## *Summary*

The long-range transport modelling is widely used for the evaluation of atmospheric transmission, deposition, and accumulation in the environment of toxic long-lived pollutants as heavy metals. Due to its peculiar character (large scale, long-time periods, and significant vagueness of input data) the modelling is connected with essential uncertainty of the output results. The present technical note is devoted to composite analysis of heavy metals modelling uncertainty due to inaccuracy of model parameters and input data. The analysis consists of four parts: sensitivity analysis, evaluation of the model uncertainty due to individual parameters, stochastic simulation of the overall model uncertainty, and comparison of the modelling results with measurements. In the first part the model sensitivity to variation of input parameters is investigated. It is shown that the modelling results are most sensitive to variation of anthropogenic emission. The model sensitivity to other parameters is estimated as well. In the second part the model uncertainties due to input parameters taken separately are considered in order to compare their relative contribution to the overall uncertainty. It is obtained that the model error due to anthropogenic emission significantly exceeds others and its value is quite close to the accepted uncertainty of emission itself (~24%). The third part of the analysis is devoted to the assessment of the overall uncertainty by means of stochastic simulation using Monte Carlo method with Halton Quasi-Random Sampling. Results of the simulation shows that the overall uncertainty of the pollutant annual total deposition mainly varies within the range 25-30% and correlates with spatial distribution of anthropogenic emission, while the uncertainty of mean annual air concentration of the pollutant amounts to 40% in the main part of the computation domain. The intrinsic model uncertainty without influence of anthropogenic emission does not exceed 10% for the total deposition and 25% for the mean air concentration over the main part of the computation domain. In the last part the modelling results are compared with available measurement data taking into account calculated uncertainties. It is deduced that inaccuracy of all model parameters considered in the analysis cannot fully account for the discrepancy of the calculated and observed values. An additional error is due to other uncertainty sources not included in the current research.

# *Contents*

# 1. Introduction

The atmospheric transmission of long-lived toxic pollutants is a topical problem arousing growing concern of general public and scientific community. Heavy metals (such as lead, cadmium, mercury etc.) are among the most harmful pollutants impacting both on the human health and the environment. Being emitted in industrial regions they are transported far from the sources and may be deposited even on remote relatively clear areas. To assess the pollutants transport, deposition, and accumulation in the environment the long-range transport (LRT) modelling is widely used.

Large scale modelling of the pollutants airborne transport generally deals with complicated atmospheric processes occurring over a large territory (a continent) during a long time period (a year). Impossibility to take into consideration all details of physical and chemical processes of the pollutant transformations and scavenging from the atmosphere as well as all meteorological and geographical conditions compels to apply some approximations and simplifications of the reality. Besides input information and model parameters destined to bind the modelling to certain conditions (such as amount and spatial distribution of anthropogenic emission etc.) are often rather uncertain themselves. All mentioned above inevitably leads to a considerable uncertainty of modelling results, and, in this aspect, it is quite important to obtain reasonable estimates for possible inaccuracy of the model.

The operation of a numerical LRT model may be in general described by the following formula:

$$\mathbf{Y} = \hat{G}(\mathbf{X}), \tag{1}$$

where **X** and **Y** are vectors of model input parameters and output variables respectively; and $\hat{G}$ is the model operator acting on **X**. In this notations the tensor of partial derivatives

$$S_{ij} = \partial Y_i / \partial X_j \tag{2}$$

characterizes sensitivity of the model at the given set of parameters values with respect to variation of the input parameters **X**, where each particular derivative represents sensitivity of a certain output variable $Y_i$ to appropriate input parameter $X_j$. Thus the sensitivity tensor makes possible to estimate relative influence of different parameters on the model output variables [e.g. *Derwent*, 1987; *Pai et. al*., 1999]. This approach supplies general, mostly qualitative information about the model dependence upon the input data. Besides, applying

information about plausible ranges of input vagueness one can assess the model error due to each input parameter taken separately in order to compare individual contributions of the parameters to the overall model uncertainty [*Alcamo and Bartnicki*, 1990]. However, to evaluate the magnitude of the overall uncertainty itself one has to consider the influence of all input parameters simultaneously. In simple linear models with non-correlated input parameters the overall uncertainty may be estimated by calculating the sum of squares of the model errors due to each individual parameter uncertainty. For more complicated models one can apply enhanced sensitivity approaches such as Decoupled Direct Method [*Dunker*, 1981 and 1984] or Fourier Amplitude Sensitivity Test [*Tilden and Seinfeld*, 1987]. In case of essentially non-linear models or models with strongly correlated input parameters even those improved methods based on the sensitivity analysis becomes hardly applicable to assessing the overall model uncertainty, but still can assist to classify input parameters in order to choose most significant of them.

Another possible approach to the problem is a stochastic simulation based on Monte Carlo method [*Sobol*, 1973; *Ang and Tang*, 1984]. It allows to take into account the collective influence of all input parameters on the overall uncertainty even in strongly non-linear models. The key point of the Monte Carlo simulation is a sampling procedure of input parameters values. Since each model run of the stochastic simulation often takes considerable computational time (from tens of minutes to several days) it becomes difficult to apply ordinary Monte Carlo method with Simple Random Sampling (SRS) requiring great number of runs [e.g. *Morgan et. al*., 1984; *Alcamo and Bartnicki*, 1987]. To reduce the runs number Hanna and co-workers [*Hanna et. al*., 1998] used SRS procedure estimating statistical tolerance limits of the output variables instead of confidence intervals. The tolerance limits confine the interval containing definite part (say 90%) of the output values population with certain probability (e.g. 95% confidence). Another method allowing to use a limited number of samples is based on the Latin Hypercube Sampling (LHS) technique [*McKay et. al*., 1979]. It was successfully applied to uncertainty analysis of various modelling systems: long-range transport and deposition of sulphur [*Alcamo and Bartnicki*, 1990] and nitrogen compounds [*Derwent*, 1987], regional mercury exposure [*Seigneur et. al*., 1998], and complex photochemical models of acid deposition [*Gao et. al*., 1996]. However, it was often found to underestimate the variance of output model variables [*NCRP*, 1996; *Hanna et. al*., 1998]. In the uncertainty analysis presented here the Halton Quasi-Random Sampling (HQRS) technique was used, which is described in Section 5 of the report.

Thus, the procedure of the composite uncertainty analysis of a LRT model appears to be as follows:

- Evaluation of vagueness ranges of the model input parameters
- Analysis of the model sensitivity
- Estimation of the model uncertainty due to individual parameters
- Stochastic simulation of the overall model uncertainty

The present research is devoted to composite analysis of heavy metals LRT modelling uncertainty due to the inaccuracy of model parameters and input data. In the analysis the operational 3D Eulerian LRT model [*Pekar*, 1996; *Ryaboshapko et. al*., 1999] is used to simulate transport and deposition of heavy metals over Europe.

The model parameters and input data treated in the uncertainty analysis are considered in Section 2 of the research as well as the model scheme simplifications being necessary for the stochastic simulation to be applied. Section 3 contains analysis of the model sensitivity to variation of most important input parameters. Both local and spatially averaged sensitivity coefficients of the model output variables are calculated and the input parameters are classified in accordance with their influence on the model results. Individual contributions of input parameters to the overall model uncertainty are investigated in Section 4. Section 5 is devoted to stochastic simulation of the overall model uncertainty. It contains detailed description of the simulation procedure in use and discussion of the results. Spatial distribution of the modelling relative error is calculated both for deposition and concentration of the pollutant in air. Influence of anthropogenic emission vagueness on the model uncertainty is estimated. In Section 6 modelling results of heavy metals transmission over Europe are compared with available measurements taking into account the error intervals. The outcome of the uncertainty analysis is shortly summarized in Conclusions.

# 2. Model parameters and simplifications

The uncertainty analysis of heavy metals LRT modelling is performed for conditions of lead airborne transport over Europe in 1996. It is assumed that lead is emitted and transported trough the atmosphere only in the particulate form. Moreover, even if it undergoes some chemical transformations it does not result in change of its aggregate state. However, it is not common situation for all heavy metals. For instance, mercury occurs in the atmosphere in various chemical and aggregate forms and is subjected to rather complicated transformation both in the gaseous and liquid phase. Nevertheless, since most of atmospheric pollutants (not heavy metals only) contain the particulate form, results of the analysis reflect the most

general model characteristics describing long-range transport and scavenging of the pollutant aerosol constituent from the atmosphere.

The model parameters and input data treated in the uncertainty analysis are presented in table 1. Appropriate set of parameters is chosen in order to take into account both possible uncertainty of input data (e.g. anthropogenic emission $Q_{anth}$, wind velocity $\boldsymbol{U}_{wind}$ etc.) and an additional error introduced due to inaccuracy of the calculation procedure of inherent model variables such as friction velocity $u_*$, mixing layer height $H_{mix}$, and dry deposition velocity $V_d$. Variance of the model parameters is simulated by multiplying their base value used in the routine computations by the *variation multiplier*. Thus, the base value $X^*$ of a parameter $X$ is multiplied by the variation multiplier $K_X$ [1]:

$$X = K_X X^* . \tag{3}$$

It allows to maintain dependence of the parameter on other parameters and, on the other hand, to assume the variation multipliers to be independent, because each of them characterises peculiar source of uncertainty. Indeed, e.g. dry deposition velocity is a function of friction velocity and roughness of underlying surface. Both of them are the sources of possible uncertainty. But calculation procedure (approximations, formulas etc.) of dry deposition velocity itself can introduce additional uncertainty independent of the first to sources. Thus, the variation multiplier of dry deposition velocity may be considered as independent of the multipliers of roughness and friction velocity. Nevertheless, the final values of these parameters may considerably correlate between each other.

Table 1 contains information on each parameter in the form of variance range of appropriate variation multipliers along with the shape of their probability distribution. For parameters normally distributed over the interval the standard deviation $\sigma$ is given instead of the variance range. Unfortunately, it is quite difficult to obtain accurate estimates of input parameters inaccuracy, especially, of the inherent model variables, dealing with physical assumptions and calculation procedures such as friction velocity, dry deposition velocity etc. In this case we tried to assign the most reasonable error intervals. The parameters are collected into two groups depending on whether their variation multipliers are changed during a computation run or not. This property of the variation multiplier is introduced to take into account independent errors of temporally varied parameters (like wind velocity) at each their input to the model and is applied only to the stochastic simulation of the overall model uncertainty.

[1] The angle $\varphi$ of wind velocity direction to the axis of abscissas of EMEP grid is an exception. In this case the variation coefficient is added to the base value

**Table 1.** Parameters of the model treated in the uncertainty analysis

| Parameters | | | | |
|---|---|---|---|---|
| Parameter | | Variation multiplier, $K_x$ | | |
| | | Distribution | Median | Range |
| Anthropogenic emission, $Q_{anth}$ | | Uniform | – | 0.75 ÷ 1.25 |
| Natural emission, $Q_{nat}$ | | Uniform | – | 0.5 ÷ 1.5 |
| Boundary concentration, $C_{bound}$ | | Uniform | – | 0.5 ÷ 1.5 |
| Friction velocity, $u_*$ | | Triangular | 1 | 0.5 ÷ 1.5 |
| Mixing layer height, $H_{mix}$ | | Triangular | 1 | 0.5 ÷ 1.5 |
| Dry deposition velocity, $V_d$ | | Normal | 1 | $\sigma$ = 0.2 |
| Changeable parameters | | | | |
| Parameter | | Variation multiplier, $K_x$ | | |
| | | Period | Distribution | Median | Range |
| Wind velocity, $\boldsymbol{U}_{wind}$ | Abs. value, $U_{wind}$ | 6 h | Normal | 1 | $\sigma$ = 0.1 |
| | Angle, $\varphi$ | 6 h | Normal | 0 | $\sigma$ = 5° |
| Temperature, $T$ | | 6 h | Normal | 1 | $\sigma$ = 0.1 |
| Precipitation intensity, $I_{prec}$ | | 6 h | Normal | 1 | $\sigma$ = 0.2 |
| Cloudiness, $N_{cloud}$ | | 6 h | Normal | 1 | $\sigma$ = 0.1 |
| Roughness, $z_0$ | | 1 month | Triangular | 1 | 0.5 ÷ 1.5 |
| Washout ratio, $W$ | | 1 month | Normal | 1 | $\sigma$ = 0.2 |

An error analysis of a numerical model (especially the stochastic simulation) is inevitably bound up with a large capacity of computations. For this reason one have to resort to some simplifications of the routine model in order to shorten the computation time. In case of the model under consideration the standard EMEP domain was transformed cutting down the Atlantic part of the domain and using rougher grid with 150 km spatial resolution. Reduction of the computation domain can only cause changes in the vicinity of the new boundaries, on the other hand, roughing the grid can have some smoothing effect and results in the additional error at points of extreme values of a pollutant deposition and its concentration in the air. The comparison of modelling results obtained with 50 km and 150 km grid resolutions has shown that the grid roughing leads to a significant error (more than 100%) only in the immediate proximity to intensive sources of the pollutant emission, while in the remained part of the domain the error does not exceed 30-40% of magnitude.

# 3. Analysis of model sensitivity

Sensitivity of the model to a certain parameter was studied by altering its variation multiplier in the range 0.25 ÷ 2 with step 0.25 and performing a complete annual computation run for each value of the multiplier. The data of anthropogenic emission and meteorology were taken for 1996. The annual values of total (dry and wet) deposition flux and mean concentration of the pollutant in the atmospheric surface layer averaged over the whole domain were chosen as outputs of the analysis. Results of the sensitivity analysis are presented in figures 1-4.

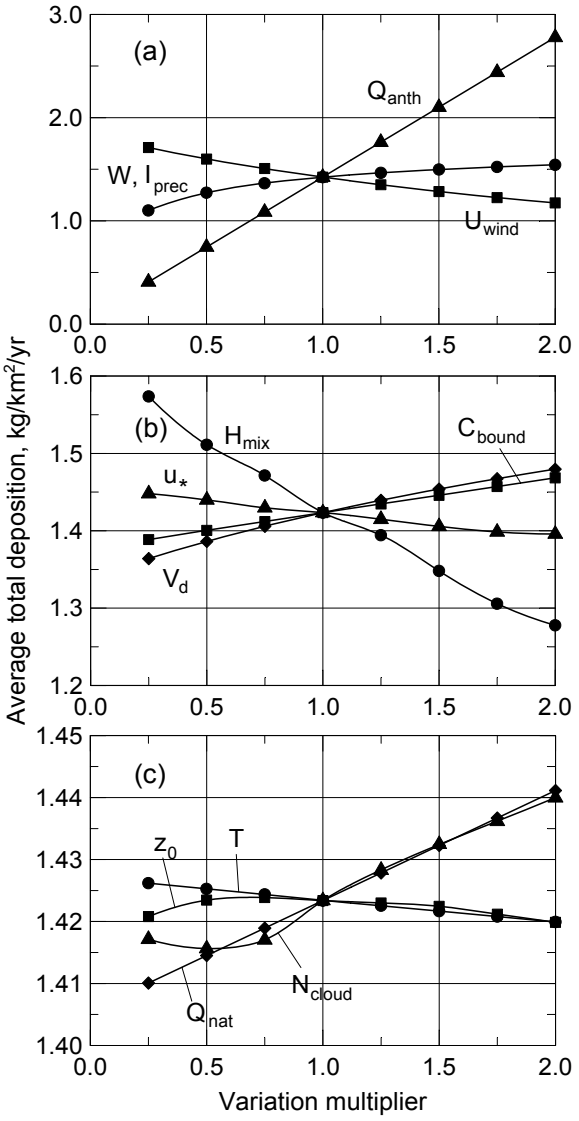


**Figure 1.** Dependence of annual total (dry plus wet) lead deposition flux averaged over the whole domain on variation of input parameters

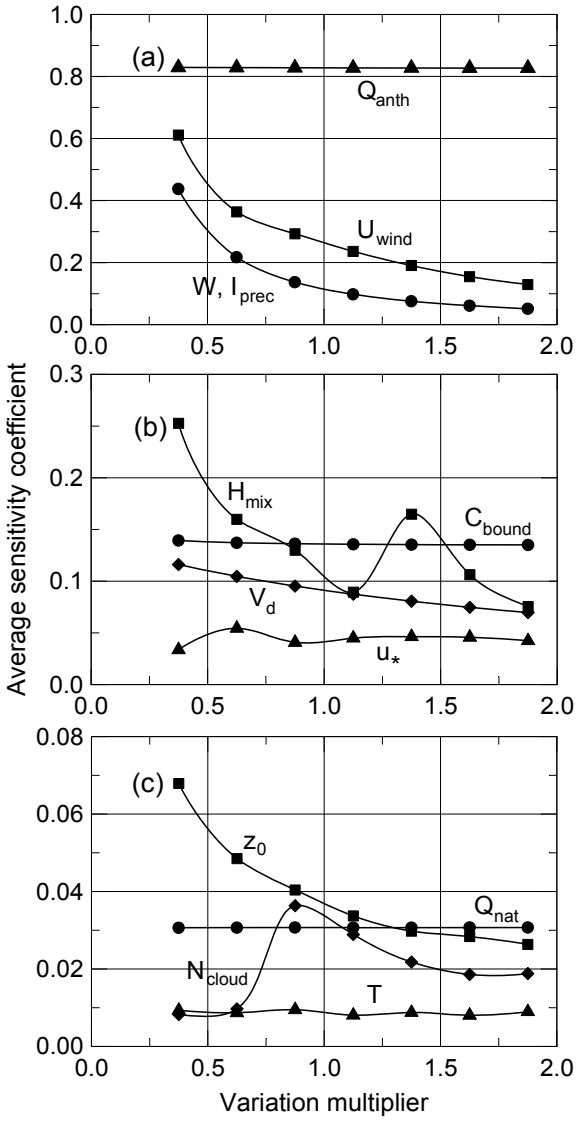


**Figure 2.** Sensitivity coefficient of average annual total (dry plus wet) lead deposition with respect to variation of input parameters

Figures 1 and 3 show the dependence of the average total deposition flux and the mean concentration of lead in the air respectively on variation of different parameters of the model. Sensitivity curves are gathered into three groups (figs. 1(a) - 1(c)) in accordance with their scales to make figures more illustrative. It should be noted that the curve for precipitation intensity $I_{prec}$ exactly coincides with that for washout ratio $W$, because these two parameters occur in the model only as the product $WI_{prec}$. Further we shall talk about the precipitation intensity meaning both of them.

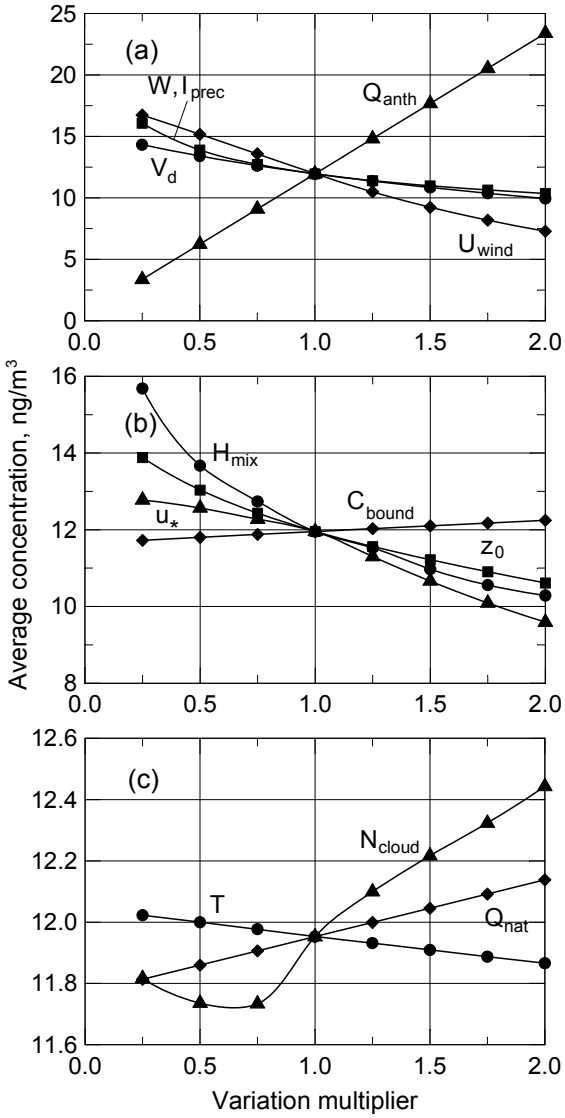


**Figure 3.** Dependence of mean annual lead concentration in the air averaged over the whole domain on variation of input parameters

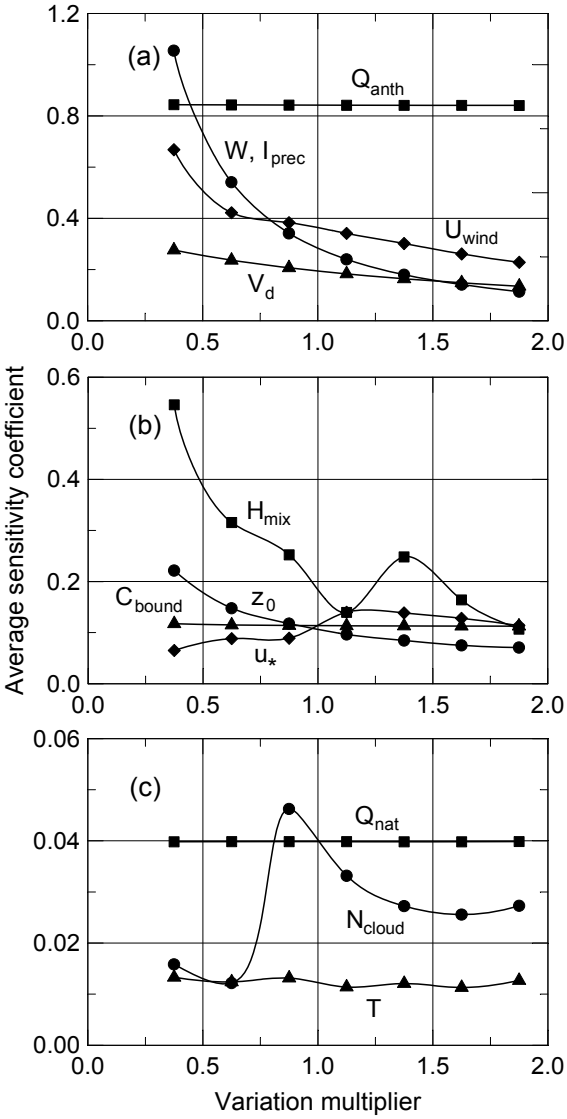


**Figure 4.** Sensitivity coefficient of mean annual lead concentration in the air with respect to variation of input parameters

In order to characterize the sensitivity of output variable $Y$ to variation of the model parameter $X$ we introduce non-dimensional normalized *sensitivity coefficient*:

$$S.c. = \left|\frac{\Delta Y}{\Delta X}\right| \frac{X^*}{Y^*} = \frac{|\Delta Y|}{Y^* \Delta K_X}, \tag{4}$$

where $\Delta X$ and $\Delta Y$ are the differences of parameter $X$ and variable $Y$ in two successive runs respectively; $X^*$ and $Y^*$ are the base values of the parameter and the variable corresponding to the variation multiplier being equal to unity; $\Delta K_X$ is the difference of the variation multiplier of the parameter $X$ in two successive runs. Moreover, to characterize sensitivity of the output distribution field as a whole one can evaluate average sensitivity coefficient:

$$\overline{S.c.} = \frac{\frac{1}{N}\sum_{i,j}\left|\Delta Y_{i,j} / Y^{*}_{i,j}\right|}{\Delta K_X}, \qquad (5)$$

where $\Delta Y_{i,j}$ and $Y^{*}_{i,j}$ are the difference in two successive runs and the base value of the output variable $Y$ in cell *(i,j)*, and summing is carried out over $N$ cells of the whole computation domain. The average sensitivity coefficients of the total deposition and the mean air concentration corresponding to sensitivity curves shown in figures 1 and 3 are presented in figures 2 and 4 respectively.

As seen from figures 1–4, both the total deposition and the mean concentration in the air are most sensitive to anthropogenic emission. They grow approximately linearly with increasing of the variation multiplier and have the largest sensitivity coefficients. In the vicinity of the base case ($K_X$ = 1) both the total deposition and the air concentration are also quite sensitive to wind velocity, precipitation intensity, mixing layer height and dry deposition velocity. Unexpected high value of the average sensitivity coefficient of the total deposition to the variation of boundary concentration (see fig. 2 (b)) can be explained by prevalent influence of the last one in the boundary regions (as it will be shown later). The total deposition somewhat grows with precipitation intensity due to the increase of wet deposition and slightly diminishes with rise of wind velocity because of blowing the pollutant outside the computation domain. Moreover, the increase of the mixing layer height causes higher raising the pollutant particles in the atmosphere assisting to transport them outside the domain, and eventually also reduces the total deposition. Obviously, the total deposition increases with rising the dry deposition velocity. On the contrary, growth of all mentioned parameters results in decrease of the air pollutant concentration. Natural emission, cloudiness and air temperature have the least influence on both output variables. Notice, that though roughness of the underlying surface considerably affects dry and, as a consequence, wet components of the deposition flux, the total deposition almost does not depend on it (see fig. 1 (c)).

As a matter of fact, the average characteristics cannot supply exhaustive information about the model sensitivity, because local values of the sensitivity coefficient can considerably differ from the averaged one. To trace variation of the model sensitivity over the computation domain we calculate local sensitivity coefficients at several control points. As the control points we select a number of monitoring stations located at one line stretched from Central Europe through Scandinavia to the boundary of the computation domain. The monitoring stations chosen for the analysis are presented in table 2. The last column of the table contains an artificial control point chosen in the boundary region. The local sensitivity coefficients of the total deposition and the air concentration to variation of six most important

parameters are shown in figures 5 and 6 respectively. As seen from the figures, the influence of anthropogenic emission slackens to the boundary regions, while that of boundary concentration significantly grows. The sensitivity to variation of other parameters is also considerably changed, thus in the region close to the boundary a relative influence of various parameters on the model output considerably differ from that one in Central Europe. Nevertheless, the results of the sensitivity analysis of average characteristics may be applied to the most part of Europe and considerably differ from local ones only in the vicinity of the boundary.

**Table 2.** Monitoring stations selected as control points for the sensitivity analysis

| Monitoring station | | Kosetice (Czech Rep.) | Neuglobsow (Germany) | Rörvik (Sweden) | Nordmoen (Norway) | Namsvatn (Norway) | Øverbygd (Norway) | Boundary region |
|---|---|---|---|---|---|---|---|---|
| Code | | CS3 | DE7 | SE2 | NO44 | NO96 | NO92 | Bnd |
| Location | Lat. | 49°35´N | 53°09´N | 57°25´N | 60°16´N | 64°59´N | 69°03´N | – |
| | Long. | 15°05´E | 13°02´E | 11°56´E | 11°06´E | 13°35´E | 19°22´E | – |
| EMEP coordinates [2] | | (24,17) | (22,18) | (19,20) | (17,22) | (16,25) | (14,28) | (12,31) |

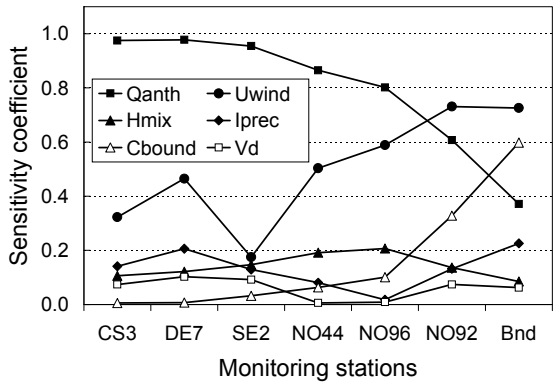


**Figure 5.** Local sensitivity coefficients of annual total lead deposition to variation of model parameters at various control points of the computation domain corresponding to selected monitoring stations

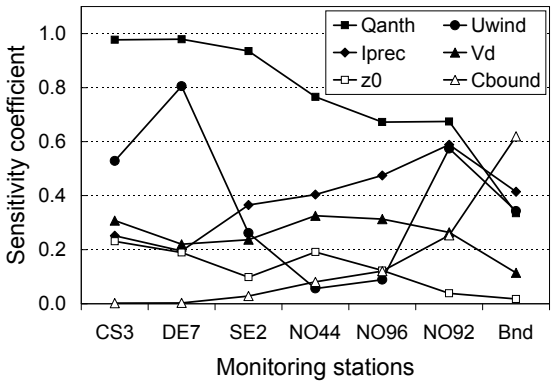


**Figure 6.** Local sensitivity coefficients of mean annual lead concentration in the air to variation of model parameters at various control points of the domain corresponding to selected monitoring stations

# 4. Model uncertainty due to individual parameters

Using the sensitivity curves shown in figures 1 and 3 it is possible to estimate an error of the model output variables due to uncertainty of input parameters taken separately. Uncertainty of the model input parameters are described in table 1. For parameters normally distributed

[2] All coordinates here and further are referred to the standard EMEP grid with 150 km spatial step

over the interval one may use 95.5% confidence interval [$a$ - $2\sigma$, $a$ + $2\sigma$], where $a$ is the median of the probability distribution.

Substituting boundary values of a variation multiplier $K_X$ from the table 1 to the appropriate sensitivity curve shown in figures 1 and 3, one can obtain minimum and maximum possible values of an output variable $Y$ and assess its relative error $\varepsilon_X$ produced by the parameter $X$ only:

$$\varepsilon_X = \frac{Y_{max} - Y_{min}}{2Y^*} 100\% . \qquad (5)$$

Relative errors of average total deposition and mean concentration of lead in the atmospheric surface layer are shown in figures 7 and 8 respectively:

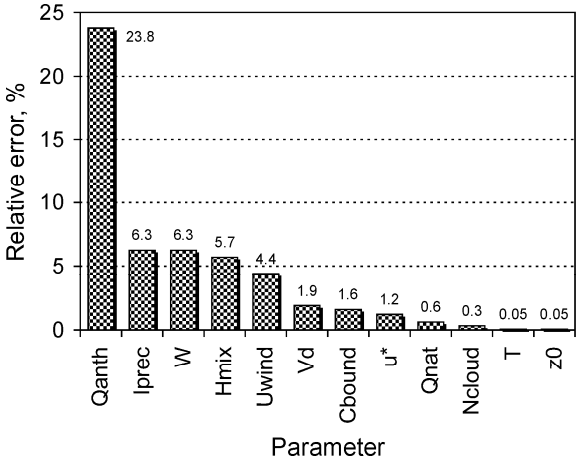


**Figure 7.** Relative error of average annual total lead deposition produced by various model parameters taken separately

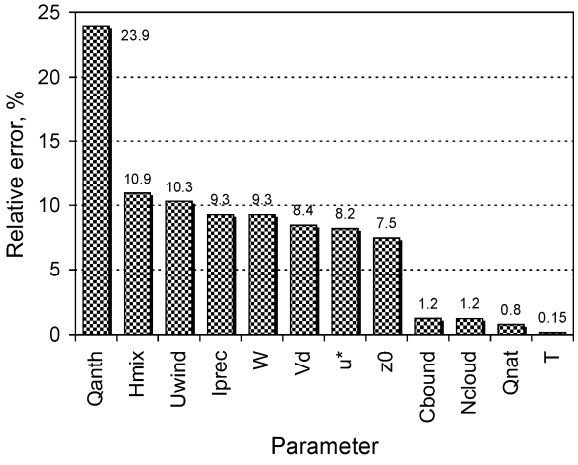


**Figure 8.** Relative error of average annual lead concentration in the air produced by various model parameters taken separately

Each bar of the charts presents an individual error produced by a certain parameter while others assumed to be accurate. As it is seen from the figures, the error due to anthropogenic emission considerably prevails over others and attains value about 24% in both cases. This value is quite close to the accepted uncertainty of anthropogenic emission. Among others the most considerable contribution to the error of the total deposition is made by precipitation intensity (and washout ratio) (6.3%), height of mixing layer (5.7%) and wind velocity (4.4%). As to the air concentration, there are seven the most significant parameters with slightly different values of the contribution – mixing layer height (11%), wind velocity (10.3%), precipitation intensity and washout ratio (9.3%), dry deposition velocity (8.5%), friction velocity (8.2%), and roughness of underlying surface (7.5%). It should be noted that parameters being changed periodically in the model operation, such as wind velocity,

precipitation intensity etc., in reality produce considerably smaller error because of essential averaging during a computation. Indeed, if some input parameter (e.g. wind velocity) is brought in the model every 6 hours of an annual run and does not contain a systematic error, then even a significant random error of its instant value will have a slight effect on the annual result.

Figures 9 and 10 present relative error introduced by various model parameters into dry and wet components of the annual lead deposition averaged over the whole domain. It is quite understandable that precipitation intensity and washout ratio have the most considerable effect (after anthropogenic emission) on the wet deposition (see fig. 9). On the other hand, as seen from figure 10 the error of dry deposition due to dry deposition velocity (32.7%) is even more than that one introduced by anthropogenic emission (24.1%). However, because of a relatively small contribution of dry deposition to the total deposition flux, dry deposition velocity slightly affects the last one (see fig. 7).

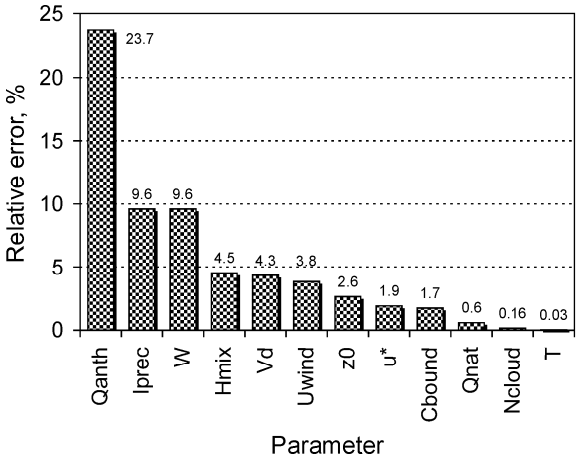


**Figure 9.** Relative error of average annual wet deposition of lead produced by various model parameters taken separately

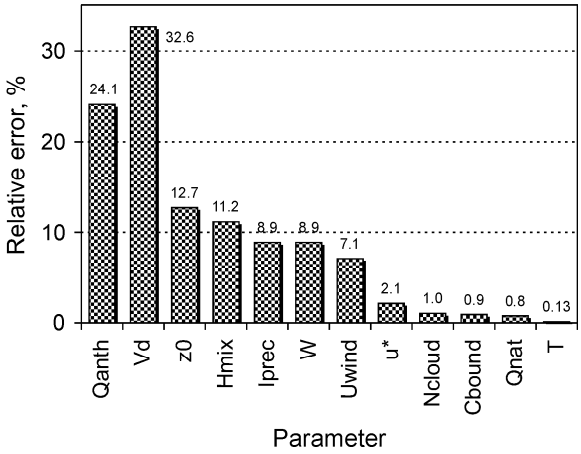


**Figure 10.** Relative error of average annual dry deposition of lead produced by various model parameters taken separately

# 5. Stochastic simulation of the overall model uncertainty

In order to assess the overall model uncertainty due to inaccuracy of all model parameters as a whole one has to consider their collective influence on the model output. A possible approach to this problem is a stochastic simulation based on the Monte Carlo method. In this way all of the treated model parameters are assigned to have a certain probability distribution and then a set of computation runs is performed stochastically sampling the parameters value. As the output of the stochastic simulation, one obtains probability distributions (or

cumulative distribution functions) of the model output variables, which characterize the model uncertainty.

Probability distributions of model parameters treated in the stochastic simulation are described in table 1. As it was said above variability of parameters is simulated multiplying each of them by the appropriate variation multiplier. Therefore all the probability characteristics relate to the variation multipliers instead of parameters themselves. Types of the probability distribution used in the stochastic simulation are shown in figure 11. For the most uncertain parameters (anthropogenic and natural emissions, boundary concentration), when we cannot choose the most probable value, the uniform distribution is used (fig. 11 (a)). The uncertainty of calculation procedure of the inherent model variables (friction velocity and mixing layer height) and that of parameters having definite limits (roughness of the underlying surface) are described by the triangular distribution (fig. 11 (b)). The parameters and input data concerned with measurements, whose uncertainty has random character, are assumed to be normally distributed around a central value (fig 11 (c)). Moreover, some of the input parameters are changed periodically during a computation run (see table 1). Sampling of their values is performed according to their probability distribution with the same frequency as they are put in.

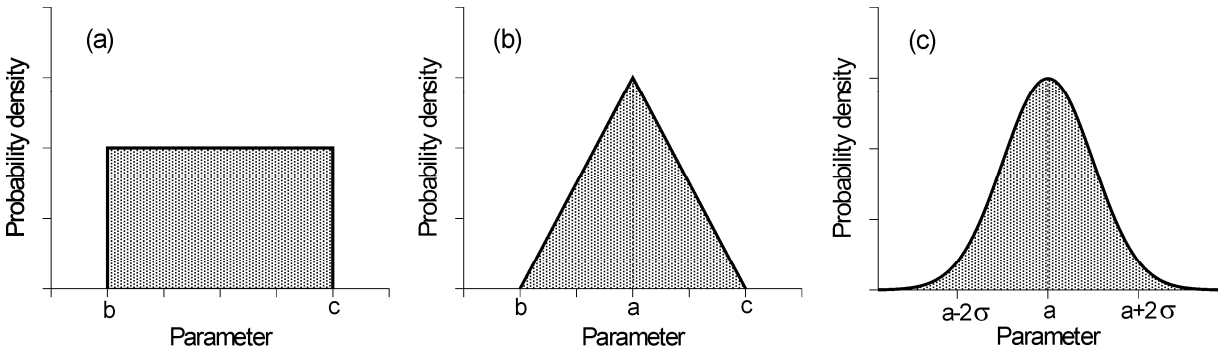


**Figure 11.** Types of probability distribution used in the stochastic simulation:
(a) – uniform; (b) – triangular; (c) – normal

## 5.1. *Simulation procedure*

The stochastic simulation is carried out using Monte Carlo method with Halton Quasi-Random Sampling (HQRS). The classical Monte Carlo method applying Simple Random Sampling (SRS) procedure is as follows. Sufficiently large number *N* of routine computation runs is performed. For each computation run a set of input parameters is generated sampling randomly a value for each parameter according to its probability distribution. The number of computations is chosen to obtain statistically reliable result of the simulation and depends on

the number of varied parameters. Unfortunately, even for a few parameters the number of computations in the simulation with SRS is estimated as many as thousands and is impermissibly expensive for the model in use. Indeed, to sample a set of $n$ values of input parameters $X_1$, $X_2$, …, $X_n$ one has to choose $n$ random numbers $\gamma_1$, $\gamma_2$, …, $\gamma_n$ uniformly distributed over the interval [0,1] and then transform them (e.g. using the inverse functions method [*Sobol*, 1973]) into random values $\xi_i = g(\gamma_i)$ satisfying the probability distributions of the appropriate parameters $X_i$. Imperfection of this technique consists in the fact that, generally speaking, the sets of random numbers $\{\gamma_1, \gamma_2, \ldots, \gamma_n\}$ are not uniformly distributed over $n$-dimensional hypercube $[0,1]^n$ of their values. Therefore one has to generate a great number of the sets to adequately simulate the aggregate probability distribution of sets of input parameters values $\{\xi_1, \xi_2, \ldots \xi_n\}$. To accelerate the convergence of the simulation procedure we use HQRS technique with the quasi-random Halton sequences [*Sobol*, 1973] instead of computer generated random numbers to sample values of the parameters (see Appendix A). The quasi-random sequences are characterized by the property that their elements $\{\gamma_1, \gamma_2, \ldots, \gamma_n\}_k$, $k = 1, \ldots, N$, are uniformly distributed over the hypercube $[0,1]^n$. It makes possible to restrict the number of computations $N$ by a quite feasible value. Numerical experiments have shown (see below), that for the model in use number of computations $N$ = 500 is quite enough for statistically reliable results.

Unchangeable input parameters related to the first group of table 1, are sampled at the beginning of each annual computation run, while those from the second group are changed with their own frequency (once per 6 hours for meteorological data and once per month for roughness and washout ratio). Some of the parameters, such as anthropogenic emission, roughness and meteorological data, have also a spatial distribution. Unfortunately, the consideration of an independent uncertainty for each cell of the grid would drastically increase the number of parameters and, in turn, the number of computations. Instead, the fields of those parameters are varied as a whole, allowing us to assess the upper limits of the uncertainty. The annual total (dry and wet) deposition and mean concentration of lead in the atmospheric surface layer in each grid cell are chosen as the model output variables.

Once performed $N$ computation runs with diverse samples of the parameters one obtains the same number of sets of the model output variables. Qualitative information on an output variable probability distribution shape can be obtained by dividing the whole range of the variable variation into $N_c$=10 cuts and counting the number of samples concerned each cut of values. A bar chart reflecting numbers of samples fallen into each cut represents the probability distribution of the variable (e.g. see fig. 14 (e)). For quantitative estimates of the overall uncertainty the *cumulative distribution function* of the variable is calculated:

$$F(Y) = \int_{-\infty}^{Y} p(x)dx \cong \sum_{i=1}^{k} \frac{i}{N}, \qquad Y_k \le Y < Y_{k+1}, \tag{7}$$

where $p(x)$ is the probability density function of the variable $Y$ described by the bar charts mentioned above; the lower indices $i$ and $k$ denote the number of sample ordered in ascending values. The cumulative distribution function of mean annual air concentration of lead averaged over the computation domain is presented in figure 12. Using the cumulative distribution function one can evaluate the upper $Y_{up}$ and lower $Y_{low}$ boundaries of the variable 95% confidence interval. The relative error characterizing the overall uncertainty of the output variable is determined as follows:

$$\varepsilon = \frac{Y_{up} - Y_{low}}{2 < Y >} 100\%, \qquad < Y >= \frac{1}{N} \sum_i Y_i \tag{8}$$

where $<Y>$ is a mean value of the variable $Y$. To verify convergence of the procedure we calculated the median and relative error of the variable during the simulation and store intermediate results corresponding to different number of computation runs. Figure 13 shows the mean value (a) and the relative error (b) of mean annual lead concentration in the atmospheric surface layer averaged over the whole domain versus the number of samples. As one can see from the figure both median and relative error of the air concentration acquires an approximately steady magnitude already after 200-300 computation runs. However, we have to perform additional number of computations in order to get the final shape of the probability distribution. Figure 14 shows bar charts of probability distribution of the air concentration for the number of computations from $N$=100 to 500 with step 100. As seen from the figure the shapes of the distribution are practically not changed during last hundred of runs.

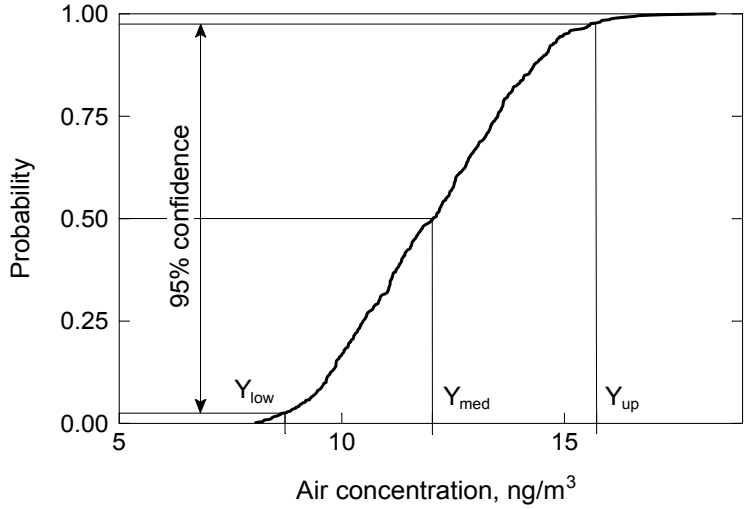


**Figure 12.** Cumulative distribution function of mean annual lead concentration in the air averaged over the whole domain

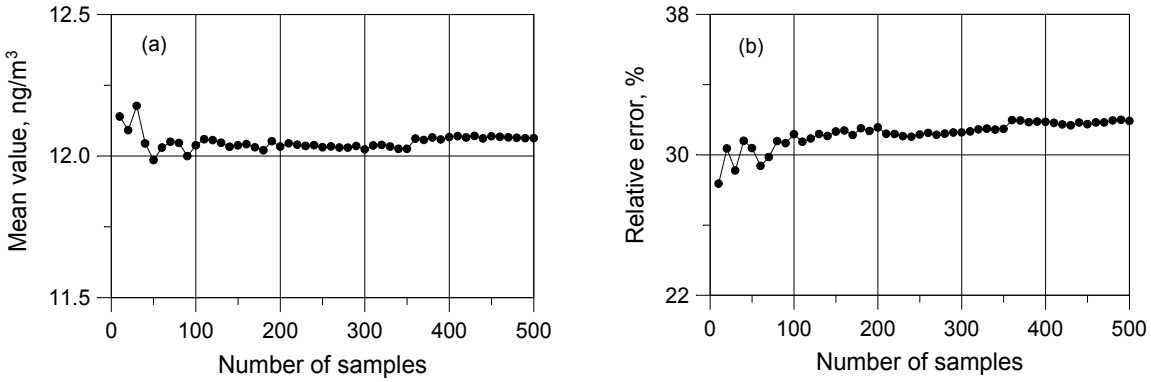


**Figure 13.** Dependence of the mean value (a) and relative error (b) of annual lead concentration in surface layer averaged over the whole domain on the number of samples

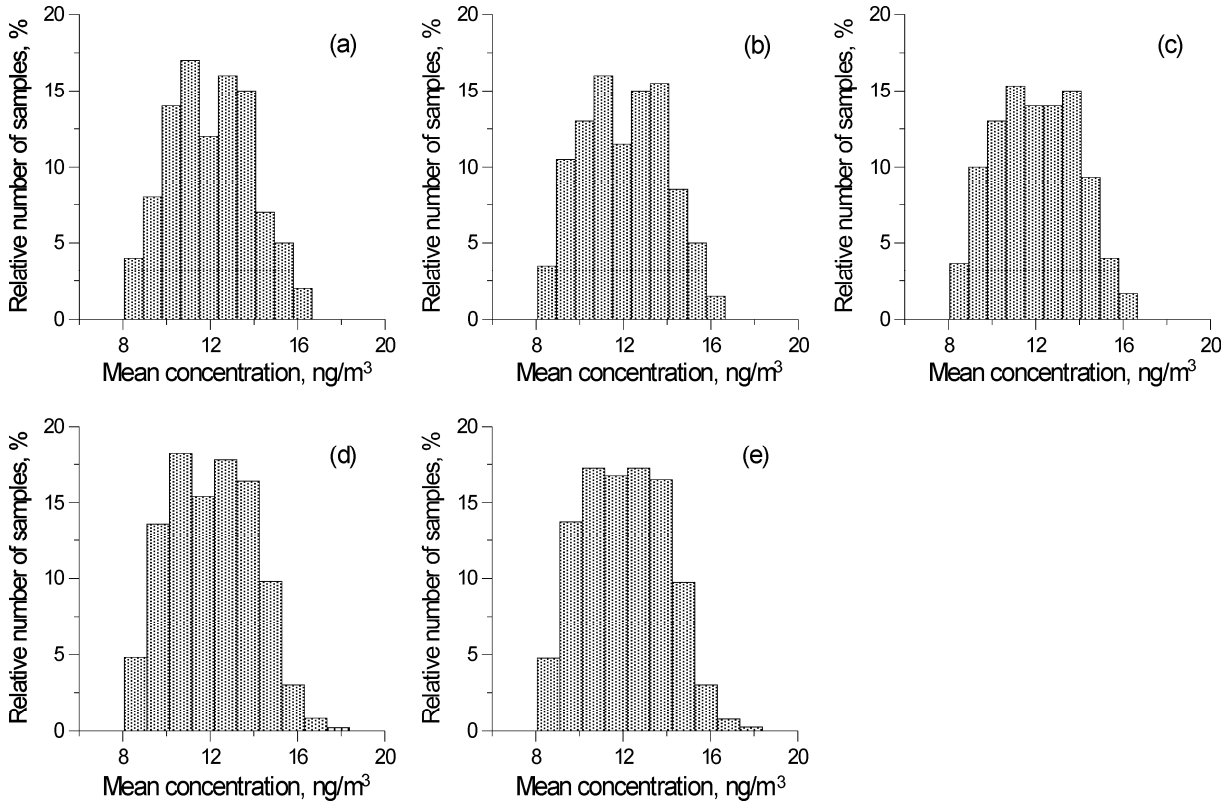


**Figure 14.** Intermediate probability distributions of annual lead concentration in air averaged over the whole domain corresponding to different number of samples: (a) – *N* = 100; (b) – 200; (c) – 300; (d) – 400; (e) – 500

## 5.2. *Spatial uncertainty distribution*

Results of the stochastic simulation are presented in figures 15-19. Figure 15 shows spatial distribution of the relative error of annual total (dry and wet) lead deposition over the computation domain. As it is seen from the figure the uncertainty of total deposition does not exceed 30% over the main part of the domain. Large values of the error up to 50% are achieved in the boundary regions. It can be explained by the determinative role of the

boundary concentration error on the total uncertainty near the boundaries. It should be noted, that even relative error of the total deposition somewhat correlates with anthropogenic emission distribution illustrated in figure 20: The error amounts to considerable values (up to 30%) in the regions with intensive anthropogenic emission. On the other hand, deposition on areas far from the emission sources but not too close to the boundaries has the minimum uncertainty (down to 24%). The variation of the total deposition uncertainty over the domain (except boundary regions) is quite small and does not exceed 6%. Figure 16 shows the distribution of relative error of mean annual lead concentration in the surface atmospheric layer. One can say that, in general, the uncertainty of the air concentration is larger than that one of the total deposition and amounts to 40% in the main part of the domain. There is no noticeable correlation between relative error of the air concentration and distribution of anthropogenic emission, except for the fact that the regions with the smallest relative error (26%) are also located far from emission sources. Surprisingly that the effect of boundary concentration is even smaller here than in case of the total deposition and is visible only in the very vicinity of the boundaries, where the uncertainty amounts to 55%.

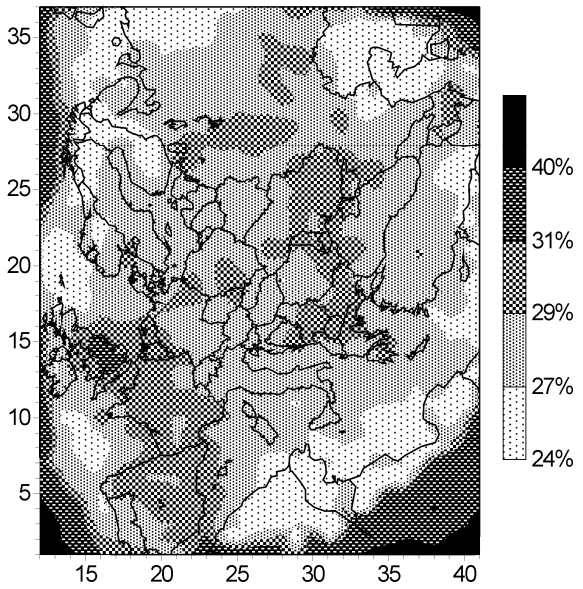


**Figure 15.** Spatial distribution of relative error of annual total (dry and wet) lead deposition

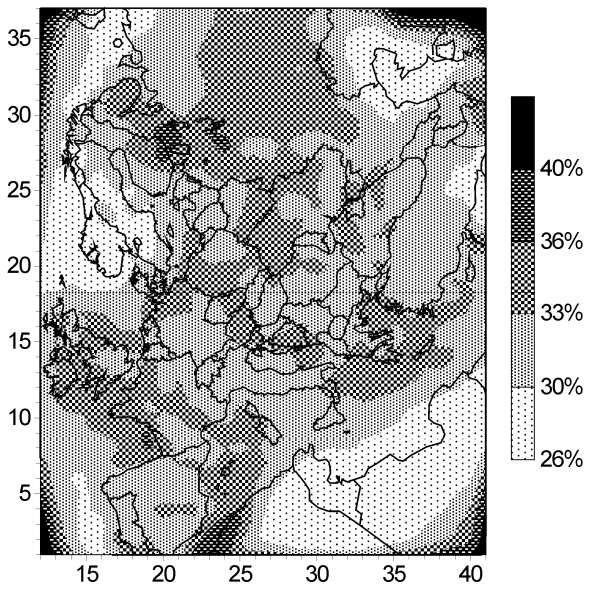


**Figure 16.** Spatial distribution of relative error of mean annual lead concentration in air

It is also useful to assess relative contributions of wet and dry components of the total deposition flux to its overall uncertainty. For this purpose we evaluate the following contribution ratio:

$$\delta = \frac{\Delta_w - \Delta_d}{\Delta_w + \Delta_d}, \qquad (9)$$

where $\Delta_w$ and $\Delta_d$ are the 95% confidence intervals of wet and dry lead deposition fluxes respectively. The ratio $\delta$ takes value 1 when uncertainty of the total deposition is determined only by wet component; value –1 when the uncertainty is only due to dry deposition; and value 0 when their contributions are equal to each other. Spatial distribution of wet-dry contribution ratio $\delta$ is shown in figure 17. It is seen that the contribution of wet component of the deposition flux to the total uncertainty predominates almost all over the whole domain. Only over African part and two European regions of the computation domain the dry deposition plays a significant role. Over marine basins the overall uncertainty of the total deposition is almost entirely determined by the vagueness of its wet component.

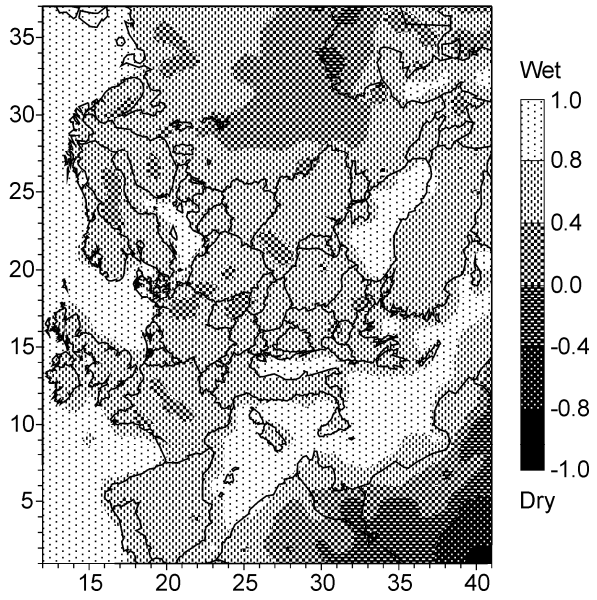


**Figure 17.** Spatial distribution of the contribution ratio of wet and dry components of the deposition flux to the total uncertainty of total annual lead deposition

## 5.3. *Influence of anthropogenic emission*

As it was said above anthropogenic emission plays a special role in the model uncertainty because of its dominating contribution to the overall error both of the total deposition and of the concentration in the air. Besides it may be referred to external data and does not characterize the model itself. Therefore it is useful to get information about the model intrinsic uncertainty without influence of anthropogenic emission. This kind of simulation has also been performed assuming that emission has the accurate value. Figures 18 and 19 show relative errors of the total deposition and the air concentration respectively without influence of anthropogenic emission uncertainty. The intrinsic model uncertainty both of the total deposition and of the air concentration is considerably lower in the main part of the domain than that taking into account anthropogenic emission. The relative error of the total

deposition (fig. 18) mostly does not exceed 10% away from boundary regions. The error distribution has a “concave shape” with the lowest values (2%) at the center of the domain and gradually increasing to the boundaries. Obviously, the boundary error (~50%) due to uncertainty of boundary concentrations does not depend on the anthropogenic emission within the domain. The intrinsic uncertainty of the air concentration (fig. 19) mainly does not exceed 25% and its distribution, in general, looks like that one in figure 16 with the influence of anthropogenic emission: The zones of maximum error are the same in both cases. It means that the uncertainty of other parameters, such as mixing layer height, precipitation intensity, wind velocity etc., also considerably affects the overall uncertainty of the air concentration. For instance, the pronounced zones of minimum intrinsic uncertainty of concentration in air somewhat correspond to regions with maximum annual precipitation amounts (compare figs. 19 and 21).

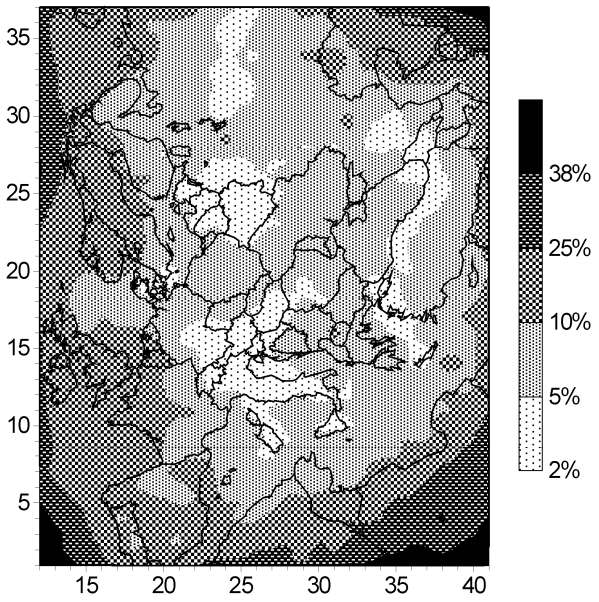


**Figure 18.** Spatial distribution of relative error of annual total (dry and wet) lead deposition without influence of anthropogenic emission uncertainty

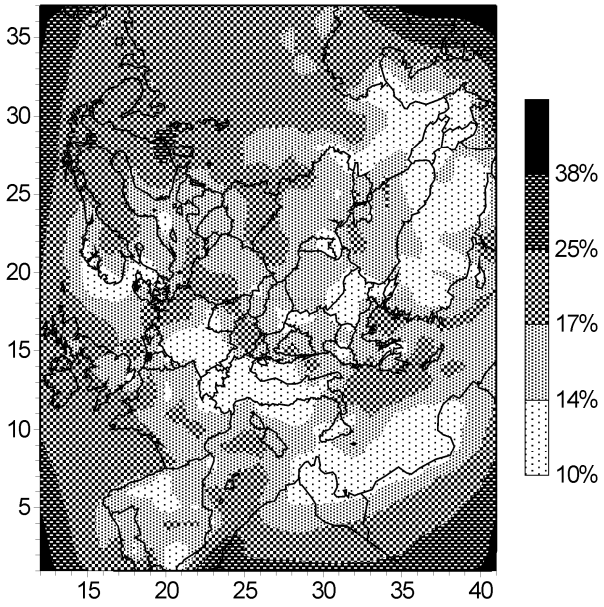


**Figure 19.** Spatial distribution of relative error of mean annual lead concentration in air without influence of anthropogenic emission uncertainty

The influence of anthropogenic emission vagueness on the overall model uncertainty is illustrated in figures 22 and 23. Figure 22 shows typical probability distribution of total lead deposition at one of the grid cells with (a) and without (b) influence of antropogenic emission uncertainty.

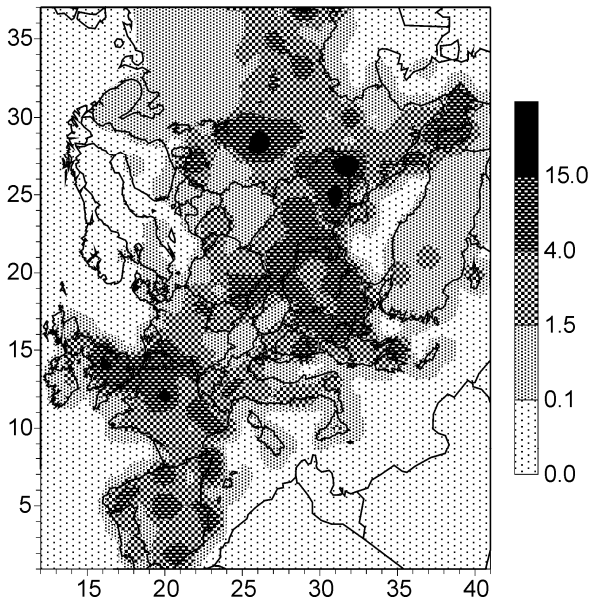


**Figure 20.** Distribution of anthropogenic emission of lead in 1996, kg/km$^2$/yr

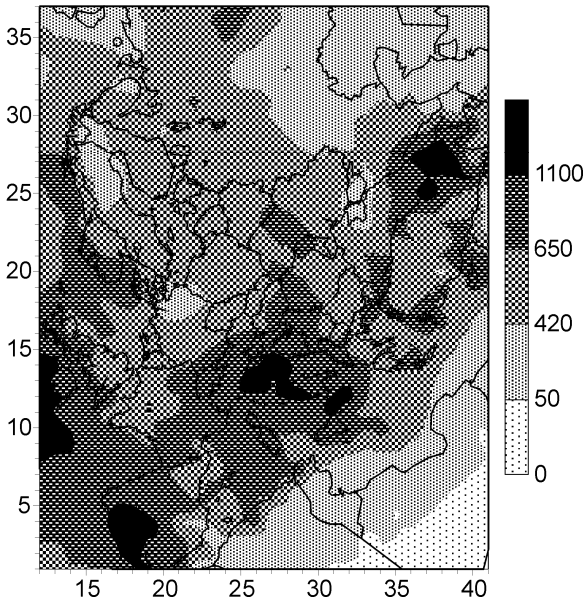


**Figure 21.** Annual amount of precipitation in 1996, mm/yr

As it is seen the effect of the emission consists not only in widening of the probability distribution but also in changing its shape: The narrow normal-shaped distribution of the intrinsic model uncertainty (fig.22 (b)) becomes almost uniform with smoothed edges due to the influence of anthropogenic emission (fig.22 (a)). Both effects lead to a significant increase of the modelling results uncertainty. Concentration of lead in the air is weaker subjected to the influence of the emission uncertainty (appropriate probability distributions are presented in fig. 23). The influence mostly results in widening of the probability distribution keeping its shape almost unchangeable.

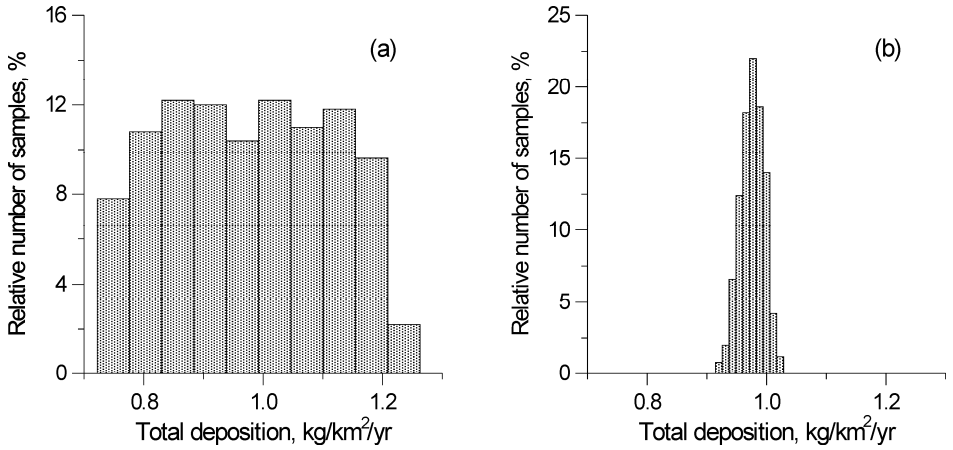


**Figure 22.** Probability distribution of annual total lead deposition with (a) and without (b) influence of anthropogenic emission uncertainty at grid cell (20,20) containing monitoring station Arup, Sweden

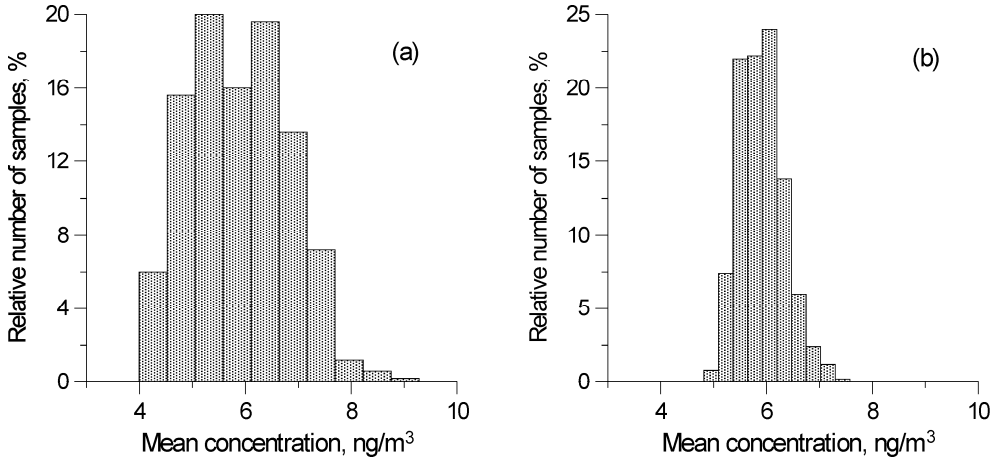


**Figure 23.** Probability distribution of mean annual concentration of lead in atmospheric surface layer with (a) and without (b) influence of anthropogenic emission uncertainty at grid cell (20,20) containing monitoring station Arup, Sweden

# 6. Comparison of modelling results with measurements

The undertaken self-consistent analysis of the model uncertainty cannot in principal comprehend all possible error sources because it is carried out within the framework of the model itself. It is concentrated on the model inaccuracy due to the vagueness of input parameters only remaining out of the consideration such essential error sources as the model formulation (physical and chemical assumptions) and its numerical realization (numerical scheme, methods etc.). Possible ways to assess the contribution of those sources to the real inaccuracy of the modelling results are the comparison of them with results of other LRT models and with available measurements. Models intercomparison is a separate important problem lying out of the current analysis. Here we present the comparison of the modelling results of lead transport and deposition in 1996 with available measurement data taking into account the calculated uncertainties. We use the most reliable measurement data obtained in the network of EMEP monitoring program [*Ryaboshapko et. al.*, 1999]. One has to realize, that the measurements themselves are often somewhat ambiguous and can contain a considerable error due to the imperfection of measuring techniques. Therefore they hardly can identically characterize the error of the modelling results but rather give an overview of their real uncertainty.

Figures 24 and 25 present the comparison of the modelling results with measurements of lead concentrations in the air and precipitation in 1996 respectively at various EMEP monitoring stations. The stations are denoted in accordance with notation accepted in the EMEP monitoring program [*Berg et. al.*, 1996]. The measurement data are shown by crosses, while filled circles with error intervals denote modelling results with calculated

uncertainties (in the form of 95% confidence intervals). As seen from the figures for a number of stations the measured values lie out of the uncertainty intervals, that is inaccuracy of model parameters considered in the analysis cannot account for discrepancy of the calculated and observed values. It could be explained, on the one hand, by insufficiently careful definition of the input parameters vagueness and, on the other hand, by additional uncertainty sources mentioned above. For lead concentration in the air an additional error is about 40-50%, while for the concentration in precipitation it amounts to 100%. In the last case the large value can be also explained by an essential ambiguity of precipitation amount data [*Ryaboshapko et. al*., 1999].

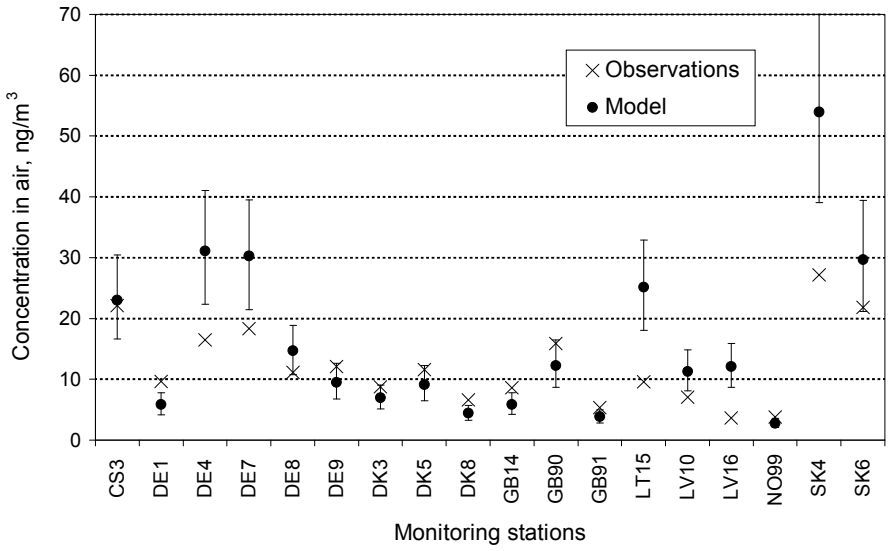


**Figure 24.** Comparison of the modelling results with measurements of mean annual lead concentration in the air in 1996 at various EMEP monitoring stations

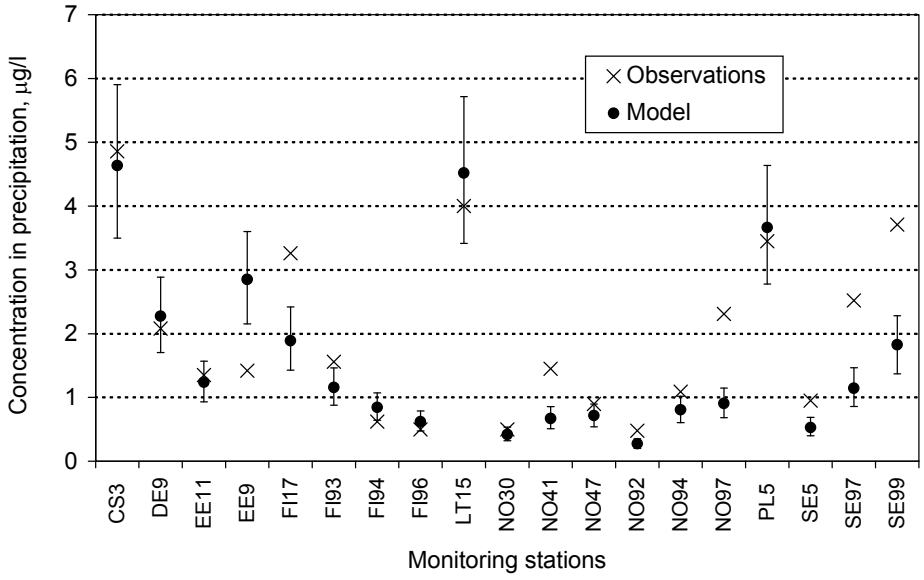


**Figure 25.** Comparison of the modelling results with measurements of mean annual lead concentration in precipitation in 1996 at various EMEP monitoring stations

## *Conclusions*

The composite uncertainty analysis of heavy metals long-range transport modelling has been performed. The investigation of the model sensitivity to the variation of various input parameters has shown, that:

- Modelling results, i.e. annual total (wet plus dry) deposition of the pollutant and its mean annual concentration in the surface atmospheric layer are the most sensitive to variation of anthropogenic emission.
- The most important parameters with respect to model sensitivity also include wind velocity, precipitation intensity, washout ratio, height of mixing layer and dry deposition velocity.

Analysis of the model uncertainty due to the inaccuracy of input parameters taken separately has demonstrated, that:

- The model error due to the vagueness of anthropogenic emission considerably prevails over others and its value is quite close to the accepted uncertainty of emission (~24%).
- The model uncertainty due to other parameters taken separately do not exceed 7% for the total deposition and 11% for the concentration in the air.

Stochastic simulation of the overall model uncertainty due to the collective influence of all parameters has shown, that:

- The overall uncertainty of annual total deposition mainly varies within rage 25-30% and correlates with distribution of anthropogenic emission.
- The uncertainty of mean annual concentration in the atmospheric surface layer does not exceed 40% in the main part of the computation domain.
- There is a significant influence of the boundary concentration vagueness in the vicinity of boundaries of the computation domain, far from the regions of practical interest, where the model uncertainty achieves 50% of the output variable value;
- Relative contribution of wet deposition into overall uncertainty of total deposition flux prevails over that of dry deposition almost all over the whole domain;
- The intrinsic model uncertainty without influence of anthropogenic emission does not exceed 10% for the annual total deposition and 25% for the air concentration over the main part of the computation domain.

The comparison of the modelling results with available measurement data taking into account calculated uncertainties has shown that the inaccuracy of model parameters

considered in the analysis cannot fully account for the discrepancy of the calculated and observed values. An additional error (40-50% for the air concentration and up to 100% for that in precipitation) is due to other uncertainty sources not included in the current research.

## *Acknowledgements*

The author is grateful to Dr. Alexey Ryaboshapko and Mr. Ilia Ilyin for fruitful discussions and immediate help in the research. He is also thankful to Mr. Sergey Dutchak, Prof. Victor Shatalov, and Dr. Alexandre Malanitchev for the real attention to the work and numerous advances. A special thanks to Mrs. Larissa Zayavlina for thorough editing of the text as well as to Ms. Irina Strizhkina and Mrs. Svetlana Yevdokimova for its preparing and publishing.

# *References*


Alcamo J. and J. Bartnicki [1987] A framework of for error analysis of a long-rage transport model with emphasis on parameter uncertainty. *Atmospheric Environment* **21**, 2121-2131.

Alcamo J. and J. Bartnicki [1990] The uncertainty of atmospheric source-receptor relationships in Europe. *Atmospheric Environment* **24A**, 2169-2189.

Ang A. H.-S. and W.H. Tang [1984] *Probability Concepts in Engineering Planing and Design, Vol. 2: Decision, Risk, and Reliability*. Wiley, New York.

Berg T., Hjellbrekke A.-G. and J.E. Skjelmoen [1996] Heavy metals and POPs within the ECE region. EMEP/CCC Report 8/96, Norwegian Institute for Air Research, Norway.

Derwent R.G. [1987] Treating uncertainty in models of the atmospheric chemistry of nitrogen compounds. *Atmospheric Environment* **21**, 2121-2131.

Dunker A.M. [1981] Efficient calculation of sensitivity coefficients for complex atmospheric models. *Atmospheric Environment* **15**, 1155-1161.

Dunker A.M. [1984] The decoupled direct method for calculating sensitivity coefficients in chemical kinetics. *Journal of Chemical Physics* **81**, 2385-2393.

Gao D., Stockwell W.R. and J.D.Milford [1996] Global uncertainty analysis of a regional-scale gas-phase chemical mechanism. *Journal of Geophysics Research* **101**, 9107-9119.

Hanna S.R., Chang J.C. and M.E. Fernau [1998] Monte Carlo estimates of uncertainties in predictions by a photochemical grid model (UAM-IV) due to uncertainties in input variables. *Atmospheric Environment* **32**, 3619-3628.

McKay M.D., Beckman R.J. and W.J. Conover [1979] A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**, 239-245.

Morgan M.G., Morris S.C., Henrion D., Amaral A.L. and W.R. Rish [1984] Technical uncertainty in quantitative policy analysis – a sulfur pollution example. *Risk Anal.* **4**, 201-216.

NCRP [1996] *A Guide for Uncertainty Analysis in Dose and Risk Assessments Related to Environmental Contamination (ed. F.O.Hoffman)*. NCRP Commentary No. 14, National Council on Radiation Protection and Measurements, 7910 Woodmont Ave., Bethesda, Maryland.

Pai P., Karamchandani P., Seigneur C. and M.A.Allan [1999] Sensitivity of simulated atmospheric mercury concentration and deposition to input model parameters. *Journal of Geophysical Research* **104**, 13855-13868.

Pekar M. [1996] Regional models LPMOD and ASIMD: Algorithms, parameterization and results of application to Pb and Cd in Europe scale for 1990. EMEP/MSCE-E Report 9/96, Meteorological Synthesizing Center – East, Moscow, Russia

Ryaboshapko A., Ilyin I., Gusev A., Afinogenova O., Berg T. and A.-G. Hjellbrekke [1999] Monitoring and modelling of lead, cadmium and mercury transboundary transport in the atmosphere of Europe. Joint report of EMEP centres: MSC-E and CCC, EMEP Report 3/99, Meteorological Synthesizing Center – East, Moscow, Russia.

Seigneur C., Lohman K., Pai P., Heim K., Mitchel D. and L. Levin [1998] Uncertainty analysis of regional mercury exposure. *Water, Air, and Soil Pollut.* **112**, 151-162.

Sobol I. M. [1973] *Numerical Monte Carlo methods*. Nauka publishing, Moscow (in Russian).

Tilden J.W. and J.H. Seinfeld [1982] Sensitivity analysis of a mathematical model for photochemical air pollution. *Atmospheric Environment* **16**, 1357-1364.

# *Appendix A*

## Halton Quasi-Random Sampling

For stochastic uncertainty simulation of a model depending on $n$ input parameters $X_1$, $X_2$, …, $X_n$ one has to sample $N$ sets of their values $\{\xi_1, \xi_2, \ldots \xi_n\}_k$, $k$ =1, …, $N$, and perform the same number of computation runs using each value set. To generate a set of the parameters values one chooses $n$ random (or quasi-random) numbers $\gamma_1$, $\gamma_2$, …, $\gamma_n$ uniformly distributed over the interval [0,1] and then transform them into random values $\xi_i = g(\gamma_i)$ satisfying the probability distributions of the appropriate parameters $X_i$. The convergence of the simulation procedure can be accelerated by using quasi-random sequences $\{\gamma_1, \gamma_2, \ldots, \gamma_n\}_k$, $k$ =1, …, $N$, uniformly distributed over the hypercube $[0,1]^n$. The procedure of Halton quasi-random sequences generation is as follows [Sobol, 1973].

Let the number of a computation run $k$ represented in the numeration system with the radix number $r_s$ has the following form

$$k_{r_s} = a_m a_{m-1} \ldots a_2 a_1, \tag{A.1}$$

where $a_i$ = 0, 1, …, $r_s$ –1 are the whole figures of the mentioned numeration system. Then in the decimal numeration the number $k$ is presented in the form

$$k = \sum_{i=1}^{m} a_i r_s^{i-1} . \tag{A.2}$$

The sequence of points in the $n$-dimensional space $K^n$ having orthogonal co-ordinates $\{\gamma_{r_1}(k), \gamma_{r_2}(k), \ldots, \gamma_{r_n}(k)\}_k$, $k$ =1, …, $N$ is called Halton sequence, if

$$\gamma_{r_s}(k) = \sum_{i=1}^{m} a_i r_s^{-i} , \quad s = 1, 2, \ldots, n \tag{A.3}$$

and $r_1$, $r_2$, …, $r_n$ are the mutually prime numbers. As it was proved the elements $\{\gamma_1, \gamma_2, \ldots, \gamma_n\}_k$ of Halton sequences are uniformly distributed over the hypercube $[0,1]^n$.