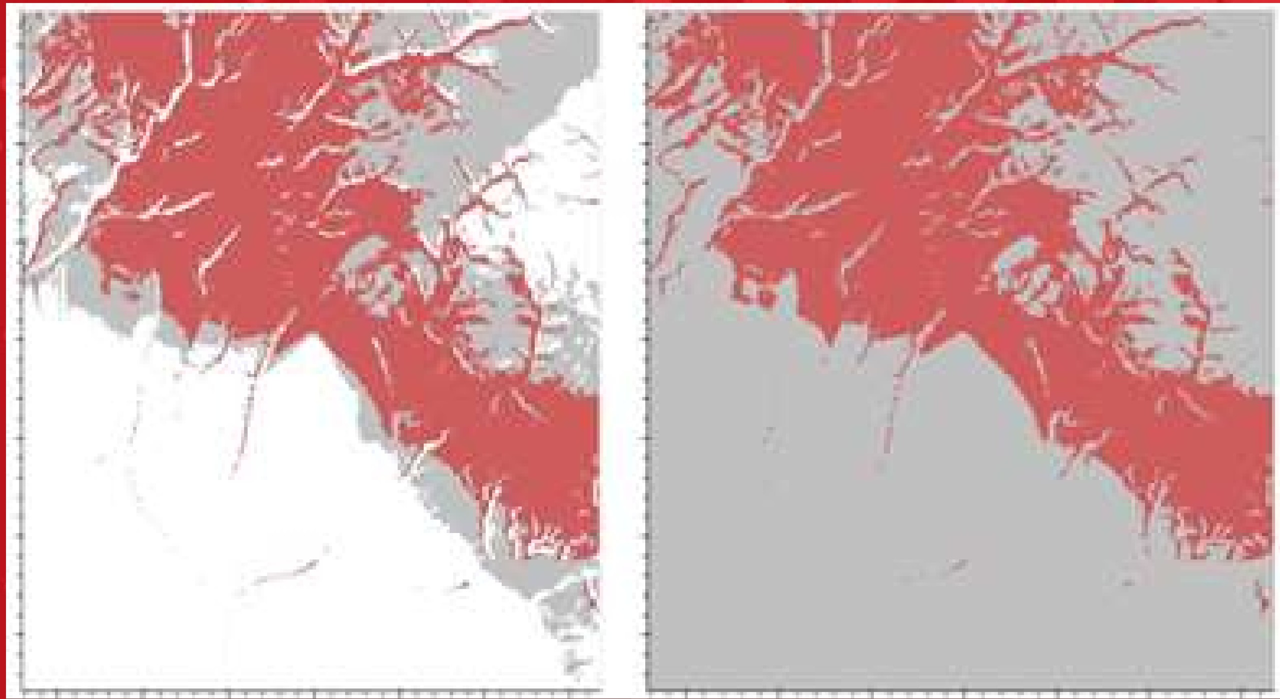
DE LA RECHERCHE À L'INDUSTRIE



Innovative probabilistic modelling of risk zones in the event of accidental atmospheric releases

Maéva CAILLAT, Valentin PIBERNUS and Sylvain GIRARD

PHIMECA Engineering, F-75012 Paris, France

Patrick ARMAND and Christophe DUCHENNE

CEA, DAM, DIF, F-91297 Arpajon, France

Harmo'21 conference 27-30 September 2022

INTRODUCTION (1/2)

- At the local scale, the flow and dispersion in the atmosphere are strongly influenced not only by the synoptic meteorology, but also by the details of topography and land-use and the presence of buildings.
- Decision-makers (e.g. operators of facilities, civil security officials or public authorities) are more and more convinced of the interest of using 3D models provided they are embedded in decision-support systems.
- However, the uncertainties associated with the results of the 3D models are most often not accounted for, neither evaluated, nor presented, while the scientific community and also the users are aware of them.
- In this research work, a methodology has been developed to accurately estimate the probability of exceeding a concentration threshold in the event of accidental or malevolent releases into the atmosphere.
- This methodology has been applied to a case study inspired by the industrial accident that occurred in January 2013, at the Lubrizol chemical plant located in Rouen (France).
- Operational mistakes and system failures in the plant resulted in extended releases of hydrogen sulfide and mercaptan, both of which are foul-smelling when they exceed a specified concentration.
- The features of the incident are very complex in several respects, namely the terrain topography and land-use, the kinetics of the releases, and the meteorological conditions during and after the releases.

INTRODUCTION (2/2)

- The atmospheric dispersion of pollutants was simulated with a modeling system whose input data related to both the meteorology and the source term are extremely uncertain.

- While epistemic uncertainties are not the only ones, taking them into account in a probabilistic framework is absolutely required for reliable decision-making (Girard et al., 2020).

- A step-by-step approach was taken to estimate the probability of exceeding a concentration threshold that might represent a certain level of danger in the context of atmospheric dispersion:

  1) Performing a deterministic simulation to estimate the area where the concentration threshold is exceeded,
  2) Performing several simulations to evaluate point by point the probability of exceedance of the threshold,
  3) Trying to estimate the confidence and credible intervals associated to a given probability of exceedance,
  4) Lowering the limit of significance by using a credible interval with a conditional spatial independence criterion.

- This methodology is developed and illustrated throughout the presentation.

FLOW AND DISPERSION SIMULATIONS WITH PMSS

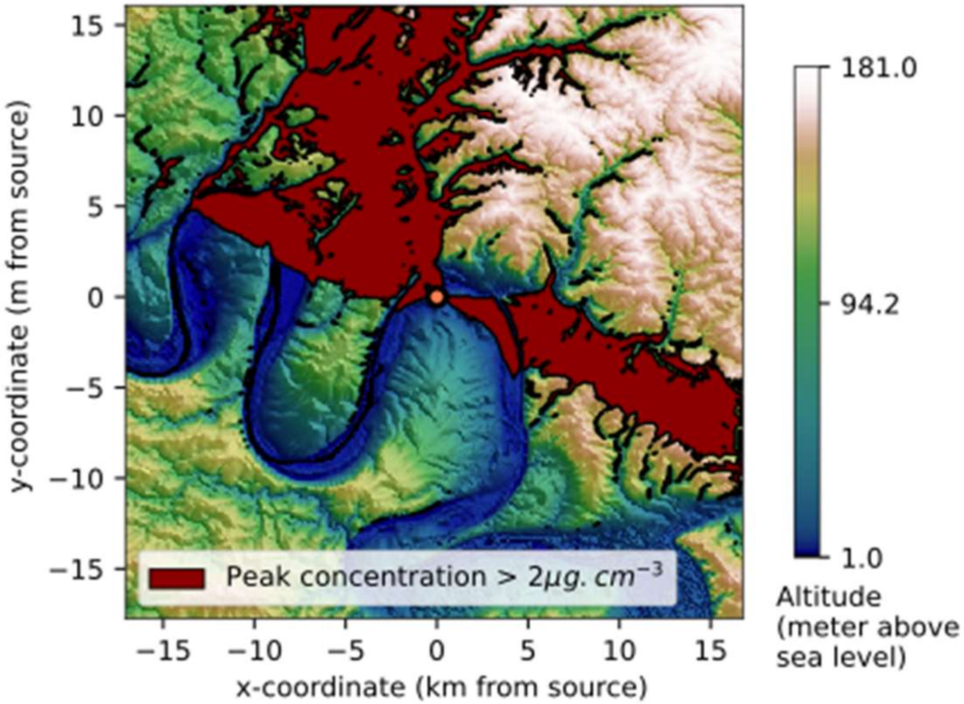
- The simulations were performed with Parallel-Micro-SWIFT-SPRAY (PMSS) which is based on MSS (Tinarelli et al., 2013) and combines the parallelized high resolution local scale versions of SWIFT and SPRAY models.
  1) SWIFT is a 3D diagnostic mass-consistent model using a terrain-following vertical coordinate ; the model uses large-scale met' data, local met' measurements, and analytical results of formulae in building-modified flow areas, which are interpolated and adjusted to generate 3D wind fields ; it also computes turbulent flow parameters.
  2) SPRAY is a Lagrangian particle dispersion model able to take into account the presence of obstacles ; the dispersion of the release is simulated by following the trajectories of a large number of numerical particles ; these trajectories are obtained by integrating in time the particle average and fluctuating velocity components.

- SWIFT and SPRAY were parallelized across time, space and numerical particles (Oldrini et al., 2017) ; the parallelism was shown to be very efficient, both on a multi-core laptop and on clusters of several hundreds of cores in a high-performance computing center (Oldrini et al., 2019) (Armand et al., 2021).

- PMSS was systematically validated against experimental wind tunnel and field campaigns for short and prolonged releases (Trini Castelli et al., 2018) ; in all configurations, the PMSS results complied with the statistical acceptance criteria of Hanna and Chang (2012) used for validating dispersion models.

DETERMINISTIC DECISION MAP

- The mercaptan atmospheric concentration generated by the accident that happened on January 2013 at the Lubrizol plant in Rouen (France) was simulated AT LOCAL SCALE (horizontal resolution = 2 m).
- We considered the 35-kilometer wide square area centered on the incident site ; the simulation covered a 35-hour period, so that all hazardous materials were either deposited or left the domain at the end.
- The input met' data were obtained from the meso-scale modelling system WRF (Skamarock et al., 2005) ; the source term was adapted from data established by Ismert and Durif (2014).
- We aimed to predict whether a concentration threshold was exceeded on the area over the whole period ; (here, we chose an arbitrary threshold value, 2 µg.cm^{-3}, to create a fictitious restricted area at risk).

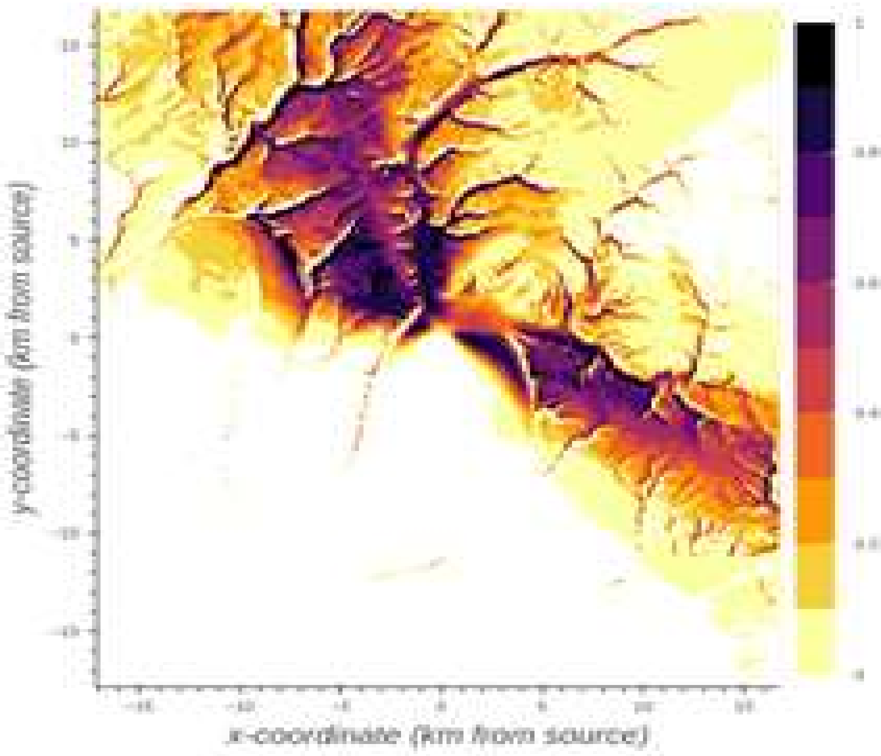


Concentration map of the chemical with a given concentration threshold

Deterministic simulation… While the inputs are substantially uncertain, and the results too!

Next, we accounted for uncertainty in the input parameters (wind speed and direction, rain intensity, and chemical release rate) and carried out 100 simulations in parallel using PMSS (duration ~1 hr).

- Random vector of uncertainties at a point s and time t: $X(s,t)$
- Maximum concentration at each s: $Y(s) = \sup_{t \in [t_0, t_{\text{sim}}]} Y(s,t)$
- Denoting the dispersion model by f: $Y(s,t) = f\big(X(s,t)\big)$
- Probability that the concentration at s is higher than the concentration threshold ζ: $p_X(s) = \Pr(Y(s) > \zeta)$
- The variable $Z(s) = I_{\{Y(s) > \zeta\}}$ follows a Bernoulli distribution of parameter p_X: $Z \sim \mathcal{B}(p)$.
- Let's consider n independent and identically distributed random variables $Z_i(s)$ following a Bernoulli distribution of parameter p_X: $(s) = \sum_{i=1}^{n} Z_i\,(s) \sim \mathcal{B}\big(n, p_X(s)\big)$
- We used the sample mean estimator: $\hat{P}_n(s) = S_n / n$

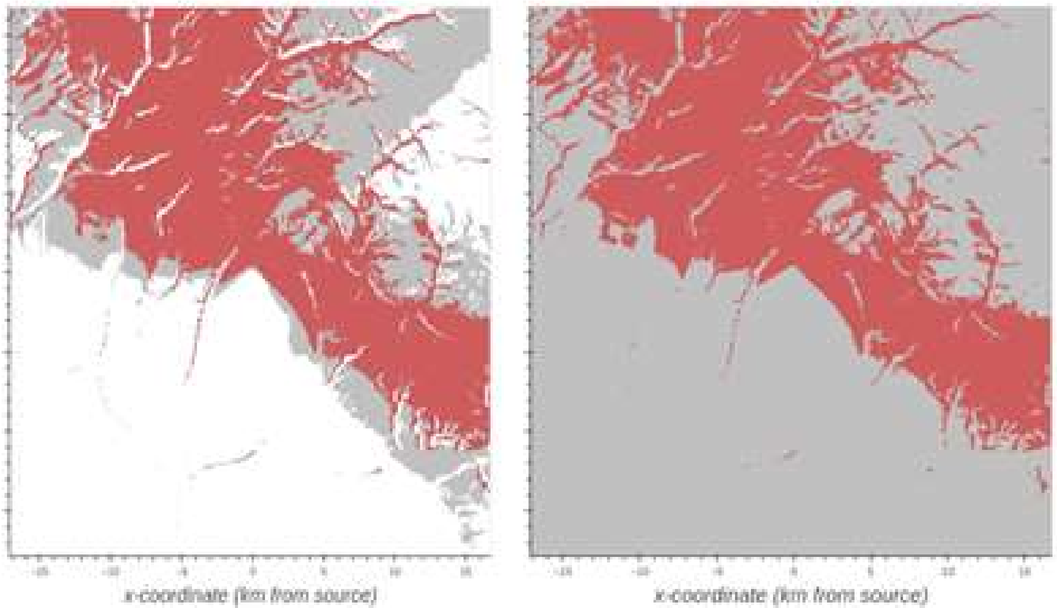


Probability of exceeding the concentration threshold with the sample mean : colors are from 0 (yellow) to 1 (black)

Unfortunately, a decision based on a point estimator does not inform us about the estimation uncertainty, contrary to confidence intervals setting lower $L_X(S_n(s), \alpha)$ and upper bounds $U_X(S_n(s), \alpha)$ to $p_X(s)$.

- Let $I_X(n, s, \alpha)$ be an interval which contains $p_X(s)$ with a confidence level of $1 - \alpha$ with α the risk accepted.
- The actual coverage probability at a fixed value p_X is an estimate of the probability that the interval actually contains p_X. Here, we focused on mean correct intervals, as they are narrower than conservative ones.
- For $n \geq 40$, Brown et al. (2001) recommend the adjusted Wald interval or Add 4 (Agresti and Coull, 1998).

- **We focus on $\{p_X(s) > p_{lim}\}$ to make decision!**
- A confidence interval is divided into three zones:
  1) In the red area where $L_X > p_{lim}$, there is strong evidence that the critical level is exceeded.
  2) In the white area where $U_X < p_{lim}$, there is strong evidence that we are below the critical level.
  3) In the grey area where $L_X < p_{lim} < U_X$, it is not possible to compare p_X and p_{lim}.
- **The left and right decision maps on the figure illustrate the loss of significativity: when p_{lim} goes under a certain value, the map becomes useless to the decider!**



Decision maps accounting for confidence intervals computed with the same parameters except p_{lim} going from 5% (left) to 4% (right)

INTRODUCING SPATIAL CORRELATION IN A BAYESIAN APPROACH

In the previous model, the sample $Z_1(s)\dots Z_n(s)$ was supposed to be independent for every location $s \in \mathbb{R}^2$.
However, the spatial structure of the data may help improving the estimation in nearby points $s + h$ of point s.

- Let's consider Bayesian statistics and the hierarchical model inspired from Diggle and Ribeiro (2007) which assumes the conditional independence of $S_n(s)$ which is the number of times the concentration threshold is exceeded at location s :

$$S_n(s) \mid P_X(s) = \sum_{i=1}^{n} Z_i(s) \mid P_X(s) \sim \mathcal{B}\big(n, p_X(s)\big)$$

$$logit\big(P_X(s)\big) \mid \beta, \tau, \lambda \sim \text{GaussianProcess}\left(X^T(s)\,\beta, \gamma(h) = \tau\, exp\left(-\frac{\|h\|}{\lambda}\right)\right)$$

With $X(s)$ the design matrix and β, τ and λ the mean, variance and scale parameters.

- A isotropic spatial Gaussian process is a stochastic process of which the joint distribution is multivariate normal for every set of positions ; it is completely defined by its mean function and its covariance function.

$$SGP_X = \{SGP_X(s_1) \dots SGP_X(s_K)\} = \big\{logit\big(P_X(s_1)\big) \dots logit\big(P_X(s_K)\big)\big\} \sim \mathcal{N}_K(\mu, \Sigma)$$

with $\mu = \big(\mu(s_j)\big)_{j=1}^{K} = \big(X^T(s_j)\beta\big)_{j=1}^{K} = X^T\beta$ and $\Sigma_{ij} = \gamma(s_i, s_j) = \tau exp\left(-\frac{\|s_i - s_j\|}{\lambda}\right)$.

- $X(s) = [1 \quad X^{(1)}(s) \quad X^{(2)}(s) \quad X^{(3)}(s)]^T$ contains explanatory variables (x and y coordinates, distance to source term).

- Prior distributions encode our initial knowledge about the parameters of the model:

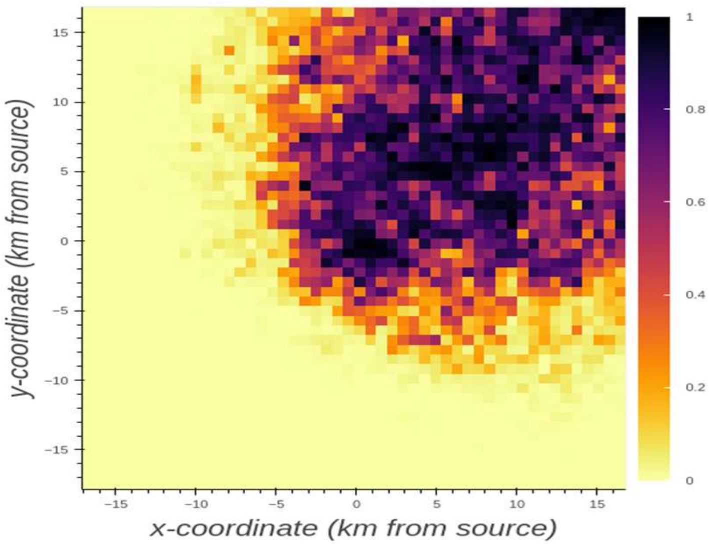
$$\forall i \in [\![1,4]\!]\ \beta_i \sim \mathcal{N}\left(\mu_{\beta_i}, \sigma_{\beta_i}^2\right) \qquad \tau \in \mathbb{R}^+, \tau \sim InvGamma(\delta_\tau, \phi_\tau) \qquad \lambda > 0, \lambda \sim \Gamma(k_\lambda, \theta_\lambda)$$

- To build the posterior distribution of $P(s_j)$ from n observations of $Z(s_j)$, we use Monte Carlo Markov chains for β_i, τ, λ and $SGP_X(s_j)$ whose stationary distribution corresponds to the posterior distribution in Bayesian statistics.

RESULTS OF THE BAYESIAN APPROACH (1/3)

Test on synthetic data and validation of the spatial Gaussian process

- We generated a simulation SGP_X^{sim} of the spatial Gaussian process $SGP_X \sim \mathcal{N}\left(X^T\beta^{true}, \Sigma(\tau^{true}, \lambda^{true})\right)$ using "true" values of the parameters ($\beta^{true} = [-8\,, 0.2\,, 0.2\,, -0.3]$, $\tau^{true} = 1$, $\lambda^{true} = 1$ and $n = 100$).
- The number of times a threshold is exceeded at s is $S_n^{\text{sim}}(s) \sim \mathcal{B}\left(n, P_X^{\text{sim}}(s) = expit\left(SGP_X^{\text{sim}}(s)\right)\right)$.



Target probability map generated for testing the Bayesian algorithm

- In the next step, we considered uninformative or low-informative prior distributions of β_i, τ and λ and initialized β_i between -1 and 1, τ at 1 and λ at 1 (in a real case, expert knowledge would be used).
- Markov chains were output for 10,000 iterations, removing the burning period (2,000 terms) and one term out of two to reduce the temporal dependence of the chain.
  - As the Markov chains look like a Gaussian noise, it shows that the MCMC algorithm has good mixing properties.
  - All the chains are centered on their retrieved true parameters:

$$\beta_{mean} = [-7.95, 0.20, 0.21, -0.31], \tau_{mean} = 1.06, \lambda_{mean} = 1.03$$

Performances of the Bayesian approach

While the Bayesian approach slightly improves point estimation, its main interest lies in interval estimation.

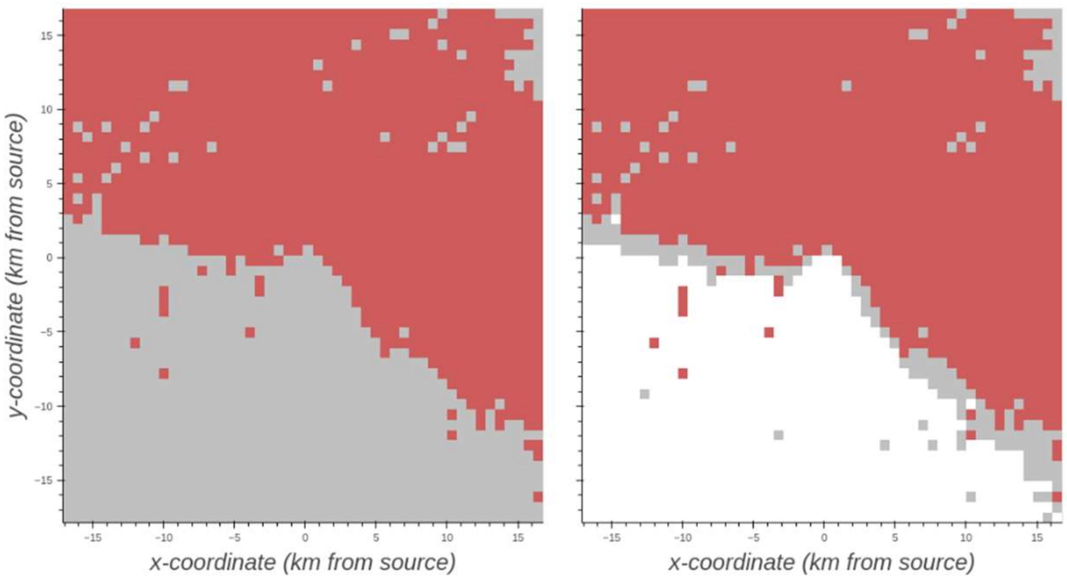
- We generated 1,000 maps of size 10 x 10 and drew $n = 100$ realizations of $\mathcal{B}\big(P(s)\big)$; from this sample, we estimated the uncertainty on $P_X(s)$ by building Bayesian and Add 4 confidence intervals for each location.
- Then, we computed the average coverage probability by assessing how many intervals contained the actual value $P_X(s)$ among the 100,000 different locations.
- Bayesian intervals have smaller average coverage probability and thus are slightly less conservative, but they achieve smaller expected widths (20% on average) than those computed with the Add 4 method.

Interval	Average coverage probability	Expected width
Bayesian at 95%	94.3% $\pm$ (8.2% x 10^{-2})	(1.29 x 10^{-1}) $\pm$ (5.9 x 10^{-4})
Bayesian at 99%	98.3% $\pm$ (4.4% x 10^{-2})	(1.64 x 10^{-1}) $\pm$ (7.5 x 10^{-4})
Add 4 at 95%	95.6% $\pm$ (6.6% x 10^{-2})	(1.55 x 10^{-1}) $\pm$ (5.0 x 10^{-4})
Add 4 at 99%	99.2% $\pm$ (2.9% x 10^{-2})	(2.04 x 10^{-1}) $\pm$ (6.6 x 10^{-4})

Average coverage probability and expected width of the Bayesian and the Add 4 intervals

Bayesian intervals as part of a decision-making

- Bayesian intervals bring two improvements when drawing decision maps as those for the Lubrizol case study
  1) It reduces the width of the gray zone of a significant amount.
  2) It counteracts the loss of significativity by preventing the grey zone to spread when considering small probability threshold or small risk, what is the main interest of our Bayesian estimator.
- Unlike the Add 4 interval, the Bayesian estimator is able to produce maps actually usable by decision-makers.
- Unfortunately, the MCMC algorithm is very time-consuming and the main limitation in the implementation of the Bayesian hierarchical model in comparison with the Add 4 method.



Decision maps obtained with the Add4 interval (left) and the Bayesian estimator (right) for the Lubrizol data set

CONCLUSION

- In this work, the classical frequentist approach and a novel Bayesian approach were compared for building decision maps from interval estimation of the probability of exceeding a concentration threshold.
  1) While confidence intervals are associated with a controllable nominal risk and useful to construct decision maps, they have a limit of significance when the sample size, the accepted risk and/or the probability threshold are small.
  2) The Bayesian model based on spatial Gaussian processes encodes the spatial dependence of the probabilities of exceeding a concentration threshold between nearby points in the probabilistic model ; it is more accurate and narrower than the Add 4 confidence intervals, and able to lower the significance limit of the estimate.

- The Bayesian approach was validated on synthetic data, then used in a case study inspired by a real accident.

- In the future, we plan to (1) reduce the computational time of the Bayesian model by running iterations of the MCMC algorithm simultaneously and making points independent after a certain distance, and (2) develop a user-friendly tool that could help better grasp the concept of estimation uncertainty on decision maps.

- Thus, the Bayesian approach will have the potential to provide information of real help on the confidence level of concentration (or exposure) maps in view of decision-making in an emergency involving atmospheric releases ; we argue that it is a good way to encourage the use of 3D modelling.

REFERENCES

Agresti, A. and B.A. Coull, 1998: Approximate is better than "exact" for interval estimation of binomial proportions. The American Statistician, 52 (2), 119–126.

Armand, P., O. Oldrini, C. Duchenne and S. Perdriel, 2021: Topical 3D modelling and simulation of air dispersion hazards as a new paradigm to support emergency preparedness and response. Environmental Modelling & Software, 143, 105–129.

Brown, T.T., L.D.Cai and A. Das Gupta, 2001: Interval estimation for a binomial proportion. Statistical Science, 16 (2), 101–117.

Geman, S. and D. Geman, 1984: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE PAMI-6.

Girard, S., P. Armand, C. Duchenne and T. Yalamas, 2020: Stochastic perturbations and dimension reduction for modelling uncertainty of atmospheric dispersion simulations. Atmospheric Environment, 117313.

Hanna, S. and J. Chang, 2012: Acceptance criteria for urban dispersion model evaluation. Meteorology and Atmospheric Physics, 116, 133–146.

Ismert, M. and M. Durif. 2014: Accident de Lubrizol du 21 janvier 2013. Couplage entre dispersion du nuage odorant et plaintes et appréciation des risques sanitaires associés. Rapport INERIS-DRC-13-137709-03375 B.

Oldrini, O., P. Armand, C. Duchenne and S. Perdriel., 2019: Parallelization performances of PMSS flow and dispersion modelling system over a huge urban area. Atmosphere, 10 (7), 404–420.

Oldrini, O., P. Armand, C. Duchenne, C. Olry, J. Moussafir and G. Tinarelli, 2017: Description and preliminary validation of the PMSS fast response parallel atmospheric flow and dispersion solver in complex built-up areas. Environmental Fluid Mechanics, 17 (5), 997–1014.

Tinarelli, G., L. Mortarini, S. Trini Castelli, G. Carlino, J. Moussafir, C. Olry, P. Armand and D. Anfossi, 2013: Description and preliminary validation of the PMSS fast response parallel atmospheric flow and dispersion Solver in complex built-up areas, American Geophysical Union (AGU), No. 200 (A), 311–327.

Skamarock, W.C., J.B. Klemp, J. Dudhia, D.O. Gill, D.M. Barker, W. Wang and J.G. Powers, 2005: A description of the advanced research WRF version 2. National Center For Atmospheric Research, Boulder (CO), USA.

Trini Castelli, S., P. Armand, G. Tinarelli, C. Duchenne and M. Nibart, 2018: Validation of a Lagrangian particle dispersion model with wind tunnel and field experiments in urban environment. Atmospheric Environment, 193, 273–289.

Thank you for your attention.

Questions ?

Corresponding author: patrick.armand@cea.fr