EFFICIENT METHODS FOR ASSESSING UNCERTAINTIES AND SENSITIVITIES IN ENVIRONMENTAL MODELS

Tilo Ziehn and Alison S. Tomlin
Energy and Resources Research Institute (ERRI), University of Leeds, UK

INTRODUCTION

There has been significant progress in the development of suitable modelling strategies for predicting pollution dispersion at the local scale in recent years. A particular feature that has received attention is the influence of local building topologies on the flow and turbulence patterns that are established within networks of urban streets (Dixon et al., 2006). Although empirical models may provide representation of some of these types of features, there is a growing acknowledgement that in order to properly resolve 3-dimensional flow structures, computational fluid dynamics (CFD) models will be required. Such models range from those using detailed representation of turbulent processes such as Large Eddy Simulation (LES), to Reynolds Averaged Navier Stokes (RANS) models which include parameterisations of turbulent mixing e.g. MISKAM (Eichhorn, 1996). It is fair to say that at the present time only RANS provides the rapid simulation times required by operational models. Assumptions made in defining a RANS model may however, have an impact on model predictions. If decisions are to be derived from such predictions, it is of concern whether or not the model can produce reliable results. For example, where an Air Quality Management Area has been defined, strategic plans are drawn up to attempt improve air quality. If the pollution dispersion model used to test these plans predicts a concentration change, it is important to establish whether this change is statistically significant, when including uncertainties in the modelling.

Overall, there is a growing interest in incorporating uncertainty analysis into the overall modelling structure for environmental applications. Due to the computational expense of many models there is a need to develop global sensitivity and uncertainty methods with minimal computational requirements which are capable of determining sensitivity indices that can be used for importance ranking of potentially large numbers of input parameters in an automatic way. These indices should be capable of representing nonlinear responses to changes in input parameters over broad input ranges, as well as parameter interactions. Recently the method of high dimensional model representation (HDMR) (Rabitz et al., 1999) has been developed to provide such global sensitivity estimates. It provides a detailed mapping of the input variable space to selected outputs which is fundamental to global sensitivity analysis. Due to its formulation as a set of hierarchical component functions, it also provides a possibility to determine sensitivity indices in an automatic way that can then be directly used in importance ranking and to explore parameter interactions.

MODEL

A large field study was conducted in 2003 in the City of York (Boddy et al., 2005) with the aim to investigate the influence of background meteorology and building topologies on flow and turbulence patterns within urban street canyons. One canyon under consideration was Gillygate which has an aspect ratio of approximately 0.8 with high traffic flows and significant periods of congestion. The street canyon flow field was simulated using the micro scale k-e model MISKAM in Dixon et al. (2006). MISKAM consists of a 3-dimensional non-hydrostatic flow model and an Eulerian dispersion model. Within the MISKAM code the RANS equations are solved using k-e turbulence closure. Figure 1 shows the grid and the building configuration of Gillygate and the surrounding area that were used for the simulation with MISKAM. The building heights are indicated in the legend in m. The measurement points are marked as G3, G4 (anemometer in-street data, one on each side of the road) and Mast (background wind speed and direction at a height of 19m). A non-equidistant grid was used to enable a higher resolution within the area of interest (Gillygate). The compass in figure represents direction from and is oriented with respect to the street canyon.

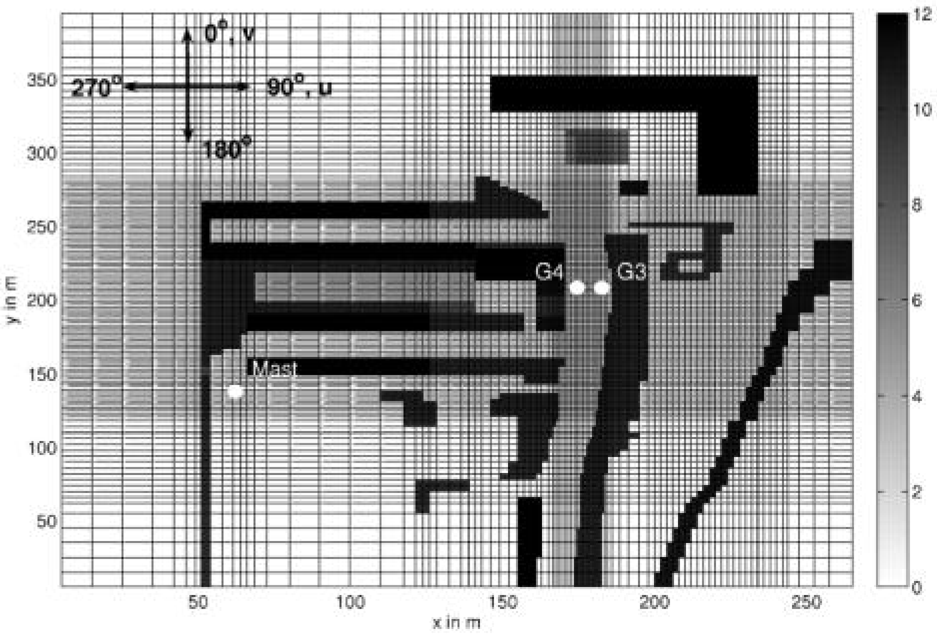

Fig. 1. Grid and building configuration around Gillygate as used in MISKAM.

The input parameters under consideration, with ranges shown in brackets, are x_1, the inflow roughness length [5...50]cm, x_2, the surface roughness length [0.5...50]cm, x_3, the wall roughness length [0.5...10]cm and x_4, the background wind direction [80...100]°. Uniform distributions are assumed for all parameters. The selected outputs chosen for illustration of the methods are y_1 the turbulent kinetic energy (tke) at G3, y_2 the tke at G4 and y_3 the horizontal wind component u at G4. Consideration of other outputs and physical interpretation of the sensitivities is given in the companion paper Benson et al. (2007).

HIGH DIMENSIONAL MODEL REPRESENTATION

The high dimensional model representation (HDMR) method is a set of tools explored by Rabitz et. al (1999) in order to express the input-output relationship of complex models with a large number of input variables. The mapping between the input variables $x_1,\ldots,x_n$ and the output variables $f(\mathbf{x})=f(x_1,\ldots,x_n)$ in the domain R^n can be written in the following form:

$$f(\mathbf{x})=f_0+\sum_{i=1}^{n} f_i(x_i)+\sum_{1\le i<j\le n} f_{ij}(x_i,x_j)+\ldots+f_{12\ldots n}(x_1,x_2,\ldots,x_n) \quad (1)$$

Here f_0 denotes the mean effect (zeroth order), which is a constant. The function $f_i(x_i)$ is a first order term giving the effect of variable x_i acting independently (although generally nonlinearly) upon the output $f(\mathbf{x})$. The function $f_{ij}(x_i,x_j)$ is a second order term describing the cooperative effects (pair correlated contribution) of the variables x_i and x_j upon the output $f(\mathbf{x})$. The higher order terms reflect the cooperative effects of increasing numbers of input variables acting together to influence the output $f(\mathbf{x})$. The HDMR expansion is computationally very efficient if higher order input variable correlations are weak and can therefore be neglected. For many systems a HDMR expression up to second order already provides satisfactory results and a good approximation of $f(\mathbf{x})$.

There are two commonly used HDMR expansions. Cut-HDMR depends on the value of $f(\mathbf{x})$ at a specific reference point $\bar{\mathbf{x}}$ and random sampling (RS) HDMR depends on the averaged value of $f(\mathbf{x})$ over the whole domain. Here, we have applied RS-HDMR, where the higher order component functions are approximated by orthonormal polynomials:

$$f_i(x_i) \approx \sum_{r=1}^{k} a_r^i j_r(x_i) \quad (2)$$

$$f_{ij}(x_i, x_j) \approx \sum_{p=1}^{l} \sum_{q=1}^{l'} \beta_{pq}^{ij} j_p(x_i) j_q(x_j) \quad (3)$$

where, k, l, l' represent the order of the polynomial expansion, $j_r(x_i)$, $j_p(x_i)$ and $j_q(x_j)$ are the orthonormal basis functions and a_r^i and β_{pq}^{ij} are constant coefficients to be determined (Li et al., 2002). The standard RS-HDMR approach has been extended by an optimisation method (Ziehn and Tomlin, 2007), which automatically chooses the best polynomial order for the approximation of each of the component functions.

The partial variances D_i and D_{ij} for sensitivity analysis are calculated from (Li et al., 2002a):

$$D_i \approx \sum_{r=1}^{k_i} \left(a_r^i\right)^2, \quad D_{ij} \approx \sum_{p=1}^{l_i} \sum_{q=1}^{l'_j} \left(b_{pq}^{ij}\right)^2 \quad (4)$$

Once the partial variances are determined sensitivity indices can be calculated as follows:

$$S_i = \frac{D_i}{D}, \quad S_{ij} = \frac{D_{ij}}{D} \quad (5)$$

where D is the total variance. The 1st order sensitivity index S_i measures the fractional contribution of x_i to the variance of $f(\mathbf{x})$. The 2nd order sensitivity index S_{ij} measures the interaction effect of x_i and x_j on the output and so on.

RESULTS

A fully functional model replacement is first constructed using the RS-HDMR approach. On the basis of this model replacement, uncertainty and sensitivity analysis can be performed in a computationally efficient way. Only one set of random samples is required in order to approximate all HDMR component functions by orthonormal polynomials. In this study a sample size of N=1024 using Sobol's quasi-random sampling method (Sobol, 1976) is applied to determine the coefficients for the orthonormal polynomials. The optimisation approach produces the optimal orders for these polynomials (table 1). The first order component functions are mainly approximated by fourth order polynomials. However the second order component functions are mainly approximated by second order polynomials (table 1). This indicates already that most of the component functions are nonlinear.

Table 1. Optimal order of polynomials for approximation of RS-HDMR component functions.

Output	f_1	f_2	f_3	f_4	f_{12}	f_{13}	f_{14}	f_{23}	f_{24}	f_{34}
y_1 G3 turbulent kinetic energy	4	4	4	4	1	2	2	2	2	2
y_2 G4 turbulent kinetic energy	2	4	2	3	1	2	2	2	2	2
y_3 G4 wind component *u*	4	2	4	4	1	2	3	2	2	3

To prove the accuracy of the second order RS-HDMR model replacement, the relative error (RE) between the approximated output and the output response of the full model was calculated for another set of 1000 random points. 100% of the approximated output y_2 and y_3 and 99.4% of the approximated output y_1 were in the 5% RE range. The statistics of both the full model and the model replacement also show very good agreement. This demonstrates that the model replacement can be used instead of the full MISKAM flow model in, for example, Monte Carlo (MC) type analysis. A large sample size can be applied because the model replacement is much less expensive to run than the full model.

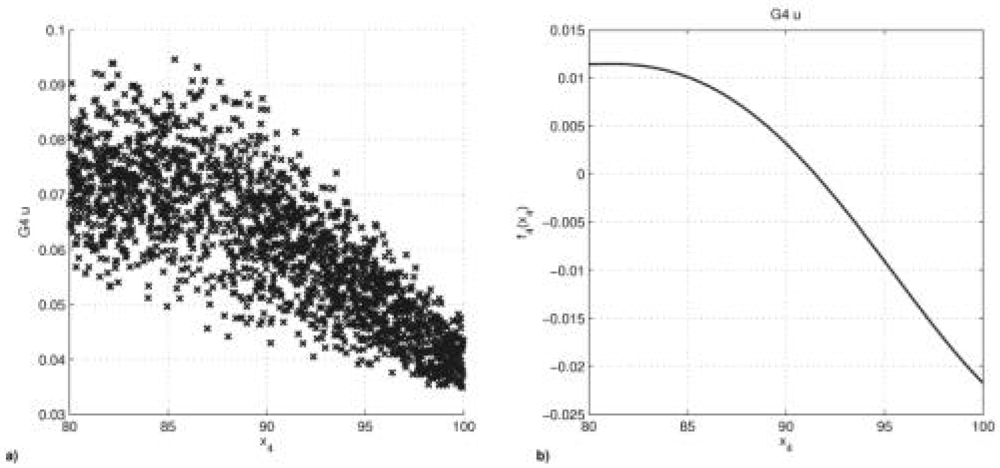

Fig. 2; Comparison scatter plot (a) and first order component function (b) for output G4 u.

The plots of the HDMR component functions reveal useful information about the input-output relationship of the model and can be used for sensitivity analysis instead of widely used scatter plots. For example, figure 2a shows an MC scatter plot using the model replacement. The amount of scatter indicates that the considered output *u* at G4 is also influenced by input parameters other than the one of interest (x_4 , background wind direction). The exact shape of the first order response is not always easy to assess within the scatter. However, figure 2b shows the corresponding first order RS-HDMR component function (here approximated by a fourth order polynomial). The component function describes the impact of the input parameter x_4 upon the output acting independently. Thus, the shape (in this case nonlinear) of the relationship becomes visible. This illustrates how much easier the component functions are to interpret than standard scatter plots. In addition, partial variances and sensitivity indices can be calculated on the basis of the HDMR component functions as shown in equations (4) and (5). The sensitivity indices are given in table 2 and can be ranked to show which input parameter (or interaction of input parameters) contributes most to the overall variance.

For example, output y_3 , G4 *u*, is mainly influenced by input parameter x_4 , wind direction (rank 1). In fact, 67.2% of the overall output variance is caused by this parameter. Output y_3

is further influenced by input parameter x_3, wall roughness length (rank 2) and x_1, inflow roughness length (rank 3). The input parameter x_2, surface roughness length contributes only by 4% to the overall output variance (rank 4). There are only very few second order effects, indicating that all outputs are mainly influenced by parameters acting independently. Further discussions and physical interpretations of the results can be found in Benson et al. (2007).

Table 2. Sensitivity indices first and second order.

Output	S_1	S_2	S_3	S_4	$\sum S_i$	S_{12}	S_{13}	S_{14}	S_{23}	S_{24}	S_{34}	$\sum S_{ij}$
y_1	0.253	0.426	0.115	0.161	0.955	0.000	0.012	0.009	0.004	0.002	0.017	0.044
y_2	0.123	0.103	0.612	0.131	0.969	0.000	0.001	0.013	0.001	0.001	0.004	0.020
y_3	0.093	0.040	0.176	0.672	0.981	0.000	0.001	0.009	0.000	0.004	0.009	0.023

CONCLUSIONS

The complexity of the street scale turbulent flow model restricts its application in connection with traditional global sensitivity analysis methods such as Monte Carlo analysis since a large number of model runs are required. One run of the full MISKAM flow model can take up to 40 min on a 3 GHz PC. Local sensitivity analysis methods are not suitable, because of the high nonlinearity of the input output relationship and parameter interactions. The RS-HDMR method has been shown to provide a straightforward approach to explore the input-output mapping and to calculate sensitivity indices in a very efficient way. Because HDMR methods treat the model as a black box, they could potentially be used for a wide range of applications in environmental modelling. The HDMR method is especially suitable for computationally expensive models with a large input space dimension.

REFERENCES

Benson, J.F., Dixon, N. S., Ziehn, T., Tomlin, A. S., 2007: The sensitivity of a 3D street canyon CFD model to uncertainties in input parameters. In: Proceedings of the 11th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes (HARMO).

Boddy, J.W.D., Smalley, R. J., Goodman, P. S., Tate, J. E., Bell, M. C., Tomlin, A. S., 2005: The spatial variability in concentrations of a traffic-related pollutant in two street canyons in York, UK Part I: The influence of background winds. Atmos. Environ. 39, 3147-3161.

Dixon, N. S., Boddy, J. W. D., Smalley, R. J., Tomlin, A. S., 2006: Evaluation of a turbulent flow and dispersion model in a typical street canyon in York, UK. Atmos. Environ. 40, 958-972.

Eichhorn, J., 1996: Validation of a Microscale Pollution Dispersal Model. In: Proceedings of the 21st International Meeting on Air Pollution Modeling and its Application.

Li, G., Wang, S.-W., Rabitz, H., 2002: Practical Approaches To Construct RS-HDMR Component Functions. J. Phys. Chem. A 106, 8721-8733.

Li, G., Wang, S.-W., Rabitz, H., Wang, S., Jaffé, P., 2002: Global uncertainty assessments by high dimensional model representations (HDMR). Chem. Eng. Sci. 57, 4445-4460.

Rabitz, H., Alis, Ö. F., Shorter, J., Shim, K., 1999: Efficient input--output model representations. Comput. Phys. Commun. 117, 11-20.

Sobol, I. M., 1976: On the distribution of points in a cube and the approximate evaluation of integrals. USSR Computational Mathematics and Mathematical Physics 7, 86-112.

Ziehn, T., Tomlin, A.S., 2007: Global sensitivity analysis of a 3D street canyon model – Part I: The development of high dimensional model representations. In preparation.