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

Integrating regional and local modeling to create a high-resolution air quality forecasting system for Hong Kong

Christina M. Hood*, Jenny R. Stocker, David J. Carruthers, William Grayson, Jonathan Handley
Cambridge Environmental Research Consultants, Cambridge, UK

Jimmy Fung and David Yeung
Hong Kong University of Science and Technology, Hong Kong

1. INTRODUCTION

The air quality in urban areas shows complex variations in both time and space, due to influences on many scales from regional pollutant transport to individual road emissions and street canyon geometries. Access to street-scale air quality forecasting data allows residents to make informed choices about transport modes and routes in order to minimise their exposure to air pollutants as they travel around the urban area.

The only air quality data currently publicly available for Hong Kong is an air quality health index value for each of 16 monitoring stations, based on real-time measurements. The Personalised Real time Air quality Informatics System for Exposure (PRAISE-HK) project aims to develop a mobile app that will provide high-resolution forecast pollution data, which will be combined with route planning and personal health alerts.

Several empirical forecasting approaches have previously been described for Hong Kong, all limited to relatively few specific pollutants and none run as publicly available operational forecasts. The synoptic climatological approach (Cheng and Lam 2000) cannot account for local variations of emissions or fine-scale meteorology. Statistical approaches, for example fuzzy logic (Pokrovsky et al. 2002), neural networks (Lu et al. 2004) and regressions (Goyal et al. 2006) all require large volumes of input observational data and are limited to specific output locations.

For more general forecasting applications, it is preferable to use models which take into account the transport and chemistry of emitted pollutants. Regional photochemical models such as CMAQ (Byun and Schere 2006) can be used to forecast average pollutant concentrations on neighbourhood scales, but this does not give an accurate representation of the local concentrations, which can vary significantly over distances of a few metres from a road. Local models such as ADMS-Urban (Owen et al. 2000) are able to capture the small-scale dispersion and chemistry processes which occur close to individual sources but have limitations when accounting for the longer-term transport and chemistry processes affecting emissions from further afield.

Coupling a regional and a local model creates a computationally efficient system for calculating high-resolution predictions of pollutant concentrations, although care is required to avoid double-counting emissions. Such a system, the ADMS-Urban Regional Model Link (RML), was described in Stocker et al. (2014), with all components running in the Windows operating system. For the current work, the system has been rebuilt to run entirely on the same Linux High Performance Computing (HPC) facility as the regional modeling and also includes increased automation for operational forecasting.

Section 2 describes the ADMS-Urban RML concept while Section 3 gives information about the Hong Kong forecasting implementation. Section 4 outlines the system performance for an example domain over an initial 4-week trial period and Section 5 describes the planned development of the system during the remainder of the project.

2. SYSTEM CONCEPT

The original ADMS-Urban RML nesting system concept was described in Stocker et al. (2012). The local and regional influences on output concentrations are separated using a mixing time, which represents the time taken for emissions from explicitly-modeled sources in an urban area to become well-mixed on the scale of the regional model grid. For longer times, the regional model is used to represent dispersion and chemistry on large scales. For shorter times, the local model calculates details of dispersion and fast chemistry for emissions from individual sources. In order to avoid double-counting

Corresponding author: Christina M. Hood, 3 King's Parade, Cambridge CB2 1SJ, UK; e-mail: chood@cerc.co.uk

emissions, the local model is configured with gridded emissions in order to estimate the concentrations generated in the regional model during the mixing time, which are subtracted from the full regional model concentrations before the concentrations from the local model run with explicit sources are added.

Specific output locations can be defined by the user, for example to match measurement locations. Additionally, a regular grid of output points can be defined, with additional points automatically added close to road sources in order to create high-resolution contours of concentration.

The Windows-based system was described in Stocker et al. (2014) alongside its application in Hong Kong. Subsequently the system has been re-built on a Linux operating system in order to make it available for use on HPC systems. The level of system automation has been increased, with the regional model grid information being read automatically by one utility and transferred through the system to other utilities. Support for the WRF-Chem (Grell et al. 2005) and CHIMERE (Schmidt et al. 2001) regional models has been added, in addition to the existing support for CMAQ, CAMx (ENVIRON, 2016) and EMEP4UK (Vieno et al. 2010).

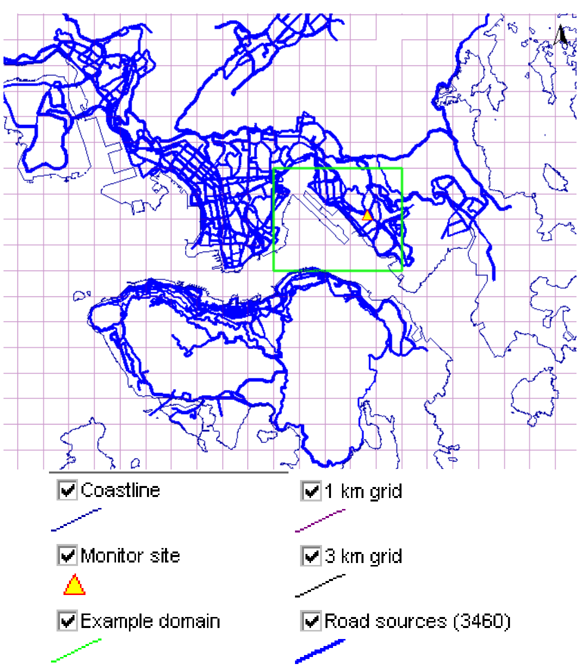
A new user option has been added to apply bilinear interpolation to regional model cell centre concentrations in order to calculate the regional model concentration at an arbitrary output point. For consistency, this has also required changes to the ADMS-Urban run settings to run additional grid cells and apply similar interpolation.

The version of the local model ADMS-Urban included in the system (4.1.2) has been updated to include automatic modeling of emissions from road tunnel portals as volume sources with dimensions depending on hourly wind speed as well as traffic speeds and tunnel dimensions.

3. FORECASTING IMPLEMENTATION

An example 5x4 km local domain, shown in Figure 1, has been used in the initial development of the high-resolution forecast system. This area (Kowloon East) is of particular current interest due to redevelopment of the former Kai Tak airport site. High resolution air quality forecasts are calculated for three days. The output species are the regulated pollutants NO_2, NO_x, O_3, $PM_{2.5}$, PM_{10}, SO_2 and CO. The value of overall ‘Air Quality Health Index’ is also calculated from the concentrations as a post-processing step.

The regional model inputs are as described in Li et al. (2013), although applied to CMAQ instead of CAMx. Note that the initial implementation of the system uses the local and regional emissions inventory for 2010 as in the previous work described in Stocker et al. (2014). This is known to include some inconsistencies between the local and regional emissions. The explicit road locations are also outdated relative to the current road network due to the changing character of the example area. The three road tunnel portals included in the example domain are modeled using the old approach of manually-defined



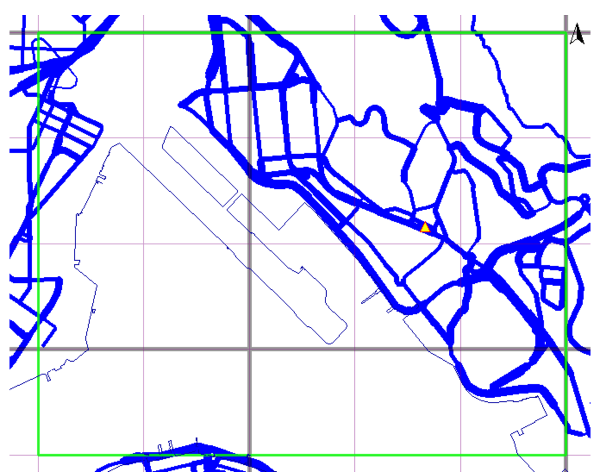


Fig. 1. Maps of example domain in the context of wider Hong Kong island and Kowloon (left), and detail showing the Kwun Tong monitoring site, roads included in the local model and regional model 1 km and 3 km grid cell boundaries (above).

volume sources with fixed dimensions. The emissions inventory and model configuration will be fully updated and further model verification carried out prior to public release of forecasts from the system.

The HPC facility used to run the forecast consists of a cluster of dual CPU servers, mainly Intel E52690 and E52699, which provide 28 cores and 44 cores, respectively. There is 128GB RAM and 4TB local RAID disk. This facility is used to run both the regional model and the ADMS-Urban RML system. The ADMS-Urban RML forecast run is triggered at around 4 am local time each day, following the completion of the regional model forecast run.

The output netCDF concentration files are processed to create contour image files for each hour, which are then converted into tiles using the Geospatial Data Abstraction Library (GDAL). This allows the high-resolution concentration forecast to be displayed through the Google Maps API.

4. SYSTEM PERFORMANCE

The high-resolution forecasting system has been assessed in terms of the computational performance of the operational system and the scientific performance in terms of the forecast concentrations at the single monitoring site within the model domain. Three sets of results are compared: the regional model CMAQ 3 km and 1 km resolution domains and the coupled ADMS-Urban RML outputs, in order to investigate the effects of increased resolution. The performance analysis has been carried out for an operational trial period covering the four weeks from 5th September to 2nd October 2017 inclusive. Concentrations are analysed from the first forecast day, with a focus on the pollutants NO_2 and $PM_{2.5}$.

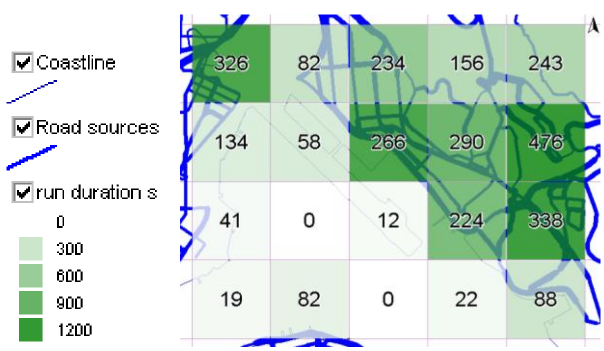


Fig. 2. Spatial variation of explicit ADMS-Urban run time duration in s (shading scale on left) for 24 modeled hours and number of straight-line road segments within each regional model 1km grid cell (numeric label).

Table 1. Output file size (absolute and normalized by domain area) per forecast day for regional model CMAQ 3 km and 1 km domain and high resolution ADMS-Urban RML.

	Extent *cells*		**Cell size *km***	**File size**	
Model domain	**x**	**y**		**Total *MB***	**Norm *MB/km²***
CMAQ 3 km	152	110	3	213	0.001
CMAQ 1 km	179	125	1	278	0.012
ADMS-Urban RML	5	4	(1)	66	3.300

4.1 Computational Performance

There was no system downtime during the trial period. The regional model forecasting system has a run time of 1.8 hours per forecast day for WRF and 2.5 hours for CMAQ, of which 43% is used for the 1 km domain, while the additional run time for the ADMS-Urban RML system for the example domain is 25 minutes per forecast day. The ADMS-Urban RML system run time is dominated by the run time for the ADMS-Urban runs with explicit sources, as the time required by all the other processes is less than one minute. The example domain covers 20 grid cells, so each grid cell can be run in parallel on a separate core of an individual node and the total run time is controlled by the slowest cell. There is a strong relationship between the run time and the number of straight-line road segments modeled within each cell, as illustrated in Figure 2. This is due to both the additional dispersion calculations associated with additional sources and additional output points added near each road segment, leading to a non-linear relationship.

The output file sizes from the system per forecast day for the 3 km and 1 km resolution CMAQ domains and the example ADMS-Urban RML domain are given in Table 1. When normalized by the domain area, the file sizes increase with the resolution of the model, as expected. The average density of the high-resolution output points is 4577 points/km^2 but this varies depending on the local density of road segments. The file size per output point is smaller for the high resolution forecast than for CMAQ due to a smaller number of stored concentration variables.

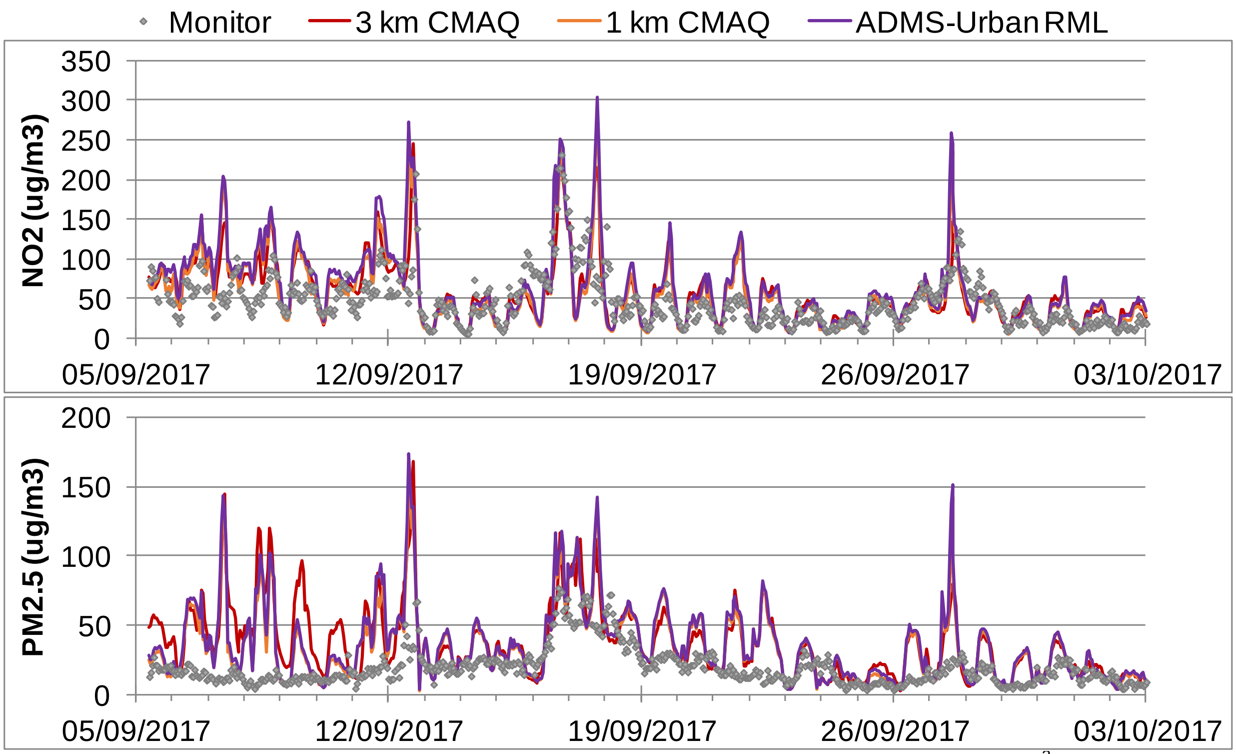


Fig. 3. Time series of observed and forecast NO_2 (upper) and $PM_{2.5}$ (lower) concentrations in µg/m^3 at the Kwun Tong monitoring site, from the CMAQ 3 km and 1 km domains and the ADMS-Urban RML system.

4.2 Scientific performance

The forecast concentrations of NO_2 and $PM_{2.5}$ from the 3 km and 1 km CMAQ domains and the ADMS-Urban RML are compared with measured concentrations at the Kwun Tong monitoring site in the time series plots shown in Figure 3. The location of this monitoring site is shown in the domain map in Figure 1 and it has an intake height of 25 m above local ground level. For NO_2, the forecast systems give concentrations of a similar range and structure to the observations, in particular correctly identifying episodes of higher concentrations on 12th, 16th and 27th September. There is however a tendency for all of the models to over-predict peak concentrations. The ADMS-

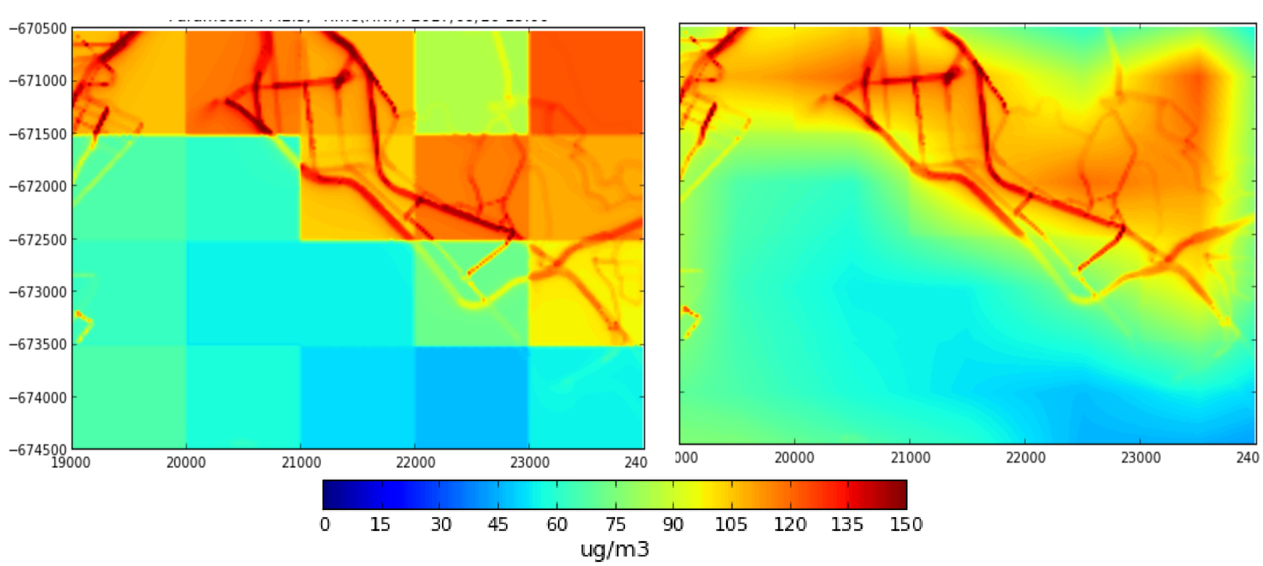


Fig. 4. Forecast PM2.5 concentration contours from ADMS-Urban RML for the example domain without (left) and with (right) interpolation option, for 3 pm local time on 16th September 2017.

Urban RML forecast concentrations are relatively similar to the CMAQ concentrations, suggesting that the nearby roads do not have a significant impact on this elevated receptor. For $PM_{2.5}$, CMAQ is substantially over-predicting concentrations for the first week of the period, especially in the 3 km domain, but correctly identifies the episode of higher concentrations 16th – 18th September. The ADMS-Urban RML forecast concentrations are very similar to the CMAQ 1 km domain concentrations, again suggesting that local roads are not significant for this receptor.

There was a public holiday (day following National Day) on 2nd October 2017 which is not taken into account in the emissions inventory. This may lead to slightly reduced monitored NO_2 concentrations during the last few days of the trial period, which are not reflected in the forecast concentrations.

Examples of the forecast contours of $PM_{2.5}$ concentration with and without interpolation of regional model concentrations are given in Figure 4. The contours with the interpolation option are substantially smoother across the regional model grid cell boundaries as well as more continuous along roads with the interpolation option in use.

5. DISCUSSION AND FUTURE WORK

The variation of run times for different cells within the example domain is due to the highly heterogeneous road density in this area, with some cells over water containing no road segments and some containing several hundred segments. These characteristics are also present in the wider Hong Kong domain to which the system will be expanded. The cell with the highest road density over the wider Hong Kong domain contains 587 straight-line road segments, 23% more than the densest cell in the example domain, so the maximum ADMS-Urban explicit run time is expected to increase to around 40 minutes for each forecast day. The total overall run time will depend on the available number of cores relative to the number of grid cells containing a high density of road segments.

The ADMS-Urban RML option to interpolate regional model concentrations gives substantially smoother contour images and more continuous concentrations along major roads. If this approach is taken for validation, interpolation should also be applied to regional model concentrations for consistency, however it is not clear whether this is a suitable depiction of the regional model results relative to the gridded calculation method.

The additional computational expense of running the high resolution forecast is relatively modest for the small example domain, but gives substantially more information about the local variations of concentration, especially where there are major roads. This additional information is of interest for identifying locations of high pollutant exposure and allowing individual or regulatory mitigation measures.

The current comparisons between the model forecast concentrations and the single monitoring site within the example domain are dominated by overestimates of both NO_2 and $PM_{2.5}$ by the CMAQ regional modeling, which also strongly influences the ADMS-Urban RML results. This is likely to be due to the use of an old emissions inventory, which does not include recent reductions in emissions.

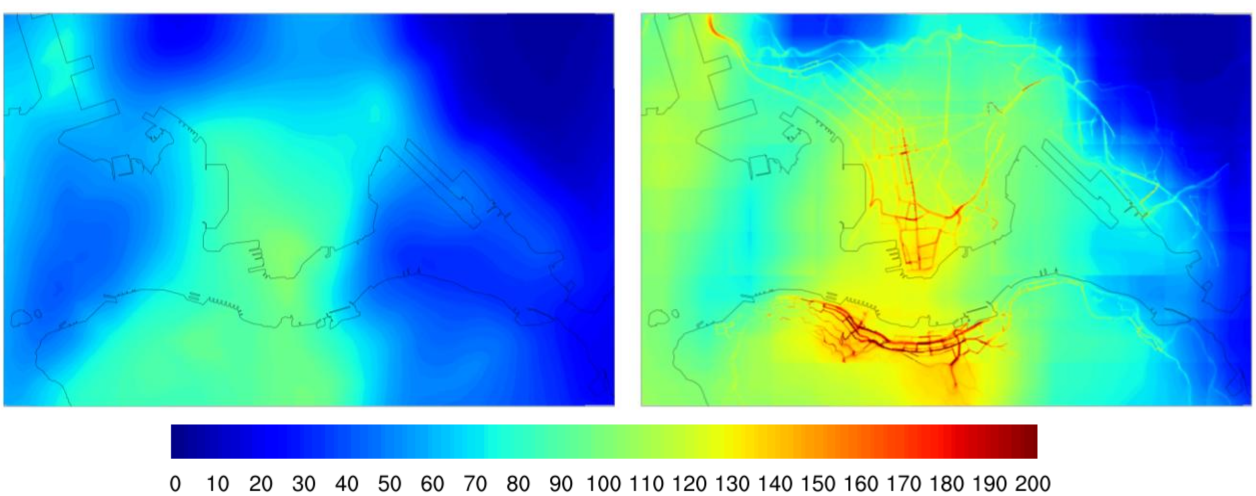


Fig. 5. Forecast NO_2 concentration contours from CMAQ (left) and ADMS-Urban RML (right) for the main urban areas of Hong Kong and Kowloon, for 3 pm local time on 13th October 2017.

An update to the forecasting system to expand the domain to cover the main urban areas of Hong Kong and Kowloon is currently in progress. Forecast concentrations from the CMAQ 1 km domain and the ADMS-Urban RML are shown for an example hour in Figure 5. It is clear that the background concentrations follow those calculated by the regional model, but the effects of explicitly modeled roads are also clearly visible.

Further work under the PRAISE-HK project will cover a thorough update of the emissions inventory used in the forecasting system. This will include the use of a 3D grid source in ADMS-Urban with automatic conversion from regional to local model and the use of time-variation profiles specific to individual explicit road sources. Once the emissions update is completed, further model verification will be carried out using all 16 permanent air quality monitoring sites in Hong Kong, as well as project-specific campaign measurements.

The public release of the mobile app is planned for late 2018. Subsequent development will incorporate real-time monitoring and traffic flow data to adjust the output concentrations to reflect current traffic and atmospheric conditions more accurately. Algorithms will also be developed to calculate indoor pollutant concentrations for typical environments derived from outdoor forecast concentrations.

8. REFERENCES

Byun, D. and K.L. Schere, 2006: Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modelling system. *Appl. Mech. Rev.* **59**, 51-77, doi: 10.1115/1.2128636.

Cheng, S. and K.-C. Lam, 2000: Synoptic typing and its application to the assessment of climate impact on concentrations of sulfur dioxide and nitrogen oxides in Hong Kong. *Atmos. Environ.*, **34**, 585-594, doi: 10.1016/j.atmosenv.2005.11.041.

ENVIRON, 2016: User's Guide to the Comprehensive Air Quality Model with Extensions Version 6.40. http://www.camx.com/files/camxusersguide_v6-40.pdf (accessed October 2017)

Goyal, P., A. T. Chan and N. Jaiswal, 2006: Statistical models for the prediction of respirable suspended particulate matter in urban cities. *Atmos. Environ.*, **40**, 2068-2077, doi: 10.1016/j.atmosenv.2005.11.041.

Grell, G. A., S. E. Peckham, R. Schmitz and S. A. McKeen, 2005: Fully coupled 'online' chemistry in the WRF model. *Atmos. Environ.* **39**, 6957 – 6976, doi: 10.1016/j.atmosenv.2005.04.027.

Lu, W.-Z., W.-J. Wang, X.-K. Wang, S.-H. Yan and J. C. Lam, 2004: Potential assessment of a neural network model with PCA/RBF approach for forecasting pollutant trends in Mong Kok urban air, Hong Kong. *Environ. Res.,* **96**, 79-87, doi: 10.1016/j.envres.2003.11.003.

Owen B., H.A. Edmunds, D.J. Carruthers and R.J. Singles, 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. *Atmos. Environ.* **34**, 397-406, doi: 10.1016/S1352-2310(99)00332-5.

Pokrovsky, O.M., R.H. Kwok and C.N. Ng, 2002: Fuzzy logic approach for description of meteorological impacts on urban air pollution species: a Hong Kong case study. *Comput. Geosci.*, **28**, 119-127, doi: 10.1016/S0098-3004(01)00020-6.

Schmidt, H., C. Derognat, R. Vautard, and M. Beekmann, 2001: A comparison of simulated and observed ozone mixing ratios for the summer of 1998 in Western Europe. *Atmos. Environ.* **35**, 6277-6297, doi: 10.1016/S1352-2310(01)00451-4.

Stocker, J., C. Hood, D. Carruthers, and C. McHugh, 2012: ADMS-Urban: developments in modeling dispersion from the city scale to the local scale. *Int. J. Environ. Pollut.*, **50**, 308-316, doi: 10.1504/IJEP.2012.051202.

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. https://www.cmascenter.org/conference/2014/abstracts/jenny_stocker_development_evaluation_2014.pdf

Vieno, M. and Coauthors, 2010: Modelling surface ozone during the 2003 heat-wave in the UK. *Atmos. Chem. Phys.* **10**, 7963-7978, doi: 10.5194/acp-10-7963-2010.