HARMO'17

17th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes

Probabilistic assessment of danger zones using a surrogate model of CFD simulations

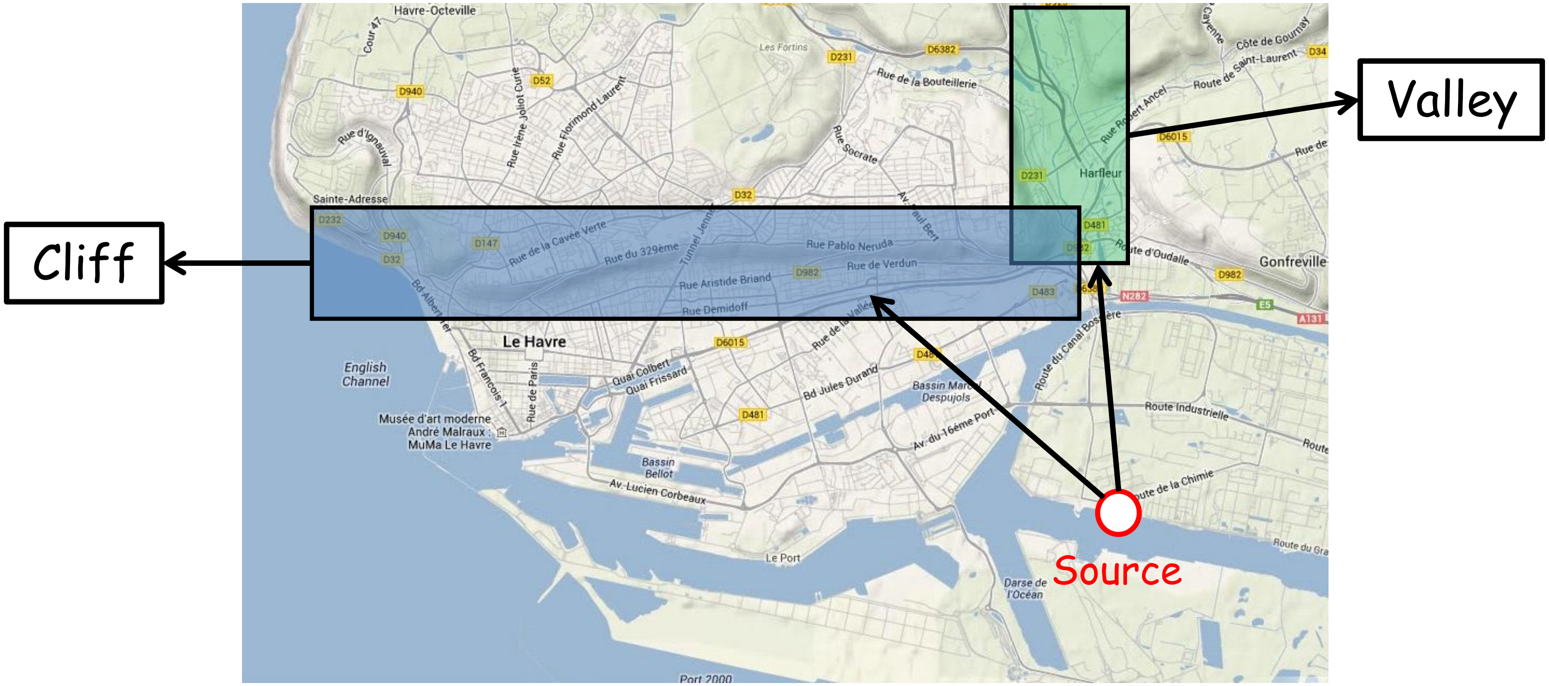
Felipe Aguirre Martinez, Yann Caniou, Christophe Duchenne, **Patrick Armand** and Thierry Yalamas

9-12 May 2016, Budapest, Hungary

Solutions for robust engineering

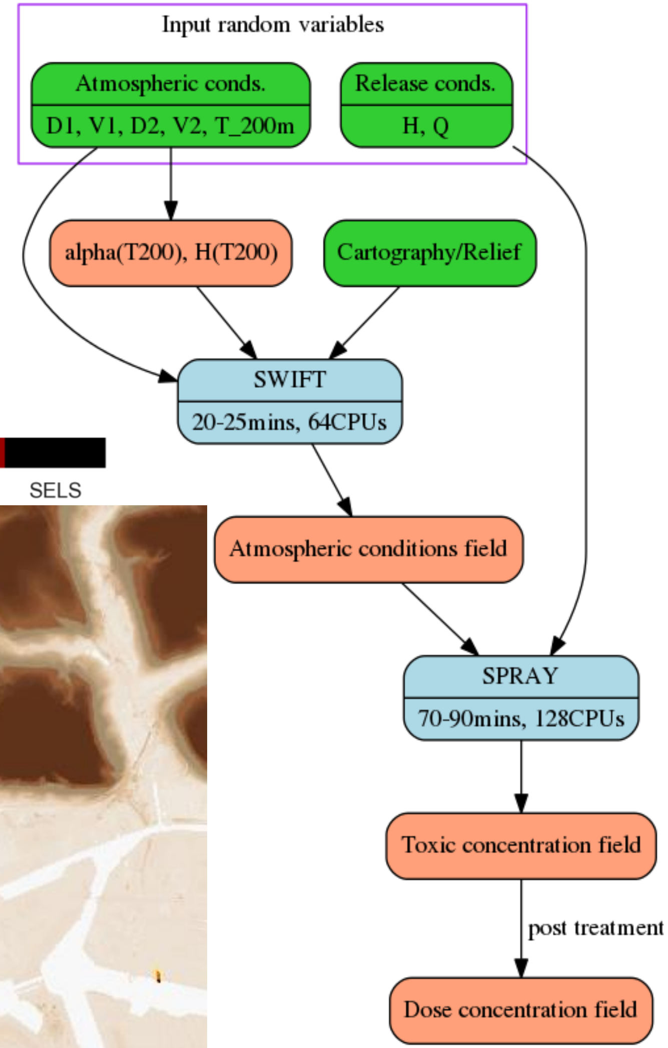
Presentation of the case-study

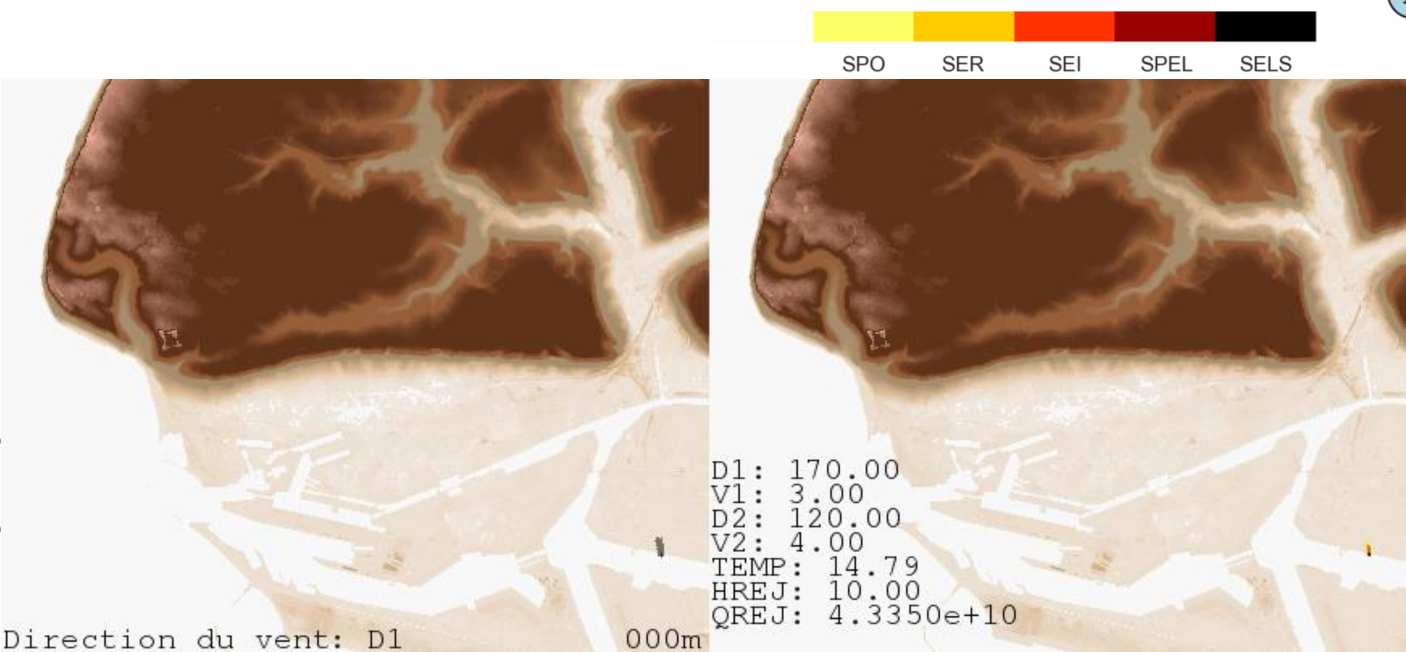
- The scenario is a fictitious accident in an industrial port
- Ammoniac is released for a period of 45 minutes
- Atmospheric and release conditions are uncertain
- The wind takes two main directions during the time period
  - In the first 30 minutes, the wind blows towards a valley on the north
  - Afterwards, the wind directs the pollutant towards the city and the cliff



Flow and dispersion modelling

- PSWIFT-PSPRAY computational chain
  - 1 hr 40 min on average per simulation
  - Distributed over 128 CPUs
  - Domain divided on 63 tiles
  - 60 time steps
- Deterministic or "reference" simulation





(a) Instantaneous concentration (b) Threshold exceedance

PHIMECA Solutions for robust engineering

Evaluation of the health consequences

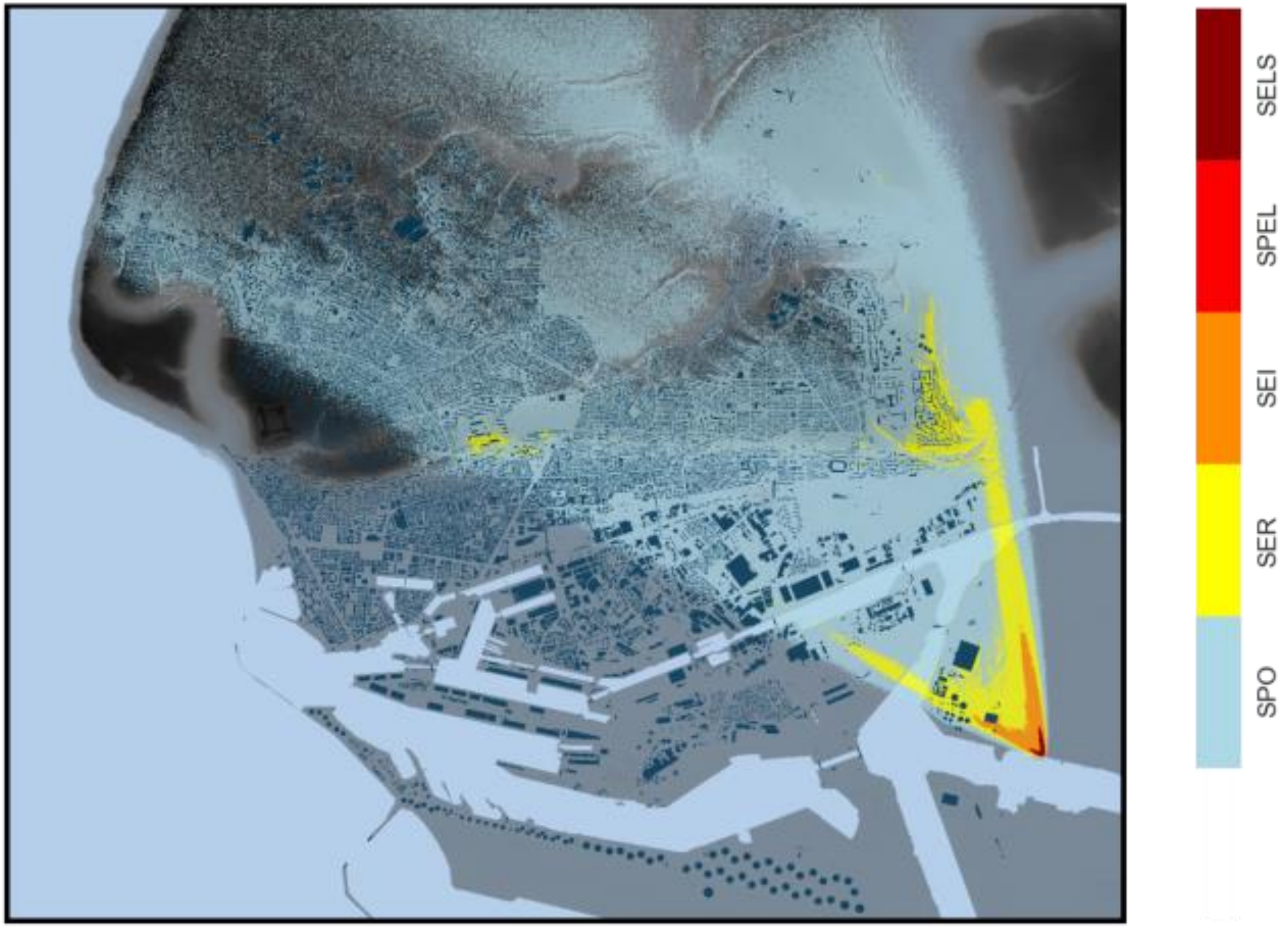
- Threshold concentrations of ammonia for more or less severe effects

Concentration (mg/m^3)	Exposure time (min)					
	60	180	600	1200	1800	3600
SELS	19623		6183	4387	3593	2543
SPEL	17710	10290	5740	4083	3337	2380
SEI	1050	700	606	428	350	248
SER	196	140	105	84	77	56

- Threshold doses (aka "toxic loads") defined as follows:
  - $D_{seuil}(\Delta t) = C_{seuil}(dt)\,\Delta t$ where:
  - seuil corresponds to SELS, SPEL, SEI or SER
  - Δt is the exposure duration
  - $C_{seuil}(\Delta t)$ is the threshold concentration for the exposure duration Δt
- Computation of the dose and comparison with the threshold values

Deterministic map of the danger zones

- "Reference" simulation (mode values of the parameters)

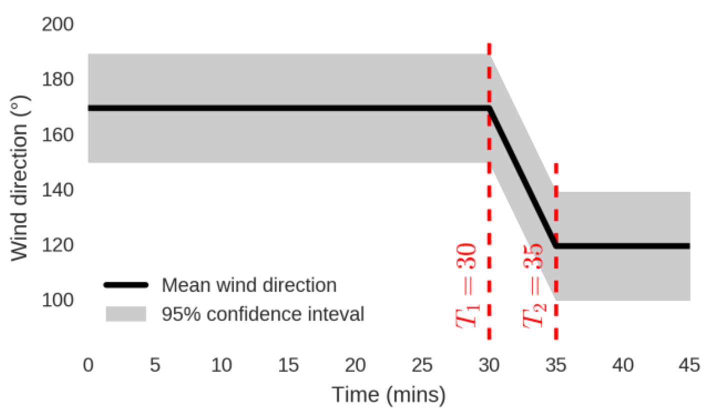


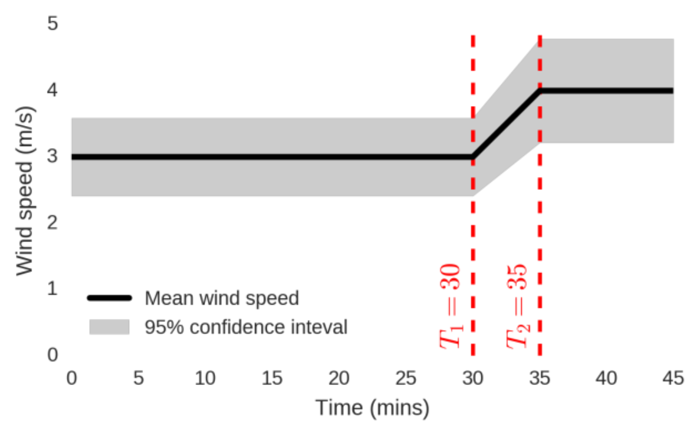
Modelling of the input data uncertainty

Probability laws

Variable	Distribution
Wind speed #1 (°)	$N(170, \sigma_1 = 10)$
Wind speed #1 (m/s)	$N(3, 0.3)$
Wind direction #2 (°)	$N(120, \sigma_2 = 10\ )$
Wind direction #2 (m/s)	$N(4, 0.4)$
Temperature gradient (°C/100 m)	$LN(\mu = -0.1, \sigma = 0.7, \gamma = -2)$
Release height (m)	$U[0,20]$
Amount of pollutant (mg)	$LN(\mu = 7.65\ E9, \sigma = 5.1\ E9, \gamma = 0)$

Emphasis on the wind conditions





Uncertainty modelling and risk assessment

- Quantity of interest >> Probability that the dose received by an individual at a given point in the domain exceeds a critical threshold for his health

$$P[D(\mathbf{X};\mathbf{p},t) > SEI] \approx \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{D(p,t)>SEI}(x^{(i)};\mathbf{p},t)$$

Where

- $\mathbf{X}$ is the random vector of uncertain parameters
- $D(\mathbf{X};p,t)$ is the dose computed with PSWIFT-PSPRAY at (p,t) for $\mathbf{X}$
- $\mathbf{1}_{D(p,t)>SEI}$ is the indicator function that is equal to one if the SEI threshold is exceeded for a given realization $x^{(i)}$ of X and zero otherwise
- N is the size of the Monte-Carlo experiment

Uncertainty propagation methodology (1/3)

Brute-force approach

- The probability of exceedance field can be estimated with Monte Carlo sampling

$$\hat{P}(\mathbf{p},t) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{D(p,t)>SEI}(x^{(i)};\mathbf{p},t)$$

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

$$\delta = \frac{\sqrt{\mathrm{Var}[\hat{P}(\mathbf{p},t)]}}{\hat{P}(\mathbf{p},t)} = \sqrt{\frac{1-p(\mathbf{p},t)}{Np(\mathbf{p},t)}}$$

- Hence, a minimum of 10 000 samples is required in order to achieve a reasonable coefficient of variation of 10% for a probability of 1%

Uncertainty propagation methodology (2/3)

We propose to replace PSWIFT-PSPRAY 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 the input x and output y of the model

Fit

* Choose a family of surrogate models amongst artificial neural networks (ANN), support vector machine (SVM), Gaussian processes (GP), generalized linear models (LM), etc.
* Compute the surrogate model parameters from the dataset $\mathcal{D} = \left((\mathbf{x}^{(i)}, \mathbf{y}^{(i)}), \mathrm{i} = 1, \dots, \mathrm{m} \right)$

Validate

* Compute statistics of the relative error between the original and approximate models
* 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 the uncertainty quantification and / or to optimize the post-processing of the results

Uncertainty propagation methodology (3/3)

Family of surrogate models >> Gaussian Process predictors (aka “kriging”)

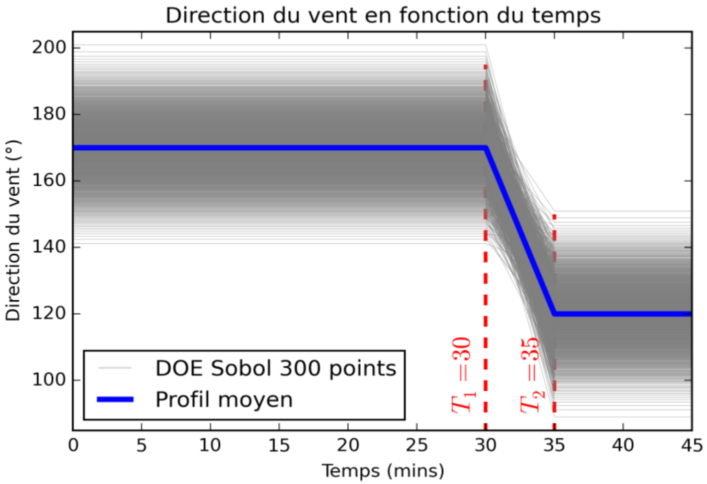
- Prediction of $y = D(\mathbf{x}; \mathbf{p}_0, t_0)$ at point $\mathbf{p}_0$ and time t_0 for any parameters $\mathbf{x}$
- Kriging is a Bayesian prediction technique that uses the joint distribution of the observations and the unobserved values of the dose Y
- The conditional distribution of the dose is Gaussian: $\widehat{Y}(\mathbf{x}) \sim \mathcal{N}_1\left(\mu_{\widehat{Y}}(\mathbf{x}), \sigma^2_{\widehat{Y}}(\mathbf{x})\right)$
- Expected value of the dose for $\mathbf{x}$ and probability that the dose exceeds a given threshold with respect to the uncertainty in the surrogate model

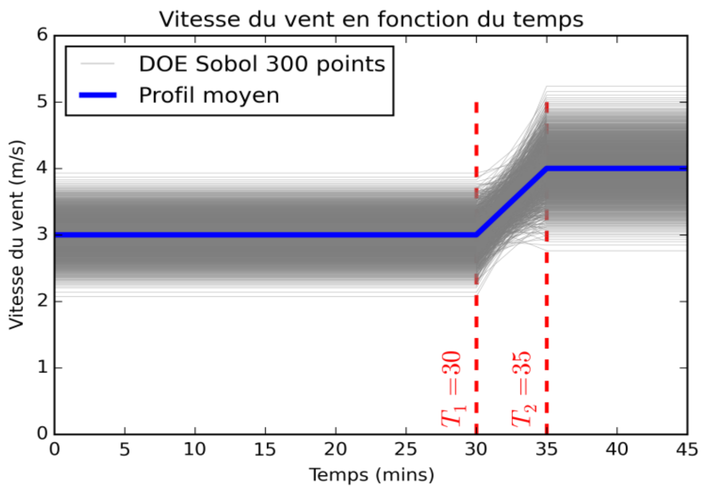
Dimension reduction using Principal Component Analysis (PCA)

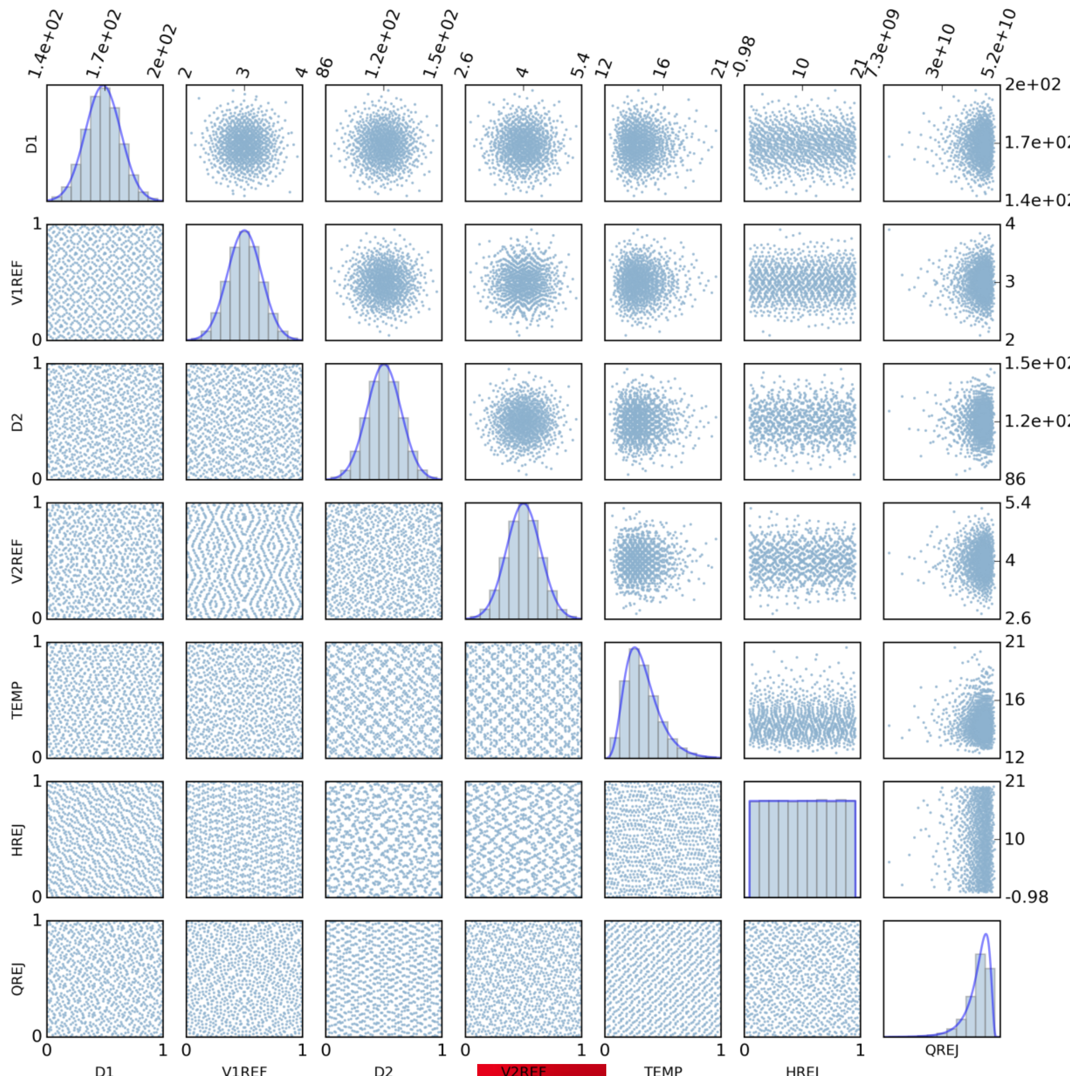
- Kriging for all $\mathbf{p}$ and t >> $N_{tiles} \times N_x \times N_y \times N_t = 63 \times 430 \times 430 \times 60 \approx 700\ 10^6$!
- Significant spatio-temporal correlation (coherence) in the output is used for reducing the dimension to a minimal vector of principal components.
- GP Predictors are applied to each component of the reduced vector
- An inverse transform is used to restore the response on the original vector

Results - Monte Carlo approach (1/3)

- Reference DOE generated from a Sobol' sequence of size 1'000



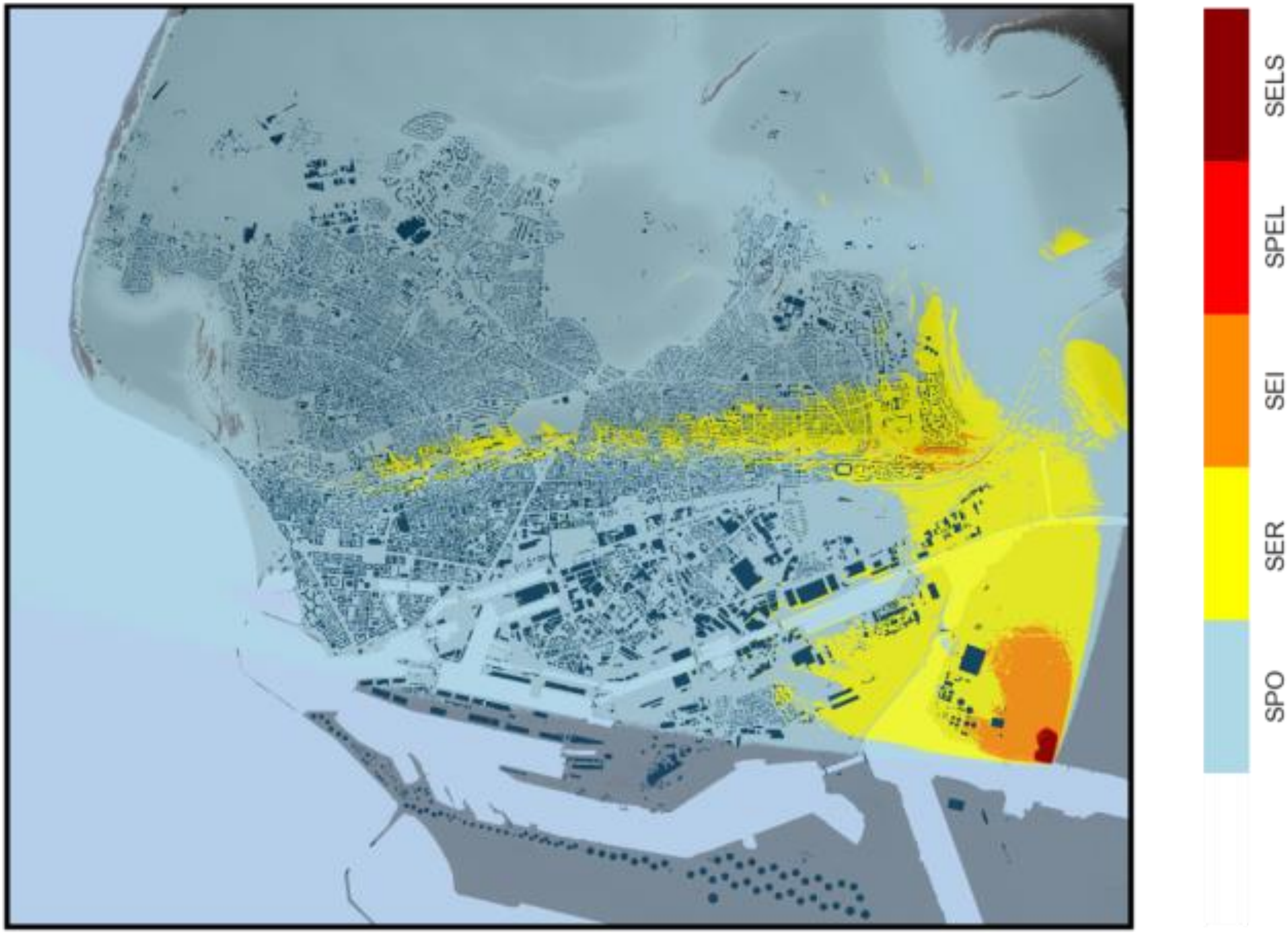




PHIMECA Solutions for robust engineering

Results – Monte Carlo approach (2/3)

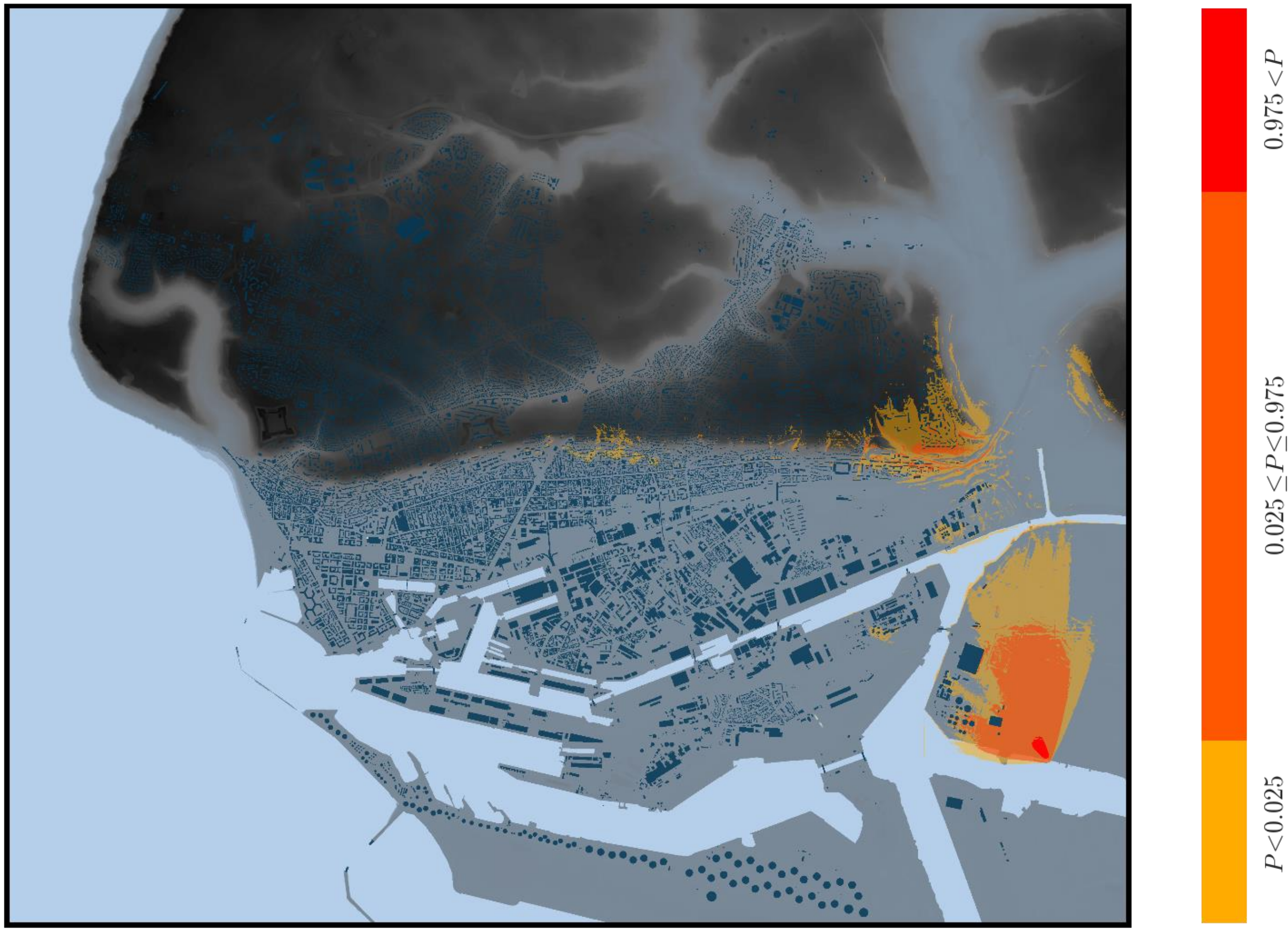
- Map of danger zones with a probability of exceedance larger than 2.5%



PHIMECA Solutions for robust engineering

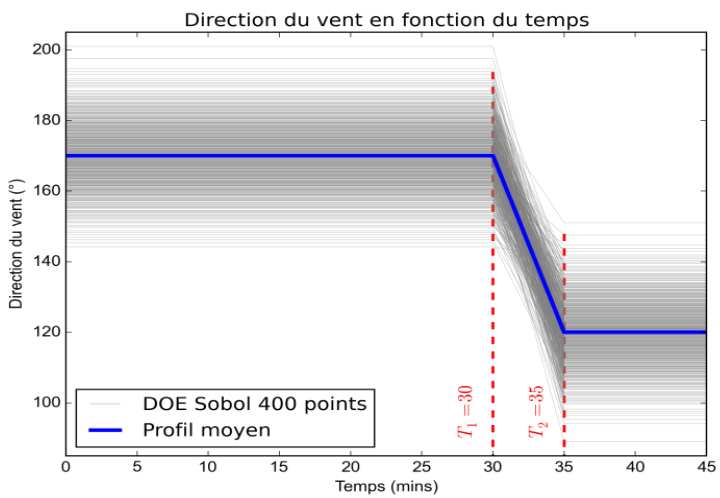
Results – Monte Carlo approach (3/3)

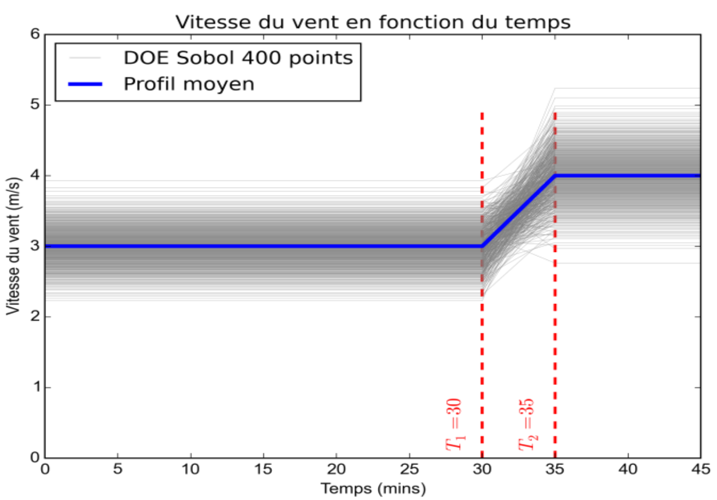
- 95% confidence interval of the boundary of the SEI danger zone (SEI: irreversible effects threshold)

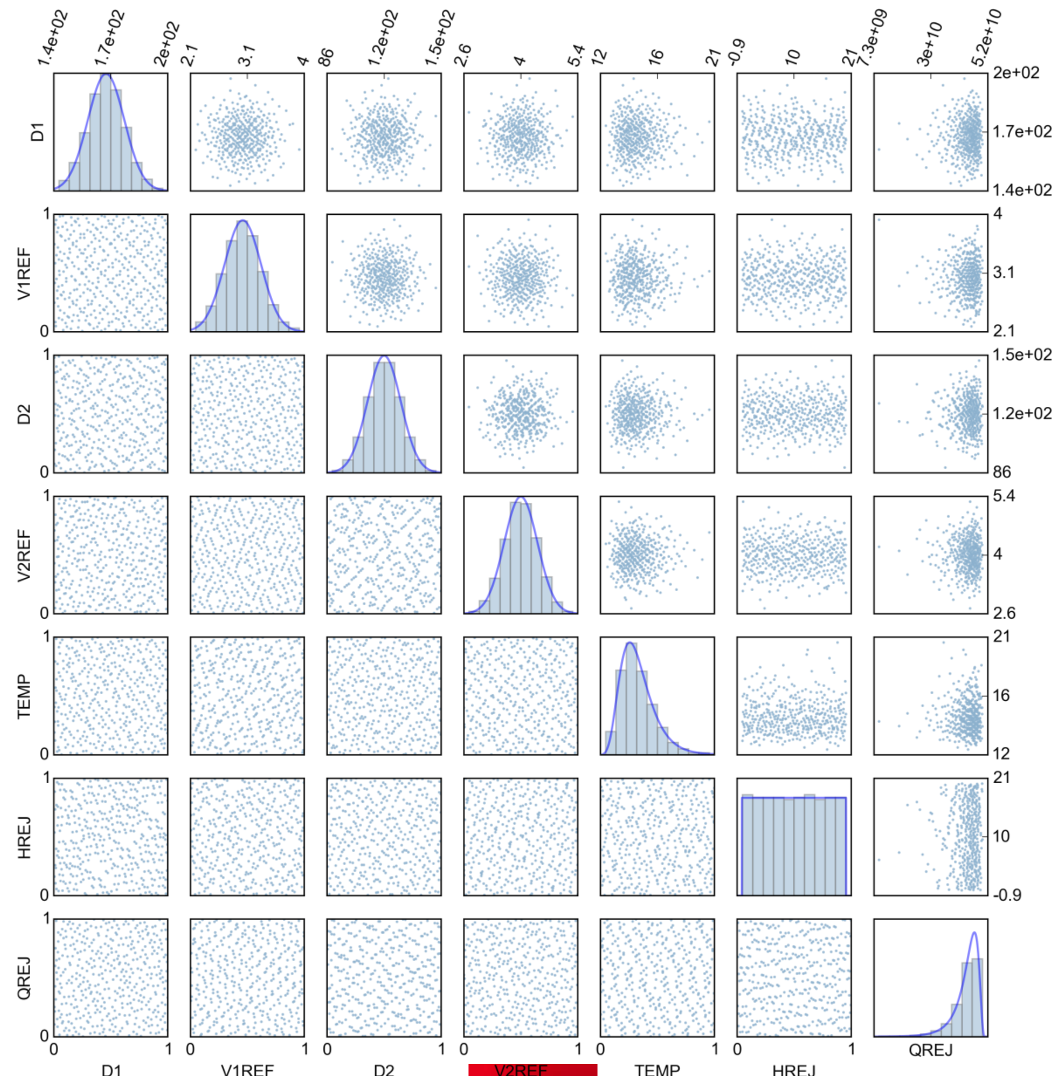


Results - Surrogate modeling approach (1/3)

- Training DOE » First 400 points of the reference DOE





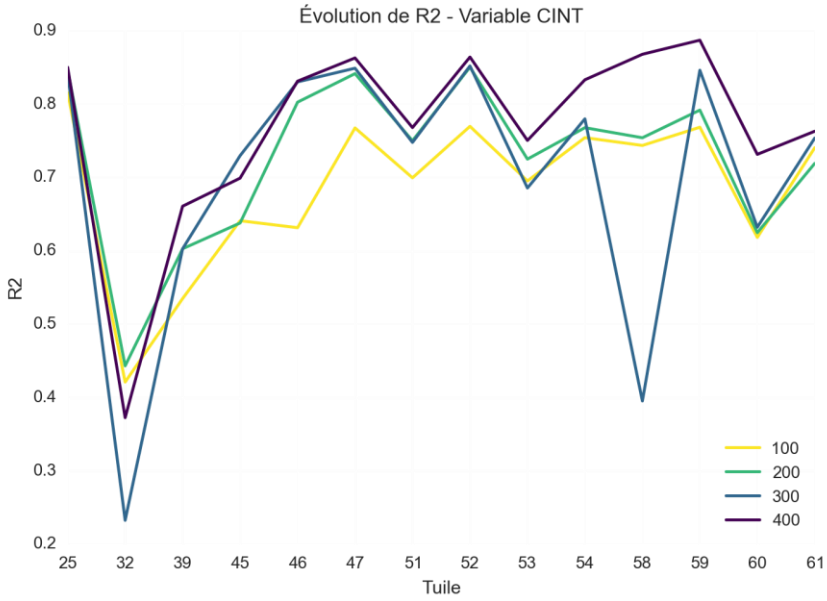


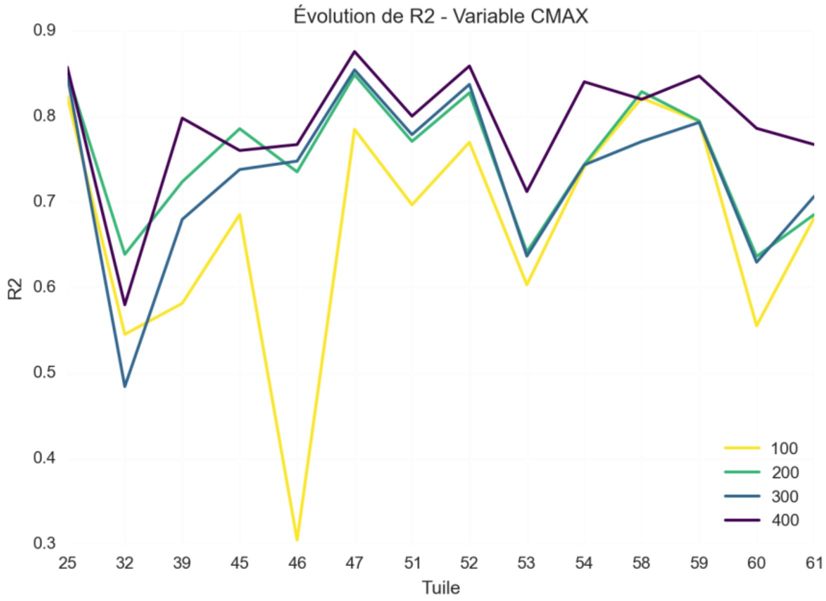
cea

PHIMECA Solutions for robust engineering

Results - Surrogate modeling approach (2/3)

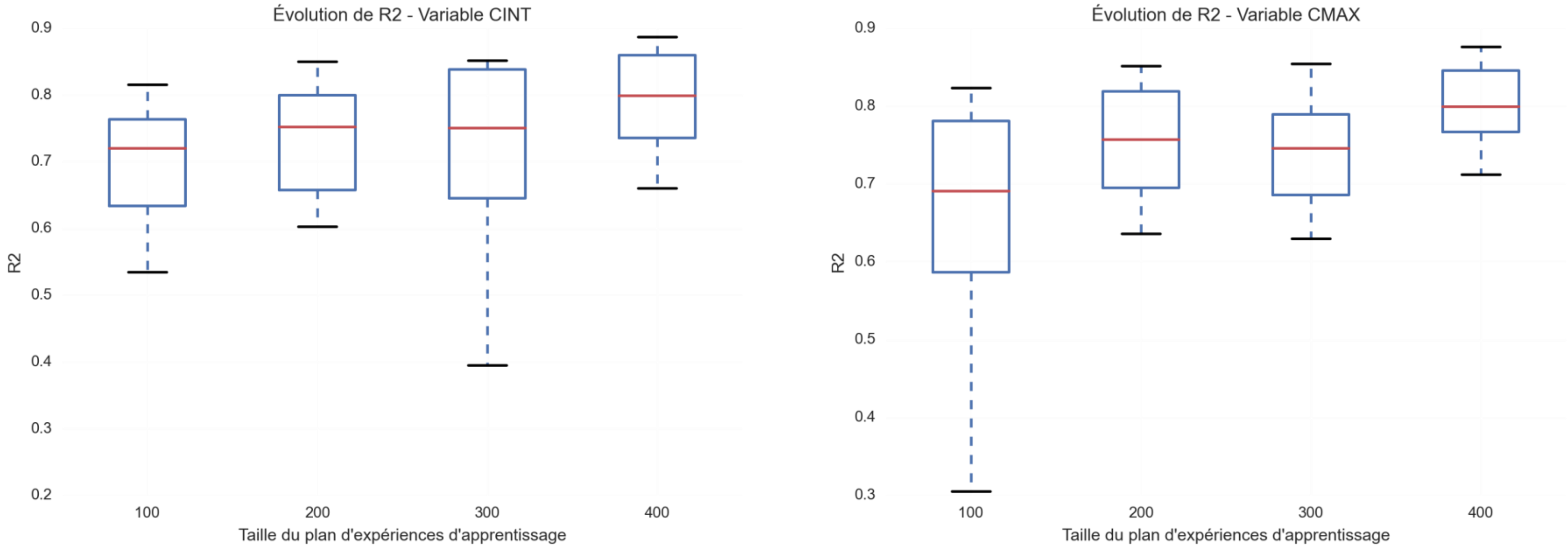
- Coefficient of determination of the surrogate models for each of the tiles and for training DOE of size 100, 200, 300 and 400





Results - Surrogate modeling approach (3/3)

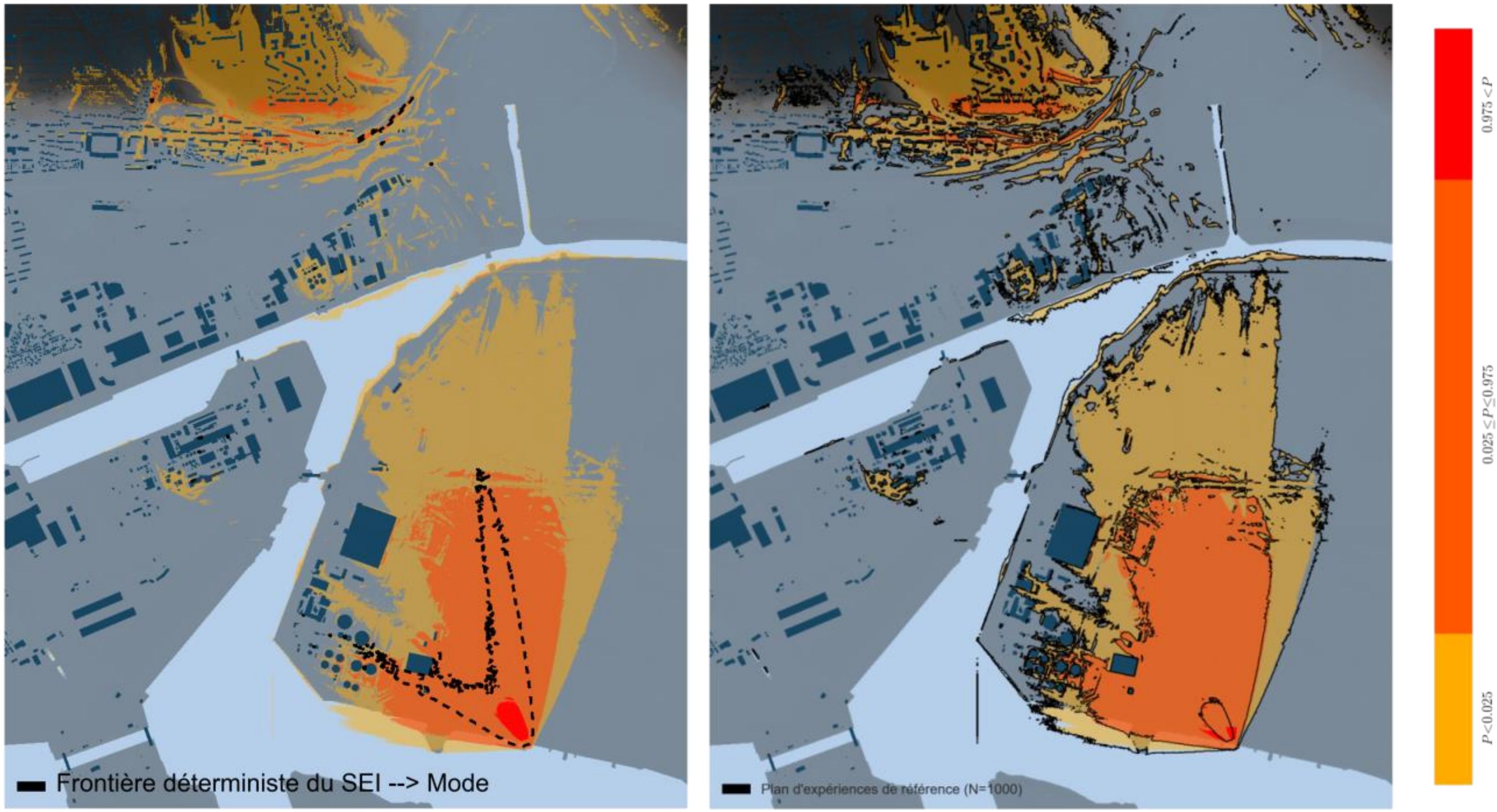
- Distribution of the coefficient of determination of the surrogate models for all tiles and training DOE of size 100, 200, 300 and 400



PHIMECA Solutions for robust engineering

Results – Comparison of various approaches

- 95% confidence interval of the boundary of the SEI danger zone (SEI: irreversible effects threshold)



Deterministic approach in black superposed to the Monte Carlo reference

Monte-Carlo reference in black superposed to the Gaussian Predictor approach

Conclusions

- Wind conditions have a great influence on the boundaries of the danger zones, that is why the 95% confidence interval is considerably wide
- Taking uncertainty into account reveals danger zones not identified by the deterministic approach (especially along the cliff)
- Even when applying a worst-case-scenario deterministic approach, maps may not reflect the impact of uncertainty on the wind conditions and some potential danger zones may be neglected
- With “only” 400 PSWIFT-PSPRAY simulations and surrogate modelling, we were able to obtain a reliable estimate of the 95% confidence interval of the boundary of the danger zones
- Surrogates are fitted in ~ 40 minutes and predictions take between 1 and 5 minutes depending on the number of simulations
- Further studies will be focused on sensitivity analysis and take account of real meteorological conditions