RESEARCH ARTICLE

Key Points:
- We performed a Sobol' sensitivity analysis of an atmospheric dispersion model on the Fukushima case
- The computational cost was drastically reduced using a Gaussian process emulator of the model
- Aggregated outputs are controlled by emitted amounts and local outputs by wind perturbations

Correspondence to:
S. Girard,
pro@sylvaingirard.net

Emulation and Sobol' sensitivity analysis of an atmospheric dispersion model applied to the Fukushima nuclear accident

Sylvain Girard[1], Vivien Mallet[2], Irène Korsakissok[1], and Anne Mathieu[1]

[1]Institut de Radioprotection et de Sûreté Nucléaire, Fontenay-aux-Roses, France, [2]INRIA, Paris, France

Abstract Simulations of the atmospheric dispersion of radionuclides involve large uncertainties originating from the limited knowledge of meteorological input data, composition, amount and timing of emissions, and some model parameters. The estimation of these uncertainties is an essential complement to modeling for decision making in case of an accidental release. We have studied the relative influence of a set of uncertain inputs on several outputs from the Eulerian model Polyphemus/Polair3D on the Fukushima case. We chose to use the variance-based sensitivity analysis method of Sobol'. This method requires a large number of model evaluations which was not achievable directly due to the high computational cost of Polyphemus/Polair3D. To circumvent this issue, we built a mathematical approximation of the model using Gaussian process emulation. We observed that aggregated outputs are mainly driven by the amount of emitted radionuclides, while local outputs are mostly sensitive to wind perturbations. The release height is notably influential, but only in the vicinity of the source. Finally, averaging either spatially or temporally tends to cancel out interactions between uncertain inputs.

1. Introduction

The French Institute for Radiation Protection and Nuclear Safety (IRSN) uses atmospheric dispersion models, both at the early stage of a nuclear accident for emergency decisions and after the release, to provide an evaluation of its environmental and sanitary impact in complement to measurements. These simulations involve large uncertainties, mainly due to input data and also to approximations in physical schemes used in the models. A detailed account of these uncertainties is essential to assess the trustworthiness of such a model as a tool for decision making.

The first source of uncertainty concerns model inputs, namely, meteorological data fields and source term (released quantities and emission timing for each radionuclide). They are still uncertain even a long time after the accident, as illustrated by the Chernobyl and Fukushima disasters. In the case of the Fukushima accident, several studies highlighted the uncertainty of meteorological inputs, resulting in a high variability in deposition patterns [*Draxler et al.*, 2015; *Arnold et al.*, 2015; *Leadbetter et al.*, 2015]. For the source term, the total emitted amount, for a given radionuclide, also varies sometimes by a factor of 2 or more. There is a still larger variability in the estimated emission kinetics, as illustrated in *Saunier et al.* [2013].

Another source of uncertainty stems from unknown coefficients in parametrizations used in the model. These can be included in the pool of model inputs and thus be treated together with the uncertain input data. For instance, *Brandt et al.* [2002] compared the deposition due to Chernobyl accident, simulated with several dry and wet deposition schemes. Several parametrizations for vertical diffusion were also tested in *Brandt* [1998]. The sensitivity to wet deposition schemes was investigated in the Fukushima case by *Arnold et al.* [2015] and *Leadbetter et al.* [2015].

To our knowledge, only *local* sensitivity studies have been conducted in atmospheric dispersion modeling. They consist of changing one uncertain parameter at a time, and computing the resulting difference with a reference configuration, to induce the model's sensitivity to this parameter. This has been done in the aforementioned studies in the Chernobyl and Fukushima cases as well as in *Korsakissok et al.* [2013] and *Groëll et al.* [2014]. This approach does not account for interactions between variables and is highly dependent on the chosen reference configuration. Our strategy of uncertainty characterization begins with *global* sensitivity analysis, that is, the study of the relative influence of each input on each model output. A comprehensive
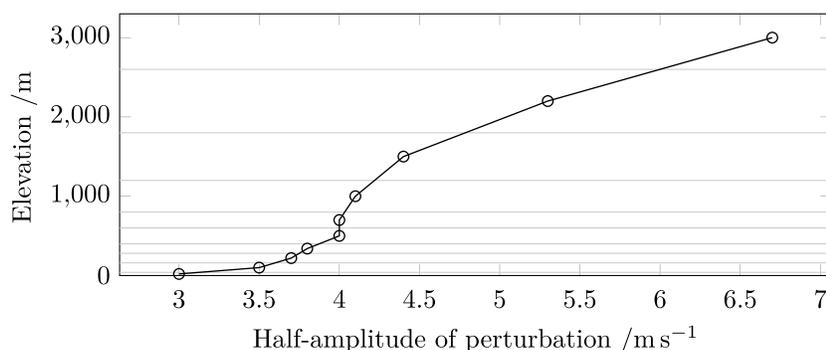
Figure 1. Standardized mean squared errors (SMSE) of the emulators for time-integrated dose rates at measurement stations. The circle diameters are proportional to SMSE. Among the 64 emulators, 30 have an SMSE below 0.15 and are represented by empty black circles. Among the others represented by orange disks, those whose SMSE exceeds 0.25 are marked by a cross. The maximal SMSE, reached just westward of the source, is equal to 0.38. The green triangle indicates the location of the source. The map corresponds to the spatial domain of the simulation. The large grey circle labeled 80 km indicates the zone that was excluded for the second set of emulators.

characterization of the parameters' sensitivity would require to explore the whole input space, that is, sampling all possible combinations of uncertain parameters. This is not an easy task because both the inputs and the outputs of the model are high dimensional and comprise complex correlation structures. The dispersion model used in the present study has 23 uncertain inputs and a high computational cost. This restricts the spectrum of directly available methods to the most simple ones. Indeed, the number of model evaluations required by most sensitivity analysis methods is directly proportional to the number of inputs. Hence, we conducted a first study with the simple *screening* method of *Morris* [1991] [*Girard et al.*, 2014]. It yielded some useful classifications of the inputs depending on their importance regarding a selection of low dimensional outputs.

The variance-based method of *Sobol'* [1993] produces more precise and detailed sensitivity indices. It quantifies the share of the output variance due to each input's uncertainty, as well as their interactions. This method, however, requires large sample sizes to achieve convergence. Hence, we decided to employ model emulation, a common expedient to too high computational costs which consists in approximating the physical model by mathematical functions whose computational cost is negligible. Such a tool, once available, can be used for sensitivity analysis and for any application that would require numerous model evaluations. It was successfully applied by *Lee et al.* [2011] for a global aerosol model.

In this study, we emulated the Eulerian model Polair3D from Polyphemus system on the Fukushima case (see simulations description in section 2) with Gaussian process emulation (section 3). Then, we applied the variance-based method of *Sobol'* [1993] for sensitivity analysis (section 4).

2. Atmospheric Dispersion Simulations

The dispersion of radionuclides following the Fukushima disaster was simulated using the Eulerian transport model Polair3D from the air quality modeling system Polyphemus [*Mallet et al.*, 2007]. An operational version of Polair3D, called ldX, is included in IRSN's operational platform C3X used for emergency evaluations. It has also been used since the Fukushima accident to produce simulations for comparisons to measurements, dose evaluations, and source term reconstruction [*Mathieu et al.*, 2012; *Saunier et al.*, 2013; *Winiarek et al.*, 2014]. The predictive performance of the model on the Fukushima case was assessed by *Mathieu et al.* [2012] and *Saunier et al.* [2013].

The spatial resolution was 0.125°, and the numerical time step was 10 min. The simulations were carried out with 10 vertical layers, whose center altitudes were 20 m, 100 m, 220 m, 340 m, 500 m, 700 m, 1000 m, 1500 m, 2200 m, and 3000 m. The spatial domain of the simulation is represented by the map in Figure 1. The computational cost for one simulation is approximately 2 h on a regular workstation.

2.1. Design of Experiment

We considered three kinds of uncertain inputs: meteorological data, source term data, and model parameters. Each input was assigned a uniform random variable on an interval deduced from the literature and

Table 1. Input Variables, the Bounds of Their Range of Variation, and the Symbols That Represent Them

Variable	Lower Bound	Upper Bound	Symbol
Emission factors[b]	0.333	3	e_{Cs} (caesium), e_I (iodine), and e_{Xe} (xenon)
Emission delay[c] (h)	−6	6	t
Source altitude[a]	first layer	fourth layer	z
Scavenging factors[a] (s^{-1})	$1 \cdot 10^{-7}$	$1 \cdot 10^{-4}$	a_g (below-cloud gas), a_p (below-cloud particle), α_g (in-cloud gas), and α_p (in-cloud particle)
Scavenging exponent[a]	0.6	1	b_g (below-cloud gas), b_p (below-cloud particle), β_g (in-cloud gas), and β_p (in-cloud and particle)
Precipitation[b]	0.5	2	r
Clouds base height[b]	0.667	2	c_b
Clouds thickness[b]	0.5	2	c_t
Dry deposition velocity[a] (mm s^{-1})	0.5	5	d_g (gas) and d_p (particle)
Horizontal diffusion[a] (10^4 m^2 s^{-1})	0	1.5	k_u (zonal) and k_v (meridional)
Vertical diffusion[b]	0.333	3	k_z
Winds[c,a] (m s^{-1})	−2.975	2.975	w_u (zonal), w_v (meridional)
	⋮	⋮	The range of winds perturbation depends on the layer.
	−6.72	6.72	

[a]Variables simply varying inside the prescribed bounds.
[b]Variables perturbed by a multiplicative factor inside the prescribed bounds.
[c]Variables perturbed by an additive increment inside the prescribed bounds.

consultation with experts. The choice of these random distributions is discussed in a previous paper [*Girard et al.*, 2014]. These intervals are reported in Table 1.

The meteorological fields were produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). The fields are from the 12 h forecast cycles starting from the analyzed fields at 00:00 UTC. They have a resolution of 0.36° horizontally, 60 sigma levels vertically, and a time step of 3 h. Simple constant perturbations were applied to the rain and wind fields in order to keep the physical consistency. The rain intensity perturbation is multiplicative and homogeneous on the whole spatial domain. The wind perturbation is additive and homogeneous in each layer, with values distributed according to the profile displayed in Figure 2. Details about the determination of this profile are given in *Girard et al.* [2014].

The first three rows of Table 1 relate to perturbations of the source term issued by *Mathieu et al.* [2012] restricted to the species responsible for the major part of the dose: ^{134}Cs, ^{136}Cs, ^{137}Cs, ^{137}Ba, ^{131}I, ^{133}I, ^{132}Te, and ^{133}Xe. These eight species were grouped into three families to which were added the species in secular equilibrium: the *caesium family* including caesium and barium, the *iodine family* including iodine and tellurium, and the *noble gases* family restricted to xenon.

The source term was perturbed by an additive time shift and multiplicative factors on the emission rates for each of the three families. The source altitude was assigned a discrete uniform random variable with four levels corresponding to the four first vertical layers, since the plume is instantly diluted within the layer where it is emitted. The time and vertical shifts were the same for all species.

We used two sets of wet scavenging coefficients and dry deposition velocities to distinguish between gaseous and particulate species. The gaseous species are actually restricted to iodine because xenon is not deposited. The particulate species is restricted to the caesium family. Indeed, iodine is emitted both in particulate and gaseous forms in unknown proportions, and the description of the dynamic chemical equilibrium between them is beyond the scope of our model. This additional source of uncertainty was not modeled separately but was included into the variability of the coefficients and velocity for gaseous species which apply to the whole iodine family. This is an acceptable approximation since the partitioning between gas and aerosol is only used in the deposition process in our model, and the transition from one phase to another is not modeled. Finally, the range of scavenging factors was lowered compared to our previous study [*Girard et al.*, 2014] from 10^{-7} s^{-1} to 10^{-4} s^{-1} following recommendation of experts from IRSN.
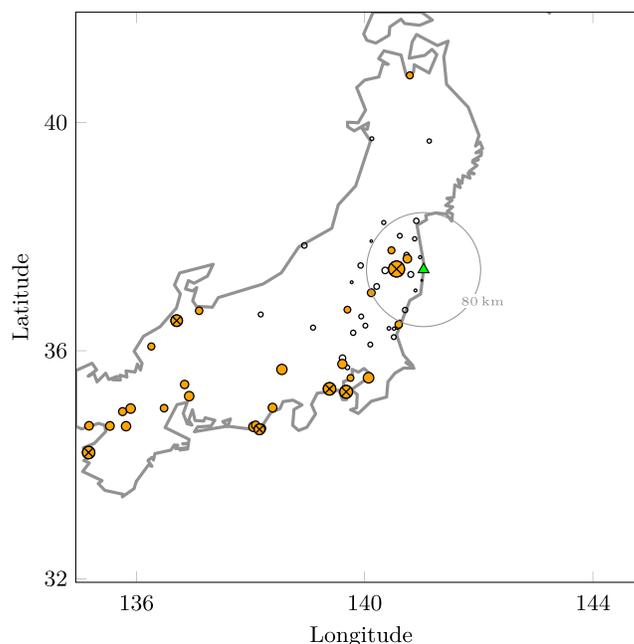
Figure 2. Vertical profile of the half amplitude of wind components perturbation. The grey lines indicate the boundaries of the vertical layers of the model.

2.2. Output Variables

The raw outputs of Polair3D are spatiotemporal fields of gamma dose rates and species concentrations. We aggregated these into more macroscopic outputs because the statistical methods used in this study only apply to scalar outputs. Atmospheric gamma dose rates are the result of gamma rays emitted by all radionuclides within the plume (*plume shine*). Ground gamma dose rates represent the contribution of radionuclides deposed on the ground (*ground shine*). The first aggregated output we consider is the spatial average on the first vertical layer of the atmospheric gamma dose rates, integrated over time. The second one is the spatial average of the maximum of the deposit gamma dose rates. These are very macroscopic variables that convey global information about the model responses. The spatial averaging was carried out both over the whole simulation domain, shown in Figure 1, and with exclusion of data within 80 km of the source. This limit is a conservative estimate of the domain of validity of the model. Indeed, subgrid-scale phenomena are not adequately reproduced by Eulerian models such as Polair3D. The common limitation of instant mixing within grid cells does not fit with point source emissions which usually stay confined in a restricted area for an extended period of time [*Maryon et al.*, 1996; *Brandt et al.*, 1998; *Korsakissok et al.*, 2013]. The region delimited by this criterion can be seen on the map in Figure 1.

As hinted in section 1, this sensitivity analysis of Polair3D is a preliminary work for uncertainty analysis which raises the difficult issue of specifying stochastic descriptions of the input uncertainties. One possible approach would be to use observations to draw Bayesian inferences about the input distributions. Hence, we also considered scalar outputs that could be later compared with observations. In the Fukushima case, gamma dose rates were recorded at many measurement stations over Japan, with a fair spatial and temporal coverage. We considered here the time integrals of these signals. This integrated value is used for estimations of dose received by local populations, which are used to infer emergency response and, later, to assess the sanitary impact of an accident. Thus, it is of particular importance to evaluate the sensitivity of this output to uncertain parameters.

3. Gaussian Process Emulation

One simulation with Polair3D and the parameters given in section 2 takes approximately 2 h on a regular workstation. Given the number of inputs, this cost is too high to allow for a direct estimation of sensitivity indices with the method of Sobol' described in section 4.1. This problem is relatively common in sensitivity

analysis, and a popular expedient is to use an emulator, also known as metamodel, surrogate model, or response surface [*Iooss et al.*, 2006]. An approximation of the model can be built from a modest-sized sample of simulations covering the input domain of interest.

3.1. Principle of Gaussian Process Emulation

The use of Gaussian process models for data interpolation is popular for model emulation because it often leads to robust results and enjoys a rich theoretical background [*Rasmussen and Williams*, 2006]. A Gaussian process is defined as "a collection of random variables, any finite number of which have a joint Gaussian distribution" [*Rasmussen and Williams*, 2006]. We settled here for the framework akin to the "ordinary Kriging" method from the geostatistical realm [*Chilès and Delfiner*, 1999; *Roustant et al.*, 2012]. The response surface, namely, the surface described by the output when the inputs vary, is modeled as the sum of a constant term and a realization of a centered Gaussian process. We denote $y = g(\mathbf{x})$ the output of the model Polair3D, g, evaluated at inputs $\mathbf{x}$. The emulator, f, is itself the conditional expectation of a Gaussian process and assumes the following form:

$$f(\mathbf{x}) = C + \sum_{j=1}^{n} w_j(\mathbf{x}) \left(g(\mathbf{x}^{(j)}) - C \right), \qquad (1)$$

where C is a constant term, $\mathbf{x}^{(j)}$ constitute a learning sample of size n, and $w_j(\mathbf{x})$ are interpolation weights. The constant trend C is determined by least squares on the training sample. The formula for the weights $w_j(\mathbf{x})$ is derived by minimizing the variance of the prediction error, which requires to model the covariance of the Gaussian process. This is usually done parametrically by choosing beforehand a family of covariance kernels. This choice reflects assumptions about the expected regularity and smoothness of the actual response surface. Common covariance kernels are usually written so as to use a characteristic length scale as parameter. It describes the way correlation decreases between the responses evaluated at two input points, as a function of their distance. The problem of estimating these parameters is considerably simplified by using *separable* kernels which are multivariate kernels built as tensor products of one-dimensional kernels. With this restriction, the covariance is a product of one-dimensional functions of the distance, each with one parameter. This is the approach retained by the developers of the DiceKriging R package [*Roustant et al.*, 2012] that we used for this study. The actual inference of the parameters was done using a maximum likelihood estimation algorithm implemented in the aforementioned package.

3.2. Sampling Scheme for Emulation

The training sample must be carefully devised for the emulation techniques presented above to be successful [*Pronzato and Müller*, 2012]. For instance, simple Monte Carlo sampling often results in poor performance, especially in high dimensional settings. It suffers from a tendency to form clusters of points which results in uneven sampling density and leaves some regions of the input space unexplored. We will say that it is poorly *space filling*.

Because our computational budget is limited and the input dimension, 23, is high, it appears that the training sample must be very space filling for the interpolation to be reliable. Actually, should all the inputs show strong and complex influence, it would be probably impossible to fill densely the input space. Indeed, the volume of a hypercube grows exponentially with its dimension. We know, however, from our previous study [*Girard et al.*, 2014] that the chosen model outputs are mostly driven by a few inputs. Hence, we are actually interested in lower dimensional subspaces of much smaller volume.

This calls attention to another desirable property for the design of experiment, namely, that it should be *resilient to projection*. Contrary to independent uniform sampling, a regular or staggered grid is very space filling. However, its projections onto lower dimensional subspaces have catastrophic space filling properties: all points that are aligned along the directions of projection collapse onto one another and the *effective* sampling size dwindles dramatically [*Sobol'*, 1979]. Simple Monte Carlo samples also share, to a lesser extent, this poor resilience to projection. Therefore, we need a training sample strongly resilient to projection.

Among the sampling schemes designed to address these issues, Latin hypercubes [*McKay et al.*, 1979] enjoy a strong popularity for emulation problems. The first step to build a Latin hypercube of size n and dimension p is to divide the p-dimensional unit hypercube into n^p regularly arranged cells of equal volume. Then, a set of n of those cells is selected so that no two cells are aligned in any direction of the canonical basis. Finally, a point is sampled uniformly in each of the selected cells. The regularity and nonalignment property of Latin

hypercubes makes them both space filling and resilient to projection. Yet not all Latin hypercubes are equally efficient, and in difficult cases, it might be desirable to invest some time to pick a good one. We followed here recommendations by *Iooss et al.* [2010] and *Damblin et al.* [2013] to use a Latin hypercube whose *centered L²* *discrepancy* [*Hickernell*, 1998] was minimized using the *enhanced stochastic evolutionary* algorithm devised by *Jin et al.* [2005]. This ensures excellent space filling properties and resilience to projection on subspaces of dimension up to 2. Finally, the input variables are scaled to the appropriate intervals given in Table 1. The source altitude is a categorical variable. The vertical levels were assigned equal probabilities by binning the corresponding input interval into segments of equal length.

3.3. Assessing the Quality of Emulators

Due to our limited simulation budget and despite our efforts to optimize the sampling scheme, it may happen that the emulation fails. Because the emulator will be used later as a surrogate for the physical model, it is essential to be able to assess the quality of the approximation. Cross validation is often used for that purpose, but in our case, its cost was deterring because we had to fit numerous emulators. Additionally, the cross-validation process breaks the good structural properties of the training sample and thus may result in too pessimistic assessments.

Another approach is to build a sample of simulations devoted to testing the emulators. Here again, the sampling scheme is determinant. In addition to being poorly space filling, simple Monte Carlo may generate test points close to training points which results in too optimistic assessment. Using a sampling strategy similar to the one used for training may marginally improve stability but still does not provide any safeguard against proximity between the two samples. A solution proposed by *Iooss et al.* [2010] is to build the test sample sequentially by picking points from a long low-discrepancy sequence while minimizing the centered L^2 discrepancy of the union of the training and test sample. Low-discrepancy sequences are deterministic sequences that mimic the uniform distribution while ensuring space filling properties. We used here a Sobol' sequence [*Sobol'*, 1998] generated with BRODA's *Sobol'Seq51* algorithm [*BRODA*, 2013]. Informally, the algorithm of *Iooss et al.* [2010] builds a test sample that "complements" the training sample by selecting points in regions of low sampling density.

Once a test sample is available, several measures of the quality of the approximations produced by the emulators can be used. One of the most straightforward measures is the mean squared error between emulator predictions and physical model response. In order to ease interpretation, we used the physical model response variance as a normalizing constant to compute *standardized mean squared errors* (SMSE) [*Rasmussen and Williams*, 2006]

$$\text{SMSE} = \frac{\overline{(y_* - f(\mathbf{x}_*))^2}}{\text{Var } y_*}, \tag{2}$$

where $y_* = g(\mathbf{x}_*)$ and $\mathbf{x}_*$ are the test response and sample, f represents the emulator, and the bar indicates average over the test sample. A model that would always predict the mean of the training set would have an SMSE approximately equal to 1. Another quality estimate popular in the machine learning community is the *mean standardized log loss* (MSLL). The log loss is the negative log likelihood

$$- \ln(P(y_*|\mathbf{x}_*, y, \mathbf{x})) = \frac{1}{2} \ln\left(2\pi\sigma_*^2\right) + \frac{(y_* - f(\mathbf{x}_*))^2}{2\sigma_*^2}, \tag{3}$$

where y and $\mathbf{x}$ are the training response and sample and σ_*^2 is the predictive variance [*Rasmussen and Williams*, 2006]. The MSLL is obtained by averaging the difference of the emulator's log loss as defined by equation (3) with the log loss of the trivial model that always predicts the training sample mean. Hence, the lower the MSLL, the better the model, and a model with a 0 MSLL is quite bad.

The SMSE depends solely on the predictive mean of the fitted Gaussian process. It measures how close the emulator predictions are from the original model response. The MSLL additionally takes into account the predictive variance so as to penalize wrong estimations of the accuracy of the approximation.

3.4. Results: Emulation of Fully Aggregated Outputs

There was a total of $p = 23$ inputs (see Table 1). We sampled their space by a Latin hypercube of size 2048 optimized with the procedure described in section 3.2 and used it to train Gaussian process models of the fully aggregated outputs described in section 2.2. The Gaussian process models were fitted using separable Matérn

Table 2. Quality Assessment of the Gaussian Process Emulators of Fully Aggregated Outputs[a]

Aggregated Dose Rate	SMSE		MSLL	
	Gaussian Process	Benchmark	Gaussian Process	Benchmark
Atmospheric	0.014	0.347	−2.143	−0.529
Ground	0.047	0.147	−1.184	−0.954
Distant atmospheric	0.023	0.108	−1.910	−1.113
Distant ground	0.104	0.259	−0.060	−0.697

[a]Simple linear regression models were used as benchmarks. SMSE stands for standardized mean squared errors. MSLL stands for mean standardized log loss.

kernels with parameter 5/2, a very common choice in machine learning [*Rasmussen and Williams*, 2006]. This choice corresponds to the assumption that the model response is twice differentiable. We assessed the quality of these emulators with a second sample of size 2048 built according to the method outlined in section 3.3.

The quality estimates of the emulator for aggregated gamma dose rates are reported in the first two rows of Table 2 along with those obtained with a simple linear model used as a benchmark. The last two rows of Table 2 report the quality estimates obtained when excluding data within 80 km of the source. The performance of the Gaussian process emulators is always excellent except for the MSLL of the emulator for distant ground dose rates. The SMSE of the latter is also not as good as those of the other emulators, but it is still satisfying for our purpose.

3.5. Results: Emulation of Dose Rates Integrated at Measurement Stations

The pair of samples used to emulate and test fully aggregated outputs was then used to fit and test Gaussian process models of the time-integrated dose rates at observation stations. Figure 1 shows the geographic distribution of the corresponding SMSE. Among the 64 emulators, 30 have an SMSE above 0.15 and 6 have an SMSE above 0.25. The maximal SMSE, reached just westward of the source, is equal to 0.38. It appears that stations in the southern region are more difficult to emulate.

Adaptive sampling could be an option to increase the performance of the local emulators while keeping the computational burden sufficiently low for operational use. It seems that the method proposed by *Busby* [2009] based on adaptive griding could be adequate because it identifies the regions of the input space where difficulties arise.

4. Variance-Based Sensitivity Analysis

Sensitivity analysis is the study of how variations in the inputs of a model affect its outputs. There is a wide spectrum of methods dedicated to this problem with different objectives and requirements. Variance-based methods aim at attributing to each input a share of the overall variance of each output. They are among the most computationally demanding sensitivity analysis methods, but they deliver precise estimates both of sensitivities and interactions. Interactions are effects caused by simultaneous variations of several inputs that are not observed when they vary independently. The emulator technique described in section 3 allows to achieve high sample size, which enabled us to use the variance-based method of Sobol' described below.

4.1. Sobol' Sensitivity Indices

Let f be a function that represents the model (in our case, the emulator), $\boldsymbol{x}$ the input vector of size p, and y a scalar output. Each input is assigned a probability distribution whose support coincides with its domain of definition. Here we used uniform variables because we are equally interested in the whole input space. In an uncertainty analysis context, these distributions would be tailored to represent the actual uncertainty of each input.

The expected variance reduction when one of the inputs x_i is fixed to a given value is equal to

$$\text{Var}(y) - E[\text{Var}(y|x_i)] = \text{Var}(E[y|x_i]). \tag{4}$$

The definition of the *first-order Sobol' index* for x_i is obtained by normalizing this quantity by the overall output variance Var (y):

$$S_i = \frac{\text{Var }(\text{E}[y|x_i])}{\text{Var }(y)}.$$ (5)

Hence, first-order indices account for the effect of each input independently, but not of their interactions. Higher-order indices involving more than one input can be defined [*Sobol'*, 2001], but the number of combinations to consider is rapidly overwhelming as the input dimension grows. Let us consider instead the first-order effect of the family $(\mathbf{x}_{\neg i}) = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p)$ constituted of all the inputs except x_i:

$$\text{E}[\text{Var }(y|\mathbf{x}_{\neg i})] = \text{E}\left[\text{Var }\left(y|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p\right)\right].$$ (6)

The complementary of this quantity is the effect of all the families of inputs that include x_i. This leads, after normalization by the overall variance, to the definition of the *total Sobol' indices* for x_i [*Homma and Saltelli*, 1996]:

$$S_i^{\text{tot}} = \frac{\text{E}[\text{Var }(y|\mathbf{x}_{\neg i})]}{\text{Var }(y)} = 1 - \frac{\text{Var }(\text{E}[y|\mathbf{x}_{\neg i}])}{\text{Var }(y)}.$$ (7)

Total Sobol' indices measure the effect of an input, including all its interactions with other inputs. Because non-Singleton families appear in the expression of more than one total index, the sum of all total indices is always equal to or greater than 1. The equality is met when all total indices are equal to their first-order counterpart, namely, when there are no interactions. Indeed, the defining equations (5) and (7) imply that the difference between the first-order and the total index of x_i measures the interactions that x_i is involved in.

4.2. Computation of Sobol' Indices

Assuming that the model can be represented by a square integrable function, the numerators in equations (5) and (7) defining sensitivity indices can be expressed as multidimensional integrals [*Sobol'*, 2001]. Contrary to quadrature techniques which become inefficient in high dimension, the convergence rate of Monte Carlo techniques is independent of the dimension. Equations (5) and (7) actually involve nested integrals: either variances of expectations or expectations of variances. These can be "flattened" by using two independent Monte Carlo samples, $\mathbf{A} \in \mathbb{R}^{N \times p}$ and $\mathbf{B} \in \mathbb{R}^{N \times p}$ where N is the sample size and p the input dimension. Here we used a sample size of 4096. Sample $\mathbf{B}$ is used in combination with sample $\mathbf{A}$ to carry out integration over one variable and, using the partial symmetry between the definition of first-order and total indices, integration over all but one variable. Let $\mathbf{A}_{\mathbf{B}}^{(i)}$ denote the new sample made from $\mathbf{A}$ by replacing its ith column by the ith column of $\mathbf{B}$. We used here the Monte Carlo estimators recommended by *Saltelli et al.* [2010]. For first-order indices, they proposed

$$\widehat{S}_i = \frac{1}{N} \sum_{j=1}^{N} f(\mathbf{B})_j \left[f\left(\mathbf{A}_{\mathbf{B}}^{(i)}\right)_j - f(\mathbf{A})_j \right],$$ (8)

where $f(\mathbf{X})_j$ denotes the response of the model when evaluating the jth row of sample $\mathbf{X}$. They suggest to follow *Jansen* [1999] for total indices:

$$\widehat{S_i^{\text{tot}}} = \frac{1}{2N} \sum_{j=1}^{N} \left[f(\mathbf{A})_j - f\left(\mathbf{A}_{\mathbf{B}}^{(i)}\right)_j \right]^2.$$ (9)

Our sensitivity estimates are subjected to two sources of error. One relates to the convergence of the Monte Carlo estimates of the integral formulation of equations (5) and (7) and was called *sampling error* by *Janon et al.* [2014]. The other stems from the use of an emulator as an imperfect approximation for the physical model. Here we addressed only sampling error using a *bootstrap method* [*Archer et al.*, 1997; *Janon et al.*, 2014]. The bootstrap is a resampling method that can be used to derive confidence intervals for any sampling-based statistical estimator [*Efron and Tibshirani*, 1993]. Here we used 2000 bootstrap replications of the sensitivity indices.
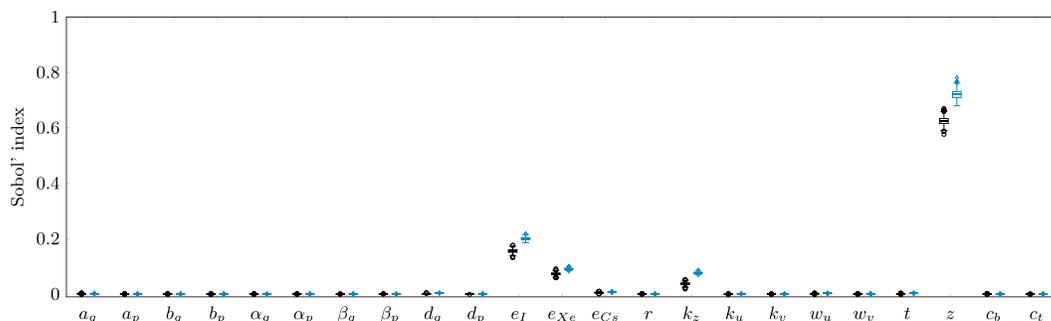
Figure 3. First-order (black) and total (blue) Sobol' indices for the aggregated atmospheric gamma dose rates.

4.3. Results: Sobol' Indices for Fully Aggregated Dose Rates

Figures 3–6 display the first-order and total Sobol' indices for the aggregated atmospheric and ground gamma dose rates, both with and without inclusion of the source's neighborhood. Each pair of box plots corresponds to an input variable. Black box plots on the left-hand side of each pair correspond to first-order Sobol' indices, while blue box plots on the right-hand side represent total indices. The box plots represent the distributions of bootstrap replications thus illustrating sampling error for the Sobol' indices. The box edges indicate the 0.25 and 0.75 quantiles. The whiskers extend to 1.5 times the width of the boxes.

Figure 3 shows that globally aggregated atmospheric dose rates are clearly dominated by the source altitude, followed by the emission factors for iodine and caesium and the vertical diffusion coefficient. All other inputs are almost noninfluential. The differences between pairs of first-order and total Sobol' indices indicate that these four inputs are mildly interacting. Figure 4 shows that ground dose rates are mostly impacted by the emission factor for iodine with almost no interactions. These results are consistent with those previously obtained with the method of Morris [*Girard et al.*, 2014]. The contrast between influential and noninfluential variables is even more pronounced. This is partly due to the fact that the Sobol' indices are defined with variances that involve squares of the outputs, contrary to elementary effects computed with the Morris method.

Excluding the neighborhood of the source dramatically reduces the importance of the source altitude which becomes secondary compared to the emission factors for iodine and xenon and cancels the interactions, as shown in Figure 5. This indicates that the preeminence of the source altitude in the full field case is mostly due to the very high dose rates in the close vicinity of the source. The release height is only important at short distance, since vertical mixing is efficient and rapidly homogenizes the concentrations within the boundary layer when distance from the source increases. The ranking of inputs does not change for deposit gamma dose rates between Figures 4 and 6, but we observe more pronounced interactions for emission time shift and zonal wind.

4.4. Results: Sobol' Indices for Integrated Dose Rates at Measurement Stations

The fully aggregated outputs studied in the previous section give useful hints on the model's global sensitivity, but spatial averaging might reduce the influence of some parameters by smoothing local variations. In practice, atmospheric dispersion simulations are mainly used to infer the contamination at particular areas
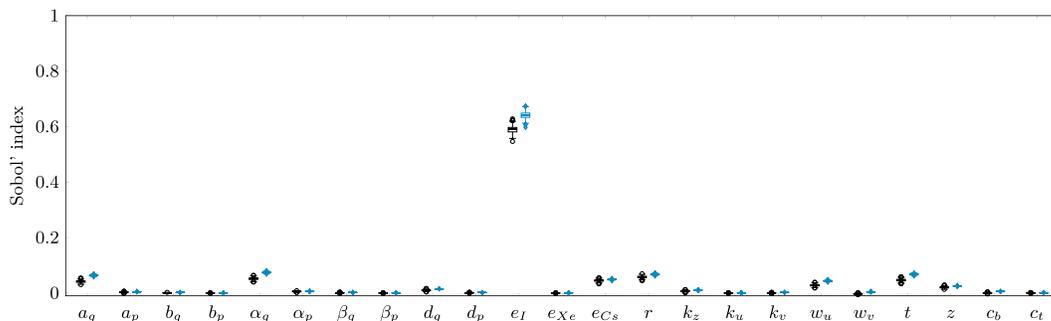

Figure 4. First-order (black) and total (blue) Sobol' indices for the aggregated ground gamma dose rates.
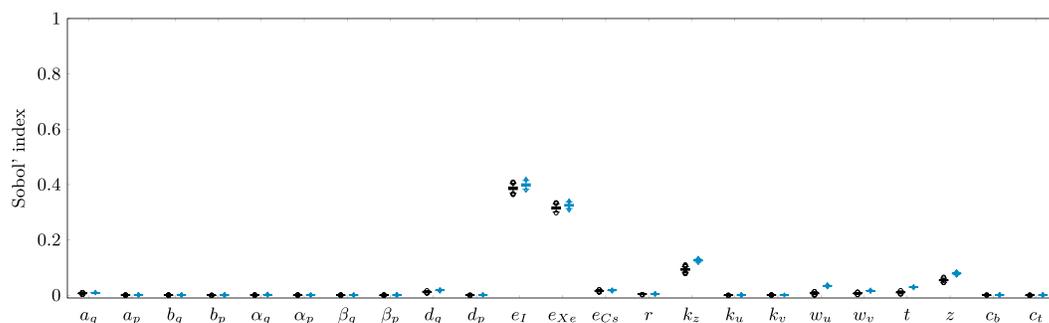
Figure 5. First-order (black) and total (blue) Sobol' indices for the aggregated atmospheric gamma dose rates, without the source's neighborhood.

of interest, as stated in section 2.2. Figure 7 gives an overview of the total Sobol' indices for time-integrated gamma dose rates at measurement stations. For each input, the vertical blue bar indicates the range spanned by its total Sobol' indices over the 64 stations. The two wind perturbations have the highest total Sobol' indices overall, reaching up to 0.9 at some stations, and their indices are nowhere equal to 0. The time shift also has a preeminent influence at some stations, with total indices almost equal to 0.6. Other variables with notable influence include the emission factor for iodine and diffusion coefficients. It must be noted that while the dose rates considered here stem from both the deposit and lower atmosphere, the influence of the former is predominating. Indeed, the signals were time integrated, which lessens the influence of the short-timed atmospheric dose rate peaks compared to the lasting influence of deposit dose rate.

Local Sobol' indices of time-integrated gamma dose rates are displayed in Figure 8. The locations of the measurement stations are indicated by black dots. The Sobol' indices themselves are represented by rays emanating from the dots. For legibility reasons, the indices at some stations behaving similarly to neighboring stations were omitted. Additionally, the indices for only the five most influential inputs are displayed. Each ray bears two ticks: the outermost indicates the value of the total index and the inward one the value of the first-order index. Hence, the gap between ticks measures interaction. The legend in the lower right corner of the figure indicates which ray corresponds to which input and the length corresponding to an index equal to 1.

The strong prevalence of the winds appears clearly on this map. The influence of the winds can be very high because they often determine whether the plume will reach a station or not. A trajectory change can have consequences at long distances. Another factor for the wind preeminence is the horizontal dilution at the source which is driven by the wind velocity. It can be noted, however, that at some stations, the perturbation in only one direction is influential. This is the case, for instance, of the northernmost inland station where zonal wind perturbation has a secondary influence compared to meridional wind perturbation. Results on this station are considered reliable since the emulator's performance was good (see Figure 1), so this is not an artifact due to the emulator.

The input ranking varies gradually as one travels across the map, but there are a few very localized phenomena. In particular, a few stations, most of them being close to the source, are notably influenced by the time
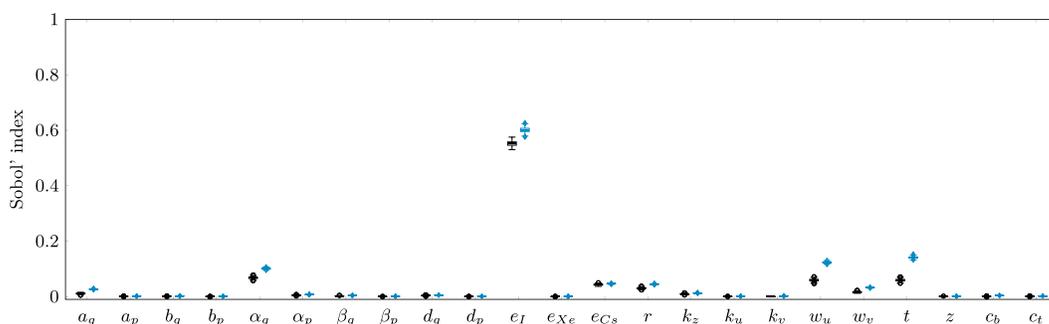

Figure 6. First-order (black) and total (blue) Sobol' indices for the aggregated ground gamma dose rates, without the source's neighborhood.
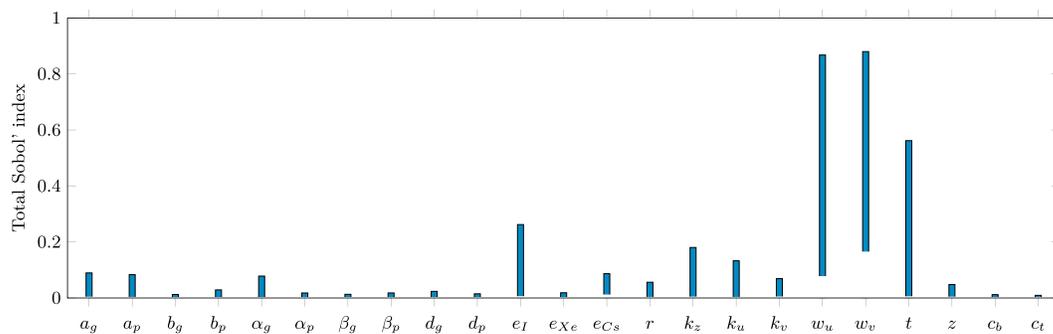
Figure 7. Spread of the total Sobol' indices for time-integrated gamma dose rates at measurement stations. The bottom (top) of the bars indicate the minimum (maximum) values of total indices across the set of 64 measurement stations.

shift. These are stations where most of the contamination occurred during a rain event, mainly on 15 March, resulting in important wet deposition. Changing the emission time would prevent the plume from crossing the rain event, leading to a spectacular drop in gamma dose rates simulated on these stations. These are stations where the emulators perform the best with an SMSE below 0.15. Stations nearby with a higher SMSE showed a similar behavior in the sensitivity study. Other inputs have much less influence.

Finally, interactions are important almost everywhere with many occurrences of important total indices associated to almost null first-order indices.

5. Synthesis and Perspectives

We built Gaussian process emulators of global and local outputs of the Polyphemus/Polair3D atmospheric dispersion model on the Fukushima case. The emulators of global outputs, namely, time- and space-aggregated
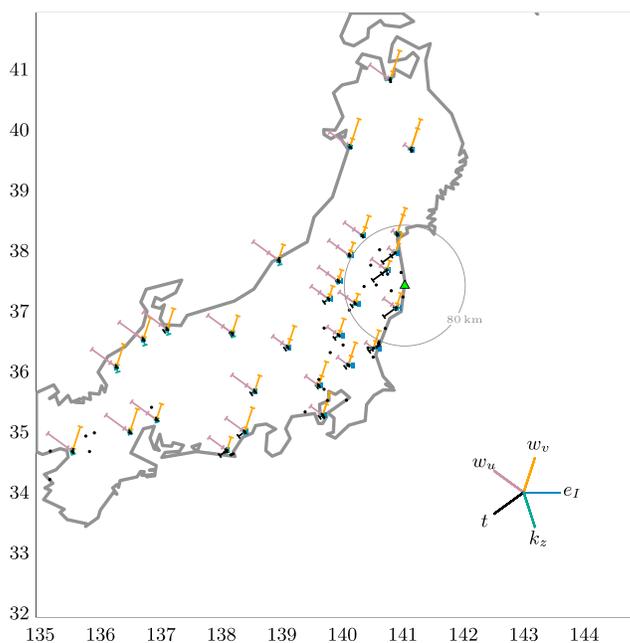

Figure 8. Geographic distribution of the Sobol' indices of the most influential inputs for time-integrated gamma dose rates at measurement stations. Measurement stations are marked by black dots. The rays radiating from the stations represent Sobol' indices. Each ray corresponds to an input, as indicated by the legend in the lower right corner. The length of a ray is proportional to the corresponding total Sobol' index, and the length of the rays in the legend correspond to a unitary total Sobol' index. The ticks at the end of the rays correspond to total indices, while those halfway across the rays correspond to first-order indices. The indices that were elided for legibility reason are all comparable to those of nearby stations.

gamma dose rates, are very close to the original model. Local approximations are harder to obtain: we could only achieve limited fidelity when approximating time-integrated gamma dose rates at some of the measurement stations. Sampling is definitely one of the key ingredients for successful emulation, especially when the input dimension is high. As doubling the sample size brought only limited improvement, we think that further refinement of the emulators could only be achieved through a different sampling strategy, for instance, using an adaptive approach. Another possible option would be to use a more generic structure for the emulator by considering nonstationary models of covariance. However, such approaches [*Romary et al.*, 2015] are still in the early phase of their development.

We used the emulators as surrogates of the original model to perform a variance-based sensitivity analysis with the method of Sobol'. The outputs we studied are mainly determined by a few inputs, but these are not the same for each output. This is in line with the conclusions we drew from applying the Morris screening method to the same model in our previous study [*Girard et al.*, 2014]. Aggregated outputs are mainly driven by perturbations on emitted quantities, whereas local outputs are mostly sensitive to wind perturbations. In addition, the release height has influence on outputs averaged over the whole simulation domain, but its sensitivity index drops when excluding the neighborhood of the source. Thus, high values in the vicinity of the source are predominant when looking at the global influence on the whole domain. It also appeared that averaging tends to cancel interactions. There are indeed lots of interactions for local outputs, while global ones are mostly driven by first-order effects.

The findings of this study imply that priority has to be given to the study of input uncertainties. For meteorological fields, a promising approach is to use meteorological ensembles in order to properly calibrate their uncertainty. Using several meteorological models and configurations, as done in Draxler15, does not ensure that the input's uncertainty is accurately represented. A proper meteorological ensemble has to be designed in order to be representative of the uncertainty of wind and rain fields in the boundary layer and calibrated using meteorological observations. The uncertainties associated to the source term are trickier to assess. Using all available source terms in the literature for the Fukushima accident would give a first idea of the a posteriori uncertainty, since all of them were reconstructed after the accident, using environmental observations. Another approach would rely on experts' knowledge of the events, to define the range of emitted quantities and possible timings for each release period. As shown in this study, the focus could be made on very few radionuclides, namely, iodine, which dominates the dose outputs on the short term, and caesium for long-term impacts. Concerning model parameters, research efforts are currently being made to improve physical schemes related to dry and wet deposition (A. Quérel et al., Hints to discriminate the choice of wet deposition models applied to an accidental radioactive release, submitted to *International Journal of Environment and Pollution*, 2015). However, this study showed that the influence of all deposition-related parameters on the studied outputs is of a second order, compared to inputs uncertainty. This has to be weighted by the fact that only spatial and/or temporal averages were used.

Now that we have emulators at our disposal, many statistical methods whose computational costs were prohibitive with the original model become available. Among them are Bayesian inference schemes, for instance, based on Monte Carlo Markov chain algorithms. Such methods could be used in the future to make the transition from sensitivity analysis to the actual study of the model uncertainty. Indeed, one must specify realistic stochastic representation of input uncertainties to be able to carry out a meaningful uncertainty analysis. The Bayesian approach could be a means to take advantage of the available gamma dose rate observations to determine appropriate probability distributions for the inputs. We saw, however, that uncertainty in time-averaged gamma dose rates at measurement stations is almost exclusively driven by the wind perturbations and emission time shift. Hence, we will probably need other outputs to be able to calibrate all input uncertainties. We could, for instance, try to retrieve the information carried by the dynamics of the model which, in the present study, was concealed for the most part by time averaging.

References

Archer, G. E. B., A. Saltelli, and I. M. Sobol' (1997), Sensitivity measures, ANOVA-like techniques and the use of bootstrap, *J. Stat. Comput. Simul.*, 58, 99–120.

Arnold, D., C. Maurer, G. Wotawa, R. Draxler, K. Saito, and P. Seibert (2015), Influence of the meteorological input on the atmospheric transport modelling with {FLEXPART} of radionuclides from the Fukushima Daiichi nuclear accident, *J. Environ. Radioact.*, 139, 212–225.

Brandt, J. (1998), Modelling transport, Dispersion and Deposition of Passive Tracers from Accidental Releases, National Environmental Research Institute, PhD thesis.

Brandt, J., A. Bastrup-Birk, J. H. Christensen, T. Mikkelsen, S. Thykier-Nielsen, and Z. Zlatev (1998), Testing the importance of accurate meteorological input fields and parameterizations in atmospheric transport modelling using dream-validation against ETEX-1, *Atmos. Environ.*, *32*(24), 4167–4186.

Brandt, J., J. Christensen, and L. Frohn (2002), Modelling transport and deposition of caesium and iodine from the Chernobyl accident using the DREAM model, *Atmos. Chem. Phys.*, *2*(5), 397–417.

BRODA (2013), Sobolseq Sobol sequence generator. [Available at: http://www.broda.co.uk/software.html.]

Busby, D. (2009), Hierarchical adaptive experimental design for Gaussian process emulators, *Reliability Eng. Syst. Saf.*, *94*(7), 1183–1193.

Chilès, J.-P., and P. Delfiner (1999), *Geostatics: Modeling Spatial Uncertainty*, Wiley.

Damblin, G., M. Couplet, and B. Iooss (2013), Numerical studies of space filling designs: Optimization of Latin Hypercube Samples and subprojection properties, *J. Simul.*, *7*, 276–289.

Draxler, R., et al. (2015), World Meteorological Organization's model simulations of the radionuclide dispersion and deposition from the Fukushima Daiichi nuclear power plant accident, *J. Environ. Radioact.*, *139*, 172–184.

Efron, B., and R. Tibshirani (1993), *An Introduction to Bootstrap*, Chapman and Hall/CRC.

Girard, S., I. Korsakissok, and V. Mallet (2014), Screening sensitivity analysis of a radionuclides atmospheric dispersion model applied to the Fukushima disaster, *Atmos. Environ.*, *95*, 490–500.

Groëll, J., D. Quélo, and A. Mathieu (2014), Sensitivity analysis of the modelled deposition of ^{137}Cs on the Japanese land following the Fukushima accident, *Int. J. Environ. Pollut.*, *55*(1–4), 67–75, doi:10.1504/IJEP.2014.065906.

Hickernell, F. J. (1998), A generalized discrepancy and quadrature error bound, *Math. Comput.*, *67*(221), 299–322, doi:10.1090/S0025-5718-98-00894-1.

Homma, T., and A. Saltelli (1996), Importance measures in global sensitivity analysis of nonlinear models, *Reliability Eng. Syst. Saf.*, *52*(1), 1–17.

Iooss, B., F. Van Dorpe, and N. Devictor (2006), Response surfaces and sensitivity analyses for an environmental model of dose calculations, *Reliability Eng. Syst. Saf.*, *91*(10–11), 1241–1251.

Iooss, B., L. Boussouf, A. Marrel, and V. Feuillard (2010), Numerical studies of the metamodel fitting and validation processes, *Int. J. Adv. Syst. Meas.*, *3*(1–2), 11–21.

Janon, A., M. Nodet, and C. Prieur (2014), Uncertainties assessment in global sensitivity indices estimation from metamodels, *Int. J. Uncertainty Quantification*, *4*(1), 21–36.

Jansen, M. J. (1999), Analysis of variance designs for model output, *Comput. Phys. Commun.*, *117*(1–2), 35–43.

Jin, R., W. Chen, and A. Sudjianto (2005), An efficient algorithm for constructing optimal design of computer experiments, *J. Stat. Plan. Inference*, *134*(1), 268–287.

Korsakissok, I., A. Mathieu, and D. Didier (2013), Atmospheric dispersion and ground deposition induced by the Fukushima Nuclear Power Plant accident: A local-scale simulation and sensitivity study, *Atmos Environ.*, *70*, 267–279.

Leadbetter, S., M. Hort, A. Jones, H. Webster, and R. Draxler (2015), Sensitivity of the modelled deposition of Caesium-137 from the Fukushima Dai-ichi nuclear power plant to the wet deposition parameterisation in NAME, *J. Environ. Radioact.*, *139*, 200–211.

Lee, L. A., K. S. Carslaw, K. J. Pringle, G. W. Mann, and D. V. Spracklen (2011), Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters, *Atmos. Chem. Phys.*, *11*(23), 12,253–12,273, doi:10.5194/acp-11-12253-2011.

Mallet, V., et al. (2007), Technical note: The air quality modeling system Polyphemus, *Atmos. Chem. Phys.*, *7*(20), 5,479–5,487, doi:10.5194/acp-7-5479-2007.

Maryon, R., J. Saltbones, D. Ryall, J. Bartnicki, H. Jakobsen, and E. Berge (1996), An intercomparison of three long range dispersion models developed for the UK Meteorological Office, DNMI and EMEP, *Tech. Rep.*

Mathieu, A., et al. (2012), Atmospheric dispersion and deposition of radionuclides from the Fukushima Daiichi nuclear power plant accident, *Elements*, *8*(3), 195–200.

McKay, M. D., R. J. Beckman, and W. J. Conover (1979), Comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, *21*(2), 239–245.

Morris, M. D. (1991), Factorial sampling plans for preliminary computational experiments, *Technometrics*, *33*, 161–174.

Pronzato, L., and W. G. Müller (2012), Design of computer experiments: Space filling and beyond, *Stat. Comput.*, *22*(3), 681–701, doi:10.1007/s11222-011-9242-3.

Rasmussen, C. E., and C. K. I. Williams (2006), Gaussian process for machine learning.

Romary, T., N. Desassis, and F. Fouedjio (2015), A full scale, non stationary approach for the kriging of large datasets, in *Spatial Stat. Conf. 2015*, Avignon, France.

Roustant, O., D. Ginsbourger, and Y. Deville (2012), DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization, *J. Stat. Software*, *51*(1), 1–55.

Saltelli, A., P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola (2010), Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index, *Comput. Phys. Commun.*, *181*(2), 259–270.

Saunier, O., A. Mathieu, D. Didier, M. Tombette, D. Quélo, V. Winiarek, and M. Bocquet (2013), An inverse modeling method to assess the source term of the Fukushima nuclear power plant accident using gamma dose rate observations, *Atmos. Chem. Phys.*, *13*(22), 11,403–11,421.

Sobol', I. M. (1979), On the systematic search in a hypercube, SIAM, *J Numer. Anal.*, *16*(5), 790–793, doi:10.1137/0716058.

Sobol', I. M. (1993), Sensitivity estimates for nonlinear mathematical models, in Matem. Modelirovanie, 2(1)(1990)112–118, *English Transl.: MMCE*, *1*(4), 407–414.

Sobol', I. M. (1998), On quasi-Monte Carlo integrations, *Math. Comput. Simul.*, *47*, 103–112.

Sobol', I. M. (2001), Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Math. Comput. Simul.*, *55*, 271–280.

Winiarek, V., M. Bocquet, N. Duhanyan, Y. Roustan, O. Saunier, and A. Mathieu (2014), Estimation of the caesium-137 source term from the Fukushima Daiichi nuclear power plant using a consistent joint assimilation of air concentration and deposition observations, *Atmos. Environ.*, *82*, 268–279.