Wind Engineering Joint Usage/Research Center
FY2016 Research Result Report

Research Field: Outdoor Environment
Research Year: FY2016
Research Number: 152010
Research Theme: Preliminary investigation on the influence of planning factors on the near field PM2.5 dispersion within urban areas

Representative Researcher: Prof. Xiaofeng Li

Budget [FY2016]: 310,000 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

The release of hazardous material into the atmosphere is a major threat to public safety. The fast and accurate retrieval of source information, e.g., location and strength, is a crucial technique to allow emergency preparedness efforts to make appropriate responses and to reduce further impairment. This identification of unknown sources is a typical source term estimation (STE) problem, which addresses the retrieval of emission source information, including location and strength, based on available information. STE can be viewed as an assimilation process of the observed concentration data measured by a sensor network and the predicted concentration data provided by a dispersion model. When considering emissions in complex urban areas, computational fluid dynamics (CFD) approaches are generally used to provide building-resolving results; however, the value of a key parameter, the turbulent Schmidt number Sc_t, has remained an arbitrary choice. Therefore, it is important to investigate the role of Sc_t in STE problems and determine its optimum value for the purpose of obtaining better estimation results. In this study, the impact of Sc_t on STE problems is examined, and Bayesian inference is used to improve estimation accuracy by treating Sc_t as an extra unknown parameter.

2. Research Method

2.1 Introduction of STE and turbulent Schmidt number

STE is a process for assimilating observed concentration data measured by a sensor network and predicted concentration data calculated by an atmospheric dispersion model. This process, however, is usually faced with several challenges, for example, (i) the sensors are sparsely distributed in space and are mostly heavily outnumbered by possible source locations, which can produce multiple release scenarios that match the same observations; (ii) the observations contain errors due to sensor noise and information loss introduced by the potential averaging process; and (iii) the predicted concentration data contain errors due to the uncertainty in the dispersion models. Overall, the STE problem is viewed as an ill-posed inverse problem that is characterized by its non-uniqueness and unstable solutions.

To make the STE problem more tractable, multiple methods have been developed using various approaches, as described in the state-of-the-art reviews of Singh et al. (2015) and Hutchinson et al. (2017). One intuitive solution is to directly inverse the transport process. Another more general alternative is the optimization method of minimizing a cost function, which quantifies the discrepancy between observed and predicted concentrations. The third approach, Bayesian inference considers the problem of STE in a probabilistic logical manner. All parameters are regarded as random variables, rather than as constants, with certain probability distributions. Bayesian inference provides not only the point estimations of unknown parameters but also their probability distributions, thus providing a more natural method of uncertainty estimation. Besides, Bayesian inference is also known for its flexibility in handling extra unknowns, and is thus used in this study.

The STE problem, as mentioned earlier, is an ill-posed inverse problem. Poor model predictions of concentrations often cause heavy deviations in estimated values. The accuracy of the dispersion model is crucial. When dealing with urban dispersion, computational fluid dynamics (CFD) modeling is often employed to reproduce realistic flow and pollutant transport in built-up areas.

In the vast majority of STE investigations performed using CFD, the Reynolds-averaging approach is employed to model turbulent flow and turbulent passive scalar transport, assuming the gradient diffusion hypothesis in order to close the turbulent scalar-flux term $-\overline{u'c'}$, as follows

$$-\overline{u'c'} = D_t \nabla c \tag{1}$$

where D_t is the turbulent mass diffusivity and ∇c is the mean mass gradient. Combest et al. (2011) published a comprehensive review of the above content. By far, the simplest and most popular way to account for D_t is to assume that there is a similarity between turbulent mass diffusivity and turbulent momentum diffusivity (i.e., eddy viscosity ν_t) by assigning a global turbulent Schmidt number

$$Sc_t = \nu_t / D_t \tag{2}$$

Sc_t plays an important role in scalar transport modeling, as its specific value has a significant impact on the accuracy of predictions. Tominaga and Stathopoulos (2007) reviewed relevant investigations performed over the past several decades and concluded that the optimum Sc_t value is problem dependent and should be selected carefully. In STE studies, despite the fact that it may cause deviations in dispersion model predictions and consequently lead to biased estimations, the influence of Sc_t has been rarely addressed. In addition, to date, the selection of Sc_t has been quite arbitrary; surprisingly, in several papers, the value of Sc_t is not even reported.

Therefore, it is important to investigate the role of Sc_t in STE problems and the determination of its optimum value in order to obtain better estimation results. In this study the impact of Sc_t in STE problems is examined, and the possibility of using Bayesian inference to improve the estimation of the source location and strength by treating Sc_t as an extra unknown parameter is explored.

2.2 Bayesian inference

Bayesian inference was first applied to atmospheric STE problems Keats et al. (2007), who provided the fundamentals of the method. In this paper, based on their work, the influence of Sc_t is taken into consideration in the Bayesian framework.

Bayesian inference addresses parameter estimation problems in a probabilistic way. Assume that we are interested in estimating a parameter set, $\boldsymbol{\theta}$, which may include source location, strength, or other unknown quantities, given the measurement information, $\boldsymbol{\mu}$, obtained from a network of sensors. Bayesian theory provides a rigorous way to make an inference based on all of the information given in this problem. More specifically, the estimation result is obtained as the posterior probability, which is given by Bayes' theorem as

$$p(\boldsymbol{\theta}|\boldsymbol{\mu}) = \frac{p(\boldsymbol{\mu}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{\mu})} \propto p(\boldsymbol{\mu}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \quad (3)$$

where the terms embodied can be interpreted as follows. $p(\boldsymbol{\mu}|\boldsymbol{\theta})$ is the likelihood function, which can be interpreted as the probability that we observe the measurement data $\boldsymbol{\mu}$ given the parameter set $\boldsymbol{\theta}$. $p(\boldsymbol{\theta})$ is the prior probability, encoding all a priori information about the unknown parameters known before the measurement. $p(\boldsymbol{\mu})$ is the evidence, acting as a normalizing constant in order to obtain the posterior distribution. $p(\boldsymbol{\theta}|\boldsymbol{\mu})$, which is the posterior probability, is the quantity of interest in the STE problems and encapsulates all of the relevant information required for the inference.

To calculate $p(\boldsymbol{\theta}|\boldsymbol{\mu})$ requires assigning the appropriate function forms for the likelihood function $p(\boldsymbol{\mu}|\boldsymbol{\theta})$ and the prior probability $p(\boldsymbol{\theta})$. In this study, we consider the STE problem of a single point source releasing with a constant strength and an undefined Sc_t, represented by $\boldsymbol{\theta} = (\boldsymbol{x_s}, q, Sc_t)$ where $\boldsymbol{x_s}$ is the source location and q is the source strength. We use the simplest and probably most frequently used form of the likelihood function, Gaussian distribution. Uniform prior, Jeffreys prior, and binomial prior are assigned to the source location $\boldsymbol{x_s}$, the source strength q, and the turbulent Schmidt number Sc_t, respectively.

2.3 Markov chain Monte Carlo

By substituting the above likelihood function and prior distributions into Eq. (3), the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{\mu})$ can be expressed explicitly. Although this distribution can be directly calculated by numerical integration, it represents a huge computational load due to its multidimensional parameter space. To efficiently obtain $p(\boldsymbol{\theta}|\boldsymbol{\mu})$, various MCMC methods have been widely applied to Bayesian inference (Andrieu et al., 2003; Brémaud, 2013), generating a set of sampling points with the desired distribution as its stable distribution. Here, the hybrid of the Gibbs sampler and Metropolis-Hastings algorithm is used.

2.4 Source-receptor relationship

The source-receptor relationship is the sensitivity of the concentration at each sensor to a given source location and Sc_t. This relationship, which contains all of the information of the dispersion model, is obtained by a CFD model. In this study an open source software program (OpenFOAM 2.2.1) is utilized to perform three-dimensional steady-state isothermal flow simulations to reproduce the airflow in complex urban structures. COST Action 732 (Franke and Baklanov, 2007) is applied as the guidelines for CFD settings. The Reynolds-averaged Navier-Stokes (RANS) equations are solved using the standard k-ε turbulence model, which models the eddy viscosity $\nu_t = C_\mu k^2/\varepsilon$, where $C_\mu = 0.09$ is a model constant, k is the turbulent kinetic energy, and ε is the dissipation rate. The second-order TVD discretization scheme is applied to all governing equations, which are solved using the semi-implicit method for pressure-linked equations (SIMPLE) algorithm.

3. Research Result

3.1 Case description

To validate the proposed method, we consider a 3-dimensional dispersion scenario with a ground-level source, here, $\boldsymbol{x_s} = (x, y)$, using a wind tunnel dispersion experiment conducted by the authors at Tokyo Polytechnic University. An open-circuit wind tunnel with a test section 1.2 m wide, 1.0 m high and 14 m long was used to measure the concentration distributions of a continuous point tracer release scenario in an urban-like geometry. As shown in Fig. 1, a 3×3 array of an urban mock-up was formed by cubic blocks and intervals equal to the block height (H=0.09 m). The source was located leeward of the first row with wind approaching perpendicularly. The mean concentrations measured at the 16 centers of the intervals are used to form the measurement data $\boldsymbol{\mu}$. The region shown in Fig. 1 is selected as the possible source area.

The wind tunnel reproduced the wind profile of an urban area. The profiles of horizontal velocity and turbulent kinetic energy were measured to fit the inlet boundary conditions in the CFD simulation. A tracer gas of pure C2H4 (ethylene) was released from a hole (with a diameter of 3 mm) at a flow rate of q=1.67×10-6 m3/s=1.667 mL/s. The concentration measurements were performed by using a fast flame ionization detector (fFID) at a sampling frequency of 150 Hz to provide time-averaged concentrations over a sampling period of 120 s.

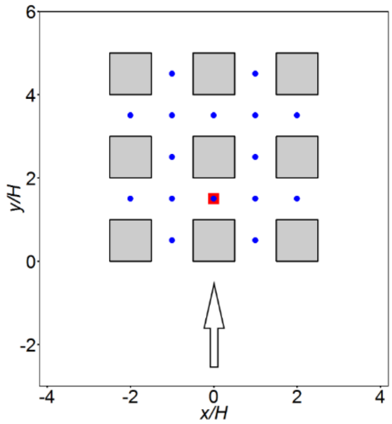

Fig. 1. Block array, source and sample point configuration. Note that the coordinates are normalized by the height of the blocks (H). The gray squares represent cubic building models. The ground level source is denoted by the red square. The 16 blue dots depict sensors located at a height of 0.5H. The arrows indicate the inflow direction of 90°. The shown region is used as the possible source area.

3.2 The impact of Sc_t on estimation results

To clarify the role of Sc_t in STE problems, the source parameters are first estimated using the conventional Bayesian inference approach, which uses a pre-assigned Sc_t value. Different Sc_t values, ranging from 0.2 to 1.3, are then assigned to determine the impact of Sc_t on estimation results. Time-averaged concentrations are used to form the measurement vector $\boldsymbol{\mu}$. The standard deviations of noise σ_i are given by the standard deviations of measurements at corresponding sensors.

The estimation results obtained with different Sc_t values are summarized in Fig. 2. High accuracy can be observed in all estimations of the x coordinates. There is barely any difference within each trial with respect to the posterior mean x coordinate, as all estimated values match the true value. Although larger uncertainties are obtained with increasing Sc_t, the 50% credible intervals maintain accuracy on a rather small scale (under 0.2H, less than 3 grids). This high accuracy is the result of the symmetric layout. With symmetric concentration measurements, it is easy to determine that the source is located on the centerline.

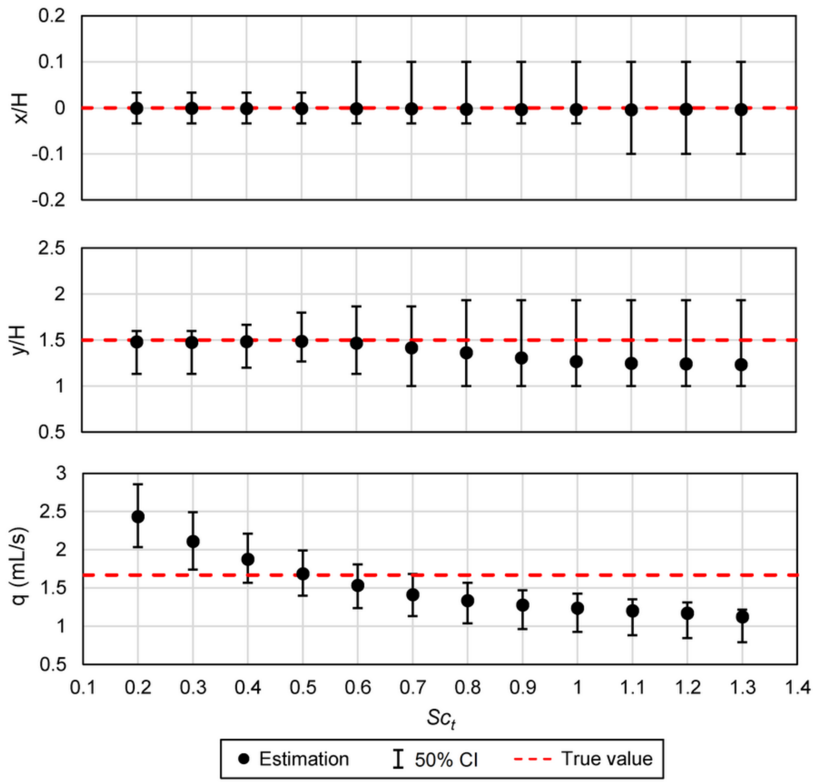

Fig. 2. Estimation results of source parameters (x/H, normalized x-coordinate, y/H, normalized y-coordinate, and q (mL/s), emission strength) with different Sc_t values. Points denote the point estimations, i.e., posterior mean values, with whiskers representing the 50% credible intervals. True values are indicated by dashed lines.

For y coordinates, it should be noted that, although the 50% credible interval tends to increase with increasing Sc_t, it always remains within the range of $1H$ to $2H$, which represents the wake region between the first and second rows of blocks due to the rapid mixing of tracer gas by the vortices in the wake region and the fact that only one sensor is installed inside. Thus, this estimation indicates that the source is located somewhere between the two blocks; however, it is difficult to make a more accurate estimation. In terms of posterior mean values, fairly close estimations (within $0.04H$) are obtained when using Sc_t values of less than 0.6. However, with Sc_t values that are greater than 0.6, the estimated location strays upwind with increasing Sc_t because when Sc_t increases, the turbulent diffusivity ($D_t = \nu_t/Sc_t$) decreases, leading to a narrower plume compared with that in the real scenario. Therefore, to compensate for the underestimated diffusivity, the estimated location tends to move upwind, resulting in a wider plume. In addition, it is worth noting that the vortex still dominates, thus, despite the fact that the estimation shifts towards the upwind direction, it will not exceed the wake region.

In terms of emission strength q, the value of Sc_t has a considerable impact. Within the investigated Sc_t range (0.2-1.3), the estimated strength varies by more than a factor of two. With increasing Sc_t, the posterior mean strength decreases, while the uncertainty remains at a similar scale due to the influence of Sc_t on turbulent diffusivity. Smaller Sc_t values lead to the overestimation of turbulent diffusion, thus causing the plume to be more dispersed and resulting in smaller predicted concentrations at the sensor locations compared to their true values. Therefore, with smaller Sc_t values, the Bayesian inference produces larger estimations of emission strength to obtain matching concentrations with respect to measured values, and vice-versa.

Overall, by comparing the estimation results with the dashed lines, i.e., the true source parameters, it is clear that $Sc_t = 0.5$ provides the best estimation. In other words, if we assign a value of $Sc_t = 0.5$ when running the Bayesian inference, we can obtain a highly accurate estimation. Unfortunately, it is impossible to select this optimum Sc_t simply by looking at the estimation results unless the Sc_t is optimized according to the true source parameters, which are exactly the unknown values that are being estimated in an STE problem. This predicament leads to the arbitrary choice of Sc_t, resulting in possible deviations in estimations. In order to remedy this, the Bayesian inference is further explored to address this difficulty in the next section.

3.3 Estimation by treating Sc_t as an unknown parameter

The conventional approach used to determine the optimum Sc_t value is to compare measured and predicted concentrations based on given source information (Combest et al., 2011; Tominaga and Stathopoulos, 2007). This process bears a close resemblance to the STE method, which simply contains more unknowns (i.e., source parameters). Luckily, Bayesian inference provides a powerful tool to cope with the estimations of multiple parameters.

Assuming that Sc_t, source location and strength are all unknowns, the parameter estimation is performed. The posterior mean of Sc_t is 0.613, which is closer to the optimum value of 0.5. The estimation results of the source parameters are illustrated in Fig. 3 (b). To show the impact of this method, the results are compared with those of the conventional estimation method obtained using a pre-assigned value of $Sc_t = 0.7$, which is a commonly used value (Tominaga and Stathopoulos, 2007); these results are shown in Fig. 3 (a).

In terms of point estimates, the accuracy of all three parameters show improvement, although on different scales, compared with the estimation performed using the pre-assigned Sc_t value. To better quantify the errors of the estimation results, we define two error indices here, namely, location error (i.e., the distance between the estimated location and the true source location) and strength error (i.e., the relative error of the estimated emission strength relative to the true strength). Thus, it is found that by treating Sc_t as an unknown, the location error is reduced by 22%, decreasing from

0.118H to 0.092H, and the strength error is reduced by 60%, changing from -16.2% to -6.5%. Therefore, regarding Sc_t as an extra unknown parameter improves the estimation accuracy.

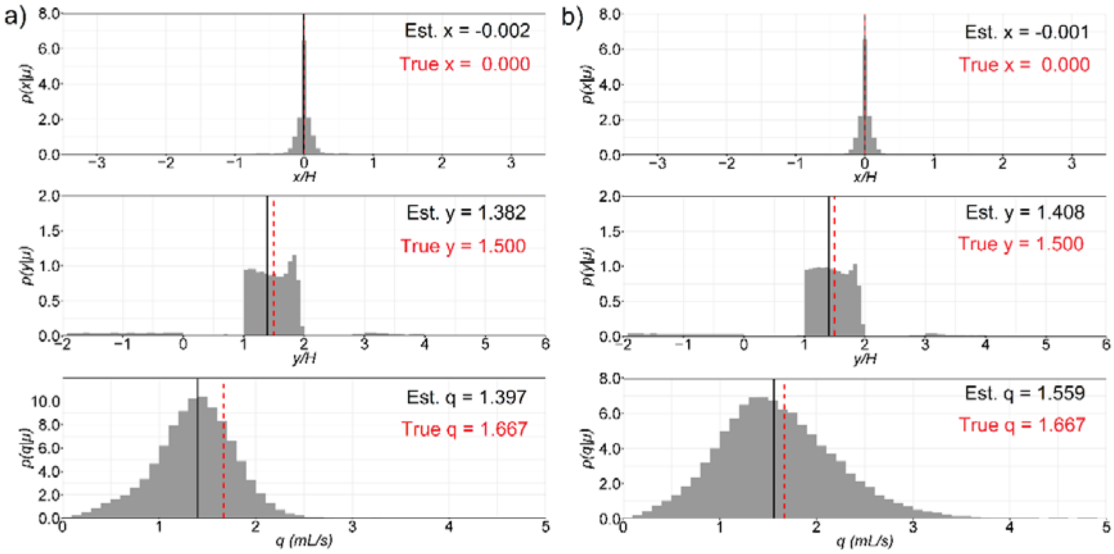

Fig. 3. Estimation results of source location (x- and y-coordinates) and strength q, (a) using pre-assigned $Sc_t = 0.7$ and (b) treating Sc_t as an extra unknown parameter. Gray histograms depict the marginal posterior probability distributions. Solid lines and dashed lines denote the point estimates (posterior means) and true parameters, respectively, with values written in the top right corner of each panel.

In terms of posterior probability distributions, their x-distributions are almost identical; although there is some small difference in the shapes of their y-distributions, they are still bounded in the wake region between the first and second rows (from 1H to 2H). This is in accordance with the impact of Sc_t on estimation results discussed in Section 5.1, i.e., in this specific case, Sc_t shows little influence on the estimation of x and a larger influence on the estimation of y. With respect to the source strength, on which the value of Sc_t shows a major impact, there is a significant difference in the posterior distributions. A flatter probability distribution can be obtained by treating Sc_t as an unknown. The 50% credible interval is almost 2 times wider. Longer tails indicate larger uncertainty in the estimation of strength, which is a deficiency of this method. This is due to the inherent characteristics of parameter estimation problems, namely, that the uncertainty always tends to increase when unknown variables are added into an inference model. To overcome this weakness, more measurement data would be helpful.

3.4 Conclusions

In this study, we investigated the problems concerning the turbulent Schmidt number Sc_t in Bayesian source term estimation. A Bayesian inference method, combined with the CFD method, adjoint equations and MCMC, was used to retrieve the source location and strength of a wind tunnel experiment with a continuous point tracer source in an urban-like geometry. First, the impact of Sc_t on estimation results was analyzed. The following conclusions are obtained:

1. With larger values of Sc_t, the estimated location tends to shift towards the upwind direction. This is because turbulent diffusion decreases with increasing Sc_t, resulting in a narrower plume. Thus, the estimated location moves upwind to compensate for the underestimation of the plume width.
2. Compared with the source location, strength is more sensitive to Sc_t. With larger values of Sc_t, we obtain smaller estimated strength values. This is also due to the effect of Sc_t on turbulent diffusivity.
3. By simply analyzing the estimation results, it is quite difficult to select the optimum value of Sc_t or to determine the best estimation results.

To remedy this, the Bayesian inference method was extended by treating Sc_t as an extra unknown parameter, in addition to source location and strength. The proposed method was performed and the results were compared with those obtained by performing the existing estimation method using a pre-assigned value of $Sc_t = 0.7$. This showed that:

4. With respect to point estimates, the proposed method yields better estimation results. The degree of improvement is positively correlated with the sensitivity of Sc_t. In the demonstrated case, the change of the strength estimation is significantly larger than that of the location estimation.
5. The uncertainty of the model increases when the proposed method introduces an extra unknown parameter. Similar to point estimates, the increment shows a positive correlation with the sensitivity of Sc_t. In the demonstrated case, the uncertainty of the strength estimation increases, while no distinct difference is seen for the location estimation.

References

Andrieu, C., De Freitas, N., Doucet, A., Jordan, M.I., 2003. An introduction to MCMC for machine learning. Machine learning 50, 5-43.

Brémaud, P., 2013. Markov chains: Gibbs fields, Monte Carlo simulation, and queues. Springer Science & Business Media.

Combest, D.P., Ramachandran, P.A., Dudukovic, M.P., 2011. On the gradient diffusion hypothesis and passive scalar transport in turbulent flows. Industrial & Engineering Chemistry Research 50, 8817-8823.

Franke, J., Baklanov, A., 2007. Best practice guideline for the CFD simulation of flows in the urban environment: COST action 732 quality assurance and improvement of microscale meteorological models. Meteorological Inst.

Hutchinson, M., Oh, H., Chen, W.-H., 2017. A review of source term estimation methods for

atmospheric dispersion events using static or mobile sensors. Information Fusion 36, 130-148.
Keats, A., Yee, E., Lien, F.S., 2007. Bayesian inference for source determination with applications to a complex urban environment. Atmospheric Environment 41, 465-479.
Singh, S.K., Sharan, M., Issartel, J.P., 2015. Inverse modelling methods for identifying unknown releases in emergency scenarios: an overview. International Journal of Environment and Pollution 57, 68-91.
Tominaga, Y., Stathopoulos, T., 2007. Turbulent Schmidt numbers for CFD analysis with various types of flowfield. Atmospheric Environment 41, 8091-8099.

4. Published Paper etc.

[Underline the representative researcher and collaborate researchers]

[Published papers]

1. Fei Xue, Xiaofeng Li, Weirong Zhang. Bayesian identification of a single tracer source in an urban-like environment using a deterministic approach. *Atmospheric Environment*. 2017. (under review).

5. Research Group

1. Representative Researcher

Xiaofeng Li	Tsinghua University, China, Associate Professor

2. Collaborate Researchers

Weirong Zhang	Tokyo Polytechnic University, Japan, Associate Professor
Fei Xue	Tsinghua University, China, PhD Candidate
Jie Ma	Tsinghua University, China, PhD
Hideki Kawabata	Tokyo Polytechnic University, Japan, Master's candidate
Atsumu Nagatomo	Tokyo Polytechnic University, Japan, Master's candidate
Zhengfeng Huang	Tokyo Polytechnic University, Japan, Master's candidate
Takumi Tachibana	Tokyo Polytechnic University, Japan, Master's candidate