Emulation and Sensitivity Analysis of the Community Multiscale Air Quality Model for a UK Ozone Pollution Episode

Andrew V. Beddows,* Nutthida Kitwiroon, Martin L. Williams, and Sean D. Beevers

King’s College London, Waterloo, London, SE1 8WA, United Kingdom

Supporting Information

ABSTRACT: Gaussian process emulation techniques have been used with the Community Multiscale Air Quality model, simulating the effects of input uncertainties on ozone and NO2 output, to allow robust global sensitivity analysis (SA). A screening process ranked the effect of perturbations in 223 inputs, isolating the 30 most influential from emissions, boundary conditions (BCs), and reaction rates. Community Multiscale Air Quality (CMAQ) simulations of a July 2006 ozone pollution episode in the UK were made with input values for these variables plus ozone dry deposition velocity chosen according to a 576 point Latin hypercube design. Emulators trained on the output of these runs were used in variance-based SA of the model output to input uncertainties. Performing these analyses for every hour of a 21 day period spanning the episode and several days on either side allowed the results to be presented as a time series of sensitivity coefficients, showing how the influence of different input uncertainties changed during the episode. This is one of the most complex models to which these methods have been applied, and here, they reveal detailed spatiotemporal patterns of model sensitivities, with NO and isoprene emissions, NO2 photolysis, ozone BCs, and deposition velocity being among the most influential input uncertainties.

INTRODUCTION

The adverse health effects of elevated ozone and NO2 concentrations are well documented. Consequently, legislation requires that concentrations of pollutants are measured at fixed site monitoring stations, and models such as the Community Multiscale Air Quality (CMAQ) model are used to estimate pollutant concentrations between sparsely distributed monitoring locations. The output of such models is subject to uncertainty, caused both by the model’s imperfect representation of reality and by uncertainties in inputs which are propagated through to the outputs. This work concerns methods to ascertain the relative contributions of different input uncertainties to the total output uncertainty, revealing where efforts to improve the accuracy of input data will yield the greatest improvement in model performance.

The response of CMAQ output to input variability has been widely studied, for example, to assess the effect of changing emissions scenarios, the effect of changing emissions in combination with different chemical mechanisms, and the effect of varying ozone deposition to vegetation. The validity of such studies may be assessed by evaluating the model’s ability to reproduce known real-world behavior in a dynamic evaluation. A closely related task is to change the components of the modeling system or its inputs, for example, the meteorological model, vertical resolution, boundary conditions (BCs), and chemical mechanism, in order to find the best configuration for prediction in a particular domain.

The above examples can be described as informal sensitivity analysis (SA) methods. An alternative approach is to compute the “local” sensitivity to small variations in the baseline value of an input, equivalent to taking a partial derivative of the output with respect to that input. Such local sensitivity coefficients can be computed directly with a set of coupled auxiliary equations or separately as the Decoupled Direct Method (DDM). Extending this to compute second derivatives takes into account nonlinearity in the model sensitivity, and “cross sensitivities” may be calculated to two or more inputs perturbed together. Cohan et al. applied this to CMAQ version 4.3, extrapolating further from the base case model run by using first and second derivatives in a second order Taylor series expansion, thus extending direct methods into somewhat more than local SA. In the context of sensitivity to input uncertainty, however, it is more informative to perform a “global” SA, one in which all inputs are perturbed simultaneously but independently over their full uncertainty ranges.

The prohibitively large number of runs required for the global SA, and related tasks like Monte Carlo uncertainty analysis, of models with many uncertain inputs means that some kind of meta-modeling technique is needed to produce a quick to run model surrogate. This idea is not new; Fang et al. describe a number of methods for producing meta-models, including polynomial regression, splines, kriging, and neural networks. Other methods which have been used in moderately...
complex applications include “Stochastic Response Surface Methods”\textsuperscript{19} and “High Dimensional Model Representations”.\textsuperscript{19} All of these methods require the computer model to be run a number of times at different input settings to provide training data to build the meta-model. The Taylor series expansions of DDM sensitivity coefficients mentioned above have been used as meta-models of CMAQ\textsuperscript{20,21} but differ in that only one or two runs are required. This, however, inherently means that considerable extrapolations are involved in their construction.

Kriging metamodels of deterministic computer models were first proposed by Sacks et al.\textsuperscript{22} who observed that the Kriging predictor can be made to exactly interpolate the output of the training runs, mirroring the way the computer model produces identical output if run again at the same input settings. This “Gaussian process emulation” method was developed further in a Bayesian framework, with the aim of incorporating observational data to calibrate the model input factors by Kennedy and O’Hagan.\textsuperscript{23} The same statistical framework can be used for global sensitivity analysis\textsuperscript{24} and, in theory, allows properties of the model output uncertainty distribution to be derived analytically, but Oakley and O’Hagan\textsuperscript{25} concede that this is best done by using the emulator as a model surrogate in a Monte Carlo analysis.

Most examples of the use of emulation in the environmental sciences are found in climate modeling, with a number of case studies described by Kennedy et al.\textsuperscript{26} Gaussian process emulators of several climate models have been used in a history matching process with observational data,\textsuperscript{27–29} which calibrates model inputs in a conservative manner to reduce their uncertainty ranges, and Sexton et al.\textsuperscript{30} used emulators to attempt the more ambitious task of producing calibrated probabilistic climate projections.

When perturbing inputs simultaneously in a global SA, the sensitivity to each can be quantified by estimating how much the associated output variance would be reduced if that input were fixed.\textsuperscript{13} Various techniques for calculating this are described by Saltelli et al.,\textsuperscript{13} all requiring considerable numbers of model runs, so the use of emulators provides obvious benefits. One such method, the Fourier Amplitude Sensitivity Test (FAST), was used by Lee et al.\textsuperscript{31} with a global aerosol model. Modeled concentration of cloud condensation nuclei was emulated with respect to eight uncertain parameters with a separate emulator and FAST for each grid cell. Lee et al.\textsuperscript{32} extended this study to include 28 input variables, this time also using the emulators for Monte Carlo uncertainty estimation. The work presented here is similar to that carried out by Lee et al.,\textsuperscript{32} in that emulation is used in conjunction with FAST to allow global SA, but differs in that a screening method was used prior to emulation in order to select which of the hundreds of uncertain inputs in the CMAQ model would be included in the analysis.

\section*{Methods}

\header{Base-Case Model Run.} July 2006 was modeled with CMAQ 4.7.1, using the Carbon Bond 5 (CB-05) chemical mechanism.\textsuperscript{33} Newer CMAQ releases exist, but ozone chemistry is largely unchanged and the CB-05 mechanism is still widely used. A 10 day spin-up allowed the SA to be performed for 21 days covering a significant ozone pollution episode and several days on either side. This was a typical UK summer pollution episode with hot dry weather and anticyclonic conditions bringing air masses from central Europe. Figure 1 shows the outer spatial domain with a grid size of 81 km, an intermediate domain with a grid size of 27 km, and an inner domain producing the output presented here, covering the UK with a grid size of 9 km.

The model was used in off-line mode, with the Weather Research and Forecasting model\textsuperscript{34} providing meteorological inputs. BCs for the outer domain were derived from data provided as part of the Air Quality Model Evaluation International Initiative,\textsuperscript{35} emissions for the outer and middle domains from the European Monitoring and Evaluation Programme (EMEP, http://www.ceip.at/), and emissions for the inner domain from the UK National Atmospheric Emissions Inventory (NAEI).\textsuperscript{36} A full description of this base-case run, along with a comprehensive performance evaluation, is contained in the DEFRA regional model evaluation report.\textsuperscript{37}

\header{Input Variable Screening.} Typically, in complex models, only a small subset of inputs have significant influence on a particular output,\textsuperscript{13} so screening the inputs to identify this subset makes subsequent analyses more tractable. Following the method described by Morris,\textsuperscript{38} the inputs were perturbed one at a time within a range of one-half to double their baseline values, with perturbations of magnitude 0.75 times that of the baseline value. This was repeated several times for each input with a different starting point on each occasion, and the remainder of the inputs were held at different constant values on each occasion, producing a number of “elementary effects” on the model output for each input, the means and standard deviations of which together indicate the overall importance and linearity/monotonicity of that input. More detail on this method is given in pages S2–S3 of the Supporting Information. For the emissions inputs, which take the form of spatiotemporally varying fields, the whole field was perturbed by the same amount in order to keep the analysis tractable. Reaction rate constants were perturbed by changing the pre-exponential factor, allowing normal spatiotemporal variation due to temperature fluctuations across the domain, and photolysis rates were perturbed before the model modified them according to cloud cover.

\begin{figure}[ht]
\centering
\includegraphics[width=\textwidth]{Figure1.png}
\caption{81 km (green), 27 km (red), and 9 km (blue) grid cell domains used in the CMAQ model run.}
\end{figure}
A useful addition to the method, suggested by Campolongo et al., uses the mean of the absolute values of the elementary effects, $\mu_i$, as a single metric to rank the inputs in order of importance. In the same paper, Campolongo et al. enhance the method to improve its space filling properties. This space filling method is implemented in the function morris of the package sensitivity, in the R statistical computing language, and is used in the screening exercises described here.

It was expected that the results of this screening would differ across the domain, which contains 9360 grid cells, but performing the screening for all of these cells would have been prohibitive; so, a subset was chosen comprising the 22 grid cells containing the monitoring sites used in the DEFRA regional model evaluation. A total of 223 inputs were screened: 40 gas-phase species from the BCs and 27 from the emissions and the 156 reactions from the CB-05 core chemical mechanism. This was performed at 4 p.m. and midnight, on both the 10th of May and 19th of July (during the same study, some SA was also performed for May 2006) with a ten day spin-up. In this total of 188 screening exercises, the highest $\mu_i$ value attained by each input factor was used to discard those factors which fell below a threshold of 2 ppb influence on both ozone and NO$_2$ output. In such a screening exercise, it is fortuitous if a natural partitioning appears between influential and noninfluential inputs. Unfortunately, this was not the case, and the 2 ppb threshold was chosen somewhat arbitrarily, although it turned out to be justified, as we show in the Results section. At the end of this process, 12 reactions were retained, along with CO and ozone BCs, and eight emissions variables. In the analyses which follow, the emissions inputs for the 9 km domain were perturbed separately from those for the 81 and 27 km domains, and ozone deposition velocity was added to make a total of 31 variables. They are listed in the section describing the FAST methods, where their uncertainty distributions are also assigned.

**Gaussian Process Emulation.** Emulators were constructed using the R DiceKriging package, described by Roustant et al., who also give advice on its use. A great deal of information is also available in the references cited in the Introduction to this paper and in pages S4–S8 of the Supporting Information, which include brief mathematical details. The package was used with default settings which implement universal Kriging with a Matérn 5/2 covariance function.

The input values at which to make the CMAQ training runs were chosen using a Latin hypercube sampling (LHS) plan. LHS is a stratified sampling technique, having desirable properties for Monte Carlo uncertainty analysis, but here is simply used to efficiently sample the input variable space without introducing correlations between inputs. This provided factors by which to multiply input variables, sampling each from one-half to double its baseline value. This range was larger than the ranges of the input uncertainties to be used in the sensitivity analyses and so provided scope to widen them or to perform other experiments as required. A separate emulator was made for ozone or NO$_2$ concentration at every location and time step used in the analyses which follow.

Emulation gives a point prediction of the model output at untried input values and a variance which describes the uncertainty in the emulation. The aim was to construct emulators which were accurate enough to allow the point predictions alone to be used with confidence in place of CMAQ. All of the results presented use emulators with 31 input variables, made using Latin hypercubes with 576 points, equating to 18.6 runs per variable. Accuracy of emulation was evaluated with leave-one-out cross validation (LOOCV), where the output of each of the training runs is predicted using an emulator trained on the remaining runs. The mean absolute error (MAE) of the predictions was calculated as follows,

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i|$$  \hspace{1cm} (1)

where $y_i$ is the output of the left-out model run and $\hat{y}_i$ is the value of that output predicted by the emulator. This gives an indication of the average amount by which an emulator will be in error when predicting CMAQ output at untried input values. The MAEs for emulators of ozone and NO$_2$ concentrations were calculated for 125 randomly chosen time steps and locations. The distributions of these MAE statistics are shown in Figure S2. The median MAE for ozone emulators is around 1.3 ppb, and three-quarters have MAEs below 1.9 ppb, while for NO$_2$ over three-quarters of the emulators have a MAE of below 0.5 ppb. These results were considered adequate to proceed with using the emulators for the analyses which follow.

**FAST.** A variance-based global SA entails decomposing the model output variance into a series of terms of increasing dimensionality and dividing each of those terms by the total variance. Such a series for a model with $k$ inputs would have $k$ terms giving the first-order effects, or “main effects”, for each input, the next $\binom{k}{2}$ terms giving the second-order effects, those involving combinations of two inputs, and so on. Each of these effects can be thought of as the amount by which the output variance would be reduced if the inputs involved in that term were fixed. The number of terms in this decomposition is $2^k - 1$ and hence becomes prohibitive to calculate. Instead, a “total effect” index may be used, which is the sum of all the terms in the series involving a particular factor, equating to the main effect plus interactions. A clear exposition of the mathematical details can be found in Saltelli et al. and further information is also given in pages S8–S10 of the Supporting Information.

The Fourier Amplitude Sensitivity Test (FAST) provides a computationally efficient procedure for estimating the main effects by relating the probability distribution representing the uncertainty in each input to a different frequency. The input variable space is then explored by a search curve which traverses each dimension at a rate proportional to the frequency assigned to the corresponding input. This curve defines a set of values of the inputs at which to run the model, which induce a periodicity in the output. Fourier analysis of this output then gives Fourier coefficients from which the main effect for each factor can be derived.

The extended FAST estimates main and total effects indices at the same time by choosing two frequencies for each input, one of which is assigned to that input, as in the classic FAST, and the other to all the remaining inputs. This induces two periodicities in the model output, the first of which can be used to calculate the main effect, and the second to calculate the total effect. More detail is given on pages S11–S12 of the Supporting Information, and the extended FAST is implemented in the function fast99 of the R package sensitivity.

In order for these indices to correctly represent the proportion of output variance attributable to each input, those inputs should be independent, and while an analysis which takes correlations between them into account is theoretically possible, the extra complexity involved means that a first SA is generally performed with the assumption of
independent inputs.\textsuperscript{48} While there is a likelihood that some of the CMAQ inputs are not independent, for example, NO and NO\textsubscript{2} emissions from road transport, this does not mean that there are easily quantifiable correlations between the uncertainties in those inputs which could be included in an analysis.

The variables retained following the screening process are given in Table 1, along with their uncertainty distributions, which are centered on 1.0, so that values sampled from them equate to a factor by which to multiply the input. The reactions are labeled following the CB05 definition,\textsuperscript{33} and rate uncertainties are taken from the NASA/Jet Propulsion Laboratory Atmospheric Chemical Data Evaluation.\textsuperscript{49} Emissions uncertainties are from a NAEI uncertainty estimation\textsuperscript{50} which gives two estimates created using different methods. The larger estimate of ±30\% for NO\textsubscript{x} and VOCs is used here, as there are major assumptions in the smaller estimate of ±7−10\%, such as the conjecture that official statistics are “subject to very limited uncertainty” and that missing sources of emissions would not contribute significantly to the total. Each of the listed emissions species was treated as a separate input for the 9 km domain (labeled “UK” in the results) and as another input for the 81 and 27 km domains taken together (labeled “EU” in the results). In the absence of available information, the same level of uncertainty was assigned to BCs and to ozone deposition velocity as to emissions.

These uncertainty distributions were used in a separate FAST, with a separate emulator of either ozone or NO\textsubscript{2} output, for each of the 504 hourly time steps in the 21 days from the 11th to the 31st of July 2006 at a number of locations. Brief descriptions of these locations are given in pages S13−S15 of the Supporting Information.

### RESULTS

The base case modeled and measured ozone concentrations\textsuperscript{51} at the grid squares containing the Harwell monitoring site in Oxfordshire and the North Kensington site in Greater London for the 11th to the 31st of July are shown in Figure 2. Overall, the model captures the observed behavior quite well but fails to reproduce some of the afternoon peaks in both magnitude and timing. The model performance appears slightly better at Harwell than London, possibly due to the fact that Harwell is a rural background monitoring location, sited so as to be representative of a large area, and so more comparable to the CMAQ 9 km grid squares than an urban monitoring site.

Figure 3 shows total effects at the above locations as time series of stacked bars, as suggested by Saltelli et al.,\textsuperscript{52} who

---

**Table 1. Distributions Used to Characterize Input Uncertainties**

<table>
<thead>
<tr>
<th>reaction</th>
<th>distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1: NO\textsubscript{2} photolysis</td>
<td>normal, sd = 0.2</td>
</tr>
<tr>
<td>R3: O\textsubscript{3} + NO</td>
<td>normal, sd = 0.1</td>
</tr>
<tr>
<td>R7: NO\textsubscript{2} + O\textsubscript{3}</td>
<td>normal, sd = 0.15</td>
</tr>
<tr>
<td>R9: O\textsubscript{3} photolysis</td>
<td>normal, sd = 0.3</td>
</tr>
<tr>
<td>R10: O1D + M</td>
<td>normal, sd = 0.5</td>
</tr>
<tr>
<td>R11: O1D + H\textsubscript{2}O</td>
<td>normal, sd = 0.08</td>
</tr>
<tr>
<td>R28: NO\textsubscript{2} + OH</td>
<td>normal, sd = 0.4</td>
</tr>
<tr>
<td>R30: HO\textsubscript{2} + NO</td>
<td>normal, sd = 0.3</td>
</tr>
<tr>
<td>R66: OH + CH\textsubscript{4}</td>
<td>normal, sd = 0.1</td>
</tr>
<tr>
<td>R74: HCHO photolysis</td>
<td>normal, sd = 0.4</td>
</tr>
<tr>
<td>R87: C\textsubscript{2}O\textsubscript{3} + NO</td>
<td>normal, sd = 0.5</td>
</tr>
<tr>
<td>R112: PAR + NO</td>
<td>normal, sd = 0.3</td>
</tr>
<tr>
<td>NO emissions</td>
<td>uniform, ±0.3</td>
</tr>
<tr>
<td>NO\textsubscript{2} emissions</td>
<td>uniform, ±0.3</td>
</tr>
<tr>
<td>ISOP emissions</td>
<td>uniform, ±0.3</td>
</tr>
<tr>
<td>PAR emissions</td>
<td>uniform, ±0.3</td>
</tr>
<tr>
<td>XYL emissions</td>
<td>uniform, ±0.3</td>
</tr>
<tr>
<td>ETH emissions</td>
<td>uniform, ±0.3</td>
</tr>
<tr>
<td>CO emissions</td>
<td>uniform, ±0.3</td>
</tr>
<tr>
<td>OLE emissions</td>
<td>uniform, ±0.3</td>
</tr>
<tr>
<td>O3 BCs</td>
<td>uniform, ±0.3</td>
</tr>
<tr>
<td>CO BCs</td>
<td>uniform, ±0.3</td>
</tr>
<tr>
<td>O3 deposition velocity</td>
<td>uniform, ±0.3</td>
</tr>
</tbody>
</table>

“All centered on 1.0.

---

![Figure 2](image_url) Measured and modeled ozone concentrations at (a) Harwell and (b) London, the 11th to the 31st of July 2006.
normalize the sensitivity indices so the height of the bars represent a true fraction of the total variance. Here, the indices are not normalized, meaning that the total height of the bars gives an indication of the size of the input interaction effects, as the effect of an interaction between several inputs is included in the total effect of each of them.

The FASTs were performed with all 31 variables identified in the screening process, but only those accounting for more than 1% of the variance at any time step are shown on the time series plots. For example, CO boundary conditions were retained after screening but accounted for less than 1% of the variance in the FASTs. These plots and those for five additional locations in the Supporting Information have between 16 and 18 variables, demonstrating that the screening threshold was not too high; it is better to allow some variables through the screening, which later turn out to be unimportant, than to set the threshold too high and risk ruling out potentially important inputs.

Of particular note on the plot for Harwell in Figure 3a is the dominance of ozone BCs, shown in blue, at the start and end of the period, when wind speeds were relatively high in comparison to those in the middle of the period when deposition velocity, in green, becomes more important. Also, as the weather became warmer in this middle period chemical processes start to dominate; BCs are important when there is little photochemical activity and vice versa. Diurnal variation in sensitivity to NO\textsubscript{2} photolysis, in yellow at the bottom of the plot, is clearly evident. It is important to note that this does not go to zero, even at night, because although the value of the total effect gives the sensitivity at just that particular time step, this sensitivity is to uncertainty in the rate for the whole period, indicating that modeled ozone concentration at night is still sensitive to photolysis occurring earlier in the day.

Figure 3a shows the results of the SA for Harwell. In comparison to Harwell, uncertainty in NO emissions is much more important, and these are likely to be dominated by local traffic emissions. What is also clear is that deposition velocity is less important in this urban location than in the rural background at Harwell. Dry deposition to vegetation is an important ozone sink in the CMAQ model, thus, an urban grid square with less vegetation than a rural area has a smaller absolute amount of deposition, and uncertainty in the deposition velocity of this smaller amount is less influential on the model output.

The only VOC emission appearing on these plots (and hence the only one accounting for more than 1% of the variance) is isoprene, of which biogenic emissions increase markedly in the kind of hot weather conditions experienced during this episode. Isoprene emissions are widespread across rural areas of the UK, and isoprene has strong ozone creation potential. Ethene, xylene, and carbon bond species, of which anthropogenic emissions comprise a large part, did come through the screening process but were not significant in the final analysis, providing an illustration of the nature of global SA. It does not necessarily mean that the model is insensitive to these inputs in isolation, but when many inputs are perturbed at the same time to which the model is more sensitive, their influence is effectively “drowned out”, indicating that the accuracy of those other inputs is more important in determining the accuracy of the model output.

A further noteworthy feature in Figure 3b is the more “spiky” nature of the top surface of the plot, indicating that interactions between input uncertainties become important for short, isolated periods of time. Closer inspection reveals that these points are shortly before midnight in every case, and given that this would also be shortly after sunset at this time of year, these interactions could be related to the transition from the daytime photochemical regime to night time chemistry. One of these periods is examined in more detail in Figure 4, where the overall height of the bars represents the total effect as the sum

Figure 3. Time series of total effects on modeled ozone concentrations at (a) Harwell and (b) London, the 11th to the 31st of July.
of the main effect and interactions. Uncertainty in UK NO emissions can be seen to be dominant, and the high degree of interaction with the other factors is clearly evident. For some inputs, NO$_2$ photolysis and ozone BCs, for example, the main effect is a small proportion of the total effect. This means that uncertainty in that input by itself is not causing a great deal of uncertainty in the model output; it is only when there is also uncertainty in other inputs with which it interacts that it is able to induce considerable uncertainty in modeled ozone concentration.

Five other locations are shown in the Supporting Information, and the progression through the Figures S4 to S8 is from urban to gradually more rural locations, finishing at Strath Vaich, a remote site in Northern Scotland. Here, uncertainty in BCs is clearly important, and emissions have little influence on the model output. The quenching of O(1D), shown in red, is an influential input, as is ozone photolysis immediately below and in combination with NO$_2$ photolysis; it can be seen that the chemistry normally associated with the tropospheric background is driving the model behavior at this location. These inputs are also influential at the next most remote location, Yarner wood in the Dartmoor national park, and moving through the other less remote rural sites and then into Manchester Picadilly and London, the transition from the tropospheric background regime to the urban regime is apparent.

Nitrogen Dioxide. Plots showing FAST time series for NO$_2$ output at London, Harwell, and Strath Vaich are shown in Figures S9–S11. In all cases, the analyses are dominated by UK NO emissions and appear less interesting than the ozone results, but there are a few points worthy of note. Again, Strath Vaich is greatly influenced by the chemistry we would expect in the tropospheric background, and emissions become more important moving toward urban areas. The influence of NO$_2$ emissions is generally small but is largest in London plot and nonexistent at Strath Vaich. The emissions inputs only have significant amounts of NO$_2$ in major conurbations and transport routes, and the speed of the reactions which cycle NO, NO$_2$, and ozone is such that those emissions rapidly decline in influence with distance from the source. In common with the ozone plot for London in Figure 3b, the London NO$_2$ plot shows more interactions than the plots for the other areas but this time with a less strict temporal pattern. There are some similarities, however, with both plots showing peaks in interactions shortly before midnight on the 11th and 12th which are associated with sharp peaks in the influence of reaction rates.

**DISCUSSION**

This analysis has revealed a complex spatiotemporal pattern of the sensitivity of modeled ozone concentrations to input uncertainties, showing that the most effective strategy for improving prediction accuracy varies greatly by time and place. Rural background predictions may be improved for significant periods of time with more accurate BCs; but, during ozone episodes, particularly in urban areas, this will have little effect, and improved NO emissions and reaction rate data will be of greater benefit. Photolysis reactions are among the most important; thus, accurate modeling of solar radiation and cloud cover will also help to improve predictions.

Studies for US domains using the DDM to calculate model sensitivities have also shown that CMAQ ozone concentrations are more sensitive to emissions of NO$_x$ than VOCs$^{16,21}$ Also using DDM-based methods in a US domain, Napelenok et al.$^{56}$ find as we do that ozone sensitivity to NO$_x$ and VOC emissions and ozone BCs has high temporal variability throughout their study period. They find greater sensitivity to VOC emissions than we do, despite using the same chemical mechanism. This could be due to differences in absolute values of emissions in different inventories but also may be attributable to the fact that the global SA perturbs many more variables at the same time; the VOC sensitivity is effectively drowned out by the fact that the model is more sensitive to so many other variables. Studies of sensitivity to the modeled chemistry have tended to focus on switching chemical mechanisms, for example,$^{11,57}$ rather than perturbing individual reaction rates, so are not comparable to the results presented here. The key difference in the methods used here to other SA methods used in air quality modeling is that perturbing many inputs simultaneously gives a more reliable estimate of which has the most influence on the model output at any given time.

The NO$_2$ analysis is dominated by uncertainty in NO emissions, except for remote locations where NO$_2$ pollution is not generally a problem, so improving the accuracy of emissions inventories would appear to be the only way of significantly improving NO$_2$ predictions, at least when only the variables which have been included in this analysis are considered. However, it is likely that errors in meteorology play a part in driving uncertainty here and are also a significant driver of ozone prediction errors.

Meteorological inputs have not been included and to do so would represent significant progress. This presents specific difficulties, however, such as violating conservation of mass between the model grid cells when perturbing variables such as wind speed and direction. It has been suggested that varying wind speed uniformly across the whole domain can alleviate this problem.$^{38}$ However, there are other variables which are deterministically linked to wind speed, such as planetary boundary layer height, so simply changing the wind speed without a means of modifying those variables accordingly would lead to modeling a physically unrealistic scenario.
An alternative proposition would be to remove meteorological variability from the analysis to some degree by grouping the results at each location by back trajectory and general synoptic conditions and only including time steps when modeled meteorology was in good agreement with observations. It may also be useful to group some inputs, such as anthropogenic VOCs, to make the results easier to interpret and more policy relevant. This analysis could not have been achieved without the use of emulation as CMAQ is simply too computationally expensive to allow sufficient numbers of runs to be made; over 2.5 million emulated “runs” were required to produce each of the time series plots in the Results and Supporting Information. The model is one of the most complex of those which have been emulated in the literature, and the emulators are among the largest in terms of number of inputs. Despite this, the LOOCV demonstrates good accuracy in emulation, showing that global SA of complex air quality models is a feasible proposition. Readers interested in using these techniques in their own work should consult references cited earlier, in particular Campolongo et al., Pujol et al., Roustant et al., and Saltelli et al. The extra information on the methods which can be found in pages S2–S11 of the Supporting Information is intended as a readable entry point into this literature.

**ASSOCIATED CONTENT**

2 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.est.6b05873.

Additional details of the methods; descriptions of locations referred to in the Article; time series plots for NO3, SA; additional locations for ozone (PDF)

**AUTHOR INFORMATION**

Corresponding Author

*E-mail: andrew.beddows@kcl.ac.uk; phone: +44 (0)20 7848 4009; fax: +44 (0)20 7848 4047.

ORCID

Andrew V. Beddows: 0000-0002-7455-4087

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

This work was supported by a MRC-PHE Centre for Environment and Health PhD studentship awarded to the first author who also thanks Dr. Marta Blangiardo of Imperial College London for advice during the course of the project.

**REFERENCES**


(37) DEFRA Phase 2 Regional Model Evaluation; King’s College London: London, UK, 2013; http://uk-air.defra.gov.uk/library.


(51) DEFRA. UK-Air Data Archive; Available at: http://uk-air.defra.gov.uk/data/ [accessed 20/01/2015].


