

Meta-Modelling for Fast Analysis of CFD-Simulated Vapour Cloud Dispersion Processes

Ke Wang^a, Tao Chen^{b,*}, Soo Tin Kwa^c, Yifei Ma^c, Raymond Lau^{a,*}

^a *School of Chemical and Biomedical Engineering, Nanyang Technological University, 62 Nanyang Drive, Singapore 637459, Singapore*

^b *Department of Chemical and Process Engineering, University of Surrey, Guildford GU2 7XH, UK*

^c *DSO National Laboratories, 20 Science Park Drive, Singapore 118230, Singapore*

Abstract

Released flammable chemicals can form an explosible vapour cloud, posing safety threat in both industrial and civilian environment. Due to the difficulty in conducting physical experiments, computational fluid dynamic (CFD) simulation is an important tool in this area. However, such simulation is computationally too slow for routine analysis. To address this issue, a meta-modelling approach is developed in this study; it uses a small number of simulations to build an empirical model, which can be used to predict the concentration field and the potential explosion region. The dimension of the concentration field is reduced from around 43,421,400 to 20 to allow meta-modelling, by using the segmented principal component transform-principal component analysis. Moreover, meta-modelling-based uncertainty analysis is explored to quantify the prediction variance, which is important for risk assessment. The effectiveness of the methodology has been demonstrated on CFD simulation of the dispersion of liquefied natural gas.

Keywords: Computational fluid dynamics; Design of experiments; Gaussian process regression; Kriging; Surrogate modelling; Vapour cloud dispersion.

1. Introduction

The dispersion of vapour cloud, formed by accidental or intentional release of flammable chemicals, is of significant concern to industrial safety and civilian security. When flammable chemicals leak, it vaporises to form a cloud. In the presence of an ignition source, the vapour cloud may explode if its concentration is between lower explosive limit (LEL) and upper explosive limit (UEL). Therefore, vapour cloud can cause catastrophic explosions, and the potential explosion zone is dependent on the concentration of the flammable chemical. Given a certain release scenario (e.g. the amount of chemical released) with the corresponding environmental parameters (e.g. wind direction and speed), it is crucial to estimate the chemical concentration within the vapour cloud and thus the potential explosion zone. Such estimation forms important information to support risk analysis and

* Corresponding authors. Tel.: +44 1483686593; Email: t.chen@surrey.ac.uk (T. Chen). Tel.: +65 6316 8830; Email: wmlau@ntu.edu.sg (Raymond Lau).

decision making with respect to mitigating strategies and evacuation planning. However, physical experiments of vapour dispersion processes are dangerous, difficult and costly, particularly when a large amount of chemical is involved and/or multiple environmental parameters need to be considered. Consequently, computer simulation has become the major tool to investigate large scale vapour cloud dispersion processes. The most commonly used are Gaussian dispersion models (Barratt 2001), which however are too simplistic to simulate complex, real world scenarios. Alternatively, computational fluid dynamic (CFD) modelling has been adopted to better represent the underlying physical and chemical mechanisms of vapour cloud dispersion (Gavelli et al. 2008, Qiao & Zhang 2010, Graham & Simon 2012).

There are many choices of commercially or freely available CFD tools, such as FLACS (FLame ACceleration Simulator), ANSYS Fluent, and OpenFOAM. Among these, FLACS is widely used in industrial and academic institutions due to its reliable and satisfactory simulation results (Hanna et al. 2009, Middha et al. 2009); it has been approved to be used for LNG vapour dispersion modelling under the US federal regulations and is also used in this study. However, no matter what software is used, one of the major challenges that have hampered the practical application of CFD is the computational cost. For example in this study, a single run of FLACS to simulate the vapour cloud dispersion, due to the release of several tons of liquefied natural gas (LNG), takes several days using a high performance cluster with peak performance up to 28 TFLOPs; it may take even months if the mass of LNG released increases to several kilo-tons. This is too computationally demanding for routine analysis, and it is a generic problem for complex computer simulations. In order to address this challenge, the meta-modelling technique has been reported in the literature, in which the main concept is to use “computer simulation data” to develop an empirical model (termed meta-model, surrogate model or response surface) (Sacks et al. 1989, Chen et al. 2011). In the field of CFD simulation, meta-modelling has been recognised as a useful tool to help reduce the computation demand (Marjavaara et al. 2007, Coetzee et al. 2012). To study vapour cloud dispersion processes, a meta-model can relate the concentration field to the input variables of the simulation (e.g. the mass of chemical released, the wind speed and weather condition), and thus is able to predict the concentration for any value of these inputs. Nevertheless, to the best knowledge of the authors, meta-modelling has not been explored with respect to its application to CFD simulation of vapour cloud dispersion, which is the primary aim of this work.

The concept of meta-modelling can be realised in different ways, depending on the choice of how the computer “experiments” (simulations) are designed (i.e. design of experiments, DoE) and what empirical models are adopted. DoE aims to determine the proper values of simulation inputs, so that the simulation data are representative of the entire input space, and thus the model developed gives accurate predictions of the outputs. In this study, DoE is through the Hammersley sequence sampling (HSS) method (Kalagnanam & Diwekar 1997, Yan et al. 2011). HSS is a “space-filling” method to allocate design points to be uniformly distributed within the input space. Compared with classical DoE (e.g. factorial designs (Myers et al. 2009)), space-filling methods tend to give better coverage of the input space and thus more reliable empirical models (Fang et al. 2000). In this work, HSS is used to determine proper combinations of the mass of the released LNG and wind speed, as

inputs, for FLACS simulation. Second, the meta-model can have different forms, such as polynomial functions, neural networks, and Gaussian process regression (GPR, also known as kriging) (Palmer & Realf 2002, Chen et al. 2006). Of particular focus in this work is GPR, for its reliable prediction results when compared with alternative methods (O'Hagan 2006, Wang et al. 2011).

In principle, when studying vapour dispersion, the chemical concentration at different spatial locations and time points could be directly used as the output variables for modelling. However, the dimension of the concentration field from CFD simulation can be extremely high. For example, the 4D simulation (3D space plus time) in this work gives approximately 43,421,400 concentration values, which cannot be directly used as outputs in meta-modelling. This issue can be addressed by using dimension reduction techniques, such as principal component analysis (PCA) (Geladi & Kowalski 1986, Jolliffe 2005). PCA can represent the large number of concentration values with very few score variables, which are used as the outputs for meta-modelling. The model prediction is thus also in the form of PCA scores, which are converted to the prediction for the original concentration field. In our former work, PCA was used for a small-scale CFD simulation with up to 182,250 output variables (Chen et al. 2011). However, when the output dimension further increases, the conventional PCA algorithm becomes undesirable or even infeasible in terms of computer memory. In order to solve this problem, a special PCA algorithm, the segmented principal component transform-principal component analysis (SegPCT-PCA) (Barros & Rutledge 2005), is used in this study. In SegPCT-PCA, the high dimensional data are divided into lower dimensional segments, and PCA is applied to each segment to obtain the segment-specific scores (the so called SegPCT step). These segment-specific scores, already in dramatically lower dimension than the original data, go through another PCA (thus the overall acronym SegPCA-PCA) to obtain the final scores. This “divide-and-conquer” methodology has been demonstrated to produce numerically indistinguishable results when compared with direct PCA (Barros & Rutledge 2005), whilst dramatically reducing the memory use.

Another important task of this work is to analyse the prediction uncertainty of the meta-model. In the event of chemical release, a single prediction (usually the mean prediction) of the potentially explosible region may not be very useful. Instead, uncertainty analysis identifies a range of plausible scenarios, such as the regions with 50% (the mean), 10%, and 1% (the worst case) probability of explosion; such information will be crucial for the authority to conduct risk assessment and make suitable decisions accordingly (Smith 2006).

In the context of meta-modelling, the main sources of prediction uncertainty include: (i) the mismatch between the real process and the CFD simulation, (ii) the mismatch between the CFD simulation and the meta-model, and (iii) the variability in the input variables. The validation between the real process and CFD simulation has been reported in the literature (Middha, Hansen et al. 2009, Hansen et al. 2010, Graham & Simon 2012). However in this study, it is not possible to compare simulation with real data, because we aim to mimic the actual weather conditions and potential release scenario in Singapore. No such experimental data are available (to our best knowledge), and it is infeasible for us to conduct such large-scale experiments with tons of chemicals released. In addition, the primary purpose of this study is to investigate the effectiveness of meta-modelling for fast analysis of CFD simulations. Therefore, although the reality-CFD mismatch may well be the primary source of

uncertainty, it is not explored in this work, and we focus on the latter two types of uncertainty. In more detail, the meta-model uncertainty, due to the mismatch between the CFD simulation and meta-model, is reflected in the prediction variance that is automatically provided by GPR (Chen et al. 2011). Furthermore, the input uncertainty refers to the uncertainty in the mass of chemical released and the wind speed (available from historical data), and it propagates through the meta-model to the prediction uncertainty in the output concentration field (Oakley & O'Hagan 2002). The Monte Carlo (MC) method is used to quantify the prediction uncertainty due to the input variability (Girard et al. 2003, O'Hagan 2006). Finally, the model and input uncertainty are combined through a mixture distribution model (Shi et al. 2003, Shi et al. 2005). It should be noted that uncertainty analysis with respect to input variability only becomes feasible with the use of meta-model; otherwise the MC method would require hundreds of simulation runs, a single of which takes days to months to complete.

In summary, the primary contribution of this work is to explore the use of meta-modelling in medium-to-large scale, 4D CFD simulations of LNG vapour cloud dispersion processes. The main challenge is the extremely large number of concentration variables (approximately 43,421,400), which is reduced by the SegPCT-PCA algorithm prior to being used as output for meta-modelling. In addition, this work investigates the prediction uncertainty, as a result of model and input uncertainties, of meta-models with very high dimensional outputs. Existing uncertainty analysis with meta-modelling has been focused on univariate output (Eisenhower et al. 2012, Manfren et al. 2013). As such, this study contributes a method to quantify the variability of meta-model-predicted concentration field (and other high dimensional variables), which is potentially crucial information to aid risk assessment and decision making in practice.

2. Methods

In this study, this first step is to use FLACS to simulate the dispersion of LNG vapour cloud under a range of released mass and wind speed. The values of these two input variables are determined by HSS, in order to provide proper coverage of the input space. After the simulations are completed, the high dimensional concentration variables are fed through the SegPCT-PCA algorithm to give the low dimensional PCA score variables. Subsequently, a GPR meta-model is developed to relate the scores (outputs) to the LNG mass and wind speed (inputs). Finally, the meta-model's prediction uncertainty, due to the model and input uncertainties, is quantified by using the MC method. An overview of HSS, CFD simulation using FLACS, SegPCT-PCA, GPR and uncertainty analysis methods is presented in the rest of this section. A flowchart is given in Figure 1 to illustrate the overall methodology.

(Figure 1 about here)

2.1 Hammersley sequence sampling (HSS)

The principle of HSS (Kalagnanam & Diwekar 1997, Yan, Chen et al. 2011) is that any integer n can be re-written in a radix notation of another integer R as

$$n \equiv n_0 n_1 n_2 \dots n_{m-1} n_m = n_m + n_{m-1} R + n_{m-2} R^2 + \dots + n_1 R^{m-1} + n_0 R^m \quad (1)$$

where m is the integral part of $\log_R n$. A function of n , called inverse radix number, can be constructed by reversing the order of the digits of n and concatenating them behind a decimal point:

$$\phi_R(n) = 0.n_m n_{m-1} \dots n_2 n_1 n_0 = n_m R^{-1} + n_{m-1} R^{-2} + \dots + n_1 R^{-m} + n_0 R^{-m-1} \quad (2)$$

Let the input variable \mathbf{x} (i.e. LNG mass and wind speed in this study) be a D dimensional vector. We select the first $D - 1$ prime numbers, denoted R_1, R_2, \dots, R_{D-1} , as the integer R in Eq. (1). Then the N design points (N combinations of LNG mass and wind speed as input variables for simulation), each being a vector of order D , are given by

$$\mathbf{x}_n = \mathbf{1} - \left(\frac{n}{N}, \phi_{R_1}(n), \phi_{R_2}(n), \dots, \phi_{R_{D-1}}(n) \right)^T \quad (3)$$

where $n = 1, 2, \dots, N$, and $\mathbf{1}$ is a unity vector.

This HSS method gives N design points within the range $[0, 1]$ for each input variable. In general, if the required range of an input variable is $x \in [a, b]$, then the design points for that input need to be transformed by multiplying $(b - a)$ and then adding a .

In this study, the CFD simulation across open sea requires three inputs, namely the wind speed, the current speed, and the mass of LNG released. It is reported that the surface drift (current speed) is approximately 3% of the wind speed under turbulent condition (Shemdin 1971). For simplicity, this dependency is adopted. As a result, only the wind speed and LNG mass are considered in HSS.

2.2 CFD simulation using FLACS

Once N sets of values are determined using HSS, they are used as the input variables in FLACS to simulate the dispersion of vapour cloud. FLACS is a comprehensive CFD tool for exploring the consequences of different types of flammable and toxic release scenarios. The governing equations include the conservation of mass and momentum, transport equations for enthalpy, fuel mass fraction, the mixture fraction, turbulent kinetic energy, and dissipation rate of turbulent kinetic energy, which are explained in details in the FLACS manual. The governing equations for mass and momentum are solved by using the unsteady Reynolds-averaged Navier-Stokes approach on a non-uniform Cartesian grid, closed using the standard $k - \varepsilon$ model (Hansen, Gavelli et al. 2010). The $k - \varepsilon$ is an eddy viscosity model that solves two additional transport equations of turbulent kinetic energy and dissipation of turbulent kinetic energy.

In FLACS simulation, the geometry, grid and scenario parameters need to be determined first. In this work, the domain is oriented that the X-direction and Y-direction are horizontal and the Z-direction is vertical; the X-direction is parallel to the wind, and the Y-direction is perpendicular to the wind. The X-coordinate extends from -100 m to 700 m, the Y-coordinate from -150 m to 150 m, and the Z-coordinate from 0 m to 50 m. The simulation domain is divided into small box grid cells. FLACS uses the distributed porosity concept to handle sub-grid objects that are smaller than the grid

cell size, making the dispersion simulation in complex geometries feasible (Hansen, Gavelli et al. 2010). The scenario parameters define the chemical, boundary conditions, monitoring point locations, ignition position and ignition time, and so on. For LNG composition considered in the context of this work, the methane, ethane and propane content by mole percentage was taken to be 88.07 %, 9.15 % and 2.78% respectively. In FLACS, wind direction cannot be changed and remains constant in a single simulation. In this work, the dispersion is simulated offshore, and no obstacles are in the domain. Therefore, the wind direction is taken to be constant in the course of the simulation. The pool model in FLACS is employed in this work. The mass of LNG is assumed to leak onto open water surface of an ambient temperature of 31 °C and the dispersion characteristics is studied.

The simulations terminates when all the concentrations in the domain drop to less than 10^{-13} . (All concentration values are in volume fraction in this paper.) The simulation time depends on several factors, such as grid and domain size, the mass of released LNG, and wind speed. When other scenario parameters are specified, the simulation time is up to the mass of released LNG, and wind speed: lower mass with faster wind gives shorter time. In this work, each simulation takes 2-4 days.

As mentioned before, LEL and UEL are important parameters to determine the explosion potential of flammable chemicals. The LEL and UEL for LNG are 4.6% and 14.6%, respectively, which are obtained using the Le Chatelier's mixing rule (Le Chatelier 1891), taking into account the mole percentage of methane, ethane and propane used for simulations. The plume with concentration between the LEL and the UEL is of interest, since it has the potential of explosion. Once the leakage occurs, plume within flammable limits could be observed near the source term. At the beginning, areas within the plume can have concentrations higher than UEL and decreases over time. Although the region above UEL is not explosible at a particular time, it is still considered hazardous in practice. Therefore, it is the contour of LEL that is of primary interest in this study.

2.3 Segmented principal component transform-principal component analysis (SegPCT-PCA)

From each simulation run, the concentration of the chemical is a function of 3D spatial location and time, and thus of extremely high dimension. PCA is a common method for dimension reduction. Let the concentration values from the n -th run are organised into a vector \mathbf{y}_n ($n = 1, \dots, N$), each having dimension of P , and the data from all the N runs are concatenated into a matrix of order $N \times P$: $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]'$. In PCA, the concentration matrix \mathbf{Y} is projected into scores and loadings as (Wold et al. 1987, Jolliffe 2005)

$$\mathbf{Y} = \mathbf{TP}' \quad (4)$$

where \mathbf{T} ($N \times Q$) is the scores, and \mathbf{P} ($P \times Q$) the loadings. If the original variables are highly correlated, the dimension of \mathbf{T} can be dramatically smaller than that of \mathbf{Y} (i.e. $Q \ll P$), whilst still containing the essential information of \mathbf{Y} . Recall that the P variables in \mathbf{Y} are concentration values across spatial and temporal coordinates, which are indeed highly correlated (Chen et al. 2011). Therefore, the PCA can be an effective approach to dimension reduction.

However, the classical PCA algorithm requires many matrix multiplications, which can be very demanding in terms of memory. Direct use of PCA is not feasible, on usual desktop computers, for

dealing with the extremely large dimension in this study. In order to solve this problem, the SegPCT-PCA algorithm (Barros & Rutledge 2005) is adopted. The main principle of SegPCT-PCA is “divide-and-conquer”, that is, to divide the matrix \mathbf{Y} into column-wise segments as $\mathbf{Y} = [\mathbf{Y}_1 | \mathbf{Y}_2 | \dots | \mathbf{Y}_q | \dots]$. For each segment, the scores \mathbf{T}_q and loadings $\mathbf{P}_{\text{PCT}_q}$ are extracted by using PCA as usual. The extracted scores are then concatenated as $\mathbf{T}_C = [\mathbf{T}_1 | \mathbf{T}_2 | \dots | \mathbf{T}_q | \dots]$, and a second PCA is applied to this \mathbf{T}_C to obtain the final score matrix \mathbf{T} and a loading matrix \mathbf{P}_{PCA} . The final loading matrix of each segment can be computed as $\mathbf{P}_q = \mathbf{P}_{\text{PCT}_q} \mathbf{P}_{\text{PCA}}$, and the overall loading matrix is: $\mathbf{P} = [\mathbf{P}_1 | \mathbf{P}_2 | \dots | \mathbf{P}_q | \dots]$. It has been demonstrated that the final scores and loadings from SegPCT-PCA are numerically equivalent to those from direct application of PCA on the initial large matrix (Barros & Rutledge 2005). The computation time of SegPCT-PCA depends on the choice of the number of segments: the more segments, the more computation time required. The number of segments is thus kept as large as possible, whilst allowed by the available computer memory, so that less computation time is needed.

2.4 Gaussian process regression (GPR)

After dimension reduction, a GPR model is developed to relate the input variables, \mathbf{x} , to the PCA scores as output. Since the scores are of dimension Q , this is a multi-output regression problem. It is possible to model the multiple outputs directly in GPR (Rasmussen & Williams 2006); however such methods can give complex models in terms of implementation and computation. Instead, we choose to build multiple GPR models, one for each dimension of the score; this approach has been shown to give good prediction accuracy in practice (Chen et al. 2011). For brevity, a single score vector from the matrix \mathbf{T} is denoted \mathbf{t} , which is of the order $N \times 1$ (N being the number of simulation runs). Such an output variable is modelled by a joint Gaussian distribution with zero mean as $\mathbf{t} = (t_1, \dots, t_N)' \sim G(\mathbf{0}, \mathbf{C})$. The ij -th element of the covariance matrix \mathbf{C} is defined in terms of the following covariance function:

$$C_{ik} = C(\mathbf{x}_i, \mathbf{x}_k) = a_0 + a_1 \sum_{j=1}^D x_{ij} x_{kj} + v_0 \exp\left(-\sum_{j=1}^D w_j (x_{ij} - x_{kj})^2\right) + \sigma^2 \delta_{ik} \quad (5)$$

where x_{ij} is the input of the i -th ($i = 1, \dots, N$) simulation at the j -th ($j = 1, \dots, D$) dimension and $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{iD}]$. The first two terms of the covariance function represent the constant bias and linear correlation, respectively. The third term is similar to the form of the radial basis function, and the fourth term corresponds to the random error. The model parameters in the covariance matrix $\boldsymbol{\theta} = (a_0, a_1, v_0, w_1, \dots, w_D, \sigma^2)$ can be estimated by using the maximum likelihood method (Rasmussen & Williams 2006). After finding the model parameters, for a new input variable \mathbf{x}_* , the prediction for the output is also normally distributed with mean \hat{t}_* and variance $\hat{\sigma}_*^2$ given by

$$\hat{t}_* = \mathbf{k}'(\mathbf{x}_*) \mathbf{C}^{-1} \mathbf{t} \quad (6)$$

$$\hat{\mathbf{S}}_*^2 = C(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}'(\mathbf{x}_*)\mathbf{C}^{-1}\mathbf{k}(\mathbf{x}_*) \quad (7)$$

where $\mathbf{k}(\mathbf{x}_*) = [C(\mathbf{x}_*, \mathbf{x}_1), \dots, C(\mathbf{x}_*, \mathbf{x}_N)]'$. The capability of automatically providing the prediction uncertainty (i.e. the variance in Eq. (7)) is a special feature of the Bayesian regression methods, of which GPR is a widely used model. This prediction variance reflects the un-modelled mismatch between the meta-model and the simulation data, and can be regarded as originating from model uncertainty.

Following this procedure, a total of Q GPR models are developed for all Q score vectors, and the prediction for a new variable \mathbf{x}_* has Q independent normal distributions given in Eqs. (6)(7): $\{\hat{t}_{*,q}, \hat{\mathbf{S}}_{*,q}^2; q = 1, \dots, Q\}$. The prediction of the concentration variables, also being normally distributed, can be obtained by applying the following transformation (Chen et al. 2011):

$$\hat{\mathbf{y}}_* = [\hat{t}_{*,1}, \dots, \hat{t}_{*,Q}]\mathbf{P}' \quad (8)$$

$$\hat{\boldsymbol{\sigma}}_*^2 = \text{diag} \left(\mathbf{P} \begin{bmatrix} \hat{\mathbf{S}}_{*,1}^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \hat{\mathbf{S}}_{*,Q}^2 \end{bmatrix} \mathbf{P}' \right) \quad (9)$$

2.5 Uncertainty analysis

The aim of uncertainty analysis is to quantify the variance of the prediction due to the uncertainties in the meta-model and the input variables. The effect of model uncertainty has already been given in Eq. (9). The input uncertainty refers to the variability about the input variables, which is usually represented by a probability density function (*pdf*) of the inputs. The *pdf* can be determined based on the experience of the domain experts, past release scenario and historical data with respect to the LNG mass and wind speed. Then, the Monte Carlo (MC) method (Kalos & Whitlock 2008) can be used to draw random samples from this *pdf*. For each MC sample, the meta-model will give an prediction of the outputs in the form of normal distributions with mean $\hat{\mathbf{y}}_*^{(h)}$ and variance $\hat{\boldsymbol{\sigma}}_*^{2,(h)}$, $h = 1, \dots, H$, where H is the number of MC samples. According to the principle of MC method, the prediction becomes a mixture of H normal distribution. For simplicity, we approximate this mixture distribution for prediction by a single normal distribution: $\mathbf{y}_* \sim G(\hat{\mathbf{y}}_*, \hat{\boldsymbol{\sigma}}_*^2)$, where the mean and variance are (Shi, Murray-Smith et al. 2003, Shi, Murray-Smith et al. 2005):

$$\hat{\mathbf{y}}_* = \frac{1}{H} \sum_{h=1}^H \hat{\mathbf{y}}_*^{(h)} \quad (10)$$

$$\hat{\boldsymbol{\sigma}}_*^2 = \frac{1}{H} \sum_{h=1}^H \hat{\boldsymbol{\sigma}}_*^{2,(h)} + \frac{1}{H} \sum_{h=1}^H \left(\hat{\mathbf{y}}_*^{(h)} \right)^2 - \hat{\mathbf{y}}_*^2 \quad (11)$$

where in Eq. (11), the first term is the prediction variance due to model uncertainty, averaged over MC samples. The second and third terms represent the prediction variance due to input uncertainty. From the prediction, the probability of a particular spatial location at a particular time point being

potentially explosible is $p(\mathbf{y}_* > \text{LEL})$, which can be calculated from the probability function of the normal distribution. Given a designated threshold for the explosion probability (e.g. 5% or 50%), the corresponding LEL contour can be established. Note that using a smaller probability threshold results in a larger LEL contour, since the region that is less likely to be explosible will be included. In a real release scenario, the authority tends to be conservative and uses a small explosion probability (say 5% or even less), which will result in a larger potential impact zone to be considered.

3. Results and Discussion

The first task is to decide the range of the LNG mass and wind speed, as input variables, for the investigation. The LNG mass varies from 2 to 5 tons. According to the weather statistics data in 2011, provided by the Singapore National Environment Agency, the wind speed has been determined to be between 0.5 and 7.3 m/s. Given the available computational resources, a total of 24 simulations have been designed using HSS (see Figure 2). It can be observed that the designed wind speeds and LNG mass cover the input space very well. Subsequently, FLACS is used to obtain the 4D concentration values due to the dispersion of the LNG vapour cloud.

(Figure 2 about here)

For the development and validation of the meta-model, the 24 simulation runs need to be divided into modelling and validation data. The method of leave-one-out cross-validation (LOOCV) is used. In LOOCV, one simulation is reserved for validation, whilst the remaining 23 runs are used to develop the meta-model. This step is repeated until every simulation has been used once as validation data. When applying SegPCT-PCA, twenty principle components are retained and they explain over 99% of the variance of the original data containing 43,421,400 variables. It indicates that 20 principle components are sufficient to capture the essential information of the high-dimensional concentration variables. Therefore, these 20 PCA score variables are used as output for meta-modelling, and the predicted scores can be transformed, according to Eqs. (8)(9), to give the prediction for the high-dimensional concentration. The root mean square error (RMSE) is used to evaluate the predictive accuracy of the model. The RMSE is given in Table 1 for each run reserved as validation data, whereby the maximal value is 0.0033 (the 8th simulation). Even for the 8th simulation, the meta-model prediction satisfactorily agrees with the CFD simulation (shown later in Figure 3). Overall, the prediction accuracy for all 24 simulations appears to be satisfactory.

(Table 1 about here)

Out of the 24 simulations, the 23rd (wind speed: 6.73 m/s, LNG mass: 4.25 ton) and the 8th (wind speed: 2.48 m/s, LNG mass of 3.41 ton) runs are selected for demonstrating the results, for they give the lowest and highest RSME, respectively. For the 23rd simulation, the LEL contours predicted by the meta-model and from simulation are displayed in Figure 3. Notice that the original concentration data are within a 3D space. However, in practice, the vertical coordinate is less important, and only the

potential explosible zone in horizontal level is of concern, for example to help decide the evacuation range. Therefore, the 3D concentration data are projected onto 2D LEL contours with X, Y coordinates only. Specifically, if any concentration along the vertical Z-coordinate is greater than LEL, the corresponding X-Y location is deemed explosible. So far, the LEL contour is calculated from the mean prediction of the concentration; neither model nor input uncertainty is considered. In addition, the LEL contour, directly calculated from the concentration values, appears to be somewhat irregular in space (e.g. the FLACS simulation in Figure 3(b)). In principle, such irregular contours may be smoothed by using filtering methods as post-processing, which is outwith the scope of this study and thus not explored.

(Figure 3 about here)

The dispersion process of the 23rd simulation lasts about 145 s. It can be observed from Figure 3 that because of the west wind, the plume mainly moves to the east. However, at the initial stage (e.g. 20s as in Figure 3(a)), the plume also expands to other directions (even slightly to the west), due to the rapid and strong dispersion from the leakage location. Approximately at 100 s, the size of the plume reaches the maximum and begins to reduce. Visual inspection reveals that most plumes (Figure 3(a)-(e)) are very well predicted. However at 130 s (Figure 3(f)), the predicted contour does not match well with the CFD simulation, especially at the east end. Nevertheless, the prediction accuracy in Figure 3(f), in terms of RMSE of concentration, is similar to other plumes. This is because that at 130 s, the concentration in the plume tends to be close to the LEL, and thus small prediction errors in concentration cause significant error of the LEL contour.

Similar LEL contours for the 8th simulation are illustrated in Figure 4. Although the mass of the released LNG in the 8th simulation is similar to that of the 23rd, the dispersion process lasts longer to about 360 s due to lower wind speed. The predicted LEL contours match reasonably well with the CFD simulation until towards the end of the process (from around 250s). This phenomenon, and its underlying cause, is similar to that for the 23rd simulation as discussed above.

(Figure 4 about here)

Next, uncertainty analysis is carried out to investigate the prediction variability, which can be important information to aid subsequent risk assessment and mitigation decisions. The results for the 23rd simulation are used for illustration. The standard deviation of the inputs is assumed to be 10% of the corresponding values, reflecting significant input uncertainty. One thousand MC samples are generated for the inputs, shown in Figure 5.

(Figure 5 about here)

By using the GPR meta-model, one prediction can be calculated for each MC sample of the inputs, resulting in a total of 1000 predictions of the 4D concentration field. These 1000 predictions are then combined as in Eqs. (10)(11), whereby Eq. (11) gives the prediction variance due to input uncertainty

and model uncertainty. In majority of the cases, input uncertainty contributes more, than model uncertainty, to the prediction variance; this is illustrated in Figure 6(a). Since it is not possible to display all 43,421,400 concentration values, only 50 concentrations at time 40 s and spatial range $\{0 < X < 100, Y = 0, Z = 0\}$ are shown in Figure 6(a). The corresponding 2D LEL contour of FLACS, input uncertainty, model uncertainty and combined uncertainty at 5% probability is give in Figure 6(b). However in a small number of cases, the model uncertainty does contribute more to the prediction variance (not displayed in Figure 6).

(Figure 6 about here)

Figure 7 gives the predicted LEL contour by meta-model, against that from FLACS simulation, for the 23rd simulation at 60 s. As expected, the predicted LEL region increases with decrease in explosion probability. In this case, the prediction matches the simulation reasonably well when no uncertainty is considered. Nevertheless, the inclusion of uncertainty can provide more confidence in the predicted potential explosive area.

(Figure 7 about here)

In addition, it is found that uncertainty analysis can compensate for under-predicted LEL contour towards the end of dispersion, when the concentration is close to LEL. As an example, Figure 8 illustrates the 23rd simulation at 130 s. When no uncertainty is considered, the meta-model under-predicts the concentration in the plume, resulting in smaller LEL contour than that from simulation. By including uncertainty, a decrease in explosion probability provides larger LEL contour, and eventually covers the entire explosible region as calculated by FLACS. Compared with Figure 7, it appears that the uncertainty at 130 s has more influence on the prediction than it does at 60 s. This may be explained by the overall higher concentration in the middle of the dispersion process than that towards the end; a snapshot is given in Figure 9. Therefore, the variance does not affect the concentration very much at 60 s, whilst small variance results in significant change in the prediction at 130 s.

(Figure 8 about here)

(Figure 9 about here)

As a further note, the current use of meta-modelling does not involve complex wind or temperature fields, but only fixed wind and temperature values. For this particular scenario, since the simulation time is relatively short (around 180 s), stable wind and temperature may be a reasonable approximation. Arguably for such a scenario, satisfactory simulation results may be obtained by using simpler models, such as SLAB (Ermak 1990), HEGADAS (Colenbrander 1980), DEGADIS (Spicer et al. 1989), HGSYSTEM (Witlox & McFarlane 1994), PHAST (DNV 2007), ALOHA (EPA 2006), SCIPUFF (Sykes et al. 2007) and TRACE (Systems 2009). Nevertheless, the aim of this study is not

to find the best simulation tool for the particular vapour dispersion scenario; instead it is to find an *enabling method* (i.e. meta-modelling) so that *CFD* can be used for fast analysis of vapour dispersion. In this sense, the present work is an initial step towards this aim, and future development is being planned for meta-modelling of more complex scenario with spatial and temporal variations in input variables. As such, comparisons between meta-modelling and other simplified simulation models are not within the scope of this paper. We refer the readers to a comprehensive validation of CFD model for dispersion (Middha et al. 2009), and comparisons between CFD and above-mentioned simplified models in simulating complex scenarios (Hanna, Hansen et al. 2009, Tauseef et al. 2011).

4. Concluding remarks

In this work, a meta-modelling method using Gaussian process regression is applied to an LNG dispersion process. Through investigation, it is demonstrated that the predictions of vapor dispersion process by meta-modelling via Gaussian process regression are successfully coordinated with the simulations by FLACS in most situations. Thus, given new mass and wind speeds, the concentration fields can be estimated in a very short time. In order to conduct risk assessment, uncertainty analysis is also investigated. Through uncertainty analysis, it is more efficient to analyze and decide the potential explosion range. Uncertainty analysis can also solve the under-estimation problem at the end of dispersion process when the concentrations are close to LEL. Therefore, the proposed uncertainty analysis method is a powerful tool to identify the explosible regions with different degrees of confidence.

List of Abbreviations

CFD	Computational Fluid Dynamic
DoE	Design of Experiments
FLACS	FLame ACceleration Simulator
GPR	Gaussian Process Regression
HSS	Hammersley Sequence Sampling
LEL	Lower Explosive Limit
LNG	Liquefied Natural Gas
MC	Monte Carlo
PCA	Principal Component Analysis
<i>pdf</i>	probability density function
RMSE	Root Mean Square Error
SegPCT-PCA	Segmented Principal Component Transform-Principal Component Analysis
UEL	Upper Explosive Limit

Acknowledgement

This work was supported by the Singapore DSO National Laboratories. The authors are grateful to DSO for providing the simulation scenarios in FLACS and the weather data, which were purchased from the Singapore National Environment Agency.

References

- Barratt, R. (2001). *Atmospheric dispersion modelling: an introduction to practical applications*. London, UK, Earthscan.
- Barros, A. S. & Rutledge, D. N. (2005). Segmented principal component transform–principal component analysis. *Chemometrics and Intelligent Laboratory Systems*. 78(1–2): 125-137.
- Chen, T., Hadinoto, K., Yan, W. J. & Ma, Y. F. (2011). Efficient meta-modelling of complex process simulations with time-space-dependent outputs. *Computers & Chemical Engineering*. 35(3): 502-509.
- Chen, V. C. P., Tsui, K.-L., Barton, R. R. & Meckesheimer, M. (2006). A review on design, modeling and applications of computer experiments. *IIE Transactions*. 38(4): 273-291.
- Coetzee, W., Coetzer, R. L. J. & Rawatlal, R. (2012). Response surface strategies in constructing statistical bubble flow models for the development of a novel bubble column simulation approach. *Computers & Chemical Engineering*. 36(0): 22-34.
- Colenbrander, G. (1980). A mathematical model for the transient behaviour of dense vapour clouds. 3rd International Symposium on Loss Prevention and Safety Promotion in the Process Industries, Basel, Switzerland.
- DNV (2007). PHAST professional user manual. *Oslo: DNV Software*.
- Eisenhower, B., O'Neill, Z., Narayanan, S., Fonoberov, V. A. & Mezić, I. (2012). A methodology for meta-model based optimization in building energy models. *Energy and Buildings*. 47(0): 292-301.
- EPA, N. (2006). ALOHA user's manual, US.
- Ermak, D. L. (1990). *User's manual for SLAB: An atmospheric dispersion model for denser-than-air releases*, Lawrence Livermore Laboratory.
- Fang, K.-T., Lin, D. K. J., Winker, P. & Zhang, Y. (2000). Uniform Design: Theory and Application. *Technometrics*. 42(3): 237-248.
- Gavelli, F., Bullister, E. & Kytomaa, H. (2008). Application of CFD (Fluent) to LNG spills into geometrically complex environments. *Journal of Hazardous Materials*. 159(1): 158-168.
- Geladi, P. & Kowalski, B. R. (1986). Partial Least-Squares Regression - a Tutorial. *Analytica Chimica Acta*. 185: 1-17.
- Girard, A., Rasmussen, C. E., Quinero-Candela, J. & Murray-Smith, R. (2003). Gaussian Process priors with uncertain inputs? Application to multiple-step ahead time series forecasting. *Neural Information Processing Systems*. Vancouver, Canada, MIT Press.
- Graham, A. & Simon, C. (2012). Vapour cloud formation: Experiments and modelling. *HSE Research report 908*.
- Hanna, S. R., Hansen, O. R., Ichard, M. & Strimaitis, D. (2009). CFD model simulation of dispersion from chlorine railcar releases in industrial and urban areas. *Atmospheric Environment*. 43(2): 262-270.
- Hansen, O. R., Gavelli, F., Ichard, M. & Davis, S. G. (2010). Validation of FLACS against experimental data sets from the model evaluation database for LNG vapor dispersion. *Journal of Loss Prevention in the Process Industries*. 23(6): 857-877.
- Jolliffe, I. (2005). Principal Component Analysis. *Encyclopedia of Statistics in Behavioral Science*, John Wiley & Sons, Ltd.

- Kalagnanam, J. R. & Diwekar, U. M. (1997). An efficient sampling technique for off-line quality control. *Technometrics*. 39(3): 308-319.
- Kalos, M. H. & Whitlock, P. A. (2008). *Monte Carlo Methods [electronic resource]*, Hoboken : Wiley-VCH, 2008.
- Le Chatelier, H. (1891). Estimation of Firedamp by Flammability Limits. *Ann. Mines*. 19(8): 388-395.
- Manfren, M., Aste, N. & Moshksar, R. (2013). Calibration and uncertainty analysis for computer models – A meta-model based approach for integrated building energy simulation. *Applied Energy*. 103(0): 627-641.
- Marjavaara, B. D., Lundstrom, T. S., Goel, T., Mack, Y. & Shyy, W. (2007). Hydraulic Turbine Diffuser Shape Optimization by Multiple Surrogate Model Approximations of Pareto Fronts. *Journal of Fluids Engineering*. 129(9): 1228-1240.
- Middha, P., Hansen, O. R. & Storvik, I. E. (2009). Validation of CFD-model for hydrogen dispersion. *Journal of Loss Prevention in the Process Industries*. 22(6): 1034-1038.
- Myers, R. H., Montgomery, D. C. & Anderson-Cook, C. M. (2009). *Response surface methodology : process and product optimization using designed experiments*, Hoboken, N.J. : Wiley, c2009.
- O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering & System Safety*. 91(10–11): 1290-1300.
- Oakley, J. & O'Hagan, A. (2002). Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika*. 89(4): 769-784.
- Palmer, K. & Realf, M. (2002). Metamodeling Approach to Optimization of Steady-State Flowsheet Simulations: Model Generation. *Chemical Engineering Research and Design*. 80(7): 760-772.
- Qiao, A. & Zhang, S. (2010). Advanced CFD modeling on vapor dispersion and vapor cloud explosion. *Journal of Loss Prevention in the Process Industries*. 23(6): 843-848.
- Rasmussen, C. E. & Williams, C. K. I. (2006). Gaussian processes for machine learning. *Gaussian Processes for Machine Learning*, MIT press: Cambridge, MA.
- Sacks, J., Welch, W. J., Toby, J. M. & Wynn, H. P. (1989). Design and Analysis of Computer Experiments. *Statistical Science*. 4(4): 409-423.
- Shemdin, O. H. (1971). Wind-generated current and phase speeds of wind waves. *Bulletin of the American Meteorological Society*. 52(9): 537-554.
- Shi, J. Q., Murray-Smith, R. & Titterton, D. M. (2003). Bayesian regression and classification using mixtures of Gaussian processes. *International Journal of Adaptive Control and Signal Processing*. 17(2): 149-161.
- Shi, J. Q., Murray-Smith, R. & Titterton, D. M. (2005). Hierarchical Gaussian process mixtures for regression. *Statistics and Computing*. 15(1): 31-41.
- Smith, E. (2006). Uncertainty Analysis. *Encyclopedia of Environmetrics*, John Wiley & Sons, Ltd.
- Spicer, T., Havens, J. & Guinnup, D. (1989). *User's guide for the DEGADIS 2.1 dense gas dispersion model*, US Environmental Protection Agency.
- Sykes, R. I., Parker, S. F., Henn, D. S. & Chowdhury, B. (2007). SCIPUFF version 2.3, technical documentation. *Princeton: L-3 Communications Titan Corporation*.

- Systems, S. (2009). Engineering-grade chemical emergency response management solution for risk analysis, prediction, and evaluation, BR0004-0807-3. *Camarillo: SAFER Systems LLC*.
- Tauseef, S. M., Rashtchian, D. & Abbasi, S. A. (2011). CFD-based simulation of dense gas dispersion in presence of obstacles. *Journal of Loss Prevention in the Process Industries*. 24(4): 371-376.
- Wang, K., Chi, G. Y., Lau, R. & Chen, T. (2011). Multivariate calibration of near infrared spectroscopy in the presence of light scattering effect: a comparative study. *Analytical Letters*. 44(5): 824 - 836
- Witlox, H. & McFarlane, K. (1994). Interfacing dispersion models in the HGSYSTEM hazard-assessment package. *Atmospheric Environment*. 28(18): 2947-2962.
- Wold, S., Esbensen, K. & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*. 2(1-3): 37-52.
- Yan, W. J., Chen, Y. T., Yang, Y. H. & Chen, T. (2011). Development of high performance catalysts for CO oxidation using data-based modeling. *Catalysis Today*. 174(1): 127-134.

Table 1. Root mean square error (RMSE) for the concentration prediction of 24 simulations. The range for the inputs are: wind speed (0.5-7.5 m/s), LNG weight (2-5 ton).

Run No.	Wind Speed	LNG Weight	RMSE	Run No.	Wind Speed	LNG Weight	RMSE
1	0.50	4.72	0.0026	13	3.90	4.44	0.0029
2	0.78	2.28	0.0027	14	4.18	2.56	0.0017
3	1.07	3.78	0.0032	15	4.47	4.06	0.0022
4	1.35	3.03	0.0019	16	4.75	3.31	0.0013
5	1.63	4.53	0.0026	17	5.03	4.81	0.0024
6	1.92	2.66	0.0027	18	5.32	2.38	0.0017
7	2.20	4.16	0.0032	19	5.60	3.88	0.0014
8	2.48	3.41	0.0033	20	5.88	3.13	0.0014
9	2.77	4.91	0.0028	21	6.17	4.63	0.0012
10	3.05	2.19	0.0021	22	6.45	2.75	0.0012
11	3.33	3.69	0.0025	23	6.73	4.25	0.0010
12	3.62	2.94	0.0021	24	7.02	3.50	0.0014

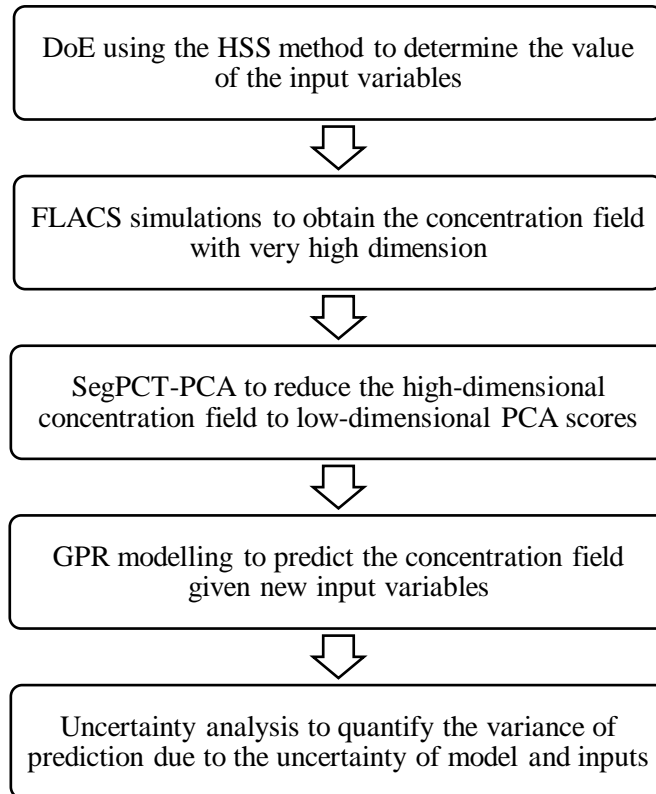


Figure 1. The flowchart of the overall methodology.

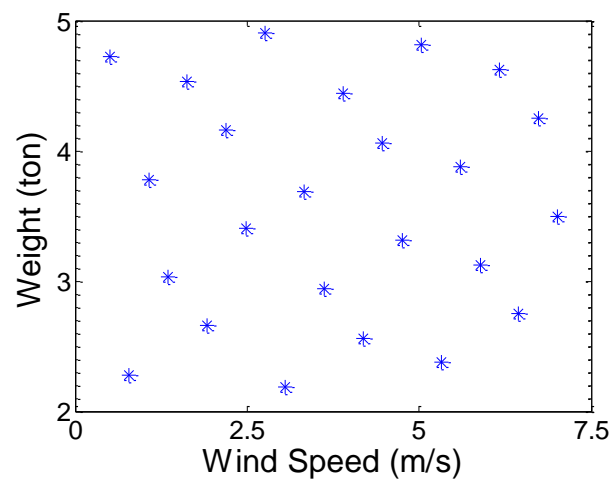


Figure 2. The designed simulation runs, in terms of the combination of wind speed and weight as input variables, based on the HSS method.

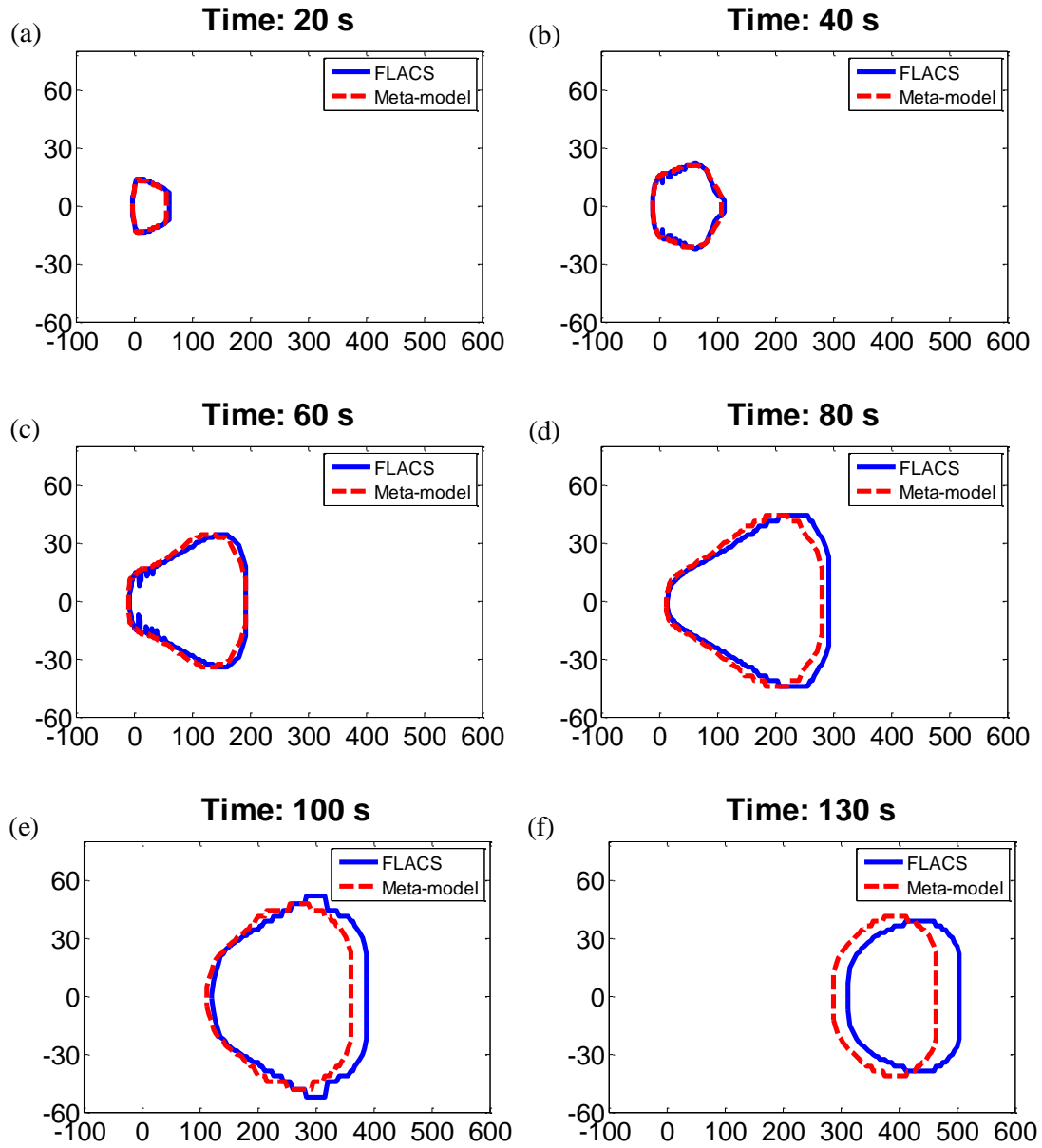


Figure 3. The meta-model predicted, and CFD simulated, LEL contours for the 23rd simulation (wind speed: 6.73 m/s, LNG weight: 4.25 ton).

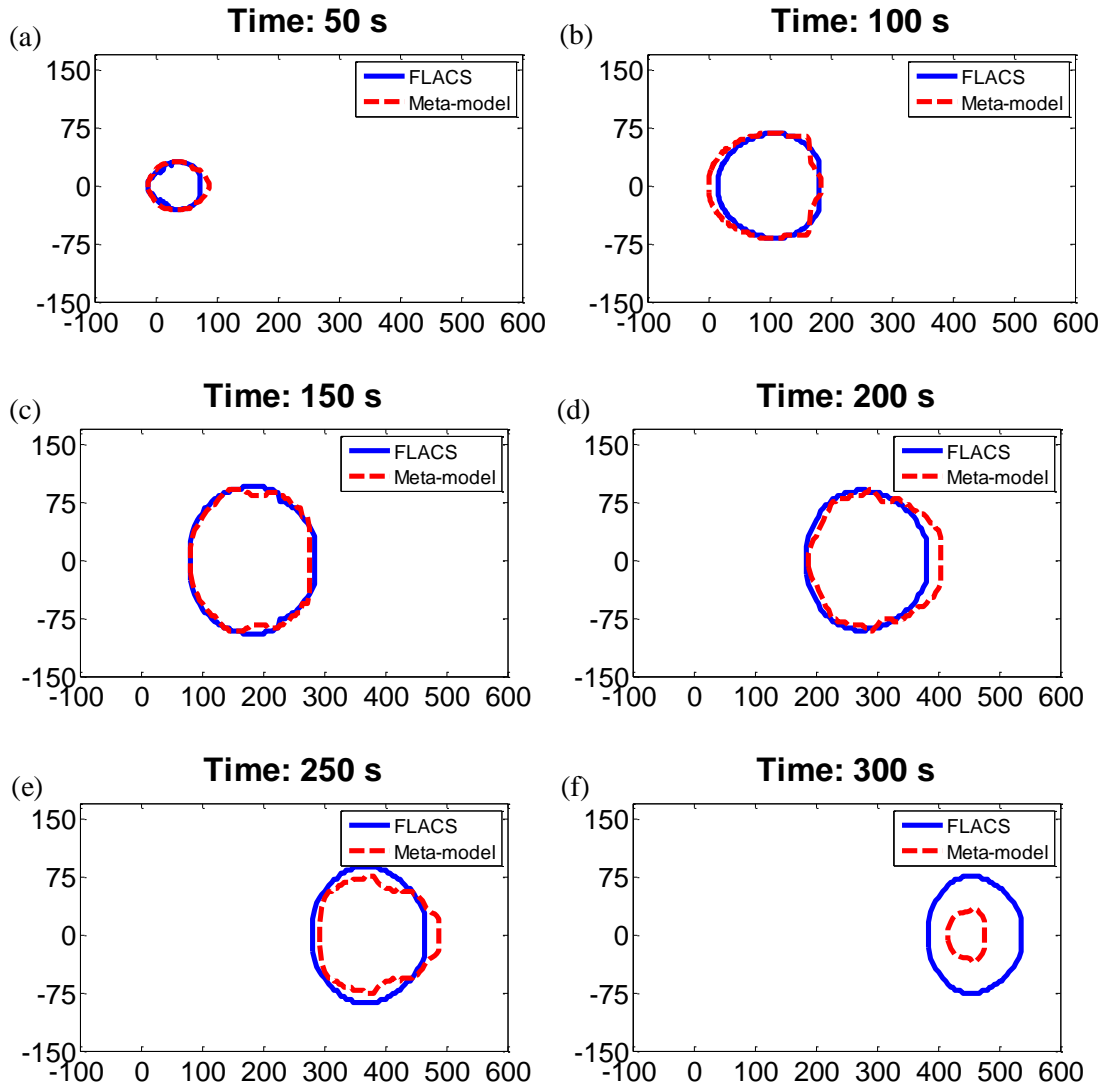


Figure 4. The meta-model predicted, and CFD simulated, LEL contours for the 8th simulation (wind speed: 2.48 m/s, LNG weight: 3.41 ton).

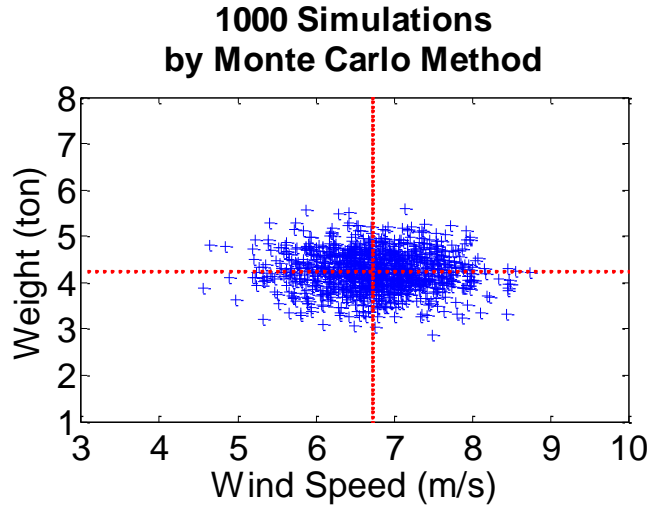


Figure 5. The 1000 MC samples generated to represent the input uncertainty for the 23rd simulation.

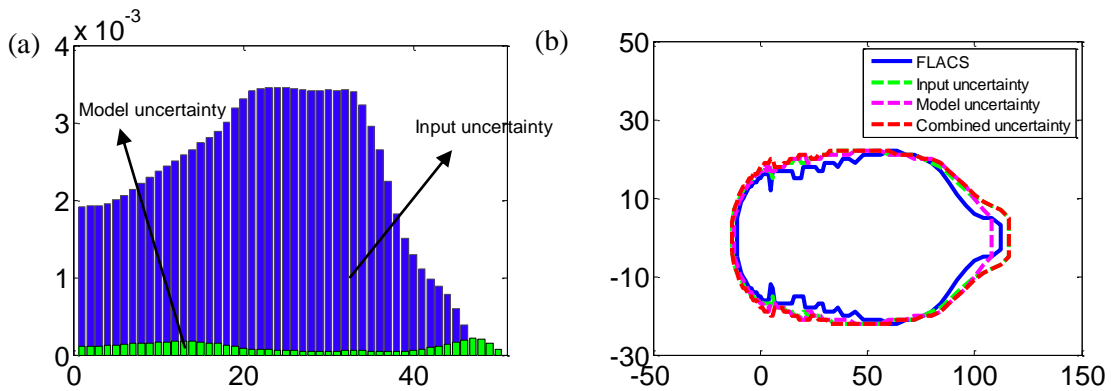


Figure 6. Uncertainty analysis for the 23rd simulation at 40 s. (a): Input uncertainty of 50 concentrations at time 40 s and spatial range of $\{0 < X < 100, Y = 0, Z = 0\}$; (b): LEL contours from FLACS, input uncertainty, model uncertainty and combined uncertainty at 5% probability.

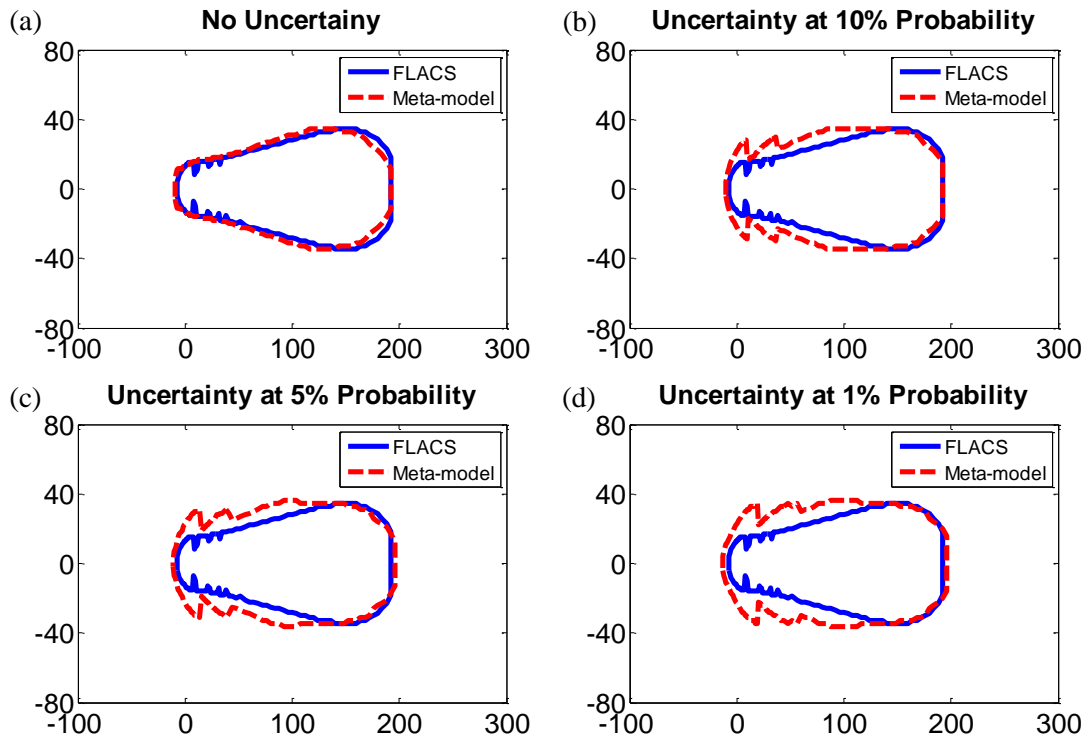


Figure 7. The uncertainty analysis at different probabilities for the 23rd simulation at 60 s.

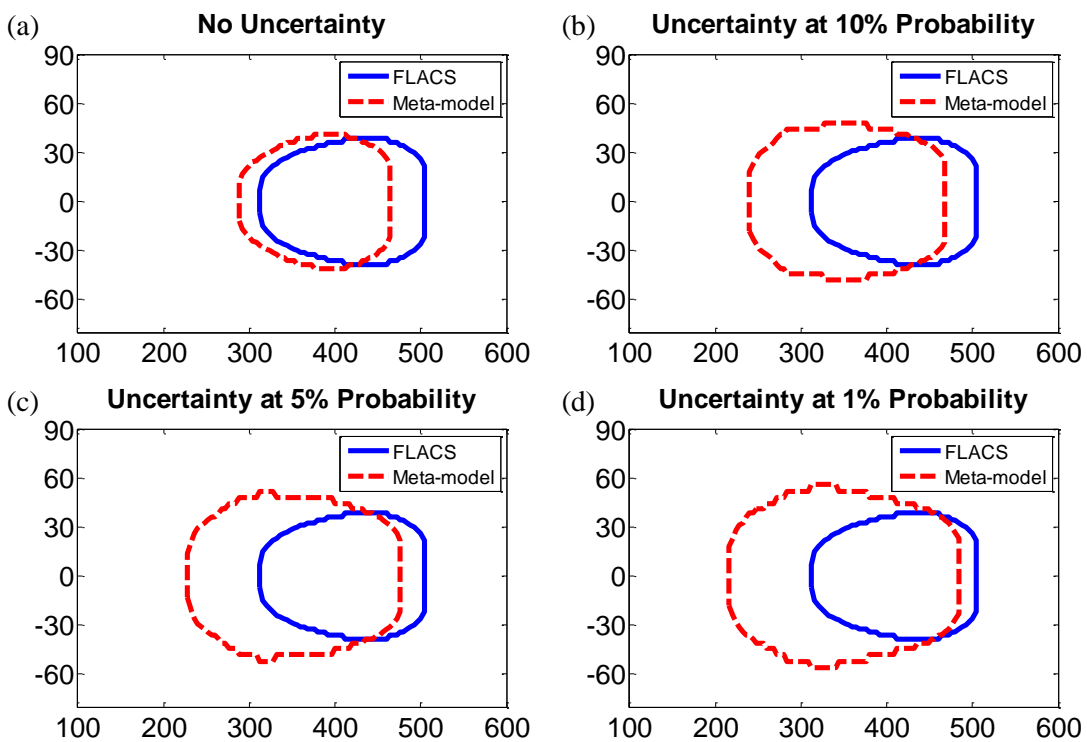


Figure 8. The uncertainty analysis at different probabilities for the 23rd simulation at 120 s.

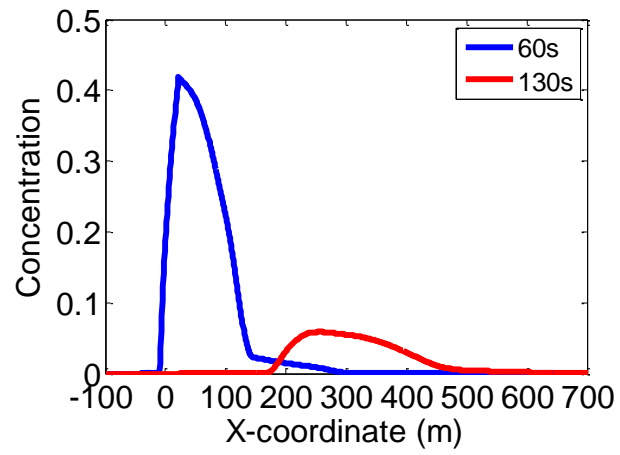


Figure 9. The comparison of concentration at 60 s and 120 s along the X-coordinate (Y=0, Z=0).