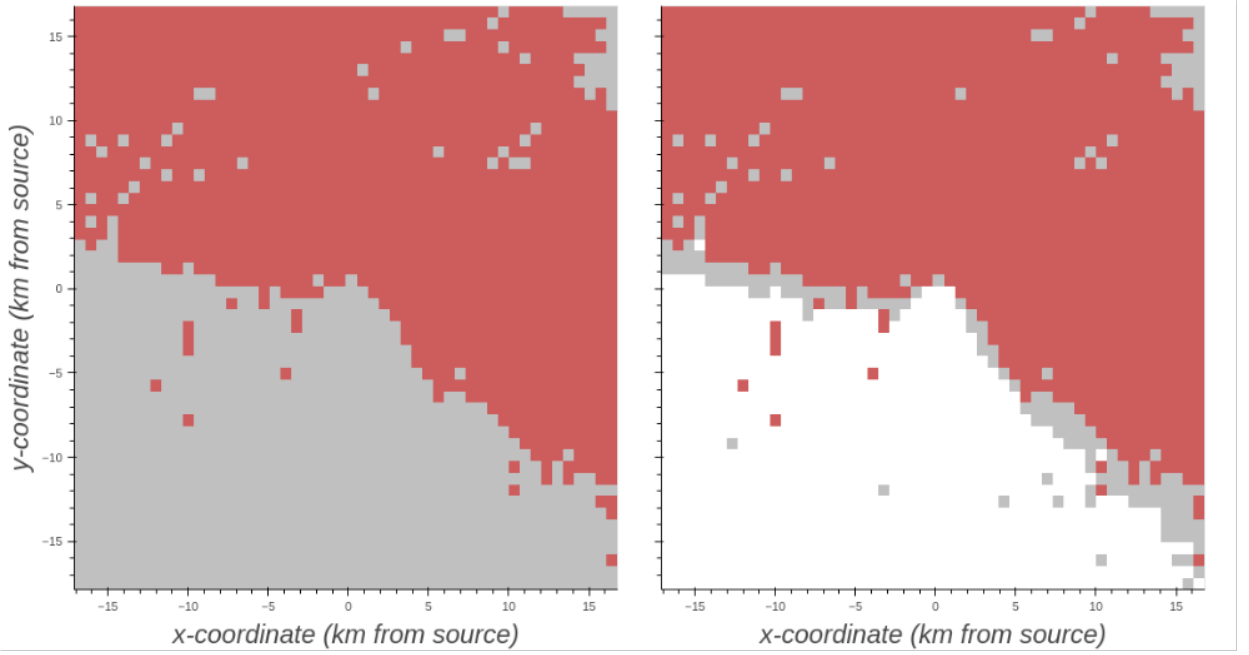
Graphical Abstract

Adaptive probabilistic modeling to support decision-making in the event of accidental atmospheric releases.

Maéva Caillat,Valentin Pibernus,Sylvain Girard,Mathieu Ribatet,Patrick Armand,Christophe Duchenne



Graphical abstract : Comparison of decision maps obtained with the Add4 frequentist interval (left) and the Bayesian estimators (right). The decision maps have three zones, red zone with a strong evidence that action is required, white zone with a strong evidence that action is not necessary, and grey zone where decision is not possible without exceeding the chosen risk. We show that the Bayesian estimator is robust against loss of significativity while the uncertain grey area covers the decision map of the frequentist estimator.

Highlights

Adaptive probabilistic modeling to support decision-making in the event of accidental atmospheric releases.

Maéva Caillat,Valentin Pibernus,Sylvain Girard,Mathieu Ribatet,Patrick Armand,Christophe Duchenne

- Assessing uncertainty for atmospheric dispersion is essential for decision making.
- Spatially independent confidence intervals have a limit of significance for low probability areas.
- Bayesian hierarchical model encode the spatial dependence of the probabilities of exceeding a concentration threshold.
- A spatially dependent Bayesian model lower the significance limit of the exceedance probability estimate.

Adaptive probabilistic modeling to support decision-making in the event of accidental atmospheric releases.

Maéva Caillat[a], Valentin Pibernus[a,*,1], Sylvain Girard[a], Mathieu Ribatet[b], Patrick Armand[c] and Christophe Duchenne[c]

[a]Phimeca Engineering, 18, Boulevard de Reuilly, Paris, 75012, France
[b]Ecole Centrale de Nantes, Département Informatique et Mathématiques, 1, rue de la Noë, Nantes, 44300, France
[c]CEA, DAM, DIF, F-91297, Arpajon, France

ARTICLE INFO

Keywords:
Atmospheric dispersion
Decision support system
Uncertainty estimation
Spatial Gaussian process
Bayesian hierarchical modelling
Markov chain Monte Carlo

ABSTRACT

In the event of accidental or malevolent atmospheric releases, decision-makers have to swiftly implement mitigating measures. Decisions are often based on the determination of danger zones and safe zones in which the concentration levels of substances emitted into the air are respectively above or below a given hazardous threshold. However, the maps representing the danger zones are established from atmospheric dispersion models whose input data on meteorology and the source term are uncertain. In addition, these maps are drawn from a limited number of simulations of atmospheric dispersion. Thus, if we consider confidence or credible intervals on low probabilities of exceeding concentration threshold, the "grey zone" in which no decision is possible can extend considerably. In this paper, we deal with this issue by developing a methodology to accurately estimate the probability of exceeding a concentration threshold of a substance adversely released in the atmosphere. Confidence or credible intervals associated with the probability of exceeding a given concentration are determined by taking into account the spatial correlation of the concentration field modelled by Gaussian processes. This methodology proves its effectiveness in lowering the significance limit of the probability estimates and allows for a more accurate estimate associated with a lower risk, especially in low probability areas. Moreover, it is applicable to various situations in terms of concentration threshold, accepted estimation risk and number of simulations. Finally, it appears promising for building maps of danger zone actually useful for decision-makers and will be implemented in a numerical decision-support tool following this work.

*Corresponding author
caillat.maeva@gmail.com (M. Caillat); pibernus@phimeca.com (V. Pibernus); girard@phimeca.com (S. Girard); mathieu.ribatet@ec-nantes.fr (M. Ribatet); patrick.armand@cea.fr (P. Armand); christophe.duchenne@cea.fr (C. Duchenne)
ORCID(s):
[1]phone : +33 6 37 75 93 45

1. Introduction

When hazardous materials are released into the atmosphere, first responders must act promptly and implement mitigation actions like sheltering or evacuating the population. To face an emergency, practitioners like firefighters need methods generally implemented in computational tools to help them evaluating the seriousness of the situation and to support decisions. In this paper, we present an approach to determine danger zones using scientifically rigorous decision criteria.

Indeed, a review of the literature in this field shows a need for the development of rigorous, reliable and robust tools for decision support. This development faces the difficulties of simulating transport and dispersion phenomena notably due to the intrinsic fluctuations of the atmosphere and the high level of uncertainty on both the models and their input data (Gunawardena et al., 2021). Quantitative probabilistic hazard mapping is used for mitigating the health risk related to the release of hazardous materials into the atmosphere in support to effective crisis response (Barsotti, 2020). However, some occurrence of underestimation can make this method insufficient in the case of atmospheric dispersion when one seeks to be conservative, limit exposure as much as possible and control the risk associated with decision making. As early as the 1950's, authors were interested in the propagation of uncertainty in atmospheric dispersion models (Strom and Ingram (1951), Gifford (1959)). Since then, numerous methods have been developed for accounting for the uncertainty inherent to those simulations. Restricting ourselves to recent years, Girard et al. (2020) present two contributions for accounting for the uncertainty inherent to those simulations. One possible approach to modelling those uncertainties is to use ensembles of input data, in particular ensembles of meteorological data. The other approach is to apply stochastic perturbations to the input conjecture.

The meteorological conjecture is usually obtained from simulations with meteorological models. Several authors suggested to apply uncertainty propagation to those models, or to use sets of different models to build ensembles of input meteorological conditions (Galmarini and Bianconi (2004), Warner et al. (2002)). However, as Dabberdt and Miller (2000) points out, it is rather unlikely that such an ensemble of conjectures might be available in a crisis context. Indeed, considerable efforts must be spent to calibrate such ensembles (Garaud and Mallet, 2011), (Zhang et al., 2012), and the calibration process requires reference data that are most often not available when dealing with accidental releases. Moreover, the calibration carried out at large scale may be inadequate to accidental situations at local scale. For instance, LE et al. (2021) make use of weather forecast ensembles to take into account meteorological uncertainties and propagate the input uncertainties through a dispersion model. While trying to cover all the possible range of variation of the uncertain variables, they do not manage to reproduce the air concentration measurements due to the coarse resolution of the weather simulations.

The alternative method chosen in this paper is to apply perturbations to the most influential parameters of the flow and dispersion models. Girard et al. (2016) study the relative influence of a set of uncertain inputs on several outputs

from a dispersion model approximated by Gaussian process emulation. While aggregated outputs are mainly driven by the amplitude of the release, they show that local outputs are extremely sensitive to wind perturbations. Leadbetter et al. (2020) examine the sources of uncertainty in a large number of numerical studies based on dispersion modelling both in the near field and in the far field. They conclude that the imperfect knowledge of the wind direction and release rate is highly influential, followed by the wind speed and release timing, and then precipitation. Consistently, these studies illustrate the necessity to consider the uncertainties associated with meteorology and release conditions. For this purpose, one can replace the commonly used deterministic map by a probabilistic risk map (Armand et al., 2014). However, the probabilistic risk map does not take into account the uncertainties on the probability of exceeding a threshold. Thus, a probabilistic risk map constructed on a small set of simulations, as it is the case in a crisis situation, is highly uncertain, especially at the decision boundary.

In our approach, we simulated the dispersion of pollutants in the atmosphere using a modeling system whose input data relating to the meteorological conditions and the source term, were 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). In this study, our objective was to estimate the probability of exceeding a concentration threshold that may represent a certain level of danger for human health or the environment. To make the theoretical considerations that we developed concrete, we considered a case study inspired by the industrial incident that occurred in January 2013, at the Lubrizol chemical plant located in Rouen, France. This builds up on our previous paper Girard et al. (2020) and the state of the art in which danger zones are determined from point estimation of the exceedance probabilities.

First, we estimated the probability of exceeding a concentration threshold using a deterministic approach. This technique does not take uncertainty into account. To overcome this, we then used a Monte Carlo approach with uncertain input data and point estimation. However, this method leads to a single value estimate of the probability of exceeding a concentration threshold . Thus, we tried to associate confidence and credible intervals to this probability. Even if this technique grasps uncertainty well, it has a limit of significance. To lower this limit, we took advantage of the spatial structure of the concentration results and finally used a credible interval with a conditional spatial independence criterion, which constitutes the final goal of the strategy presented in this paper.

The paper is organized as follows. In section 2, we describe the accidental situation, and the flow and dispersion models used to simulate it by conjecturing the input data. We thus produce a deterministic map of the area where a concentration threshold is exceeded corresponding to a potential hazard. Note that the method developed in the following is not adherent to the dispersion model. In section 3, we present our first approach to generate a probabilistic map by propagating uncertain input data and post-processing the results of the simulations to determine the fraction of exceedance of the concentration threshold at each point of the map. In section 4, we take into account the uncertainties

on our estimates by considering confidence intervals, namely Wald or Add4 mean correct intervals. This method reveals zones with a strong evidence that the concentration threshold respectively is or is not exceeded, but also a "grey zone" where decision is not possible. We also show that the combination of a small sample of simulations with low concentration threshold and/or low probability of exceedance and/or low accepted risk level leads to an extension of the grey zone making the probabilistic map of the danger area useless for decision-makers. Thus, it is the motivation to go further in the production of a probabilistic danger map. In section 5, we present a Bayesian hierarchical model based on Gaussian processes able to reproduce the spatial correlation underlying a 2D concentration field. After giving the definition of a spatial Gaussian process, we explain how the posterior distributions of its parameter may be built from their prior knowledge and Monte Carlo Markov Chains using the Metropolis-Hastings within Gibbs algorithm. In section 6, we validate the hierarchical model by creating concentration data from a realization of a Gaussian process, then reconstructing these synthetic data from non informative priors and Markov Chains converging to the values of the Gaussian process parameters. We also show that maps of the probability to exceed a concentration threshold are similar when they are drawn using the mean of realizations of the concentration field, the mean of realizations of the Gaussian processes (underlying the concentration fields), and the mean of Markov Chains simulating the Gaussian processes. Using the same synthetic data, we demonstrate that the Bayesian credible intervals determined with Gaussian processes are notably narrower than the Add4 intervals computed for the concentration threshold exceedance. Thus, the grey zone becomes smaller taking advantage of the spatial correlation. In section 7, we eventually apply the Bayesian hierarchical model to the danger map associated to the accidental situation described in section 2. Again, we show that the grey zone is strongly diminished when Gaussian processes are used to account for the spatial structure of the concentration field in comparison to simple point estimates. Adversely, we have to point out that our method is hampered by lengthy computational times. In section 8 we conclude on future developments intended to reduce these times and make our original approach operational. The figure 1 summarizes the techniques and methods discussed in the paper, and situates them in the context of the study. The innovation of our approach lies in the consideration of the risk associated with the danger maps.

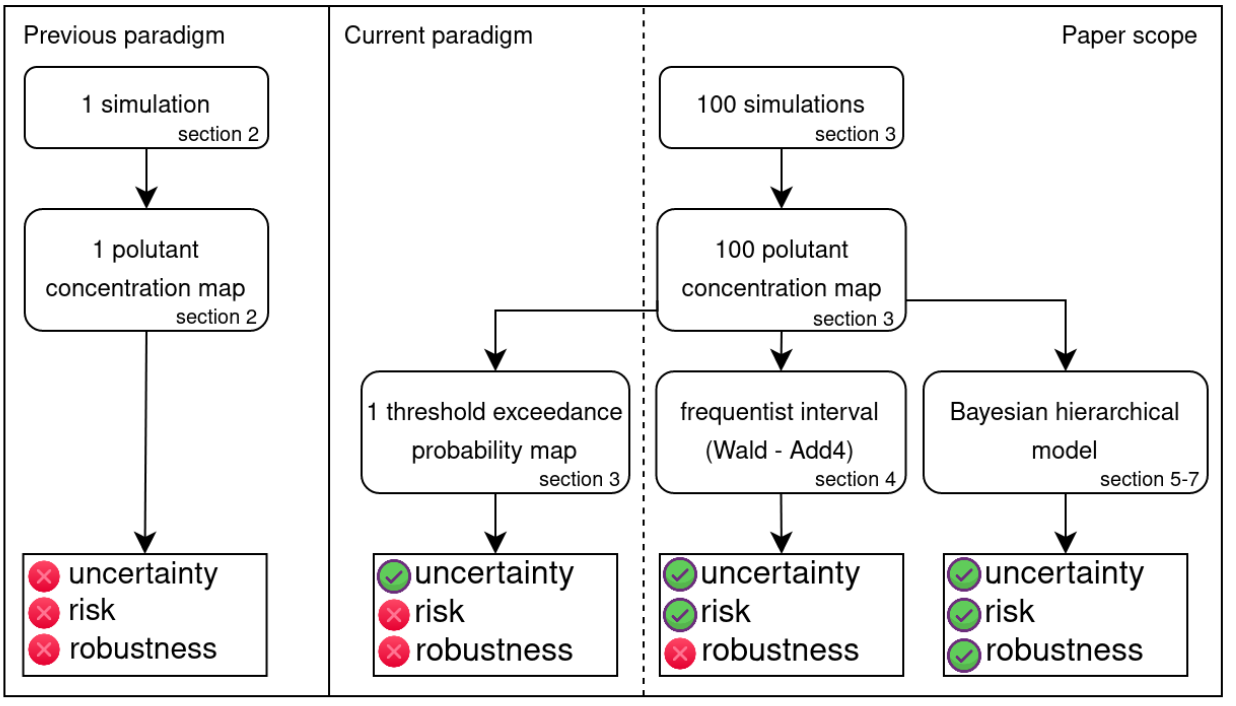


Figure 1: Graphical summary of the methods and concepts studied in the paper.

2. Deterministic simulation of an accidental situation

In this paper, we considered a real industrial accident that happened on January 2013, at the Lubrizol chemical plant (Rouen, France). Operational mistakes and system failures in the plant resulted in extended releases from the plant stacks of hydrogen sulfide and mercaptan, both of which are foul-smelling when they exceed a specified concentration. Consequently, thousands of people were able to smell the chemicals, some of them suffering from nausea and headaches. The features of the incident are very complex in several respects, namely the terrain characterized by rugged topography and alternating industrialized, urban and natural areas, the long-lasting and variable kinetics of the releases, and the highly variable meteorological conditions during and after the releases. Therefore, we selected this incident to link the theoretical considerations developed in our paper to a real and very challenging situation. Nevertheless, we make no claim to indicate how this situation should have been handled or how the same type of situation should be handled if it were to occur again. For purely practical reasons, we studied the chemical releases over a 35 kilometers wide square area, centered on the accident site. The mesh is 3D with a horizontal resolution of 2 meters and a vertical resolution of the same order close to the ground and decreasing with the height up to the top of

the domain at 1000 meters. The simulation covers a 35-hour period, so that all hazardous materials have either been deposited or have left the simulation domain by the end of the period.

The dispersion simulations were carried out using Parallel-Micro-SWIFT-SPRAY (PMSS). Originally, Micro-SWIFT-SPRAY (MSS) (Tinarelli et al., 2013) was developed to provide a simplified but rigorous CFD solution of the flow and dispersion in built-up environments in a limited amount of time. MSS is composed of the high resolution local scale versions of the SWIFT and SPRAY models:

- SWIFT is a 3D diagnostic and mass-consistent model using a terrain-following coordinate. Large scale meteorological data, local meteorological measurements, and analytical formulae in building-modified flow areas are interpolated and adjusted to generate 3D wind fields. Other meteorological data such as temperature or humidity are also interpolated. Eventually, the turbulent flow parameters are computed by SWIFT to be used by SPRAY.

- SPRAY is a Lagrangian particle dispersion model (LPDM) 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 fictitious particles. Trajectories are obtained by integrating in time the particle velocity which is the sum of a transport component defined by the local averaged wind generally provided by SWIFT, and a stochastic component, representing the dispersion due to atmospheric turbulence.

Both SWIFT and SPRAY can handle complex terrains and changing meteorological conditions, as well as specific release features, such as heavy or light gases. More recently, SWIFT and SPRAY were parallelized across time, space, and numerical particles, resulting in the PMSS system (Oldrini et al., 2017). The parallelism was shown to be very efficient, both on a multi-core laptop and on clusters of several hundreds or thousands of cores in the case of a high-performance computing center (Oldrini et al., 2019) (Armand et al., 2021). PMSS was systematically validated over numerous experimental wind tunnel and field campaigns for both short and extended releases (Castelli et al., 2018). In all the investigated configurations, PMSS results comply with the statistical acceptance criteria defined by Hanna and Chang (Hanna and Chang, 2012) commonly used for validating dispersion models in built-up environments.

The deterministic simulation was based on a conjectured input data based on the best of the knowledge about the situation. The conjectured rate of chemical emissions (also called source term) was adapted from data established by Ismert and Durif (Ismert and Durif, 2014). The conjectured rain and wind fields were obtained from the community reconstruction weather and forecast meso-scale modelling system WRF (Skamarock et al., 2005). Our objective was to predict whether an arbitrary concentration threshold of a chemical, namely $2\mu g.cm^{-3}$, was exceeded on the studied area and time period. It is worth noticing that this value is below the olfactory detection limits of the chemicals involved in the incident in order to create a sufficiently large area where the population and first responders could be hypothetically

at risk. One can observe the results of this simulation on figure 2. Even if this model enabled us to make some predictions on the concentrations, it did not take uncertainty on the input data into account. Yet, these inputs are substantially uncertain, and the results are too.

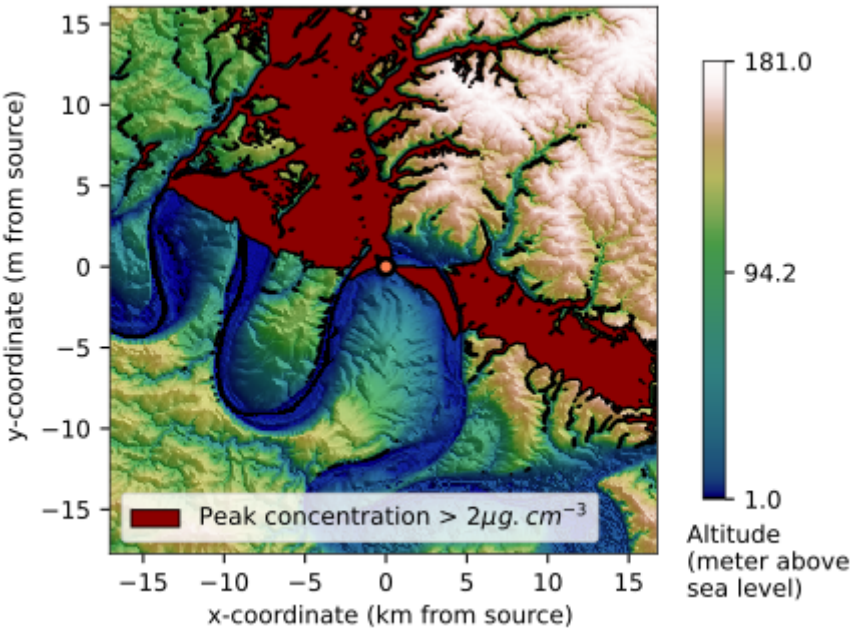


Figure 2: Deterministic map showing the area where the chemical concentration threshold is exceeded

3. Probabilistic simulations of the accidental situation using uncertain input data

Next, we accounted for uncertainty in the input parameters focusing on the source term and on the meteorological conditions known as mainly influencing the dispersion of the chemicals, namely the wind velocity, the wind direction, and the rain intensity (Stockie, 2011). We carried out 100 simulations using PMSS with different sets of input parameters. This sample has a quite small size due to the duration of each simulation of the order of one hour. It is representative of the number of simulations that could be performed in a real emergency, even though it is worth noting that simulations can be run in parallel on suitable computing resources. The sets of input parameters are generated by perturbing the wind velocity, wind direction, and rain intensity as presented in Girard et al. (2020).

Despite the aforesaid deterministic presentation of the flow and dispersion model, uncertainty can be taken into account in the modeling system. Let $Y(s,t)$ be the term of propagation of uncertainty in space and time. In this case, it represents the concentration of the chemical in $\mu g.cm^{-3}$ at a given point in space and time. In this paper, we assume that Y is continuous over $\mathbb{R}^3$. Symptoms induced by exposure to the chemical are triggered when the concentration near the ground reaches given values. We therefore focused on the maximum temporal concentration at each location

throughout the simulation: $Y(s) = \sup_{t\in[t_0,t_{simu}]} Y(s,t)$, with t_0 and t_{simu} being respectively 2013-01-21 8:00 am and 2013-01-22 7:00 pm. Let us notice that another interesting output variable would be the maximum of the time derivative of concentration as human olfaction is more sensitive to changes than to absolute values of concentration.

Let X represent the real random vector of uncertainties. For time t and point s, $X(s,t)$ can contain any kind of information, such as continuous variables like spatial coordinates or distance to the source term, or time-dependent variables like source term, rain intensity, wind speed, wind direction. Let f be a measurable function called numerical model of atmospheric dispersion. One may want to assess the distribution of $Y(s) = \sup_t f(X(s,t))$. Let $\zeta \in \mathbb{R}^+$ be a fixed concentration threshold and $Z(s) = I_{\{Y(s)>\zeta\}}$ be a variable worth 1 if the concentration exceeds the concentration threshold at a point s, and 0 otherwise. Let Z follows a Bernoulli distribution of parameter p_X: $Z \sim \mathcal{B}(p_X)$. $p_X(s) = \Pr(Y(s) > \zeta)$ represents the probability that the concentration at s is higher than the concentration threshold. Let $p_{lim} \in [0,1]$ be the threshold of the probability of exceeding a concentration. We focus on the event $\{p_X(s) > p_{lim}\}$, which enables us to make a decision at point s.

Let's consider the sampling model $(\mathcal{B}(p_X))_{p_X\in\Theta=[0,1]}$. Let $Z_1(s)$, :., $Z_n(s)$ be a n-sample of $Z(s)$, that is to say n independent and identically distributed (i.i.d.) random variables that follow a Bernoulli distribution of parameter $p_X(s)$. In our application, this sample is constituted by the 100 simulations of flow and dispersion around the Lubrizol site using 100 different sets of parameters.

Let $S_n(s) = \sum_{i=1}^n Z_i(s) \sim \mathcal{B}(n, p_X(s))$ be the number of times the concentration at s exceeds the threshold. $\mathcal{B}(n, p_X(s))$ is a binomial distribution with parameters n and $p_X(s)$.

In a previous work Girard et al. (2020), we estimated $p_X(s)$ with the sample mean estimator: $\hat{P}_n(s) = \frac{S_n}{n}$. The expected value of this estimator is $\mathbb{E}(\hat{P}_n(s)) = p_X(s)$, and its variance is $\mathbb{V}(\hat{P}_n(s)) = \frac{p_X(s)(1-p_X(s))}{n}$. The resulting estimated probabilities are shown in the figure 3.

4. Introduction of confidence and credible intervals in the danger map

Until now, the danger map built with the sample mean estimator ignored the level of confidence in the prediction. An area where the probability of exceeding the threshold $p_X(s)$ is lower than the threshold p_{lim} by 0.1% is therefore announced as "safe". A decision based on the sample mean does not take the estimation uncertainty into account, contrary to credible and confidence intervals. In order to introduce the level of confidence at each point of the danger map, we decided to present the results with credible or confidence intervals. The difference between these intervals is that credible intervals account for the actual observed sample, here dispersion simulations, which is not the case for confidence intervals. More information on the difference between credible and confidence intervals are available in (Lu et al., 2012). These intervals set limits to $p_X(s)$ thanks to two estimators, the lower bound of the interval $L_X(S_n(s),\alpha)$ and its upper bound $U_X(S_n(s),\alpha)$. Let $I_X(n,s,\alpha) = [L_X(S_n(s),\alpha), U_X(S_n(s),\alpha)]$ be a confidence or a credible interval

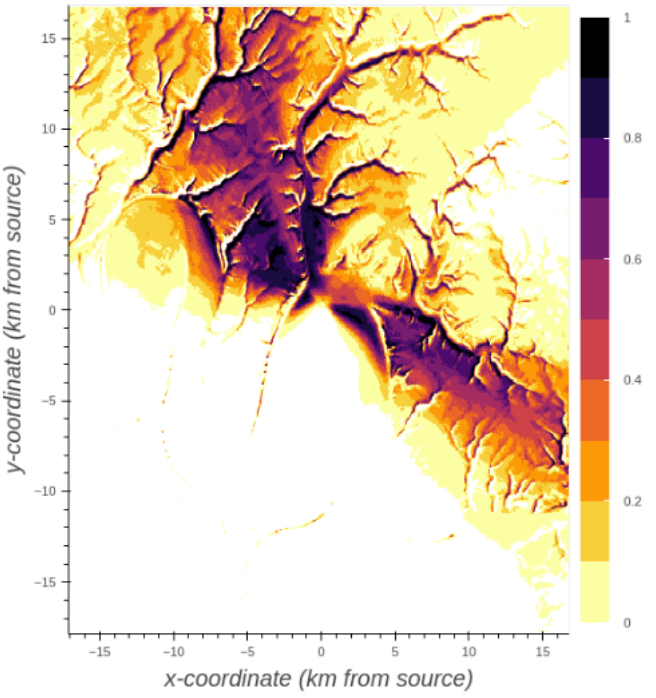


Figure 3: Probabilistic map estimated with the sample mean showing the area where the chemical concentration threshold is exceeded. Color indicates probability from 0 (white) to 1 (black).

which contains $p_X(s)$ with a confidence level of $1-\alpha$, with α the risk accepted to be as low as possible. Estimating a binomial proportion thanks to an interval is thoroughly studied, and many credible and confidence intervals have been suggested in the literature. Before deciding between them, let us introduce a few definitions first.

The actual coverage probability at a fixed value of p_X is an estimate of the probability that an interval actually contains p_X. The nominal coverage probability is the nominal confidence level $1-\alpha$. The nominal coverage probability is the "target" coverage probability: the one we try to attain when we derive a method providing a confidence interval. The actual coverage is the "true" coverage. The expected width is the expectation of the length of an interval. It is a good measure of the estimation performance for a given interval with an actual coverage probability of $1-\alpha$. Indeed, in an emergency, we want to minimize uncertainty, that is to say to have the lowest expected width possible for a given risk.

Two categories of intervals are opposed: strictly conservative intervals and mean correct intervals. A conservative interval has its actual coverage probability greater or equal to the nominal confidence level. Conventionally, the Clopper-Pearson interval (Clopper and Pearson, 1934) is considered to be the best exact confidence interval to use when strict conservativeness is mandatory. Conversely, mean correct intervals have a mean coverage probability of at least $1-\alpha$, but their actual coverage probability can be lower. The actual coverage probability is notably lower than

the confidence level for extreme values of p_X, close to 0 or 1. For our problem, we chose to only focus on mean correct intervals, as they are narrower than conservative ones.

Historically, one of the first confidence intervals introduced was the Wald interval (Laplace, 1812). It is still widely used, despite its poor performance in terms of coverage probabilities. Based on the comparative studies of Vollset (1993), Newcombe (1998), Brown et al. (2001) and Pires and Amado (2008), we only retained two mean correct intervals. For $n \leq 40$, Brown et al. (2001) recommends to use the score interval (Wilson, 1927), also called Wilson's interval. For $n \geq 40$, Brown et al. (2001) recommends to employ the adjusted Wald interval, also called Add 4 (Agresti and Coull, 1998) what we did in our application.

Figure 4 shows example of decisions maps accounting for confidence or credible intervals. The map is divided into three zones, each one associated with a different decision-making strategy:

- In the red zone, the lower bound of the interval is greater than p_{lim}. There is strong evidence that the concentration threshold is exceeded at each location s. More formally, p_X is asymptotically larger than p_{lim} with a confidence level of $1 - \alpha$. Thus, an action is expected.

- In the white zone, the upper bound of the interval is lower than p_{lim}. There is strong evidence that the concentration threshold is not exceeded at any of the location s. More formally, p_X is asymptotically smaller than p_{lim} with a confidence level of $1 - \alpha$. Thus, an action is not necessary.

- In the grey zone, p_{lim} is enclosed by the interval bounds. p_X and p_{lim} cannot be easily compared and the result is non significant.

The left and right decision maps of figure 4 are computed with the same data and parameters, except for the probability threshold p_{lim}. It illustrates the loss of significativity phenomenon: when p_{lim} goes under a certain value (dependent on the sample size and the confidence level), the grey zone tends to occupy all the map which becomes useless to the decider.

The upper bound $U_X(S_n(s), \alpha)$ of the confidence interval depends on the number n of samples. For instance, for a given sample size n, an accepted risk α of 5%, and no realization of concentration exceedance among the n simulations ($S_n(s) = 0$), the upper bound of the interval is: $U_X(0, 5\%) = \frac{1}{n+4}\left(2 + z_{0,975}\sqrt{\frac{2(n+2)}{n+4}}\right)$. Details to obtain this result are given in Appendix 9.1. Figure 5 shows the upper bound of the Add 4 interval (used in our application) versus the sample size for this example. Whenever the probability threshold p_{lim} is under the upper bound, we face a loss of significativity: with all non-red zones colored in grey. In the following of our paper, we present a strategy which aims at avoiding the loss of significativity and reducing the grey zone.

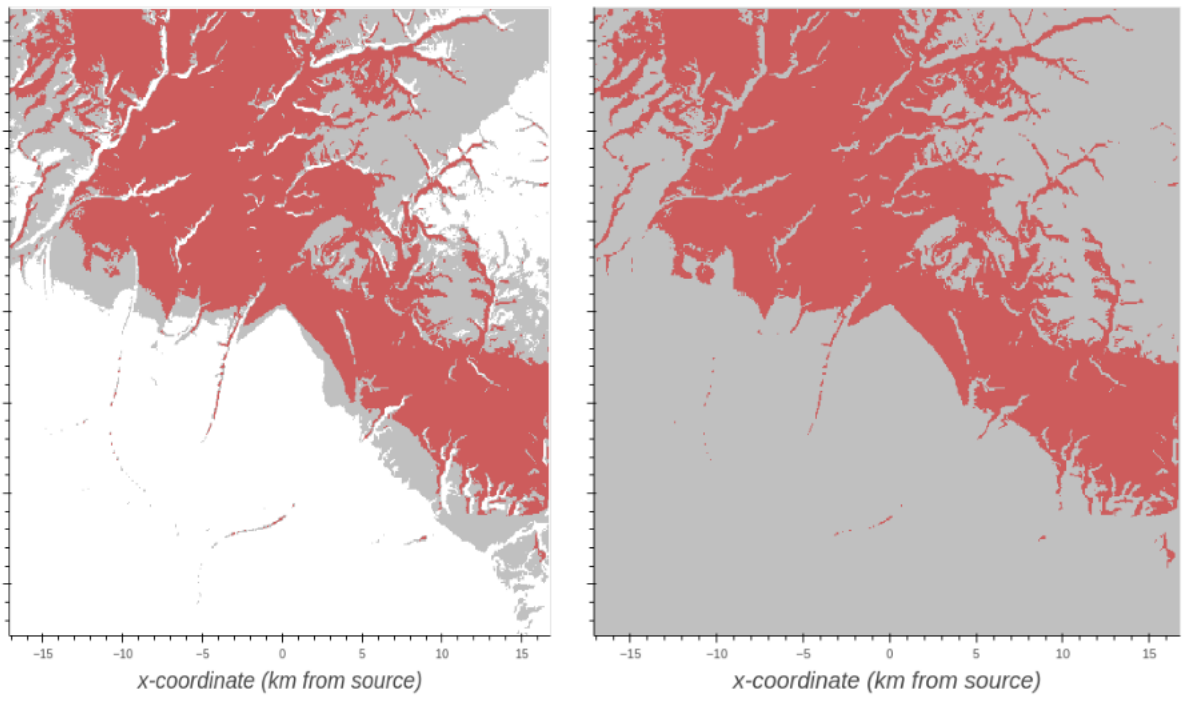


Figure 4: Two probabilistic maps accounting for confidence intervals of the area where the concentration threshold is exceeded. Both maps are computed with the same data and parameters, except for the probability threshold p_{lim} of 5% (on the left) and 4% (on the right).

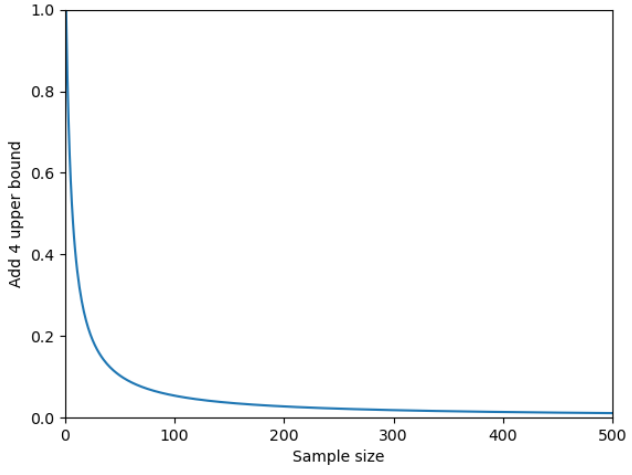


Figure 5: Upper bound of the Add 4 confidence interval with no realization of concentration exceedance ($S_n = 0$) and and accepted risk $\alpha = 5\%$ as a function of the sample size n.

5. Accounting for spatial correlation in a Bayesian framework

5.1. Definition of a hierarchical model based on a Gaussian process

In our previous model, the n-sample $Z_1(s), ..., Z_n(s)$ of $Z(s)$ was supposed to be independent for every location $s \in \mathbb{R}^2$, and the spatial structure of the data was not taken into account. However, atmospheric concentration data have a continuously varying response in space. Therefore, the information gathered at location s may help improving the estimations in nearby points $s + h$, with h low.

From now on, we place ourselves in the framework of Bayesian statistics and the parameter p_X becomes the random variable P_X. Let's assume the conditional independence of $S_n(s)$ (the number of times the concentration threshold is exceeded at location s) and consider the following hierarchical model, inspired from Diggle and Ribeiro (2007) with $SGP_X(s)$ denoting a spatial Gaussian process:

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

With $logit(P_X(s)) \mid \beta, \tau, \lambda \sim SGP_X(s)$. The logit function $x \to ln(\frac{x}{1-x})$ for $x \in (0, 1)$, sends the set $[0, 1]$ to $\mathbb{R}$. Thus, to go from $P_X(s)$ to $SGP_X(s)$, one needs to apply the logit function to $P_X(s)$.

A spatial Gaussian process $SGP_X(s) : s \in \mathbb{R}^2$ is a stochastic process of which the joint distribution $SGP_X = \{SGP_X(s_1), \dots, SGP_X(s_K)\}$ is multivariate normal for every set of positions $s_1, \dots, s_K$ with $s_j \in \mathbb{R}^2$. Any such process is completely defined by its mean function $\mu(s) = \mathbb{E}[SGP_X(s)]$, and its covariance function $\gamma(s, s') = \text{cov}(SGP_X(s), SGP_X(s'))$.

For a set of positions $s_1, \dots, s_K$, the spatial Gaussian process SGP_X reads:

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

$\mu = (\mu(s_j))_{j=1}^{K} = (X^T(s_j)\beta)_{j=1}^{K} = X^T\beta$ is a K-dimensional mean vector with $X(s)$ the design matrix and β a mean parameter. The design matrix gathers all the information (coordinates, distance to the source term, terrain topology) specific to position s, which could be correlated to the probability of exceeding the threshold. For example, the design matrix may be $X(s) = [1, X_1(s), X_2(s), X_3(s)]^T$ with the y-coordinate $X_1(s)$, the x-coordinate $X_2(s)$ and the distance to the source term $X_3(s)$ as explanatory variables. Σ such that $\Sigma_{ij} = \gamma(s_i, s_j) = \tau exp(-\frac{||s_i - s_j||}{\lambda})$ is a KxK-dimensional matrix with τ a variance parameter and λ a scale parameter. This spatial Gaussian process is supposed to be isotropic, that is to say that the covariance $\gamma(s_i, s_j)$ depends only on the distance between the points s_i and s_j.

The covariance function $\gamma : \mathbb{R}^2 \times \mathbb{R}^2 \to [0, \tau]$ is such that:

$$\forall (s_i, s_j) \in \mathbb{R}^2 \times \mathbb{R}^2 \; \gamma(s_i, s_j) = \tau exp(-\frac{||s_i - s_j||}{\lambda})$$

The covariance function decreases exponentially as the Euclidean distance between s_i and s_j increases, and is equal to τ if $s_i = s_j$. τ is the variance of $SGP_X(s)$ for any location $s \in \mathbb{R}^2$. For $||s_i - s_j|| = 3\lambda$, $\gamma(s_i, s_j) = \text{cov}(SGP_X(s_i), SGP_X(s_j)) \leq 0.05\tau$. The covariance is negligible with respect to τ at a distance of more than 3λ.

5.2. Prior and posterior distributions of the Gaussian process parameters

Prior distributions encode our initial knowledge about the parameters of the Gaussian process. The parameters $(\beta_i)_{i \in [\![1;4]\!]}$ can take any value in $\mathbb{R}$, so we chose a normal distribution a priori: $\forall i \in [\![1;4]\!] \; \beta_i \sim \mathcal{N}(\mu_{\beta_i}, \sigma^2_{\beta_i})$. The parameter τ is a variance, so we typically chose an inverse gamma distribution because it is defined on $\mathbb{R}^+$ and induces conjugate distributions: $\tau \sim \text{InvGamma}(\delta_\tau, \phi_\tau)$. The parameter λ must be strictly positive, so we chose a gamma distribution a priori: $\lambda \sim \Gamma(k_\lambda, \theta_\lambda)$.

The corresponding directed acyclic graph (DAG) is given in figure 6. A DAG is a graphical model that represents a hierarchical dependence structure, i.e. $\forall i \in \mathcal{V}$, with $\mathcal{V}$ being the vertices of the graph, W_i and its non descendants are conditionally independent given the parents of W_i, W_i being the random variables of the problem. For example, on figure 6, one can see that $\forall i \in [\![1;n]\!] \forall j \in [\![1;K]\!] Z_i(s_j) \sim \mathcal{B}(P(s_j))$.

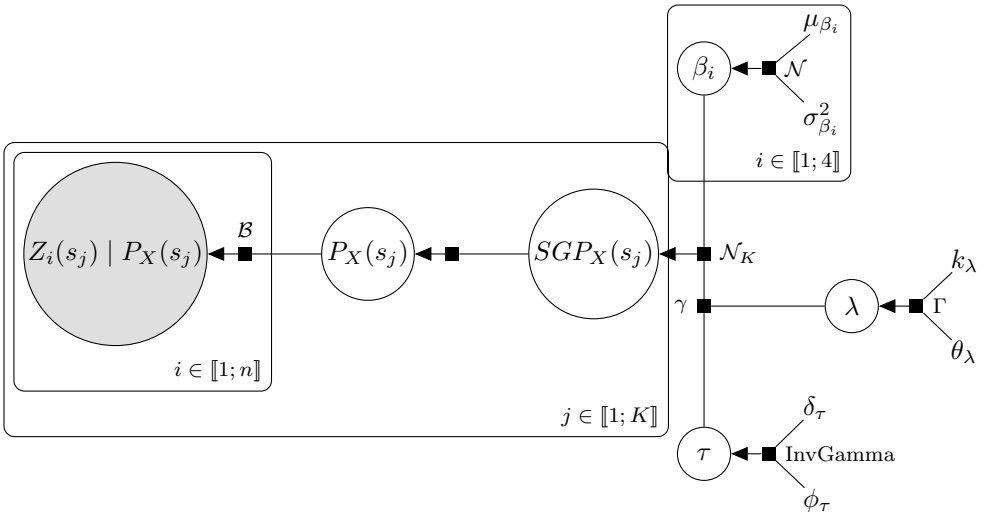


Figure 6: Directed Acyclic Graph (DAG) of the model. We use the standard symbols.

To build the posterior distribution of $\{P(s_j) : j \in [\![1;K]\!]\}$ from n observations of $Z(s_j)$, we used the class of algorithms known as Markov chain Monte Carlo methods (MCMC). MCMC techniques are a set of sampling methods that produce a Markov chain whose stationary distribution is the posterior distribution in Bayesian statistics.

The posterior distributions are presented in Appendix 9.2. As it is not possible to access directly the distributions of β_i, λ and $SGP_X(s)$, we used a Metropolis (Metropolis et al., 1953) - Hastings (Hastings, 1970) algorithm within a Gibbs sampler (Geman and Geman, 1984) to evaluate the posterior distributions. The proposal kernel for the MCMC algorithm are given in Appendix 9.3 Finally, we built a Markov chain for β, τ, λ and $SGP_X(s_j)$, that is to say a chain of dimension $lenbeta$ (the dimension of β) + 2 + K (the number of points of the map).

6. Validation of the Gaussian process model

6.1. Generation of data

First, we tested our method on simulated data representing concentrations. Let β^{true}, λ^{true} and τ^{true} be the real parameters that we look for. We first generated one realization SGP_X^{simu} of the spatial Gaussian process $SGP_X \sim \mathcal{N}(X^T\beta^{true}, \Sigma(\tau^{true}, \lambda^{true}))$. SGP_X is a vector of dimension K, which we can represent as an exceedance probability map of dimension $\sqrt{K}\times\sqrt{K}$, by applying to it the expit function (also called logistic sigmoid function) $x \mapsto expit(x) = sigmoid(x) = \frac{exp(x)}{1+\exp(x)}$ for $x \in \mathbb{R}$, which sends the set $\mathbb{R}$ to the set $[0, 1]$. The expit function is the inverse of the logit function. We used this realization to simulate $S_n^{\text{simu}}(s) \sim \mathcal{B}(n, P_X^{\text{simu}}(s) = expit(SGP_X^{\text{simu}}(s))$.

We set the parameters as follows: $\beta^{true} = [-8, 0.2, 0.2, -0.3]$, $\tau^{true} = 1$, $\lambda^{true} = 1$, and $n = 100$. We could have set them to other values; these remain just fictive values for assessing the model. Figure 7 shows a realization of this process, that we used as the target for the probability estimation algorithm.

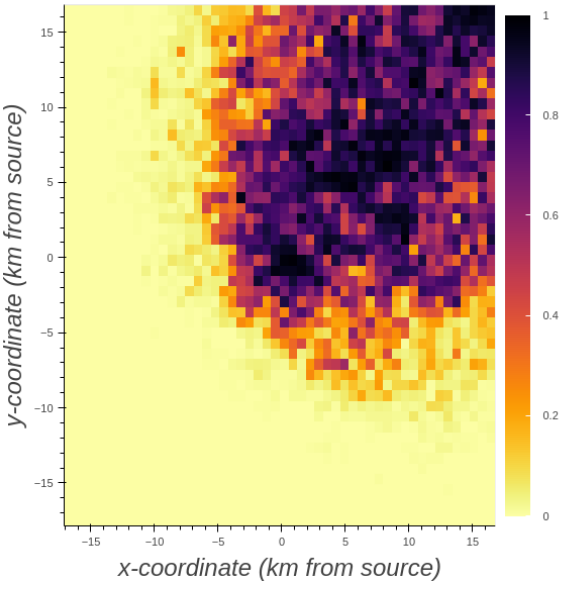


Figure 7: Target probability map generated for testing the estimation algorithm.

6.2. Reconstruction of the synthetic data

We assume that we have no or little prior knowledge about the data to reconstruct. Thus, we consider uninformative or low-informative prior distributions as follows:

$\forall i \in [\![1; 4]\!]\ \beta_i \sim \mathcal{N}(\mu_{\beta_i} = 0, \sigma^2_{\beta_i} = 100)$, $\tau \sim \text{InvGamma}(\delta_\tau = 1, \phi_\tau = 1)$, $\lambda \sim \Gamma(k_\lambda = 2, \theta_\lambda = 0.5)$.

β_i is randomly initialized between -1 and 1. We arbitrarily initialize τ at 1 because it is the order of magnitude of the parameters that interest us: $\tau \leq 0.1$ is almost trivial and $\tau \geq 10$ renders very unstructured, white noise type

maps. We also arbitrarily initialize λ at 1 so that it adapts to the size of the map. Expert knowledge could allow a better initialization on real data. Since we suspect that the true value of $S_n^{\text{simu}}(s)$ is around the sample mean, we initialize $SGP_X(s)$ at its sample mean. We fix the variances of the proposal kernels so as to have an acceptance rate close to 0.234: $\sigma_{\text{prop}_\beta} = [0.25, 0.01, 0.01, 0.01]^T$, $\sigma_{\text{prop}_\tau} = 0.13$, $\sigma_{\text{prop}_\lambda} = 0.18$ and $\sigma_{\text{prop}_{SGP_X(s_j)}} = 2.53$.

Figure 8 represents the output Markov chain of the MCMC algorithm for 10,000 iterations. The first 2,000 terms, notably the burning period, were removed from it, as well as one term out of two to reduce the temporal dependence of the chain. Figure 8 shows that this MCMC algorithm has good mixing properties, since the output Markov chain looks like a Gaussian noise. All the chains are centered on their true parameters: $\beta_{mean} = [-7.95, 0.20, 0.21, -0.31]$, $\tau_{mean} = 1.06$, and $\lambda_{mean} = 1.03$. One can observe that τ and λ are highly correlated, as expected. This validates the reconstruction of our synthetic data representing concentrations in the atmosphere.

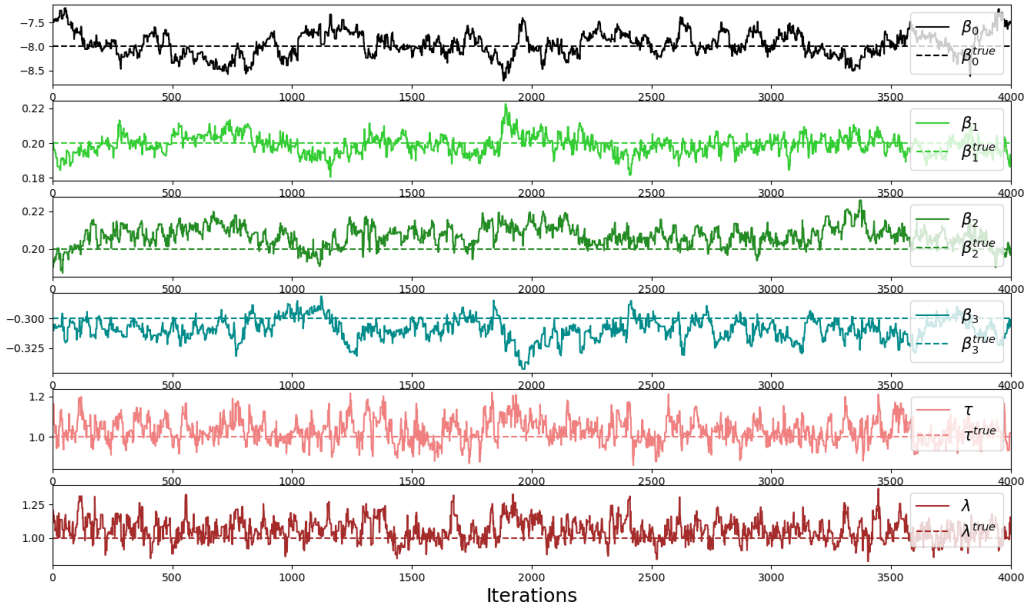


Figure 8: Markov chains of the parameters β, τ and λ for $N = 10,000$ iterations, $n = 100$ observations, $K = 2,500$ points, $\beta^{true} = [-8, 0.2, 0.2, -0.3]$, $\tau^{true} = 1$, and $\lambda^{true} = 1$

6.3. Comparison of point estimates ($P_X(s_j)$)

$SGP_X(s)$ Markov chain simulates the distribution of $SGP_X(s_j)$. So, we estimated $P_X(s_j)$ using the expit of the mean of the Markov chain of the distribution of $SGP_X(s_j) = logit(P_X(s_j)\ \forall j \in [1; k]$ built with MCMC methods. From now on, we refer to this estimator as the Bayesian estimator for simplicity.

Figure 9 compares point estimations with the Bayesian and sample mean estimator to a target map. The differences are very tenuous between the target map, namely the sample mean of n realizations of a Bernoulli distribution of P_X^{simu}, the map obtained with the estimator $\frac{S_n(s_j)}{n}$ and the map obtained with the Bayesian estimator. The three probability maps (target, sample mean estimation, and Bayesian estimation) are very similar. The rightmost map shows the difference of mean absolute error (MAE), the average of the absolute difference between estimation and actual observation: $MAE = \frac{1}{n}\sum_{i=1}^{n} |\hat{P}_i^{estimed} - P_i^{observed}|$. Blue cells are the ones for which the Bayesian estimator performs betters, and the red cells are where the sample mean performs better. There is no clear pattern, namely no particular region where one estimator beats the other.

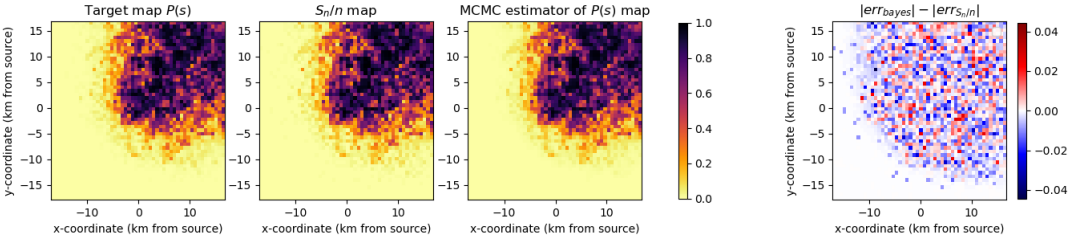


Figure 9: Probability maps comparing the sample mean and Bayesian point estimation. The differences are very tenuous. Left: the target map, namely the sample mean of n realizations of a Bernoulli distribution of P_X^{simu}. Middle left: map obtained with the sample mean estimator $\frac{S_n(s_j)}{n}$. Middle right: map obtained with the Bayesian estimator. Right: difference of the absolute errors between the Bayesian and sample mean estimators. The Bayesian (resp. sample mean) estimator performs better in blue (resp. red) cells.

Numerically the Bayesian estimator performs slightly better overall: the MAE of the Bayesian estimator, equal to 0.0154, is below the MAE of $\frac{S_n(s)}{n}$, equal to 0.0167. Although small, this difference is found on the whole of the tests performed.

6.4. Empirical validation based on the coverage probability

We wish now to assess the uncertainty on the estimation of $P_X(s_j)$ by building either Add 4 or Bayesian credible intervals. In the Bayesian approach, we use the distribution of $SGP_X(s_j)$ to determine for instance the 95% credible interval by considering the 2.5% and 97.5% quantiles of the distribution.

To compare our Bayesian interval with that of Add 4, we computed the mean coverage probability for the risks $\alpha = 0.01$ and $\alpha = 0.05$. We generated 1,000 maps of size 10 x 10 with their associated random parameters β, τ and λ for a total of 1,000 x 10 x 10 = 100,000 different locations. Keeping in mind that we want to estimate the uncertainty on $P_X(s)$ for each of these locations, we drew $n = 100$ realizations of $\mathcal{B}(P(s))$. From this sample, we estimated the corresponding Bayesian and Add4 intervals for each location. We could then compute the average coverage probability by assessing how many Bayesian and Add4 intervals contained the actual value $P_X(s)$ among the 100,000 different locations.

Table 1
Comparison table of the average coverage probability and the expected width of the Bayesian interval and the Add 4 interval.

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

Table 1 compares the performances of the Bayesian and Add 4 intervals, computed with the test case illustrated in figure 7. With smaller average coverage probability, Bayesian intervals are slightly less conservative but they achieve significantly smaller expected widths (20% on average) than those computed with the Add 4 method, which make them an attractive choice. We obtained similar results in all our other experiments (not shown here).

6.5. Improvement of significativity using the Gaussian process

While the Bayesian approach slightly improves point estimation, its main interest lies in interval estimation as we show hereafter. More precisely, we consider the grey zones on the probabilistic maps obtained by using alternatively Add 4 and Bayesian intervals.

Figure 10 represents the decision maps drawn with the sample mean, the confidence interval Add 4, and our Bayesian hierarchical model, computed with the test case illustrated in figure 7. Let's recall that this case considers $n = 100$ simulations, $\alpha = 5\%$, $p_{lim} = 5\%$, $K = 2\,500$ points, $N_{\text{chain}} = 10\,000$ iterations, $\beta^{true} = [-8, 0.2, 0.2, -0.3]$, $\tau^{true} = 1$, and $\lambda^{true} = 1$. The grey zone obtained from Bayesian intervals (right) is smaller than with Add 4 intervals (middle). It is reduced from "both sides", by the growth of both the red and white zone on the right map.

Figure 11 displays the size of the red, gray, and white areas of figure 10. By taking into account the spatial dependence of the data, the Bayesian model managed to reduce the uncertainty associated with the estimation in comparison with Add 4. Here the reduction comes mostly from coloring gray cells into white. However, one can notice that the converse was observed in other experiments that we conducted (not shown here).

7. Probabilistic maps for the real accident using the hierarchical model

Finally returning to the study of the accidental situation, the figure 12 represents the decision maps drawn with the Add 4 confidence interval and our Bayesian hierarchical model for the Lubrizol data set. This case considers $n = 100$ simulations, $\alpha = 5\%$, $p_{lim} = 4\%$, $K = 2\,500$ points and $N_{chain} = 10\,000$. It clearly illustrates the capacity of the Bayesian estimator to overcome loss of significativity exemplified in figure 4.

Thus, Bayesian intervals bring two improvements when it comes to drawing decision maps:

- It reduces the width of the grey area of a small but significant amount.

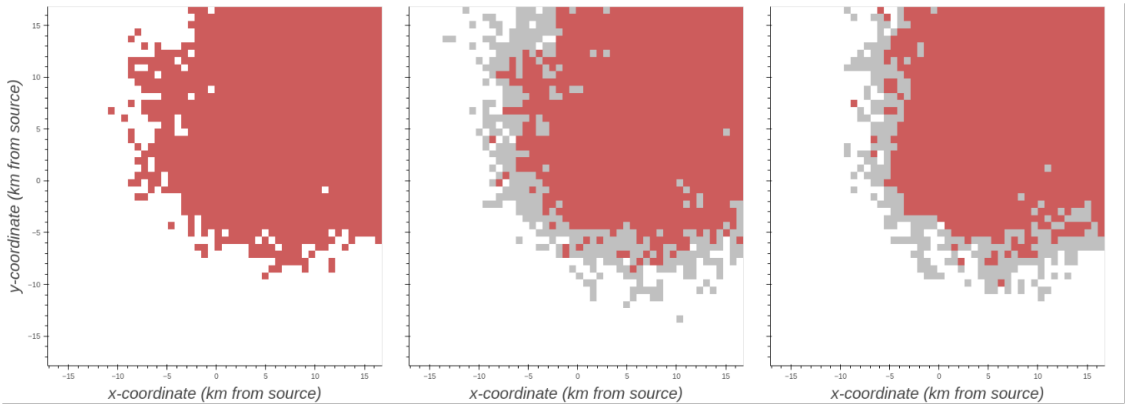


Figure 10: Probabilistic maps for the synthetic data. On the left: point estimate; in the middle: use of Add 4 intervals; on the right: use of Bayesian intervals. $p_{lim} = 5\%$

Figure 11: Histogram of the size of the three probabilistic areas corresponding to the synthetic data.

- Above all, it counteracts the loss of significativity phenomenon, namely it prevents the grey zone to spread everywhere when considering small probability threshold or small risk. This is probably the main interest of our Bayesian estimator, and is indeed illustrated above with the Lubrizol test case

The Bayesian model produces a map that can be used by decision-makers, unlike the Add 4 interval.

Sometimes, we cannot approximate a map using the design matrix with a mean parameter. For instance, one can imagine scenario with two source terms, where the mean parameter does not model the situation well, as only the distance to one source term would be taken into account. According to Diggle and Ribeiro (2007), even in

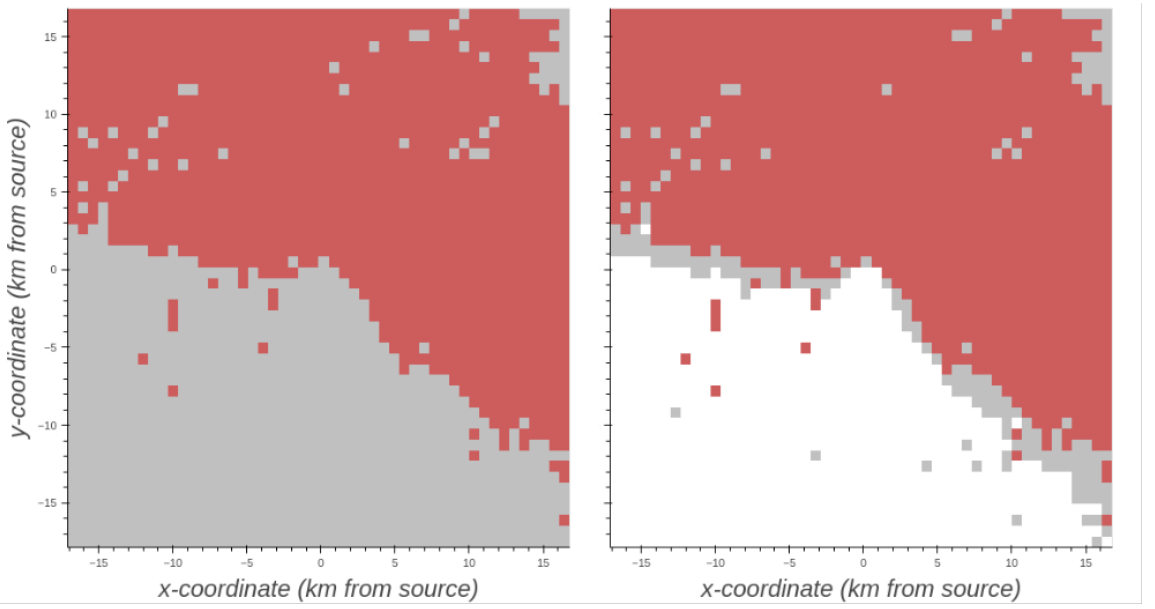


Figure 12: Comparison of decision maps obtained with the Add4 interval (on the left) and the Bayesian estimators (on the right). The Bayesian estimator is robust against loss of significativity ($p_{lim} = 4\%$).

this situation, if the data are spatially dependent, this Bayesian framework manages to grasp the spatial dependence through the variance and scale parameters, and to give a reliable estimation of the map. In this way, in any case, our Bayesian framework gives more accurate estimations of the probability of exceeding a concentration threshold than the confidence intervals studied earlier.

7.1. Limit of the Bayesian model

Table 2 represents the computation time for different number of locations K, for 10,000 MCMC iterations. We used the Python language, version 3.9.12, and a computer with an Intel Core i7-10810U Processor, 8 GB of memory, and a processor speed between 1.1GHz and 4.9GHz. Figure 13 represents the log-scaled computation time depending on different numbers of locations for the confidence interval Add 4 and the credible interval computed with the MCMC algorithm. Table 2 and figure 13 both show that our MCMC algorithm is very time-consuming, especially when compared to the Add 4 interval. In this way, the main limitation in the implementation of our Bayesian hierarchical model is the computation time.

In order to overcome the computational time barrier and reduce the "grey area", we can consider two options. We can increase the number of simulations in the sample by using existing models. The width of the frequentist interval being of the order of magnitude of $\frac{1}{\sqrt{n}}$, doubling the number of simulations would allow to divide the width by $\sqrt{2}$. In this way, the grey area of the Add4 interval will be reduced. We can also try to reduce the computation time of the

Table 2
Computation time table depending on the number of locations $K \times K$ for $N_{\text{chain}} = 10,000$ iterations.

Number of locations	MCMC computation time	Add 4 computation time
25	9s	0.01s
100	1min 10s	0.04s
225	3min 11s	0.05s
400	8min 6s	0.07s
900	28min 7s	0.11s
1600	2h 36min 58s	0.15s
2500	12h 48min 01s	0.22s

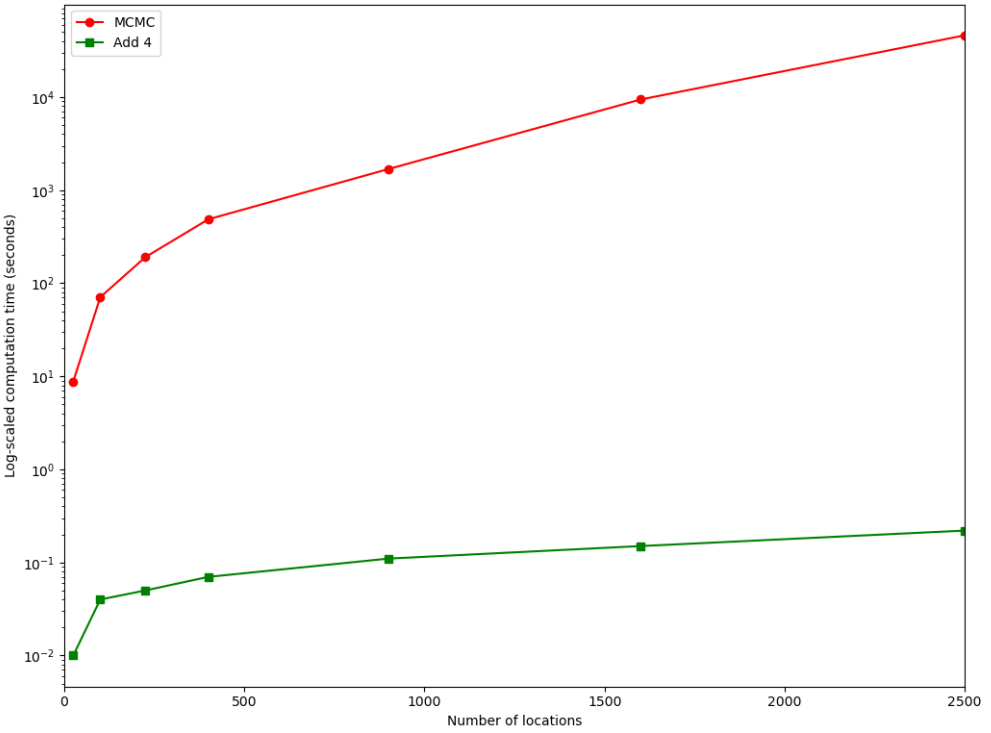


Figure 13: Log-scaled computation time versus number of locations for Add 4 and the MCMC framework

Bayesian intervals. These two options can probably be carried out together, but the choice of one or the other method is decisive and could be the subject of a further study not addressed here.

8. Conclusion

The objective of this work was to better evaluate the danger zones that could be generated in an accidental situation implying the atmospheric dispersion of hazardous materials. To deal with this issue, we have developed methods to generate probabilistic maps of the area where a concentration threshold associated with adverse consequences is exceeded. These maps display a red zone with strong evidence of exceeding the concentration threshold, a white zone with strong evidence of not exceeding the concentration threshold, and in between a grey zone where nothing can be claimed. In the red zone, decisions would certainly be taken like, for example, sheltering or evacuating the population. Contrarily, in the white zone, no action would certainly be taken. The grey zone is the most difficult to take decision. While caution is required, action would be recommended. We presented and compared different strategies for building probabilistic danger maps from interval estimation of the probability of exceeding a concentration threshold: the frequentist approach and a Bayesian framework. We conducted extensive tests with synthetic data, and illustrated the results with a case study inspired by the Lubrizol accident that occurred in 2013 in Rouen (France).

First, we studied confidence intervals whose estimation is associated with a controllable nominal risk. We compared several mean correct intervals to estimate a binomial proportion. We stipulated this rule based on the paper of Brown and DasGupta (Brown et al., 2001): for a number of simulations below 40, we recommend using the Score interval; for a number of simulations above 40, we advise using the Add 4 interval (also called Agresti and Coull). Although these intervals allow the construction of very useful decision maps, they have a limit of significance. Indeed, when the sample size is small as are the concentration threshold, the probability of exceeding threshold and/or the accepted risk level, the probabilistic map may be of no use for decision-making.

To solve this problem, we encoded the spatial dependence of the probabilities of exceeding a concentration threshold between nearby points in a probabilistic model. We achieved this by implementing a Bayesian hierarchical model, based on spatial Gaussian processes. We compared the performance of the Bayesian model and the above-mentioned confidence intervals. We showed on simulated data that the Bayesian model was more accurate and led to narrower interval than the Add 4 interval. We also showed that the Bayesian model was able to significantly lower the significance limit of the estimate. In this way, it was able to draw informative probabilistic maps when frequentist interval was of no help to decision-makers. However, the computation time of the Bayesian model was considerably longer than Add 4, especially for an increasing number of points on maps. On the contrary, the computation time is not impacted by the number of simulations since the computation time of the sample mean is negligible compared to that of the Bayesian model.

In light of these results, in an emergency, we recommend plotting decision maps with a frequentist estimator. For a given concentration threshold, risk, number of simulations, and exceedance probability limit, if the uncertain zone of the probabilistic map becomes dominant over the zone of "certainty", the Bayesian model should be used to obtain

more accurate results. We also recommend, for a given concentration threshold and a number of simulations, to change the risk and probability threshold and see what lower bound on these parameters can be used to evaluate the probability of exceeding a concentration threshold.

To the best of our knowledge, our scientifically reliable method is the first attempt to provide information on the level of confidence attributed to concentration results in case of atmospheric releases. In the future, we plan to reduce the computational time of the Bayesian hierarchical model, for instance by doing a few steps of the MCMC algorithm simultaneously or by marking points independent after a certain distance, what is called Gaussian Markov random fields (Rue and Held, 2005). Indeed, currently, the evaluation of each complete conditional distribution requires the inversion of a K x K-dimensional covariance matrix. Thanks to Gaussian Markov random fields, we could ignore a large part of this matrix composed of 0, and only invert sub-matrices much smaller. This should speed up the computation and reduce the time of the Bayesian algorithm to that of the frequentist intervals. Later on, we also plan to implement our Bayesian method into an interactive and user-friendly tool that could help decision-makers better grasp the concept of estimation uncertainty on probabilistic danger maps and make this probabilistic tool understandable to users without expertise in statistics or physical modeling.

9. Appendices

9.1. Upper bound of the Add 4 interval

The confidence interval Add 4 (Agresti and Coull, 1998) is:

$$\hat{\pi}_n \pm z_{1-\alpha/2}\sqrt{\frac{\hat{\pi}_n(1-\hat{\pi}_n)}{n+4}},$$

with $\hat{\pi}_n = \frac{S_n+2}{n+4}$ and $z_{1-\alpha/2}$ the $1-\alpha/2$ Gaussian quantile.

For a given sample size n, $\alpha = 5\%$, and zero realization of concentration exceedance among the 100 simulations, the upper bound of Add 4 is:

$$U_X(0,\alpha) = \frac{2}{n+4} + z_{1-\alpha/2}\sqrt{\frac{\frac{2}{n+4}(1-\frac{2}{n+4})}{n+4}} = \frac{1}{n+4}\left(2 + z_{0,975}\sqrt{\frac{2(n+2)}{n+4}}\right).$$

9.2. Posterior distributions

Let $s_1, \dots, s_K \in \mathbb{X} \subseteq \mathbb{R}^2$ be a set of K locations. $P_X = (P_X(s_1), \dots, P_X(s_K))$, $S_n = (S_n(s_1), \dots, S_n(s_K))$, and $SGP_X = (SGP_X(s_1), \dots, SGP_X(s_K))$. $X(s_j) = (1, \text{covariables}(s_j))$ and $X = (X(s_1), \dots, X(s_K))$ is the design matrix. $\beta = (\beta_1, \dots, \beta_{lenbeta-1}) \in \mathbb{R}^{lenbeta}$, $\tau \in \mathbb{R}^+$ and $\lambda \in \mathbb{R}^+$ are the parameters of our model. β_{-i} is the vector β without the term β_i.

Let's consider $C_X = SGP_X - X^T\beta$. $|.|$ represents the determinant of a matrix. For example, $|\Sigma(\tau,\lambda)| = \det(\Sigma(\tau,\lambda))$. The implementation of the a posteriori distributions is done using a logarithmic scale. So we calculate the log-distributions below.

$$\pi(\beta_i \mid \beta_{-i}, \tau, \lambda, SGP_X) \propto \pi(SGP_X \mid \beta, \tau, \lambda) \times \pi(\beta_i) \text{ and } \pi(\beta_i) \sim \mathcal{N}(\mu_{\beta_i}, \sigma^2_{\beta_i}) \text{ a priori}$$

Thus, the posterior log-distribution of β_i is:

$$\log(\pi(\beta_i \mid \beta_{-i}, \tau, \lambda, SGP_X)) \propto -\frac{1}{2}{C_X}^T\Sigma^{-1}C_X - \frac{1}{2}(\frac{\beta_i - \mu_{\beta_i}}{\sigma_{\beta_i}})^2 \tag{1}$$

$$\pi(\tau \mid \beta, \lambda, SGP_X) \propto \pi(SGP_X \mid \beta, \tau, \lambda) \times \pi(\tau) \text{ and } \pi(\tau) \sim \text{Inv-Gamma}(\delta_\tau, \phi_\tau) \text{ a priori}$$

Thus, the posterior log-distribution of τ is:

$$\log(\pi(\tau \mid \beta, \lambda, SGP_X)) \propto (-K/2 - \delta_\tau - 1)\log(\tau) - \frac{1}{2\tau}(C_X^T\Sigma_\tau^{-1}C_X + 2\phi_\tau) \tag{2}$$

$$\pi(\lambda \mid \beta, \tau, SGP_X) \propto \pi(SGP_X \mid \beta, \tau, \lambda) \times \pi(\lambda) \text{ and } \pi(\lambda) \sim \text{Gamma}(k_\lambda, \theta_\lambda) \text{ a priori}$$

Thus, the posterior log-distribution of λ is:

$$\log(\pi(\lambda \mid \beta, \tau, SGP_X)) \propto -\frac{1}{2}\log(|\Sigma(\tau, \lambda)|) + (k_\lambda - 1)\log(\lambda) - \frac{\lambda}{\theta_\lambda} - \frac{1}{2}C_X^T \Sigma^{-1} C_X \tag{3}$$

$$\pi(SGP_X(s_j) \mid \beta, \tau, \lambda, SGP_X^{-j}, S_n(s_j)) \propto \Pr(S_n(s_j) \mid SGP_X(s_j)) \times \pi(SGP_X(s_j) \mid \beta, \tau, \lambda, SGP_X^{-j})$$

$$\text{and } \pi(SGP_X(s_j) \mid \beta, \tau, \lambda) \sim \mathcal{N}_K(X^T\beta, \Sigma) \text{ a priori}$$

Thus, the posterior log-distribution of $SGP_X(s_j)$ is:

$$\begin{aligned}\log(\pi(SGP_X(s_j) \mid \beta, \tau, \lambda, SGP_X^{-j}, S_n(s_j))) \propto\ & S_n(s_j)\log(\text{expit}(SGP_X(s_j))) \\ & + (n - S_n(s_j))\log(1 - \text{expit}(SGP_X(s_j))) \\ & - \frac{1}{2}C_X^T\Sigma(\tau, \lambda)^{-1}C_X\end{aligned} \tag{4}$$

9.3. Proposal kernels for the MCMC algorithm

Let Ker : $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ be a proposal kernel for the MCMC algorithm. We consider the following kernels for the parameters of the model:

- The kernel of the β_i is normally distributed, i.e. $\text{Ker}(\beta_i, .) \sim \mathcal{N}(\beta_i, \sigma_{prop_\beta})$. σ_{prop_β} is the proposal variance of Ker. It directly influences the acceptance rate of the proposal β_i'. The choice of σ_{prop} is made empirically, so as to have an acceptance rate close to 0.234. Indeed, Roberts, Gelman and Gilks (Roberts et al., 1997) stated that for high-dimensional target distributions formed of independent and identically distributed components, for Metropolis algorithms with Gaussian proposals, the asymptotically optimal acceptance rate is 0.234.

- For τ, we already know how to simulate according to an inverse gamma distribution, so there is no need for a proposal kernel.

- The parameter λ is defined on $\mathbb{R}^+$. We used a proposal kernel Log-$\mathcal{N}(log(\lambda), \sigma_{prop_\lambda})$. This kernel allows performing a log-scale random walk in Metropolis-Hastings.

- For the $SGP_X(s_j)$, we used the normal distribution as before.

CRediT authorship contribution statement

Maéva Caillat: Methodology - Investigation - Software - Formal analysis - Writing - Original Draft. **Valentin Pibernus:** Methodology - Investigation - Software - Formal analysis - Writing - Original Draft. **Sylvain Girard:** Supervision - Project administration - Funding acquisition - Writing - Review & Editing. **Mathieu Ribatet:** Conceptualization - Writing - Review & Editing. **Patrick Armand:** Methodology - Supervision - Writing - Review Editing. **Christophe Duchenne:** Methodology - Software.

References

N. Gunawardena, G. Pallotta, M. Simpson, and D. D. Lucas. Machine learning emulation of spatial deposition from a multi-physics ensemble of weather and atmospheric transport models. *Atmosphere*, 12(8), 2021. ISSN 2073-4433. doi: 10.3390/atmos12080953. URL https://www.mdpi.com/2073-4433/12/8/953.

Sara Barsotti. Probabilistic hazard maps for operational use: the case of so2 air pollution during the holuhraun eruption (bárðarbunga, iceland) in 2014–2015. *Bulletin of Volcanology*, 82(7):56, Jun 2020. ISSN 1432-0819. doi: 10.1007/s00445-020-01395-3. URL https://doi.org/10.1007/s00445-020-01395-3.

G. H. Ph.D. Strom and W. T. M.P.H. Ingram. The development of prototype model research on atmospheric pollution problems. *American Industrial Hygiene Association Quarterly*, 12(2):59–63, 1951. doi: 10.1080/00968205109344328. URL https://doi.org/10.1080/00968205109344328.

Frank Gifford. Statistical properties of a fluctuating plume dispersion model. volume 6 of *Advances in Geophysics*, pages 117–137. Elsevier, 1959. doi: https://doi.org/10.1016/S0065-2687(08)60099-0. URL https://www.sciencedirect.com/science/article/pii/S0065268708600990.

S. Girard, P. Armand, C. Duchenne, and T. Yalamas. Stochastic perturbations and dimension reduction for modelling uncertainty of atmospheric dispersion simulations. *Atmospheric Environment*, page 117313, 2020. doi: 10.1016/j.atmosenv.2020.117313. URL https://sylvaingirard.net/pdf/girard20-perturbation_aam.pdf.

S. Galmarini and R. Bianconi. Ensemble dispersion forecasting—part i: concept, approach and indicators. *Atmospheric Environment*, 38(28):4607–4617, 2004. ISSN 1352-2310. doi: https://doi.org/10.1016/j.atmosenv.2004.05.030. URL https://www.sciencedirect.com/science/article/pii/S1352231004004959.

T. T. Warner, R. Sheu, J. F. Bowers, R. I. Sykes, G. C. Dodd, and D. S. Henn. Ensemble simulations with coupled atmospheric dynamic and dispersion models: Illustrating uncertainties in dosage simulations. *Journal of Applied Meteorology and Climatology*, 41(5):488–504, 2002.

W. F Dabberdt and E. Miller. Uncertainty, ensembles and air quality dispersion modeling: applications and challenges. *Atmospheric Environment*, 34(27):4667–4673, 2000.

D. Garaud and V. Mallet. Automatic calibration of an ensemble for uncertainty estimation and probabilistic forecast: Application to air quality. *Journal of Geophysical Research: Atmospheres*, 116(D19), 2011.

Y. Zhang, M. Bocquet, V. Mallet, C. Seigneur, and A. Baklanov. Real-time air quality forecasting, part ii: State of the science, current research needs, and future prospects. *Atmospheric environment*, 60:656–676, 2012.

N. B. T. LE, I. Korsakissok, V. Mallet, R. Périllat, and A. Mathieu. Uncertainty study on atmospheric dispersion simulations using meteorological ensembles with a monte carlo approach, applied to the fukushima nuclear accident. *Atmospheric Environment: X*, 10:100112, 2021.

ISSN 2590-1621. doi: https://doi.org/10.1016/j.aeaoa.2021.100112. URL https://www.sciencedirect.com/science/article/pii/S2590162121000125.

S. Girard, V. Mallet, I. Korsakissok, and A. Mathieu. Emulation and sobol' sensitivity analysis of an atmospheric dispersion model applied to the fukushima nuclear accident. *Journal of Geophysical Research: Atmospheres*, 121(7):3484–3496, 2016. doi: https://doi.org/10.1002/2015JD023993. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015JD023993.

S.J. Leadbetter, S. Andronopoulos, P. Bedwell, K. Chevalier-Jabet, G. Geertsema, F. Gering, T. Hamburger, A.R. Jones, H. Klein, I. Korsakissok, A. Mathieu, T. Pázmándi, R. Périllat, Cs. Rudas, A. Sogachev, P. Szántó, J.M. Tomas, C. Twenhöfel, H. de Vries, and J. Wellings. Ranking uncertainties in atmospheric dispersion modelling following the accidental release of radioactive material. *Radioprotection*, 55:S51–S55, 2020. doi: 10.1051/radiopro/2020012. URL https://doi.org/10.1051/radiopro/2020012.

P. Armand, F. Brocheton, D. Poulet, F. Vendel, V. Dubourg, and T. Yalamas. Probabilistic safety analysis for urgent situations following the accidental release of a pollutant in the atmosphere. *Atmospheric Environment*, 96:1–10, 2014. ISSN 1352-2310. doi: https://doi.org/10.1016/j.atmosenv.2014.07.022. URL https://www.sciencedirect.com/science/article/pii/S1352231014005433.

G. Tinarelli, L. Mortarini, S. Trini Castelli, G. Carlino, J. Moussafir, C. Olry, P. Armand, and D. Anfossi. Description and preliminary validation of the pmss fast response parallel atmospheric flow and dispersion solver in complex built-up areas. *American Geophysical Union (AGU)*, (200 (A)):311–327, 2013.

O. Oldrini, P. Armand, C. Duchenne, C. Olry, J. Moussafir, and G. Tinarelli. 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, 2017.

O. Oldrini, P. Armand, C. Duchenne, and S. Perdriel. Parallelization performances of pmss flow and dispersion modeling system over a huge urban area. *Atmosphere*, 10(7), 2019. doi: 10.3390/atmos10070404.

P. Armand, O. Oldrini, C. Duchenne, and S. Perdriel. 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, 2021. doi: 10.1016/j.envsoft.2021.105129.

S. T. Castelli, P. Armand, G. Tinarelli, C. Duchenne, and M. Nibart. Validation of a lagrangian particle dispersion model with wind tunnel and field experiments in urban environment. *Atmospheric environment*, 193:273–289, 2018. doi: 10.1016/j.atmosenv.2018.08.045.

S. Hanna and J. Chang. Acceptance criteria for urban dispersion model evaluation. *Meteorology and Atmospheric Physics*, 116:133–146, 2012. doi: 10.1007/s00703-011-0177-1.

M. Ismert and M. Durif. Accident de Lubrizol du 21 janvier 2013 – couplage dispersion du nuage odorant/plaintes et appreciation des risques sanitaires associes. Technical Report INERIS-DRC-13-137709-03375B, Ineris, 2014.

W. C. Skamarock, J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers. A description of the advanced research wrf version 2. Technical report, National Center For Atmospheric Research Boulder Co Mesoscale and Microscale Meteorology Div, 2005.

John M. Stockie. The mathematics of atmospheric dispersion modeling. *SIAM Review*, 53(2):349–372, 2011. doi: 10.1137/10080991X. URL https://doi.org/10.1137/10080991X.

D. Lu, M. Ye, and M. C. Hill. Analysis of regression confidence intervals and bayesian credible intervals for uncertainty quantification. *Water Resources Research*, 48(9), 2012. doi: https://doi.org/10.1029/2011WR011289. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011WR011289.

C. J. Clopper and E. S. Pearson. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4):404–413, 1934.

P.S. Laplace. *Théorie Analytique des Probabilités*. Courcier, 1812.

S. E. Vollset. Confidence intervals for a binomial proportion. *Statistics in Medicine*, 12(9):809–824, 1993.

R. G. Newcombe. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in Medicine*, 17(8):857–872, 1998.

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

A. M. Pires and C. Amado. Interval estimators for a binomial proportion: Comparison of twenty methods. *Revstat*, 6(2):165–197, 2008.

E. B. Wilson. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 1927. doi: 10.1080/01621459.1927.10502953. URL https://doi.org/10.1080/01621459.1927.10502953.

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

P.J. Diggle and P. J. Jr. Ribeiro. *Model-based Geostatistics*. Springer, 2007.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, and A. H. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21, 1953. doi: 10.1063/1.1699114.

W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57:97–109, 1970. doi: 10.1093/biomet/57.1.97.

S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE*, PAMI-6, 1984. doi: 10.1109/TPAMI.1984.4767596.

H. Rue and L. Held. *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall/CRC, 2005. doi: 10.1201/9780203492024.

G. O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scaling of random walk metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120, 1997.