acp-2018-690: Advanced methods for uncertainty assessment and global sensitivity analysis of a Eulerian atmospheric chemistry transport model

by Aleksankina et al.

Response to reviewer #1

General Comments

The authors have made a commendable effort to apply uncertainty and sensitivity analysis methods which have a long theoretical history in the stats literature but have only in recent years begun to be applied to complex models such as this.

Response: We thank the reviewer for this supportive comment.

1) *Given that they only look at sensitivity to emissions, and conclude that the model is not particularly sensitive to those inputs, it is perhaps a shame that they didn't attempt to include more input variables into the analysis, as there have been a number of published studies which demonstrate that these methods can be used with significantly larger numbers of inputs.*

Response: There is still a practical trade-off between the desirability of investigating a large number of inputs and the computational costs associated with running a complex atmospheric chemistry transport model (ACTM) for that number of inputs. Hence in this study we concentrated on the input emissions, because input emissions have been reported to strongly affect uncertainty in the modelled surface concentrations of various air pollutants and because simulation of the effects of changes in emissions is a principal application for an ACTM (Introduction, p2 L17). It is only by deploying formal sensitivity methods as we have done here that robust statements can ultimately be made concerning the sensitivities of a given ACTM – in our illustrative case here, the sensitivities of the EMEP4UK model to input emissions.

2) *There are a few concerns regarding the implementation of the methods and the effect that this may have on the validity of the results. In particular please see the points below concerning sample size and emulator validation, which should be addressed before the paper is recommended for publication.*

Response: We address these points where they are raised in more detail below.

Specific Comments

3) *P5,L2 How many emissions inputs does the model have, and why were the ones used chosen?*

Response: The model uses both anthropogenic and biogenic emissions as inputs – a full description of the model inputs can be found in Simpson et al. (2012). In this study, emissions of all of the major primary anthropogenic pollutant compounds were investigated, as described in the paper. The biogenic emission sources were not perturbed, as these are computed within the model itself and are not therefore classed as input datasets. This decision was made based on the fact that one of the main applications of the EMEP model is in providing scientific

support for policy-making regarding impacts of interventions leading to anthropogenic emissions reductions (see Introduction). In the short to medium term at least, the potential future changes in emissions driven by environmental and climate change policies are not likely to affect biogenic emissions as much as anthropogenic emissions, hence it was decided to investigate model response to the changes in anthropogenic emissions.

4) *P5,L5 Why were the shipping emissions not split by pollutant type when the other emissions were? Does this not make the results harder to interpret?*

Response: This goes back to the trade-off between how many input variables should be used in the analysis to obtain the most benefit from it and the computational cost of the analysis. The decision was made to concentrate on the land-based input emissions. The interpretation of sensitivities to shipping emissions was made harder by our approach only in places particularly impacted by shipping. The majority of this impact takes place in the sea-based grid cells where the assessment of the impact of air pollutants on human health and ecosystems is not relevant.

5) *P6,L19 Could the authors include their reasoning for choosing only 84 design points. The normal recommended minimum number for constructing Gaussian process emulators is 10 per input variable, which in this case would be 130. See, for example Loeppky et al, 2009, Choosing the Sample Size of a Computer Experiment: A Practical Guide, Technometrics.*

Response: We were aware of this paper cited by the reviewer. The paper states that an empirical $n = 10d$ rule (where n is the sample size and d is the number of variables under investigation) provides reasonable accuracy when approximating model response with a Gaussian process emulator. This rule is applicable to the model response of any complexity. The paper indicates this is an empirical rule and does not state that $10d$ is the minimum sample size for a Gaussian process emulator to perform well. In our particular study, the input-output response function is expected to be smooth, hence there is no need for a very large number of sample points as it will not reduce error of the emulator.

Again, it is a balance on return for computational resource. Ideally, in the situation where model runs are computationally expensive a sequential sampling technique can be applied to track the improvement of emulator performance with increase in the sample size.

6) *P6,L30 The authors state that the choice of a linear mean function incorporates "prior beliefs" – could they explain what prior beliefs motivated this choice.*

Response: We are referring here to the fact that whilst one can choose the form of $\boldsymbol{h(.)}$ to use, the choice incorporates assumptions, i.e. prior beliefs, as to the most appropriate form to use. In this study the choice was a linear mean function. We assumed that the response of the surface concentration of air pollutants to the changes in the input emissions is likely to be smooth (i.e. with no discontinuities and no sharp fluctuations). Therefore, a prior linear trend was chosen to be more suitable compared to using a constant (mean) or a quadratic (or higher polynomial) function as a trend. In the revised paper we have now expanded the relevant sentence as follow (p7, L12) to explain the choice:

"In this study, the mean function was chosen to have a linear form $\beta_o + \sum_{i=1}^{13} \beta_i x_i$, on the basis that the response of the surface concentration to changes in input emissions is expected to be smooth with no discontinuities or fluctuations."

7) *P7,L10 The authors state that the emulator error was estimated using cross validation and this is presented in the SI. However very little detail is given there except a reference to a paper describing the Matlab package used to construct the emulators (Lataniotis, 2017), which says that the package uses cross validation for parameter estimation, not emulator validation. The authors also say in the paragraph above that they used cross validation for parameter estimation. A clear statement is required as to whether or not the same cross validation was used for both parameter estimation and emulator validation. The accuracy of the emulators is of such key importance to everything that follows that summary statistics of either a separate cross validation, or a validation with a held-out data set, must be presented in the main text.*

Response: We have checked the cross-validation error values by performing a separate cross-validation calculation. The results of explicitly-performed cross-validation do not differ from that which was originally presented in the supplementary material. We accept that the reference we originally cited may have caused some confusion, as the formula for cross-validation is reported in the parameter estimation section. The same formula was used for calculating the cross-validation error for the emulator. We now instead explicitly state in the supplementary information the cross-validation equation used. The three figures in the supplementary information have now been replaced with the updated cross-validation versions. The code written to perform these independent cross-validations is also added to the data repository for this paper.

8) *P8,L2 Could the authors comment on the validity of describing the 5km grid square containing Marylebone Rd as 'urban traffic' and the one containing N. Kensington as 'urban background'. Surely at this resolution both grid squares must be considered as urban background – this is demonstrated by the almost identical sensitivity of NO_2 and O_3 concentrations to NO_x emissions from road transport shown in figure 8.*

Response: The descriptors assigned to the grid squares are the descriptors used by Defra and the UK air quality community for the national network air quality monitoring stations in those grid squares. Whilst we agree that the 5 km resolution of the model causes some averaging of modelled pollutant output compared with a point measurement at those monitoring stations, we do not agree that our results are almost identical for the two grids. The modelled concentrations of NO_2 and O_3 as well as the uncertainty values are different for the two grid squares, and the sensitivity indices (fig. 8) are similar but not identical. It is not claimed that the "Marylebone Rd" patterns of sensitivity are applicable to all 'urban traffic' sites.

9) *P8,L5 The authors state that the "sensitivity indices were estimated". There are a numbered of published methods for doing this so could they say which method they used and why.*

Response: The references provided in the paper describe the numerical methods used for the estimation of the first and total order indices. These methods for estimation of the sensitivity

indices were chosen as they are widely used and well-established methods. However, to clarify explicitly the methods we used for these indices we have now added the following sentence to section 2.4.2: "The first and total-order sensitivity indices were estimated following the methods described by Sobol' (1993) and Janon et al. (2014) respectively."

10) *P14,L9 Given that O_3 concentration is known to be highly non-linear and nonmonotonic in response to changes in NOx and VOC emissions, the contention that these emulators could be used to test emission reduction scenarios is questionable as the small number of training runs used to construct the emulators means they would be unlikely to be able to accurately predict the emissions levels at which the chemical regime changes between being NOx sensitive and VOC sensitive.*

Response: The response of O_3 to changes in NO_x and VOC concentrations although non-linear is still smooth. The effect on O_3 is analysed and presented on annual and monthly timescales, hence additional hourly and daily fluctuations are smoothed out.

11) *P14,L19 Could the very low sensitivity of annual mean ozone to emissions be a result of the strong diurnal variation and photochemical nature of the production of this pollutant? Could the authors comment on whether annual average 8-hour maximum might have been a more informative metric to emulate.*

Response: The annual average 8-hour metric and the annual average are very strongly correlated metrics (see, for example, page 34 of AQEG (2009)); so whilst the absolute values will change we do not believe the nature of the O_3 response to the input emission perturbations will change.

12) *P14,L20 As the authors findings that variation in emissions does not cause substantial variation in the outputs contrasts with their statement in the introduction that it has been previously found that uncertainties in input emissions are major contributors to the uncertainty in the ACTM outputs, could they comment on why this might be – do they think it is a feature of the EMEP model or a result of the analysis methods?*

Response: As was noted in the paper, the previous studies we refer to were performed with different models. The lack of sensitivity here could be due both to the fact that annual and monthly averaged surface concentrations were investigated and to the possible strong effect of background concentrations of pollutants as well as pollutants transported from outside the UK.

The point of this study was not to show how good or not the EMEP4UK model is at reproducing observed pollutant concentrations but to demonstrate a method that can be used to improve the understanding of model response and/or to potentially point out the aspects that may need attention or improvement in future model development.

13) *P14,L25 Given the authors assertion in the previous paragraph that the uncertainty in model output is likely to be driven mainly by variables that they have not included in their analysis, would they concede that their uncertainty estimates are likely to underestimate by a*

large degree the real uncertainty in the model output, i.e. that caused by uncertainty in all of the input variables plus the model discrepancy.

Response: Yes, we agree that the total uncertainty in the estimated concentration of the pollutants is higher than the uncertainty propagated from the input emissions only as there are other uncertain model inputs and parameters. Ideally sensitivity analysis should be incorporated as a part of the model development process (which could aid in both model simplification/reduction and calibration). In that approach the effect of all uncertain inputs and parameters could be assessed without having to do it retrospectively. In addition, screening techniques, e.g. the Morris method (Morris, 1991), could be applied to identify the inputs and parameters that most drive variation in the model outputs and which need to be investigated further. However, here we concentrate on presenting the application of the method itself and on a subset of inputs which previously was shown to drive uncertainty in the model output values.

References:

AQEG: Ozone in the United Kingdom. Fifth report of the Air Quality Expert Group., UK Department for Environment, Food and Rural Affairs, London. PB13216. ISBN 978-0-85521-184-4. http://uk-air.defra.gov.uk/library/reports?report_id=544, 2009.

Janon, A., Klein, T., Lagnoux, A., Nodet, M. and Prieur, C.: Asymptotic normality and efficiency of two Sobol index estimators, ESAIM Probab. Stat., 18, 342–364, doi:10.1051/ps/2013040, 2014.

Morris, M. D.: Factorial Sampling Plans for Preliminary Computational Experiments, Technometrics, 33(2), 161, doi:10.2307/1269043, 1991.

Simpson, D., Benedictow, a., Berge, H., Bergström, R., Emberson, L. D., Fagerli, H., Flechard, C. R., Hayman, G. D., Gauss, M., Jonson, J. E., Jenkin, M. E., Nyíri, a., Richter, C., Semeena, V. S., Tsyro, S., Tuovinen, J.-P., Valdebenito, Á. and Wind, P.: The EMEP MSC-W chemical transport model – technical description, Atmos. Chem. Phys., 12(16), 7825–7865, doi:10.5194/acp-12-7825-2012, 2012

Sobol', I. M.: Sensitivity estimates for nonlinear mathematical models, Math. Model. Comput. Exp., 1(4), 407–414, 1993.)

acp-2018-690: Advanced methods for uncertainty assessment and global sensitivity analysis of a Eulerian atmospheric chemistry transport model

by Aleksankina et al.

Response to reviewer #2

This discussion paper by Aleksankina et al. documents a global sensitivity and uncertainty analyses for the regional chemical transport model EMEP4UK, with the objective of quantifying the uncertainty in surface concentrations of air pollutants (ozone, nitrogen dioxide, and particulate matter below 2.5 um in diameter) and the contribution to that uncertainty from uncertainties in UK-only emissions. No uncertainties associated with model transport and/or chemical processes, or the lateral boundary conditions or driving meteorology were considered. I found the paper to be well organised, well written, and a really nice example of applying powerful statistical approaches to understanding model behaviour and uncertainties. The discussion on the sensitivity analysis itself was very interesting and shows how insightful this technique is. The paper will add to the growing literature base on the use of Gaussian emulation in quantifying uncertainties in geophysical models. I wholeheartedly recommend that the paper is accepted and published in Atmos. Chem. Phys.

Response: We much appreciate the reviewer's comments on the merits of the paper and their enthusiastic recommendation for its acceptance and publication. Thank you.

However, I have a few comments which I hope the authors will consider when submitting a revised manuscript:

1) *Intro: For the non-specialist, I think it would be worthwhile to include some basic introductory material on what you mean by sensitivity analysis versus uncertainty analysis.*

Response: The following text has been added to the Introduction (p2, L26):

"The main distinction between uncertainty and sensitivity analysis is that uncertainty analysis is performed to quantify model output uncertainty arising from the uncertainty in a single or multiple inputs, whilst sensitivity analysis is performed to investigate input–output relationships and to apportion the variation in model output to different inputs. Hence the sensitivity analysis allows conclusions to be drawn on the extent to which the overall variation in the modelled values is driven by variation in different inputs (Saltelli, 2002)"

2. Can you include some discussion on structural uncertainty?

Response: The following text has been added to the Introduction (p2 L17).

"There are various sources of uncertainty in a model; the sources range from structural or conceptual uncertainties about how well a given model represents reality to uncertainties in the model input data and physical and chemical constants, which have an effect on calculation results of the model."

3. *Intro: Note that aerosols affect climate through aerosol-cloud interactions and not only aerosol-radiation interactions*

Response: The first paragraph of the Introduction has been amended to now read as follows:

"Additionally, particulate matter and O_3 contribute to climate change through radiative forcing and aerosol-cloud interactions (for PM) (IPCC, 2013; Stevenson et al., 2013) and O_3 has an adverse impact on natural and semi-natural vegetation and crop yields (Teixeira et al., 2011)."

4. *Intro: Meta models have also been used in exploring climate sensitivity/climate response e.g. Murphy et al. (2004)*

Response: From our search of the literature we are assuming the reviewer is referring to the following paper: "Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth. Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature, 430, 768–772, 204." If so, we do not think it appropriate to include as an example of meta-model application in sensitivity analysis because the methodology for sensitivity analysis described in this paper is one-at-a-time (OAT) and is based on an ensemble approach.

5) *Section 2.1: Full names for SO2, NH3 etc.*

Response: Full names of chemical species are now added in the methods section where they first appear (p4, L19).

6. *Section 2.1: Can you include details of bvoc emissions scheme, and parameterisations for sea salt and dust emissions?*

Response: The following two blocks of text have now been added to section 2.1 (p4):

"Biogenic emissions of monoterpenes and isoprene are calculated by the model for every grid cell and time step according to the methodology of Guenther et al. (1993, 1995), using near-surface air temperature and photosynthetically active radiation as well as aggregated land-cover categorisations, as described in Simpson et al. (2012)."

And

"The details of the sea-salt generation parameterisation scheme used in the model are presented in Monahan et al. (1986) and Mårtensson et al. (2003). The boundary condition monthly average concentrations of fine and coarse dust are calculated with the global chemical transport model of the University of Oslo (Grini et al., 2005). The detailed parametrisation of dust mobilisation is presented in Simpson et al. (2012)."

7) *Table 2: Slight error with SNAP sectors for NH3_O (i.e. 10 should not be included!)*

Response: Thank you for spotting this typo which we have now corrected.

8) *Results Section 3.1: You say that there is a "substantial contribution of hemispheric background O_3 to UK ambient concentrations"? Can you be more quantitative here?*

Response: We cannot estimate the first-order effect of the background O_3 on the UK surface concentrations of O_3 as this input was not one of the perturbed inputs in this study. (We focused on perturbation of primary anthropogenic emissions.) It is not at all straightforward to quantify the background contribution to a secondary pollutant such as O_3 because not only is there import of O_3 and of O_3 precursors into the UK, but the UK is also a surface sink for O_3. The statement in our paper was based on Simpson et al. (2012) who state that "ambient ozone levels in Europe are typically not much greater than the Northern hemispheric background ozone". Additionally, in EMEP4UK a "Mace-Head" adjustment is applied to monthly boundary condition values of the O_3 concentrations. Hence in this paper the contribution of hemispheric background O_3 to the UK ambient concentrations is offered as a possible explanation of the lack of sensitivity of the surface O_3 to changes in the precursor emissions.

We have now added the citation to Simpson et al. (2012) to the end of the sentence in question.

9) *Results Section 3.1: You refer to the 'compensation of errors' as one explanation why the surface response is weak given the input uncertainties. Can you point to the literature for evidence of this statement? I've only seen "compensation of errors" only referred to in the context of process representation in models.*

Response: The phrase comes from Skeffington et al. 2007. In that paper the reason for narrowing of confidence limits for critical loads compared to those of the input parameters was explained to be due to a "compensation of errors" mechanism, but no further explanation was provided. Here, by compensation of errors we mean a situation when the variation in the output is less than expected. This could be caused by multiple inputs having an opposite effect on the magnitude of change in the output of interest.

We have changed the phrasing in our paper to "so-called compensation of errors".

10. *Results Section 3.3: One potential explanation for the seasonal change in sensitivity at Harwell to shipping emissions is the seasonal change in the wind direction which results in more NOx from shipping emissions being transported to the site. Can this be verified from the WRF meteorology used to drive the model?*

Response: The seasonal wind speed and direction for the year 2012 is shown in Figure A below, using the meteorology supplied from the AURN data as extracted using the openair package. It could be argued that there is some correlation between sensitivity index patterns in Figure 8 of our paper and the wind direction; however most likely the seasonality in NO_x sensitivity to shipping emissions is due to a combination of interacting processes within the model.

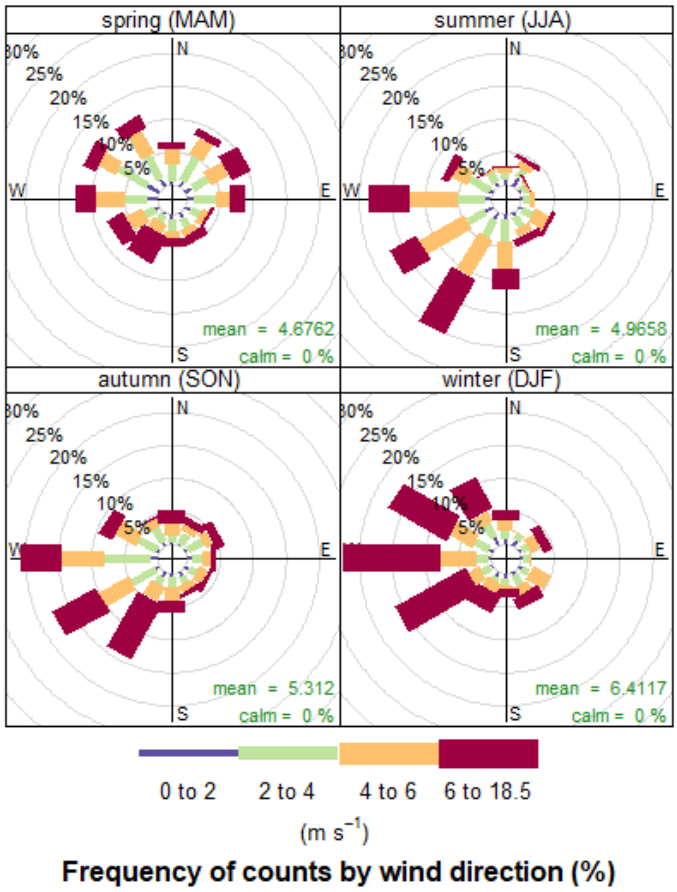

Figure A. Wind rose, Harwell AURN site, 2012.

References:

Grini, A., Myhre, G., Zender, C. S. and Isaksen, I. S. A.: Model simulations of dust sources and transport in the global atmosphere: Effects of soil erodibility and wind speed variability, J. Geophys. Res., 110(D2), D02205, doi:10.1029/2004JD005037, 2005.

Guenther, A., Hewitt, C. N., Erickson, D., Fall, R., Geron, C., Graedel, T., Harley, P., Klinger, L., Lerdau, M., Mckay, W. A., Pierce, T., Scholes, B., Steinbrecher, R., Tallamraju, R., Taylor, J. and Zimmerman, P.: A global model of natural volatile organic compound emissions, J. Geophys. Res., 100, 8873-8892, 1995.

Guenther, A. B., Zimmerman, P. R., Harley, P. C., Monson, R. K. and Fall, R.: Isoprene and Monoterpene Emission Rate Variability - Model Evaluations and Sensitivity Analyses, J. Geophys. Res., 98, 12609-12617, 1993.

Mårtensson, E. M., Nilsson, E. D., de Leeuw, G., Cohen, L. H. and Hansson, H.-C.: Laboratory simulations and parameterization of the primary marine aerosol production, J. Geophys. Res. Atmos., 108(D9), n/a-n/a, doi:10.1029/2002JD002263, 2003.

Monahan, E. C., Spiel, D. E. and Davidson, K. L.: A Model of Marine Aerosol Generation Via Whitecaps and Wave Disruption, pp. 167–174, Springer, Dordrecht., 1986.

Simpson, D., Benedictow, A., Berge, H., Bergström, R., Emberson, L. D., Fagerli, H., Flechard, C. R., Hayman, G. D., Gauss, M., Jonson, J. E., Jenkin, M. E., Nyiri, A., Richter, C., Semeena, V. S., Tsyro, S., Tuovinen, J. P., Valdebenito, A. and Wind, P.: The EMEP MSC-W chemical transport model - technical description, Atmos. Chem. Phys., 12, 7825-7865, 2012.

Skeffington, R. a., Whitehead, P. G., Heywood, E., Hall, J. R., Wadsworth, R. a. and Reynolds, B.: Estimating uncertainty in terrestrial critical loads and their exceedances at four sites in the UK, Sci. Total Environ., 382(2–3), 199–213, doi:10.1016/j.scitotenv.2007.05.001, 2007.

Advanced methods for uncertainty assessment and global sensitivity analysis of a Eulerian atmospheric chemistry transport model

Ksenia Aleksankina,[1,2] Stefan Reis[2,3], Massimo Vieno[2], and Mathew R. Heal[1]

[1] School of Chemistry, University of Edinburgh, Edinburgh, UK
[2] NERC Centre for Ecology & Hydrology, Penicuik, UK
[3] University of Exeter Medical School, European Centre for Environment and Health, Knowledge Spa, Truro, UK

Correspondence to: Ksenia Aleksankina (k.aleksankina@sms.ed.ac.uk) and Mathew Heal (m.heal@ed.ac.uk)

Abstract. Atmospheric chemistry transport models (ACTMs) are extensively used to provide scientific support for the development of policies to mitigate against the detrimental effects of air pollution on human health and ecosystems. Therefore, it is essential to quantitatively assess the level of model uncertainty and to identify the model input parameters that contribute the most to the uncertainty. For complex process-based models, such as ACTMs, uncertainty and global sensitivity analyses are still challenging and are often limited by computational constraints due to the requirement of a large number of model runs. In this work, we demonstrate an emulator-based approach to uncertainty quantification and variance-based sensitivity analysis for the EMEP4UK model (regional application of the European Monitoring and Evaluation Programme Meteorological Synthesizing Centre-West). A separate Gaussian process emulator was used to estimate model predictions at unsampled points in the space of the uncertain model inputs for every modelled grid cell. The training points for the emulator were chosen using an optimised Latin hypercube sampling design. The uncertainties in surface concentrations of O_3, NO_2, and $PM_{2.5}$ were propagated from the uncertainties in the anthropogenic emissions of NO_x, SO_2, NH_3, VOC, and primary $PM_{2.5}$ reported by the UK National Atmospheric Emissions Inventory. The results of the EMEP4UK uncertainty analysis for the annually averaged model predictions indicate that modelled surface concentrations of O_3, NO_2, and $PM_{2.5}$ have the highest level of uncertainty in the grid cells comprising urban areas (up to ± 7%, ± 9%, and ± 9% respectively). The uncertainty in the surface concentrations of O_3 and NO_2 were dominated by uncertainties in NO_x emissions combined from non-dominant sectors (i.e. all sectors excluding energy production and road transport) and shipping emissions. Additionally, uncertainty in O_3 was driven by uncertainty VOC emissions combined from sectors excluding solvent use. Uncertainties in the modelled $PM_{2.5}$ concentrations were mainly driven by uncertainties in primary $PM_{2.5}$ emissions and NH_3 emissions from the agricultural sector. Uncertainty and sensitivity analyses were also performed for five selected grid sells for monthly averaged model predictions to illustrate the seasonal change in the magnitude of uncertainty and change in the contribution of different model inputs to the overall uncertainty. Our study demonstrates the viability of a Gaussian process emulator-based approach for uncertainty and global sensitivity analyses, which can be applied to other ACTMs. Conducting these analyses helps to increase the confidence in model predictions. Additionally, the emulators created for these analyses can be used to predict the ACTM response for any other combination of perturbed input emissions within the ranges set for the original Latin hypercube sampling design without the need to re-run the ACTM, thus allowing fast exploratory assessments at significantly reduced computational costs.

1 Introduction

Air pollution has a wide range of detrimental impacts. Exposure to air pollutants such as nitrogen dioxide (NO_2), ozone (O_3), and particulate matter ($PM_{2.5}$) is associated with increased risk of stroke, cardiovascular disease, and chronic and acute respiratory diseases (WHO, 2006, 2013). Additionally, particulate matter and O_3 contribute to climate change through radiative forcing and aerosol-cloud interactions (for PM) (IPCC, 2013; Stevenson et al., 2013) and O_3 has an adverse impact on natural and semi-natural vegetation and crop yields (Teixeira et al., 2011).

To reduce the harmful impact of air pollution, various policies and directives have been implemented. For example, in the European Union, the Ambient Air Quality Directive (EC Directive, 2008) sets limit values on ambient concentrations of air pollutants, whilst other directives set source-specific emissions limits. Atmospheric Chemistry Transport Models (ACTMs) play an essential role in the evaluation of the potential outcomes of different management options aimed at improvement of future air quality.

The majority of existing ACTMs are deterministic, meaning that the output variables are presented as a single value without any indication of the expected uncertainty around this value. The uncertainty estimate for the modelled value is critical because it provides an assessment of confidence in the model predictions and the confidence range may encompass different recommendations that can be drawn from the model (Frost et al., 2013; Rypdal and Winiwarter, 2001). There are various sources of uncertainty in a model; the sources range from structural or conceptual uncertainties about how well a given model represents reality to uncertainties in the model input data and physical and chemical constants, which have an effect on calculation results of the model. It has been previously found that uncertainties in input emissions are major contributors to the uncertainty in the ACTM outputs (Hanna et al., 2007; Rodriguez et al., 2007; Sax and Isakov, 2003). Therefore, this study concentrates on implementing a systematic approach for ACTM output uncertainty quantification and on determining the extent to which different input emissions drive the uncertainty in the output variables.

Analytical uncertainty propagation is not feasible for complex models such as ACTMs because it requires an exact function for input-output mapping. Consequently, Monte Carlo based methods for uncertainty assessment have to be used. Uncertainty analysis should be performed in tandem with sensitivity analysis to maximise the knowledge gained. The main distinction between uncertainty and sensitivity analysis is that uncertainty analysis is performed to quantify model output uncertainty arising from the uncertainty in a single or multiple inputs, whilst sensitivity analysis is performed to investigate input–output relationships and to apportion the variation in model output to the different inputs. Hence the sensitivity analysis allows conclusions to be drawn on the extent to which the overall variation in the modelled values is driven by variation in different inputs (Saltelli, 2002).

For computationally demanding models, such as ACTMs, a local one-at-a-time (OAT) sensitivity analysis is the most commonly used approach (Ferretti et al., 2015). However, unlike global sensitivity analysis, the local OAT approach does not take into account the non-linearities in the model response and the interactions between the input parameters (Aleksankina et al., 2018; Saltelli and Annoni, 2010).

The computational cost of running ACTMs to explore the entire parameter space of the uncertain inputs using Monte Carlo based uncertainty and sensitivity analyses is typically prohibitively high because the analyses require a large number of points

in parameter space which translates to thousands of model simulations. To tackle this issue, the use of meta-models has been increasing in recent years (Gladish et al., 2017; Iooss and Lemaître, 2015; Ratto et al., 2012; Yang, 2011). A meta-model (or emulator) is a statistical approximation of the original simulation model that can be evaluated many times at a lower computational cost relative to the original model (Castelletti et al., 2012; O'Hagan, 2006). This approach allows the output of an ACTM for a large number of points in parameter space to be estimated efficiently making uncertainty and sensitivity analyses feasible.

Different meta-modelling approaches have been used for uncertainty and sensitivity analysis; these techniques include regression smoothers (Storlie et al., 2009; Storlie and Helton, 2008), Gaussian process emulator (Oakley and O'Hagan, 2004), high-dimensional model representation (Rabitz and Alış, 1999; Ziehn and Tomlin, 2009), and polynomial chaos expansion (Sudret, 2008). Meta-models have been applied for uncertainty and sensitivity analyses in earth science fields such as ecological modelling (Luo et al., 2013; Parry et al., 2013), hydrological modelling (Asher et al., 2015; Gladish et al., 2017), and atmospheric aerosol modelling (Carslaw et al., 2013; Chen et al., 2013; Christian et al., 2017; Lee et al., 2011).

In this study, a Gaussian process is used for emulation because of its desirable properties and available implementations (i.e. Matlab based software UQLab or R package DiceKriging). Gaussian process emulators are non-parametric statistical models that use the principles of conditional probability to estimate model outputs. The beneficial properties are the curve that fits through the training points (for deterministic models) and a measure of the uncertainty for the estimated points when using an emulator in place of the original model for the estimation of new points.

The efficiency of the emulator compared to the original model is determined by how smooth and continuous the model response is to input perturbations. For a smooth and continuous input-output relationship, the high correlation between the inputs and the simulated points means a lower uncertainty in predictions made using the emulator further away from the training points (i.e. resulting in a good emulator performance with a small number of training points) (Lee et al., 2011).

The design of computer experiments for deterministic models differs from the designs for physical experiments. As there is no random error involved in computer experiments, replication is not required (Jones and Johnson, 2009). Hence sampling techniques that have good space-filling properties and the ability to maintain uniform spacing when projected into a lower-dimensional space are used (Dean et al., 2015; Jones and Johnson, 2009). Latin hypercube sampling (LHS) introduced by (McKay et al., 1979) meets these desirable criteria. Additionally, advances have been made to optimise the space filling properties of LHS including maximin sampling (Johnson et al., 1990; Morris and Mitchell, 1995) and the ability to add extra design points to the parameter space if necessary (Sheikholeslami and Razavi, 2017) which makes it well suited for multi-dimensional designs that may require the addition of extra points.

The aim of this study is to demonstrate the method for uncertainty assessment and global sensitivity analysis for computationally demanding ACTMs. The ACTM to which the method is applied here is the WRF-EMEP4UK model (Vieno et al., 2010, 2014, 2016a), and the outputs of interest are the modelled surface concentrations of O_3, NO_2, and $PM_{2.5}$, but the methodology is generic for model and output variable. The analyses described here investigated sensitivities and uncertainties of model output to emissions from UK land-based sources and from surrounding shipping. Additionally, we identify which

model inputs drive uncertainty in the output variables, and to what extent; as well as discuss how the uncertainty ranges that are obtained affect current predictions/scenario analysis outcomes (i.e. confidence in model outputs).

2 Methods

2.1 Model description

The EMEP4UK model is a regional application of the EMEP MSC-W (European Monitoring and Evaluation Programme Meteorological Synthesizing Centre-West) open source ACTM (www.github.com/metno/emep-ctm, version rv4.8, last access: 11 June 2018). The detailed description of EMEP MSC-W is available from Simpson et al. (2012), and the EMEP4UK model is described by (Vieno et al., 2010, 2014, 2016a).

EMEP4UK is a 3-D one-way nested Eulerian model with a horizontal resolution of 5 km × 5 km over the British Isles nested within an extended European domain with 50 km × 50 km resolution. The extent of the inner domain is shown in Figure 1. The model has 20 vertical levels, extending from the ground to 100 hPa with the lowest vertical layer of ~90 m. The model time-step is 20 s for chemistry, 5 min for the advection in the inner domain, and 20 min for the advection in the outer domain. The meteorological fields were computed using Weather Research and Forecast model version 3.1.1 (www.wrf-model.org, last access: 15 November 2017) (Skamarock et al., 2008). The WRF model initial and boundary conditions are derived from the US National Center for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) Global Forecast System (GFS) at 1◦ resolution, including Newtonian nudging every 6 h (NCEP, 2000).

The anthropogenic emissions of sulfur dioxide (SO_2), nitrogen oxides (NO_x), ammonia (NH_3), fine and coarse primary particulate matter ($PM_{2.5}$, PM_{coarse}), carbon monoxide (CO), and non-methane volatile organic compounds (NMVOC) for the UK were derived from the National Atmospheric Emissions Inventory (http://naei.beis.gov.uk/, last access: 15 October 2015). For the outer domain, the emissions are provided by the Centre for Emission Inventories and Projections (CEIP, http://www.ceip.at/, last access: 15 October 2017). All emissions are split across a set of emission source sectors defined by the Selected Nomenclature for Air Pollutants (SNAP) described in Table 1. The hour-of-day, day-of-week and monthly emission factors are used to distribute the annual total emissions temporally to hourly resolution as described in Simpson et al. (2012). The international shipping emissions were derived from ENTEC UK Ltd. (now Amec Foster Wheeler). Biogenic emissions of dimethyl sulfide in addition to monthly in-flight aircraft, soil, and lightning NO_x emissions are included as described in Simpson et al. (2012). Biogenic emissions of monoterpenes and isoprene are calculated by the model for every grid cell and time step according to the methodology of Guenther et al. (1993, 1995), using near-surface air temperature and photosynthetically active radiation as well as aggregated land-cover categorisations, as described in Simpson et al. (2012). The emissions of sea salt and wind-blown dust are also included. The details of the sea salt generation parameterisation scheme used in the model are presented in Monahan et al. (1986) and Mårtensson et al. (2003). The boundary condition monthly average concentrations of fine and coarse dust are calculated with the global chemical transport model of the University of Oslo (Grini et al., 2005); the detailed parametrisation of dust mobilisation is presented in Simpson et al. (2012).

The chemistry, aerosol formation, and wet and dry deposition schemes are as described in Simpson et al. (2012). The chemistry scheme has 72 species, 137 reactions, and the gas/aerosol partitioning is described by the MARS formulation. A detailed evaluation of model performance is discussed elsewhere (Dore et al., 2015; Lin et al., 2017; Vieno et al., 2010, 2016b). In our study, all model runs were executed using meteorology and emissions data for the year 2012.

2.2 Input variables and their uncertainty ranges

For this study, emissions of five pollutants (NO_x, SO_x, VOC, NH_3, primary $PM_{2.5}$) were split into 13 model input variables based on the contributions from different emission source sectors to total annual emissions; the emissions from the dominant sector (the sector with the highest relative contribution to total emissions) for every pollutant were treated as a separate variable, while the emissions from the rest of the sectors were grouped and treated as another input variable. Shipping emissions were treated as a separate variable and were not split by the pollutant type. The description of the Selected Nomenclature for Air Pollution (SNAP) sectors is shown in Table 1, and the definitions of the input variables for the uncertainty and sensitivity analyses in this work are presented in Table 2, where variables marked with D represent emissions from a single dominant sector (D1 and D2 in case of multiple dominant sectors) and variables marked with O indicate the grouped 'other' emissions from the rest of the sectors. Emissions from 'natural' sources (e.g. lightening, soil, ocean) were not part of the uncertainty and sensitivity analyses.

Uncertainty ranges for the input emissions from UK anthropogenic land-based sources were assigned according to data in the UK Informative Inventory Report (IIR) (Wakeling et al., 2017). In the IIR, uncertainties are defined as upper and lower limits of the 95% confidence interval relative to the central estimate. There is no information on uncertainty ranges for different source sectors available for the emissions for 2012 because uncertainties split by the emission source sector were first presented in the IIR that included 2014 emissions (Wakeling et al., 2016). Hence, for this study, the most recently published data for the uncertainty ranges of pollutants split by source sector were used.

Equation 1 was used to aggregate uncertainties for multiple emission source sectors for the grouped-source input variables, where x is the quantity of interest and U is the uncertainty of that quantity, taken from the EMEP/EEA air pollutant emission inventory guidebook (Pulles and Kuenen, 2016).

$$U_{total} = \frac{\sqrt{(U_1 x_1)^2 + (U_2 x_2)^2 + \cdots + (U_n x_n)^2}}{x_1 + x_2 + \cdots + x_n} \quad (1)$$

The shipping emission variable in this study combines all emissions of all relevant pollutants, hence a 'best estimate' range for the uncertainty was chosen. The range was estimated based on the available published information. Some recently published sources (Corbett, 2003; Scarbrough et al., 2017) state that the uncertainty in shipping emissions is significant, but do not provide quantitative estimates. The most recent source of quantitative information on the uncertainty in shipping emissions is the report for the European Commission (Entec, 2002) which presents the estimates of uncertainties for emission factors of NO_x, SO_2, PM, VOC, for the ships' emissions 'at sea', 'manoeuvring', and 'in port'. The uncertainties are presented for the emissions for the year 2000 as 95% CI with the lowest values of uncertainty presented for 'at sea' emission factors (± 10-20%)

and highest values for 'manoeuvring' emission factors (± 30-50%). For the total pollutant emissions for the year 2000 the percentage uncertainties around the estimates are ± 21% for NO_x, ± 11% for SO_2, ± 11% for CO_2, ± 28% for VOC and ±45% for PM. Additionally, in Moreno-Gutiérrez et al. (2015) the uncertainty in the emission factors for all pollutant compounds was estimated to be ± 20%. Using the above data, an overall uncertainty of ± 30% was assigned to the shipping emissions variable in this study (Table 2). It was applied to all shipping emissions within the inner British Isles domain of the EMEP4UK model.

2.3 Gaussian process emulator for EMEP4UK

A Gaussian Process emulator was used to estimate model predictions at unsampled points in the space of the uncertain model inputs. The UQLab, a MATLAB-based software framework for uncertainty quantification (Lataniotis et al., 2017; Marelli and Sudret, 2014), was implemented to build the emulators for the uncertainty propagation and the following sensitivity analysis. The comprehensive description of the statistical theory of Gaussian process applied to uncertainty and sensitivity analysis with full mathematical details can be found in O'Hagan (2006) and Oakley and O'Hagan (2002, 2004).

The uncertainty values and sensitivity indices were calculated for three EMEP4UK model outputs (O_3, NO_2, and $PM_{2.5}$ surface concentrations) with annual and monthly temporal resolution. For the annually-averaged outputs, an emulator was created for each modelled grid cell in the EMEP4UK domain (n = 59 400). The first and total-order sensitivity indices were calculated for the land-based grid cells only (n > 10 000). For the monthly mean model outputs, uncertainty and sensitivity analysis were performed for five selected grid cells. The five grid cells were selected to contain a UK national-network air pollution monitoring station to aid classification according to the environment (i.e. rural background, urban background, and urban traffic) and also to provide geographically representative coverage across the UK.

LHS maximin design, which maximises the minimum distance between the points in the parameter space to provide the optimum space-filling properties was used. The design was previously demonstrated suitable for Gaussian process emulators by Jones and Johnson (2009). The design with 84 data points was created for the scaling coefficients that were subsequently applied to the input emissions. This means that emissions corresponding to a particular input variable were perturbed homogeneously throughout the whole of the UK model domain. The ranges of scaling coefficient used for the sampling design are presented in Table 2.

In this study, the surface concentration of O_3, NO_2, and $PM_{2.5}$ for every grid cell is defined as a scalar output $Y = f(\boldsymbol{X})$ where $\boldsymbol{X}$ is the vector of input values $\{X_1, \ldots, X_{13}\}$.

A Gaussian process emulator utilises a Bayesian approach; the training data is used to update the selected prior to produce posterior mean and covariance functions. The Gaussian process is specified by its mean function and covariance function. The mean function is given by Eq. 2:

$$\mathbb{E}[f(\boldsymbol{x})|\boldsymbol{\beta}] = \boldsymbol{h}(\boldsymbol{x})^{\boldsymbol{T}}\boldsymbol{\beta} \quad (2)$$

where $\boldsymbol{h(\cdot)}$ is a vector of regression functions and $\boldsymbol{\beta}$ is a vector of unknown coefficients. The choice of $\boldsymbol{h(\cdot)}$ incorporates any prior beliefs about the form of $f(\cdot)$. In this study, the mean function was chosen to have a linear form $\beta_o + \sum_{i=1}^{13} \beta_i x_i$ on the basis that the response of the surface concentration to changes in input emissions is expected to be smooth with no discontinuities or fluctuations.

The covariance function between $f(\boldsymbol{x})$ and $f(\boldsymbol{x}')$ is given by Eq. 3:

$$cov\{f(\boldsymbol{x}), f(\boldsymbol{x}')|\sigma^2\} = \sigma^2 c(\boldsymbol{x}, \boldsymbol{x}') \quad (3)$$

where σ^2 is the hyperparameter that represents the variance of the Gaussian process and $c(\boldsymbol{x}, \boldsymbol{x}')$ is the correlation function. The correlation function increases as the distance between $\boldsymbol{x}$ and $\boldsymbol{x}'$ decreases and equals one when $\boldsymbol{x} = \boldsymbol{x}'$. In this study Matérn 5/2 (Eq. 4) was used, where h is the absolute distance between $\boldsymbol{x}$ and $\boldsymbol{x}'$ and $\boldsymbol{\theta}$ is a vector of range parameters or length-scales, which define how far one needs to move along a particular axis in the input space for the function values to become uncorrelated.

$$c(x, x') = \left(1 + \frac{\sqrt{5}|h|}{\theta} + \frac{5\,h^2}{3\theta^2}\right) \exp\left(-\frac{\sqrt{5}|h|}{\theta}\right) \quad (4)$$

A number of emulators were built with the EMEP4UK simulation data using other available covariance functions; however, little difference was found in the performance of the emulators. The hyperparameters $\boldsymbol{\beta}$, σ^2, and $\boldsymbol{\theta}$ were estimated using a cross-validation approach.

The emulator error was estimated by implementing k-fold cross-validation (Gladish et al., 2017; Urban and Fricker, 2010). The original sample was randomly partitioned into k = 10 sized subsamples which allowed approximately 90% of data to be used as a training set and 10% as a validation set. Spatial distribution of cross-validation errors is presented in the supplementary information (Figure S1).

2.4 Uncertainty and sensitivity analysis

2.4.1 Uncertainty propagation

The uncertainties for the EMEP4UK output variables were estimated using a Monte Carlo approach (also described in the IPCC guidelines (IPCC, 2006) as a Tier 2 approach). The specific uncertainty ranges assigned to the input emission variables were used to constrain the input sampling space. All inputs were assigned normal distributions with baseline value as the mean and the standard deviation derived from the corresponding confidence interval (Table 1). For every grid cell, the emulator was used to predict model values of surface concentrations of O_3, NO_2, and $PM_{2.5}$ at the new set of input points (n = 5,000). The resulting probability distributions for each grid cell were evaluated, and the resulting uncertainty was estimated as a half of the 95% confidence interval relative to the central estimate (i.e. the mean for normally distributed values) of the output value, as described in the EMEP/EEA and IPCC Guidebooks (IPCC, 2006; Pulles and Kuenen, 2016). The uncertainty for the monthly

average modelled surface concentrations of O_3, NO_2, and $PM_{2.5}$ was calculated for five grid cells using the same approach as above. The locations of the grid cells within the UK are shown in Figure 1. The five grid cells selected were assigned the following environment types – the names and environment type reflect those of the national-network monitoring site within that grid cell: Auchencorth Moss and Harwell - rural background, Birmingham Acocks Green and London N. Kensington - urban background, and London Marylebone Road - urban traffic.

2.4.2 Global sensitivity analysis; first- and total-order indices

A variance-based global sensitivity analysis was conducted to apportion overall uncertainty in modelled variables to the uncertainty in the input emissions. Sobol' first and total-order sensitivity indices were estimated (Homma and Saltelli, 1996; Janon et al., 2014; Sobol', 2001, 1993). The first-order indices represent the fraction of total variance of the output (i.e. the proportion of the overall uncertainty in Y) explained by the variance in an input X_i while total-order indices show the sum of the effects due to an input X_i and all of its interactions with other inputs ($X_{\sim i}$). Therefore, the values of first and total-order indices can be compared to identify the presence of interactions between input X_i and all other model inputs.

Unlike an OAT sensitivity coefficient, a first-order sensitivity index accounts for the non-linear response of a model output to a parameter across the specified parameter variation range. Sensitivity indices in this context are also indicators of importance for the input variables.

The first-order sensitivity index is defined as the ratio of the variance of the mean of Y when one input variable is fixed, $V_{Xi}(E_{X\sim i}(Y|X_i))$, to the unconditional variance of Y, $V(Y)$ (Eq. 5).

$$S_i = \frac{V_{X_i}(E_{\boldsymbol{X}_{\sim i}}(Y|X_i))}{V(Y)} \tag{5}$$

The total order sensitivity index measures the total effect of a variable, which includes its first-order effect and interactions with any other variables (Eq. 6).

$$S_{Ti} = 1 - \frac{V_{\boldsymbol{X}_{\sim i}}\left(E_{X_i}(Y|\boldsymbol{X}_{\sim i})\right)}{V(Y)} = \frac{E_{X_{\sim i}}\left(V_{X_i}(Y|\boldsymbol{X}_{\sim i})\right)}{V(Y)} \tag{6}$$

where $\boldsymbol{X}_{\sim i}$ denotes the matrix of all variables but X_i. In $E_{X\sim i}\,(V_{Xi}(Y \mid \boldsymbol{X}_{\sim i}))$ the inner variance of Y is taken over all possible values of X_i while keeping $\boldsymbol{X}_{\sim i}$ fixed, while the output expectation E is taken over all possible values $\boldsymbol{X}_{\sim i}$ (Ghanem et al., 2017). The first and total-order sensitivity indices were estimated following the methods described by Sobol' (1993) and Janon et al. (2014) respectively.

For the annual average modelled surface concentrations of O_3, NO_2 and $PM_{2.5}$, the sensitivity indices were calculated for the UK land-based grid cells for the whole domain. For the monthly average modelled concentrations, sensitivity indices for five

selected grid cells (discussed above) were estimated to determine whether seasonality affects the magnitude of the sensitivity indices.

3 Results and discussion

3.1 Uncertainty propagation

Figure 2 shows the spatial distribution of annual average surface concentrations of O_3, NO_2, and $PM_{2.5}$ modelled with EMEP4UK and their absolute and relative uncertainties given the uncertainties in UK pollutant emissions for each source sector shown in Table 2. The uncertainties are presented as a range of ± the baseline value and represent the 95% confidence interval. The maps represent the uncertainty in surface concentrations propagated from the uncertainties reported in the UK emissions (Wakeling et al., 2017) and estimated uncertainties in shipping emissions in the EMEP4UK model domain (Entec, 2002; Moreno-Gutiérrez et al., 2015). The uncertainties in surface concentration do not incorporate any uncertainties in the spatial and temporal aspects of the input emissions because no data on these aspects of uncertainty are provided by the compilers of the emissions inventories.

For O_3 and NO_2 the areas with the highest uncertainty coincide with the location of the shipping lanes. This is due to assigning all shipping emissions an uncertainty of ± 30%, which causes high variability in the corresponding NO_x emissions. The uncertainty in O_3 surface concentrations for the land-based grid cells is generally low (median relative uncertainty is ± 0.6%) with values of uncertainty up to ± 7% or ± 1.4 ppb occurring in the grid cells containing major UK cities. The overall low uncertainty in the modelled O_3 concentrations can be attributed to the combination of a low uncertainty in precursor emissions and the substantial contribution of hemispheric background O_3 to UK ambient concentrations, the concentrations of which are not part of this analysis of uncertainty with respect to the UK-only emissions (Simpson et al., 2012).

The relative uncertainty of NO_2 has a homogeneous spatial pattern (median relative uncertainty for all land-based grid cells is ± 7.4%) while absolute uncertainty is found to be higher (up to ± 3.5 μg m^{-3} or ± 9%) in the areas with the major UK cities. The magnitude of uncertainty in NO_2 is determined by the combination of two factors: i) NO_2 uncertainty is driven by NO_x emission inputs which have low levels of uncertainty associated with them; ii) low overall variation in O_3 surface concentrations affects the reactions between NO, NO_2 and O_3 that are linked through the photolysis of NO_2 to give NO and the reaction of NO with O_3 to produce NO_2.

The spatial pattern of $PM_{2.5}$ surface concentrations and the corresponding absolute and relative uncertainties differ from those for O_3 and NO_2. The concentration gradient indicates the presence of transboundary $PM_{2.5}$ transport into the UK. This is consistent with findings reported by AQEG (2013) that only about half of the $PM_{2.5}$ annual average concentrations have a UK origin. The spatial pattern of uncertainty in $PM_{2.5}$ concentrations shows higher uncertainty, both relative and absolute, in the grid cells with major cities; median relative uncertainty for all land-based grid cells is ± 4.6% with up to ± 9% (± 0.9 μg m^{-3}) in the grid cells with major cities. The surface concentrations of $PM_{2.5}$ are dominantly comprised of primary $PM_{2.5}$ emissions and inorganic aerosols resulting from chemical reactions between SO_2, NO_x, and NH_3. Hence the spatial pattern of uncertainty can be explained by the fact that the main contribution to primary $PM_{2.5}$ comes from emissions from sources such as stationary

combustion (e.g. residential heating) and road transport. The pattern of decreasing uncertainty from the land-based grid cells (centre) towards the edges of the domain indicates the change in variation due to the transport of $PM_{2.5}$ away from the sources of emitted pollutants.

The overall uncertainty in the output variables (O_3, NO_2, and $PM_{2.5}$) was found to be lower compared to the uncertainty of the model input emissions. This can be explained by the overall weak response of surface concentrations to changes in the emission originating from the UK which leads to the conclusion that the surface concentrations are affected by the transport of pollutants from elsewhere. Another explanation is the 'so-called compensation of errors' whereby a positive effect of one or multiple input variables on the output is compensated by a negative effect of another input variable(s). This leads to the narrower confidence intervals associated with the EMEP4UK outputs.

An important observation from this uncertainty analysis is that the areas with the highest uncertainty coincide with the most populated areas. Given that O_3, NO_2, and $PM_{2.5}$ are associated with adverse health effects, it is particularly important to have an estimate for the confidence level of the modelled values in the more densely-populated regions. This work has shown that the highest uncertainty is precisely in these regions. The reason for the increased levels of uncertainty in the grid cells coinciding with urban areas is discussed below.

3.2 Sensitivity analysis

In addition to quantitative uncertainty estimates, it is of interest to know how the uncertainty of each input contributes to the overall uncertainty and whether there are interactions between inputs that potentially affect the magnitude of overall uncertainty. This was achieved by conducting a variance-based sensitivity analysis.

Figures 3, 4, and 5 show the spatial distribution of the first-order sensitivity indices that represent the fractional contribution of the uncertainty of each input variable to the overall uncertainty in the output. Only the variables with $S_i > 0.03$ are presented here. First-order indices with values less than 0.03 were omitted as the method used for computation of sensitivity indices is prone to numerical errors when the analytical sensitivity index values are close to zero (Saltelli et al., 2006). The threshold was estimated by examining the noise in first-order sensitivity indices calculated for unimportant input variables. Excluding $S_i < 0.03$ does not have an effect on the results presented because a relative contribution of less than 3% to the overall uncertainty can be considered negligible.

Difference between total and first-order sensitivity is used to highlight interactions between variable X_i and all other input variables. For the sensitivity coefficients computed for the annual-averaged model outputs, there was no substantial difference found between first and total- order sensitivity indices, hence no between-input interactions were identified on the annual timescale (Fig. S2).

Figure 3 shows the spatial distribution of first-order sensitivity indices for the input variables affecting modelled O_3 concentrations. It is predominantly the NO_x input emissions that drive the uncertainty in modelled O_3 surface concentrations. The greatest contribution to O_3 surface concentration uncertainty in the areas with higher levels of overall uncertainty is from the input variable NO_x_O, which represents NO_x emissions from all the other SNAP sectors apart from SNAP 1 (combustion in energy and transformation industries) and SNAP 7 (road transport). The NO_x emissions combined into this input variable

account for 27% of total NO_x emissions and the uncertainty range for this variable is ± 19%. The input variable NO_x_D1 (emissions from combustion in energy and transformation industries) does not contribute substantially to output uncertainty despite making up 41% of total NO_x emissions, with a relative uncertainty of ± 7%. This is explained by the height at which these emissions occur; the emissions are injected into the vertical layers at heights of >184 m above ground level. This leads to NO_x being dispersed and transported away from these elevated sources without affecting ground-level O_3 concentrations locally. The NO_x emissions from input variable NO_x_D2 (road transport) account for the remaining 32% of total NO_x emissions. The spatial distribution of corresponding sensitivity indices indicates that uncertainty in road transport emissions affects overall uncertainty in O_3 surface concentrations in the grid cells closest to the emission sources (i.e. major roads). A large proportion (>80%) of overall uncertainty in O_3 concentrations in areas adjacent to the south and south-east coasts of England is apportioned to the uncertainty in shipping emissions.

In Scotland, most of the overall uncertainty in O_3 surface concentration is apportioned to the variables VOC_D and VOC_O that respectively represent VOC input emissions from the dominant VOC source sector (solvent and other product use) and emissions from the rest of the source sectors grouped into a single input. A small proportion is apportioned to the variable NH3_D that represents NH_3 emissions from agricultural sources. The effect of these input variables manifests in Scotland because of low levels of locally-emitted NO_x. The overall uncertainty in this area is very low.

In summary, the uncertainty in modelled surface concentrations of O_3 in the densely populated areas can be apportioned to the uncertainty in NOx emissions from non-dominant sources and uncertainty in shipping emissions.

The uncertainty in surface concentration of NO_2 was found to be driven mostly by uncertainty in NO_x emissions (variables NO_x_D1, NO_x_D2, NO_x_O) and shipping emissions (Fig. 4). Similarly to O_3, NO_2 is most sensitive to NO_x emissions combined from all SNAP sectors apart from SNAP 1 (combustion in energy and transformation industries) and SNAP 7 (road transport). There is almost no sensitivity to NO_x emissions from SNAP 1, for the same reason given above that these are elevated emissions. The sensitivity to NO_x emissions from SNAP 7 is most pronounced close to the source of emissions (i.e. major roads and cities).

The similarity in spatial distribution of sensitivity indices for O_3 and NO_2 model outputs results from the concentrations of these pollutants being inversely correlated, as their chemical transformation reactions are interlinked. In the same way as for O_3, uncertainty in the NO_2 concentrations along the south and south-east coasts of England is mostly driven by the uncertainty in the shipping emissions. In fact, uncertainty in shipping emissions contributes approximately 30% of uncertainty in NO_2 concentrations even well inland, in areas away from major roads and cities.

Figure 5 shows the spatial distribution of first-order sensitivity indexes for the model inputs that contribute to the uncertainty in modelled surface concentrations of $PM_{2.5}$. Modelled $PM_{2.5}$ is sensitive to all emissions of NH_3 (dominant sector is agriculture) and to primary $PM_{2.5}$ (dominant sectors D1 is residential combustion and D2 is road transport), and to shipping emissions. In the areas with lower surface $PM_{2.5}$ concentrations such as Scotland, Wales, northern England and south-west England the uncertainty is mainly driven by NH_3 emissions from agriculture (NH_3_D). The spatial pattern of emissions sensitivity indices for $PM_{2.5}$ mirrors the spatial distribution of $PM_{2.5}$ emission sources. From Figure 2 and Figure 5 it can be

seen that in the areas with the highest levels of uncertainty the model output is most sensitive to the emissions of primary $PM_{2.5}$. Similar to the results for O_3 and NO_2, the areas with the highest uncertainty coincide with the most populated areas.

The pattern in calculated sensitivity indices partially agrees with a previous study of changes in $PM_{2.5}$ surface concentrations in response to 30% reduction in emissions of $PM_{2.5}$, NH_3, SO_x, NO_x, and VOC by Vieno et al. (2016). In the study by Vieno et al. (2016) surface concentrations of $PM_{2.5}$ were found to be sensitive to reductions in each of the five pollutants individually (the same reduction was applied to a pollutant's emissions from all SNAP sectors simultaneously), with highest sensitivity to NH_3 and $PM_{2.5}$ emissions (up to ~6% reduction in surface concentration in response to 30% reduction in emissions). In comparison, our study the uncertainty in $PM_{2.5}$ surface concentrations is not affected by the perturbations of SO_x, NO_x, and VOC. This is likely to be due to i) the difference in ranges of variation (i.e. uncertainty ranges) in this study (SO_x, NO_x and VOC input variables have narrower ranges of variation compared to $PM_{2.5}$ and NH_3), and ii) the presence of non-additivity and non-linearity in the model response to perturbations in the inputs.

3.3 Uncertainty propagation and sensitivity analysis for monthly averaged model outputs

The uncertainty assessment and sensitivity analysis for monthly averaged surface concentrations of NO_2, O_3, and $PM_{2.5}$ were performed for five different grid cells that were assigned the following environment types based on the national-network monitoring site within that grid cell: Auchencorth Moss and Harwell - rural background, Birmingham Acocks Green and London N. Kensington - urban background, and London Marylebone Road - urban traffic.

Monthly average concentrations with error bars representing the absolute uncertainty values (as a 95% CI) are presented in Figure 6. Figure 7 shows corresponding values of the relative uncertainty. Figure 8 shows how the magnitude of first-order sensitivity indices estimated for five different grid cells changes on a monthly timescale. If all first-order sensitivity coefficients add up to 1 then there are no interactions between inputs and all model variance can be apportioned to the variance in the individual inputs.

The NO_2 surface concentrations show a seasonal trend of lower concentrations occurring during summer months with the exception of the Auchencorth Moss grid cell where NO_2 concentrations are low throughout the year. The magnitude of uncertainty in NO_2 is proportional to the modelled concentration and changes relative to the concentration, which can be seen from the monthly relative uncertainty values (Fig. 7). The first-order sensitivity indices for NO_2 show that only NO_x emissions (across all sectors) and shipping emissions influence the modelled surface NO_2 concentrations. Hence it can be concluded that the uncertainty in modelled concentrations of NO_2 directly depends on the uncertainty in NO_x emissions and is not affected by the uncertainties in the emissions of any other pollutant. The change in the magnitude of sensitivity coefficients for the Harwell grid cell indicates increasing influence of shipping emissions on NO_2 concentrations during the summer months. Potential explanation for this is seasonal change in the wind direction which results in more NO_x from shipping emissions being transported to the grid cell during the summer months.

The uncertainties in the O_3 modelled surface concentrations show an inverse seasonal trend compared to the uncertainties in modelled NO_2. Unlike the uncertainty in NO_2 concentration, the uncertainty in O_3 concentration is influenced by the grid cell environment type; the highest level of uncertainty is observed for the London Marylebone Road grid cell (urban traffic). The

relative uncertainty in O_3 concentrations for the Auchencorth Moss grid cell (rural background) is small and close to the median relative uncertainty in O_3 for annual average concentrations, which as discussed above is ± 0.6%. This indicates that perturbations in the input emissions do not substantially affect O_3 concentration in this grid cell. Although the magnitude of uncertainty in O_3 is very small in this grid cell, the inputs that drive it differ noticeably throughout the year; during May-August the variance is mostly explained by VOC emissions (explains 77% of uncertainty for July) and during November-February NO_x emissions drive the uncertainty. The magnitude of O_3 concentrations and corresponding uncertainties in the Birmingham Acocks Green and Harwell grid cells are very similar. The trends in sensitivity indices are also similar; during the April-September period some variance in the model output is explained by uncertainty in VOC emissions. However, in the Harwell grid cell shipping emissions play a more important role. For the London-based grid cells, the level of uncertainty is the highest and it is mainly driven by the uncertainty in NO_x and shipping emissions.

For the $PM_{2.5}$ monthly average concentrations, London-based grid cells show the highest values of absolute uncertainty and Auchencorth Moss - the lowest. The relative uncertainty in London based grid cells is also the highest. From Figure 7 it can be seen that the contribution to the overall uncertainty from the uncertainty due to NH_3 emissions for these grid cells is not as important as for other three, the majority of uncertainty is explained by the uncertainty in the primary $PM_{2.5}$ emissions with $PM_{2.5}$ from road transport being the dominating variable. In Birmingham Acocks Green and Harwell, the effect of NH_3 emissions from agricultural sources is more pronounced; from 30% to 70% of overall uncertainty in $PM_{2.5}$ can be apportioned to uncertainty coming from agricultural emissions of NH_3 during spring and summer months.

3.4 Wider implications of our study

There are published studies that apply global sampling-based uncertainty and sensitivity analyses as well as derivative based methods (methods that do not have limitations of local OAT, i.e. linearity assumption) to ACTMs. However, the results reported by these studies are mostly of use for model development and calibration purposes and not the assessment of confidence in the model predictions/outputs. This is mainly because the simulations are performed for a short period ranging from days (Beddows et al., 2017; Chen and Brune, 2012; Rodriguez et al., 2007) to weeks (Cohan et al., 2010; Shrivastava et al., 2016).

Additionally, in some studies, commercial software or packages with a graphical user interface (GUI) are used for global sensitivity and uncertainty analysis (Chen and Brune, 2012; Christian et al., 2017; Lee et al., 2011). These tools are well designed for a specific purpose but lack the option to scale up and to automate the analysis, i.e. ability to calculate sensitivity indices and uncertainty ranges for thousands of grid squares automatically.

Our study addresses both of the shortcomings. We demonstrate sensitivity and uncertainty analyses for the ACTM for a whole year for the UK domain as well as investigate variations in sensitivity and uncertainty on the monthly timescale for multiple locations with different environmental characteristics. Additionally, the package used to create Gaussian process emulators and to conduct uncertainty and sensitivity calculations is fully customisable and can be adapted for any application.

The model runs generated for the global sensitivity and uncertainty analysis can be utilised for other purposes provided that the sampling range for all inputs of interest is wide enough. For example, in our study the training points for the Gaussian

emulator were selected to cover a wider range of input perturbations compared to the corresponding uncertainty range (Table 2). For all input emissions of SO_x, NO_x, VOC, and NH_3 the ranges of variation for the LHS design were set to ± 40% of their baseline value, for primary $PM_{2.5}$ emissions the range was set to ± 75% and for shipping emissions from – 40% to + 100%. Hence the emulators created in this study using the model runs within the aforementioned input space can be used to investigate other scenarios of the model response to input emission perturbations with no extra computational cost. Hence, alternative ranges and probability distributions can be assigned to the model inputs to estimate the resulting output uncertainty or the emulator can be used for various emission reduction scenario analyses.

4 Conclusions

In this study, we have conducted global sensitivity and uncertainty analyses for the EMEP4UK Eulerian atmospheric chemistry transport model to quantify the uncertainty in surface concentrations of O_3, NO_2, and $PM_{2.5}$ and to identify the input emission variables that contribute the most to the uncertainty in each of the outputs. The uncertainty for model outputs was estimated from the uncertainties assigned to the UK emissions of SO_2, NO_x, NH_3, VOC, and primary $PM_{2.5}$ and documented in the UK National Atmospheric Emissions Inventory. The benefit of conducting global sensitivity analysis in addition to uncertainty assessment is that it allows to determine how a model responds to the input perturbations within the ranges set by the input uncertainty estimates and consequently to identify the inputs which cause the variation in the model outputs (i.e. drive the uncertainty). The median values of the overall uncertainty calculated for the UK land-based grid cells for annual average surface concentrations of O_3, NO_2, and $PM_{2.5}$ were found to be in the ranges of ±0.6%, ±7.4%, and ±4.6% respectively. This indicates that the variation in the input data (i.e. emissions) does not cause a substantial variation in the outputs. Our results indicate, that this can likely be explained by variations in the other model input parameters such as chemical reaction rates, deposition velocities or physical constant values which might cause more variation in the model outputs. Alternatively, surface concentrations of the modelled pollutants in the UK may be dominated by the precursor emissions and long-range transport from outside the UK and are therefore relatively insensitive to changes in the UK emissions.

As a consequence, our results can provide more clarity about the confidence in modelled surface concentrations of pollutants that affect human health, especially in densely-populated urban areas. The results of our analysis indicate that modelled surface concentrations of O_3, NO_2, and $PM_{2.5}$ have the highest level of uncertainty in the grid cells comprising dense urban areas. The uncertainties of O_3, NO_2, and $PM_{2.5}$ in these grid cells reach ± 7%, ± 9%, and ± 9% respectively.

In addition to obtaining a quantitative estimate of the overall uncertainty, the input emissions that have the greatest influence on the uncertainty in the modelled outputs were identified by performing a global variance-based sensitivity analysis. It was found that in urban areas uncertainty in $PM_{2.5}$ concentrations are driven by the uncertainty in primary $PM_{2.5}$ emissions. In contrast, in more remote areas NH_3 emissions had a stronger influence. Emissions of NO_x combined from non-dominant sectors (i.e. all sectors excluding energy production and road transport) were found to contribute the most to the uncertainty in both O_3 and NO_2 surface concentrations. Along the south and east coasts of England the uncertainty in shipping emissions contributed the most to the overall uncertainty in O_3 and NO_2 concentrations.

The comparison between first and total-order sensitivity indices did not indicate substantial interactions between the input variables for the model response on the annual timescale.

In our study we also demonstrated how the degree of uncertainty changes throughout the year by calculating uncertainty ranges for monthly-averaged surface concentrations of O_3, NO_2, and $PM_{2.5}$ for five selected grid cells. The global sensitivity conducted for monthly-averaged values showed seasonal trends in the type of input emissions that drive uncertainty in the surface concentrations.

The ability to estimate uncertainty in the predictions produced by a model is vital, because even low levels of uncertainty could be important in areas where the model yields predictions of surface concentrations that are close to limit values. This can lead to instances of exceedance due to the binary nature of limit value exceedance calculations, i.e. concentration is either over or under the limit. The sensitivity analysis should be an integral part of the assessment process applied *ex-ante* for the implementation of policy interventions, as it is also important to know which of the inputs contribute to the uncertainty in model outputs the most.

This work has demonstrated a global sensitivity and uncertainty analyses application for a Eulerian ACTM. The emulator-based approach used here is applicable to any other complex model and any type of model inputs such as emissions, physical constants or chemical reaction rate constants. The results of the analyses provide useful insights into the level of confidence in modelled predictions. Additionally, the Gaussian process emulators created for this analysis can be used with very little computational cost for any other scenario exploration purposes or assessment of overall uncertainty given different uncertainty ranges and probability distributions assigned to the model inputs.

Data availability. The EMEP MSC-W model code is available from www.github.com/metno/emep-ctm. The uncertainty and sensitivity data presented in this paper are available from https://doi.org/10.5281/zenodo.2213633 together with the analysis scripts.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. Ksenia Aleksankina acknowledges studentship funding from the University of Edinburgh and the NERC Centre for Ecology & Hydrology. The work by Stefan Reis and Massimo Vieno has been funded under the Natural Environment Research Council National Capability funding stream (Grant Reference ceh020011; Atmospheric Chemistry and Effects) and the Natural Environment Research Council Long-Term Science Single Centre funding stream UK-SCaPE.

References

Air Quality Expert Group: Mitigation of United Kingdom PM 2.5 Concentrations. [online] Available from: https://uk-air.defra.gov.uk/assets/documents/reports/cat11/1508060903_DEF-PB14161_Mitigation_of_UK_PM25.pdf (Accessed 15 May 2018), 2013.

Aleksankina, K., Heal, M. R., Dore, A. J., Van Oijen, M. and Reis, S.: Global sensitivity and uncertainty analysis of an atmospheric chemistry transport model: the FRAME model (version 9.15.0) as a case study, Geosci. Model Dev., 11(4), 1653–1664, doi:10.5194/gmd-11-1653-2018, 2018.

Asher, M. J., Croke, B. F. W., Jakeman, A. J. and Peeters, L. J. M.: A review of surrogate models and their application to groundwater modeling, Water Resour. Res., 51(8), 5957–5973, doi:10.1002/2015WR016967, 2015.

Beddows, A. V., Kitwiroon, N., Williams, M. L. and Beevers, S. D.: Emulation and Sensitivity Analysis of the Community Multiscale Air Quality Model for a UK Ozone Pollution Episode, Environ. Sci. Technol., 51(11), 6229–6236, doi:10.1021/acs.est.6b05873, 2017.

Carslaw, K. S., Lee, L. A., Reddington, C. L., Pringle, K. J., Rap, A., Forster, P. M., Mann, G. W., Spracklen, D. V., Woodhouse, M. T., Regayre, L. A. and Pierce, J. R.: Large contribution of natural aerosols to uncertainty in indirect forcing, Nature, 503(7474), 67–71, doi:10.1038/nature12674, 2013.

Castelletti, A., Galelli, S., Ratto, M., Soncini-Sessa, R. and Young, P. C.: A general framework for Dynamic Emulation Modelling in environmental problems, Environ. Model. Softw., 34, 5–18, doi:10.1016/j.envsoft.2012.01.002, 2012.

Chen, S. and Brune, W. H.: Global sensitivity analysis of ozone production and O3–NOx–VOC limitation based on field data, Atmos. Environ., 55, 288–296, doi:10.1016/j.atmosenv.2012.03.061, 2012.

Chen, S., Brune, W. H., Lambe, A. T., Davidovits, P. and Onasch, T. B.: Modeling organic aerosol from the oxidation of α-pinene in a Potential Aerosol Mass (PAM) chamber, Atmos. Chem. Phys., 13(9), 5017–5031, doi:10.5194/acp-13-5017-2013, 2013.

Christian, K. E., Brune, W. H. and Mao, J.: Global sensitivity analysis of the GEOS-Chem chemical transport model: ozone and hydrogen oxides during ARCTAS (2008), Atmos. Chem. Phys., 17(5), 3769–3784, doi:10.5194/acp-17-3769-2017, 2017.

Cohan, D. S., Koo, B. and Yarwood, G.: Influence of uncertain reaction rates on ozone sensitivity to emissions, Atmos. Environ., 44(26), 3101–3109, doi:10.1016/j.atmosenv.2010.05.034, 2010.

Corbett, J. J.: Updated emissions from ocean shipping, J. Geophys. Res., 108(D20), 4650, doi:10.1029/2003JD003751, 2003.

Dean, A., Morris, M., Stufken, J. and Bingham, D.: Handbook of Design and Analysis of Experiments, Chapman and Hall/CRC, New York., 2015.

Dore, A. J., Carslaw, D. C., Braban, C., Cain, M., Chemel, C., Conolly, C., Derwent, R. G., Griffiths, S. J., Hall, J., Hayman, G., Lawrence, S., Metcalfe, S. E., Redington, A., Simpson, D., Sutton, M. A., Sutton, P., Tang, Y. S., Vieno, M., Werner, M. and Whyatt, J. D.: Evaluation of the performance of different atmospheric chemical transport models and inter-comparison of nitrogen and sulphur deposition estimates for the UK, Atmos. Environ., 119(606), 131–143, doi:10.1016/j.atmosenv.2015.08.008, 2015.

EC Directive: Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe, [online] Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32008L0050&from=en (Accessed 1 May 2018), 2008.

Entec: Quantification of emissions from ships associated with ship movements between ports in the European Community. [online] Available from: http://ec.europa.eu/environment/air/pdf/chapter2_ship_emissions.pdf (Accessed 2 May 2018), 2002.

Entec: UK Ship Emissions Inventory. [online] Available from: http://uk-air.defra.gov.uk/reports/cat15/1012131459_21897_Final_Report_291110.pdf (Accessed 1 July 2018), 2010.

Eurostat: NAMEA for Air Emissions Compilation Guide., 2004.

Ferretti, F., Saltelli, A. and Tarantola, S.: Trends in sensitivity analysis practice in the last decade, Sci. Total Environ., 568, 666–670, doi:10.1016/j.scitotenv.2016.02.133, 2015.

Frost, G. J., Middleton, P., Tarrasón, L., Granier, C., Guenther, A., Cardenas, B., Denier van der Gon, H., Janssens-Maenhout, G., Kaiser, J. W., Keating, T., Klimont, Z., Lamarque, J. F., Liousse, C., Nickovic, S., Ohara, T., Schultz, M. G., Skiba, U., Van Aardenne, J. and Wang, Y.: New Directions: GEIA's 2020 vision for better air emissions information, Atmos. Environ., 81(2013), 710–712, doi:10.1016/j.atmosenv.2013.08.063, 2013.

Ghanem, R., Higdon, D. and Owhadi, H.: Handbook of Uncertainty Quantification, edited by R. Ghanem, D. Higdon, and H. Owhadi, Springer International Publishing., 2017.

Gladish, D. W., Pagendam, D. E., Peeters, L. J. M., Kuhnert, P. M. and Vaze, J.: Emulation Engines: Choice and Quantification of Uncertainty for Complex Hydrological Models, J. Agric. Biol. Environ. Stat., 23(1), 39–62, doi:10.1007/s13253-017-0308-3, 2017.

Grini, A., Myhre, G., Zender, C. S. and Isaksen, I. S. A.: Model simulations of dust sources and transport in the global atmosphere: Effects of soil erodibility and wind speed variability, J. Geophys. Res., 110(D2), D02205, doi:10.1029/2004JD005037, 2005.

Guenther, A., Hewitt, C. N., Erickson, D., Fall, R., Geron, C., Graedel, T., Harley, P., Klinger, L., Lerdau, M., Mckay, W. A., Pierce, T., Scholes, B., Steinbrecher, R., Tallamraju, R., Taylor, J. and Zimmerman, P.: A global model of natural volatile organic compound emissions, J. Geophys. Res., 100(D5), 8873, doi:10.1029/94JD02950, 1995.

Guenther, A. B., Zimmerman, P. R., Harley, P. C., Monson, R. K. and Fall, R.: Isoprene and monoterpene emission rate variability: Model evaluations and sensitivity analyses, J. Geophys. Res., 98(D7), 12609, doi:10.1029/93JD00527, 1993.

Hanna, S. R., Paine, R., Heinold, D., Kintigh, E. and Baker, D.: Uncertainties in air toxics calculated by the dispersion models AERMOD and ISCST3 in the Houston ship channel area, J. Appl. Meteorol. Climatol., 46(9), 1372–1382, doi:10.1175/JAM2540.1, 2007.

Homma, T. and Saltelli, A.: Importance measures in global sensitivity analysis of nonlinear models, Reliab. Eng. Syst. Saf., 52(1), 1–17, doi:10.1016/0951-8320(96)00002-6, 1996.

Iooss, B. and Lemaître, P.: A Review on Global Sensitivity Analysis Methods, in Uncertainty management in Simulation-Optimization of Complex Systems: Algorithms and Applications, pp. 101–122, Springer US., 2015.

IPCC: IPCC Guidelines for National Greenhouse Gas Inventories, General Guidance and Reporting. [online] Available from: https://www.ipcc-nggip.iges.or.jp/public/2006gl/pdf/1_Volume1/V1_3_Ch3_Uncertainties.pdf (Accessed 2 February 2018), 2006.

IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge, United Kingdom and New York, NY, USA,., 2013.

Janon, A., Klein, T., Lagnoux, A., Nodet, M. and Prieur, C.: Asymptotic normality and efficiency of two Sobol index

estimators, ESAIM Probab. Stat., 18, 342–364, doi:10.1051/ps/2013040, 2014.

Johnson, M. E., Moore, L. M. and Ylvisaker, D.: Minimax and maximin distance designs, J. Stat. Plan. Inference, 26(2), 131–148, doi:10.1016/0378-3758(90)90122-B, 1990.

Jones, B. and Johnson, R. T.: Design and analysis for the Gaussian process model, Qual. Reliab. Eng. Int., 25(5), 515–524, doi:10.1002/qre.1044, 2009.

Lataniotis, C., Marelli, S. and Sudret, B.: Gaussian process modelling using UQLab, [online] Available from: http://arxiv.org/abs/1709.09382, 2017.

Lee, L. A., Carslaw, K. S., Pringle, K. J., Mann, G. W. and Spracklen, D. V.: Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters, Atmos. Chem. Phys., 11(23), 12253–12273, doi:10.5194/acp-11-12253-2011, 2011.

Lin, C., Heal, M. R., Vieno, M., MacKenzie, I. A., Armstrong, B. G., Butland, B. K., Milojevic, A., Chalabi, Z., Atkinson, R. W., Stevenson, D. S., Doherty, R. M. and Wilkinson, P.: Spatiotemporal evaluation of EMEP4UK-WRF v4.3 atmospheric chemistry transport simulations of health-related metrics for NO2, O3, and PM10 for 2001-2010, Geosci. Model Dev., 10(4), 1767–1787, doi:10.5194/gmd-10-1767-2017, 2017.

Luo, Z., Wang, E., Bryan, B. A., King, D., Zhao, G., Pan, X. and Bende-Michl, U.: Meta-modeling soil organic carbon sequestration potential and its application at regional scale, Ecol. Appl., 23(2), 408–420, doi:10.1890/12-0672.1, 2013.

Marelli, S. and Sudret, B.: UQLab: A Framework for Uncertainty Quantification in Matlab, in Vulnerability, Uncertainty, and Risk, pp. 2554–2563, American Society of Civil Engineers, Reston, VA., 2014.

Mårtensson, E. M., Nilsson, E. D., de Leeuw, G., Cohen, L. H. and Hansson, H.-C.: Laboratory simulations and parameterization of the primary marine aerosol production, J. Geophys. Res. Atmos., 108(D9), n/a-n/a, doi:10.1029/2002JD002263, 2003.

McKay, M. D., Beckman, R. J. and Conover, W. J.: Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, Technometrics, 21(2), 239–245, doi:10.1080/00401706.1979.10489755, 1979.

Monahan, E. C., Spiel, D. E. and Davidson, K. L.: A Model of Marine Aerosol Generation Via Whitecaps and Wave Disruption, pp. 167–174, Springer, Dordrecht., 1986.

Moreno-Gutiérrez, J., Calderay, F., Saborido, N., Boile, M., Rodríguez Valero, R. and Durán-Grados, V.: Methodologies for estimating shipping emissions and energy consumption: A comparative analysis of current methods, Energy, 86, 603–616, doi:10.1016/j.energy.2015.04.083, 2015.

Morris, M. D. and Mitchell, T. J.: Exploratory designs for computational experiments, J. Stat. Plan. Inference, 43(3), 381–402, doi:10.1016/0378-3758(94)00035-T, 1995.

NCEP: NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, CO., 2000.

O'Hagan, A.: Bayesian analysis of computer code outputs: A tutorial, Reliab. Eng. Syst. Saf., 91(10–11), 1290–1300, doi:10.1016/j.ress.2005.11.025, 2006.

Oakley, J. and O'Hagan, A.: Bayesian inference for the uncertainty distribution of computer model outputs, Biometrika, 89(4),

769–784, 2002.

Oakley, J. E. and O'Hagan, A.: Probabilistic sensitivity analysis of complex models: A Bayesian approach, J. R. Stat. Soc. Ser. B Stat. Methodol., 66(3), 751–769, doi:10.1111/j.1467-9868.2004.05304.x, 2004.

Parry, H. R., Topping, C. J., Kennedy, M. C., Boatman, N. D. and Murray, A. W. A.: A Bayesian sensitivity analysis applied to an Agent-based model of bird population response to landscape change, Environ. Model. Softw., 45, 104–115, doi:10.1016/j.envsoft.2012.08.006, 2013.

Pulles, T. and Kuenen, J.: EMEP/EEA air pollutant emission inventory guidebook. [online] Available from: https://www.eea.europa.eu/publications/emep-eea-guidebook-2016, 2016.

Rabitz, H. and Alış, Ö. F.: General foundations of high-dimensional model representations, J. Math. Chem., 25(2–3), 197–233, doi:10.1023/A:1019188517934, 1999.

Ratto, M., Castelletti, A. and Pagano, A.: Emulation techniques for the reduction and sensitivity analysis of complex environmental models, Environ. Model. Softw., 34, 1–4, doi:10.1016/j.envsoft.2011.11.003, 2012.

Rodriguez, M. A., Brouwer, J., Samuelsen, G. S. and Dabdub, D.: Air quality impacts of distributed power generation in the South Coast Air Basin of California 2: Model uncertainty and sensitivity analysis, Atmos. Environ., 41(27), 5618–5635, doi:10.1016/j.atmosenv.2007.02.049, 2007.

Rypdal, K. and Winiwarter, W.: Uncertainties in greenhouse gas emission inventories — evaluation, comparability and implications, Environ. Sci. Policy, 4(2–3), 107–116, doi:10.1016/S1462-9011(00)00113-1, 2001.

Saltelli, A.: Making best use of model evaluations to compute sensitivity indices, Comput. Phys. Commun., 145, 280–297, doi:10.1016/S0010-4655(02)00280-1, 2002.

Saltelli, A. and Annoni, P.: How to avoid a perfunctory sensitivity analysis, Environ. Model. Softw., 25(12), 1508–1517, doi:10.1016/j.envsoft.2010.04.012, 2010.

Saltelli, A., Ratto, M., Tarantola, S. and Campolongo, F.: Sensitivity analysis practice: A guide to scientific models., 2006.

Sax, T. and Isakov, V.: A case study for assessing uncertainty in local-scale regulatory air quality modeling applications, Atmos. Environ., 37(25), 3481–3489, doi:10.1016/S1352-2310(03)00411-4, 2003.

Scarbrough, T., Tsagatakis, I., Smith, K., Wakeling, D., Smith, T., Hauerhoff, E. and Murrells, T.: A review of the NAEI shipping emissions methodology. [online] Available from: https://uk-air.defra.gov.uk/assets/documents/reports/cat07/1712140936_ED61406_NAEI_shipping_report_12Dec2017.pdf (Accessed 15 March 2018), 2017.

Sheikholeslami, R. and Razavi, S.: Progressive Latin Hypercube Sampling: An efficient approach for robust sampling-based analysis of environmental models, Environ. Model. Softw., 93(2017), 109–126, doi:10.1016/j.envsoft.2017.03.010, 2017.

Shrivastava, M., Zhao, C., Easter, R. C., Qian, Y., Zelenyuk, A., Fast, J. D., Liu, Y., Zhang, Q. and Guenther, A.: Sensitivity analysis of simulated SOA loadings using a variance-based statistical approach, J. Adv. Model. Earth Syst., 8(2), 499–519, doi:10.1002/2015MS000554, 2016.

Simpson, D., Benedictow, a., Berge, H., Bergström, R., Emberson, L. D., Fagerli, H., Flechard, C. R., Hayman, G. D., Gauss, M., Jonson, J. E., Jenkin, M. E., Nyíri, a., Richter, C., Semeena, V. S., Tsyro, S., Tuovinen, J.-P., Valdebenito, Á. and Wind,

P.: The EMEP MSC-W chemical transport model – technical description, Atmos. Chem. Phys., 12(16), 7825–7865, doi:10.5194/acp-12-7825-2012, 2012.

Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Barker, D., Duda, M., Huang, X., Wang, W. and Powers, J.: A Description of the Advanced Research WRF Version 3. NCAR technical note NCAR/TN-475+STR., 2008.

Sobol', I. .: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Math. Comput. Simul., 55(1–3), 271–280, doi:10.1016/S0378-4754(00)00270-6, 2001.

Sobol', I. M.: Sensitivity estimates for nonlinear mathematical models, Math. Model. Comput. Exp., 1(4), 407–414, 1993.

Stevenson, D. S., Young, P. J., Naik, V., Lamarque, J. F., Shindell, D. T., Voulgarakis, A., Skeie, R. B., Dalsoren, S. B., Myhre, G., Berntsen, T. K., Folberth, G. A., Rumbold, S. T., Collins, W. J., MacKenzie, I. A., Doherty, R. M., Zeng, G., Van Noije, T. P. C., Strunk, A., Bergmann, D., Cameron-Smith, P., Plummer, D. A., Strode, S. A., Horowitz, L., Lee, Y. H., Szopa, S., Sudo, K., Nagashima, T., Josse, B., Cionni, I., Righi, M., Eyring, V., Conley, A., Bowman, K. W., Wild, O. and Archibald, A.: Tropospheric ozone changes, radiative forcing and attribution to emissions in the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP), Atmos. Chem. Phys., 13(6), 3063–3085, doi:10.5194/acp-13-3063-2013, 2013.

Storlie, C. B. and Helton, J. C.: Multiple predictor smoothing methods for sensitivity analysis: Description of techniques, Reliab. Eng. Syst. Saf., 93(1), 28–54, doi:10.1016/J.RESS.2006.10.012, 2008.

Storlie, C. B., Swiler, L. P., Helton, J. C. and Sallaberry, C. J.: Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models, Reliab. Eng. Syst. Saf., 94(11), 1735–1763, doi:10.1016/j.ress.2009.05.007, 2009.

Sudret, B.: Global sensitivity analysis using polynomial chaos expansions, Reliab. Eng. Syst. Saf., 93(7), 964–979, doi:10.1016/j.ress.2007.04.002, 2008.

Teixeira, E., Fischer, G., van Velthuizen, H., van Dingenen, R., Dentener, F., Mills, G., Walter, C. and Ewert, F.: Limited potential of crop management for mitigating surface ozone impacts on global food supply, Atmos. Environ., 45(15), 2569–2576, doi:10.1016/j.atmosenv.2011.02.002, 2011.

Urban, N. M. and Fricker, T. E.: A comparison of Latin hypercube and grid ensemble designs for the multivariate emulation of an Earth system model, Comput. Geosci., 36(6), 746–755, doi:10.1016/j.cageo.2009.11.004, 2010.

Vieno, M., Dore, a. J., Stevenson, D. S., Doherty, R., Heal, M. R., Reis, S., Hallsworth, S., Tarrason, L., Wind, P., Fowler, D., Simpson, D. and Sutton, M. a.: Modelling surface ozone during the 2003 heat-wave in the UK, Atmos. Chem. Phys., 10(16), 7963–7978, doi:10.5194/acp-10-7963-2010, 2010.

Vieno, M., Heal, M. R., Hallsworth, S., Famulari, D., Doherty, R. M., Dore, a. J., Tang, Y. S., Braban, C. F., Leaver, D., Sutton, M. a. and Reis, S.: The role of long-range transport and domestic emissions in determining atmospheric secondary inorganic particle concentrations across the UK, Atmos. Chem. Phys., 14(16), 8435–8447, doi:10.5194/acp-14-8435-2014, 2014.

Vieno, M., Heal, M. R., Williams, M. L., Carnell, E. J., Nemitz, E., Stedman, J. R. and Reis, S.: The sensitivities of emissions reductions for the mitigation of UK PM2.5, Atmos. Chem. Phys., 16(1), 265–276, doi:10.5194/acp-16-265-2016, 2016a.

Vieno, M., Heal, M. R., Twigg, M. M., MacKenzie, I. A., Braban, C. F., Lingard, J. J. N., Ritchie, S., Beck, R. C., Móring, A.,

Ots, R., Di Marco, C. F., Nemitz, E., Sutton, M. A. and Reis, S.: The UK particulate matter air pollution episode of March–April 2014: more than Saharan dust, Environ. Res. Lett., 11(4), 044004, doi:10.1088/1748-9326/11/4/044004, 2016b.

Wakeling, D., Passant, N. R., Murrells, T. P., Pang, Y., Thistlethwaite, G., Walker, C., Garcia, J. M. R., Webb, J., Brown, P., Vento, S. del, Misra, A., Hobson, M., Pridmore, A., Dore, C. and Misselbrook, T.: UK Informative Inventory Report (1990 to 2014), 2016.

Wakeling, D., Passant, N., Murrells, T., Pang, Y., Thistlethwaite, G., Walker, C., Brown, P., Vento, S. del, Hunter, R., Wiltshire, J., Broomfield, M., Watterson, J., Pearson, B., Rushton, K., Hobson, M., Smith, H. and Misselbrook, T.: UK Informative Inventory Report (1990 to 2015)., 2017.

WHO: Air quality guidelines. Global update 2005. Particulate matter, ozone, nitrogen dioxide and sulfur dioxide, World Health Organisation Regional Office for Europe, Copenhagen., 2006.

WHO: Health risks of air pollution in Europe – HRAPIE project, Copenhagen. [online] Available from: http://www.euro.who.int/__data/assets/pdf_file/0017/234026/e96933.pdf?ua=1 (Accessed 2 December 2017), 2013.

Yang, J.: Convergence and uncertainty analyses in Monte-Carlo based sensitivity analysis, Environ. Model. Softw., 26(4), 444–457, doi:10.1016/j.envsoft.2010.10.007, 2011.

Ziehn, T. and Tomlin, A. S.: GUI–HDMR – A software tool for global sensitivity analysis of complex models, Environ. Model. Softw., 24(7), 775–785, doi:10.1016/j.envsoft.2008.12.002, 2009.

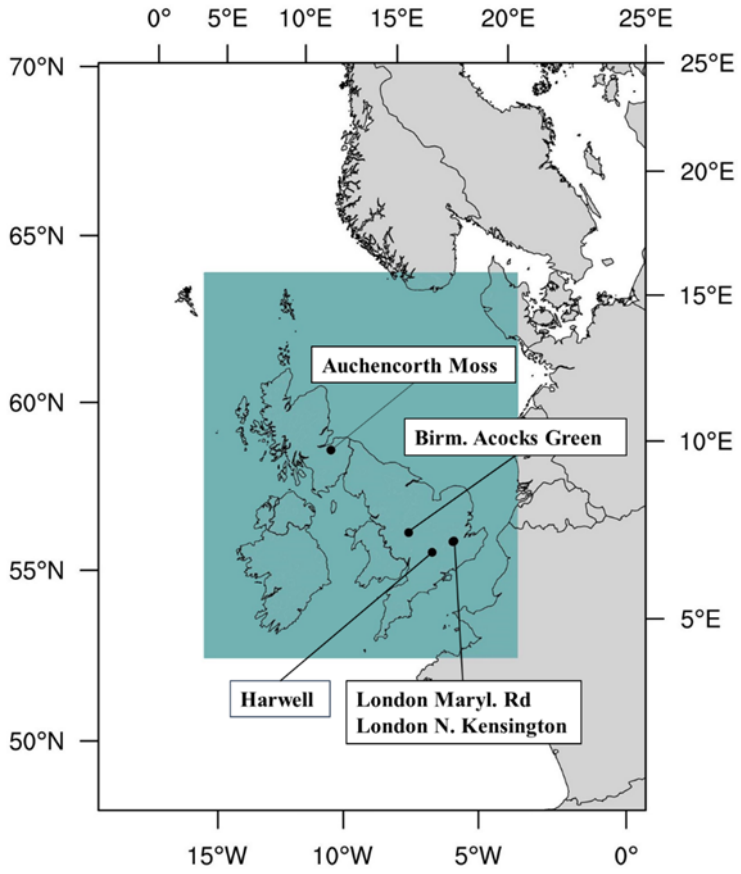

Figure 1 The inner shaded box illustrates the EMEP4UK model British Isles domain, which is modelled at 5 km× 5 km horizontal resolution. The location of five grid cells used for uncertainty quantification and sensitivity analysis for monthly average modelled concentrations of O_3, NO_2, and $PM_{2.5}$ are shown.

5

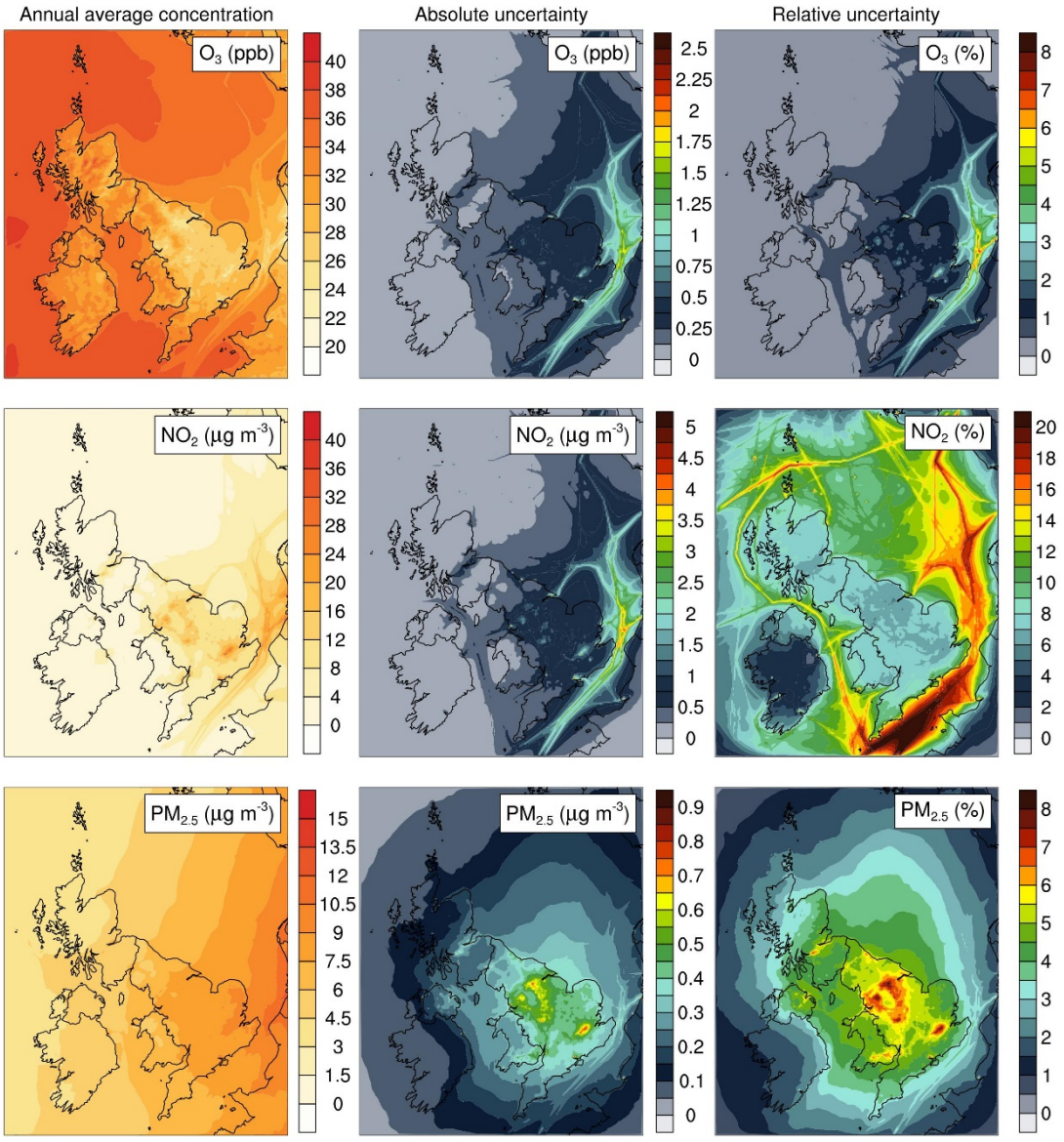

Figure 2 Baseline surface concentrations of O_3, NO_2, and $PM_{2.5}$, and their respective spatial distributions of the absolute and relative uncertainties (at the 5 km × 5 km model grid resolution, year 2012) for the specified uncertainties in UK emissions. The uncertainty values are represented as a range of ± the baseline value and represent the 95% confidence interval.

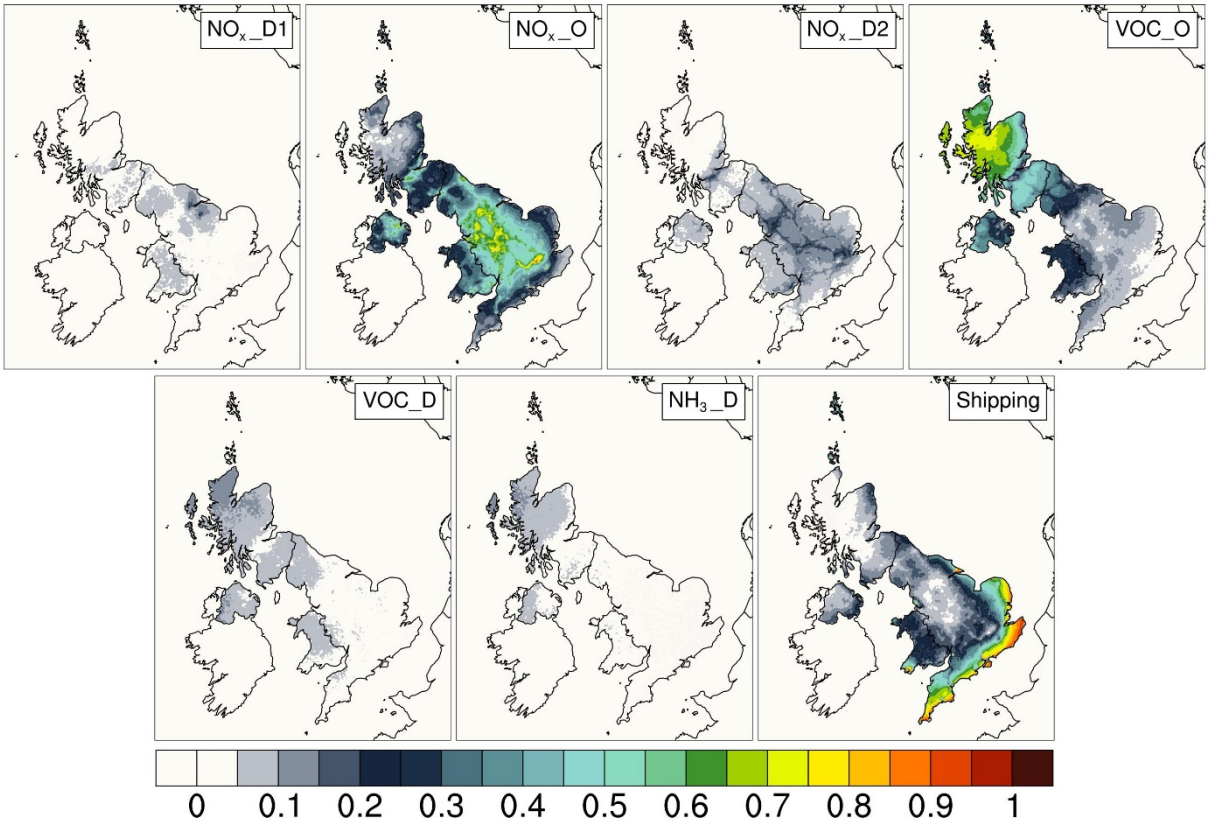

Figure 3 Spatial distributions (at the 5km×5km model grid resolution) of the first-order sensitivity indices for modelled surface concentrations of O_3. D indicates emissions from a dominant sector and O indicates grouped emissions from the rest of the sectors. For NO_x emissions dominant sectors are energy production (D1) and road transport (D2), for VOC emissions – solvent use, and for NH_3 – agriculture. Shipping emissions variable combines emissions of all relevant pollutants.

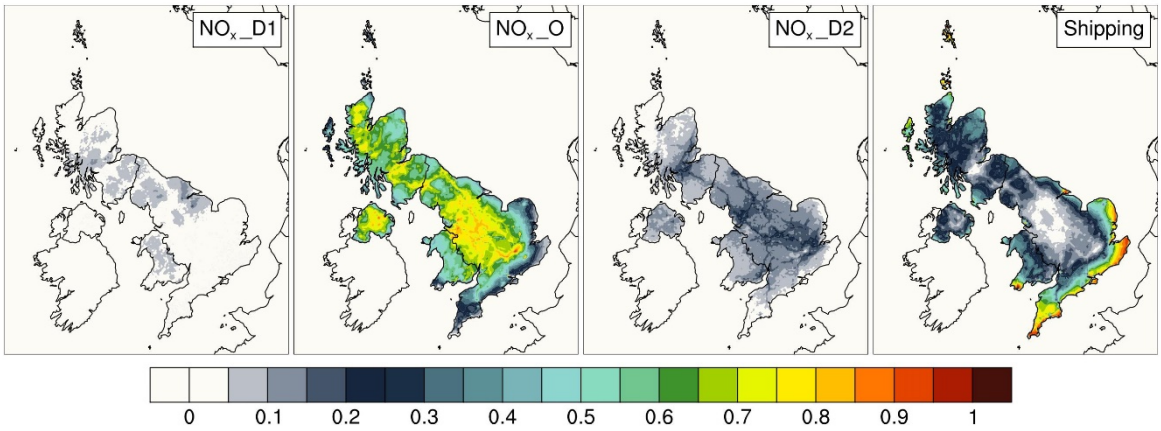

Figure 4 Spatial distributions (at the 5km×5km model grid resolution) of the first-order sensitivity indices for modelled surface concentrations of NO_2. D indicates emissions from a dominant sector and O indicates grouped emissions from the rest of the sectors. For NO_x emissions dominant sectors are energy production (D1) and road transport (D2). Shipping emissions variable combines emissions of all relevant pollutants.

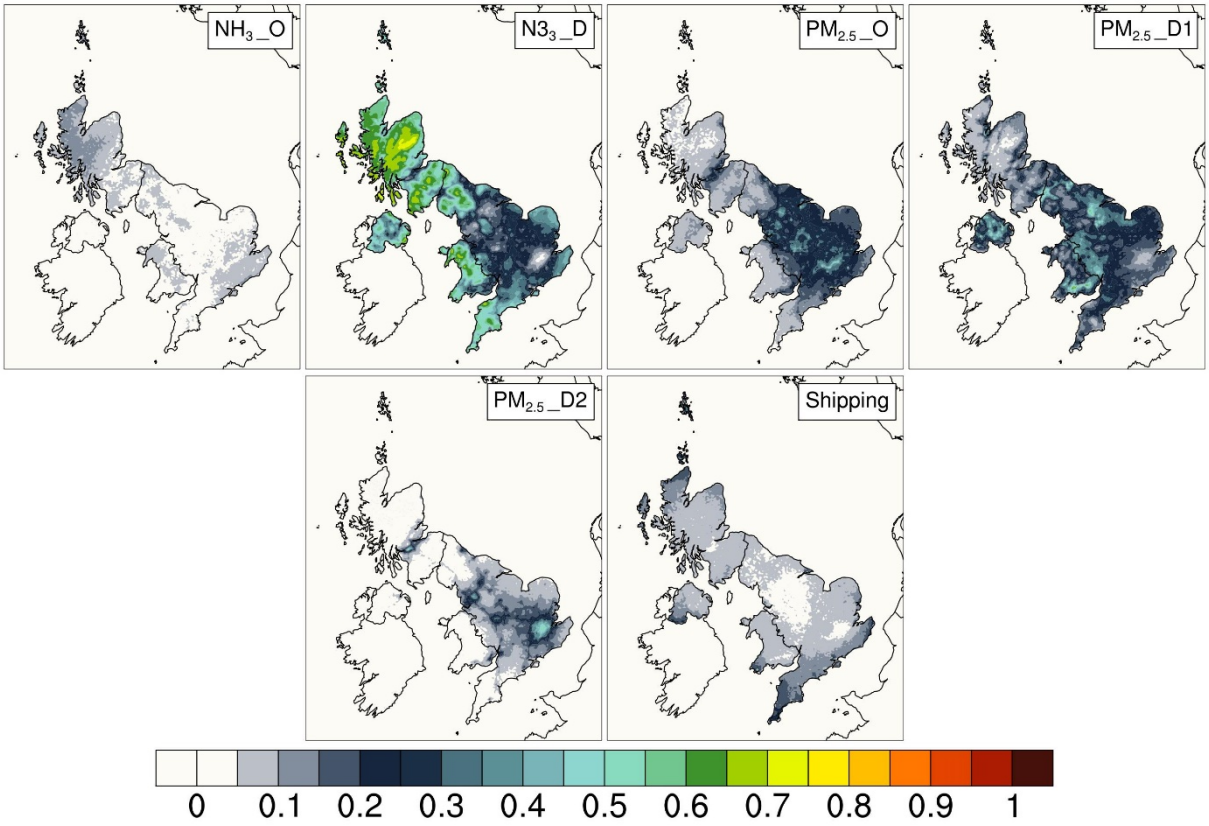

Figure 5 Spatial distributions (at the 5km×5km model grid resolution) of the first-order sensitivity indices for modelled surface concentrations of $PM_{2.5}$. D indicates emissions from a dominant sector and O indicates grouped emissions from the rest of the sectors. For NH_3 emissions dominant sector is agriculture, for $PM_{2.5}$ dominant sectors are residential and non-industrial combustion (D1) and road transport (D2). Shipping emissions variable combines emissions of all relevant pollutants.

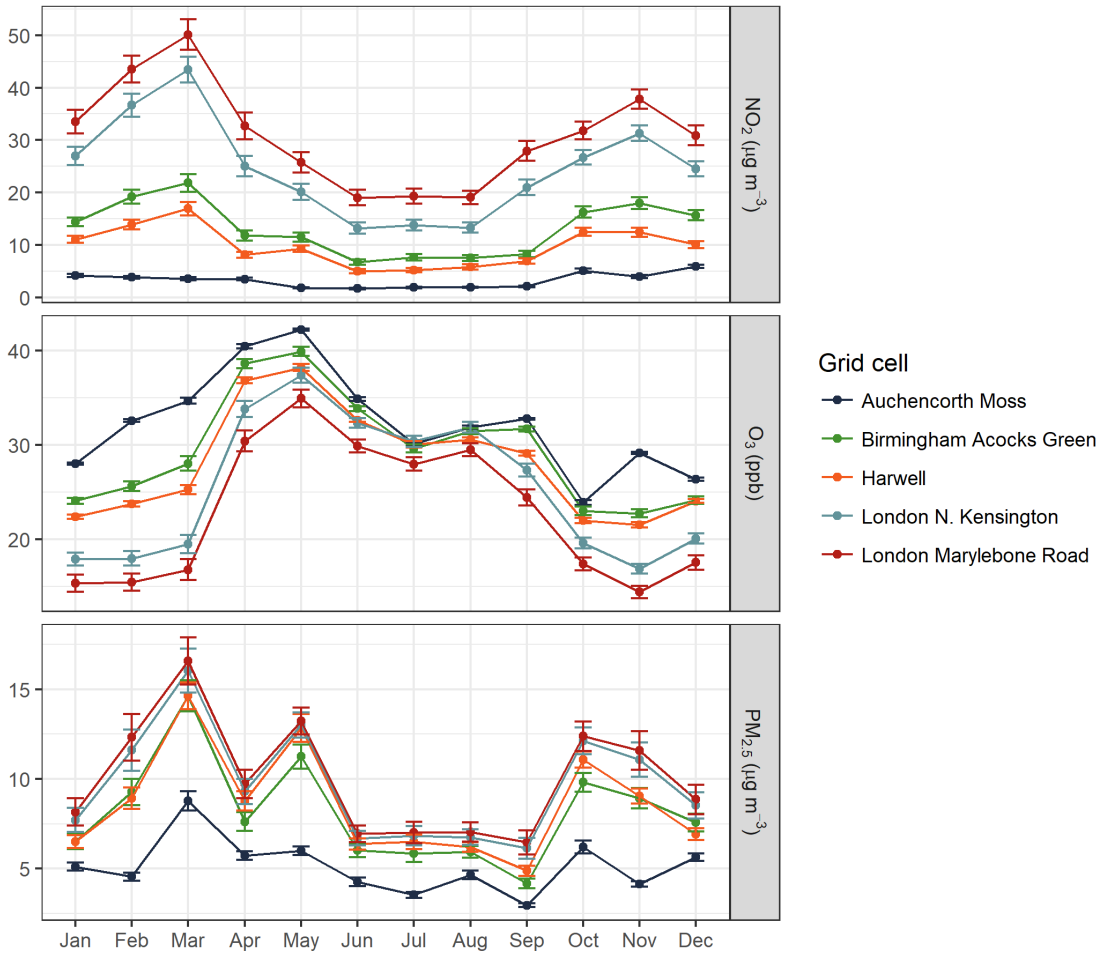

Figure 6 Monthly average surface concentrations of NO_2, O_3 and $PM_{2.5}$ with error bars showing (absolute) uncertainty, for five grid cells across the UK representing a spread of geographical locations and environment types. The environment types are assigned as follows: Auchencorth Moss and Harwell - rural background, Birmingham Acocks Green and London N. Kensington - urban background, and London Marylebone Road - urban traffic.

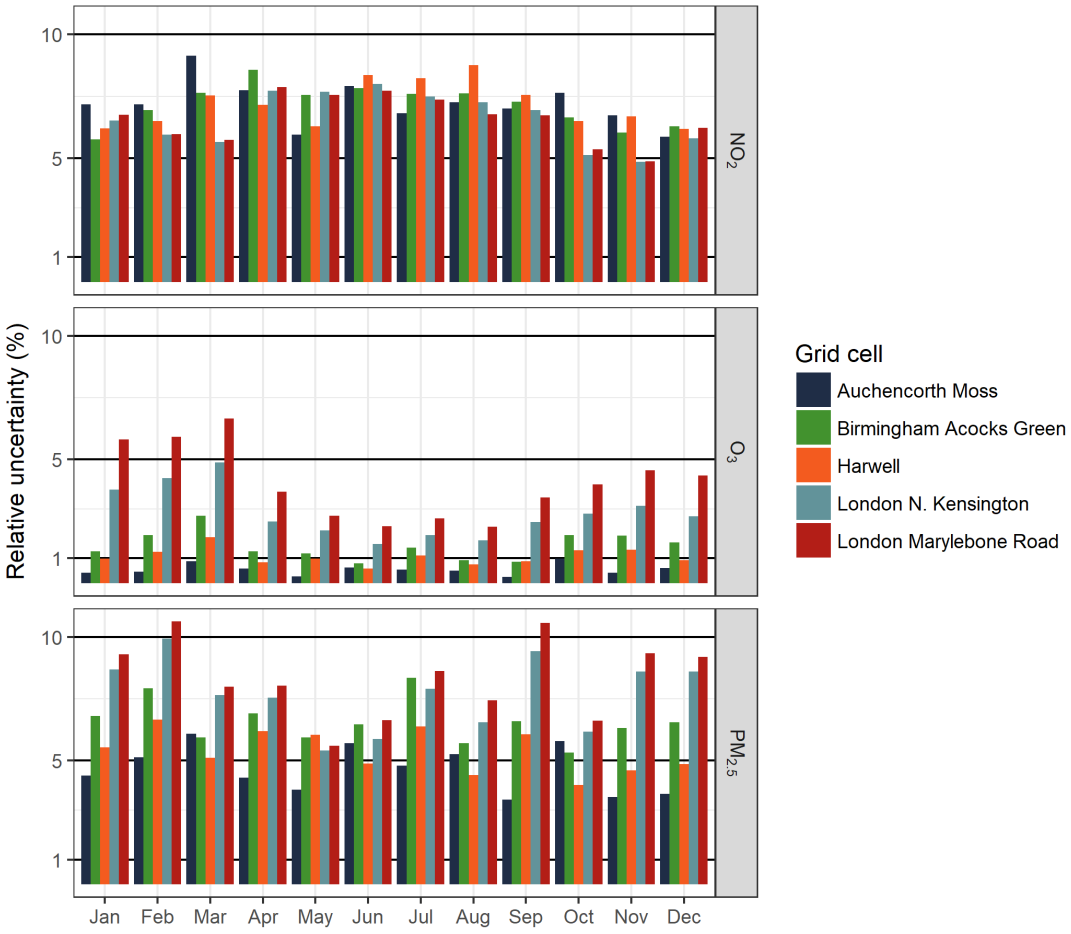

Figure 7 Magnitude of relative uncertainty in monthly average surface concentrations of NO_2, O_3, and $PM_{2.5}$ for five grid cells across the UK representing a spread of geographical locations and environment types. The environment types are assigned as follows: Auchencorth Moss and Harwell - rural background, Birmingham Acocks Green and London N. Kensington - urban background, and London Marylebone Road - urban traffic.

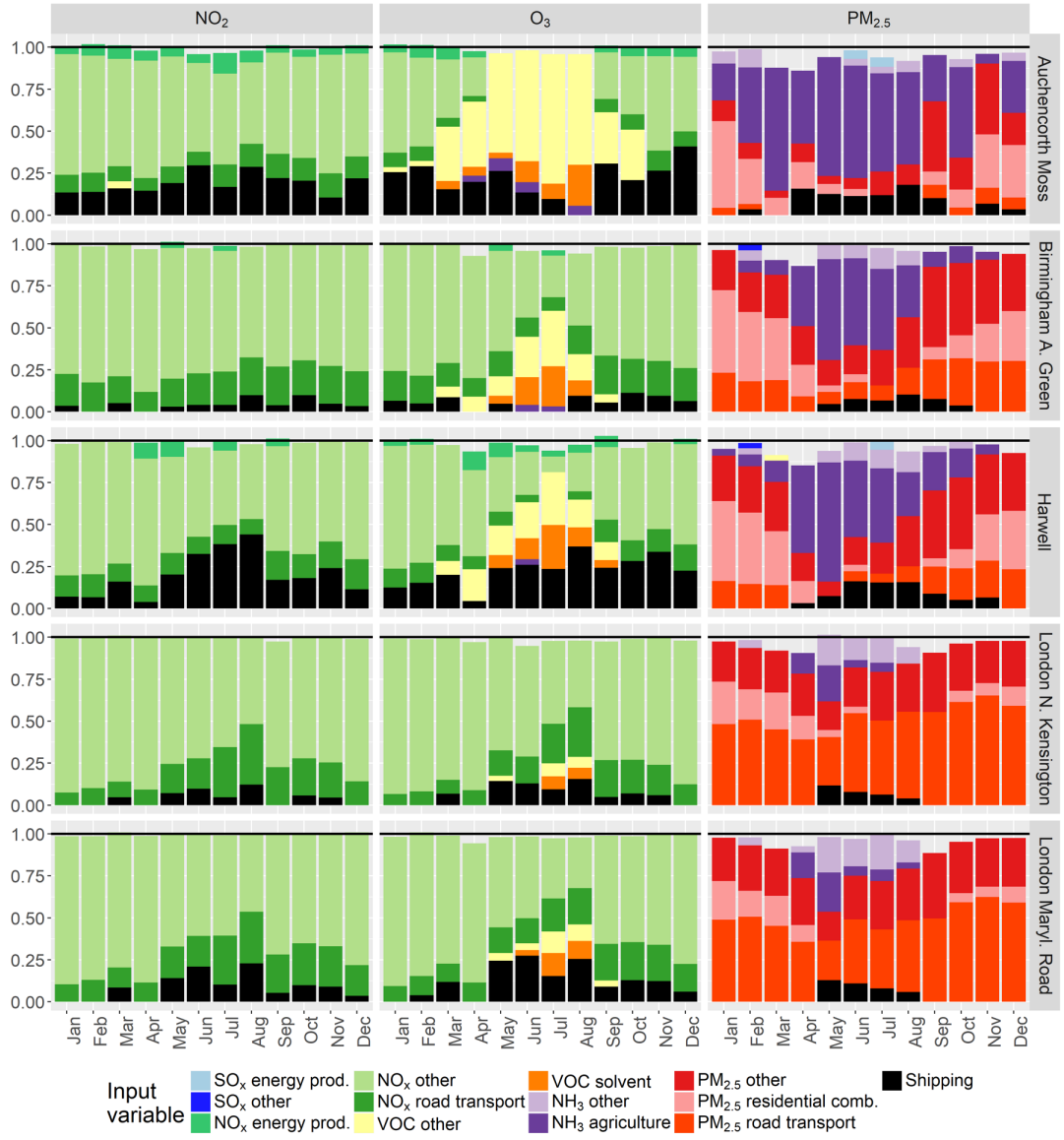

Figure 8 Monthly variation in the first-order sensitivity indices for five grid cells across the UK representing a spread of geographical locations and environment types. Based on the monitoring station classification grid squares are assigned the following environment types: Auchencorth Moss and Harwell - rural background, Birmingham Acocks Green and London N. Kensington - urban background, and London Marylebone Road - urban traffic.

Table 1 SNAP source sectors (Eurostat, 2004).

SNAP 1	Combustion in energy and transformation industries
SNAP 2	Residential and non-industrial combustion
SNAP 3	Combustion in manufacturing industry
SNAP 4	Production processes
SNAP 5	Extraction and distribution of fossil fuels
SNAP 6	Solvent and other product use
SNAP 7	Road transport
SNAP 8	Other mobile sources and machinery
SNAP 9	Waste treatment and disposal
SNAP 10	Agriculture

Table 2 Input variable definitions for the EMEP4UK uncertainty propagation and apportionment. The quoted uncertainties for emission sources are for UK annual totals. See main text for information on the sources of these values.

Variable used for sampling design	SNAP source sector	Contribution of source sector to total land-based emissions of that pollutant (%)	Uncertainty (as a 95% CI)	Ranges of scaling coefficients for the input emissions used in the LHS design
SO_x_D	1	80	± 12 %	0.6 – 1.4
SO_x_O	2-10	20	± 17 %	0.6 – 1.4
NO_x_D1	1	41	± 7 %	0.6 – 1.4
NO_x_D2	7	32	± 7 %	0.6 – 1.4
NO_x_O	2-6, 8-10	27	± 19 %	0.6 – 1.4
VOC_D	6	39	± 22 %	0.6 – 1.4
VOC_O	1-5, 7-10	61	± 24 %	0.6 – 1.4
NH_3_D	10	88	± 33 %	0.6 – 1.4
NH_3_O	1-9	12	± 35 %	0.6 – 1.4
$PM_{2.5}$_D1	2	33	± 59 %	0.25 – 1.75
$PM_{2.5}$_D2	7	21	± 59 %	0.25 – 1.75
$PM_{2.5}$_O	1, 3-6, 8-10	46	± 58 %	0.25 – 1.75
Shipping	N/A	N/A	± 30 %	0.6 – 2.0