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

EVALUATION OF LOCAL AND REGIONAL AIR QUALITY FORECASTS FOR LONDON

Amy Stidworthy, Mark Jackson, Kate Johnson, David Carruthers, Jenny Stocker

Cambridge Environmental Research Consultants, UK

Abstract: This paper presents key results from the evaluation of a local-scale air quality forecasting system, *air*TEXT, alongside the CAMS regional-scale forecast. The CAMS forecast, which is used to account for the long-range transport of pollutants within *air*TEXT, is adjusted to account for the apparent bias in concentrations predicted within the south-east of England. Forecasts of the UK Daily Air Quality Index metrics for NO_2, O_3, PM_{10} and $PM_{2.5}$ in London have been assessed using the Model Evaluation Toolkit and the DELTA Tool for a five month period during 2017. For NO_2, where air pollutant concentrations are primarily a result of local emissions, the local forecast performs significantly better than the regional forecast as the steep roadside concentration gradients are resolved by ADMS-Urban, the dispersion model used within *air*TEXT. Although regional O_3 and particulate forecasts dominate the local forecasts, accounting for emissions, dispersion and chemistry at high spatial resolution means that *air*TEXT also performs better than CAMS for these pollutants.

Key words: *air quality, evaluation, forecast, airTEXT, CAMS, ADMS-Urban*

INTRODUCTION

One of the most important applications of atmospheric dispersion models is their use within air quality forecasting systems. Providing the public with advance warning of air quality episodes allows sensitive individuals to take action to alleviate their symptoms. Forecasting systems must account for long-range transport of pollutants in addition to local emissions, chemical processes and urban morphology; thus it is common practice to couple local air dispersion models with regional models to account for pollutant emissions, transport and chemistry at a range of scales.

London's *air*TEXT[1] air quality forecasting system has operated since 2007. The current implementation uses CAMS regional ensemble air quality forecast data[2] as boundary conditions for the local urban air dispersion model, ADMS-Urban (Owen *et al.*, 2000); previous versions of the system used Prev'air, a service operated by INERIS using CHIMERE. In recent years, *air*TEXT has been extended to include additional areas within south-east England (Cambridge, Colchester and Chelmsford) and a new implementation in Riga (Latvia) will be launched in 2018. A range of atmospheric pollutants are modelled within the system: nitrogen dioxide (NO_2), particulates (PM_{10}, $PM_{2.5}$) and ozone (O_3). In the UK, forecast pollutant concentration levels are assessed according to the Government's Daily Air Quality Index (DAQI, Connolly *et al.*, 2013) which is a 10-level indexing system with four air quality bands: low, moderate, high and very high.

The regional models included in the CAMS ensemble assimilate observational data. However, to best represent air quality for a particular region it is still necessary to adjust model predictions to account for historical regional model bias. This paper presents results of a model evaluation exercise where three forecasting datasets have been evaluated: *air*TEXT predictions; 'raw' CAMS ensemble model output; and 'adjusted' CAMS ensemble model output for south-east England. Forecast concentrations have been evaluated against measurements recorded at 66 stations within Greater London. The exercise has been undertaken using both the Model Evaluation Toolkit (originally developed as part of the EU FP7 PASODOBLE project, Stidworthy *et al.*, 2013) and the forecast mode within FAIRMODE's DELTA Tool (Miglietta *et al.*, 2012).

[1] http://www.airtext.info/
[2] http://atmosphere.copernicus.eu/

DATASETS FOR EVALUATION

Three forecasting datasets have been evaluated during this exercise:

- a street-scale *air*TEXT forecast;
- a 'raw' CAMS forecast; and
- an 'adjusted' CAMS forecast, used as input to *air*TEXT to represent regional air quality.

The datasets comprise forecast air quality data time series for the five months (February – July 2017) following the update of the *air*TEXT system to use CAMS regional model forecasts. The measured data used in the model evaluation exercise were obtained from the extensive London Air Quality Network. Table 1 summarises the number of measurement sites used in the evaluation exercise, categorised according to pollutant and site type, specifically: roadside and non-roadside (suburban, urban background and industrial). The data capture threshold is 75%.

Table 1. Summary of the number of measurement sites used in the evaluation exercise, by pollutant and site type

Site type	**Pollutant**			
	NO_2	**O_3**	**PM_{10}**	**$PM_{2.5}$**
Roadside	29	8	32	10
Suburban, urban background & industrial	21	8	19	6
Total	**60**	**16**	**51**	**16**

*air*TEXT is a free service for the public providing air quality alerts by SMS text message, email and voicemail and 3-day forecasts of air quality, pollen, UV and temperature across Greater London, Cambridge, Colchester and Chelmsford. The street-scale *air*TEXT air quality maps (Figure 1) are generated using the urban air quality dispersion modelling system ADMS-Urban. Separate maps showing the DAQI pollution indices for NO_2, O_3, PM_{10} and $PM_{2.5}$ in addition to an overall pollution index are displayed on the website. ADMS-Urban is a deterministic model that requires emissions for the domain of interest as input in addition to estimates of the long-range transport of pollutants advected into the model domain. When ADMS-Urban is run in forecasting mode, the long-range transport component of the system is obtained from the CAMS Regional Ensemble Forecast product.

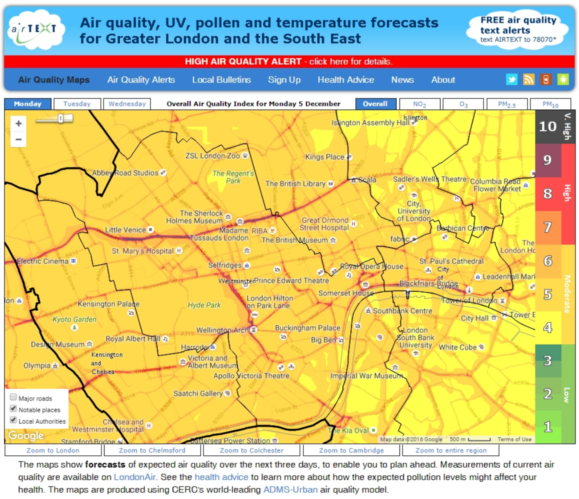

Figure 1. Example *air*TEXT map showing the overall DAQI during an air quality episode in central London (December 2016)

Table 2. Factors used to calculate the 'adjusted' CAMS forecast which is more suitable for use within southeast England, by pollutant

Pollutant	**'adjusted' CAMS concentration = $A_0 + A_1$ 'raw' CAMS concentration**	
	A_0 (μg/m³)	**A_1 (-)**
NO_2	1.40	0.77
O_3	0.22	0.89
PM_{10}	1.80	1.20
$PM_{2.5}$	3.70	1.20

The CAMS Regional Ensemble Forecast product is freely available from the EU's Copernicus Programme; the service provides hourly 96-hour forecasts of pollutants including NO_2, NO, O_3, PM_{10}, and $PM_{2.5}$ at 0.1° resolution on a domain covering all of Europe. This is a regional-scale forecast derived from an ensemble of models that use varying degrees of data assimilation (in-situ and satellite). However, model evaluation at rural monitoring sites for the south-east of England indicates that the 'raw' CAMS ensemble model forecast includes some bias within this domain. Regression analyses using historical datasets have allowed the calculation of factors that may be used to linearly adjust the 'raw' CAMS forecast to derive an 'adjusted' CAMS dataset; details are given in Table 2. In summary, for south-east England, the 'raw' CAMS NO_2 and O_3 forecasts appear to be over-predicting by approximately 30% and 12% respectively, and the 'raw' CAMS particulate forecasts are under-predicting by approximately 17%.

In addition to demonstrating the accuracy of the street-scale *air*TEXT forecasts within Greater London, the evaluation presented in this paper indicates how well the adjustments shown in Table 2 improve the skill of the CAMS forecasts for south-east of England.

EVALUATION METHODLOGY

A number of tools exist for air dispersion model evaluation, but fewer have been developed for the specific purpose of evaluating the predictions made by air quality forecasting systems. Forecasting system accuracy needs to be assessed in terms of how well the model is able to predict exceedances of air quality threshold values, although evaluation of overall model performance is also useful. For this work, two tools have been used to evaluate the three forecasting datasets:

- the Model Evaluation Toolkit; and
- the DELTA Tool[3].

Both tools generate statistics relating to the number of correct alerts, false alerts and missed alerts, and graphs that indicate, for instance, the probability of a correct forecast, the probability of detection and the false alarm ratio. The Model Evaluation Toolkit additionally produces forecast accuracy graphs and statistics relating to forecast indices (in addition to threshold statistics). The DELTA Tool has been developed to take measurement uncertainty into account within model evaluation. However, this is challenging to do when considering a series of threshold values. That is, when the difference between the alert threshold concentration and the measured concentration is less than the observation uncertainty, it is not possible to definitively say whether or not there has been an exceedance of the threshold and therefore not possible state whether or not a model forecast is correct. This issue is currently being discussed within the FAIRMODE community, but at the current time, the forecasting mode within the DELTA Tool includes options that take 'conservative' and 'cautious' approaches to the definition of an alert. Under the 'conservative' approach, if accounting for measurement uncertainty results in the possibility of a threshold exceedance, then it is assumed that an exceedance occurred; conversely under the 'cautious' approach if there is the possibility that an exceedance did not occur, then it is assumed that it did not.

RESULTS

A selection of results from the model evaluation exercise is presented in this section. Firstly, overall model performance is evaluated and secondly statistics relating to threshold exceedances are considered. Table 3 summarises the model evaluation statistics for the five-month period under consideration.

Table 3. Summary statistics for non-roadside and all sites, by pollutant and model: FAC2 – proportion of modelled values within a factor of two of the observed; NMSE – normalised mean square error; R – correlation coefficient; best value / statistic shown in bold

Pollutant (daily statistic)	Sites	Average concentration: Obs. (μg/m^3)	Modelled (μg/m^3): *air*TEXT	Modelled: 'raw' CAMS	Modelled: 'adjusted' CAMS	Model evaluation statistics: FAC2 *air*TEXT	FAC2 'raw' CAMS	FAC2 'adjusted' CAMS	NMSE *air*TEXT	NMSE 'raw' CAMS	NMSE 'adjusted' CAMS	R *air*TEXT	R 'raw' CAMS	R 'adjusted' CAMS
NO_2 (max 1-hour)	non-road	56.3	**61.7**	39.6	31.9	**0.85**	0.78	0.65	**0.23**	0.46	0.74	**0.54**	0.43	0.43
	all	73.6	**77.3**	39.8	32.1	**0.87**	0.56	0.45	**0.20**	0.82	1.23	**0.63**	0.38	0.38
O_3 (max 8-hour rolling)	non-road	62.4	**58.1**	75.0	66.9	**0.95**	0.93	**0.95**	0.08	0.08	**0.06**	0.69	**0.73**	**0.73**
	all	53.5	**54.1**	74.6	66.7	**0.92**	0.79	0.85	**0.09**	0.19	0.13	**0.69**	0.62	0.62
PM_{10} (average)	non-road	21.3	**19.3**	13.7	18.2	**0.92**	0.76	0.89	**0.19**	0.48	0.24	**0.63**	0.56	0.56
	all	22.8	**22.5**	13.7	18.3	**0.93**	0.71	0.89	**0.16**	0.52	0.24	**0.62**	0.59	0.59
$PM_{2.5}$ (average)	non-road	13.9	**14.4**	9.3	14.9	**0.97**	0.83	0.96	**0.14**	0.40	0.14	**0.80**	**0.80**	**0.80**
	all	13.9	16.0	9.3	**14.9**	0.92	0.82	**0.93**	0.17	0.41	**0.15**	0.77	**0.80**	**0.80**

[3] Note that the forecast mode of the DELTA Tool is still undergoing development

The CAMS forecasts are regional scale with a resolution of approximately 10 km and as such they are not at sufficiently high resolution to predict concentrations at roadside sites. Consequently, it is helpful to evaluate all non-roadside sites together in order to assess how well the forecast performs at urban background, suburban and industrial sites; results are also presented for all sites together. Statistics relating to daily maximum hourly NO_2, 8-hour rolling average O_3 and average PM_{10} and $PM_{2.5}$ are presented; average concentrations are given in addition to the proportion of values within a factor of two of the observed (FAC2), normalised mean square error (NMSE) and the Pearson's correlation coefficient (R).

Figure 2 shows the pollutant time series for the five-month period considered averaged over all stations. The observed concentrations are shown in black, *air*TEXT predictions in green and CAMS 'raw' and 'adjusted' in light and dark blue respectively. The alert threshold concentrations are shown on the figures, with the moderate, high and very high DAQI thresholds being indicated by orange, red and purple dashed lines respectively. The number of monitoring stations is also indicated on these figures using a grey line.

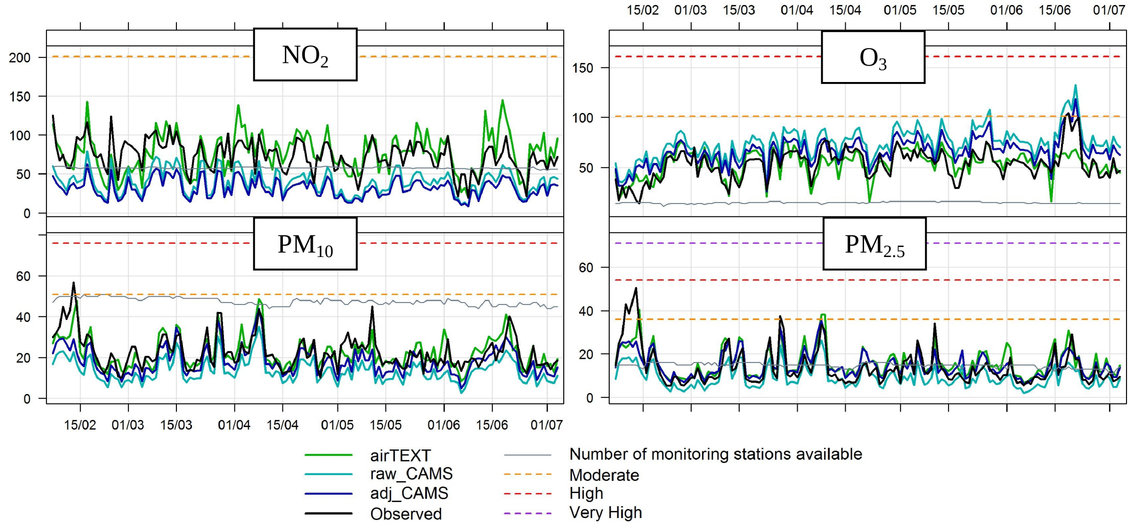

Figure 2. Pollutant time series for five-month period

When run in forecast mode, the DELTA Tool generates a number of plots and statistics that account for measurement uncertainty. Figure 3 presents a graphical summary of PM_{10} moderate alert statistics relating to all sites when no measurement uncertainty is accounted for, alongside values that allow for measurement uncertainty, using both the 'conservative' and 'cautious' approach to the definition of an exceedance of the alert threshold. Currently $PM_{2.5}$ is not an option in the DELTA Tool forecast mode.

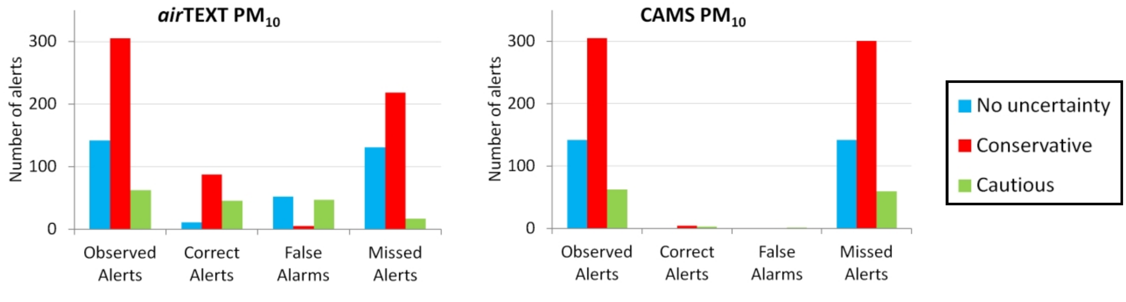

Figure 3. Number of observed, correct, false and missed PM_{10} moderate alerts for *air*TEXT (left) and CAMS (right) with ('Conservative' & 'Cautious') and without ('No uncertainty') accounting for measurement uncertainty; data for all sites presented

DISCUSSION

No major air pollution episodes occurred during this period in the south-east of England so it is difficult to draw firm conclusions regarding the overall skill of the forecasting systems in predicting air quality episodes. However, the evaluation exercise has provided an example of how well the models are able to predict concentrations in terms of the health-related DAQI concentration metrics and exceedances in terms of the DAQI alert thresholds (Table 1, Figure 2, Figure 3). Specifically, *air*TEXT performs better than the regional-scale CAMS forecasts for all pollutants considered.

NO_2 concentrations are much better predicted by *air*TEXT than by either version of CAMS (Table 1, Figure 2). This is unsurprising because the main source of NO_x concentrations in urban areas is road traffic. Roads need to be modelled at high resolution to account for steep emissions gradients and fast NO_x chemistry. CAMS significantly under-predicts NO_2 at all site types. Both observed and modelled NO_2 concentrations lie well below the DAQI 'Moderate' alert threshold during the evaluation period.

O_3 being solely a secondary pollutant is strongly influenced by long-range transport, thus O_3 *air*TEXT predictions are strongly dependent on CAMS. However, in the vicinity of busy roads, O_3 titration occurs, which the regional-scale forecast is unable to resolve and this explains why *air*TEXT performs better than CAMS when all sites are considered (Table 2). When only non-road sites are considered, the 'adjusted' version of CAMS performs similarly to *air*TEXT; the 'raw' CAMS model has a tendency to over-predict O_3. There is one short O_3 episode towards the end of the evaluation period (Figure 2) which occurred in hot conditions; the 'adjusted' CAMS represents this well, but the *air*TEXT predictions are low showing that local O_3 titration by NO within the city is not being sufficiently offset by local generation of ozone occurring in the prevailing hot conditions. A likely explanation is that the current implementation of *air*TEXT does not consider the large increases in biogenic emissions occurring in hot conditions; in addition the 'Generic Reaction Set' chemistry scheme (Azzi *et al.*, 1992) included within ADMS-Urban may underestimate the effective VOC reactivity in such conditions.

PM_{10} concentrations are also strongly influenced by long-range transport so the PM_{10} *air*TEXT prediction is dependent on the CAMS forecast. The 'raw' CAMS forecast under-predicts but the calculated adjustment leads to good performance for *air*TEXT when all sites are considered. Non-road site concentrations are under-predicted, indicating that there may be an issue with the balance between road and non-road emissions relating to the resuspension of dust from various sources, for instance construction activities. $PM_{2.5}$ concentrations are influenced both by long-range transport and local emissions sources. The 'raw' CAMS forecasts are lower than measured values, but the current 'adjusted' CAMS forecast is a slight over-prediction, which leads to a small over-prediction of *air*TEXT. There is one particulate concentration episode at the beginning of the evaluation period (Figure 2). Particulate episodes are usually associated with easterly, low wind speed conditions which result in high particulate concentrations being transported from mainland Europe in addition to the accumulation of particulates emitted from local sources; this usually occurs in springtime. CAMS missed the episode leading to an under-prediction by *air*TEXT, although the local forecast predicted some high PM_{10} values. Analysis of rural measurements recorded upwind of the urban area during this episode would indicate whether the episode was primarily driven by regional or local emissions and dispersion. When measurement uncertainty is accounted for (Figure 3), *air*TEXT correctly predicts 29% and 73% of alerts, using 'conservative' and 'cautious' approaches respectively.

REFERENCES

Azzi, M., Johnson, G.M. and Cope, M., 1992: An introduction to the generic reaction set photochemical smog mechanism. *Proceedings of the 11th International Clean Air and Environment Conference.*

Connolly, E., Fuller, G., Baker, T. and Willis, P. 2013: Update on Implementation of the Daily Air Quality Index. Department for Environment, Food and Rural Affairs,

Miglietta, M.M., Thunis, P., Georgieva, E., Pederzoli, A., Bessagnet, B., Terrenoire, E. and Colette, A., 2012: Evaluation of WRF model performance in different European regions with the DELTA-FAIRMODE evaluation tool. *International Journal of Environment & Pollution*, **50(1-4)**, 83-97.

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.

Stidworthy, A., Carruthers, D., Stocker, J., Balis, D., Katragkou, E. and Kukkonen, J., 2013: MyAir Toolkit for Model Evaluation. 15th International Conference on Harmonisation, Madrid, Spain.