UNCERTAINTY MAPPING FOR AIR QUALITY MODELLING AND DATA ASSIMILATION

Bruce Denby[1], Ana Margarida Costa[2], Alexandra Monteiro[2], Agnes Dudek[1], Sam Erik Walker[1], Sef van den Elshout[3] and Bernard Fisher[4]
[1] Norwegian Institute for Air Research (NILU), Kjeller, Norway
[2] CESAM & Department of Environment and Planning, University of Aveiro, Portugal
[3] DCMR EPA Rijnmond, Schiedam, The Netherlands
[4] Environmental Agency, Reading, UK

INTRODUCTION

There are a number of methodologies available for assessing air quality model uncertainty. For the most part these involve direct comparison between models and observation. However, since models are generally applied to produce spatially resolving maps, even at points without observations, it is equally important to indicate model uncertainty as a spatial map. The method used and the uncertainty parameter displayed may vary between applications. Uncertainty maps may be used to indicate the quality of the model, to support decisions on the placement of monitoring sites, for data assimilation purposes or as information for decision makers in integrated risk assessment.

Air quality models produce temporal and/or spatial concentration data based on various inputs. Air quality models may be prognostic, statistical or a combination of these in nature. Data assimilation also involves the use of Bayesian statistical methods to combine models and observations. There are, as a result, a large number of methods available for producing air quality maps and all of these have, in principle, an associated uncertainty. It is thus useful to use an uncertainty parameter that can be used for almost all applications and methodologies so that mapping uncertainties are comparable.

This paper looks at uncertainty parameters and the methodologies for spatially calculating these in air quality models, based on experience gained in a number of case studies during the EU FP6 project Air4EU (www.air4eu.nl). For air quality purposes uncertainty is rarely, if ever, presented in the form of maps but some examples do exist. *Van de Kassteele and Velders* (2006) present maps of standard deviation and probability of exceedance for NO_2, *Fuentes and Raftery* (2005) present maps of bias and standard deviation for SO_2. *Horalek et al.* (2007) present maps of standard deviation derived from the residual kriging variance. To promote the use of uncertainty maps the Air4EU project developed a mapping tool (www.air4eumaps.info) that displays a number of assessment and uncertainty maps.

SELECTION OF UNCERTAINTY PARAMETERS FOR USE IN MAPPING

There are a number of statistical parameters used for the assessment of model uncertainty that may be appropriate for spatial mapping of uncertainty. These may include EC directive related uncertainty assessment parameters (e.g. *EC*, 1999) such as relative maximum error (RME) or its alternative relative percentile error (RPE) (*Stern and Flemming*, 2004). More statistically based error estimates such as standard deviation (SD), root mean square error (RMSE), bias (BIAS) or coefficient of determination (r^2) may also be applied. Discussions and applications concerning these can be found in *Borrego et al*. (2007).

When considering uncertainty there are basically two different types that can be considered. The *intrinsic* model uncertainty and the *predictive* model uncertainty. The first of these refers to the models uncertainty in terms of input data, model formulation and numerical description. The second type, predictive model uncertainty, refers to the models ability to predict a measurement made at some point in space and will include both the intrinsic uncertainty as well as the uncertainty due to spatial and temporal representativeness. Representativeness uncertainty can be large when the scale of spatial variation is smaller than the model resolution.

For uncertainty mapping a general concept is useful. For this purpose Bayesian approaches using probability distributions, represented by probability density functions (PDFs), provide a basis for describing the uncertainty. Associated with PDFs are the typical uncertainty parameters of SD and BIAS that can be used to describe the uncertainty in a normally distributed PDF. The assumption of normal distributions is often not met and other distributions, such as log-normal or more general discretised distributions, may be more appropriate. The SD uncertainty concept may still be applied for these though the description is more complex.

Relative maximum and percentile error

The present European legislation defines the Modelling Quality Objectives as an acceptability measure to guarantee good model performance and reliable modelling results for decision makers. In this context, the uncertainty for modelling and objective estimation is defined and should be estimated in each air quality modelling assessment. When modelling uncertainty is required for directive purposes then the most relevant parameter is RPE for daily or hourly averages and RME for annual means, evaluated at monitoring sites (*Air4EU*, 2007; M2).

Standard deviation

In general it is recommended to use uncertainty parameters that are indicative of the SD of the assessment uncertainty. This uncertainty indicator is thus useful because it is similar to other uncertainty parameters that may be calculated such as RMSE, kriging variance, SD resulting from Monte Carlo ensembles, or covariance matrices required for data assimilation techniques. However, SD is a good uncertainty indicator only when the system is unbiased. It is thus also useful to indicate uncertainty in maps using bias as well.

Bias

Bias can be directly calculated at observational points or interpreted from linear regression. Bias can be removed from models by regression or spatially through optimal interpolation or kriging techniques. Though the mapping of bias may be useful for understanding the model system it is recommended to try to remove bias, through one of these methods, or through improvements in the model description.

Probability distribution functions

In their most general form PDFs can be used to indicate uncertainty. When the PDF is normally distributed then this can be replaced by the uncertainty parameters of SD and BIAS. This will generally not be the case so it is possible to construct discrete PDFs based on direct comparisons with observations. This is achieved by calculating the relative frequency distribution, i.e. by binning the ratio of observed to modelled concentrations *(O/M)*. These discrete PDFs can be used to estimate confidence intervals, SD or probability of exceedance.

SPATIAL DISTRIBUTION OF THE UNCERTAINTY PARAMETERS

In this section methodologies are described that will allow the spatial representation of the uncertainty with emphasis on the representation of SD.

Indicative uncertainty maps

This type of uncertainty map treats the spatial uncertainty, using SD as the uncertainty parameter, as a function of the absolute uncertainty s_A, the relative uncertainty s_R and the model concentration *M(x,y)* in the following form

$$s_M(x,y) = \sqrt{s_A^2 + s_R^2 M(x,y)^2} \qquad (1)$$

The estimated values of s_R and s_A may be calculated in various ways, e.g. based on experience with the model, based on the normalised RMSE, based on the SD of PDFs or by fitting equation 1 to scatter plots representing SD as a function of model concentration.

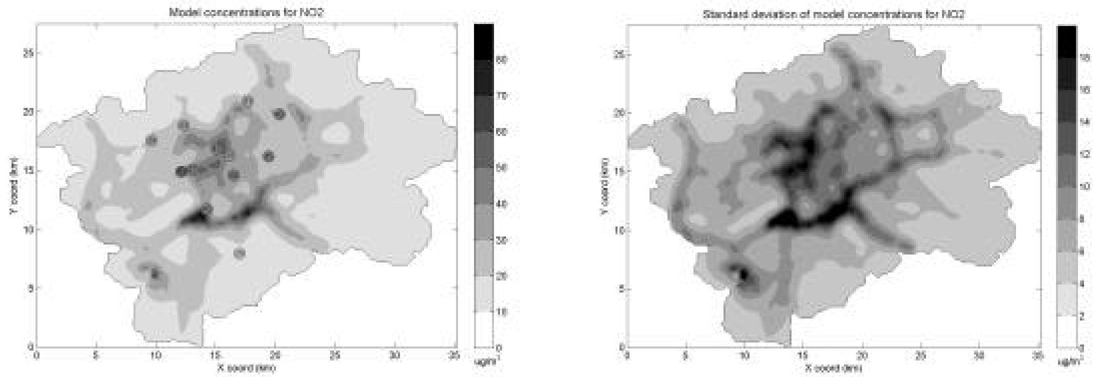

Figure 1: Example of an assessment (left) and uncertainty map (right) of the annual mean NO_2 concentrations in Prague, 2003 (Air4EU, 2007; D7.1.6). The uncertainty map is constructed using s_R based on the normalised RMSE of 12 stations where s_R = 27%.

Spatial representation of model error

When model assessment is carried out at spatially distributed monitoring sites, model error can be determined at these points. This error may be represented as points on the map or, when there is a sufficient density of stations and the pollutant varies on scales larger than the typical distance between stations, it may be interpolated to represent the spatial uncertainty of the model. Such an interpolation will tend to reflect the spatial positions of the stations.

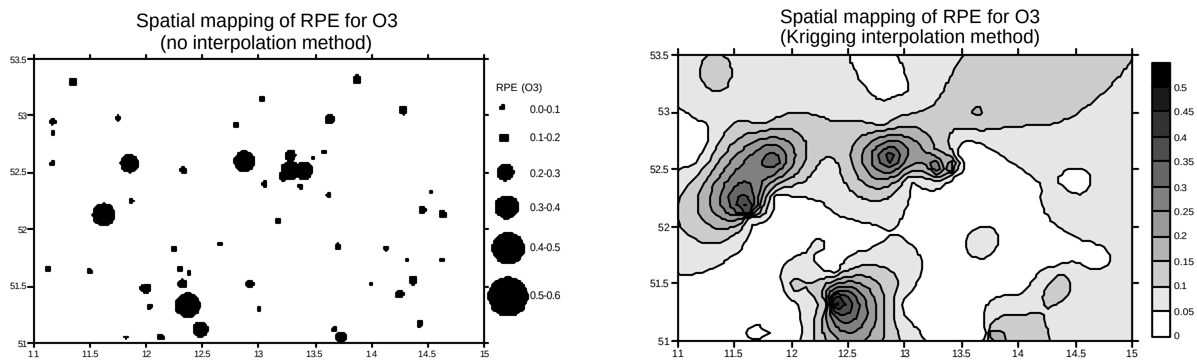

Figure 2: Mapping of RPE of an O_3 simulation (concerning the 26th maximum 8h running average percentile) with the RGC model over Berlin (Air4EU, 2007;D7.1.14). Left: RPE at the 49 individual monitoring sites is represented by the size of the circles. Right: RPE is interpolated using kriging in the same domain.

Maps based on data assimilation, statistical and kriging variance

When data assimilation or other statistically based methods are used, for example linear regression analysis, residual kriging, variational methods or ensemble techniques such as Ensemble Kalman filters, estimates of the variance/covariance are required for the assimilation technique and can be used in the uncertainty maps.

Accounting for temporal covariance in uncertainty mapping

For many modelling and data assimilation applications uncertainty may be calculated on a higher temporal resolution, e.g. hourly or daily, than the final assessment, e.g. annual, requires. To estimate the total uncertainty of the assessment based on the individual hourly or daily spatial uncertainty, the temporal covariance matrix must be determined. Such an assessment was carried out for PM_{10} in Europe (*Air4EU*, 2007; D7.1.13) where the temporal covariance was found to be extremely important for the final uncertainty assessment, due to similar meteorological conditions that lead to correlated concentration distributions.

Accounting for spatial representativeness in the uncertainty mapping of exceedances

For some applications the number of exceedances (NOE) above some threshold value is required. In order to represent the spatial uncertainty in the NOE a pragmatic approach is recommended since aspects related to bias may be involved. The uncertainty in the NOE is calculated by use of the annual mean SD percentile band, i.e. by adding and subtracting the annual mean SD from the daily mean value ($\pm\sigma$), which reflect the model and representativeness bias. The uncertainty in NOE days can be interpreted as being the maximum deviation, in number of days, from the $\pm\sigma$ calculations (*Air4EU*, 2007; D7.1.13).

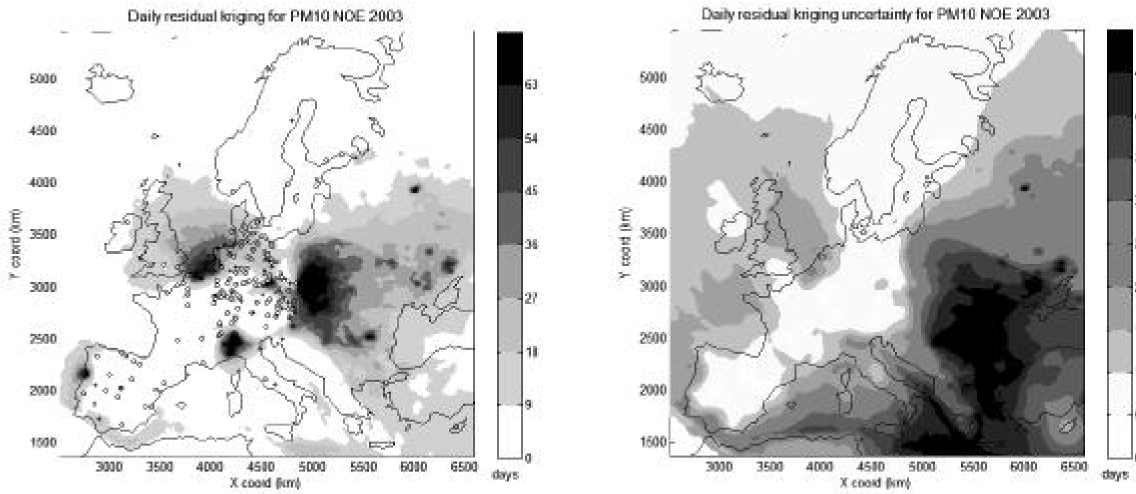

Figure 3. European map of the assimilated (left) and uncertainty (right) fields for the number of daily mean exceedances of PM_{10} above a threshold of 50 µg/m^3, 2003. The uncertainty is calculated using the annual mean percentile band (Air4EU, 2007; D7.1.13).

Model uncertainty using Monte Carlo simulations

Another method for estimating uncertainty is to carry out ensemble runs by perturbing model parameters and input data (within their uncertainty) to obtain estimates of the intrinsic model uncertainty (e.g. *Air4EU*, 2007; D7.1.4). This can be carried out using various Monte Carlo methods where the PDFs of the model parameters are used as prior distributions for the ensembles. This will give a spread of model results that can be mapped. It will not provide information on bias unless direct comparisons are made with observations. As a special case of the Monte Carlo methods some uncertainties can be analysed in a simpler framework. If input parameters, such as emissions, can be simply described by a normal distribution and the model is linear, then the uncertainty can be calculated by simply summing the uncertainties from the various emission sources.

Probability of exceedance

Given both concentration and uncertainty fields, as PDFs or in terms of SD, it is possible to determine the probability of exceedance (POE), given some limit value (LV), by integration of the PDF. POE can be seen as a parameter that includes both calculated concentrations and uncertainty information in the one parameter and may be useful in applications for risk assessment. POE is not an uncertainty parameter and a POE map does not show to what extent the map is determined by the model concentrations or the model uncertainty.

CONCLUSIONS

A number of methodologies for producing uncertainty maps are briefly described. The methods used are dependent on the methodology applied for the creation of the assessment map itself. It is recommended that standard deviation be the major indicative uncertainty parameter in the uncertainty maps. How this is mapped can vary. It can be based on the spatial distribution of the uncertainty, on a functional relationship with concentration through regression or derived by Monte Carlo simulations. Whatever the method applied, an uncertainty assessment of the derived air quality map should always be provided in some form.

ACKNOWLEDGMENTS

This research was funded by the EU FP6 project Air4EU. Contract no. (503596)

REFERENCES

Air4EU (2007): Air4EU deliverable and milestone reports. Available at: http://www.air4eu.nl/reports_products.html

Borrego, C., A. Monteiro, J. Ferreira, A. I. Miranda, A. M. Costa and M. Sousa (2007) Modeling uncertainty estimation procedures for air quality assessment. In proceedings of the 3rd International Symposium on Air Quality Management at Urban, Regional and Global Scales (AQM), 26-30 September 2005; Istanbul, Turkey. Eds. S. Topçu, M.F. Yardim, A. Bayram, T. Elbir and C. Kahya, Vol. I, pp. 210-219.)

EC, 1999: Directive 1999/30/EC of the European Parliament and of the Council relating to limit values for sulphur dioxide, nitrogen dioxide and oxides of nitrogen, particulate matter and lead in ambient air, European Commission, Brussels.

Stern, J. and Flemming, R., 2004: Formulation of criteria to be used for the determination of the accuracy of model calculations according to the requirements of the EU Directives for air quality – Examples using the chemical transport model REM-CALGRID, Freie Universität Berlin, Institut für Meteorologie.

Van de Kassteele, J. and G.J.M Velders (2006). Uncertainty assessment of local NO2 concentrations derived from error-in-variable external drift kriging and its relationship to the 2010 air quality standard. Atmos. Environ*;* 40(14):2583-2595

Fuentes, M. and A.E. Raftery (2005). Model evaluation and spatial interpolation by bayesian combination of observations with outputs from numerical models. Biometrics, 66, 36–45.

Horalek J., B. Denby, P. de Smet, F. de Leeuw, P. Kurfurst, R. Swart and T. van Noije (2007), Spatial mapping of air quality for European scale assessment. ETC/ACC Technical paper 2006/6.