H13-191

SOURCE DETERMINATION IN CONGESTED ENVIRONMENT THROUGH BAYESIAN INFERENCE

Jean-Marie Libre[(1)], Malo Le Guellec[(2)], Amita Tripathi[(2)], Thibault Mailliard[(2)], Stéphanie Guérin[(1)], Claude Souprayen[(2)], Aldo Castellari[(1)].

(1) TOTAL E&P, La Défense, FRANCE
(2) Fluidyn-Transoft, Saint-Denis, FRANCE

Abstract: The quick identification of contaminant plume sources can greatly enhance emergency response efforts. The critical questions are: Where is the leak? What is the release mass flow? How long is the release duration? Accurate estimation of the source term is essential to predict plume dispersion, manage the emergency planning and mitigate consequences in the surrounding area.
Based on limited information provided by a finite and noisy set of transient concentration measurements obtained from real-time gas sensors, the source identification is also complicated by the presence of building obstacles.
While high-resolution CFD simulations are available for predicting plume evolution in complex geometries, a Bayesian probabilistic inferential method can provide the probability density function of the source parameters (location, mass flow, release duration, turn on and turn off time) given a set of concentration measurements on the sensor network. The approach considers both measurements and forward model errors in the probability density function computation. This method is coupled with a fast Markov Chain Monte Carlo (MCMC) stochastic sampling method to determine the source term parameters. The developed algorithm avoids wasting time in generating samples from regions in parameter space which contribute very little to the probability density function. It should be noted that the probabilistic methodology can be used in many fields and is used here for time-varying release rates with complex flow conditions. Once a series of MCMC samples has been obtained, summary statistics related to each variable can be built.
The results of the event reconstruction indicate the probability of a source being at a particular location with a release rate. When the source parameters have been defined a fast complementary dispersion model can be used to estimate the impact of the toxic release and activate an emergency response.
After first validations based on a limited range of parameters but with complex flow patterns, this method is now being validated on a modeling platform through measurements campaigns. The activation of the simulation platform is triggered by the detection of above threshold concentrations at the sensors. The source term calculated is then used in forward dispersion mode to simulate the dispersion in Lagrangian puff mode. The industrial site of Lacq (France) has been chosen as a pilot and the key hazardous substance considered in this project is hydrogen sulphide (H_2S).

INTRODUCTION

On industrial installations devoted to oil and gas extraction, processing, refining or petrochemical production, accidental releases of toxic compounds represent a significant part of the risks related to these activities. When such an accident occurs, the knowledge in real time of the concentration fields resulting from the release of the hazardous substance would be extremely valuable information as support for emergency actions and impact evaluation on the industrial site itself and its vicinity.

For that purpose, a modelling platform is developed to simulate in real time the atmospheric dispersion of hazardous substances at the scale of the industrial site and also on its surroundings i.e. from few meters to 10 km in any direction. The industrial site of Lacq (France) specialized in gas production has been chosen as a pilot for this project because of a favourable infrastructure in terms of network of hydrogen sulphide (H_2S) sensors and meteorological stations availability.
A preliminary survey of existing tools (Borysiewicz M.J., M.A. Borisiewicz, 2006) showed that they are either unable to take into account the detailed structure of the site (Gaussian dispersion model) or do not provide a suitable source term determination algorithm. This suggested the need to develop a new approach.

The general principle of this development is to couple in real time the sensor network of the industrial site with appropriate algorithm for source term determination and 3D dispersion modelling.

Significant technical difficulties have to be solved to develop this tool successfully: the calculation of the dispersion must take into account the real structure of the installation to correctly simulate concentration fields on the industrial site, the location and intensity estimation of the accidental release has to be performed automatically by using the recordings at the toxic sensors and computational time must not exceed a few minutes to enable real time use of data. The characteristics of the sensors bring other issues to the project: lower and upper threshold concentrations, response time...

The project is divided in three phases: a first phase devoted to the development of suitable algorithms. A second phase concerns its implementation on the industrial modelling platform and a third phase is devoted to in situ validation of the CFD model with an atmospheric tracer (SF_6) which took place in the first part of year 2010. This paper addresses the issue of the development of a source determination algorithm in phase 1 of this project.

GENERAL DESCRIPTION

The 3D CFD (Computational Fluid Dynamic) model Fluidyn-PANEPR (Mazzoldi A., *et al.*, 2008) has been chosen, to simulate the 3D wind field pattern on the industrial site, taking into account the details of the installations (Hill R., *et al.*, 2007). This model solves the Navier-Stokes equations including mass, momentum and enthalpy conservation, state law and equations for advection-diffusion. A K-ε model is used for turbulence simulations and a micro meteorological model simulates the atmospheric temperature profile based on Monin Obukov similarity theory. This model is used in Eulerian mode to compute wind field pattern whereas the Lagrangian mode is used for the species dispersion. This approach enables a simulation as close as possible to the turbulence and flow around the buildings which could not be done with a standard Gaussian approach which tends to overestimate the concentrations and impact distances and is unable to simulate correctly concentrations at close range.

In order to identify the location and compute the characteristics of the source event, a probabilistic approach has been selected. Bayesian inference approach is particularly suited for application where scarce and noisy data are likely to be available as in an industrial context. The probability that a set of parameters values for a source term be responsible for a concentration field observed in real time is evaluated by comparing the concentration actually observed at each sensor with pre-calculated concentration fields for each possible source.

When a sensor series detects high concentration level which exceed a preset limit (threshold detection), the source reconstruction starts by an optimized research of the leak parameters (location, mass flow rate and release duration) with a Monte Carlo Markov chain (Chow F.K., B. Kosovic and S.T. Chan, 2006). Based on MCMC sampling, a probability density function is produced for each parameter and a statistic analysis of this function provides the likeliest characteristics of the real source term.

The advantage of this probabilistic approach is its robustness since it is based on a comparison between the concentration fields resulting from the most plausible accidental emission release locations from the industrial site and real time observations from the sensor network. Also, CFD calculation enables a reliable simulation of wind flow and dispersion around complex geometries, taking into account turbulence effects thus giving the proper relationship between concentration fields at short range and source term value (Neuman S., 2006).

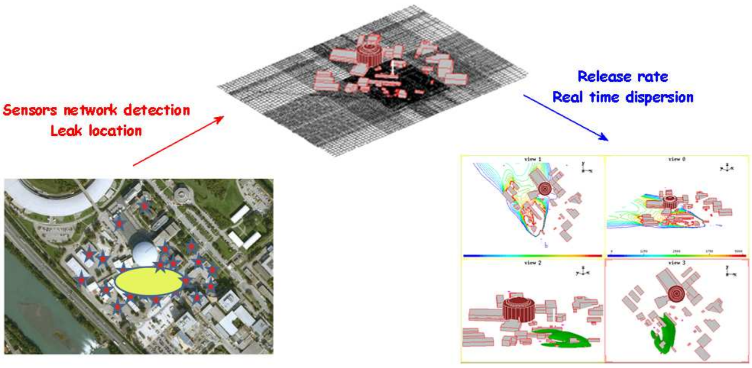


Figure 1: General principle

DESCRIPTION OF THE SOURCE RECONSTRUCTION ALGORITHM

Bayesian inference

A probabilistic theory like Bayesian inference in the source event reconstruction enables to compute the likeliest parameters on the sensors data basis (Keats A., F.-S. Lien and E. Yee, 2006).The posterior probability is a conditional probability which is the link between the hypothesis and the concentration at sensors and the prior information.
The general formulation for conditional probability is

$$\mathrm{P}(X|Y,Z) \tag{1}$$

X : proposal
Y : conditional information
Z : context

By considering the m parameter vector which includes the leak characteristics:

$$m = (loc, q_s, t_{on}, d)$$

Where loc is the leak location, q is the mass flow rate, t_{on} is the turn on time and d is the release duration. The source reconstruction algorithm computes probability density function which after an statistical analysis provides the most probable values.
If the Bayesian theory is applied to the source event determination, the posterior probability for the m vector with the concentration C at sensors and the prior data E

$$P(m|C,E) = \frac{P(m|E)P(C|m,E)}{P(C|E)} \tag{2}$$

$P(C|E)$ is a normalization constant. The posterior probability function is proportional to the product of the prior probability and $P(C|m,E)$:

$$P(m|C,E) \propto P(m|E)P(C|m,E) \tag{3}$$

$P(C|m,E)$ is the probability to get the C concentration measured at sensors for a selected m vector. This probability estimates the deviation between the C concentration recorded at sensors and the C_{mod} modelled concentration provided by the atmospheric dispersion model for the m set of parameters. This deviation includes the measurements and numeric models errors. By hypothesis, the deviation between the measured concentration at i detector and the real concentration follows a normal law distribution. The same statement is made for the theoretical concentration at I detector and the real concentration. The two noise components are supposed to get null average and σ_C andσ_{Cmod} variances.
The probability that the measured concentration be foreseen by the dispersion model for the m vector can be computed by the following relation:

$$P(C|m,E) \propto \left[-\frac{1}{2}\sum_i \frac{\left(C_i - C_{moc,i}(m)\right)^2}{\sigma_C{}^2 - \sigma_{Cmod}{}^2}\right] \tag{4}$$

$P(m|E)$ is the prior probability for the m set of parameters. In this project, this probability is set as a constant that means there is no more probable leak in one source than in another.

$$P(m|E) = constant \tag{5}$$

Nevertheless, it is obvious that the m parameters are defined in a range between 0 and the maximal realistic value (for instance the maximal flow rate based on the process characteristics in the hazard study). Then, it is supposed that the leak can't occur in a building and the probability is set at 0 inside buildings.

$P(m|C,E)$ is the posterior probability for the m vector based on the measured concentrations at sensors and the parameter space E. The relation for the probability density function (PDF) is:

$$P(m|C,E) \propto \left[-\frac{1}{2}\sum_i \frac{\left(C_i - C_{moc,i}(m)\right)^2}{\sigma_C{}^2 - \sigma_{Cmod}{}^2}\right] \tag{6}$$

Random sampling by Markov Chain

The combination of the Bayesian inference technique and a sampling method as the MCMC (Monte Carlo Markov Chain) enables a reliable leak parameters determination.

Indeed, the PDF is a huge space which must be sampled. A classical Monte Carlo sampling is not an appropriated methodology for a multidimensional function. If this sampling is done with a Markov chain which includes the PDF value, the method is highly efficient. There is no waste in time in exploring non useful parts of the parameters space which have low contribution in the PDF.

For a set of m parameters (loc, q, t_{on}, d) the posterior probability is computed by the algorithm. It is accepted if it improves the PDF value of the previous set. Each new part of the Markov chain m_k depends on the previous part m_{k-1}. The MCMC generates a point series as a chain and the distribution of these points follow the $P(m|C,E)$ distribution. Based on the MCMC results, a statistical analysis (histogram, mean value, standard deviation…) can be done for each parameter.

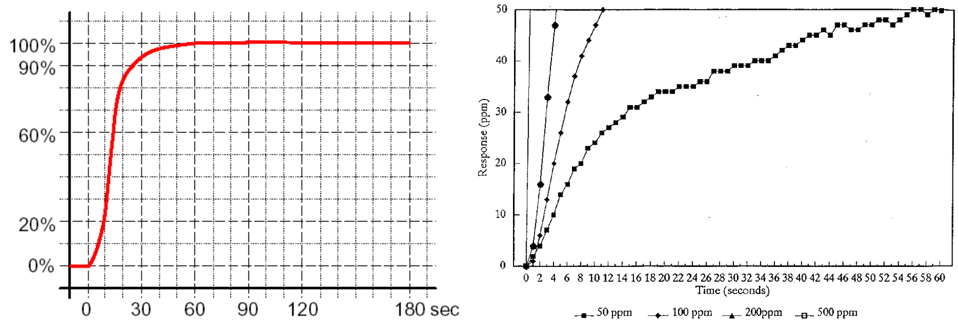
Sensors characteristics

The characteristics of the flammable gas detectors (infrared or catalytic) are normalized and require time response T50<10s and T90<30s (T50 and T90 are the times to get 50% and 90% of the real concentration). In the case of toxic sensors, there is no standard. There are two kinds of H_2S sensors: electrochemical or semiconductor.

In terms of security, a quick detection enables a quick efficient action. In case of H_2S release it is really important because low concentrations can trigger toxic effects.

Whatever measurement equipment, sensors have a response time before recording the real concentration. For a given sensor, the response curve is different for each concentration but the duration between the beginning of the recording and the stabilized measurement is constant. Thus, the T20, T50 and T90 time for the detector to record 20%, 50% and 90% of the real concentration are constants.

The response curve can vary according to the sensors (analytic method, ambient conditions, age…).



(b)

Figure 2: Response curve of H_2S sensors H_2S ((a): electrochemical sensor, (b): semiconductor sensor (*Source* : *DRÄGER Safety*)

The above characteristics have to be included in the source reconstruction. By using a mean response time of the sensor network and the standard response curves provided by the constructor, the following normalized curve has been used:

$$F(t) = 1 - \exp\left(-\frac{t}{\tau}\right) \tag{7}$$

This response time curve has been included in the algorithm by a convolution product of the modelled concentration. In order to model the real behaviour and a sort of inertia of the on-site sensors, the following standard formula has been coded:

$$\mathrm{C}_{\mathrm{unit-conv,i}}(\mathrm{t}) = \mathrm{C}_{\mathrm{unit,i}}(\mathrm{t}) \times \left(1 - \exp\left(-\frac{t}{\tau}\right)\right) \tag{8}$$

With $C_{unit,i}$ is the modelled concentration at the i sensor, τ is the response time and $C_{unit\text{-}conv,\ I}$ is the modelled convoluted concentration.

All the synthetic tests carried out for the algorithm validation include the response time of the sensors.

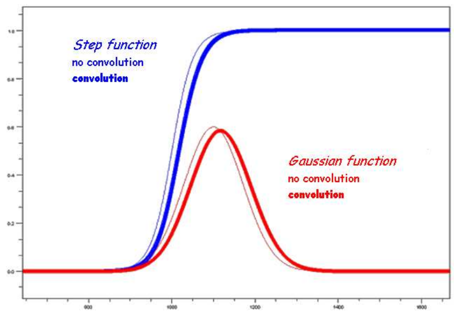


Figure 3: Convolution products examples

In addition to the inclusion in the algorithm of the response time, the H_2S sensors on the site have a 20 ppm upper threshold. All the higher concentrations are not recorded at the detector. That means only the first recording time will be used by the source inversion algorithm.

CASE STUDIES OF THE SOURCE TERM DETERMINATION ALGORITHM

Before full implementation, preliminary tests were conducted to validate the principle of the Bayesian algorithm. Initially a very simple configuration test with 5 potential sources and 3 sensors and no numerical representation of the installation was conducted successfully. A second series of tests were carried out using the detailed numerical representation of the site of Lacq on a part of the local domain. A domain of 1000 x 600 m was chosen within the local domain with 10 H_2S sensors and 5 potential sources. So far this configuration is arbitrarily determined with the objective to test the inference algorithm in a complex situation basically identical to the fully detailed configuration.

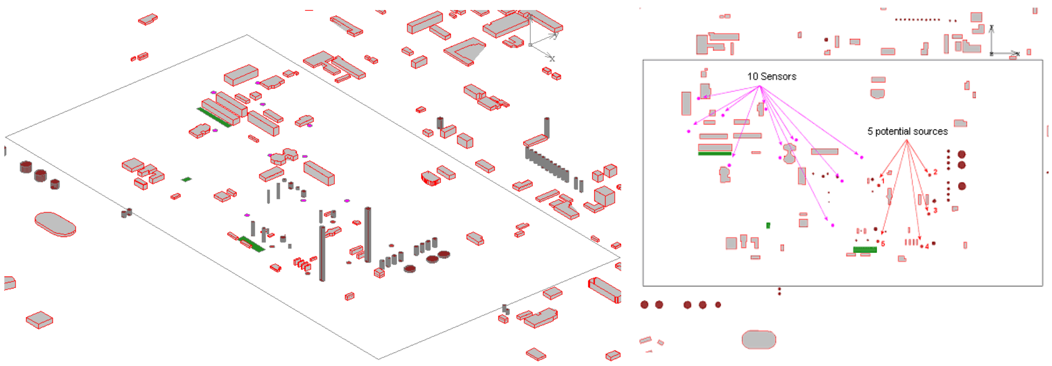
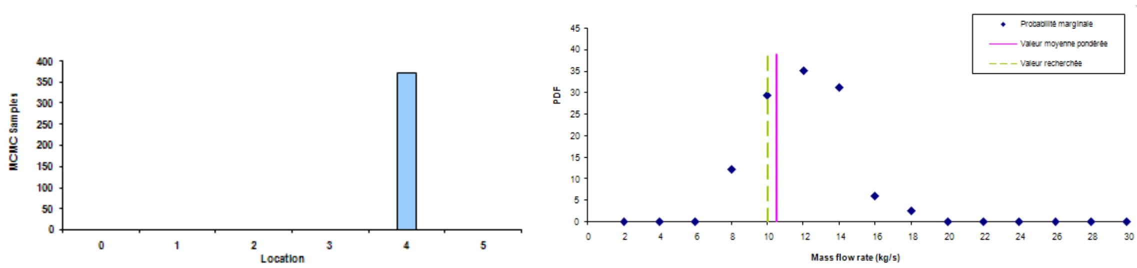


Figure 4: Numerical model of terrain for source reconstruction testing

In the case of this test the condition for the unsteady accidental release was chosen to be 10 kg/s during 30s from source n°4 with a boundary condition of 3m/s 115°N. The detection threshold is fixed at 5ppm with only two measures at the sensor for this case. A sampling frequency of the signal at each sensor has been set to every 5s. The algorithm was authorized to run for 40 000 propositions of set of parameters. The space of "posterior" i.e. source parameters tested against actual concentration fields is sampled through a Markov Monte Carlo Chain (MCMC) process and a probability density function is derived from these results. The histograms shown describe the marginal distribution of the MCMC samples.



	Mass Flow (Kg/s)	Starting time (s)
Average MCMC samples	11.9	120
Average Weighted value	10.5	121
Target	10	130

Figure 5: Marginal distributions and summary statistics

CONCLUSIONS

Some of the key elements of the modelling platform dedicated to the real time simulation of accidental release on industrial site have been tested successfully in unsteady mode. Bayesian inference theory has been used to formulate the problem of source determination in the complex geometries of the industrial site by using a CFD code and a Markov Chain Monte Carlo sampling approach. The same algorithm framework will be now expanded to the full local domain. Then, the first phase of a measurement campaign which took place in March 2010 on Lacq site will provide real data for further validations. The second phase is likely to be carried out in summer 2010.

ACKNOWLEDGEMENT

The authors thank CITEPH (Concertation pour l'Innovation Technique pour l'Exploration et la Production des Hydrocarbures) and Frédérique Michaud, R&D Delegate in TOTAL E&P for the support given to this project.

REFERENCES

Borysiewicz M.J., M.A. Borisiewicz: Atmospheric dispersions modelling for emergency management, Institute of atomic energy, 2006

Hill R. *et al.*, 2007: Field and wind tunnel evaluation of CFD model predictions of local dispersion from an area source on a complex industrial site, *Proceedings of the 11th International Conference on Harmonisation within Atmospheric dispersion models for regulatory purposes*.

Chow F.K., B. Kosovic and S.T. Chan, 2006: Source Inversion for contaminant Plume dispersion in Urban Environments using building-resolving simulations. *86th American Meteorological society annual Meeting, Atlanta, 2006.* Lawrence, Livermore National Laboratory, Livermore, CA.

Keats A., F.-S. Lien and E. Yee, 2006: Bayesian inference for source determination with applications to a complex urban environment. Atmospheric Environment 41, 465-479.

Mazzoldi, A., *et al.*, CFD and Gaussian atmospheric dispersion models: A comparison for leak from carbon dioxide transportation and storage facilities, Atmospheric Environment (2008), doi:10.1016/j.atmosenv.2008.06.038

Neuman S., L. Glascoe, B. Kosovic, K. Dyer, W. Hanley, J. Nitao and R. Gordon, 2006: Event reconstruction with the urban dispersion model. *American Meteorological society annual Meeting, Atlanta, 2006.* Lawrence Livermore National Laboratory, Livermore, CA.

US-EPA, 1998, Reference for fluidyn-PANACHE in Appendix W to Part51- Guideline on air quality models