educational outreach efforts in our lead developer's home city.

The RapidAir project has provided a platform for Ricardo's educational engagement with universities in Glasgow, UK. Some of the primary research and evaluation was done via a formal industrial sponsorship we provided to Strathclyde University, Glasgow. This engagement contributed to a recently successful PhD candidate (Nicola Masey, now of Ricardo) and we are industrial supervisors on two more projects. We also collaborated with Glasgow School of Art in providing a RapidAir model of Glasgow which was used by Trudi Hannah during a placement with us to create a physical air pollution model of the city. This was used to explore routes to better public engagement through visualization- the model was nominated for an international design award in 2016.(https://www.informationisbeautifulawards.com/showcase/1614-3d-visualisation-of-air-pollution-in-glasgow) .

5. Future directions

We are currently working on transitioning the model into an integrated platform with CMAQ or WRF-Chem providing chemical boundary conditions. Our CMAQ system is operational and produces hourly forecasts each day for the whole of the UK. We have developed a beta version of RapidAir that consumes the same WRF data via the MMIF processor to produce forecasts out to +48hrs for the city of London. However, we were not able to prepare results in time to present to the conference on this occasion.

6. References

Beevers, S.D., Westmoreland, E., de Jong, M.C., Williams, M.L., Carslaw, D.C., 2012b. Trends in NOx and NO2 emissions from road traffic in Great Britain. Atmos. Environ. 54, 107–116. doi:10.1016/j.atmosenv.2012.02.028

Berkowicz, R., 2000. OSPM - A Parameterised Street Pollution Model. Environ. Monit. Assess. 65, 323–331. doi:10.1023/A:1006448321977

Buckland, A.T., 1998. Validation of a Street Canyon Model in Two Cities. Environ. Monit. Assess. 52, 255–267. doi:10.1023/A:1005828128097

Carslaw, D., 2011. Defra urban model evaluation analysis - Phase 1. King's College London.

Carslaw, D., Beevers, S.D., Westmoreland, E., Williams, M., Tate, J., Murrells, T., Stedman, J., Li, Y., Grice, S., Dabberdt, W.F., Ludwig, F.L., Johnson, W.B., 1973. Validation and applications of an urban diffusion model for vehicular pollutants. Atmospheric Environ. 1967 7, 603–618. doi:10.1016/0004-6981(73)90019-X

DEFRA, 2017. The Data Verification and Ratification Process [WWW Document]. URL https://uk-air.defra.gov.uk/assets/documents/The_Data_Verification_and_Ratification_Process.pdf (accessed 4.1.17).

DEFRA, 2016. Local Air Quality Management Technical Guidance (TG16) (No. TG16). Defra.

Gillespie, J., Beverland, I.J., Hamilton, S., Padmanabhan, S., 2016. Development, Evaluation, and Comparison of Land Use Regression Modeling Methods to Estimate Residential Exposure to Nitrogen Dioxide in a Cohort Study. Environ. Sci. Technol. 50, 11085–11093. doi:10.1021/acs.est.6b02089

Gillespie, J., Masey, N., Heal, M.R., Hamilton, S., Beverland, I.J., 2017. Estimation of spatial patterns of urban air pollution over a 4-week period from repeated 5-min measurements. Atmos. Environ. 150, 295–302. doi:10.1016/j.atmosenv.2016.11.035

Gulliver, J., Briggs, D., 2011. STEMS-Air: A simple GIS-based air pollution dispersion model for city-wide exposure assessment. Sci. Total Environ. 409, 2419–2429.

Johnson, M., MacNeill, M., Grgicak-Mannion, A., Nethery, E., Xu, X., Dales, R., Rasmussen, P., Wheeler, A., 2013. Development of temporally refined land-use regression models predicting daily household-level air pollution in a panel study of lung function among asthmatic children. J. Expo. Sci. Environ. Epidemiol. 23, 259–267. doi:10.1038/jes.2013.1

Johnson, W.B., Ludwig, F.L., Dabberdt, W.F., Allen, R.J., 1973. An Urban Diffusion Simulation Model For Carbon Monoxide. J. Air Pollut. Control Assoc. 23, 490–498. doi:10.1080/00022470.1973.10469794

Middleton, D.R., 1999. Development of AEOLIUS for street canyon screening. Clean Air 29, 155–161.

Middleton, D.R., 1998a. Dispersion Modelling: A Guide for Local Authorities, Met Office Turbulence and Diffusion Note number 241. The Meteorologcal Office, Bracknell, Bercks.

Middleton, D.R., 1998b. A New Box Model to Forecast Urban Air Quality: Boxurb. Environ. Monit. Assess. 52, 315–335. doi:10.1023/A:1005817202196

U.S. Environmental Protection Agency, 2017. 40 CFR Part 51 Revisions to the Guideline on Air Quality Models: Enhancements to the AERMOD Dispersion Modelling System and Incorporation of Approaches To Address Ozone and Fine Particulate Matter, EPA-HQ-OAR-2015-0310; FRL-9956-23-OAR.

U.S. Environmental Protection Agency, 2015. Transportation Conformity Guidance for Quantitative Hot-spot Analysis in PM2.5 and PM10 Nonattainment and Maintenance Areas (No. EPA-420-B-15-084).

Williams, M., Barrowcliffe, R., Laxen, D., Monks, P., 2011. Review of Air Quality modelling in Defra. Defra.

World Health Organization, 2016. Ambient air pollution: A global assessment of exposure and burden of disease. Geneva, Switzerland.

World Health Organization, 2013. Review of evidence on health aspects of air pollution – REVIHAAP Project: Technical Report. Copenhagen, Denmark.

Zhu, G., Zhang, P., Tshukudu, T., Yin, J., Fan, G., Zheng, X., 2015. Forecasting traffic-related nitrogen oxides within a street canyon by combining a genetic algorithm-back propagation artificial neural network and parametric models. Atmospheric Pollut. Res. 6, 1087–1097. doi:10.1016/j.apr.2015.06.006