TRANSPHORM submitted deliverable

D4.2.2 Implementation of uncertainty methodology into software tools – development of uncertainty assessment toolbox

Partners: AUTH

Prepared by: Denis Sarigiannis, Periklis Kontoroupis, Spyros Karakitsios (AUTH)

Nature: Report

Dissemination level: Public

Delivery date from ANNEX I:

Actual delivery date: July 2014

Notes:

TABLE OF CONTENTS

ABSTRACT

This report aims to demonstrate the implementation of uncertainty tools to the Transphorm project, presented in the D.4.2.1. In particular, the software developed incorporates routines (included with this toolbox) available in the statistic toolbox in Matalb® , with scope to fit distributions to data, to generate data in accordance to pre-defined statistics, use sampling algorithms and others. These routines are used to compute the level of uncertainty, presented as the coefficient of variation, to the estimated health impact from PMx and EC exposure. In particular uncertainty to health impact is computed at different resolution, for a number of core cities (e.g. London) and for Europe at the EMEP grid level

1 UNCERTAINTY TO THE HEALTH IMPACT FROM PM AND EC FOR THE CORE CITIES

The health effects are related from the delta change in $PM_{2.5}$, EC and PNC. The resulting Gain in Life years per city are computed and the associated uncertainty to the baseline and the 2020 scenarios including, i) 10% less urban road traffic, ii) 50% electric vehicles and iii) LEZ in these cities.

1.1 The case study of London

For the case of London, the uncertainties to the health impact from the EC and $PM_{2.5}$ exposure are computed, associated with the uncertainty in concentration and the concentration response functions. In particular two health impact methods are used, a simplified health impact method (UH) developed by the University of Hertfordshire and via the use of life tables. Starting from the available information on the observed and the modelled concentrations (OSCAR model), presented in table 1, uncertainty in predictions is computed, by firstly determining the ratio modelled vs observed concentrations and secondly via fitting a distribution that best describes the data variation, as illustrated in figure 1.

Table 1. Modelled and observed concentrations for $PM_{2.5}$ (2008 data)

xs	ys	Model $PM_{2.5}$ (ug/m^3)	Observation $PM_{2.5}$ (ug/m^3)	Ratio
549975	179064	13.67	9.95	1.37
551860	176376	13.03	15.32	0.85
552615	175416	13.17	12.35	1.07
552566	175384	13.03	11.00	1.19
547323	181231	14.19	9.40	1.51
520866	185169	17.71	14.27	1.24
530123	182014	20.34	16.13	1.26
526629	184391	21.69	16.54	1.31
544084	178881	15.26	19.94	0.77
543978	174655	13.65	16.54	0.83
540169	178999	15.96	16.49	0.97
545560	178526	14.25	15.10	0.94
541879	175016	15.44	17.19	0.90
540200	178367	18.81	16.38	1.15

544997	175098	15.47	18.25	0.85
532947	182575	20.08	13.48	1.49
533891	190707	15.37	17.48	0.88
517877	192314	12.68	12.88	0.99
508300	177800	14.15	13.47	1.05
526524	178965	21.02	16.24	1.29
524045	181752	19.73	13.96	1.41
540823	188369	15.41	15.83	0.97
569356	182736	13.88	15.27	0.91
538290	181452	19.16	18.23	1.05
528125	182016	23.12	16.72	1.38
528125	182016	23.12	22.64	1.02

Figure 1, depicts two distributions fitted to the data, a normal and a lognormal distribution, benchmarked in accordance to the following statistical criteria: the Kolmogorov – Smirnov, the Anderson – Darling and the Chi-Squared method, presented in table 2. According to these criteria the normal distribution is selected, used as a correction factor to the ambient air concentration depicted in figure 2 following the form N(15.50, 0.22^2).

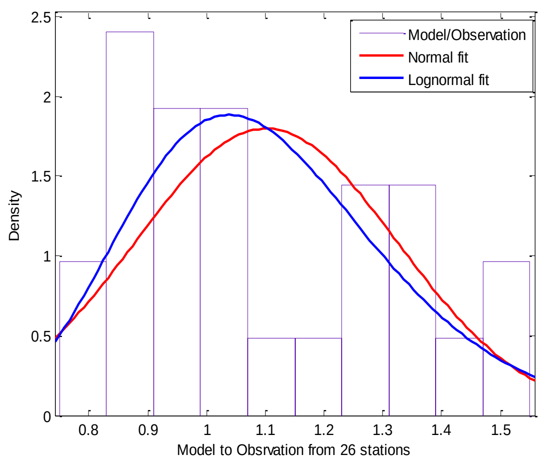

Figure 1. Population exposure to $PM_{2.5}$ per concentration band

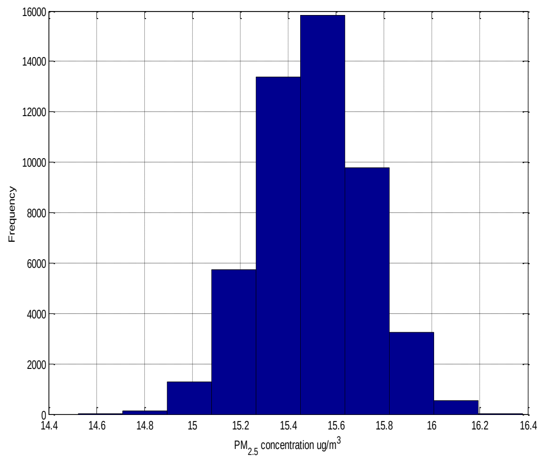

Figure 2. Visualized uncertainty to the mean city $PM_{2.5}$ ambient air concentration i.e. N(15.50, 0.22^2)

Table 2 Statistics of the distributions fitted to the data

Metric	Normal Distribution	Lognormal distribution
Log likelihood	2.73	3.44
Domain	-Inf < y < Inf	0 < y < Inf

Mean	1.10	1.10
Variance	0.05	0.05
Parameter Estimate Std. Err: mu	0.04	0.04
Parameter Estimate Std. Err sigma	0.03	0.03
Kolmogorov - Smirnov	0.12	0.11
Anderson - Darling	0.36	0.44
Chi-Squared	2.37	2.31

Furthermore, the population-weighted annual average concentration for $PM_{2.5}$ is of the form x ~ N(μ_1, σ_1^2) and R the relative risk of the form RR ~ N (μ_2, σ_2^2), Using a uniform random sampling method, the necessary input data are extracted from each of the two health impact methods. Hence, using the ambient air concentration and crossing them with the available population, it is possible to estimate the average population weighed concentrations and its associated uncertainty. The aforementioned calculation is presented in table 3 for $PM_{2.5}$ and visualized with respect to the population weighted concentration in figure 4.

Table 3. Propagation of uncertainty to the population weighted concentration for $PM_{2.5}$

Population weighted concentration Scenario year 2008	Mean estimate (ug/m^3)	Lower confidence interval (ug/m^3)	Upper confidence interval (ug/m^3)
$PM_{2.5}$ (ug/m^3)	15.40	12.40	18.50

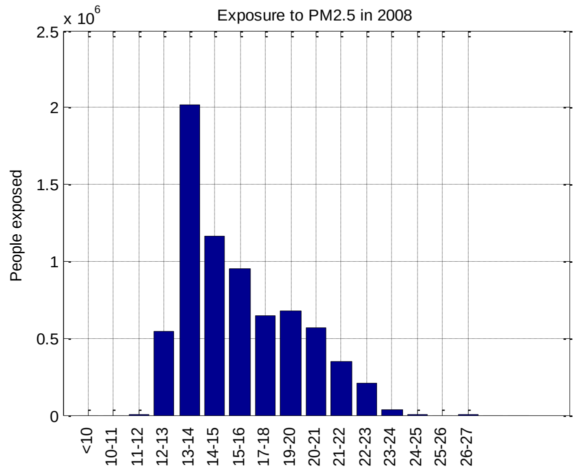

Figure 3. Population exposure to $PM_{2.5}$ per concentration band : 15.40 ug/m^3 is the average population weighted concentration

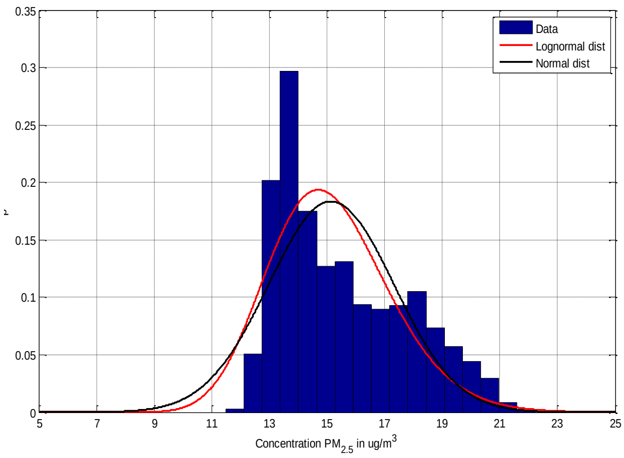

Figure 4. Visualized uncertainty to the average population weighted concentration of $PM_{2.5}$

It is noted that for the future scenarios, white noise ξ ~ N(0,0.5^2) is added to the population weighted concentration, depicted in figure 5. Hence the final population weighted concentration is presented in figure 6 and in table 4. Lastly, in Table , fitted distributions to the associated confidence intervals for the Concentration Response functions R, are presented.

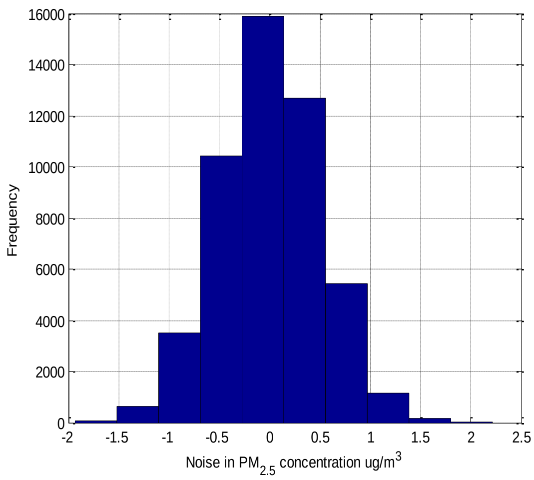

Figure 5. White noise added at posteriori $\xi \sim N(0,0.5^2)$

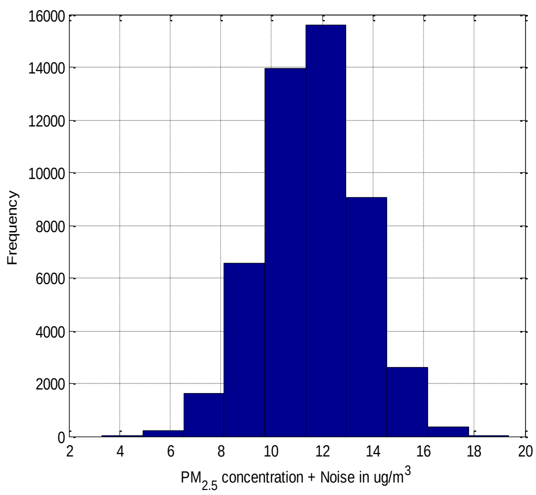

Figure 6. $PM_{2.5}$ concentration for scenario year 2020 including noise ~ $N(11.60, 1.9^2)$

Table 4. Uncertainty associated to the population weighted annual concentrations for $PM_{2.5}$

Scenario year	$PM_{2.5}$ (ug/m^3)
2008	N(15.40, 1.85 ^2)
2020	N(11.60, 1.92^2)
2020+measure: 50% electric	N(11.59, 1.92^2)
2020+measure: LEZ	N(11.59, 1.92^2)
2020+measure: 10% less traffic	N(11.60, 1.92^2)

Table 6. Fitted distribution to CRFs associated to EC, PM_{10} and $PM_{2.5}$

CRF per pollutant	Fitted distribution
$PM_{2.5}$: RR= 1.06 per U = 10 μg/m^3 $PM_{2.5}$, CI=[(1.041 – 1.084)]	LN(1.06, 0.015^2)
EC: RR= 1.06 per U = 1 μg/m^3, CI=[(1.04 - 1.09)]	LN(1.06, 0.015^2)
PNC: RR=1.003 per 1000, CI=[(1.001 – 1.012)]	LN(1.003, 0.004^2)

Uncertainty is assessed via Monte Carlo Simulation (MCS), which involves a large number of samples (typically hundreds of thousands) from the distribution of the input parameters that were combined to obtain probability distributions for the health impact rate output and thus statistically quantify the residual uncertainty. Uniform sampling is used and the computed health impact is reported in table 7.

Table 7. Fitted distribution to CRFs associated to $PM_{2.5}$ core-city of London – Simplified Health Impact version

		Simplified Health Impact		Life Tables	
Scenario	$PM_{2.5}$ (ug/m^3)	Attr Deaths	Attr Life Years	Attr Deaths	Attr Life Years
2008	15.40	4212 (CI: 2447, 6119)	48245 (CI: 28142, 70372)	4212 (CI: 2499, 6187)	48245 (CI: 31357, 77636)
2020	11.60	3139 (CI: 1769, 4798)	35961 (CI: 20347, 55172)	3139 (CI: 1834, 4769)	35961 (CI: 2301, 61093)
2020+measure:	11.59	3136	35927	3136	35927

50% electric		(CI: 1757, 4741)	(CI: 20202, 54519)	(CI:1815, 4886)	(CI: 2277, 61303)
2020+measure: LEZ	11.59	3134 (CI: 17468, 47707)	35904 (CI: 20088, 54863)	3134 (CI:1815, 4886)	35904 (CI: 2277, 61303)
2020+measure: 10% less traffic	11.60	3138 (CI: 1737, 4778)	35949 (CI: 19974, 54951)	3138 (CI: 1834, 4769)	35949 (CI: 2301, 61093)

2 UNCERTAINTY TO THE HEALTH IMPACT FROM PM EXPOSURE IN EUROPE

A pilot fully chain health impact assessment was conducted, divided in a number of steps presented next,

Step 1: Uncertainty to emission per source is computed at cell level from the emission factors and is aggregated to the country level. Here uncertainty to the emission factors from the road and rail transport (from the remaining sources a conservative assumption of 50% of the relative standard deviation -RSD to emissions per grid, were used), was estimated per cell level and reported at country level. It is noted that for the case of road transport, aggregation was possible after taking into account a correlation matrix, deduced from the vehicle category, type and technology per grid level.

Step 2: Uncertainty in concentration is deduced after comparing model output i.e. PM_{10} and $PM_{2.5}$ to the airbase observation for the year 2005, differentiated between the urban and rural areas, as determined by the population density. For the $PM_{2.5}$ concentrations $PM_{2.5}$ to PM_{10} factors are used to generate statistically significant observations for a number of countries.

Step 3: The urban increment model was used to link emissions to concentration in the urban areas and act as surrogate to the regional models. Size of the urban areas was deduced from the CORINE land cover types and yearly average meteorological data were used.

Step 4: Uncertainty to the population weighted concentrations, per country, is deduced taking into account the uncertainty in the concentration models.

Step 5: Uncertainty to the relative risk per health end point were defined by fitting distributions to the concentration response functions, in order to match the associated confidence intervals, under the maximum likelihood criteria.

Step 6: For the future scenarios, ‘white noise’ is added to 2020 population weighted concentrations.

Step 7: A uniform sampling algorithm was selected, starting from emission to heath impact per grid .

Step 8: Data per grid are collected and statistics are extracted.

3 SENSITIVITY ANALYSIS TO THE URBAN INCREMENT MODEL

An urban increment is recently been developed (Torras Ortiz and Friedrich 2013) for all German cities with over 50000 inhabitants. The urban background increment for PM_{10} is estimated using a functional relationship between emissions, city size and average wind speed, presented by equation 1.

$$C_{i,urban} = \omega_i + \phi_i \cdot \frac{E_{i,UE}}{A_{UE} \cdot u_{avg}} + \gamma_i \cdot C_{i,rural} \quad (1)$$

Where:

$C_{i,\,urban}$ = Urban increment of pollutant *i*.

$E_{,iUE}$ = Total emission of pollutant *i* within an urban entity in tons.

A_{UE} = Urban area in km^2.

u_{avg} = Urban average wind speed in m/s.

$C_{i,\,rural}$ = Rural background concentration of pollutant *i* in ug/m^3.

ω_i, ϕ_i, and γ_i = Multiple-regression parameters for pollutant *i*.

Table 8. Urban increment parameters

Pollutant	Intercept (omega)	Rural (gamma)	Ratio (phi)
PM_{10}	-1.89	1.03	6.22
NO_2	13.90	0.51	17.67

3.1.1.1 Global sensitivity tools to the urban increment model

Global sensitivity analysis methods are used here to quantify the variation in emission that is caused by model inputs such as ambient temperature, fraction of pesticide lost due to drift, fraction intercepted by the crop and lateral distance from the emission source. The Sobol method (I. M Sobol 1993) was selected, which takes into account all the input variation ranges in order to apportion emission uncertainty to the uncertainty in the input factors previously mentioned. Sobol indices are based on a probabilistic framework using Monte-Carlo, which measures the main effect of the input parametes to the output, as well as the interaction with the other input parameters. Thus, the total effect index is a more accurate measure of an input parameter on model output. In order to compute these indices, a model function is assumed for the Concentration (C_k) in a cell *k*, of the form:

$C_k = f(S7, S81, S82, S83, S84, S85, S86, S87, S88, W, C_b)$, where probability density functions could be assumed; here *S* stands for the SNAP type of emission source, *W*, is the wind speed at the particular cell k, and C_b the background concnetration in a rural area close the city. In order to rank these input factors according to the amount of variance they would explain, it is required to know the true value $S7^*, S81^*, S82^*, S83^*, S84^*, S85^* S86^*, S87^*, S88^*, W^*, C_b^*$ of a given input factoid $S7, S81, S82, S83, S84, S85, S86, S87, S88, W, C_b$. In addition, the conditional variance of C_k for one of the inputs, given that $S7 = S7^*$ defined as $V(C_k \mid S7 = S7^*)$ is computed from the variance over all factors but S7. Since the true value $S7^*$ is unknown, the average of this conditional variance for all possible values $S7^*$ of S7 is used. The same can be assumed for the remaining inputs $S81^*, S82^*, S83^*, S84^*, S85^* S86^*, S87^*, S88^*, W^*, C_b^*$. Furthermore, for all inputs, the unconditional variance of the output is $V(C_k)$ and by using the following property of the variance:

$$V(C_k) = V(E[C_k \mid (S7, S81, S82, S83, S84, S85, S86, S87, S88, W, C_b)]) + \ldots \quad (2)$$

$$\ldots + E[V(C_k \mid (S7, S81, S82, S83, S84, S85, S86, S87, S88, W, C_b))]$$

the variance of the conditional expectation is defined as

$$V_i = V(E[C_k \mid (S7, S81, S82, S83, S84, S85, S86, S87, S88, W, C_b)]) \quad (3)$$

This measure is called the 'main effect' and is used as an indicator of the importance of inputs $(S7, S81, S82, S83, S84, S85, S86, S87, S88, W, C_b)$ on the variance of C_k, i.e. the sensitivity of C_k to $(S7, S81, S82, S83, S84, S85, S86, S87, S88, W, C_b)$.

Via normalization of V_i the first order sensitivity index is computed as follows

$$S_i = \frac{V(E[C_k \mid (S7, S81, S82, S83, S84, S85, S86, S87, S88, W, C_b)])}{V(C_k)} \quad (4)$$

Equation 4 (first order sensitivity index or Sensitivity Index) provides a measure only of the main contribution of each input parameter to the emission rate variance, without taking into account interactions among the input factors.

To take into account the interactions among all four input factors the 4th order effect between the four orthogonal factors $(S7, S81, S82, S83, S84, S85, S86, S87, S88, W, C_b)$ would be:

$$V_{S7,S81,S82,S83,S84,S85,S86,S87,S88,W,C_b} = V(E[C_k \mid_{S7,S81,S82,S83,S84,S85,S86,S87,S88,W,C_b}]) - \ldots$$

$$-V_{L,\alpha} - V_{L,f_c} - V_{L,T} - V_{\alpha,f_c} - V_{\alpha,T} - V_{f_c,T} - V_L - V_\alpha - V_{f_c} - V_T$$

$$-V_{C_b,S7} - V_{C_b,S81} - V_{C_b,S82} - V_{C_b,S83_c} - V_{C_b,S84} - V_{C_b,S85} - V_{C_b,S86} - V_{C_b,S87} - V_{C_b,S88} - V_{C_b,W} - \ldots$$

$$-V_{S7,S81} - V_{S7,S82} - V_{S7,S83_c} - V_{S7,S84} - V_{S7,S85} - V_{S7,S86} - V_{S7,S87} - V_{S7,S88} - V_{S7,W} - \ldots$$

$$-V_{S81,S82} - V_{S81,S83_c} - V_{S81,S84} - V_{S81,S85} - V_{S81,S86} - V_{S81,S87} - V_{S81,S88} - V_{S81,W} - \ldots$$

$$-V_{S82,S83_c} - V_{S82,S84} - V_{S82,S85} - V_{S82,S86} - V_{S82,S87} - V_{S82,S88} - V_{S82,W} - \ldots$$

$$-V_{S83,S84}-V_{S83,S85}-V_{S83,S86}-V_{S83,S87}-V_{S83,S88}-V_{S83,W}-\dots$$

$$-V_{S84,S85}-V_{S84,S86}-V_{S84,S87}-V_{S84,S88}-V_{S84,W}-\dots$$

$$-V_{S85,S86}-V_{S85,S87}-V_{S85,S88}-V_{S85,W}-\dots$$

$$-V_{S86,S87}-V_{S86,S88}-V_{S86,W}-\dots$$

$$-V_{S86,S88}-V_{S86,W}-\dots$$

$$-V_{C_b}-V_{S81}-V_{S82}-V_{S83_c}-V_{S84}-V_{S85}-V_{S86}-V_{S87}-V_{S88}-V_{W} \quad (5)$$

\The sum of all the order effects that a factor accounts for is called the 'total' effect. Having a model with four input factors, the total sensitivity index firstly for input factor C_b would then be computed according to equation 5 and similarly for the remaining inputs.

$$ST_{C_b}=S_{C_b}+S_{C_b,S7}+S_{C_b,S81}+S_{C_b,S82}+S_{C_b,S83}+S_{C_b,S84}+S_{C_b,85}+S_{C_b,S86}+S_{C_b,S87}+S_{C_b,88}+S_{C_b,W} \quad (6)$$
$$+S_{C_b,S7,S81,S82,S83,S84,S85,S86,S87,S88,W}$$

Based on equation 7, the second order sensivity index (or Total Sensitivity Index) can be computed as follows, based on Sobol's method.

$$S_{Ti}=1-\frac{V\left(E\left[C_k \mid (S7,S81,S82,S83,S84,S85,S86,S87,S88,W,C_b)_{-i}\right]\right)}{V(C_k)} \quad (7)$$

where the $(S7,S81,S82,S83,S84,S85,S86,S87,S88,W,C_b)_{-i}$ denotes the multiple input combination.

3.1.1.2 The Athens test case

For the city of Athens, the urban increment model is implemented to a specific concentration cell (longitude 38 and latitude 24), with scope to approximate the SILAM annual PM_{10} concentration estimate of 16.7ug/m^3. Annual emission data are extract from the corresponding cells, disaggregated per emission snap code in accordance to emission source and a sigma to each emission value is allocated from the uncertainty to the emission factors used. The variation in emission and the associated uncertainty is depicted in Table .

Table 9 Variation is emission and the associated contribution to the total emission

Source type	Mean Annual Emission in ton	Sigma in ton	% Contribution to total Emission
SNAP 7	543	115	24%
SNAP 81	6	0.7	0.3%
SNAP 82	45	6	2%
SNAP 83	0	0	0
SNAP 84	387	80	17%
SNAP 85	0	0	0
SNAP 86	1151	175	51%
SNAP 87	108	10	4.8%
SNAP 88	0	0	0

Also in Table , uncertainty to the remaining inputs is presented, including variation in wind speed and background concentration.

Table 10 Definition of distributions to the remaining input parameters

Other Inputs	Mean	Sigma
Wind Speed (W)	4.85 m/s	0.40
Background Concentration (C_b)	11.80 ug/m^3	0.70
City Area	376.90 m^2	N/A

The Sobol method was used to deduce the impact of these aforementioned input variables on the urban concentration. In particular simulations show that the highest Sobol indices SI and TSI are attributed to the background concentration (SI:0.29,TSI:0.29) followed by the wind speed (SI:0.23 and TSI: 0.24) in the parameter space investigated, as summarized in Table .

Table 11 Sensitivity indices per input to the Urban increment model

Inputs	Sensitivity Index	Total Sensitivity Index
SNAP 7	0.09	0.08
SNAP 81	0.001	0.01
SNAP 82	7.931e-04	0.01
SNAP 83	0	0
SNAP 84	0.04	0.03
SNAP 85	0	0
SNAP 86	0.20	0.20
SNAP 87	0.13	0.12
SNAP 88	0	0
W	0.23	0.24
C_b	0.29	0.29

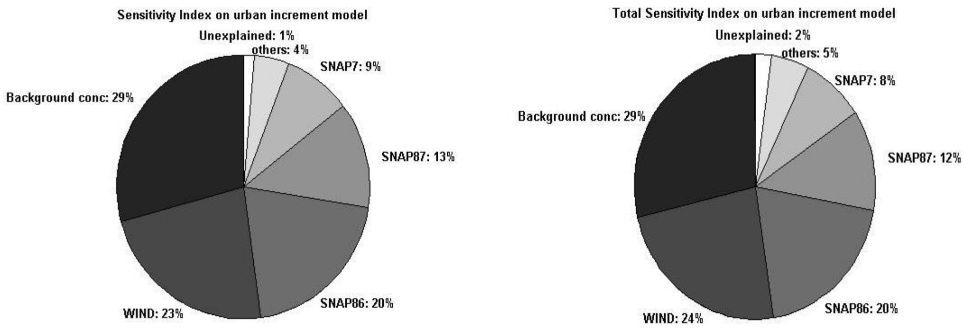

Figure 1. Sensitivity and Total sensitivity Index to the Urban Increment Model

4 CONCLUSIONS

To summarize, the method utilized is a Tier 2 Monte Carlo for all elements composing a full chain health impact assessment, as presented in D.4.2.1. In accordance for the emission source, uncertainty to the emission factors, traffic fleet composition and activity data was quantified via normal distributions. For the air quality modelling, the urban increment model was used to approximate the regional model output from the regional models (SILAM, EMEP, CMAQ and Lotus Euros). Regional model performance was evaluated against the airbase observations for both the rural and urban areas. The inherent uncertainty from this comparison was used to associate uncertainty to the estimated concentration estimates, utilized to update the population weighted concentrations at a cell or city level. For the health impact, the confidence intervals associated with the concentration-response functions (CRFs) were used to fit probability distributions, in order to demonstrate how uncertainty is propagated from exposure to health for given population at a cell or city level.

5 REFERENCES

Sobol IM. 1993. Sensitivity analysis for nonlinear mathematical models. Mathematical Modeling and Computational Experiment 1(4): 407-414.

Torras Ortiz S, Friedrich R. 2013. A modelling approach for estimating background pollutant concentrations in urban areas. Atmospheric Pollution Research 4(2): 147-156

.