19th International Conference on
Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes
3-6 June 2019, Bruges, Belgium

A GENERAL PERTURBATION MODEL OF THE METEOROLOGICAL AND SOURCE TERM PARAMETERS TO ASSESS THE UNCERTAINTIES IN 3D DISPERSION SIMULATIONS

Sylvain Girard[1], Patrick Armand[2], Christophe Duchenne[2], and T. Yalamas[2]

[1]*PHIMECA Engineering, F-75012 Paris, France*

[2]CEA, DAM, DIF, F-91297 Arpajon, France

Abstract: This paper presents a generic mathematical framework for modelling uncertainties in 3D atmospheric dispersion simulations. It is exemplified by an accidental release from a very complex rough and built-up area. The next step will be to analyse the resulting set of concentration maps by statistical learning of an underlying manifold using an auto-associative model.

Key-words: *dispersion simulation, local scale, perturbation, wind field, source term uncertainty propagation, probabilistic danger zones.*

INTRODUCTION

Modern 3D modelling systems are capable of precisely simulating atmospheric dispersion of hazardous material, notably by accounting for the effect of the topography and buildings on the flow and dispersion. They are used to evaluate the impact of hazardous materials releases on human health for both regulatory purpose and emergency preparedness or response. However, the input data describing meteorological conditions and the release itself are often highly uncertain.

This paper presents a generic mathematical framework for modelling those uncertainties by perturbation of a conjecture. It allows a thorough exploration of the possible outcomes by Monte Carlo methods, while preserving physical consistency and limiting the number of tuning parameters. We implemented conjecture perturbation in a cascade modelling chain involving WRF meteorological forecast model and PMSS flow and dispersion modelling system. It is exemplified by an accidental release from a very complex rough and built-up area.

The output of the physical model is a spatial map of a quantity, here the maximum concentration reached during the considered time interval. We present an innovative approach for analysing such set of maps by statistical learning of an underlying manifold using an auto-associative model. It allows assessing the reliability of the probabilistic diagnosis, as well as extracting physically meaningful features. The Monte Carlo simulations for the above mentioned case study are under way. The complete analysis will be presented at the conference.

MOTIVATING CASE STUDY: EXTENDED RELEASE FROM A CHEMICAL FACILITY

We consider the following idealised decision problem. Hazardous material is released in the atmosphere during a given period of time. A physical model simulates the transport and dispersion in the atmosphere, the deposition by rain, and computes concentration near the ground. We want to decide in advance if a given threshold, possibly leading to health consequences, is exceeded. In our case, the threshold is based on the maximum reached concentration. The exceedance of the threshold in a geographical area might lead to decision-taking like shelter or evacuate the population of the area.

This decision making scenario is inspired by a real industrial incident that happened on January 2013 at the Lubrizol chemical plant located in Rouen, France. Operation mistakes and minor system failures in the plant resulted in extended release of hydrogen sulphur and mercaptan, which are both foul-smelling. The major part of the material inventory was emitted between Monday 21th evening to Tuesday 22th morning (IMPEL, 2013). The wind blew the plume as far as Paris during Monday night and towards London on Tuesday. Thousands of people have complained of nausea and headaches.

Physical model

The dispersion simulations will be carried out with Parallel-Micro-SWIFT-SPRAY (PMSS). Originally, Micro-SWIFT-SPRAY (MSS) (Tinarelli et al., 2012) was developed in order to provide a simplified, but rigorous CFD solution of the flow and dispersion in built-up environments in a limited amount of time. MSS encompasses the local scale high resolution versions of the SWIFT and SPRAY models. SWIFT is a 3D terrain-following mass-consistent diagnostic model taking account of the buildings and providing the 3D fields of wind, turbulence, and temperature. SWIFT interpolates between meteorological measurements (ground stations and vertical profiles), numerical data issued by meso-scale simulations (as in the present work) and, possibly, analytical relations of the flow influenced by the buildings (displacement zone, wake zone, skimming zone, etc.). SPRAY is a 3D Lagrangian Particle Dispersion Model able to account for the presence of buildings. Both SWIFT and SPRAY can deal with complex terrains and evolving meteorological conditions and with specific features of the release (heavy gas, light gas, etc.). More recently, SWIFT and SPRAY have been efficiently parallelized in time, space, and numerical particles leading to the PMSS system (Oldrini, Armand et al., 2017).

Deterministic decision boundary

The source is located in the middle of the simulation domain, a square with side length of about 35 km. In a deterministic framework, a single simulation is run using the most credible values for the model inputs. Later on, we call this set of values the (input) conjecture, and likewise we refer to the concentration simulated using them as conjectured concentrations. We wish to seek for each location where the maximum over time of the conjectured concentration exceeds a given threshold, possibly associated with any health damage or the olfaction limit of the released hazmat. The maximum over time of the conjectured concentrations are shown in **Figure 1a**. From this field, we determine the deterministic area where the olfaction threshold is exceeded. This is the blue coloured area in **Figure 1b**.

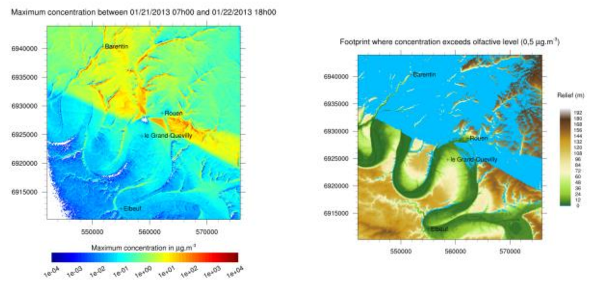

Figure 1. (a – left) *Maximum over time of the conjectured concentrations.*
(b – right) *Deterministic area where the olfaction threshold of the released hazmat is exceeded.*

Uncertain inputs

The conjecture comprises any model input that is issued from partial or imprecise observations, or from other physical model simulations, and possibly also model parameters that implicitly account for the unavoidable discrepancy between the real complex system under study, and its idealised mathematical representation. We focus here on three uncertain input data known to have a substantial impact on simulation output (Aguirre Martinez et al., 2016; Girard, Mallet et al., 2016), all of which are high dimensional and have complex physically meaningful structures:

- the rate of emission of material at the source is a time series, called later on the source term,
- the rain intensity is a scalar spatio-temporal field,
- and wind velocity and direction is a vector spatio-temporal field.

The conjectured source term was adapted from data established by Ismert and Durif (2014). All meteorological inputs (rain and wind fields in particular) were obtained from the community reconstruction and forecast meso-scale modelling system WRF whose horizontal resolution was 1 km. For the wind, we used a set of 514 vertical profiles of the horizontal components, and we kept the 21 vertical layers below 3 km above ground level (AGL), plus surface data at 10 m AGL. WRF simulations are sampled every 15 minutes, and we used the 141 time steps from 21/01/2013 07:00 to 22/01/2013 18:00.

MATHEMATICAL FORMALISATION OF THE DECISION PROBLEM

We tackle the decision problem stated in the previous section from the viewpoint of uncertainty propagation. We need the probability distribution of maximum concentrations, but cannot model it directly. Instead, we modelled the uncertainty of the most influential inputs of the physical model. The chain of functions involved is represented on **Figure 2**.

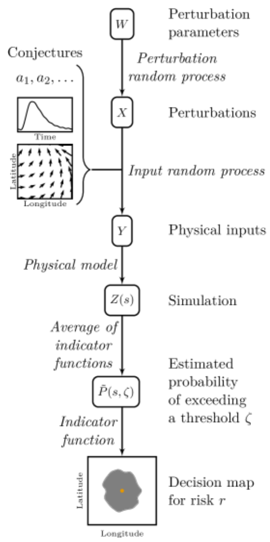

Figure 2. *Function chain of the simulation process. Capital framed letters designate random vectors. Arrows annotated in italic designate functions.*

The atmospheric dispersion model is a deterministic function denoted by f. The output of f is the maximum concentration (or the concentration greater than a threshold exceedance) z as a function of location s and time t. The random vector Y is the perturbed conjecture and models the input uncertainty. Here Y represents jointly the source term, and the rain and wind fields. We want to uncover the distribution of $Z(s)$, the random vector corresponding to $z(s)$. It represents the spatial repartition of maximum concentrations (or concentrations greater than a threshold value), which can be represented by a map such as that displayed in **Figure 1a (or 1b)**.

“Decision maps” are obtained from maximum concentration maps by setting each cell to 1 if the chosen threshold is exceeded and 0 otherwise. Averaging a set of decision maps yields an estimator of the probability of exceedance. Taking a decision implies to admit a fixed risk of committing a specified error.

For instance, evacuation is decided at locations where the average of decision map exceeds 5%.

Generic amplitude and dynamics perturbation scheme

Our probabilistic model of input uncertainty is based on perturbations of the conjecture. In **Figure 2**, conjectures of diverse dimensions are grouped by a brace, and the random perturbations are denoted X. We generalised and unified the perturbation schemes previously designed for modelling wind field uncertainty (Duchenne et al., 2017).

Let $Y(s,t)$ be the spatio-temporal random vector obtained by perturbation of a conjecture $c(s,t)$. We adopted the following generic perturbation:

$$Y(s,t) = \Gamma(s,\Theta(t))c(s,\Theta(t)) + \Lambda(s,\Theta(t)).$$

The random vector Γ (resp. Λ) is a multiplicative (resp. additive) perturbation of the amplitude of the conjecture, that we will call the “gain” (resp. “offset”) perturbation. The random function Θ is the perturbation of the dynamics of the conjecture.

The spatio-temporal structure of each element of the perturbation must be postulated. We chose to impose smooth oscillating temporal variations and have the perturbation depend on the location only through the conjecture. We formulated the temporal structure of the gain and offset perturbation using the phase rotation technique, namely the summation of cosine with random phases. The gain (resp. offset) perturbation writes:

$$\Gamma(s,t) = \sigma_c(s,t)\sum_{k=1}^{K} W_{\Gamma,k}\, cos(2\pi\omega_k t + \Psi_k).$$

When the phases Ψ_k are independent and uniformly distributed on $[0,2\pi]$, this formulation produces smoothly oscillating signals. The set of periods $\Omega = \{\omega_k: 1 \leq k \leq K\}$ controls the temporal structure of those signals and must be postulated, as well as the distribution of the set of amplitudes $W_\Gamma = \{W_{\Gamma,k}: 1 \leq k \leq K\}$. The function σ_c is arbitrary and is used mostly for either restricting the perturbation to a given time span, or make it depend on the intrinsic variability of the conjecture.

The random function Θ models the uncertainty on the dynamics of the conjecture by altering time. At any given instant t, an interval warp function ϕ expands or contract a time interval Δt by a factor $\beta(t)$ that varies in time:

$$\theta: t, \Delta t \mapsto \phi(t,\Delta t) = \beta(t)\Delta t.$$

We then define the associated time warp function θ mapping any instant t_i from $\{t_0 = 0, t_1, \ldots, t_q = T\}$ to the warped instant:

$$\theta(t_i) = \frac{T}{\sum_{j=0}^{q-1} \phi\,(t_j, t_{j+1} - t_j)} \sum_{j=0}^{j} \phi\,(t_j, t_{j+1} - t_j).$$

We further constrain the time warp functions to preserve the time origin: $\theta(t_0 = 0) = t_0 = 0$. By definition, time warp functions also preserve the total duration $\theta(t_q = T) = t_q = T$.

We used again phase rotation to define the distribution Θ of time warp functions, resulting in smooth time structure for the dynamics perturbation. In practice, the warped time series $y(s,t)$ are obtained by first applying the gain and offset perturbation to the conjecture, and then interpolating the resulting time series at the warped instants $\{\theta(t_0), \theta(t_1), \ldots, \theta(t_q)\}$.

ANALYSIS OF SET SPATIAL MAPS BY MANIFOLD LEARNING

As discussed in the previous section, the predictions of the physical model are spatio-temporal quantities, and the decision is based on a set of spatial maps. Sampling these maps at the p nodes of a grid, we obtain p random variables, depending intricately on each other. The set of n maps can then be seen as a cloud of n points in $\mathbb{R}^p$. Because the decision problem we consider is posed in a crisis context, time, and therefore sample size, is limited. Besides, the most interesting exceedance probabilities are small, 5% or lower, which require a large sample to be estimated accurately. Straightforward uncertainty propagation is thus unlikely to succeed and we have to resort to either model emulation (Aguirre Martinez et al., 2016; Aguirre and Yalamas, 2014; Sylvain Girard, Mallet et al., 2016; Mallet et al., 2018), namely mathematical approximation of the physical model, or more efficient sampling strategies such as weighted sampling or Monte Carlo Markov chain. Either of those approaches are intractable in high dimension, namely when p exceeds a few units.

Limitations of principal component analysis

Principal component analysis (PCA) is a century old method used extensively in data analysis (Jolliffe and Cadima, 2016). It can be considered the state of the art for dimension reduction in the specific field of dispersion simulation (Mallet et al., 2018; Swallow et al., 2017; Tran Le et al., 2018). Informally, PCA finds the cube of smallest dimension d_l that best contain a given point cloud in $\mathbb{R}^p$. The number $d_l \leq p$ is called the *linear dimension* of the point cloud. Each point x in $\mathbb{R}^p$ is then associated to the point of $\mathbb{R}_l^d$ whose coordinates are the orthogonal projections of x onto the sides of the cube. This defines d_l new variables, the principal components, by linear combination of the p original variables. This method is efficient when there are indeed linear relations between the p original variables that allow summarizing the point cloud. Unfortunately, it is often not the case when the p original variables are sampled on the time series or on a spatial map.

Consider for instance a set of time series consisting of a Gaussian pulse, a bell shaped spike, shifted along an abscissa. Sampling it at p evenly spaced abscissas yields a p-dimensional point cloud as before. This very simple example devised by Fukunaga and Olsen (1971) is actually representative of many situation common in atmospheric dispersion: time evolution of a concentration when a plume passes over a recording station, or comparison of circular cross sections of plumes at various angles. A single scalar, say the abscissa of the top of the bell, is sufficient to fully parametrize the set of curves. We will say that the *intrinsic dimension* of the data set d_i is equal to 1. However, the relation between the p original variables being non-linear, the linear dimension is much larger than the intrinsic dimension, $d_i \ll d_l$, and PCA fails to achieve satisfying dimension reduction.

In a previous communication (Duchenne et al., 2017), we applied PCA to sets of dispersion simulations resulting from low dimensional perturbations. While the first 2 or 3 principal components carried much more information than the subsequent ones, at least a dozen of them were required to properly encode the original dataset. This lack of efficiency even with simple perturbations means that PCA will not be adapted to analyse the output of complex perturbation scheme.

Manifold learning by auto-associative model

The auto-associative model (AAM) proposed by Stéphane Girard and Iovleff (2008) can be seen as a non-linear extension of PCA. The point cloud is assumed to hover close to a manifold instead of a cube. A manifold of dimension d is a set that is topologically locally equivalent to a d-dimensional cube. It handles non linearity and can generally achieve reduction to dimension equal or close to the intrinsic dimension while preserving the fidelity of the reconstruction.

In a recent experiment, AAM allowed to automatically extract physically meaningful structure from a set of simulated responses of a perturbed dynamic system (Gerrer and Girard, 2019). AAM applied to the two previously mentioned sets of decisions maps (Duchenne et al., 2017) yielded representations of dimension 2 and 3, which are most likely the actual intrinsic dimension of the sets. One of the main objectives of the future work will be to investigate the ability of AAM to analyse point cloud of higher dimension, like the set of decision maps resulting from the perturbation scheme described in the paper.

CONCLUSION

In this paper, we introduce a generic mathematical framework aimed at modelling uncertainties in 3D atmospheric dispersion simulations. To exemplify the practical use of the method, we consider a complex accidental situation on a rough and built-up terrain characterized by uncertain release and meteorological conditions (wind direction, wind force and precipitations) which are outputs of the WRF model. The next steps will be first to produce a set of concentration maps with the PMSS modelling system by varying the uncertain parameters, and then, to analyse the resulting set of maps by statistical learning of an underlying manifold using an auto-associative model.

REFERENCES

Aguirre Martinez, F., Y. Caniou, C. Duchenne, P. Armand, and T. Yalamas, 2016: Probabilistic assessment of danger zones associated with a hypothetical accident in a major French port using a surrogate model of CFD simulations. In Proc. of the 17th Int. Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, 9-12 May 2016, Budapest, Hungary.

Duchenne, C., P. Armand, M. Marcilhac, S. Girard and T. Yalamas, 2017: A new method for assessing the uncertainty associated with 3D dispersion simulations in any variable meteorological conditions. In Proc. of the 18th Int. Conference within Atmospheric Dispersion Modelling for Regulatory Purposes, 9-12 October 2017, Bologna, Italy.

Fukunaga, K and D. R. Olsen, 1971: An algorithm for finding intrinsic dimensionality of data. In IEEE Transactions on Computers C-20.2, pp. 176–183. doi: 10.1109/t-c.1971.223208.

Gerrer, C.-E. and Sylvain Girard, 2019: Non-linear dimension reduction of dynamic model output. In Proc. of the 13th Int. Modelica Conference.

Girard, Stéphane and S. Iovleff, 2008: Auto-associative models, non-linear Principal Component Analysis, manifolds and projection pursuit. In Lecture Notes in Computational Science and Engineering. Springer Berlin Heidelberg, pp. 202–218. doi: 10.1007/978-3-540-73750-68.

Girard, Sylvain, V. Mallet, I. Korsakissok and A. Mathieu, 2016: Emulation and Sobol' sensitivity analysis of an atmospheric dispersion model applied to the Fukushima nuclear accident. In Journal of Geophysical Research: Atmospheres. doi: 10.1002/2015jd023993.

IMPEL, 2013: Extended release of mercaptan at a chemical plant. Tech. rep. 43616. French Ministry for Sustainable Development. https://www.aria.developpement-durable.gouv.fr/wp-content/files_mf/FD_43616_rouen_GB.pdf.

Ismert, M. and M. Durif, 2014: Accident de Lubrizol du 21 janvier 2013 – Couplage dispersion du nuage odorant / Plaintes et appréciation des risques sanitaires associes.

Jolliffe, I. T. and J. Cadima, 2016: Principal component analysis: a review and recent developments. In Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374.2065, doi: 10.1098/rsta.2015.0202.

Mallet, V. et al., 2018: Meta-modeling of ADMS-Urban by dimension reduction and emulation. In Atmospheric Environment 184, pp. 37–46. doi: 10.1016/j.atmosenv.2018.04.009.

Oldrini, O., P. Armand, C. Duchenne, C. Olry and G. Tinarelli, 2017: Description and preliminary validation of the PMSS fast response parallel atmospheric flow and dispersion solver in complex built-up areas. Journal of Environmental Fluid Mechanics, Vol. 17, No. 3, 1-18.

Swallow, B et al., 2017: Parametric uncertainty in complex environmental models: a cheap emulation approach for models with high-dimensional output. In: arXiv preprint arXiv:1702.03696.

Tinarelli, G., L. Mortarini, et al., 2013: Review and validation of Micro-Spray, a Lagrangian particle model of turbulent dispersion. In Lagrangian Modeling of the Atmosphere, Geophysical Monograph, Volume 200, American Geophysical Union, AGU: 311-327.

Tran Le, N. B. et al., 2018: Metamodeling and optimization of probabilistic scores for long-range atmospheric dispersion applied to the Fukushima nuclear disaster. In EGU General Assembly Conference Abstracts. Vol. 20, p. 17209.