

# Wind Engineering Joint Usage/Research Center FY2019 Research Result Report

Research Field: Outdoor Environment  
Research Year: FY2019  
Research Number:  
Research Theme: Neighborhood scale source term estimation of atmospheric releases in urban areas using LES approach

Representative Researcher: Prof. Xiaofeng Li

Budget [FY2019]: 400,000 Yen

- \*There is no limitation of the number of pages of this report.
- \*Figures can be included to the report and they can also be colored.
- \*Submitted reports will be uploaded to the JURC Homepage.

## 1. Research Aim

Source term estimation (STE) of the releases of hazardous materials into the atmosphere refers to the identification of the source information, e.g., strength and location, based on limited and noisy concentration data. This process can be viewed as an assimilation process of the observed concentration data and the predicted concentration data. When dealing with releases in built-up areas, the predicted data are generally obtained by the Reynolds-averaged Navier-Stokes (RANS) equations, which yields building-resolving results; however, RANS-based models are outperformed by large-eddy simulation (LES) in the predictions of both airflow and dispersion. Therefore, it is important to explore the possibility of improving the estimation of the source parameters by using the LES approach. In this research, a novel source term estimation method is proposed based on LES approach using Bayesian inference.

## 2. Method

### 2.1 *Bayesian Inference*

Bayesian inference coupled with CFD approach was first used to identify urban releases by Keats, et al. [1], who provided the fundamentals of the method. Following their work, in this research, the LES approach is employed to improve the accuracy of the estimation results. The proposed method is demonstrated using a basic case of a single point source with a constant releasing strength; however, by combining this with other Bayesian STE methods, it can be generalized to address multiple point sources and variable releasing strengths.

#### 2.1.1 *Problem Formulation*

In the Bayesian inference, the STE problems are addressed by using a probabilistic

logic. Let  $\boldsymbol{\theta}$  denote the set of unknown source parameters, and  $\mathbf{d}$  denote the measurement data of a network of sensors. The estimation results can be obtained by calculating the posterior probability based on the Bayes' theorem:

$$p(\boldsymbol{\theta}|\mathbf{d}) = \frac{p(\mathbf{d}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{d})} \propto p(\mathbf{d}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \quad (1)$$

where  $p(\mathbf{d}|\boldsymbol{\theta})$  is the likelihood function,  $p(\boldsymbol{\theta})$  is the prior probability which represents all the a priori knowledge about  $\boldsymbol{\theta}$ ,  $p(\mathbf{d})$  is the marginal probability which acts as a normalizing factor and does not affect the relative probabilities, and  $p(\boldsymbol{\theta}|\mathbf{d})$  is the posterior probability which is the quantity of interest in the STE problems: the probability of the source parameters  $\boldsymbol{\theta}$  given the observations  $\mathbf{d}$ . To obtain  $p(\boldsymbol{\theta}|\mathbf{d})$ , appropriate function forms need to be assigned to  $p(\mathbf{d}|\boldsymbol{\theta})$  and  $p(\boldsymbol{\theta})$ .

### 2.1.2 Likelihood function

Here, a single point source with a constant releasing strength is considered, denoted by  $\boldsymbol{\theta} = (\mathbf{x}_s, q)$ , where  $\mathbf{x}_s$  is the source location and  $q$  is the releasing strength. Assume the concentration measurements are provided by a network of  $M$  sensors. The relationship between the observations and the predictions is then described by:

$$\mathbf{d} = q\mathbf{h}(\mathbf{x}_s) + \boldsymbol{\varepsilon} \quad (2)$$

where  $\mathbf{d} \in \mathbb{R}^M$  is the vector of concentration measurement data,  $\mathbf{h}(\mathbf{x}_s) \in \mathbb{R}^M$  is the source-receptor relationship vector representing the predicted mean concentrations at the sensors if a unitary release is made at location  $\mathbf{x}_s$ , and  $\boldsymbol{\varepsilon} \in \mathbb{R}^M$  is the error vector, which comprises both measurement and model errors.

The likelihood function describes the distribution of  $\boldsymbol{\varepsilon}$  and has different forms. Here, we assign the simplest and probably most frequently used form, the Gaussian distribution. The error vector  $\boldsymbol{\varepsilon}$  is assumed to be an independent, zero-mean Gaussian random vector, i.e.,  $\varepsilon_m \sim \mathcal{N}(0, \sigma_m^2)$ , which yields the likelihood function:

$$p(\mathbf{d}|\boldsymbol{\theta}) \propto \exp \left[ -\frac{1}{2} \sum_{m=1}^M \frac{(d_m - qh_m(\mathbf{x}_s))^2}{\sigma_m^2} \right] \quad (3)$$

### 2.1.3 Prior probability

The prior probability represents the knowledge about  $\boldsymbol{\theta}$  prior to receiving any measurements. Here, it is assumed that the source parameters are independent, indicating that one parameter implies nothing about the others:

$$p(\boldsymbol{\theta}) \equiv p(\mathbf{x}_s, q) = p(\mathbf{x}_s)p(q) \quad (4)$$

Generally, in a STE problem, the prior information of the source parameters is unavailable, so non-informative priors are used for both parameters. Namely, for  $\mathbf{x}_s$ , complete ignorance of the location parameters is to assign a uniform prior, assuming the source can be located anywhere with equal possibility in a specified potential source area,

as denoted by  $\Omega$ :

$$p(\mathbf{x}_s) \sim \mathcal{U}_\Omega(\mathbf{x}_s) \quad (5)$$

With respect to  $q$ , a scale parameter, the Jeffreys prior is assigned, which remains invariant under transformations of scale:

$$p(q) \propto 1/q, \quad q \in [q_{\min}, q_{\max}] \quad (6)$$

Using a scale-invariant prior ensures that  $p(q) = p(aq)$  for any constant  $a$ . The upper and lower bounds ensure that the prior probability density function is normalizable. In this study,  $[q_{\min}, q_{\max}] = [0.01, 100]$  L/min.

#### 2.1.4 Posterior probability and MCMC

Substituting Eq. (3-6) into Eq. (1) yields the explicit form of  $p(\boldsymbol{\theta}|\mathbf{d})$ . Despite the fact that its distribution can be directly obtained by numerical quadrature, it is computationally expensive due to the multidimensional parameter space. To reduce the computational cost, MCMC algorithms have been used [2, 3]. A set of sampling points are generated to approximate the distribution of  $p(\boldsymbol{\theta}|\mathbf{d})$  by the Metropolis-Hastings-within-Gibbs algorithm.

In this research, rejection sampling and slice sampling are employed to generate samples from the priors  $p(\mathbf{x}_s)$  and  $p(q)$ , respectively. 3 parallel Markov chains are generated to get more robust estimations. Each chain starts from a random location and has 10,000 total samples, among which the first 1,000 are discarded (known as the burn-in period). The remaining samples, 27,000 in total, are used as the approximation of the distribution of  $p(\boldsymbol{\theta}|\mathbf{d})$ .

## 2.2 Source-Receptor Relationship

The source-receptor relationship is the sensitivity of the concentration at each sensor to a given source location. This relationship contains the information of the dispersion model. Assume that the potential source area is discretized into  $N$  grid cells, and then the source-receptor relationship is a matrix composed of  $N$  source-receptor vectors:

$$\mathbf{H} = [\mathbf{h}(\mathbf{x}_{s,1}), \mathbf{h}(\mathbf{x}_{s,2}), \dots, \mathbf{h}(\mathbf{x}_{s,N})] \quad (7)$$

where  $\mathbf{H} \in \mathbb{R}^{M \times N}$  is a Jacobian matrix, where the  $(m,n)$ -th element is the predicted concentration at the  $m$ -th sensor given a unitary release at the  $n$ -th source location. By solving the advection-diffusion equations,  $\mathbf{H}$  can be calculated one column at a time,  $N$  times in total, after the simulation of the flow field. Normally, however, the sensors are heavily outnumbered by the possible locations, i.e.,  $M \ll N$ . By solving the adjoint equations,  $\mathbf{H}$  can be calculated row by row, thus the solution counts are markedly reduced from  $N$  to  $M$  [4].

### 2.2.1 Airflow model

The LES approach is used to simulate the flow field. Three-dimensional isothermal flow simulations are performed by an open source software program (OpenFOAM 2.2.1) to reproduce the airflow in built-up areas. The CFD settings follow the Architectural Institute of Japan (AIJ) guidelines for wind environment prediction [5]. In terms of the sub-grid scale (SGS) eddy viscosity model, the standard Smagorinsky model with the constant  $C_s = 0.12$  is used. The Van Driest damping function is applied in near-wall regions. The second-order central differencing scheme is applied to the spatial discretization. The second-order implicit scheme is specified for the time derivative. The governing equations are solved using the pressure-implicit split-operator (PISO) algorithm. It should be noted that although we choose version 2.1.1 of OpenFOAM instead of the latest version, all of the analysis used in this research can be performed equally with any later version.

### 2.2.2 Adjoint equation in time-averaged flow

The adjoint advection-diffusion equation is a useful approach for reducing the cost of STE problems since they only need to be solved once for each sensor to obtain the entire source-receptor relationship. Pudykiewicz [4] provided the detailed derivation of the adjoint equations. After the flow field is obtained by the airflow model, the adjoint equations in transient flow can be constructed by reversing the signs of the variation term and the advection term in the forward advection-diffusion equations. Theoretically, the adjoint equations in transient flow can be solved in a backward time-stepping; however, this requires storing the flow field of every time step, which costs an impractically huge amount of time and storage space due to the small time steps and high grid resolution in LES.

To overcome this difficulty, the adjoint equations are derived based on the time-averaged flow field of LES results, as follows

$$-\frac{\partial h_m}{\partial t} - u_i \frac{\partial h_m}{\partial x_i} - \frac{\partial}{\partial x_i} \left( D_e \frac{\partial h_m}{\partial x_i} \right) = r_m, \quad m = 1, 2, \dots, M \quad (8)$$

with the initial condition

$$h_m(t = T) = 0 \quad (9)$$

and the boundary conditions

$$\begin{aligned} h_m &= 0 && \text{at } \Gamma_1 \\ D_e \frac{\partial h_m}{\partial \mathbf{n}} + \sum_{i=1}^3 u_i \cdot \mathbf{n} h_m &= 0 && \text{at } \Gamma_2 \\ \frac{\partial h_m}{\partial \mathbf{n}} &= 0 && \text{at } \Gamma_3 \end{aligned} \quad (10)$$

where  $h_m$  is the adjoint corresponding to the  $m$ -th sensor,  $r_m$  acts like a unitary source at the  $m$ -th sensor location,  $u_i$  is the mean velocity vector,  $T$  is the latest time step in

the forward simulation,  $\Gamma_1$  represents the outflow boundary,  $\Gamma_2$  represents the inflow boundary,  $\Gamma_3$  represents the solid surfaces and the far field,  $\mathbf{n}$  is the unit vector normal to the boundaries, and  $D_e$  is the mean mass diffusivity, including the molecular diffusivity  $D$  and the turbulent diffusivity  $D_t$ , as  $D_e = D + D_t$ . In Eq. (8),  $u_i$  can be obtained by averaging the LES flow fields,  $D$  is a global parameter dependent on temperature, and thus  $D_t$  is the only parameter to be determined.

Combest, et al. [6] published a review on different approaches to model  $D_t$ , among which the simplest and most popular assumption is adopted here that there is a similarity between the turbulent mass diffusivity and turbulent momentum diffusivity (i.e., eddy viscosity  $\nu_t$ ). The similarity is characterized by assigning a global turbulent Schmidt number:

$$D_t = \nu_t / \text{Sc}_t \quad (11)$$

where the turbulent Schmidt number  $\text{Sc}_t$  is empirically assigned as 0.7 [7], and the eddy viscosity  $\nu_t$  can be estimated based on the mean velocity and the Reynolds stresses from the LES airflow results, following the method presented by Tominaga and Stathopoulos [8]. Based on the gradient diffusion hypothesis,  $\nu_t$  is given by Eq. (12-14).

$$\nu_t = - \sum_{i,j} R_{ij} S_{ij} / \sum_{i,j} 2S_{ij}^2 \quad (12)$$

where

$$R_{ij} = \langle u'_i u'_j \rangle - \frac{2}{3} k \delta_{ij} \quad (13)$$

is the deviatoric component of the Reynolds stress tensor,

$$S_{ij} = \frac{1}{2} \left( \frac{\partial \langle u_i \rangle}{\partial x_j} + \frac{\partial \langle u_j \rangle}{\partial x_i} \right) \quad (14)$$

is the mean strain rate,  $\langle u_i \rangle$  is the time-averaged velocity,  $\langle u'_i u'_j \rangle$  is the Reynold stress,  $k = \langle u'_i u'_i \rangle / 2$  is the turbulent kinetic energy,  $\delta_{ij}$  is the Kronecker delta, and the indices  $i$  and  $j$  both take on values 1, 2, or 3, denoting spatial components.

Substituting Eq. (11-14) into Eq. (8) yields the adjoint equations in the time-averaged flow which can be obtained by conducting LES without storing the results of every time step. Although the time-dependent fluctuations and the anisotropic flow behaviors are averaged out, this approach provides superior predictions of the mean flow field and the turbulent structure compared with RANS-based models, thus yielding an improved source-receptor relationship and consequently better estimation of the source parameters.

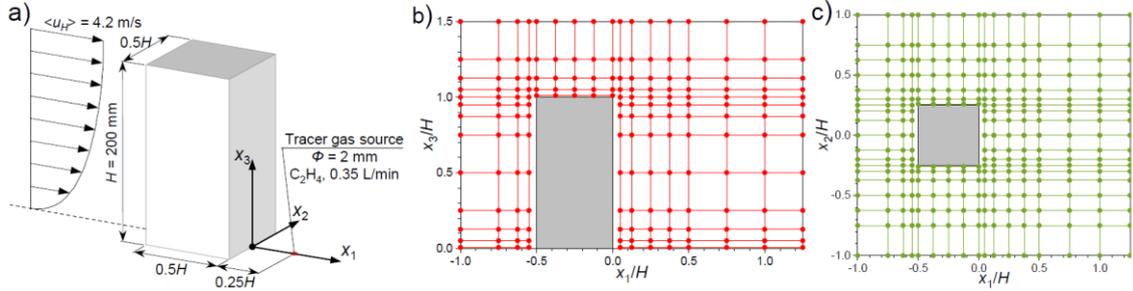
For a release with a constant strength, as considered in this research, the variation term is dropped, and Eq. (8) becomes steady in terms of time. Solving the steady-state adjoint equations gives the  $m$ -th row in the source-receptor relationship  $\mathbf{H}$ .

### 3. Research Result

#### 3.1 Case Description

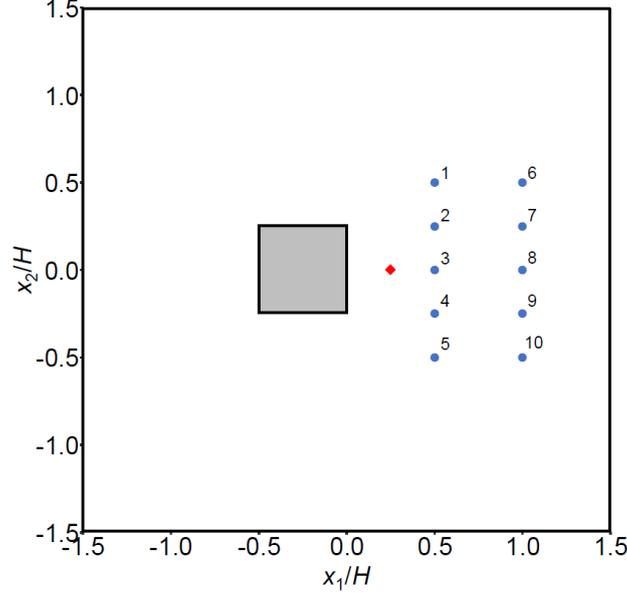
##### 3.1.1 Wind Tunnel Experiment

To validate the proposed method, we consider a three-dimensional dispersion scenario with a ground-level source, here  $\mathbf{x}_s = (x_1, x_2)$ , using an online wind tunnel dispersion experiment carried out at Tokyo Polytechnic University [9]. A wind tunnel with a test section 1.2 m wide and 1.0 m high was used to measure the concentration distributions of a continuous point tracer releasing scenario. As shown in Fig. 1 (a), a cuboid block of height  $H = 200$  mm, width  $W = 100$  mm and length  $L = 100$  mm was set as the building model. An atmospheric boundary layer flow approaching perpendicularly to the building was reproduced. Pure  $C_2H_4$  was released at a flow rate of  $q = 0.35$  L/min from a hole (with a diameter of 2 mm) on the ground in the wake region of the building, a challenging location to identify due to the dominant vortex and strong turbulence.



**Fig. 1.** Experimental setup (a) and measuring points: (b) vertical plane ( $x_2/H=0$ ) and (c) horizontal plane ( $x_3/H=0.0625$ ).

The velocity components and concentrations measured at the vertical symmetry plane ( $x_2/H = 0$ ) and a horizontal plane ( $x_3/H = 0.0625$ ) are used to evaluate the airflow models. The measuring points are shown in Fig. 1 (b) and (c). The wind velocity and gas concentration were simultaneously measured using a split film probe and a fast response flame ionization detector, respectively. The sampling frequency was set at 1,000 Hz, to obtain 120,000 data in 120 seconds. The mean concentrations measured at the 10 points on the horizontal plane are used to form the measurement vector  $\mathbf{d}$  to perform STE, as shown in Fig. 2. In addition, the profiles of velocity components of the approaching wind were measured to fit the inlet boundary conditions in the CFD simulation. The reference velocity  $\langle u_H \rangle = 4.2$  m/s is the upwind mean velocity at the building height  $H$ .



**Fig. 2.** Source term estimation configuration. The ground level source is denoted by the diamond. The 10 numbered dots depict sensors located at a height of  $0.0625H$ . The shown region is used as the possible source area.

### 3.1.2 Simulation settings

The proposed method is employed to estimate the source parameters based on LES approach, and its performance is compared to that of the existing method using RANS simulation data. The settings of both the LES and the RANS simulations are set up by referring to several best practice guidelines [5, 10] for the CFD simulations of flows and dispersion in the urban environment.

To make a fair comparison, the same computational domain and mesh are used. The size of the computational domain is  $12.5H \times 6H \times 4H$  (2.5 m  $\times$  1.2 m  $\times$  0.8 m). The distance between the inflow boundary and the building model is  $2H$ ; the distance between the outflow boundary and the model is  $10H$ ; and the distance between the lateral boundaries and the blocks is  $2.75H$ . A rectangular mesh is generated. The central domain near the blocks (the measured region) consists of the smallest elements, cubic grids with a uniform dimension of  $1/42H$ . The maximum length ratio outside the central domain is less than 1.05. The total number of computational cells is approximately 3.3 million. Grid independence is evaluated by the grid convergence index (GCI), following the procedures introduced by Hefny and Ooka [11]. The GCI values for LES and RANS are 0.41% and 0.31%, respectively, for doubled cell number, indicating that the given mesh arrangement is fine enough.

**LES settings.** The eddy viscosity model, discretization schemes, and solver follow the generic settings. The time step is set to 0.001 s. Before time averaging, a simulation

is conducted for 60 s to eliminate the impact of the initial condition; then, another 60 s simulation is performed to obtain the time-averaged flow data. The averaging period has been confirmed to be long enough to give steady values.

The lateral and upper boundaries are treated as symmetry planes. For the outflow boundary, zero gradients are specified for all variables. For the ground boundary and building surfaces, the Spalding's law of the wall is used [12]. In terms of inflow boundary, a separate LES computation is conducted to reproduce the turbulent inflow, or more specifically, the entire roughness fetch of the wind tunnel is simulated.

**RANS simulation settings.** The RANS equations are solved using the standard  $k\text{-}\varepsilon$  turbulence model. Shirasawa, et al. [13] reported the prediction accuracy of the standard  $k\text{-}\varepsilon$  model and revised  $k\text{-}\varepsilon$  models such as the Realizable, Renormalized Group, and Kato-Launder models for the flow and dispersion with a similar 2:1:1 building model. Compared with the standard  $k\text{-}\varepsilon$  model, the revised models overpredicted the size of the recirculation zone and showed lower accuracy in the wake region. Thus, the revised models are not used in the present study, because the source is located in the wake region of the building.

In the standard  $k\text{-}\varepsilon$  model, the eddy viscosity  $\nu_t$  required in Eq. (11) is modeled by:

$$\nu_t = C_\mu k^2 / \varepsilon \quad (15)$$

where  $C_\mu = 0.09$  is a model constant, and  $\varepsilon$  is the dissipation rate. The second-order total variation diminishing (TVD) discretization scheme is applied to all governing equations, which are solved using the semi-implicit method for pressure-linked equations (SIMPLE) algorithm.

Boundary conditions for the lateral, upper, and outflow boundaries are the same as those used in the LES. For the ground and building boundaries, the generalized log-law wall function is specified. The profiles of  $\langle u_1 \rangle$  and  $k$  at the inflow boundary are specified based on the wind tunnel measurements. The inflow values of  $\varepsilon$  are calculated by assuming local equilibrium between  $\varepsilon$  and the production term of  $k$ :

$$\varepsilon = C_\mu^2 k \frac{du}{dx_3} \quad (16)$$

### 3.2 Results and Discussion

In this section, the predicted airflow and concentration fields are compared with the experimental data, and the estimation results obtained by using both LES and RANS approaches are presented. The CFD simulations were conducted by a high-performance workstation and the Bayesian inference was performed on a laptop. The computational time and the computer information are summarized in Table 1. The LES approach requires more time than the RANS model, but in practice these time-consuming CFD

simulations can be completed before a pollutant release event occurs. In this pre-event stage, only simulations with different inflow directions need to be performed, since the inflow wind speed is inversely proportional to the source-receptor relationship. Thus, after a limited number of simulations, we can have a database with all of the source-receptor relationships that we need. Then, during an emergency or accidental release event, as data stream in from the sensors, the Bayesian inference can be conducted without running CFD simulations and finished in a very short time, even on a laptop.

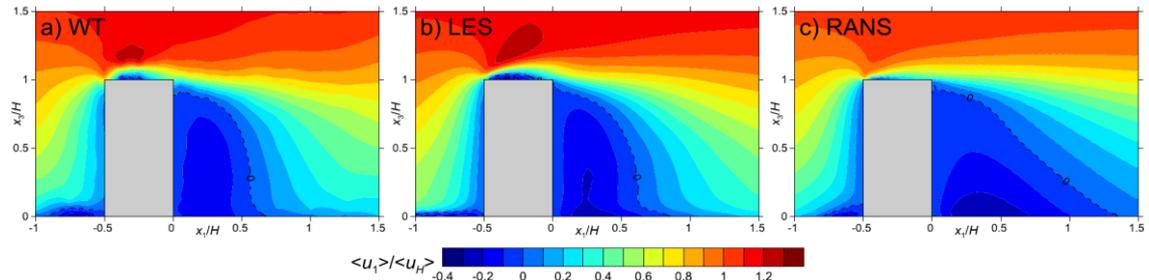
**Table 1**

Summary of the computational time and the computer information.

	LES	RANS	Computer information
Airflow simulation	67h 40min	3h 15min	Intel(R) Xeon(R) E5-2680 @ 2.70GHz (2 CPUs × 8 cores in parallel)
Adjoint equations	51min	1h 3min	
Bayesian inference	23s	23s	Intel(R) Core(TM) i7-3610QM @ 2.30GHz (3 cores in parallel)

### 3.2.1 Comparison between observed and predicted flow fields

The performance of the STE is dependent on the source-receptor relationship, which is solved based on the predicted flow field; therefore, the accuracy of the predicted flow field is critical to the success of the estimations. Here, the observed and predicted results of the mean velocity and the turbulent kinetic energy are compared.



**Fig. 3.** Mean streamwise wind velocity  $\langle u_1 \rangle$  of the vertical plane ( $x_2/H = 0$ ): (a) experiment, (b) LES, and (c) RANS. The dashed lines represent the contours of  $\langle u_1 \rangle = 0$ .

Fig. 3 depicts the vertical distributions of the mean streamwise wind velocity of the experiment and the simulations. The RANS model overestimates the size of the recirculation zone behind the building. In the near-field region, the predicted reverse flow is stronger than the observed flow, whereas the predicted flow behind the recirculation zone ( $x_1/H = 0.7 - 1.5$ ) is weaker. Compared with the RANS model, the LES provides a shorter recirculation zone with a shape similar to the one observed in the wind tunnel, and the velocity in and behind the recirculation zone is closer to the

experimental results.

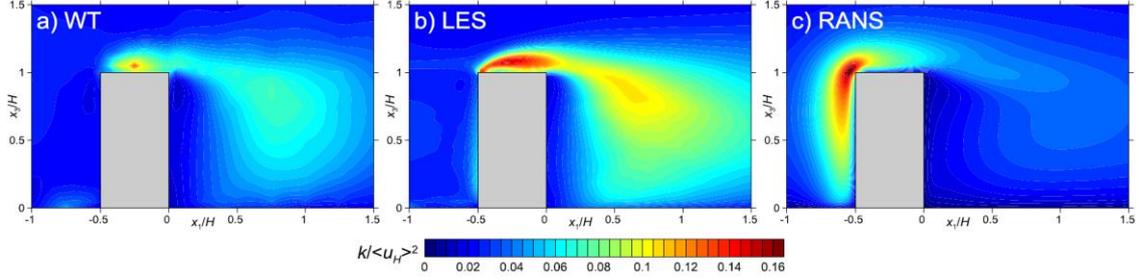
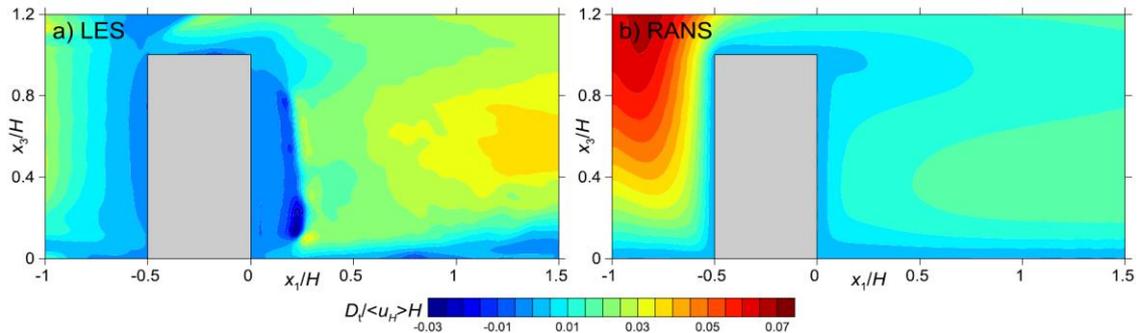


Fig. 4 shows the vertical distributions of the turbulent kinetic energy. The RANS model yields a false peak area near the windward face of the building and fails to predict the peak area in the recirculation zone behind the building. In addition, the RANS model gives a smaller value of  $k$  in the leeward of the building compared with the observed value. The LES provides closer predictions of  $k$  and the correct peak area in the downwind region, although in which an overestimation can be found.

Overall, the LES outperforms the RANS model in the predictions of both the mean flow field and the turbulent structure, especially in the wake region of the building, because the LES reproduces the periodic fluctuations due to vortex shedding, whereas the RANS model fails.

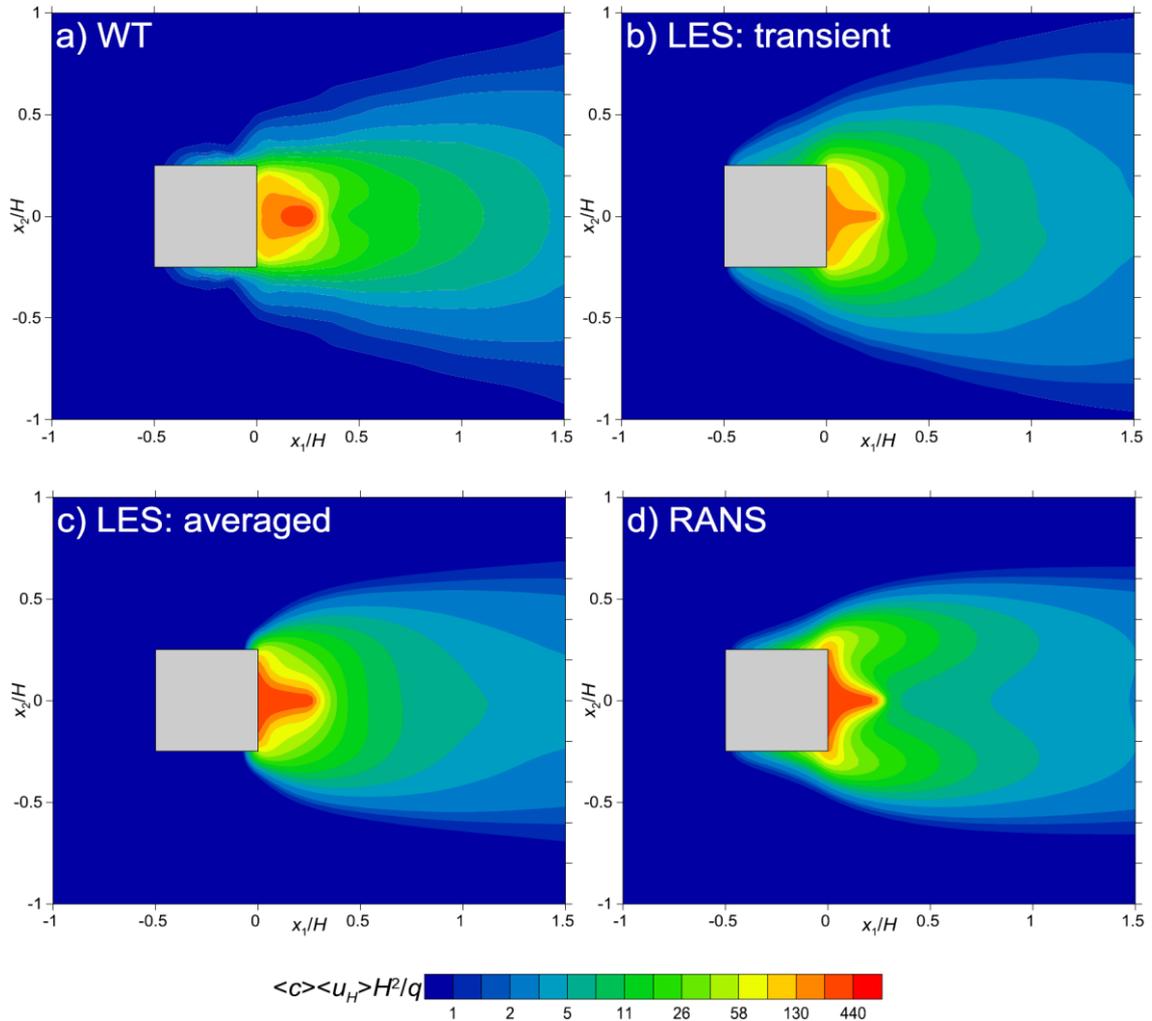
In addition, by applying the proposed method as expressed by Eq. (11-14), the LES results is used to estimate the turbulent diffusivity  $D_t$ , which is compared with the one calculated by the RANS model based on Eq. (11) and (15). The comparison is shown in Fig. 5. It is worth noting that the distribution obtained by the LES approach is not smooth because the mass transport in real scenarios is anisotropic. The peak area in the windward region of the RANS model is caused by the overestimation of  $k$  shown in Fig. 4 (c). In the leeward region, a similar trend of the distributions can be observed; however, the value given by the RANS model is smaller than that estimated by the LES approach. This underestimation is also because of the failure to reproduce the periodic fluctuations.



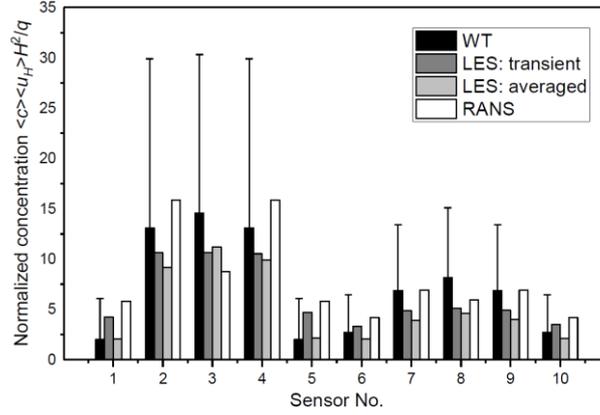
**Fig. 5.** Turbulent diffusivity  $D_t$  of the vertical plane ( $x_2/H = 0$ ): (a) estimated by using the LES data, and (b) calculated by the RANS model.

### 3.2.2 Comparison between observed and predicted concentrations

The different  $\langle u_i \rangle$  and  $D_t$  distributions lead to different predictions of concentration distributions. By setting the source properties as the true source, the predicted concentrations can be calculated by solving the forward advection-diffusion equations. The normalized concentrations on the horizontal plane obtained by different simulation methods are shown in Fig. 6, in comparison with the observed concentrations. Fig. 6 (b) shows the mean concentration distribution predicted by LES approach, solving the forward advection-diffusion equations by each time step. Fig. 6 (c) shows the concentrations obtained by solving the advection-diffusion equations constructed using the time-averaged flow simulated by the LES approach, which is the transport and dispersion model used in the proposed STE method. Fig. 6 (d) shows the concentrations predicted by the RANS model. The simulations based on LES yield better distribution patterns in the wake region compared with the RANS simulation. Besides, the horizontal dispersion is underestimated by the two approaches using time-averaged flow.



**Fig. 6.** Mean concentration  $\langle c \rangle$  of the horizontal plane ( $x_3/H = 0.0625$ ) obtained: (a) in wind tunnel experiment, (b) in transient flow of LES, (c) in time-averaged flow of LES and (d) using RANS.



**Fig. 7.** Comparison of the mean concentrations  $\langle c \rangle$  at the 10 sensor locations, obtained: (a) in wind tunnel experiment, (b) in transient flow of LES, (c) in time-averaged flow of LES and (d) using RANS. The error bars denote the standard deviations of the observed concentrations,  $\sigma_m$ .

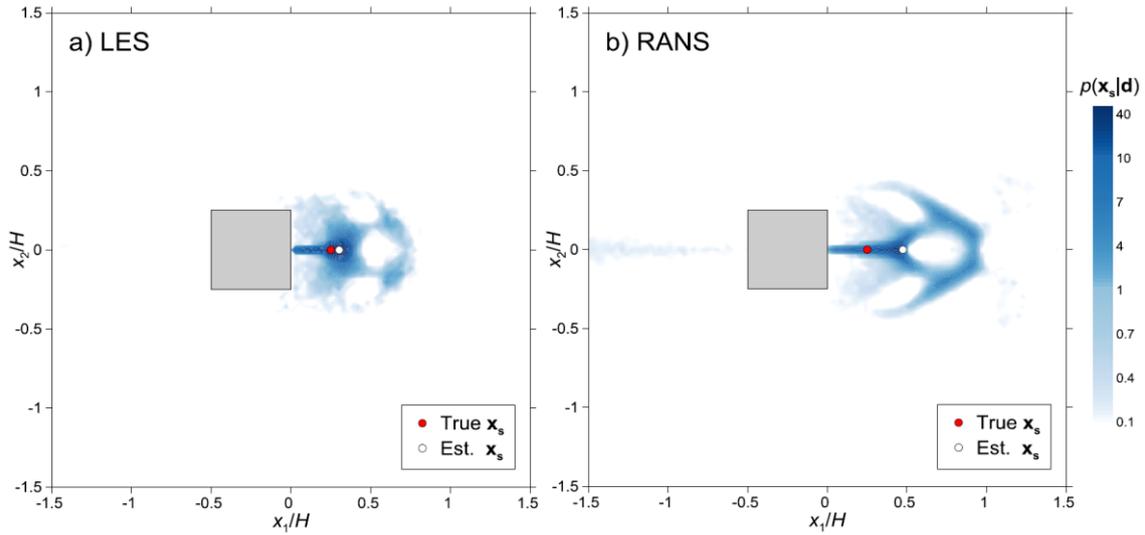
Fig. 7 shows the comparison between the observed and the predicted concentrations at the 10 sensor locations. The error bars indicate the standard deviations of the observed concentrations,  $\sigma_m$ . The concentrations near the center line ( $x_2/H = 0$ ) are underestimated in the transient flow of LES, while the lateral concentrations are overpredicted. In averaged flow of LES, the predictions are similar to the ones obtained in transient flow. With regard to RANS, the simulation yields higher concentrations at the lateral sensors, whereas the concentrations near the center line are underestimated.

### 3.2.3 Estimation results of source parameters

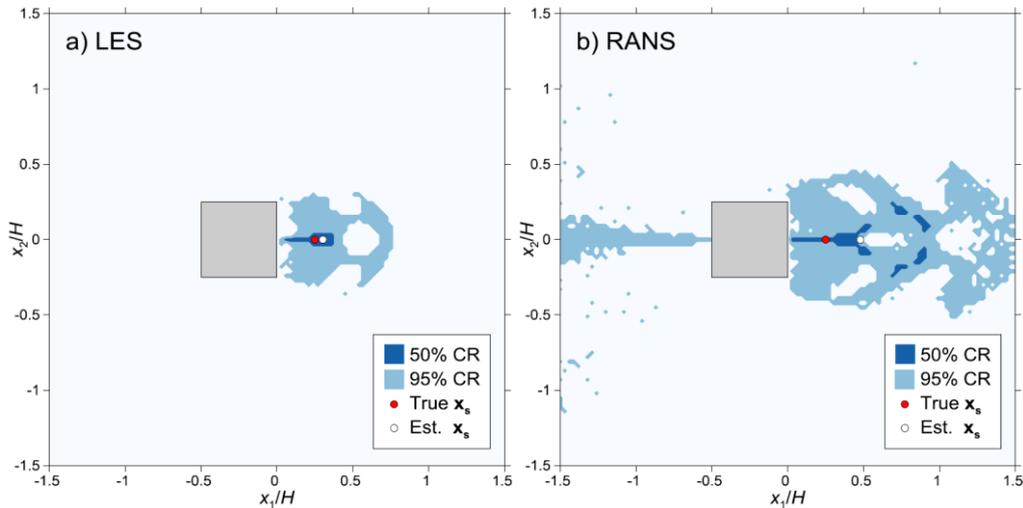
Different predictions of concentration result in different estimations of the source parameters. These differences in estimation are shown by keeping other STE inputs the same: time-averaged concentrations are used to form the measurement vector  $\mathbf{d}$ ; the standard deviations of noise  $\sigma_m$  are given by the standard deviations of measurements at the corresponding sensors, indicating that the errors are of the same magnitude as the turbulent fluctuations of the tracer concentration; and the settings of Bayesian inference are according to those described earlier. It should be noted that  $\sigma_m$  is both difficult to determine and influential for the estimation.

Fig. 8 depicts the marginal posterior distributions of the source location  $p(\mathbf{x}_s|\mathbf{d})$  obtained by using the LES and the RANS approaches, respectively. For the estimation using LES, a triangular-shaped area with high probability density centers around the true location (Fig. 8 (a)); and a more dispersed distribution can be observed for the estimation based on RANS simulation (Fig. 8 (b)). As a result, the inference using LES yields a close estimation of  $\mathbf{x}_s$ , with a slight deviation towards downwind, whereas the inference using RANS simulation provides a much farther estimation. To show the dispersions of the probability distribution, the 50% and 95% credible regions (CRs) are

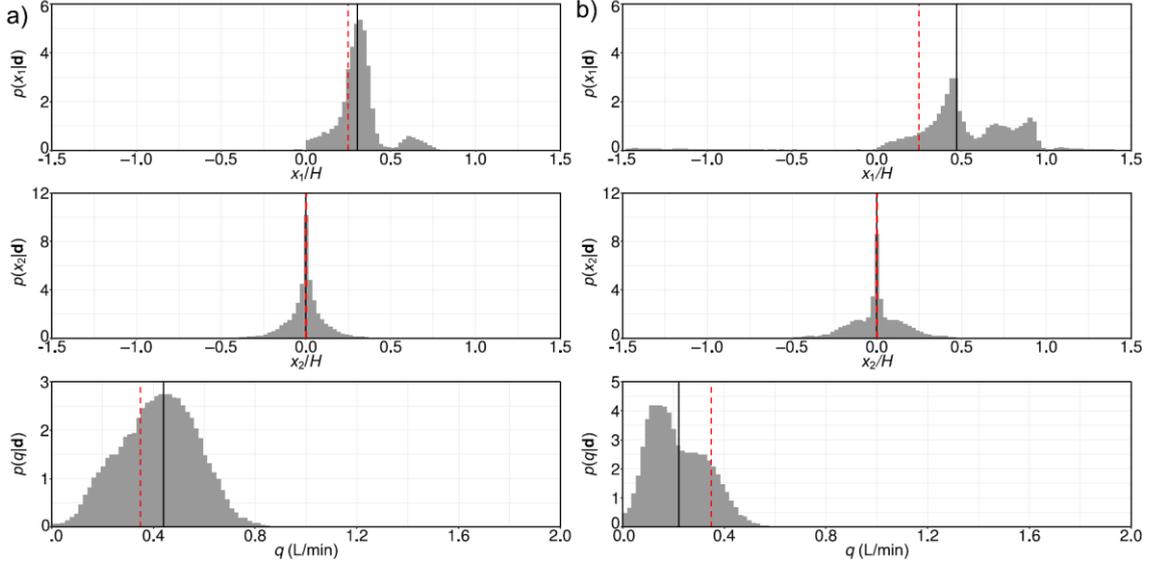
presented in Fig. 9. With respect to the estimation using LES shown in Fig. 9 (a), the 50% CR is a small area, of which the true source is located in the middle; the 95% CR consists of a bulk of probability mass and very few stray spots and is located entirely in the recirculation region, indicating that the source is most likely to be in the region, but somewhat less clear within it, due to the rapid mixing of tracer gas by the dominant vortices. In terms of the estimation using RANS simulation shown in Fig. 9 (b), the discontinuous 50% CR indicates several peaks of the probability distribution; and the 95% CR is much larger and more dispersed, with a large portion in the upwind region.



**Fig. 8.** The marginal posterior probability density distributions of the source location  $p(\mathbf{x}_s|\boldsymbol{\mu})$ : estimations using (a) LES, and (b) RANS simulation. The red and white dots denote the true and estimated source locations, respectively. Here, the posterior mean is selected as the point estimate.



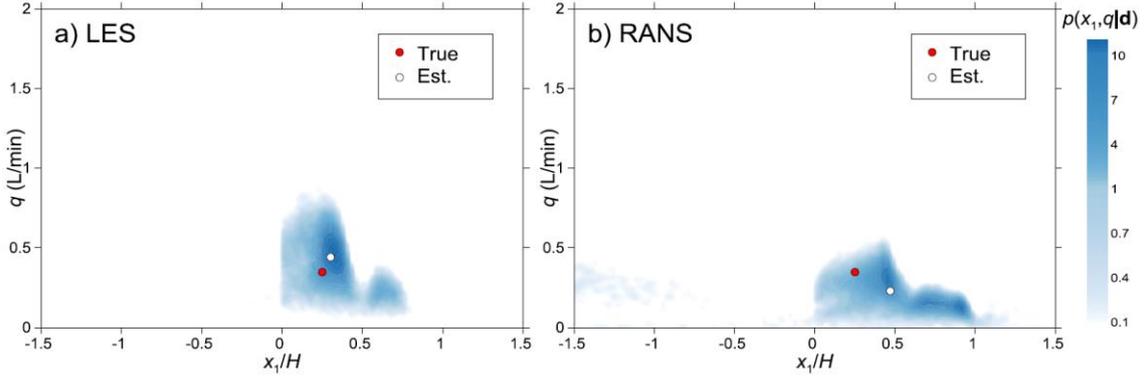
**Fig. 9.** The 50% and 95% credible regions (CRs), namely the highest posterior density regions, of the source location: estimations using (a) LES, and (b) RANS simulation. The red and white dots denote the true and estimated source locations, respectively.



**Fig. 10.** The marginal posterior probability distributions of source location ( $x_1$ - and  $x_2$ -coordinates) and strength  $q$ : estimations using (a) LES, and (b) RANS simulation. Solid lines and dashed lines denote the point estimates (posterior means) and true parameters, respectively.

Fig. 10 shows the marginal posterior distributions of the source parameters. For  $x_1$ , the distribution obtained by LES has one clear peak near the true source location, although there is another small peak around  $x_1/H = 0.60$ ; the distribution by RANS has three peaks, the largest one near  $x_1/H = 0.47$ . For  $x_2$ , both methods yield a symmetrical distribution centered at the true source location, but the distribution by LES is narrower. For  $q$ , the distribution by LES has a peak value equal to its point estimate; whereas the distribution by RANS has a longer tail, leading to a smaller peak value compared to its point estimate.

It is worth noting that the estimated  $x_1$  has a positive bias in both methods, while the estimated  $q$  has a positive bias in LES but a negative bias in RANS. The joint posterior probability density distributions between  $x_1$  and  $q$  are shown in Fig. 11 for a better understanding of such pattern. First, the source is located in the wake of the building, where the velocity direction is opposite to the inflow. Therefore, a slight positive bias of  $x_1$  means the source is farther from the sensors, so the estimated  $q$  increase in LES. However, in RANS, the positive bias of  $x_1$  becomes too large, even with a large probability of  $x_1/H > 0.5$ , which is upstream of the sensors No. 1 to 5 ( $x_1/H = 0.5$ ). As a result, the tracer is predicted to reach this sensor directly without diffusion in the wake, and a negative bias of  $q$  is occurred to meet the concentration measured at the sensors. Second, even if the true source location is given, the estimated source strength would still be positively biased in LES and negatively biased in RANS, because the concentrations are underestimated in LES and overpredicted in RANS. The above two reasons combined can be an explanation of the different bias directions of  $q$ .



**Fig. 11.** The joint posterior probability density distributions between  $x_1$  and  $q$ : estimations using (a) LES, and (b) RANS simulation.

The estimation results are summarized in Table 2. In general, the inference using LES outperforms the one using RANS simulation in the estimations of both the source location and the strength. Especially with respect to the source location, the inference using LES yields not only the better point estimation but also much smaller credible regions, indicating that the true location has been well identified with high confidence. To quantify the accuracy of the estimations, two indices are defined: the location error  $E_L$ , representing the Euclidian distance between the estimated source position and the true location; and the strength error  $E_q = q_{\text{est}}/q_{\text{true}} - 1$ , a ratio of estimated to true source strength. Compared with the one using RANS simulation, the inference using LES reduces  $E_L$  and  $E_q$  by 77% and 28.1%, respectively.

**Table 2**

Summary of the estimation results.

Method	True value		Source location estimation				Source strength estimation		
	$(x_1, x_2)^a$	$q^b$	$(x_1, x_2)^a$	50% CR <sup>c</sup>	95% CR <sup>c</sup>	$E_L^a$	$q^b$	95% CI <sup>b</sup>	$E_q$
LES	(0.25, 0.0)	0.35	(0.30, 0.00)	0.019	0.267	0.05	0.441	(0.147, 0.688)	0.261
RANS	(0.25, 0.0)	0.35	(0.47, 0.00)	0.049	1.045	0.22	0.223	(0.051, 0.479)	- 0.363

<sup>a</sup> Normalized by  $H$ .

<sup>b</sup> Unit: L/min.

<sup>c</sup> The sizes of the credible regions (normalized by  $H^2$ ).

#### 4. Conclusions

In this research, a novel STE method is proposed based on the LES approach. A Bayesian inference method, combined with the LES approach, the adjoint equations and the MCMC algorithm, is used to retrieve the source location and strength of a point source in a wind tunnel experiment, in which the source is a constant tracer release

located in the wake region of a building model. The source-receptor relationship is obtained by solving the adjoint equations constructed using the time-averaged flow simulated by LES approach. The performance of the proposed method is evaluated against the existing STE method which uses a RANS model to calculate the source-receptor relationship.

By comparing the predicted flow fields with the experimental measurements, the LES yields better results of both the mean flow field and the turbulent structure, especially in the wake region of the building, relative to the RANS simulation. One of the reasons is that the LES reproduces the periodic fluctuations due to vortex shedding, whereas the RANS model cannot capture transient behaviors.

Better airflow predictions lead to a more accurate source-receptor relationship, consequently resulting in better estimations of the source parameters. The estimation results show that the proposed inference method using LES outperforms the existing one using RANS simulation in the estimations of both the source location and the strength. By adopting the proposed method, the location error is reduced from  $0.22H$  to  $0.05H$ , and the strength error is also reduced from  $-0.363$  to  $0.261$ . In addition, the inference using LES yields not only better point estimation of the source location but also much smaller credible regions, indicating that the true source location has been well identified. Overall, the proposed STE method using LES approach can provide accurate identification of the source in the wake regions, where the conventional methods, such as the inference using RANS simulation, tend to yield less accurate estimations.

## References

- [1] A. Keats, E. Yee, F.S. Lien, Bayesian inference for source determination with applications to a complex urban environment, *Atmospheric Environment*, 41 (2007) 465-479.
- [2] C. Andrieu, N. De Freitas, A. Doucet, M.I. Jordan, An introduction to MCMC for machine learning, *Machine learning*, 50 (2003) 5-43.
- [3] P. Brémaud, *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*, Springer Science & Business Media, 2013.
- [4] J.A. Pudykiewicz, Application of adjoint tracer transport equations for evaluating source parameters, *Atmospheric Environment*, 32 (1998) 3039-3050.
- [5] Y. Tominaga, A. Mochida, R. Yoshie, H. Kataoka, T. Nozu, M. Yoshikawa, T. Shirasawa, AIJ guidelines for practical applications of CFD to pedestrian wind environment around buildings, *Journal of Wind Engineering and Industrial Aerodynamics*, 96 (2008) 1749-1761.
- [6] D.P. Combest, P.A. Ramachandran, M.P. Dudukovic, On the gradient diffusion hypothesis and passive scalar transport in turbulent flows, *Industrial & Engineering Chemistry Research*, 50 (2011) 8817-8823.
- [7] Y. Tominaga, T. Stathopoulos, Turbulent Schmidt numbers for CFD analysis with various

