HARMO15
15th International conference on harmonization within Atmospheric Dispersion Modelling for Regulatory Purposes

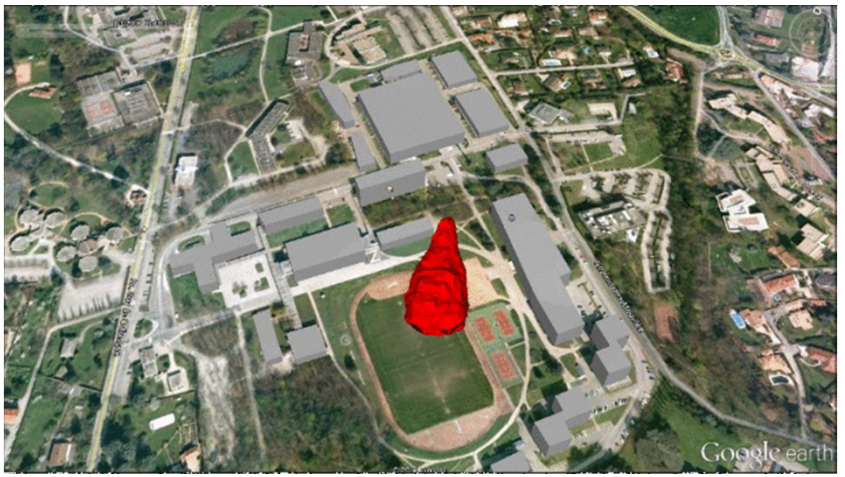
May 6-9 2013, Madrid, Spain

Uncertainty treatment in dispersion modelling of accidental releases

Vincent Dubourg, Patrick Armand, David Poulet,
Florian Vendel, Sébastien Argence, Thierry Yalamas,
Fabien Brocheton and Perrine Volta

Introduction

- *High fidelity* atmospheric dispersion modelling…



- … increasingly depends on *our knowledge of the exact environmental conditions*.
- Such conditions are unknown to some extent, especially in the case of accidental releases.

We propose *a risk assessment framework* that accounts for such uncertainty in the form of *probability distributions*.

Outline

- Dispersion modelling in the presence of uncertainty
  - Dispersion modelling
  - Uncertainty modelling
  - Quantity of interest for risk assessment
- Risk assessment methodology for urgent situations
  - Brute-force approach
  - Elements of surrogate modelling
  - Scalar-valued Gaussian process predictors
  - Dimension reduction using principal component analysis
- Results

Outline

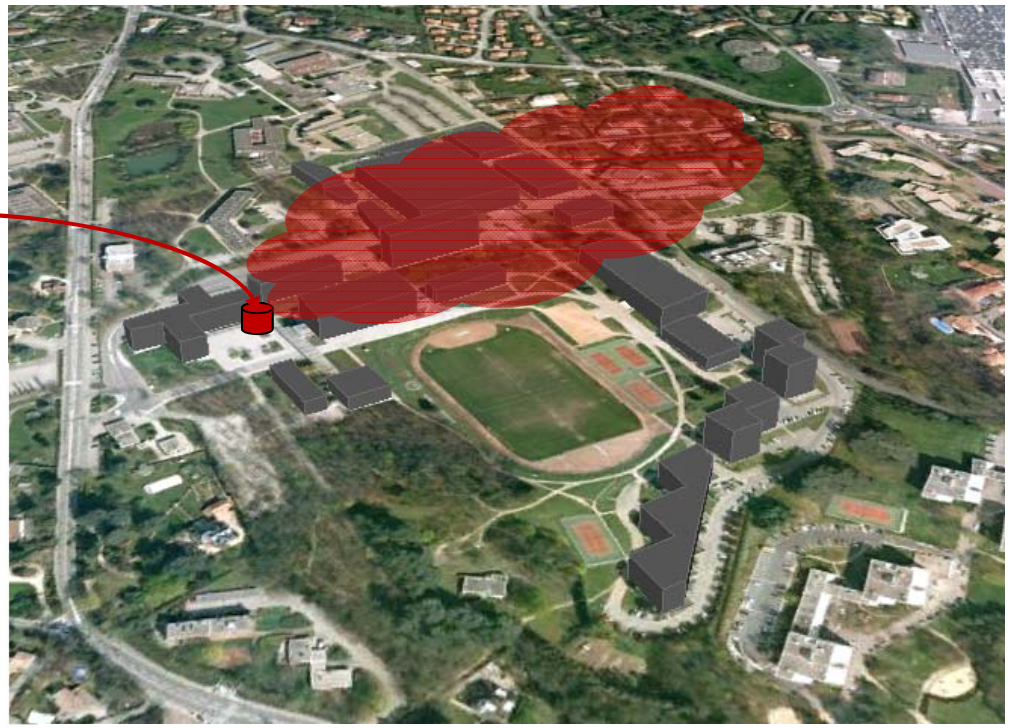
- Dispersion modelling in the presence of uncertainty
  - Dispersion modelling
  - Uncertainty modelling
  - Quantity of interest for risk assessment
- Risk assessment methodology for urgent situations
  - Brute-force approach
  - Elements of surrogate modelling
  - Scalar-valued Gaussian process predictors
  - Dimension reduction using principal component analysis
- Results

Dispersion & uncertainty modelling

Dispersion modelling

- The exact *source location* is *supposedly known*.
- The release *starts* on March 13th 2011 at 06:35:00 and *ends* at 06:40:00.
- *Meteorological conditions* (wind speed, direction, etc. …) are *uncertain (imprecise)*.

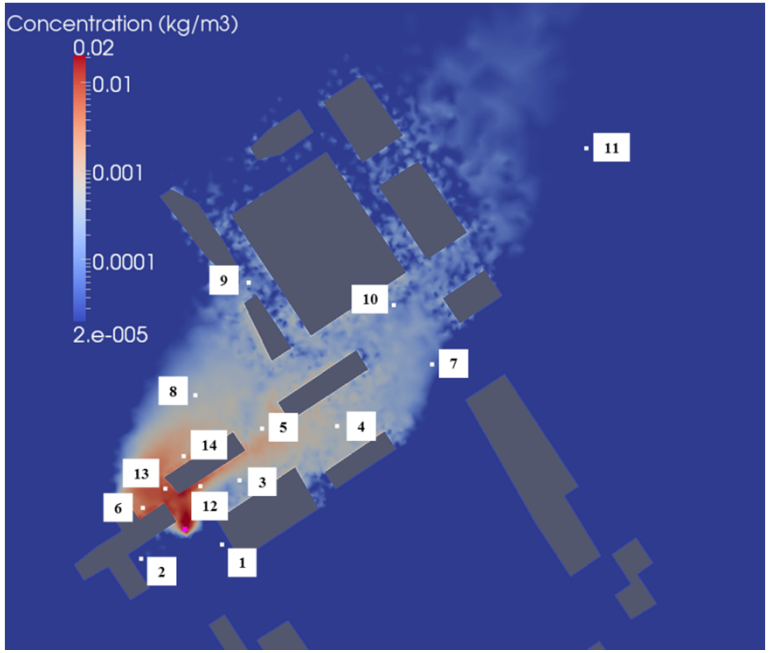
Source



Dispersion & uncertainty modelling

Dispersion modelling

- A *Lagrangian model* (SLAM) is used for simulating the dispersion of the pollutant (assuming a light gas behaviour).
- A *pre-computed CFD database* enables the calculation of the *perturbed wind field* in the constructed area in the vicinity of the source for *a large variety of incident winds* (using multi-linear interpolation).

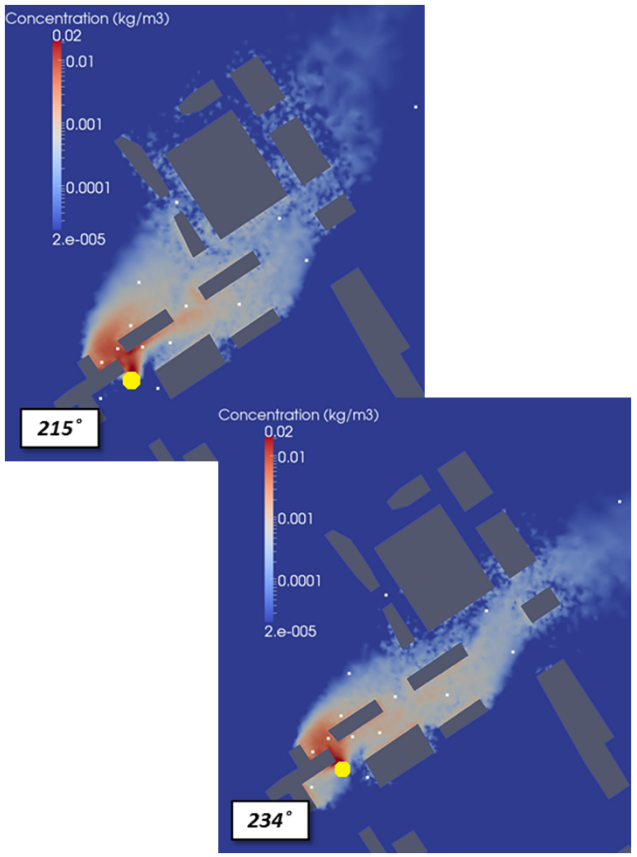


Dispersion & uncertainty modelling

Uncertainty modelling

- The *lack of knowledge* about some parameters describing the release conditions is modelled as *a probability distribution*.
- These variables are assumed independent in a first simplified approach.

Parameter	Probability distribution
Wind speed	Gaussian with mean 2 $m.s^{-1}$ and standard deviation 0.17 $m.s^{-1}$
Wind direction	Truncated Gaussian with mean 225° and standard deviation 22.15°, over [215°; 234°]
Cloud cover	Truncated Gaussian with mean 6 octas and standard deviation 1 octa, over [1 octa; 9 octas]
Temperature	Uniform over [14°C; 16°C]
Emitted quantity	Uniform over [70 $kg.s^{-1}$; 130 $kg.s^{-1}$]
Source height	Uniform over [1.75 m; 2,25 m]



Dispersion & uncertainty modelling

Quantity of interest for risk assessment

- We consider the *cumulated dose causing irreversible effects on human health* according to INERIS recommandations for *phosphine* :

$$\square E\square, \square, \square F = \int \square E\square, \square, \square F \, d\square$$

where :

- $\square$ denotes the random vector of *uncertain release conditions*
- $\square$ and $\square$ are the *position* and *exposure time* respectively
- $\square$ is the *instant phosphine concentration* calculated by SLAM
- $\square$ = 0.53 according to INERIS

- The subject is assumed *not to move* during exposure.
- The *risk analysis* consists in estimating:

$$\square = \text{Prob}\, x\square E\square, \square, \square F > \square z$$

where $\square$ = 20.10 according to INERIS.

Outline

- Dispersion modelling in the presence of uncertainty
  - Dispersion modelling
  - Uncertainty modelling
  - Quantity of interest for risk assessment
- Risk assessment methodology for urgent situations
  - Brute-force approach
  - Elements of surrogate modelling
  - Scalar-valued Gaussian process predictors
  - Dimension reduction using principal component analysis
- Results

Risk assessment methodology

Brute-force approach

- The spatio-temporal field of exceedance probabilities can be estimated using Monte Carlo sampling :

$$\hat{P}(E,F) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{E}(E, t, F > F)$$

- This estimator converges as the number of samples (the number of SLAM runs) increases.
- Convergence is measured in terms of its coefficient of variation:

$$\delta = \frac{\sqrt{Var\left[\mathbb{1}_{E}(E,F)\right]}}{P(E,F)} = \sqrt{\frac{1 - P(E,F)}{n\,P(E,F)}}$$

- Hence, a minimum of 10 000 samples is required in order to achieve a reasonable coefficient of variation of 32% on a probability of 10^{-3}.

Such a large number of SLAM runs is *incompatible with the urgency associated to accidental releases scenarii*.

Risk assessment methodology

We propose to replace SLAM by a *surrogate model* that is *much faster to evaluate*.

Elements of surrogate modelling

DOE	- Run the model $\mathcal{M}$ on a well-chosen set of input (gathered in an experimental design). - The purpose is to capture the largest amount of information about the functional relationship between its input x and output y.
fit	- Choose a family of surrogate models amongst artificial neural networks (ANN), support vector machine (SVM), Gaussian processes (GP), generalized linear models (LM). - Compute the surrogate model parameters from the dataset $\mathcal{D} = \left(\left(x^{(i)}, y^{(i)}\right), i = 1, \ldots, m\right)$.
validate	- Compute summary statistics about the relative error between the original model and its approximation. - The purpose is to qualify the surrogate model on a bounded domain of the input space.
predict	- Use the surrogate model instead of the original model to speed up uncertainty quantification or to optimize post-processings.

Risk assessment methodology

Scalar-valued Gaussian process predictors (*a.k.a.* Kriging)

- Assume we want to predict the dose □ = □ E□, □F *for some fixed point* □ *and time* □, but for any parameters value □.

- Kriging is a *Bayesian prediction technique* that starts from the following *Gaussian process prior model*:

$$\square\,E\square F = \square\,E\square F^{T}\square + \square\,E\square F$$

where:

  - □ E□F^T□ is the *trend* expressed as a *linear model*;
  - □ E□F is a *zero-mean stationary Gaussian process* with covariance function:

$$\square\,E\square, \square F = \square\square\,E\ddot{o}\square - \square'\ddot{o}, \square F$$

- Hence, the *observations* □ *in the dataset* □ and *some unobserved value* □ E□F *we would like to predict* are jointly distributed as follows:

$$E_{\square}{}^{\square}{}_{E\square F}\,F \sim \square\,E E_{\square}\,{}_{E\square F}{}^{\square\square}\,{}_{\square F}, \square\,x_{\square\,E\square, \square F}{}^{\square\,E\square F}\;{}_{1}{}^{\square(\square, \square)}{}_{zF}$$

Risk assessment methodology

Scalar-valued Gaussian process predictors (*a.k.a.* Kriging, cont'd)

- Kriging exploits *the non-zero cross-correlation term* □(□, □) in order to predict the unobserved value given the observations:

$$\ddot{u}\,E\square F = x\square\, E\square F \mid \square = E\square, \square = 1, \ldots, \square F, \square, \square z$$

- This *conditional distribution remains Gaussian*:

$$\ddot{u}\,E\square F \sim \square\, E\square\, E\square F, \square\, E\square F F$$

with *known mean* and *variance*.

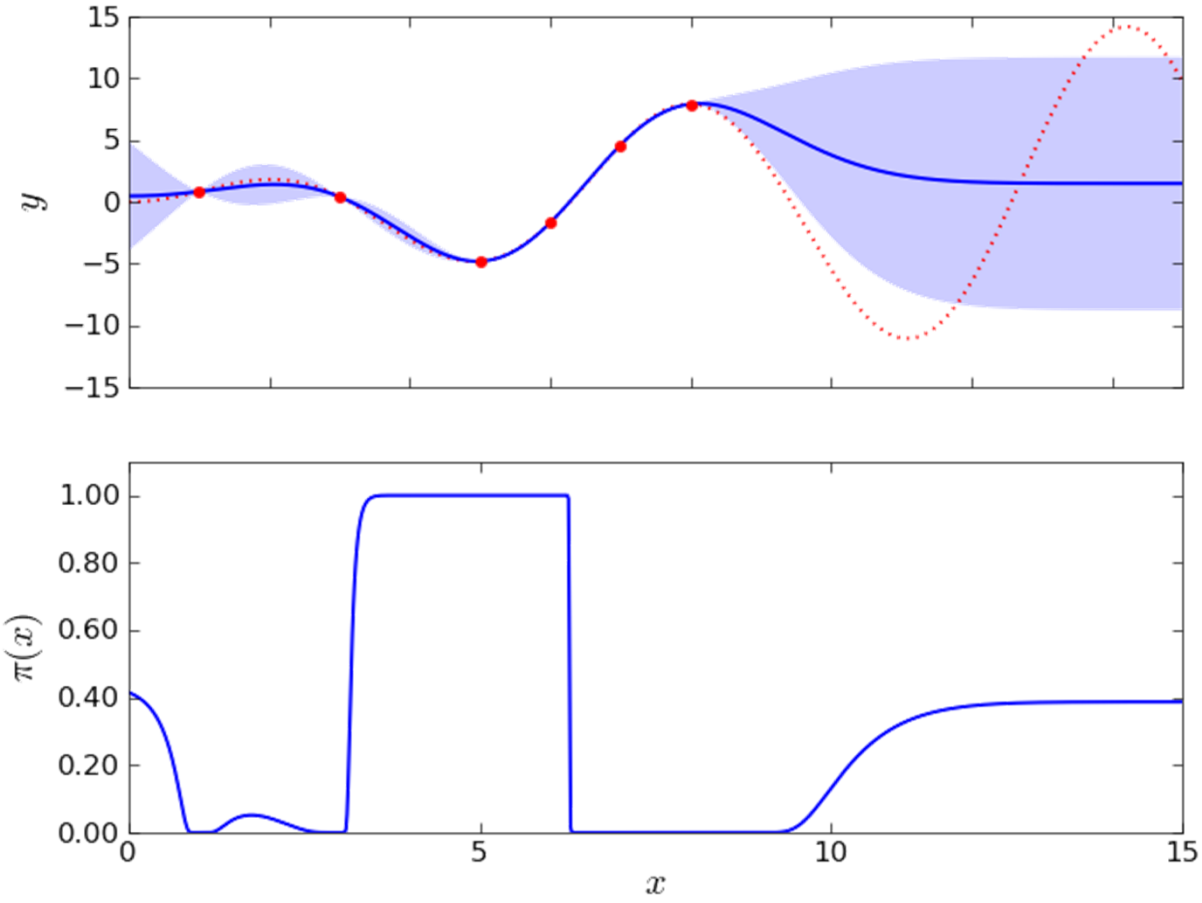
- In pratice □ and □ are determined from the dataset using *maximum likelihood estimation*.

- Eventually, we are able to predict, for any new parameters □:
  - the *expected value of the dose*: □ E□F ;
  - the probability that the value of the dose is less than some threshold □ with respect to *the uncertainty in the surrogate model*:

$$\square\, E\square F = \text{Prob}\, x\ddot{u}\, E\square F \le \square z = \Phi_{E\frac{\square - \square\, E\square F}{\square\, E\square F}F}$$

Risk assessment methodology

- Scalar-valued Gaussian process predictors (*a.k.a.* Kriging, cont'd)
  - Ex : $\mathcal{M}$: $\square \rightarrow \square$ sin($\square$)



$\square E\square F = \Phi_{E}\frac{0 - \square E\square F}{\square E\square F}_{F}$

Risk assessment methodology

Dimension reduction using principal component analysis (PCA)

- We could apply kriging *for all □ and □ over a spatio-temporal grid* in order to surrogate the whole output of SLAM.
- But this would be *heavy/long for dense grids* (□ × □ × □ = 50 × 50 × 71, for the present application)!
- It is proposed to exploit the *significant spatio-temporal correlation* (coherence) that exists in the output of SLAM for reducing its dimension to *a minimal vector of principal components*.
- The *invertible linear mapping*, known as the *Karhunen-Loève transform*, reads:

  $$\square : \ddot{o}\mathbb{R}_{\square} \overset{\mapsto}{\longrightarrow} \square = \mathbb{R}_{\square\square\square}, \qquad \mathrm{Cov}\,x\square z \approx \square\square\square$$

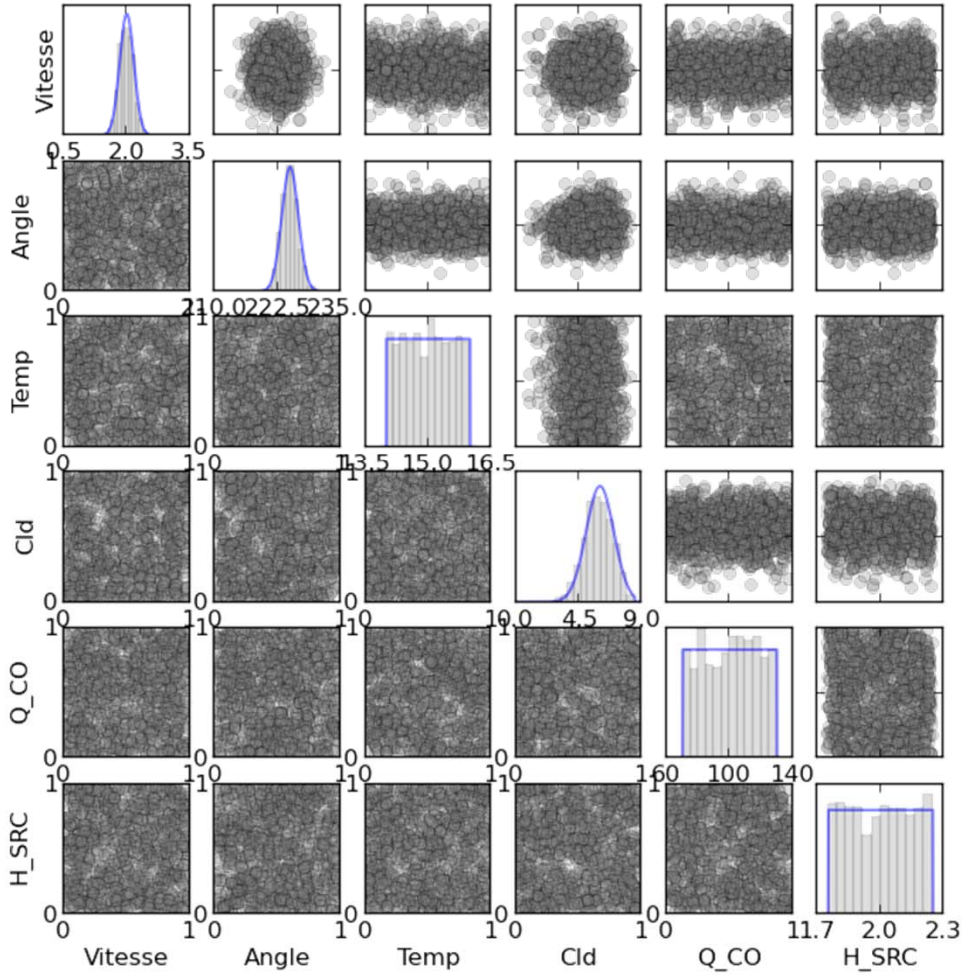
  where □ (□ × □) and □ (□ × □) are the matrices containing the eigenvalues and eigenvectors associated to *the □ ≪ □ largest eigenvalues of the output covariance matrix*.
- *Kriging is then applied to each component of the reduced vector* □ instead of the original one (the inverse transform is used at predict time).

Outline

- Dispersion modelling in the presence of uncertainty
  - Dispersion modelling
  - Uncertainty modelling
  - Quantity of interest for risk assessment
- Risk assessment methodology for urgent situations
  - Brute-force approach
  - Elements of surrogate modelling
  - Scalar-valued Gaussian process predictors
  - Dimension reduction using principal component analysis
- Results

Results

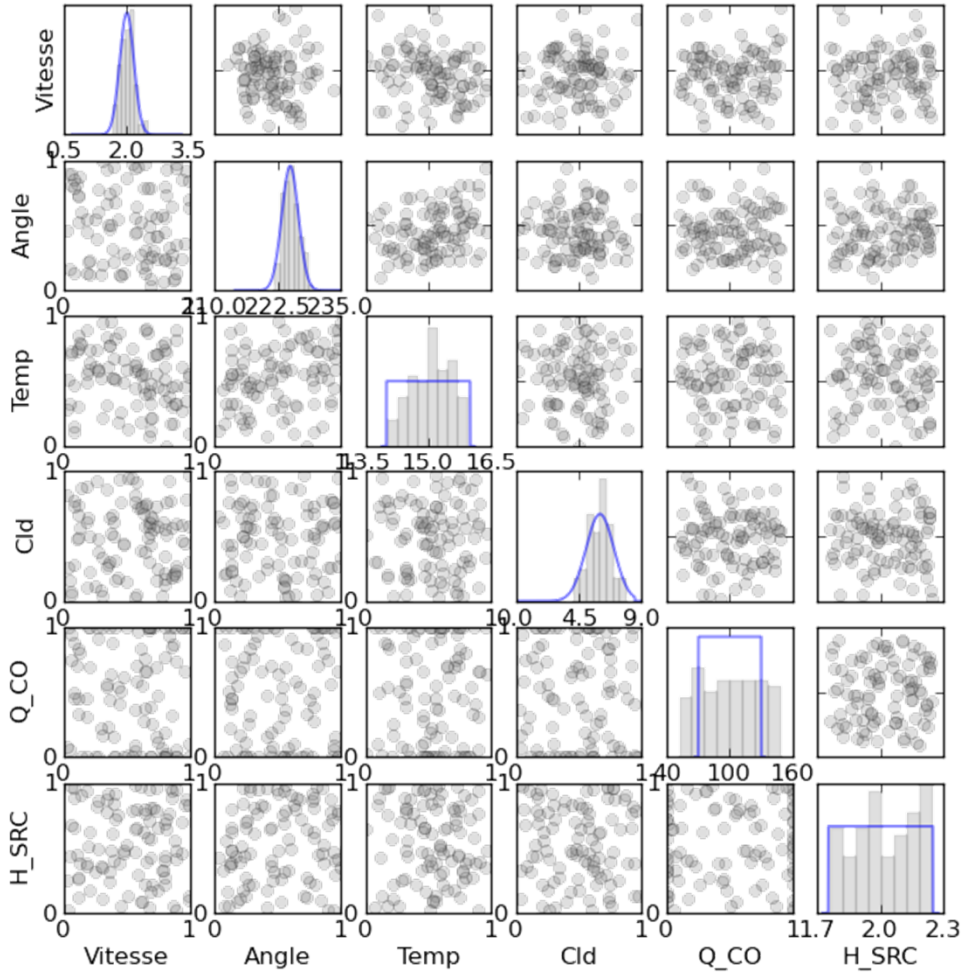
- Monte Carlo experiment used for validation of the surrogate-based approach (= 1,000)



High performance computing resources are used to distribute the 1,000 SLAM runs (over 100 CPUs)

Results

- Design of experiments used for the surrogate modelling (subset selection in the previous Monte Carlo experiment, □ = 100)



Results

Timewise surrogate modelling results (validation step)

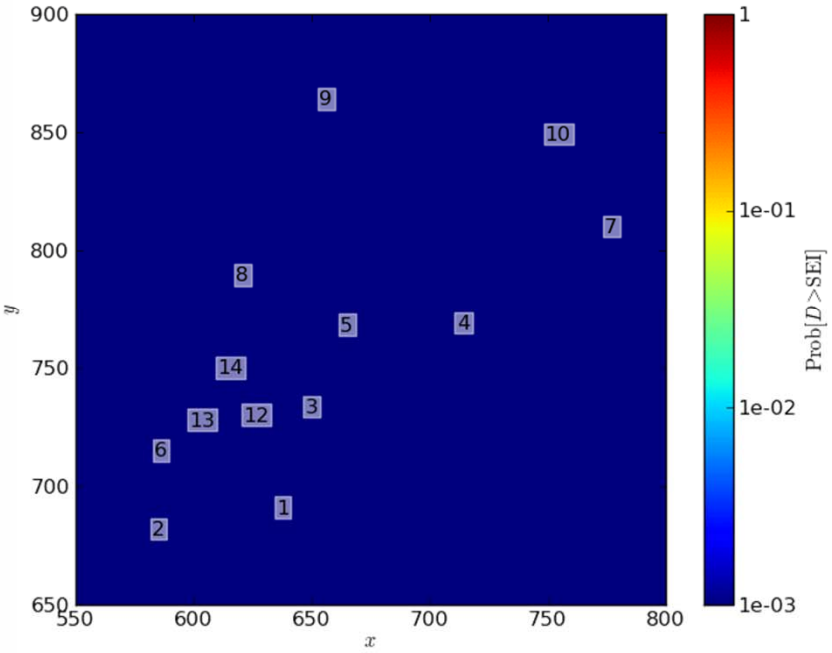
- □ is the *number of principal components* (avg = 92, min = 79, max = 94)
- □ is the *regression coefficient* estimated by means of *leave-one-out cross-validation* (avg = 0.65, min = 0.45, max = 0.92)*
- □ (test) is the *regression coefficient* estimated on the *900 other observations in the Monte Carlo experiment* (avg = 0.58, min = 0.44, max = 0.90)*
- The average fitting time is *12 seconds*.
- The average time to predict 10,000 points is *25 seconds* (which should be compared to 10,000 runs × 10 minutes per SLAM run)

Time step	r << n (n = 2,500)	Q^2	R^2 (test)	Fitting time (s)	Time to predict 10,000 points (s)
1	79	0.92	0.90	10.97	20.91
2	79	0.92	0.90	10.47	21.61
3	79	0.92	0.90	11.58	20.86
4	88	0.81	0.76	10.65	22.57
5	88	0.82	0.76	10.41	22.54
6	89	0.82	0.77	10.55	22.70
7	89	0.72	0.57	11.72	23.99
8	90	0.73	0.64	10.87	24.03
9	91	0.74	0.65	12.46	24.07
10	91	0.65	0.57	10.82	25.04
…					
66	92	0.49	0.47	12.11	27.71
67	92	0.48	0.46	12.31	27.61
68	92	0.47	0.45	12.44	27.44
69	92	0.47	0.45	12.31	27.48
70	92	0.46	0.44	12.15	27.49
71	92	0.45	0.44	12.34	27.54

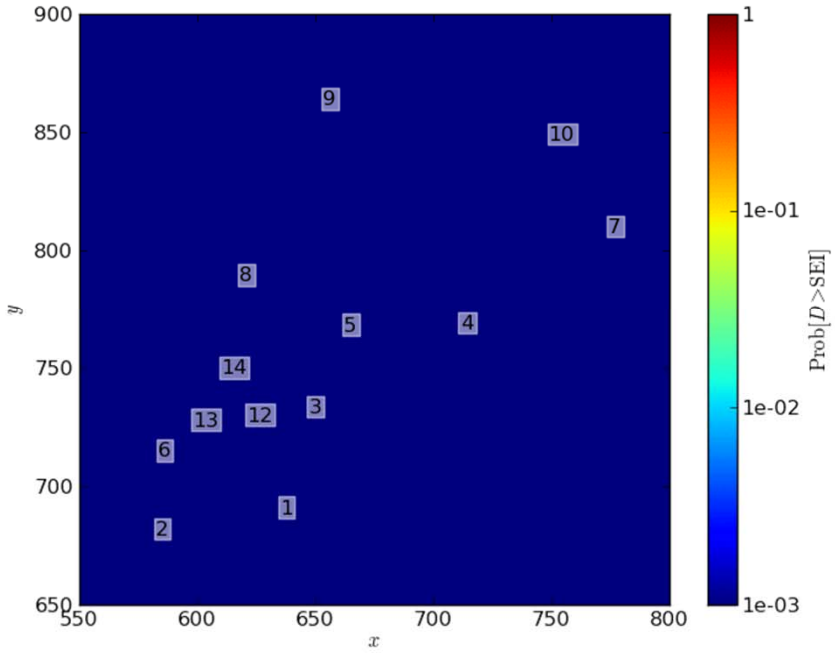
** The best regression coefficient is □ = 1 (perfect match between the model and its surrogate).*

Results

- Probability of exceeding the threshold dose of irreversible effects



Brute-force approach
(□ = 1,000)



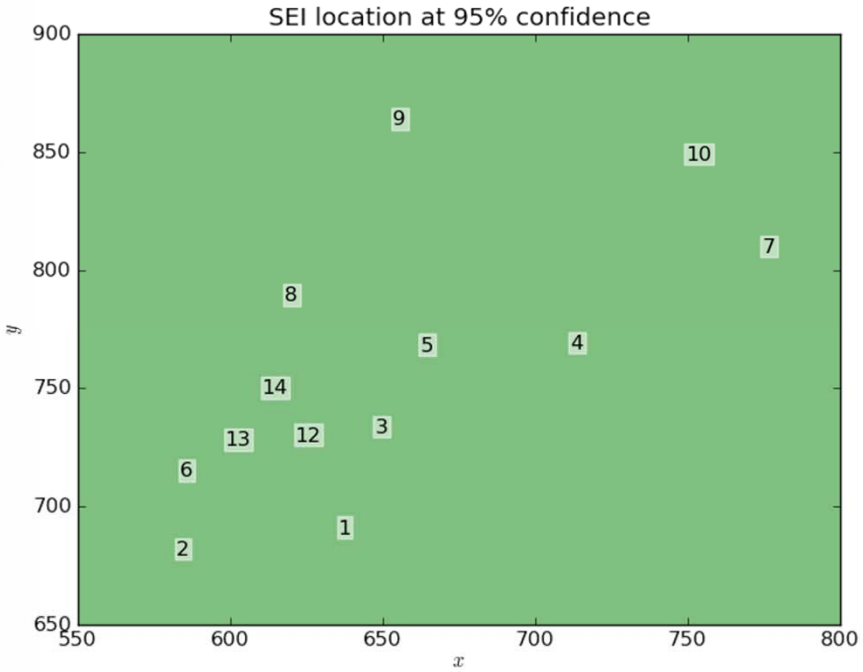
Surrogate-based approach
(□ = 10,000)

- The surrogate-based approach accounts for *the uncertainty in the kriging predictor* (Gaussian) :

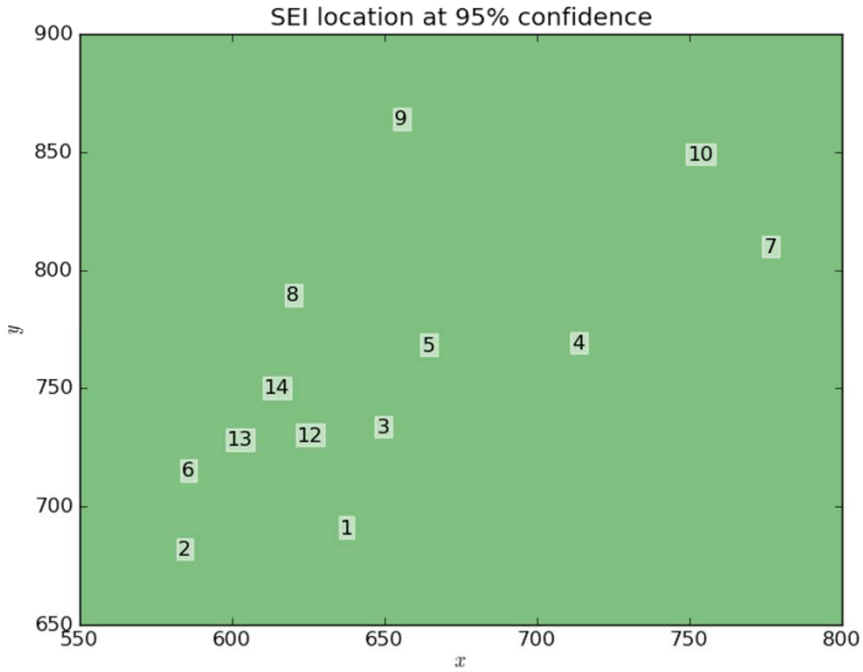
$$\text{ü}\ \square(E\square, \square F) = \frac{1}{\square}\,\text{ü}\ 1 - \Phi\left(E\frac{\square - \square(E\square, \square, \square F}{\square(E\square, \square, \square F}F\right)$$

Results

Risk map



Brute-force approach
(= 1,000)

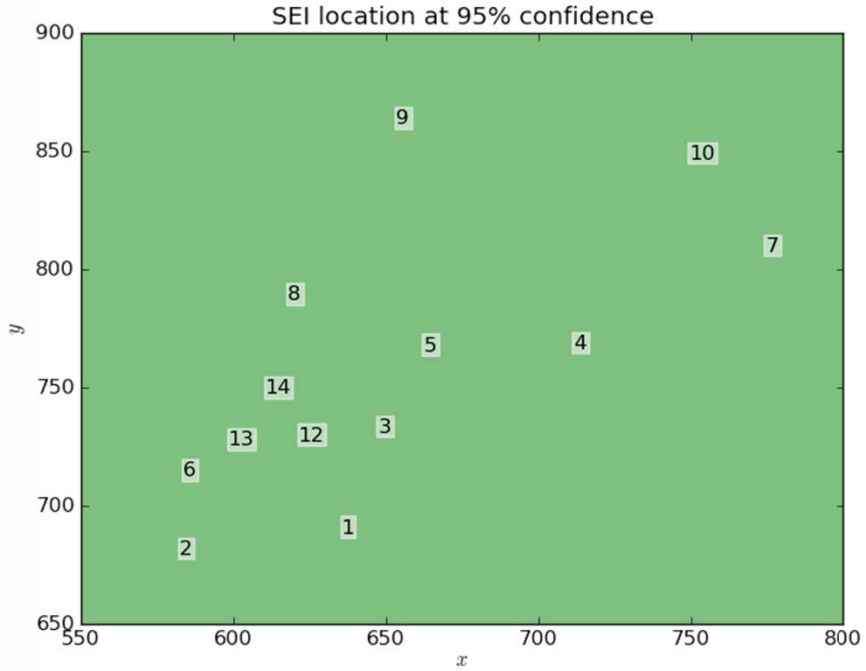


Surrogate-based approach
(= 10,000)

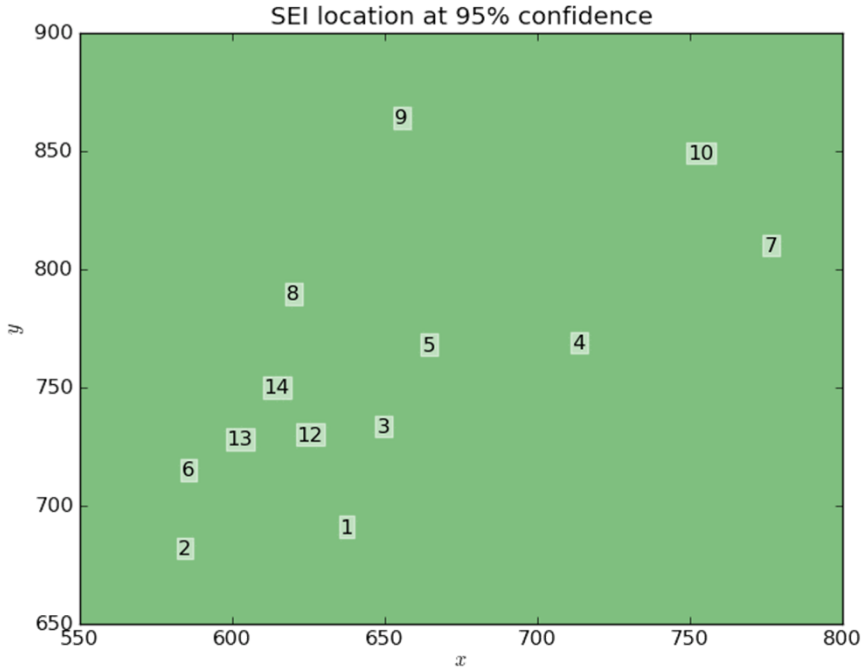
- The probability of exceeding the threshold dose of irreversible effects is :
  - less than 2.5 % in the green zone ;
  - between 2.5 % and 97.5 % in the orange zone ;
  - larger than 97.5 % in the red zone.

Results

Risk map (with different emitted quantity distributions)



□ ~ □ Ex70 kg. s; 130 kg. szF



□ ~ □ Ex7 kg. s; 13 kg. szF

- An *arbitrarily large emitted quantity distribution* was first used for *reaching the threshold of irreversible effects in the far field*.
- A *smaller emitted quantity distribution* eventually *augments the spread* of the uncertain (orange) zone.

Conclusion

- *Probabilistic modelling* is used to describe uncertain *release conditions*.
- *Risk* is assessed as the *probability of exceeding a critical dose*.
- *Surrogate modelling* enables a *drastic speed-up* in the production of risk maps :
  - provided the CFD database is already computed (for industrial sites at risk) ;
  - *20 minutes per SLAM run* in the DOE (× 100 runs, but × using HPC) ;
  - *about 12 seconds per time step* for fitting the kriging predictors ;
  - *about 25 seconds per time step* to predict the 10,000 configurations required for the final probability estimation.
- *Kriging* is a convenient surrogate for *incorporating the uncertainty about the surrogate model* in the final risk maps.
- Risk can be represented as *time-varying maps of dose exceedance probabilities*.