H14-67

PROPOSAL AND DEMONSTRATION OF A PRACTICAL APPROACH TO IDENTIFY AND PROPAGATE UNCERTAINTIES IN ATMOSPHERIC DISPERSION AND LONG-TERM IMPACT ASSESSMENT STUDIES

Sébastien Argence[1], Patrick Armand[2], Fabien Brocheton[1], Thierry Yalamas[3], François Deheeger[3] and Alexandre Micol[3]

[1]NUMTECH, F-63175 Aubière, France
[2]CEA, DAM, DIF, F-91297 Arpajon, France
[3]PHIMECA Engineering, F-75012Paris, France

Abstract: Long-term impact assessment (LIA) studies still rely on the use of deterministic dispersion modelling systems. Such a methodology does not allow quantifying uncertainties regarding the results of LIA studies while this kind of information is more and more asked by regulatory institutions. For three years, NUMTECH, CEA and PHIMECA have been collaborating to conceive an exhaustive tool dealing with the quantification of uncertainties related to LIA results. This tool is now functional and covers all aspects of a standard uncertainties study from the definition of input uncertainties to the statistical processing of concentration outputs (mean, confidence intervals, probabilities of threshold exceedance, etc.). The paper explains how this tool has been developed and gives examples of its use for a practical case study (with a unique source, a flat terrain and no plume rise) and a Gaussian plume dispersion model.

Key words: *Gaussian plume models, probabilistic modelling, long-term impact assessment studies, uncertainties propagation.*

1. INTRODUCTION

Long-term impact assessment studies generally involved the use of Gaussian plume models to evaluate the risk associated with chronic pollutant releases in the atmosphere. These models are commonly used in a deterministic framework. In other words, for one particular LIA study, only one model run is performed to evaluate the atmospheric dispersion of the pollutant releases that are taken into account. Such a methodology does not allow accounting for uncertainties that are related to the use of such numerical models. These are uncertainties associated with the model itself (physical parameterizations and assumptions that drive the evolution of modelled plumes) and with input data that are provided to run the model (meteorology, geographical site characteristics, specifications of the releases, etc.).

The problem of quantification and propagation of uncertainties in atmospheric dispersion modelling has been widely studied during the last 10 years. These studies mainly focused on the identification, quantification and propagation of uncertainties related to model input data and involved models ranging from Gaussian plume models (e.g. Hanna *et al.*, 1998; Hill *et al.*, 2002) to more complex models like computational fluid dynamics and lagrangian models (Demaël 2007 ; Mallet and Sportisse 2008). Generally, these authors applied parametric approaches or ensemble methods to characterize the impact of various sources of input uncertainties on the prediction of surface pollutant concentrations. Excepted for ensemble prediction studies that generally rely on the simulation of real cases, with three dimensional lagrangian models, most of these studies addressed the problem of uncertainties related to very simple case studies that did not reflect possible applications to LIA studies.

Since three years, NUMTECH, CEA and PHIMECA have worked on the development of a functional tool that allows assessing the impact of input uncertainties on final results of typical LIA studies. First of all, a method to identify uncertainty sources along an atmospheric dispersion modelling chain was proposed (Brocheton *et al.*, 2008). The various uncertainty sources were reviewed: collection of input data, modelling of dispersion processes and treatment of simulation outputs. Then, we developed a first prototype chain for the quantification of uncertainties associated with impact studies at local scale (Argence *et al.*, 2010). The work consisted in (i) defining all sources of uncertainties (meteorology, emission, etc.), and (ii) implementing Monte-Carlo simulations for the propagation of input uncertainties with a Gaussian plume model. However, due to the computational cost of the simulations, the chain presented limitations and some aspects of a "classical" uncertainties study, like sensitivity analysis, were not implemented. More recently, we have supplemented the uncertainties evaluation chain and succeeded in validating it in the case of a LIA study. Two major tasks were achieved: (i) study of the various ways to characterize input uncertainties with a parametric representation (*i.e.* assigning specific probability density functionto each uncertain variable) or a non-parametric one (kernel smoothing distributions); (ii) study of the feasibility of various uncertainties propagation methods (Monte-Carlo simulations, polynomial chaos or quadratic accumulation).The entire chain was assessed using a practical case study (with a unique source, a flat terrain and no plume rise) and a Gaussian plume dispersion model.

The paper is structured as follows. In section 2, the approach that has been developed is presented. In particular, this section addresses the questions of input uncertainties characterization, uncertainties propagation and statistical analysis of the results. Section 3 presents some fundamentals results obtained when applying our tool to the case under study when focusing on the prediction of ground level concentrations. The question of the graphical representation of the results regarding an uncertainties study is discussed in section 4. Some conclusions and perspectives will be drawn in the section 5.

2. DESCRIPTION OF THE APPROACH

As previously shown by Argence *et al.* (2010), a chain for the quantification of uncertainties associated with simulation outputs has been developed following the recommendations of de Rocquigny *et al.* (2008). This approach is displayed in Figure 1. Each step of the approach is described in the following subsections.

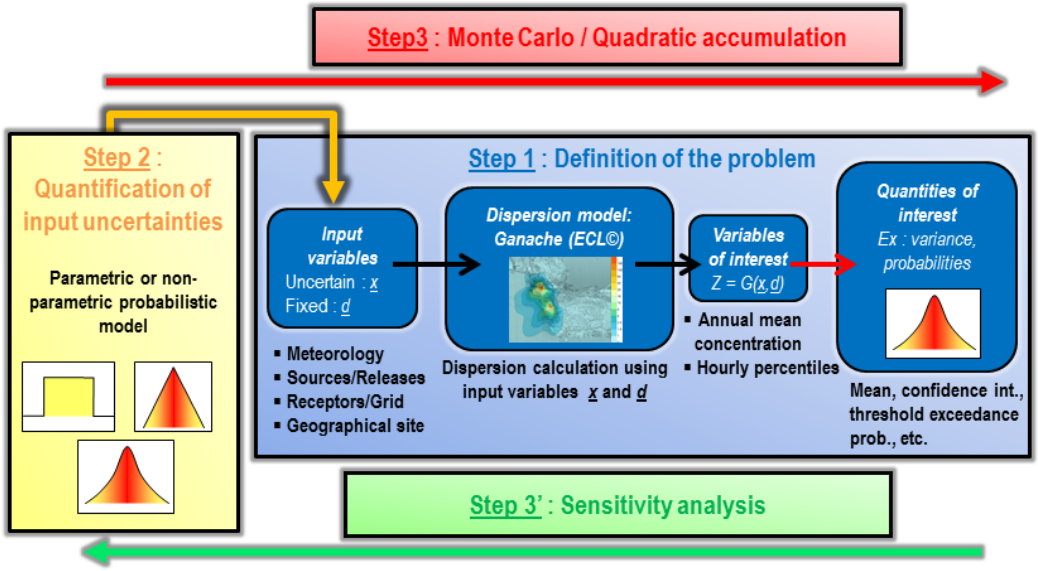

Figure 1. Schematic diagram representing the approach used for the quantification of uncertainties associated with groundlevel concentrations (adapted from de Rocquigny *et al.* (2008)).

2.1 Step 1: Definition of the problem

This step is probably the most important of an uncertainty analysis and each substep (input variables, dispersion model, etc.) must be carefully studied. Argence *et al.* (2010) described in details the framework of the study, that is: the goal of the uncertainty study, the appropriate case-study, the dispersion model to be used, the input variables, the variables and statistical quantities of interest. Here we just recall some fundamental information about the definition of the problem.

The Gaussian plume model GANACHE, which is currently developed at Central School of Lyon (France), has been used for this study and is fully integrated to the tool that has been developed. This model may use various parameterizations of atmospheric stability (similitude theory, Pasquill-Gifford and Doury classes) but in this paper we will only focus on numerical experiments using the similitude and Doury approaches. Uncertainties related to variables related to meteorology, dispersion site (rugosity), receptors and grid height as well as release characteristics have been taken into account. Our approach coming within the context of LIA studies, we used a five years period of observed meteorological data to conduct our experiments (wind speed and direction, cloud cover and temperature). Doury classes were computed using wind speed and cloud cover data. The study mainly focused on the probabilistic prediction of ground level concentrations. Here we computed annual mean concentrations and 95th percentiles.

2.2 Step 2: Quantification of input uncertainties

As described in Argence *et al.* (2010), two approacheswere carried out to derive probability distribution functions (PDF) associated with uncertain input variables describing the meteorology: a parametric one that consists in assigning PDF to each uncertain input variable and a non-parametric representation of input uncertainties that relies on the use of kernel smoothing distributions that the best fit the natural distribution of input variables. As a mean to define the best way to process input uncertainties, a comparison between these two approaches has been conducted by comparing the natural relationships of input variables with the artificial relationships between these variables when they are modelled with PDF. An example of the results we obtained when plotting observed wind speeds against boundary layer heights is given in Figure 2. Obviously, the non-parametric approach is the unique method that gives a reliable representation of the relationship between wind speed and boundary layer height measurements. Starting from this consideration, it was concluded that the non-parametric approach was the best adapted to the quantification of input uncertainties. This method was thus implemented in the tool we developed and results presented in the next sections are associated with it. Note that the PDF of input variables related to the characteristics of the release, the dispersion site, and receptors and grid heights were mainly defined from expert judgment.

2.3 Steps 3 and 3': Propagation methods and sensitivity analysis

This step of the uncertainty analysis consisted in propagating the joint PDF of each input variable through the dispersion model to generate the PDF of the model results (ground level concentrations). Several methods were considered in this study from the cheapest (in terms of computational cost) to the most expensive ones: quadratic accumulation, polynomial chaos expansion and Monte-Carlo simulations. Actually, quadratic accumulation has been applied only to a few tests of sensitivity analysis when considering only some specific hours from the whole period under study (five years). As it was decided to use a non-parametric representation of input uncertainties, the polynomial chaos approach was not adapted since it requires the use of a parametric probabilistic model of input data. Monte Carlo simulations were then used to propagate input uncertainties through the Gaussian dispersion model GANACHE. Using Monte Carlo simulations with sufficiently large random sample, it is possible to accurately determine the mean of the variable of interest (here it corresponds to annual mean concentration) but also to conduct a statistical analysis of its probabilistic distribution. It is then possible to apply the following methods:

- Resampling (like bootstrap) to compute confidence intervals associated with statistical parameters (mean, percentiles, for example).
- Extreme values analysis (analysis of the tail of the distribution) to evaluate high order percentiles.
- Statistical inference to find an empirical PDF that the best fits the predicted concentration distribution. This method then allows to perform some statistical analysis of the PDF to derive quantities of interest (mean, confidence intervals, etc.).

Some examples of the application of steps 1 to 3 are given in section 3.

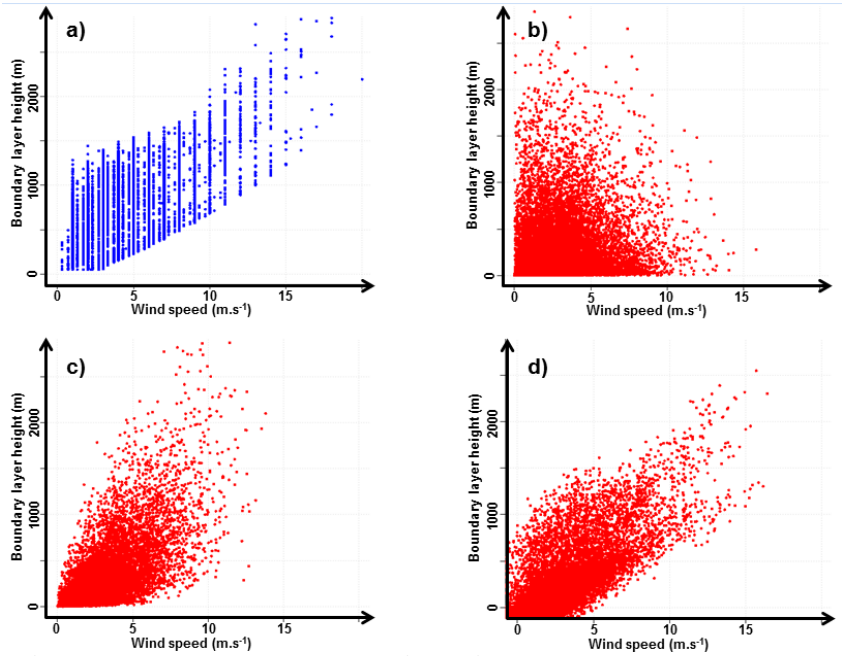

Figure 2.Scatter plot of boundary layer height against wind speed for : (a) five years observations, (b) random sample of 10000 elements from a parametric model with no dependencies between input variables, (c) random sample of 10000 elements from a parametric model with dependencies modelled by Gaussian copulas and (d) random sample of 10000 elements from a non-parametric model .

3. APPLICATION OF THE APPROACH TO A PRACTICAL CASE STUDY

3.1Case study

The case under study considered a single release of inert gas of 50 m height with a diameter of 1 m. The source is located at the centre of a domain composed of 50 x 50 points using a horizontal resolution of 100 m. It was hypothesised that the release occurred in flat terrain with uniform roughness. These simplifications implied neglecting uncertainties associated with topography, obstacles, source geometry, physical and chemical transformations of pollutants. Numerical calculations were performed considering fixed receptors only. The receptors have been placed according to two particular wind directions identified from the dataset which has been used: one rather infrequent (direction D1) and the other frequently observed (direction D2).

3.2Annual mean concentration computation

Figure 3 and Table 1 present some results when focusing on the estimation of annual mean concentration computed using both a deterministic approach and Monte Carlo simulations. Generally, atmospheric dispersion simulations obtained with both the deterministic approach and the probabilistic one give pretty similar results. For example, when using the Doury parameterization and looking at ground level concentration prediction in direction D2 (Figure 3b), it is shown that deterministic and Monte Carlo simulations exhibit very similar values for receptors located between 1000 and 5000 m from the release. Whatever the approach used to characterize atmospheric stability (Doury or similitude theory), the main differences appear in the vicinity of the source, for receptors located at 500 m of the release. These differences are more important for wind direction D1, which is rarely observed. This may be explained by the fact that the non-parametric meteorological model may not fully represent infrequent wind directions(that is with the right frequency), that are mainly associated with weak wind speed. It is also interesting to look at the size of 95% confidence intervals associated with the estimation of annual mean concentrations (Table 1) that can be interpreted as a measure of the uncertainty related to the probabilistic computation of the average values. Note that increasing the size of the sample used for Monte Carlo simulations will necessarily led to a decrease of the size of 95% confidence intervals.

Table 1. Annual mean concentration computed for receptors located at a distance of 500 m from the pollutant release. Results are given for the five years deterministic run and Monte Carlo simulations (using the non-parametric meteorological model). 95% confidence intervals are also reported in square brackets.

Approach	Direction D1	Direction D2
Doury – Deterministic (Five years)	3.57	4.83
Doury – Monte Carlo simulations (sampling size 10000)	2.50 [2.31 ; 2.68]	4.57 [4.32 ; 4.83]
Similitude – Deterministic (Five years)	1.32	2.35
Similitude – Monte Carlo simulations (sampling size 10000)	1.08 [0.99; 1.17]	2,03 [1.92 ; 2.13]

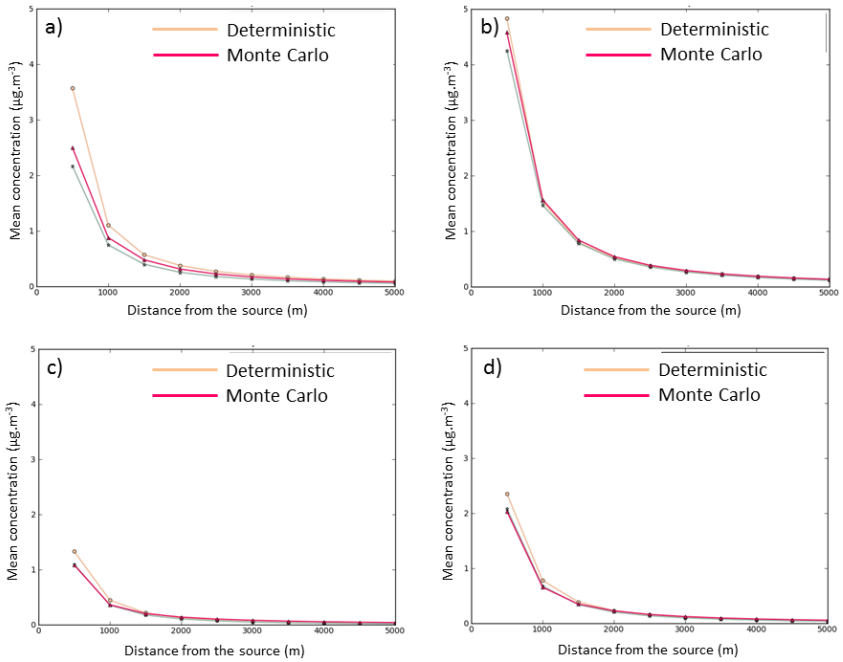

Figure 3.Mean concentration as a function of the distance from the source for (a) Doury/direction D1, (b) Doury/direction D2, (c) Similitude/direction D1 and (d) Similitude/direction D2. The blue curve is associated with an experiment that is not described in this paper.

3.2Estimation of 95th percentiles of the predicted concentration

Table 2 presents some results when focusing on the estimation of 95th percentiles computed using both a deterministic approach and Monte Carlo simulations. Note that the computation of percentiles is strongly sensitive to the physical and probabilistic models used for the study, so the following results must be considered with care. Table 2 shows that Monte Carlo simulations using similitude theory are able to reproduce 95% percentiles that are close to the deterministic ones (five years integration), or at least of the same order of magnitude. Generally, the use of the Doury approach led Monte Carlo simulations to strongly overestimate 95% percentiles values, particularly for receptors located in the wind direction D1. Studying the distribution of predicted ground level concentrations at 500 m from the release shown that histograms generally contained a large proportion of zero values (not shown). Nevertheless, small differences in the proportion of simulated values falling in classes associated with large concentrations values are observed between experiments. This main explain the variability of 95% percentiles that is observed when comparing deterministic and probabilistic approaches. For example, for the five years deterministic run using Doury classification, the 95% percentile at 500 m from the release in direction D1 is about 0.25 µg.m^{-3}while the 97% percentile at the same location and for the same experiment is greater than 6µg.m^{-3}.A detailed analysis of the results revealed the strong sensitivity of high order percentiles estimation to both the large proportion of simulated zero values and to the shape of the tail of simulated concentrations distribution.

Table 2. 95th percentile of the predicted concentration computed for receptors located at a distance of 500 m from the pollutant release. Results are given for the five years deterministic run and Monte Carlo simulations (using the non-parametric meteorological model).

Approach	Direction D1	Direction D2
Doury – Deterministic (Five years)	0.25	16.89
Doury – Monte Carlo simulations (sampling size 10000)	13.84	38.98
Similitude – Deterministic (Five years)	8.90	13.92
Similitude – Monte Carlo simulations (sampling size 10000)	2.75	14.84

4. PRESENTATION OF RESULTS ASSOCIATED WITH AN UNCERTAINTY ANALYSIS

The presentation of the results of an uncertainty analysis is fundamental and the answer must be adapted to the audience. It is then necessary to present results that may be easily understood forgetting mathematical principles that were used to produce

them. We have then proposed few parameters that may be used to present results of a probabilistic LIA study. For a given variable of interest (annual mean for example), one may presents its probabilistic estimation, boundaries of a given confidence interval (giving the possible minimum and maximum values accounting for input uncertainties) and a probability of threshold exceedance. These parameters may be computed and plotted for calculations considering both specific receptors only and full grid as well. Figure 4 presents some examples for an idealized experiment conducted over a full simulation grid.

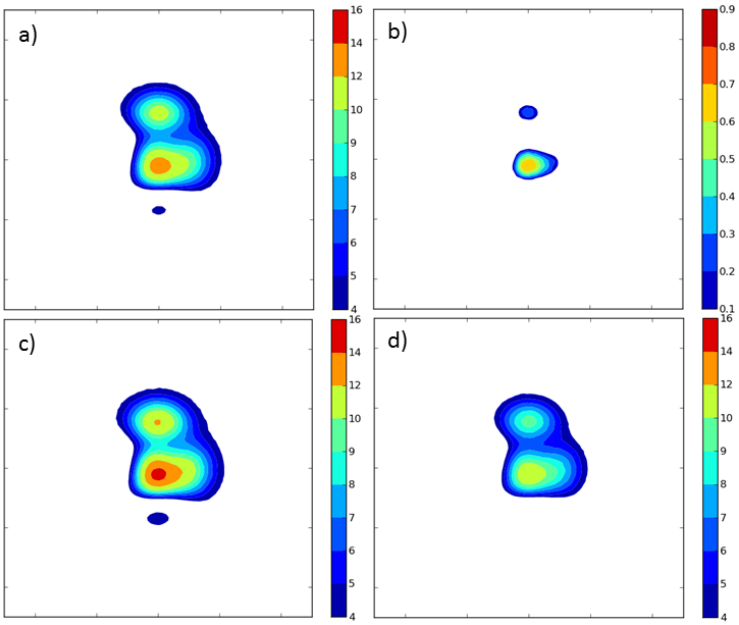

Figure 4.Examples of graphical representation of the results of an uncertainty study: (a) average of the variable of interest, (b) probability of exceeding 12 µg.m^{-3}, (c) upper and (d) lower boundaries of the 95% confidence interval of the calculated average.

5. CONCLUSIONS AND PERSPECTIVES

One of the main conclusions of this work is that we have developed a complete and operational chain for the treatment of uncertainties associated with atmospheric dispersion modelling. This chain contains all aspects of an uncertainty analysis from the quantification of uncertainties associated with input variables to the statistical postprocessing of predicted ground level concentrations (mean, percentiles, confidence intervals, threshold exceedance probabilities, etc.). The whole chain is based on Python scripts and is computationally efficient (parallelization of computation). The chain has been validated using a practical case study and now constitutes an operational and fully automatized tool that may be easily coupled to other Gaussian dispersion models (actually, the only modifications of the chain consist in modifying subroutines that deal with input and output files format of the model used). Among the various perspectives that may follow this study, we may think to apply this chain to computations on a whole simulation grid (here, only receptors were considered), extend our approach to other variables of interest (daily percentiles for example), use this tool for more complex case studies (with topography and multiple sources, for example) and coupling this chain with more complex dispersion models like CFD models (doing this, we will assess the feasibility of using this chain to conduct impact studies for both accidental and chronic releases).

REFERENCES

Argence, S.,P. Armand, T. Yalamas, F. Deheeger, F. Brocheton and E. Buisson, 2010: Investigation of different methodologies to characterize and propagate uncertainties in atmospheric dispersion modelling: application to long-term impact assessment. *Proceedings of the 13th international conference on harmonisation within atmospheric dispersion modelling for regulatory purposes, Paris, France*, 80-84.

Brocheton, F., P. Armand, L. Soulhac and E. Buisson, 2008: A methodology to characterize the sources of uncertainties in atmospheric transport modelling. *Proceedings of the 12th international conference on harmonisation within atmospheric dispersion modelling for regulatory purposes, Cavtat, Croatia*, 78-82.

De Rocquigny, E., and co-authors, 2008: Uncertainty in industrial practice: A guide to quantitative uncertainty management. Editors Etienne de Rocquigny, Nicolas Devictor and Stefano Tarantola, John Wiley & Sons publishers, 339 pp.

Demaël, E., 2007: Modélisation de la dispersion atmosphérique en milieu complexe et incertitudes associées. PhD Thesis, 273 pp.

Hanna, S.R., J.C. Chang 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. *Atm. Env.*, **32**, 3619-3628.

Hill, R., I. Lowles, V. Auld and T. Parker, 2002: Probabilistic evaluation of models for the atmospheric dispersion of effluents release for a complex site. *Proceedings of the 8th international conference on harmonisation within atmospheric dispersion modelling for regulatory purposes, Bulgarian Academy of Sciences*, 48-52.

Mallet, V. and B. Sportisse, 2008: Air quality modelling: From deterministic to stochastic approaches. *Comp. and Math.with App.*, **55**, 2329-2337.