Sensitivity Analysis of Air Pollution Models

Rayna Georgieva and Sofiya Ivanovska

IPP - Bulgarian Academy of Sciences,
Sofia, Bulgaria

NATO Grant PDD(TC)-(ESP.EAP.CLG 982641)
"Monte Carlo Sensitivity Studies of Environmental Security",

ACET, University of Reading, November 13, 2008

Outline

- Introduction
- Mathematical background
- Numerical experiments
- Concluding remarks

Introduction

- Goal

  The aim is to propose a new mechanism for sensitivity studies of the concentrations levels of some important pollutants (like ozone O_3) in real-live scenarios of air pollution transport over Europe with Unified Danish Eulerian Model (UNI-DEM).

- Sensitivity analysis studies

- Motivation

Mathematical Background

- Unified Danish Eulerian Model (UNI-DEM)
- Total Sensitivity Indices
- Sobol' Approach for Computing Global Sensitivity Indices

Mathematical Background: UNI-DEM

$$
\begin{aligned}
\frac{\partial c_s}{\partial t} = \quad & \frac{\partial (u c_s)}{\partial x} - \frac{\partial (v c_s)}{\partial y} - \frac{\partial (w c_s)}{\partial z} + \\
& + \frac{\partial}{\partial x}\left(K_x \frac{\partial (c_s)}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y \frac{\partial (c_s)}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_z \frac{\partial (c_s)}{\partial z}\right) + \\
& + E_s + Q_s(c_1, c_2, \ldots, c_q) - (k_{1s} k_{2s} cs), \qquad s = 1, 2, \ldots, q.
\end{aligned}
$$

q	- number of equations $=$ number of chemical species,
c_s	- concentrations of the chemical species,
u, v, w	- components of the wind along the coordinate axes,
K_x, K_y, K_z	- diffusion coefficients,
E_s	- emissions in the space domain,
k_{1s}, k_{2s}	- coefficients of dry and wet deposition respectively $(s = 1, \ldots, q)$,
$Q_s(c_1, c_2, \ldots, c_q)$	- non-linear functions that describe the chemical reactions between species.

Mathematical Background: Mathematical Model

- The mathematical model

$$u = f(\mathbf{x}), \quad \text{where} \quad \mathbf{x} = (x_1, x_2, \ldots, x_d) \in U^d \equiv [0,1]^d$$

  is a vector of input parameters with a joint p.d.f. $p(\mathbf{x}) = p(x_1, \ldots, x_d)$.

  - Local sensitivity: $\mathbf{x} = \mathbf{x}^* \in U^d, \quad u^* = f(\mathbf{x}^*)$.
  - Global sensitivity: $u = f(\mathbf{x}), \quad \mathbf{x} \in U^d$.

Mathematical Background: Total Sensitivity Index

- Total Sensitivity Index of input parameter x_i, $\; i \in \{1, \ldots, d\}$:

$$S_{x_i}^{tot} = S_i + \sum_{l_1 \neq i} S_{il_1} + \sum_{l_1, l_2 \neq i, l_1 < l_2} S_{il_1 l_2} + \ldots + S_{il_1 \ldots l_{d-1}},$$

where

$$S_{il_1 \ldots l_{j-1}} - j^{\text{th}} \text{ order sensitivity index for parameter } x_i \; (1 \leq j \leq d),$$

$$j = 1: \;\; S_i - \text{"the main effect" of } x_i.$$

- Classification of input parameters:

–very important:	0.8	$<$	$S_{x_i}^{tot}$		
–important:	0.5	$<$	$S_{x_i}^{tot}$	$<$	0.8
–unimportant:	0.3	$<$	$S_{x_i}^{tot}$	$<$	0.5
–irrelevant:			$S_{x_i}^{tot}$	$<$	0.3

Mathematical Background: Sobol' Approach

ANalysis Of VAriances (ANOVA) HDMR of a square integrable function $f(\mathbf{x})$:

$$f(\mathbf{x}) = f_0 + \sum_{s=1}^{d} \sum_{l_1<\ldots<l_s} f_{l_1\ldots l_s}(x_{l_1}, x_{l_2}, \ldots, x_{l_s}),$$

where

- f_0 - constant,
- $\int_0^1 f_{l_1\ldots l_s}(x_{l_1}, x_{l_2}, \ldots, x_{l_s})\,\mathrm{d}x_{l_k} = 0, \quad 1 \leq k \leq s, \quad s = 1, \ldots, d.$

I.M. Sobol, Multidimensional Quadrature Formulas and Haar Functions, Nauka, Moscow, 1969.
I.M. Sobol, Sensitivity estimates for nonlinear mathematical models, Mathematical Modelling and Computational Experiments, 1(4) (1993) 407–414.

Mathematical Background: Sobol' Approach

Therefore

- $\int_{U^d} f_{i_1,\ldots,i_s} f_{j_1,\ldots,j_l}\, \mathrm{d}\mathbf{x} = 0, \quad (i_1,\ldots,i_s) \neq (j_1,\ldots,j_l), \quad s,l \in \{1,\ldots,d\}$

and the functions in the right-hand side are defined in a unique way:

- $f_0 = \int_{U^d} f(\mathbf{x})\mathrm{d}\mathbf{x}$
- $f_{l_1}(x_{l_1}) = \int_{U^{d-1}} f(\mathbf{x}) \prod_{k\neq l_1} \mathrm{d}\mathbf{x}_k - f_0,\ l_1 \in \{1,2,\ldots,d\}$
- $f_{l_1 l_2}(x_{l_1}, x_{l_2}) = \int_{U^{d-2}} f(\mathbf{x}) \prod_{k\neq l_1,l_2} \mathrm{d}\mathbf{x}_k - f_0 - f_{l_1}(x_{l_1}) - f_{l_2}(x_{l_2})$,
  $l_1, l_2 \in \{1,2,\ldots,d\}$

Mathematical Background: Global (Sobol') Sensitivity Indices

Definition (Sobol'):

$$S_{l_1 \ldots l_s} = \frac{\mathbf{D}_{l_1 \ldots l_s}}{\mathbf{D}}, \qquad s \in \{1, \ldots, d\},$$

where

- variances $\mathbf{D}_{l_1 \ldots l_s} = \int f^2_{l_1 \ldots l_s} \mathrm{d}x_{l_1} \ldots \mathrm{d}x_{l_s}$,
- total variance $\mathbf{D} = \int_{U^d} f^2(\mathbf{x}) \mathrm{d}\mathbf{x} - f_0^2$,

and the following properties hold:

- $S_{l_1 \ldots l_s} \geq 0, \qquad \sum_{s=1}^{d} \sum_{l_1 < \ldots < l_s}^{d} S_{l_1 \ldots l_s} = 1.$

Methods for Evaluating Global Sensitivity Indices

Method	Cost (Model runs)	Sensitivities
FAST (1973)	$O(d^2)$	$S_i, \forall i$
Sobol (1993)	$N(2d+2)$	$S_i, S_{x_i}^{tot}, \forall i$
EFAST (1999)	dN	$S_i, S_{x_i}^{tot}, \forall i$
Saltelli (2002)	$N(d+2)$	$S_i, \forall i, S_{lj}, \forall l, j, l \neq i$

Mathematical Background: Sobol' Monte Carlo Algorithm

Let $\mathbf{x} = (\mathbf{y}, \mathbf{z}) \in \mathbb{R}^d, \mathbf{y} = (x_{k_1}, x_{k_2}, \dots, x_{k_m}) \in \mathbb{R}^m, K = (k_1, k_2, \dots, k_m)$.

Variance of the subset $\mathbf{y}$: $\quad \mathbf{D}_{\mathbf{y}} = \sum_{s=1}^{m} \sum_{(k_1,\dots,k_m)\in K} \mathbf{D}_{l_1\dots l_s}.$

Theorem (Sobol'). $\mathbf{D}_{\mathbf{y}} = \int f(\mathbf{x})\, f(\mathbf{y}, \mathbf{z}')\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{z}' - f_0^2 \qquad \left(f_0 = \int_{U^d} f(\mathbf{x})\mathrm{d}\mathbf{x} \right).$

Monte Carlo algorithm:

$$\frac{1}{N} \sum_{j=1}^{N} f(\xi_j) \xrightarrow{P} f_0 \qquad\qquad \frac{1}{N} \sum_{j=1}^{N} f^2(\xi_j) \xrightarrow{P} \mathbf{D} + f_0^2$$

$$\frac{1}{N} \sum_{j=1}^{N} f(\xi_j)\ f(\eta_j, \zeta_j') \xrightarrow{P} \mathbf{D}_{\mathbf{y}} + f_0^2, \qquad\qquad \xi = (\eta, \zeta).$$

$$S_{l_1} = S_{(l_1)} = \frac{\mathbf{D}_{(l_1)}}{\mathbf{D}}, \quad S_{(l_1 l_2)} = S_{l_1} + S_{l_2} + S_{l_1 l_2} = \frac{\mathbf{D}_{(l_1 l_2)}}{\mathbf{D}} \Longrightarrow S_{l_1 l_2}.$$

Approaches for Small Indices:
Reducing the Mean Value (I.M. Sobol', 1990)

Motivation: If $\mathbf{D_y} << f_0^2 \quad \Rightarrow$ a loss of accuracy.

Approach concept: Choose $c \sim f_0$ and set the function $\varphi(\mathbf{x}) = f(\mathbf{x}) - c$.

Therefore

$$\mathbf{D_y} = \int \varphi(\mathbf{x})\ \varphi(\mathbf{y}, \mathbf{z}')\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{z}' - \omega^2, \qquad \text{where} \quad \omega = \int \varphi(\mathbf{x})\mathrm{d}\mathbf{x},$$

$$\mathbf{D} = \int \varphi^2(\mathbf{x})\mathrm{d}\mathbf{x} - \omega^2, \qquad \omega = f_0 - c.$$

Variance (standard error) of the estimate:

$$V^{(1)} = V_0^{(1)} + \mathbf{O}(|\omega|), \quad V^{(1)} = \int \varphi^2(\mathbf{x})\varphi^2(\mathbf{y}, \mathbf{z}')\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{z}' - (\mathbf{D_y} + \omega^2)^2,$$ where

$$V_0^{(1)} = \int h^2(\mathbf{x})\ h^2(\mathbf{y}, \mathbf{z}')\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{z}' - \mathbf{D_y^2}, \quad h(\mathbf{x}) = f(\mathbf{x}) - f_0.$$

Approaches for Small Indices:
A Combined Approach (Saltelli, 2002, Kucherenko, 2007)

- **Approach concept**
  - Choose a constant $c \sim f_0$ and set the function $\varphi(\mathbf{x}) = f(\mathbf{x}) - c$.
  - Use $\varphi(\mathbf{x})$ rather than $f(\mathbf{x})$:

$$\mathbf{D_y} = \int \varphi(\mathbf{x})\ [\varphi(\mathbf{y},\mathbf{z}')dxdz' - \varphi(\mathbf{x}')]\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{x}',$$

$$\mathbf{D} = \int \varphi(\mathbf{x})[\varphi(\mathbf{x}) - \varphi(\mathbf{x}')]\ \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{x}'.$$

- **Variance of the estimate**

$$V^{(2)} = V_0^{(2)} + \mathbf{O}(|\omega|), \qquad V_0^{(2)} = \int h^2(\mathbf{x})\ [h(\mathbf{y},\mathbf{z}') - h(\mathbf{x}')]^2\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{x}' - \mathbf{D}_\mathbf{y}^2.$$

Comparison of estimators variances: $V_0^{(2)}$ versus $V_0^{(1)}$

ANOVA-like decomposition of the model function:

$$f(\mathbf{x}) = f_0 + g_1(\mathbf{y}) + g_2(\mathbf{z}) + g_{12}(\mathbf{x}), \qquad \text{where}$$

$$g_1(\mathbf{y}) = \int f(\mathbf{x})\mathrm{d}\mathbf{z} - f_0, \quad g_2(\mathbf{z}) = \int f(\mathbf{x})\mathrm{d}\mathbf{y} - f_0,$$

$$g_{12}(\mathbf{x}) = f(\mathbf{x}) - \int f(\mathbf{x})\mathrm{d}\mathbf{y} - \int f(\mathbf{x})\mathrm{d}\mathbf{z} + f_0.$$

Proposition (Sobol', Myshetskaya, 2007):

Denote $\delta = \sup |v(\mathbf{y})| \frac{2}{\mathbf{D_z}}$. If $\delta < 1$ and $S_{\mathbf{z}} > \frac{1}{2-\delta}$, then $V_0^{(2)} < V_0^{(1)}$,

where $\quad v(\mathbf{y}) = \int g_2(\mathbf{z}')\, g_{12}(\mathbf{y}, \mathbf{z}')\mathrm{d}\mathbf{z}'.$

Numerical Results from the Preliminary Stage

The input data:

$$r_s(\alpha) = \frac{c_s^{\alpha}(a_s^{i_{max}}, b_s^{j_{max}})}{c_s^{max}}, \qquad \text{where} \quad \alpha_i \in \{0.1, 0.2, \ldots, 2.0\}, \quad \alpha = \{\alpha_i\} \quad \text{and}$$

$i = 1, 2, \ldots, 69$	–	the time-dependent chemical reactions,
$i = 1, 2, \ldots, 47$	–	the constant chemical reactions,
$s = 1, \ldots, 35$	–	the chemical species (pollutants),
$c_s^{max} = c_s^{max}(a_s^{i_{max}}, b_s^{j_{max}})$	–	the maximum mean value of the concentration of chemical species for July 1998.

Numerical Results from the Preliminary Stage

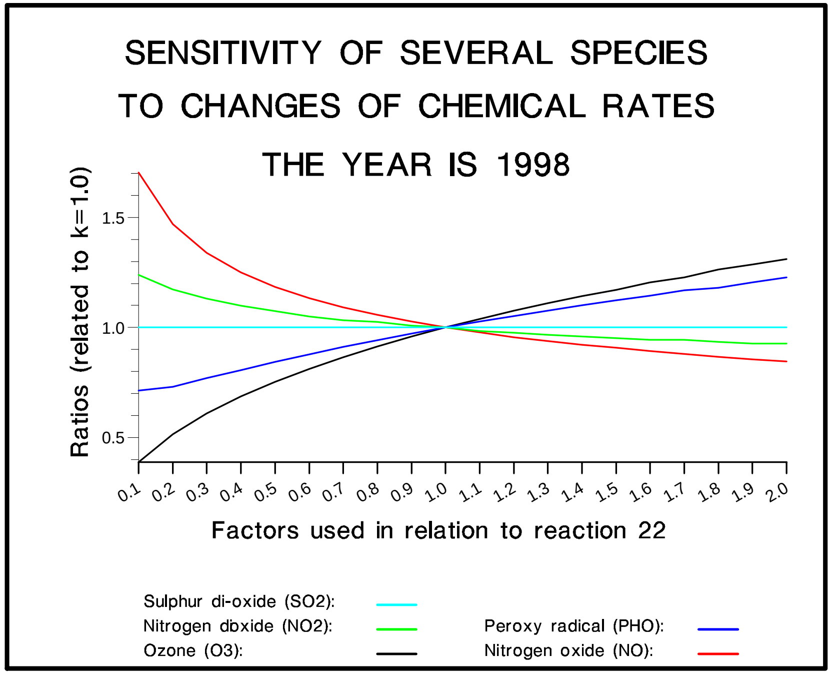

Figure 1: Sensitivity of several species to changes of chemical rates (1998).

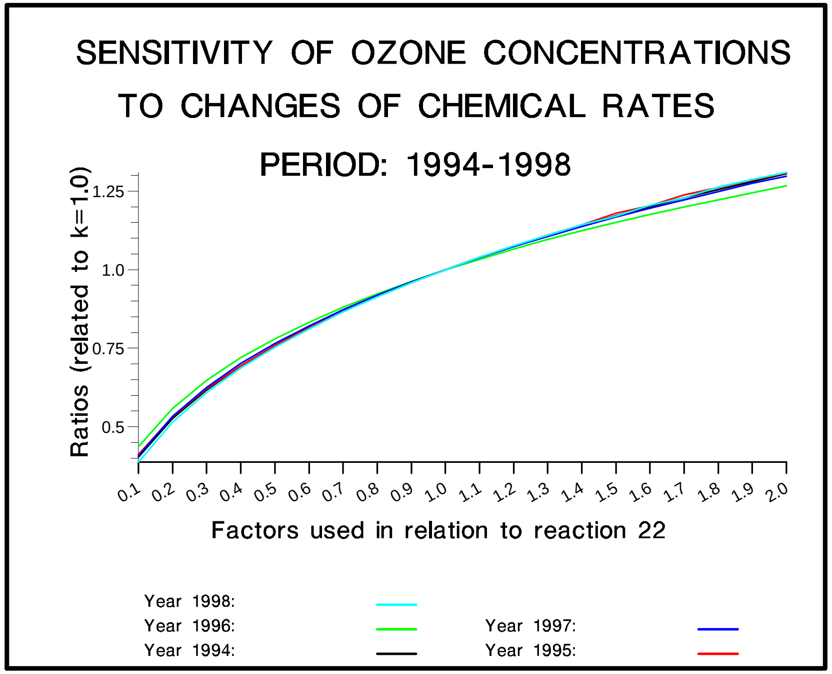

Figure 2: Sensitivity of ozone concentrations to changes of chemical rates (period 1994-1998).

Numerical Results from the Preliminary Stage

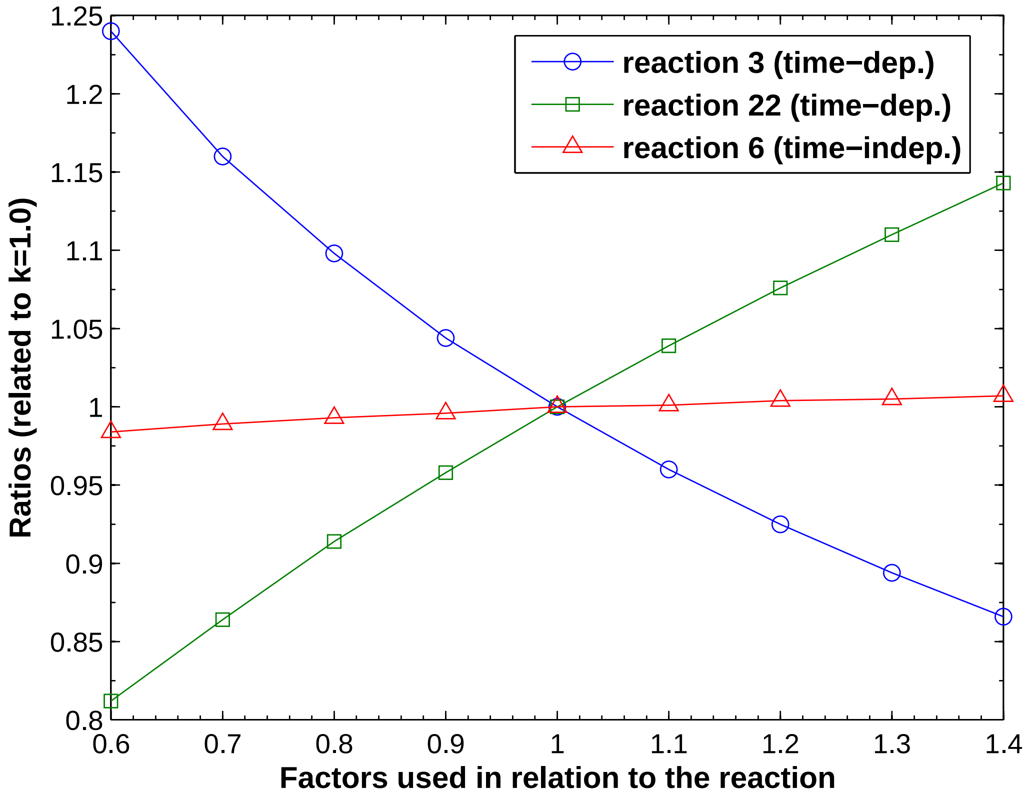

Figure 3: Sensitivity of ozone concentrations to changes of chemical rates (July 1998).

The Case Study: Approximation

Input data: $r_s(\alpha)$, $s \leftrightarrow O_3$, $i = \{3, 22, 6\}$, **Denote** $\mathbf{x} := \alpha$.

$\alpha_i \in \{0.6, 0.7, \ldots, 1.4\}$, $\alpha = (\alpha_3, \alpha_{22}, \alpha_6)$.

Approximation function: $$f_s(\mathbf{x}) = \sum_{j=0}^{k} \sum_{\substack{\nu_1, \nu_2, \ldots, \nu_d = 0 \\ \nu_1 + \ldots + \nu_d = j}}^{k} a_{\nu_1 \ldots \nu_d} \, x_1^{\nu_1} x_2^{\nu_2} \ldots x_d^{\nu_d}.$$

Error estimation: $$||f_s - r_s||_2^2 = \sum_{l=1}^{n} [f_s(\mathbf{x}_l) - r_s(\mathbf{x}_l)]^2, \qquad \mathbf{x}_l \in [0.6; 1.4]^3.$$

Table 1: Squared 2-vector norm of the residual (f_s is a polynomial of 4th degree).

$\mathbf{x} \in [0.1; 2.0]^3$	$\mathbf{x} \in [0.6; 1.4]^3$
0.016	0.00005

Numerical Results

Table 2: Total sensitivity indices of input parameters obtained using different variance-based approaches for sensitivity analysis.

approach / estimated quantity	Standard (Sobol')		Approaches for small indices	
			red. of the m.v.	combined
	$\mathbf{x} \in [0.1; 2.0]^3$	$\mathbf{x} \in [0.6; 1.4]^3$	$\mathbf{x} \in [0.6; 1.4]^3$	$\mathbf{x} \in [0.6; 1.4]^3$
integrand $g(\mathbf{x})$	$f(\mathbf{x})$	$f(\mathbf{x})$	$f(\mathbf{x}) - c$	$f(\mathbf{x}) - c$
c	-	-	0.51737	0.51737
g_0	0.51520	0.51634	0.25145	0.25145
$\mathbf{D}$	0.26181	0.26446	0.07061	0.00530
S_1	0.26386	0.26530	0.27354	0.52979
S_2	0.26447	0.26359	0.26713	0.46142
S_3	0.25348	0.25209	0.22406	0.00222
$\sum_{i=1}^{3} S_i$	0.78182	0.78097	0.76474	0.99342

Numerical Results

Table 3: Total sensitivity indices of input parameters obtained using different variance-based approaches for sensitivity analysis.

estimated quantity \ approach	Standard (Sobol')		Approaches for small indices	
			red. of the m.v.	combined
	$\mathbf{x} \in [0.1; 2.0]^3$	$\mathbf{x} \in [0.6; 1.4]^3$	$\mathbf{x} \in [0.6; 1.4]^3$	$\mathbf{x} \in [0.6; 1.4]^3$
S_{12}	0.06885	0.06941	0.07994	0.00628
S_{13}	0.06598	0.06634	0.06845	0.00009
S_{23}	0.06613	0.06592	0.06686	0.00021
$\sum_{i,j=1,i\leq j}^{3} S_{ij}$	0.20096	0.20167	0.21525	0.00658
S_{123}	0.01722	0.01736	0.02001	0.000003
$S_{x_1}^{tot}$	0.41592	0.41841	0.44195	0.53615
$S_{x_2}^{tot}$	0.41667	0.41627	0.43395	0.46791
$S_{x_3}^{tot}$	0.40281	0.40170	0.37938	0.00252

Numerical Results

Table 4: First order and total sensitivity indices of input parameters obtained using combined approaches and two approaches implemented in R package.

estimated quantity	$\mathbf{x} \in [0.1; 2.0]^3$			$\mathbf{x} \in [0.6; 1.4]^3$		
	combined approach	R Package		combined approach	R Package	
		Sobol, Saltelli	FAST		Sobol, Saltelli	FAST
S_1	0.48262	0.46645	0.47933	0.52979	0.53029	0.52783
S_2	0.51080	0.54567	0.50928	0.46142	0.47884	0.46034
S_3	0.00104	0.00288	0.00101	0.00222	0.00254	0.00221
$S_{x_1}^{tot}$	0.48807	0.48641	0.48800	0.53615	0.53760	0.53609
$S_{x_2}^{tot}$	0.51592	0.50845	0.51604	0.46791	0.45089	0.46799
$S_{x_3}^{tot}$	0.00157	-0.00043	0.00308	0.00252	0.00180	0.00365

More information about R language and environment for statistical computing: http://www.r-project.org/

Conclusion

Some of the most advanced variance-based approaches for sensitivity analysis have been applied to study the influence of chemical rates variation over the concentration levels of air pollutants using UNI-DEM.

- Applicability of the results
  - for a verification and an improvement of the model;
  - for a reliable prediction of the final output.
- Future plans
  - other approximation tools;
  - computations with 3D version of UNI-DEM;
  - studies of model sensitivity depending on emissions levels and boundary conditions.

Bibliography

- T. Ishigami, T. Homma, An Importance Quantification Technique in Uncertainty Analysis for Computer Models, in: *Proceedings of the ISUMA'90* (1990).

- A. Saltelli, S. Tarantola, F. Campolongo, M. Ratto, *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*, Halsted Press, New York (2004).

- I.M. Sobol', Global Sensitivity Indices for Nonlinear Mathematical Models and Their Monte Carlo Estimates, *Mathematics and Computers in Simulation*, **55 (1-3)** (2001) 271–280.

- I. Sobol', E. Myshetskaya, Monte Carlo Estimators for Small Sensitivity Indices, *Monte Carlo Methods and Applications* **13** (5-6) (2007) 455–465.

- Z. Zlatev and I. Dimov, *Computational and Numerical Challengies in Environmental Modelling*, Elsevier, Amsterdam (2006).