Papadopoulou, M., Raphael, B., Smith, I.F.C. and Sekhar, C. "Optimal Sensor Placement for Time-Dependent Systems: Application to Wind Studies around Buildings." J. Comput. Civ. Eng., Vol 30, no 2, 2016, 10.1061/(ASCE)CP.1943-5487.0000497, p.04015024 http://cedb.asce.org Copyright ASCE

OPTIMAL SENSOR PLACEMENT FOR TIME-DEPENDENT SYSTEMS: APPLICATION TO WIND STUDIES AROUND BUILDINGS

Maria Papadopoulou[1,2]; Benny Raphael[3]; Ian F.C. Smith, F.ASCE[4]; and Chandra Sekhar[5]

ABSTRACT

Warm climates pose challenges to building energy consumption and pedestrian comfort. Knowledge of the wind flow around buildings can help address these issues through improving natural ventilation, energy use and outdoor thermal comfort. Computational Fluid Dynamics (CFD) simulations are widely used to predict wind flow around buildings, despite

---

[1]Researcher, Singapore-ETH Centre, Future Cities Laboratory, 1 CREATE Way, CREATE Tower, Singapore 138602 (corresponding author). E-mail: papmara@gmail.com

[2]Ph.D. Candidate, Dept. of Building, School of Design and Environment, National Univ. of Singapore, 4 Architecture Drive, Singapore 117566

[3] Associate Professor, BSB 225, Building Technology and Construction Management, Dept. of Civil Engineering, IIT Madras, Chennai 600036, Tamil Nadu, India. E-mail: benny@iitm.ac.in

[4] Professor, Applied Computing and Mechanics Laboratory (IMAC), Civil Engineering Institute, School of Architecture, Civil and Environmental Engineering (ENAC), Swiss Federal Institute of Technology (EPFL), Lausanne 10115. E-mail: ian.smith@epfl.ch

[5] Professor, Dept. of Building, School of Design and Environment, National Univ. of Singapore, 4 Architecture Drive, Singapore 117566. E-mail: bdgscs@nus.edu.sg

the large discrepancies that often occur between model predictions and actual measurements. Wind speed and direction exhibit a high degree of variability that adds uncertainties in modeling and measurements. Although some studies focus on methods to evaluate and minimize modeling uncertainties, sensor placement has been mostly based on subjective judgment and intuition; no systematic methodology is available to identify optimal sensor locations prior to field measurement. This work proposes a methodology for systematic sensor placement for situations when no measurement data are available and knowledge of the wind environment around buildings is limited. Sequential sensor placement algorithms and criteria are used to identify sensor configurations based on CFD simulation predictions at plausible locations. Optimal sensor configurations are compared for their ability to improve wind speed predictions at another location where no measurements are taken. The methodology is applied to two full-scale building systems of varying size. Results show that the methodology can be applied prior to field measurement to identify optimal configurations of a limited number of sensors that improve wind speed predictions at unmeasured locations.

Author keywords: Computational Fluid Dynamics (CFD); System identification; Uncertainties; Measurement system; Sensors; Model falsification

1. INTRODUCTION

The continuous growth of urban areas has emerged as an important environmental issue around the globe. Of the many consequences of urban growth are poor air quality and thermal comfort, as well as increased energy consumption. The wind environment has a prominent role over these issues and improving knowledge of wind flow around buildings has become important. Recent studies have used computational models, such as Computational Fluid Dynamics (CFD), to assess the wind environment around buildings and address issues related to air pollution (Balczó, et al. 2009; Gousseau, et al. 2011) natural ventilation (Chen 2009), pedestrian comfort (Mochida and Lun 2008) and safety (Blocken, et al. 2012), wind-driven rain (van Hooff, et al. 2011) and convection (Defraeye, et al. 2011). Advantages of CFD simulations are i) they allow the study of complex geometries and ii) they provide detailed information on flow characteristics. However the accuracy of predictions is usually questionable (Assimakopoulos, et al. 2006; Blocken, et al. 2007), in particular when steady-state CFD analysis is performed, since predictions are very sensitive to values of the input parameters (Gousseau, et al. 2011; Murakami 1998). Moreover, wind flow around buildings is dominated by complex phenomena and a high degree of variability is expected, due to large differences in building heights and obstacles, as well as inherent climatic variations (Mochida and Lun 2008; Schatzmann and Leitl 2011).

Some researchers have recognized the uncertainty associated with geometric and climatic variations and employed computational parameterization and measurements in search of rules that can be applied to all cases (Martilli, et al. 2003; Oleson, et al. 2008). Recommendations have also been provided on the use of CFD for wind studies around buildings (Franke 2007; Tominaga, et al. 2008). The main issue encountered in previous research is that CFD models have been derived in part from experiments, carried out under specific conditions and within controlled environments. Establishing similitude is consequently a challenge.

Field measurements are essential for evaluating CFD predictions and ensuring that simulations have a sound basis. However, field measurements are difficult to perform, expensive, and result in limited quantities of data with low repeatability. More importantly, wind flow varies considerably over space and time and measurements within the urban canopy depend heavily on the location of sensors and sampling frequency (Pavageau and Schatzmann 1999). Previous research has shown that even when measurements are taken under steady ambient conditions, large discrepancies occur between measured and predicted values that are caused by low frequency variations of the flow (Schatzmann and Leitl 2011).

These factors add to the uncertainties associated with CFD modeling and field measurements. Related work has shown that using a single model with one set of input parameter values may lead to erroneous predictions (Blocken, et al. 2007; Schatzmann and Leitl 2011). In addition, the limited number of sensor locations poses challenges, since it influences the value of the measurement data and the decisions made based on the data (van Hooff and Blocken 2012).

The task of using measurements to infer the behavior of a dynamic system is known as system identification (Ljung 1988). In model-based system identification, physics-based models are used to infer model parameters that are uncertain. Among the common approaches used for inference are based on model validation, such as residual minimization and Bayesian inference. In residual minimization (also known as *model calibration*), values of model parameters are adjusted to minimize the difference between model predictions and measured data; in Bayesian inference conditional probabilities are used to update the prior knowledge on model parameters. These approaches have already been applied in dynamic structural systems. For instance, values of model parameters, such as frequencies and mode shapes, have been estimated using vibration data (Friswell and Mottershead 1995). However, the performance of these approaches depends on the knowledge of modeling errors and their correlations (Beven 2008). Therefore they cannot be applied to wind studies around buildings, since modeling errors are not well known and are associated with the time-dependent atmospheric boundary conditions (Schatzmann and Leitl 2011).

Alternatively, inference approaches based on model falsification use measurements to falsify and not validate models that are not in agreement with the data. Such an approach is called error-domain model falsification and it has been proposed for infrastructure diagnosis (Goulet and Smith 2013). Recent work has demonstrated that this approach is more robust compared with Bayesian approaches for cases when systematic modeling errors are not well known (Goulet, et al. 2013). Uncertainties related to model-parameter values are explicitly represented in error-domain model falsification through a multiple-model approach (Raphael and Smith 2013). Falsification of model instances is performed using measured data and estimated error bounds. Non-falsified candidate models are then obtained that explain the measurements and thus describe the behavior of the system through ranges of parameter values. The term *model instance* refers to a computational model in which input parameters are assigned a definite combination of values and the corresponding values of output variables are predicted using a simulation.

The success of any system-identification approach depends on measurement data and in model falsification, measurement data are important for falsifying model instances and identifying candidate models. In infrastructure diagnosis, there is a tendency to over-instrument (Brownjohn 2007), and therefore many authors have developed sensor placement methodologies that identify the optimal locations needed for identification and diagnosis of structures. Criteria used for placing sensors have typically involved information gain (Meo and Zumpano 2005; Stephan 2012) and information entropy (Kripakaran and Smith 2009; Papadimitriou 2004; Robert-Nicoud, et al. 2005b). Goulet and Smith proposed a sensor placement methodology that predicted the usefulness of monitoring through the capability to falsify candidate models and reduce measurement-system costs (Goulet and Smith 2012). Most importantly, the methodology incorporates systematic modeling and measurement errors, as well as their dependencies. None of these methodologies and criteria has been adapted to predict the behavior of time-dependent systems at unmeasured locations, such as wind flow around buildings.

Compared with infrastructure diagnosis, sensor placement in wind studies around buildings still remains a challenge (van Hooff and Blocken 2012). Sensors have been placed mostly by educated guess, intuitive judgment and common sense. Some researchers have investigated optimal sensor configurations and the information obtained from measured data, either to reduce detection time and consumption of hazardous air pollutants (Hamel, et al. 2006), or to reconstruct a close approximation of the flow field (Mokhasi and Rempfer 2004). Other studies proposed optimal sensor placement approaches based on probabilistic models called Gaussian Processes (GPs) to predict values of several indoor and outdoor environmental variables at unmeasured locations, including temperature, humidity, precipitation and soil moisture (Das and Kempe 2008; Krause, et al. 2008; Osborne, et al. 2008; Wu, et al. 2012). Such approaches are data-driven and require prior knowledge of data distributions and spatial correlations that have been obtained from denser pre-deployment of sensors. However, pre-deploying a large number of wind sensors for outdoor monitoring is costly and time-intensive. Recent work by Du *et al*. (Du, et al. 2014) proposed a mixture GP model based on historical measured data and trained it with CFD simulation predictions to learn spatial correlations. Although Du et al. used the concept of maximum entropy in a similar way to our earlier study (Papadopoulou, et al. 2013), both modeling and measurement data were assumed to be free of errors. None of these studies have presented a rational and systematic sensor placement methodology for wind predictions that includes modeling and measurement uncertainty and can be used prior to field measurement in cases when limited knowledge on wind conditions is available.

This paper proposes a methodology for systematic sensor placement to identify sensor locations that improve predictions for time-dependent systems, such as wind flow around buildings. The methodology uses a multiple-model system identification approach to account for parameter uncertainty in CFD simulation predictions. Existing sensor placement algorithms and criteria are evaluated and adapted from system identification to wind studies around buildings. Section 2 summarizes three sequential sensor placement algorithms, based on incremental addition (forward and forward-max algorithms) and removal of sensors (backward algorithm), using entropy and subset-size as placement criteria. The performance evaluation of sensor placement strategies is performed using a combination of simulated and field measurements (Section 2.3). The final sections demonstrate the applicability of the sensor-placement methodology for predicting the time-dependent behavior of wind around buildings, through testing on two full-scale case studies.

2. Systematic sensor placement methodology

Important inputs when selecting sensor locations are the objectives of measuring and how the data will be used. Then, the sensor placement strategy is chosen in order to identify the optimal number of sensor locations and their configuration. Several sensor placement criteria may be included, such as sensor characteristics, redundancy of information and cost of equipment; however, studying the effect of multiple and conflicting criteria on sensor placement is outside the context of this paper.

In this work, an optimal sensor placement methodology is developed with the objective to improve time-dependent wind predictions around buildings (Fig. 1). The methodology is applied based on the hypothesis that measurement data are best used for model falsification and not for model validation. For each set of data at a particular instant of time, model instances whose predictions are inconsistent with the data are falsified taking into account modeling and measurement uncertainties. The remaining model instances form the *candidate model set for this instant of time*. The candidate model sets identified at any given time are used to perform synchronous predictions of wind speed and horizontal direction, at locations where no measurements are taken.

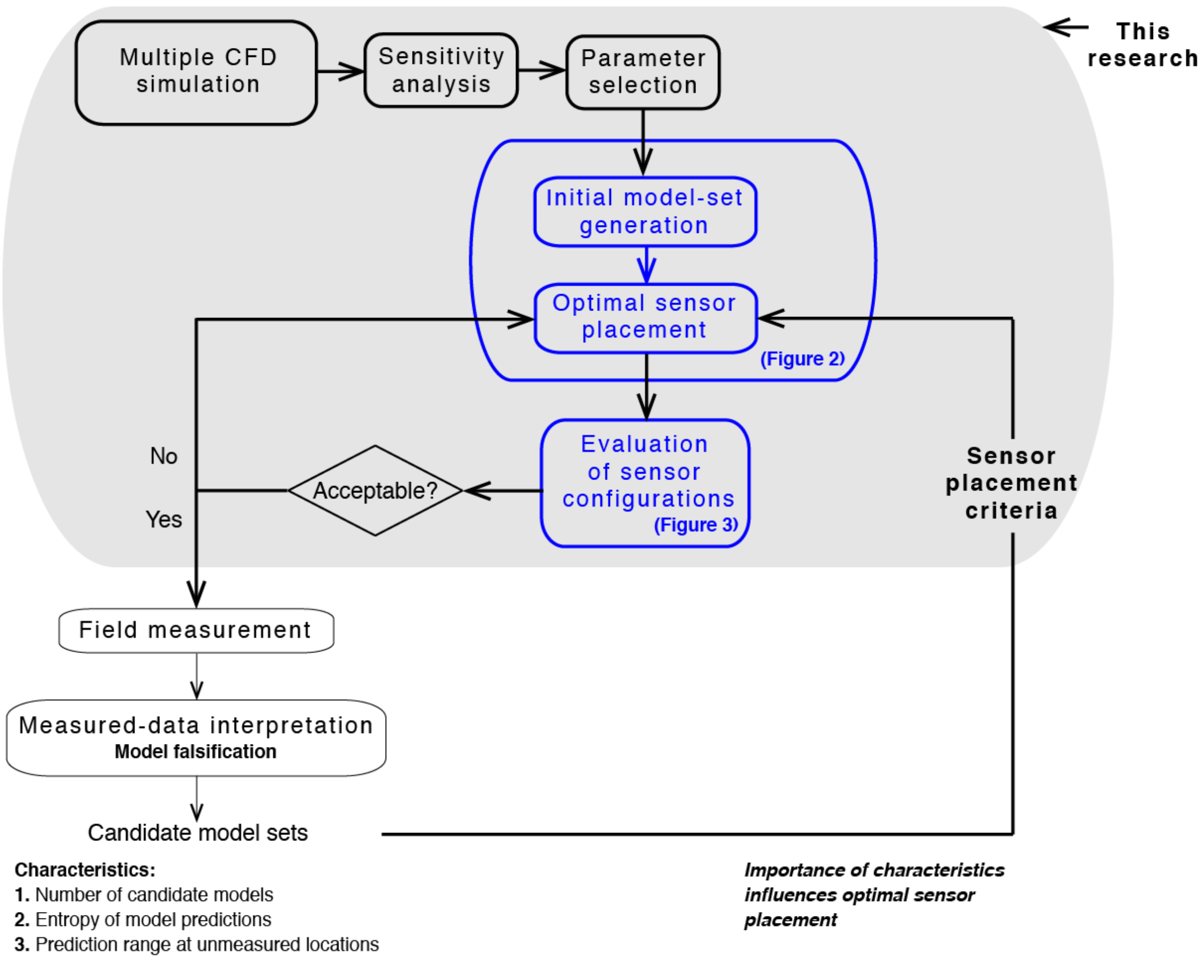

Fig. 1. The optimal sensor placement methodology for wind predictions around buildings and the scope of this work; the contexts of Figures 2 and 3 are also shown.

Although specific interpretation strategies are out of the scope of this work (see Fig.1), they influence the sensor placement criteria. More details on the individual stages of the sensor placement methodology are given below.

2.1. Multiple CFD simulation and sensitivity analysis

The multiple-model approach proposed by Robert-Nicoud et al. (Robert-Nicoud, et al. 2005a) is adapted for CFD simulations in order to include parameter uncertainties. A discrete population of predictions is generated that describes possible wind behavior around buildings. The discrete population of predictions is generated from simulations, varying values of input parameters that are not precisely known with plausible initial ranges defined by engineering judgment and literature. Since the number of possible value combinations is large, it is necessary to minimize computational cost and reduce the number of parameters. Sensitivity analysis and parameter selection are employed in order to choose a reduced set of parameters that have the highest impact on wind speed and direction predictions. The reduced set of parameters allows a simple-grid sampling through selecting values uniformly within the plausible ranges of the selected parameters; the remaining parameters are set to constant values. Multiple, steady-state CFD simulations are performed using all combinations, k, of values of the selected parameters and discrete populations of wind predictions, $y_{k,j}$, are obtained at possible sensor locations, $j = 1, \dots, n_s$, where n_s is the number of potential sensor locations that are fixed. A definite combination of input values of parameters, and the corresponding wind predictions at the potential locations, is one *model instance,* m. The generated discrete population of model instances is called the *initial model set*, M.

Since sensor placement is performed prior to field measurements, optimal sensor configurations are identified using the initial model set, taking into account modeling and measurement errors. The sensor placement criteria depend on the data interpretation approach and in this study the hypothesis is that optimal sensor configurations support multi-model falsification approaches, such as (Goulet, et al. 2013; Vernay, et al. 2014). Therefore the sensor selection criteria should increase the number of falsified model instances and reduce the number of candidate models. The sequence of sensor selection is incremental and two placement strategies are evaluated: addition of sensors from an initial state of no sensors and removal of sensors from an initial state of sensors at all locations. More details are given in the section below.

Finally, the performance of the optimal sensor configurations is evaluated for its ability to improve simulation predictions. During this stage, the candidate models identified with the optimal sensor configuration are used to update simulation predictions at an unmeasured location. The update predictions are then compared with measurements taking into account modeling and measurement errors. In this work, the performance of several sensor configurations is evaluated at the same instant in time, with a limited number of available sensors. Therefore, a combination of simulated and field measurements is used during performance evaluation. The evaluation procedure is described in section 2.3.

2.2. Sequential sensor placement strategies

Three sequential sensor placement algorithms are coded in MATLAB 8.1: the forward, the forward-max (adapted from (Papadimitriou 2004; Robert-Nicoud, et al. 2005b)) and the backward (adapted from (Goulet and Smith 2012)). The forward and the forward-max algorithms incrementally add sensors to the configuration from an initial condition of no sensors. The backward algorithm incrementally removes sensors from an initial configuration of all potential sensors. Sensor locations are selected using either of two placement criteria: entropy and subset size (adapted from (Kripakaran and Smith 2009; Papadimitriou 2004).

The individual stages of the sensor placement strategies are shown in Fig. 2. The first step is to choose the algorithm and the criterion to be evaluated. Then, the range of predicted values at the j^{th} location is divided into equal intervals $\left(I_{w,i}\right)_j$ of width W. The width W depends on estimates of modeling error, e_{model}, and measurement error, e_{measur}:

$$W = |e_{model}| + |e_{measur}| \quad (2.1)$$

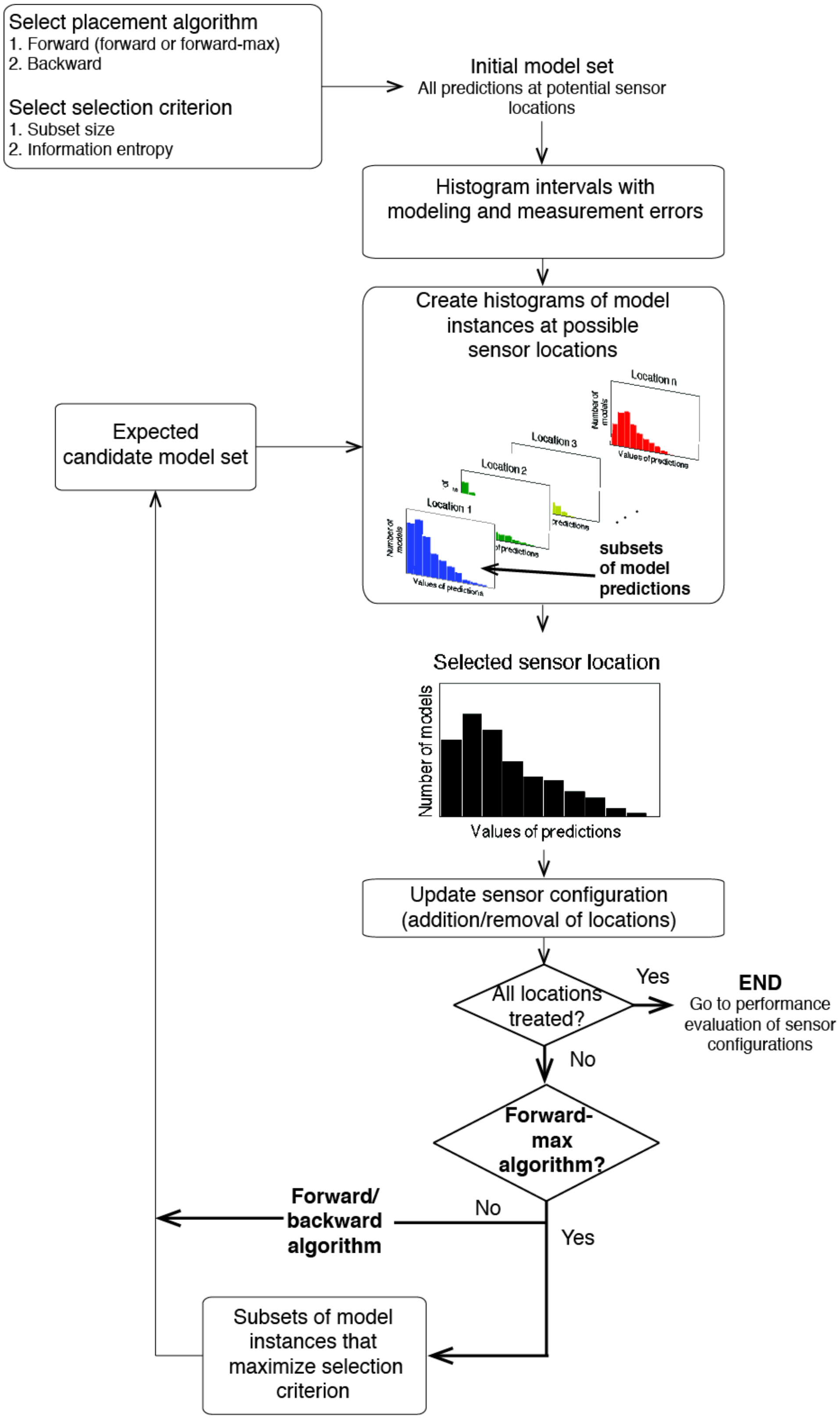

Fig. 2. The three sensor placement algorithms with incremental addition (forward and forward-max) and removal (backward) of sensors, from a configuration of sensors at potential locations.

Next, for each possible sensor location j, all model instances from the initial model set M, are distributed into subsets according to the interval bounds $\left(I_{w,i}\right)_j$, satisfying the condition that $\forall i \in \{1, \ldots, N_I\}\ and\ \forall (m_i)_j \in M : \left(I_{w,i}\right)_j \leq (m_i)_j \leq \left(I_{w,i+1}\right)_j$, where N_I is the maximum number of intervals at the j^{th} location. That is, model instances m_i that predict values within an interval belong to the same subset. Model instances are grouped into subsets depending on modeling and measurement errors, and therefore may not be further discriminated using the current sensor configuration. Histograms of model instances are then created at each location j and are used to evaluate the chosen placement criterion, entropy or subset size, which are explained below.

The subset-size criterion is a direct measure of the number of model instances in a subset, m_i. It is used to estimate the expected maximum number of candidate models. Since sensor placement is performed prior to field measurements, this number corresponds to the subset with the largest number of model instances, among all subsets of the optimal sensor locations.

On the other hand, entropy is used as an indirect measure of disorder in model instances. It is computed using the histograms of predictions of the model instances at possible sensor locations; uniform distributions have the highest entropy. Entropy refers to Shannon entropy or Information entropy and is defined as:

$$H(y)_j = -\sum_{i=1}^{N_I} p(y_i)_j \log_2(p(y_i)_j) \tag{2.2}$$

where $H(y)_j$ is the entropy of a random output variable y, such as wind speed and horizontal direction, at a sensor location j, $p(y_i)_j$ is the probability of the i^{th} interval of a variable's distribution, with $i = 1, \ldots, N_I$ and N_I is the maximum number of intervals at the j^{th} location.

The entropy at each location j is computed through first calculating the number of model instances that lie within each interval, $(m_i)_j$ and then calculating the probability of the interval as $p(y_i)_j = (m_i)_j / M$.

According to the hypothesis that measurements are best used to support multiple model falsification, optimal sensor configurations should increase the number of model instances that are falsified and reduce the number of candidate models.

For the forward algorithm, using the subset-size criterion, optimal sensor locations are incrementally selected in order to reduce the number of candidate models (the subset-size); locations that provide the minimum subset-size are selected. Using the entropy criterion, optimal locations are incrementally selected to maximize separation between model instances in order to increase the number of model instances that are falsified; locations that provide maximum entropy in model predictions are selected.

The backward algorithm is the inverse of forward algorithm and the least useful sensor locations are incrementally removed from a configuration of sensors at all possible locations. Consequently, locations are selected in order of maximum subset-size or minimum entropy.

The forward-max algorithm is essentially a forward strategy with regard to incrementally selecting sensor locations, however it works differently after the 1st optimal location is selected. The simple forward algorithm (as well as the backward) is advantageous when compared to global search algorithms with regard to computational cost (Papadimitriou 2004). Since sensor selection is based on incremental entropy calculations, mutual information between sensors is disregarded and redundant sensor locations may be selected. For example, the sensor location having the second highest value for entropy might contain the same information as provided by the first location. In general, sensor configuration is a combinatorial optimization problem. If there are N possible sensor locations, there are (2^N -1) number of combinations. Selecting the best k locations from among all these combinations would require evaluating non-redundant information content of each combination.

The forward-max algorithm deals with this issue through creating subsets of model instances that predict values within the same intervals of previously selected sensor locations. These is done as follows:

1. For each subset of model instances $(m_i)_{opt1}$ in the histogram of the 1st optimal location (*opt*1), interval bounds $\left(I_{w,i}\right)_{opt1}$ are recalculated at all possible sensor locations *j*.
2. Each subset $(m_i)_{opt1}$ is *subdivided* into smaller subsets of model instances at all locations j, so that $\forall i \in \{1, \ldots, N_I\}\ and\ \forall (m_i)_j \in (m_i)_{opt1} : \left(I_{w,i}\right)_j \le (m_i)_j \le \left(I_{w,i+1}\right)_j$ and new histograms of model instances are created. For example, if there are 10 subsets of model instances $(m_i)_{opt1}$ that cannot be separated further with the 1st sensor location, and 4 possible sensor locations, for each subset $(m_i)_{opt1}$, 4 histograms of model instances are created—one corresponding to each location (in total, 40 histograms).
3. The chosen criterion is evaluated for all histograms and for each subset the maximum criterion value and the corresponding location are recorded (in the above example, 10 criteria values would be recorded).
4. The recorded values are then compared and the location that scores the maximum value is added to the configuration as the 2nd optimal location (*opt*2). The corresponding subset is stored as, $\left(m_{opt2}\right)_{opt1}$.
5. The subset $\left(m_{opt2}\right)_{opt1}$ of the 1st optimal location is then replaced with the smaller subsets according to the *subdivisions* that location (*opt*2) provides.

6. The histogram of the 1st optimal location (*opt*1) is updated and incremental sensor selections are repeated from step 1 until all locations are treated, or no more model instances can be separated.

Although the forward-max algorithm does not directly evaluate the mutual information between sensors, the incremental sensor selection is based on the subset of model instances of previously selected sensors that maximizes the chosen criterion. This procedure has linear complexity with respect to the number of model instances and does not depend on the number of combinations of sensor locations. The maximum number of iterations required is equal to the number of possible subdivisions; the upper bound for this quantity is the maximum number of model instances of all subsets of the 1st optimal location.

For all algorithms, the maximum number of candidate models of the current sensor configuration is recorded during sensor placement and following every update. This number corresponds to the subset with the largest number of model instances among all the subsets of the sensor configuration.

2.3. Evaluation using simulated and field measurements

Each sensor placement algorithm and criterion is expected to construct a different optimal sensor configuration. In order to evaluate and compare their performance, actual measurements at those locations are needed. Since sensor placement is performed prior to field measurements, data at these locations are not currently available.

For performance evaluation, several sensor configurations need to be compared at the same time instant. This could require costly deployment of a large number of sensors. Simulated measurements are therefore generated at optimal locations and historically measured field data are used to create realistic measurements. The simulated measurements are used for making predictions at other locations in order to compute parameters such as prediction range, which indicate the capability of the algorithm to improve the quality of predictions.

The procedure to generate simulated measurements is shown in Fig. 3. First, predictions of the initial model set are combined with modeling and measurement uncertainties of random distribution using a Monte Carlo simulation. Thousands of initial values of simulated measurements are generated at the optimal locations and a random sample is extracted from the combined distribution.

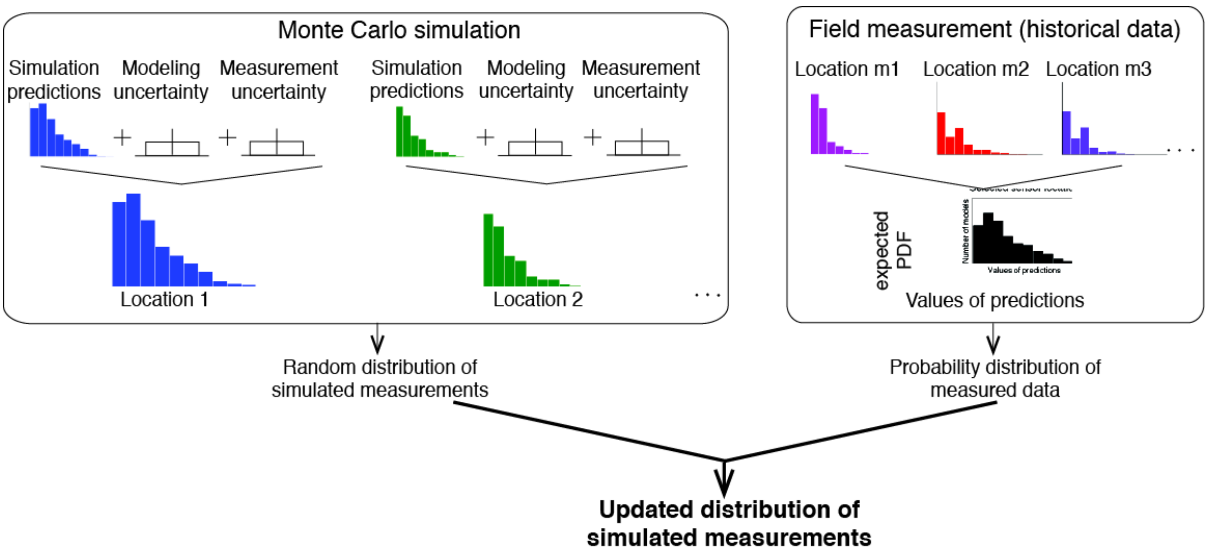

Fig. 3. Framework for generating simulated measurements based on historically measured data and through combining simulation predictions with uncertainties.

Historically measured data, available at other locations, are used to obtain more realistic values for simulated measurements. It is assumed that the sample distribution of the simulated measurements at the locations historically measured should follow the probability distribution of the measurement data. Therefore, the set of initial simulated measurements is sampled in order to obtain a similar probability distribution to the one measured. The corresponding values at other locations, where no measurement data are available, are picked from the initial values of simulated measurements. An updated sample of simulated measurements is thus obtained at the optimal locations. This sample forms the final set of simulated measurements.

Each simulated measurement from the final set is treated as an independent time step and model instances are falsified simultaneously over the optimal sensor locations. An independent candidate model set is obtained for each time step and is used to update predictions at an unmeasured location, which has been randomly selected. The resulting prediction ranges are compared with the initially generated simulated measurements at the same location.

In order for the identification to be successful, the optimal sensor configuration should not only reduce the number of candidate models and narrow prediction ranges, but the prediction ranges should also contain the simulated measurements. The performance of the sensor placement strategies is therefore assessed with respect to prediction ranges, number of candidate models and success in identification.

3. APPLICATIONS

3.1. CASE STUDY 1: BUBBLEZERO

The sensor configuration methodology was applied to BubbleZERO, which is an experimental facility of the Singapore-ETH center for Global Environmental Sustainability, located on NUS campus (Fig. 4, top). Its simple geometry, as well as the tropical climate of Singapore that is characterized by uniform temperatures and two distinct monsoon seasons, made it a good candidate for this study.

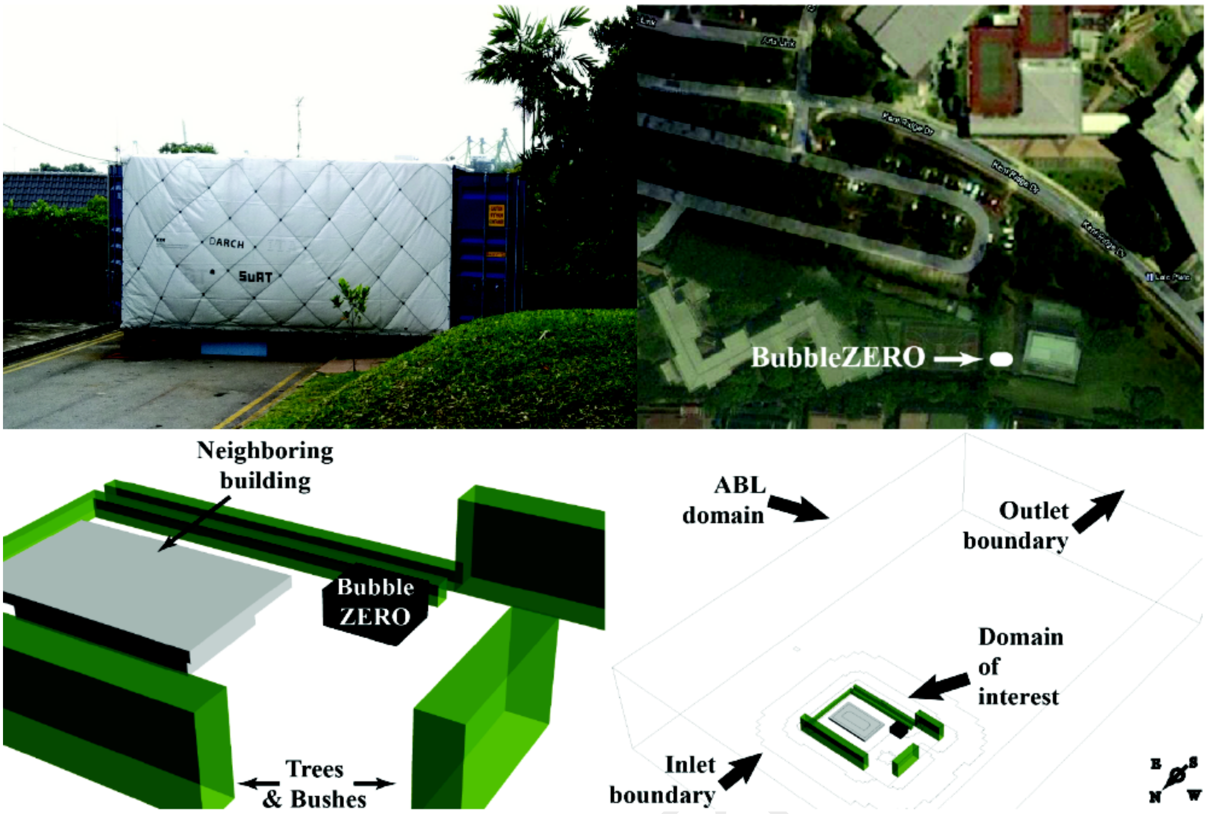

Fig. 4. (Top) Location of BubbleZERO and (bottom) 3D views of the computational domain.

CFD simulations are performed with ANSYS Workbench 14.5, which is a platform that offers a probabilistic analysis in GUI mode using design exploration tools. FLUENT is used as a solver for the equations of flow motion and the Design Exploration tool for sensitivity analysis and feature selection. The simulations require geometrical simplifications and assumptions related to the numerical methods that control the solver:

- The geometry consists of the BubbleZERO, with dimensions 5 m × 6 m × 3 m, and obstacles in proximity: a neighboring building and vegetation (Fig. 4, bottom). The orography of the area is assumed uniform and surface details of obstacles are omitted. The entire size of the computational domain (or *boundary domain*) is 220 m × 140 m × 40 m. The extent of the modeled area and the size of the domain are defined according to recommendations available in literature (Franke 2007; Tominaga, et al. 2008).

- Assumptions with regard to numerical simplifications include the CutCell Cartesian meshing that is used as a discretization method to generate a predominantly hexahedral mesh with minimum user input. The SIMPLE algorithm is employed to achieve pressure-velocity coupling and second-order discretization is used as a pressure interpolation scheme. Finally, the single-precision solver is considered sufficiently accurate for this study.

3.1.1. Numerical analysis and simulation

The behavior of wind around the BubbleZERO and the neighboring obstacles is characterized by a set of mathematical models, parameters, variables and constants that describe flow motion. The selected mathematical models are the RANS-equations, the realizable k-ε equations to represent turbulence and the standard wall-functions to treat near-wall turbulence. Steady-RANS analysis using the realizable k-ε equations is one of the most economical approaches to solve turbulent flows.

In total, 15 parameters are identified related to the geometry, the discretization and the boundary conditions. These include parameters related to the discretization method, the geometry of the boundary domain, the surface roughness of the terrain and of the buildings, the inertial resistance of the vegetation, as well as inlet boundary conditions including wind speed, horizontal direction, turbulence kinetic energy and eddy dissipation. The following equations are used to describe boundary conditions:

$$U(y) = \frac{u_* \ln\left(\frac{z+z_0}{z_0}\right)}{\kappa} \tag{3.1}$$

where $U(y)$ is the wind speed at height z, u_* is the atmospheric-boundary-layer friction (or *shear*) velocity, z_0 the surface roughness and $\kappa \cong 0.41$ the von Kármán constant.

$$k = \frac{{u_*}^2}{\sqrt{C_\mu}}, \tag{3.2}$$

where k is the turbulence kinetic energy and C_μ a model constant.

$$\varepsilon(y) = \frac{{u_*}^3}{\kappa(z + z_0)} \tag{3.3}$$

where $\varepsilon(y)$ is the turbulence eddy dissipation at height z.

In FLUENT, the surface roughness is represented by the roughness height, z_0, which is modified using the equivalent sand-grain roughness, $k_{s,ABL}$, (Equation(3.4)3.4).

$$k_{s,ABL} = \frac{9.793 z_0}{C_s} \tag{3.4}$$

where C_s is the roughness constant, set to satisfy the constraint $k_{s,ABL} \leq z_p$, and z_p is the grid resolution (the distance of the centroid of the wall-adjacent cell to the wall).

Vegetation is modeled as porous media, C, with inertial resistance set in the x- and y-direction as below (Guo and Maghirang 2012):

$$C = C_d d_{SA} \tag{3.5}$$

where C_d is the drag coefficient, varying from 0.1 to 0.5, and d_{SA} is the local leaf-area density, with range 1 to 7 (Tiwary, et al. 2006).

Sensitivity analysis is carried out with the Design Exploration tool in Workbench in order to reduce the number of parameters and minimize computational cost. An Optimal Space-Filling design (Ansys 2011) with CCD sampling (Box and Hunter 1957) is applied, resulting in 283 simulations. The wind speed, referring to the magnitude of the horizontal component of the velocity vector, and the horizontal direction, in degrees, were the output variables of the simulations. Distributions of the output variables were built as a full second-order polynomial response-surface and predictions of wind speed and horizontal direction were obtained at 63 possible sensor locations, which were fixed uniformly near the BubbleZERO, at three height levels: 0.6, 1.5, 2.7 m (Fig. 5).

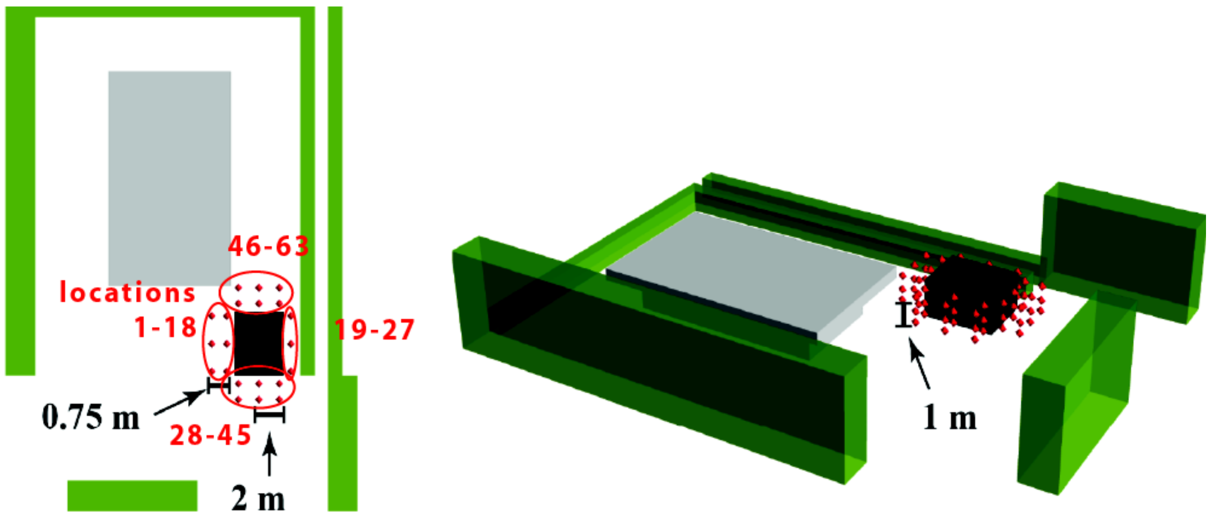

Fig. 5. Possible sensor locations: top view on the left and front view on the right

Spearman's rho correlation coefficient, ρ_j, between the 15 input parameters and the output wind speed and horizontal direction, is calculated at each potential sensor location from Equation (3.6):

$$\rho_j = -\frac{\sum_k (x_{k,j} - \bar{x}_j)(y_{k,j} - \bar{y}_j)}{\sqrt{\sum_k (x_{k,j} - \bar{x}_j)^2 (y_{k,j} - \bar{y}_j)^2}} \tag{3.6}$$

where $x_{k,j}$, $y_{k,j}$ the ranks of the input parameters and output variables respectively at each location $j \in \{1, \dots, 63\}$, with $k = 1, \dots, n$ the size of the sample and $\bar{x}_j$, $\bar{y}_j$ the mean values.

For each input parameter and output variable, the average correlation coefficient $\bar{\rho}_j$ is calculated over all potential sensor locations. The parameters with the highest $\bar{\rho}_j$ over the two output variables are identified as parameters with the highest impact on wind predictions. In order to study wind variability, computational cost had to be reduced by selecting the three parameters with the highest $\bar{\rho}_j$: wind speed, horizontal direction and turbulence kinetic energy at the inlet boundary (with coefficients 0.6, 0.4 and 0.1, respectively).

Although the selected parameters are inlet boundary conditions that might be measured, this is not always practical. Orographic constraints are present in both the pilot study and the second case study (Section 3.2). One of the major issues encountered is the difficulty in deploying sensors remote from the buildings in order to measure the undisturbed flow. In most previous experiments, reference weather stations were used to obtain the values for inlet boundary conditions. However, it is doubtful whether weather station data accurately represent inlet conditions of the simulation model because of the high spatial and temporal variability in climatic conditions. In most cases, resources are limited and only a few points can be measured, which are therefore selected near the buildings. The influence of the sensor locations on the measured data has, however, never been examined (Schatzmann and Leitl 2011).

Simulations were performed by varying values of the parameters within plausible ranges shown in Table 1. The reduced number of parameters allowed a simple-grid sampling through selecting values uniformly within the ranges. A set of 1024 combinations of values was created and simulations were carried out. A discrete population of wind-speed and horizontal-direction predictions at the 63 possible sensor locations was the output of the simulations.

Recent work (Vernay, et al. 2014) has demonstrated that the range of modeling errors associated with wind direction can vary significantly, depending on input values of boundary conditions and sensor locations. The study estimated that modeling errors associated with wind direction could be the most that is possible, up to 180 degrees both ways. This is largely due to use of a steady-state RANS analysis that uses time-averaged equations to describe flow behavior. Therefore, in this study wind direction is not considered and only wind speed predictions are employed to demonstrate the application of the methodology.

Table 1 Parameters and their ranges of values used to generate the initial model set.

Parameter	Lower	Upper	Unit	Comments

	value	value		
Horizontal wind direction	1	360	deg	The wind direction varied from 1 to 360 degrees in order to account for possible values.
Wind speed	0.5	7.2	m/s	The lower and upper bounds were set according to meteorological data obtained from the Changi WMO, Singapore.
Turbulent kinetic energy	0.08	7.23	J/kg	The lower and upper bounds were set according to (Franke 2007; Tominaga, et al. 2008).

3.1.2. Optimal sensor placement

Sensor placement strategies and criteria were evaluated using the initial model set in order to reveal optimal sensor configurations. The range of modeling errors can vary on average between -0.8 and +0.6 m/s for horizontal wind speed (Vernay, et al. 2014). In this study, a spatially uniform and constant value of modeling error is defined equal to ± 0.7 m/s. The range of measurement errors depends on the characteristics of the measurement equipment and is set to 0.1 m/s.

Fig. 6 shows a comparison of three sensor placement algorithms for wind-speed predictions using entropy as a placement criterion. The bars represent the maximum number of candidate models that is expected for a set of optimally placed sensors. For all the algorithms, the rate of change in the maximum number of candidate models is negligible after the 3rd sensor is added to the configuration: for the forward and backward algorithms it levels off and for the forward-max it drops below 5%. However, the forward-max algorithm estimates a significantly lower number of candidate models than the forward and backward algorithms.

The difference exceeds 50 candidate models for sensor configurations of four sensors and above. Overall, the forward-max algorithm has a better performance than the forward and backward algorithms in reducing the number of candidate models, while requiring the least number of sensors.

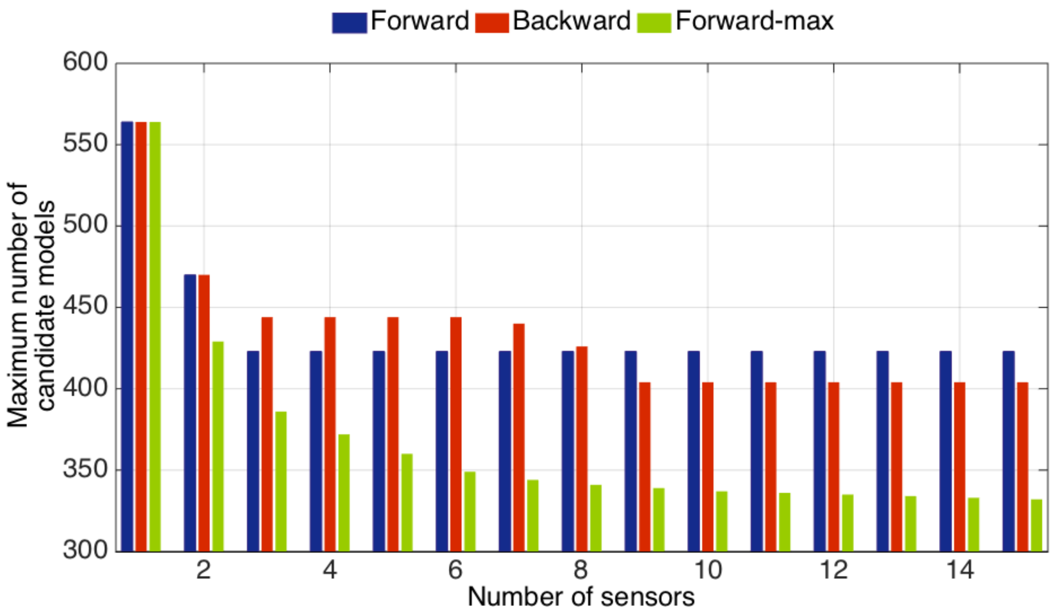

Fig. 6. Comparison of three sensor placement algorithms in estimating the expected maximum number of candidate models, using entropy as a placement criterion; a maximum set of 15 optimally placed sensors is displayed out of the possible 63.

In Fig. 7 the entropy and the subset-size placement criteria are compared for wind-speed predictions using the forward-max sensor placement algorithm. When the subset-size criterion is employed, the estimated maximum number of candidate models is consistently higher than using the entropy criterion. Although the maximum difference between the two criteria decreases with the number of sensors, it is retained above 150 candidate models when less than ten sensors are deployed. Nevertheless, the entropy-based configuration of four sensors estimates a maximum of 372 candidate models, which is 36% of the size of the initial model set; the subset-size criterion estimates a 1.5 times larger upper bound of 535 candidate models for the same number of sensors.

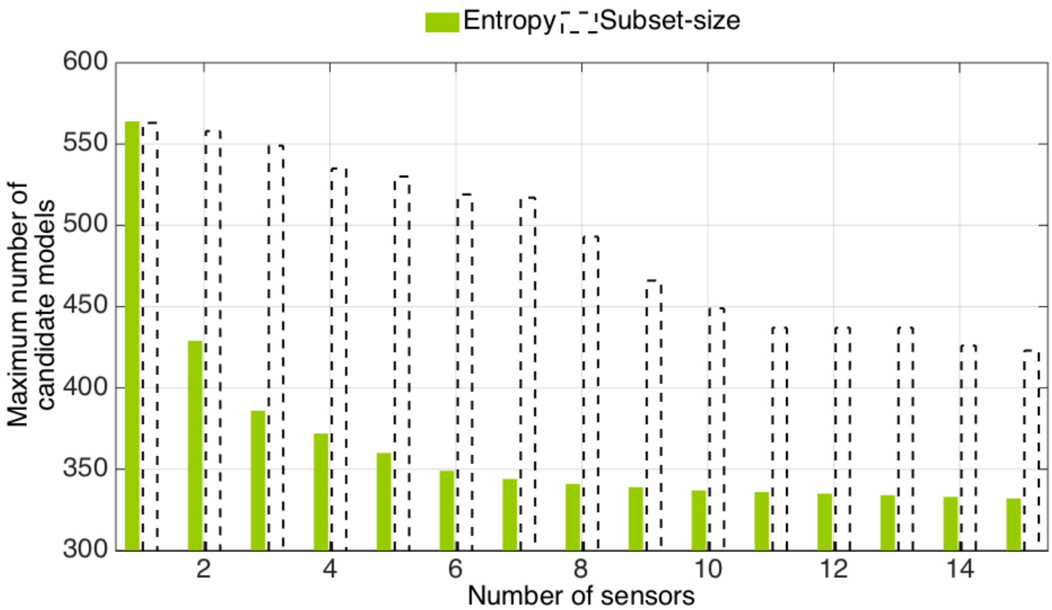

Fig. 7. Comparison of the entropy and the subset-size placement criteria in estimating the expected maximum number of candidate models, using the forward-max sensor placement algorithm; a maximum set of 15 optimally placed sensors is displayed out of the possible 63.

Fig. 8 presents the optimal sensor configurations of 4 sensors using the forward-max algorithm with the entropy (left) and the subset-size (right) placement criteria for predicting wind speed. Except the sensor location L17, the two criteria propose different optimal sensor configurations. Employing the subset-size criterion, no location is selected near the south façade of the building and with entropy, no locations is selected near the west façade of the building. Using either criterion, the 2nd and 3rd location are selected near the north façade. Overall, the optimal configurations are sensitive to the placement criterion and no common configurations are found.

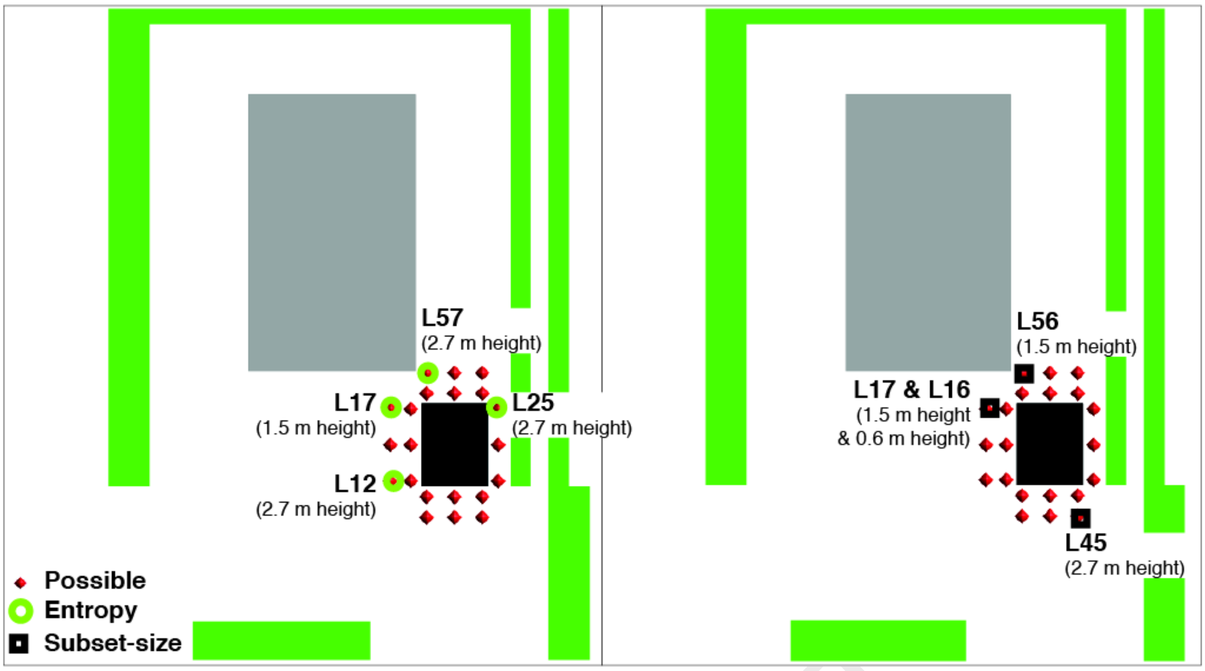

Fig. 8. Optimal sensor locations of the first four sensors for predicting wind-speed, obtained using the forward-max sensor placement with the entropy (left) and the subset-size (right) placement criteria.

3.1.3. Evaluation of sensor configurations

A measurement campaign was carried out around the BubbleZERO in order to evaluate the optimal sensor placement methodology. Four sets of Wireless Vantage Pro2™ and Vantage Pro2 Plus™ weather stations were used for testing with resolution 0.1 m/s and 22.5 deg. Measurements were taken on December 18, 2012 under rainy conditions, which justified the premise of negligible convective effects and the use of isothermal condition during modeling.

Four sensor locations were chosen at random (Fig. 9) and data were collected for a period of two hours, starting at 1pm. The sampling frequency was 2 sec for wind speed and 1 sec for wind direction, while data were recorded every 10 sec. A moving average time series was computed with a short averaging window of 60 sec. The objective is to capture short-term variations in atmospheric boundary conditions and minimize the effect of seasonal variations on flow.

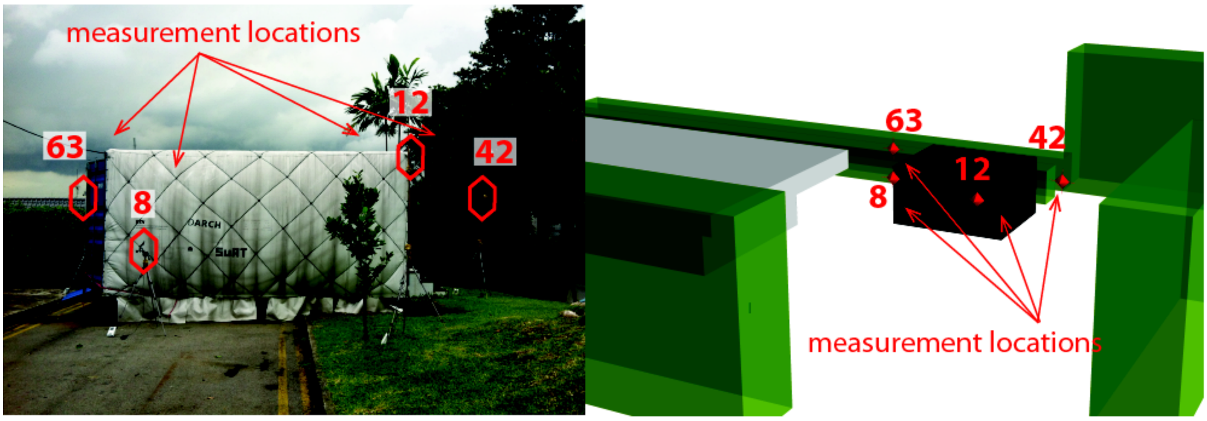

Fig. 9. On-site measurement locations (left) and the same locations in the simulation environment (right).

A sample set of simulated measurements of 2 hours is generated (as described in Section 2.3) at the potential sensor locations. In order to create realistic data, the probability distribution of the simulated measurements is updated according to the distribution of the measured data at the same locations.

Model falsification is performed independently for each time step of the simulated measurements, using the optimal configurations identified by the forward-max algorithm with the two placement criteria. The resulting candidate model sets are used to obtain ranges of wind speed predictions at a 4th unseen location. Each candidate model set represents a set of boundary conditions and wind-speed prediction ranges at that instant of time.

Fig. 10 presents a comparison of the wind-speed prediction ranges obtained using the entropy-based configuration and the subset-size-based configuration of four sensors with the forward-max placement algorithm. Although a short duration of 15 minutes is displayed, the results of the entire 2-hour prediction period are determined. There is a slight difference in the performance of the two placement criteria: the average size of the candidate model set for a 2-hour prediction period drops by 86% using the entropy criterion and by 88% using the subset-size criterion (from the initial model set of 1024). Remarkably, 95% of the simulated measurements are within the prediction range using either criterion. However, estimated prediction ranges show differences. On average, the prediction range is reduced to 2.4 m/s using the entropy-based configuration and to 3.2 m/s using the subset-size-based configuration. This difference in the performance of the two criteria is in agreement with the results in Fig. 7.

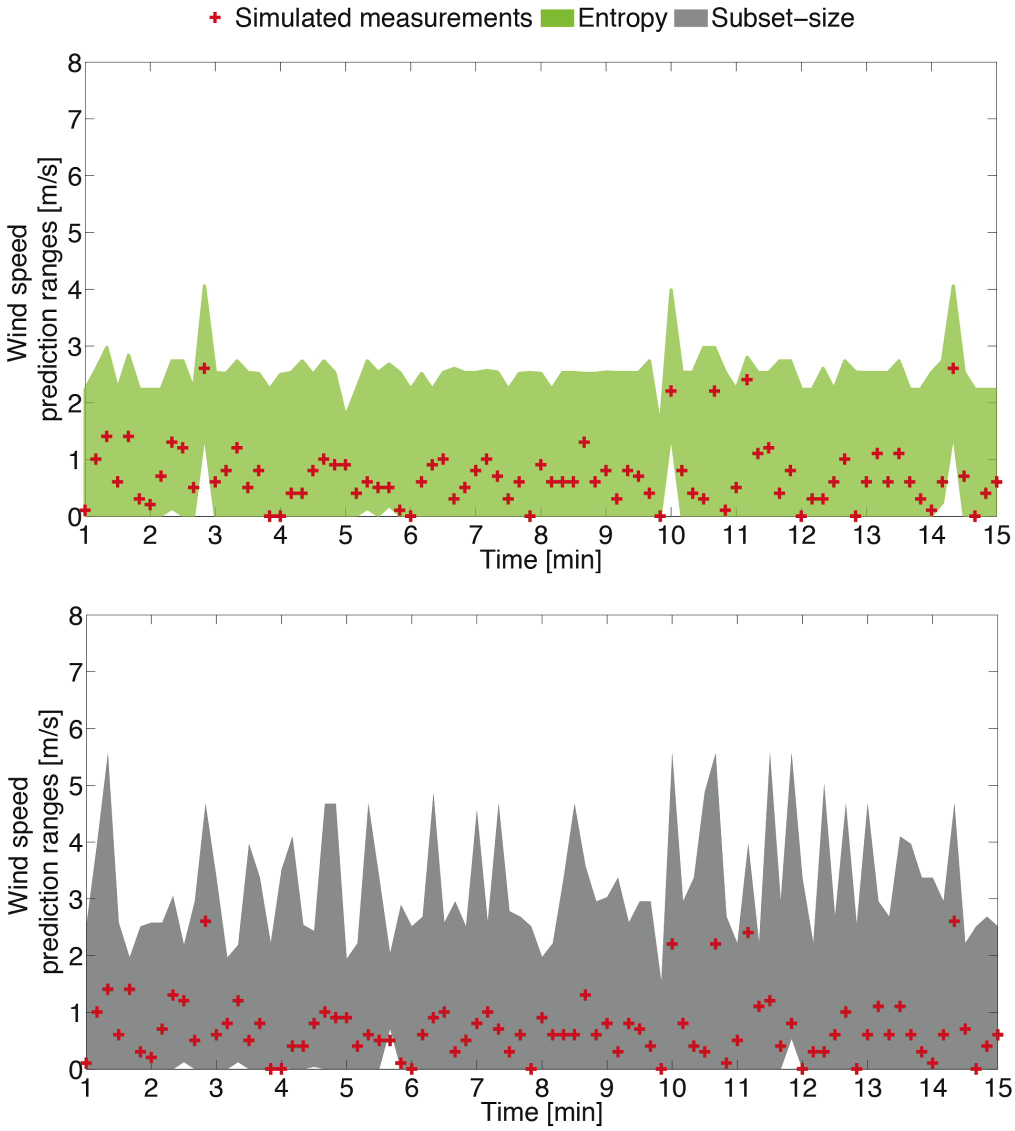

Fig. 10. Comparison of the wind-speed prediction ranges at an unseen location near BubbleZERO, using the entropy-based (top) and subset-size-based (bottom) sensor configuration of three sensors provided by the forward-max algorithm; a short duration of 15 min is displayed from a 2-hr measurement period.

3.2. Case study 2: CREATE Tower

The optimal sensor placement methodology was tested on a larger case study in order to demonstrate its applicability. The study involved the CREATE Tower, a 60 meter high office building on NUS campus in Singapore (Fig. 11).

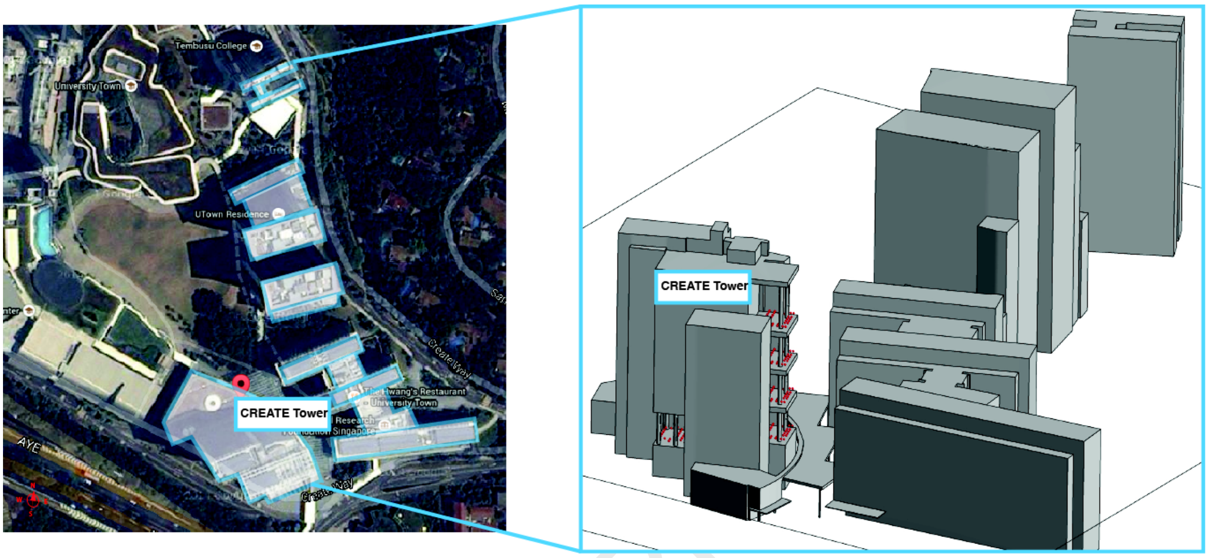

Fig. 11. (Left) Location of CREATE Tower and surrounding buildings considered in modeling and (right) their 3D views in the simulation environment.

The same procedure with Case study 1 is followed and geometric simplifications and assumptions are made during the numerical analysis. Steady-RANS analysis is employed using the realizable k-ε equations and the standard wall-functions. In total, 9 initial parameters are identified related to the geometry, the discretization and the boundary conditions. These are related to the discretization method, the geometry of the boundary domain, surface roughness, wind speed and horizontal direction at the inlet boundary; the parameters are set similar to Section 3.1.1.

Sensitivity analysis is performed and three parameters are identified to have the highest Spearman's correlation coefficient: wind speed, horizontal direction and surface roughness of buildings (with coefficients 0.4, 0.3 and 0.1, respectively).

Multiple CFD simulations are performed, using ANSYS FLUENT, varying values of these parameters within the plausible ranges [0 8.7] m/s, [1 360] deg, [8E-3 0.2] m, defined by engineering judgment and available literature (Franke 2007; Tominaga, et al. 2008). The reduced set of parameters allows simple grid sampling and in total 768 CFD simulations are performed. A discrete population of wind speed predictions is obtained at 187 possible sensor locations, which are fixed uniformly at 1.5 m height near the balconies (east and west) and the north terrace of CREATE Tower (Fig. 11, right). This initial model set was used to evaluate the sensor placement algorithms and criteria and to verify that results are in agreement with Sections 3.1.2 and 3.1.3.

Fig. 12 (a) shows a comparison of the three sensor placement algorithms using entropy as the selection criterion. After placing the second sensor, the forward-max algorithm consistently estimates lower values for the maximum number of candidate models than the forward and backward algorithms, which provide the same results. This difference levels off after the 6th sensor is selected, and is retained to around 50 candidate models.

In Fig. 12 (b), the entropy criterion is compared against the subset-size criterion for its ability to falsify candidate models, using the forward-max sensor placement algorithm. Results are similar with the BubbleZERO case study (Fig. 7), since sensor locations selected using the subset-size criterion estimate a maximum number of candidate models that is consistently higher than using the entropy criterion. Although the maximum difference levels off with the number of sensors, it is more than 100 candidate models for configurations involving less than 7 sensors.

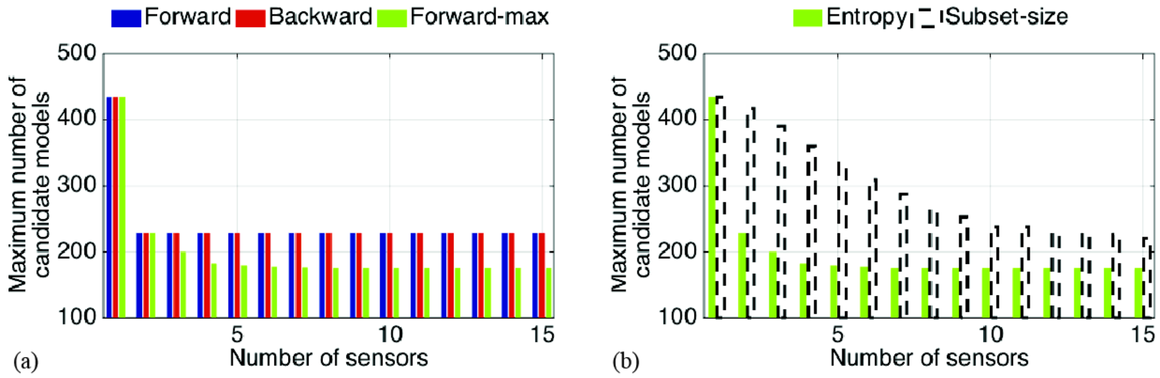

Fig. 12 Comparison of (a) the three sensor placement algorithms using the entropy criterion and (b) the two placement criteria using the forward-max algorithm, for wind-speed predictions; only the first 15 optimal sensor locations are displayed.

Fig. 13 shows the optimal locations of six sensors, using the forward-max algorithm with the entropy (left) and the subset-size (right) placement criteria for predicting wind speed. Similar with the Case study 1 (Fig. 8), the two criteria construct different optimal configurations.

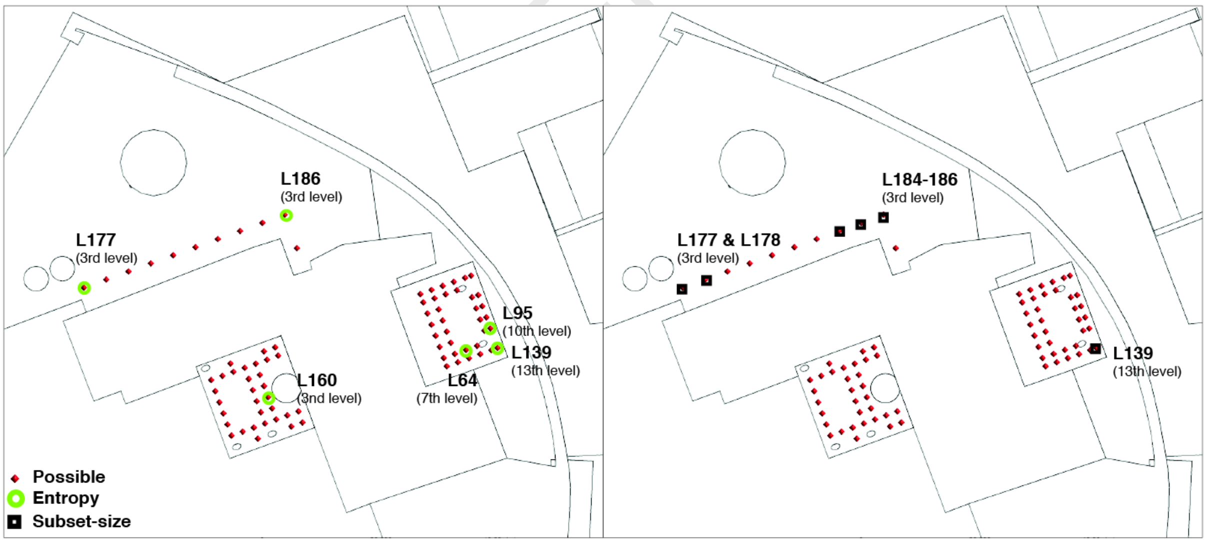

Fig. 13. Optimal sensor locations of the first six sensors for predicting wind-speed near CREATE Tower, obtained using the forward-max sensor placement with the entropy (left) and the subset-size (right) placement criteria.

Although, the above results are similar to those from the small case study (Fig. 6 and Fig. 7), the minimum number of sensors selected is increased from four to six, while the differences in the performance of the algorithms and placement criteria is reduced. These differences are attributed to the effects associated with the size of the case study. The distances between possible sensor locations are large (4 m) compared with the BubbleZERO study (Section 0) and this means that the area covered by the sensors is different. Such differences create varying sensitivities to the selection criteria.

As was done in Section 3.1.3, the optimal sensor configurations are evaluated using a realistic sample distribution of simulated measurements generated based on historical measurements taken at other locations. A total of 2 hours of simulated measurements are used to falsify model instances. A set of candidate models is obtained at each time step and used to update wind-speed predictions at an unmeasured (unseen) location.

Fig. 14 presents a comparison of the ranges of wind-speed predictions obtained at the unmeasured location using entropy-based and subset-size based configurations of six sensors, provided by the forward-max algorithm. Similar results are obtained using either criterion, consistent with the results in Fig. 12 (b) and Fig. 13. The average size of the candidate model set for a 2-hour prediction period drops by 99% (from the initial model set of 768 instances) and the prediction range reduces to almost 1.7 m/s. On average, 67% of the simulated measurements are within the prediction range, which indicates that the accuracy has not been significantly affected by the reduction in the size of candidate model set. Furthermore, it is seen that simulated measurements that lie outside the prediction range are quite close to the boundary (Fig. 14).

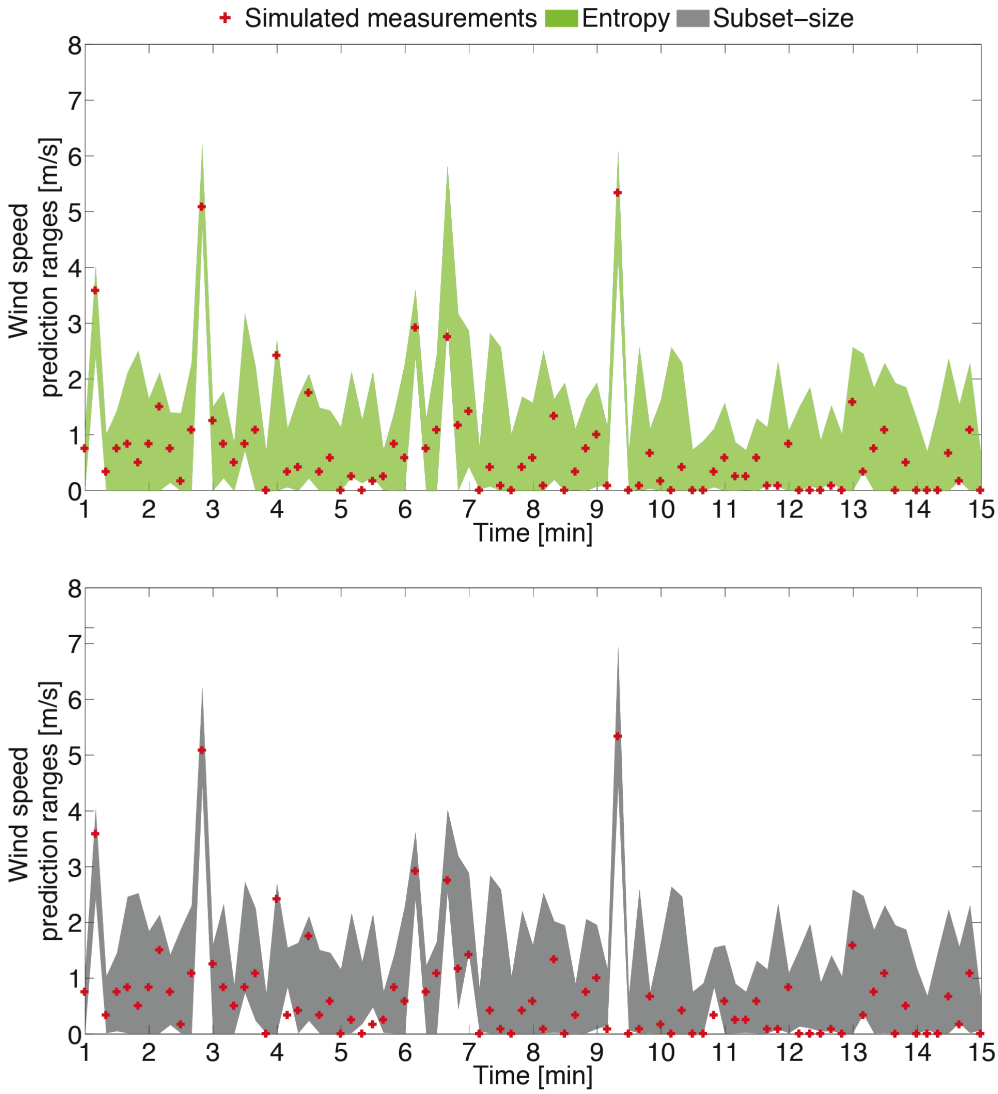

Fig. 14. Comparison of the wind-speed prediction ranges at an unseen location near CREATE Tower, using the entropy-based (top) and subset-size-based (bottom) sensor configuration of three sensors provided by the forward-max algorithm; a short duration of 15 min is displayed from a 2-hr measurement period.

Applying the methodology to a second case study with significant differences in size confirmed that the candidate models identified using an entropy-based configuration with a forward-max placement algorithm can be used most effectively to make predictions at locations where no measured data are available. Overall, it was demonstrated that the methodology could be employed to identify optimal sensor configurations that improve the accuracy of wind-speed predictions and are able to capture the variability of atmospheric boundary conditions.

4. DISCUSSION

Optimal sensor configurations were identified using three sensor placement algorithms and two criteria and were compared according to their ability to accurately predict short-term wind speed variation around buildings. Although the sensor placement strategies evaluated in this paper are similar to (Goulet and Smith 2012; Papadimitriou 2004; Robert-Nicoud, et al. 2005b), they also have important differences as follows:

- In (Goulet and Smith 2012) a backward strategy has been proposed using a similar placement criterion to the subset-size criterion included in this work. However, Goulet and Smith generated simulated measurements from random values of model predictions and used them to select sensor locations.
- The performance of forward and backward sensor placements was compared in (Papadimitriou 2004) with entropy as a placement criterion. However, Bayesian statistical methodology was employed to identify optimal configurations.
- The forward-max algorithm was inspired from (Robert-Nicoud, et al. 2005b) who combined it with the entropy criterion. However, Robert-Nicoud et al. did not explicitly use the subset-size criterion to guide the search for optimal sensor locations.

None of the above studies presented a comparison of the performance of the three algorithms and two criteria for improving the accuracy of predictions using a multiple-model falsification approach. Moreover the above studies concentrated on the identification of structural systems; sensor placement strategies have never been evaluated for time-dependent systems such as wind studies around buildings.

CFD-simulation predictions were employed during the optimal sensor placement and thus the number of assumptions during the application of CFD affected the results. Sensitivity analysis was employed to deal with this issue, since current computational means imposed a constraint on the number of parameters and variables that could be studied. Furthermore, isothermal conditions were assumed during modeling, which are justified, since measurements were taken during rainy conditions to minimize effects of convection on wind flow.

An important contribution of this work is that the effects of modeling error are explicitly incorporated in the optimal sensor placement methodology. Evaluation of the approach using measurements from a full-scale case study is another significant contribution.

A limitation of this work is that systematic errors as well as spatial correlations between errors are not considered. In this work modeling errors are also assumed to be constant. It is known that modeling errors associated with wind speed and direction may vary from location to location. This is due to the RANS-based modeling used in this work, which employs time-averaged equations of flow motion. Ongoing research in our group is studying the effects of modeling errors in terms of horizontal wind direction, input values of boundary conditions and sensor locations. Including such aspects is expected to increase further the accuracy of wind predictions. Future investigations will incorporate systematic modeling errors and spatial correlations in the sensor placement methodology to allow the examination of wind direction predictions at unmeasured locations.

5. CONCLUSIONS

A multiple-model system identification approach has been successfully employed to optimize sensor configurations that improve the accuracy of predictions of time-dependent systems, such as wind-flow around buildings. Specific conclusions are as follows:

1. Sensor placement based on an incrementally updated forward-max algorithm is better than forward and backward algorithms for falsifying model instances of wind speed.
2. Information entropy is a better sensor placement criterion than the subset size for falsifying model instances of wind speed; the degree of reduction in model instances depends on the number of sensors used for identification and the size of the case study.
3. Although information entropy provides a similar reduction in prediction ranges than the subset-size criterion, it can provide better identification of wind speed depending on the size of the case study.
4. Sensor locations that have been configured using entropy with an incrementally updated forward-max algorithm can improve predictions of wind speed at unmeasured locations while capturing time-variability.

6. ACKNOWLEDGEMENTS

This research has been conducted at the Singapore-ETH Centre for Global Environmental Sustainability (SEC), co-funded by the Singapore National Research Foundation (NRF) and ETH Zurich. The authors would like to gratefully acknowledge the support of Prof. M. Santamouris, Dr. J-A. Goulet and D.G. Vernay.

7. REFERENCES

Ansys (2011). "Design Exploration User Guide." ANSYS, Inc., Canonsburg, PA.

Assimakopoulos, V., Georgakis, C., and Santamouris, M. (2006). "Experimental validation of a computational fluid dynamics code to predict the wind speed in street canyons for passive cooling purposes." *Solar energy*, 80(4), 423-434.

Balczó, M., Gromke, C., and Ruck, B. (2009). "Numerical modeling of flow and pollutant dispersion in street canyons with tree planting." *Meteorologische Zeitschrift*, 18(2), 197-206.

Beven, K. (2008). *Environmental modelling: an uncertain future?*, Taylor & Francis.

Blocken, B., Janssen, W. D., and van Hooff, T. (2012). "CFD simulation for pedestrian wind comfort and wind safety in urban areas: General decision framework and case study for the Eindhoven University campus." *Environmental Modelling & Software*, 30(0), 15-34.

Blocken, B., Stathopoulos, T., and Carmeliet, J. (2007). "CFD simulation of the atmospheric boundary layer: wall function problems." *Atmospheric Environment*, 41(2), 238-252.

Box, G. E., and Hunter, J. S. (1957). "Multi-factor experimental designs for exploring response surfaces." *The Annals of Mathematical Statistics*, 28(1), 195-241.

Brownjohn, J. M. W. (2007). "Structural health monitoring of civil infrastructure." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851), 589-622.

Chen, Q. (2009). "Ventilation performance prediction for buildings: A method overview and recent applications." *Building and Environment*, 44(4), 848-858.

Das, A., and Kempe, D. "Sensor selection for minimizing worst-case prediction error." *Proc., Information Processing in Sensor Networks, 2008. IPSN'08. International Conference on*, IEEE, 97-108.

Defraeye, T., Blocken, B., and Carmeliet, J. (2011). "An adjusted temperature wall function for turbulent forced convective heat transfer for bluff bodies in the atmospheric boundary layer." *Building and Environment*, 46(11), 2130-2141.

Du, W., Xing, Z., Li, M., He, B., Chua, L. H. C., and Miao, H. "Optimal sensor placement and measurement of wind for water quality studies in urban reservoirs." *Proc., Proceedings of the 13th international symposium on Information processing in sensor networks*, IEEE Press, 167-178.

Franke, J. (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.

Friswell, M., and Mottershead, J. E. (1995). *Finite element model updating in structural dynamics*, Springer.

Goulet, J.-A., and Smith, I. F. (2013). "Structural identification with systematic errors and unknown uncertainty dependencies." *Computers & structures*, 128, 251-258.

Goulet, J. A., Coutu, S., and Smith, I. F. C. (2013). "Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks." *Advanced Engineering Informatics*, 27(2), 261-269.

Goulet, J. A., and Smith, I. F. C. (2012). "Performance-driven measurement system design for structural identification." *Journal of Computing in Civil Engineering*, 27(4), 427-436.

Gousseau, P., Blocken, B., Stathopoulos, T., and van Heijst, G. J. F. (2011). "CFD simulation of near-field pollutant dispersion on a high-resolution grid: a case study by LES and RANS for a building group in downtown Montreal." *Atmospheric Environment*, 45(2), 428-438.

Guo, L., and Maghirang, R. G. (2012). "Numerical simulation of airflow and particle collection by vegetative barriers." *Engineering Applications of Computational Fluid Mechanics*, 6(1), 110-122.

Hamel, D., Chwastek, M., Farouk, B., Dandekar, K., and Kam, M. "A computational fluid dynamics approach for optimization of a sensor network." *Proc., Measurement Systems for Homeland Security, Contraband Detection and Personal Safety, Proceedings of the 2006 IEEE International Workshop on*, IEEE, 38-42.

Krause, A., Singh, A., and Guestrin, C. (2008). "Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies." *The Journal of Machine Learning Research*, 9, 235-284.

Kripakaran, P., and Smith, I. F. C. (2009). "Configuring and enhancing measurement systems for damage identification." *Advanced Engineering Informatics*, 23(4), 424-432.

Ljung, L. (1988). "System Identification Toolbox™." *User's Guide*.

Martilli, A., Roulet, Y. A., Junier, M., Kirchner, F., Rotach, M. W., and Clappier, A. (2003). "On the impact of urban surface exchange parameterisations on air quality simulations: the Athens case." *Atmospheric Environment*, 37(30), 4217-4231.

Meo, M., and Zumpano, G. (2005). "On the optimal sensor placement techniques for a bridge structure." *Engineering Structures*, 27(10), 1488-1497.

Mochida, A., and Lun, I. Y. F. (2008). "Prediction of wind environment and thermal comfort at pedestrian level in urban area." *Journal of Wind Engineering and Industrial Aerodynamics*, 96(10-11), 1498-1527.

Mokhasi, P., and Rempfer, D. (2004). "Optimized sensor placement for urban flow measurement." *Physics of fluids (1994-present)*, 16(5), 1758-1764.

Murakami, S. (1998). "Overview of turbulence models applied in CWE–1997." *Journal of Wind Engineering and Industrial Aerodynamics*, 74, 1-24.

Oleson, K. W., Bonan, G., Feddema, J., Vertenstein, M., and Grimmond, C. (2008). "An urban parameterization for a global climate model. Part I: Formulation and evaluation for two cities." *Journal of Applied Meteorology and Climatology*, 47(4), 1038-1060.

Osborne, M. A., Roberts, S. J., Rogers, A., Ramchurn, S. D., and Jennings, N. R. "Towards real-time information processing of sensor network data using computationally efficient multi-output Gaussian processes." *Proc., Proceedings of the 7th international conference on Information processing in sensor networks*, IEEE Computer Society, 109-120.

Papadimitriou, C. (2004). "Optimal sensor placement methodology for parametric identification of structural systems." *Journal of sound and vibration*, 278(4), 923-947.

Papadopoulou, M., Raphael, B., Sekhar, C., and Smith, I. F. (2013). "Sensor placement for predicting airflow around buildings to enhance natural ventilation." *ASHRAE IAQ 2013 Proceedings: Environmental Health in Low Energy Buildings*(EPFL-CONF-195670), 240-249.

Pavageau, M., and Schatzmann, M. (1999). "Wind tunnel measurements of concentration fluctuations in an urban street canyon." *Atmospheric Environment*, 33(24), 3961-3971.

Raphael, B., and Smith, I. F. C. (2013). *Engineering Informatics: Fundamentals of Computer-Aided Engineering*, Wiley.

Robert-Nicoud, Y., Raphael, B., Burdet, O., and Smith, I. F. C. (2005a). "Model identification of bridges using measurement data." *Computer - Aided Civil and Infrastructure Engineering*, 20(2), 118-131.

Robert-Nicoud, Y., Raphael, B., and Smith, I. F. C. (2005b). "Configuration of measurement systems using Shannon's entropy function." *Computers & structures*, 83(8), 599-612.

Schatzmann, M., and Leitl, B. (2011). "Issues with validation of urban flow and dispersion CFD models." *Journal of Wind Engineering and Industrial Aerodynamics*, 99(4), 169-186.

Stephan, C. (2012). "Sensor placement for modal identification." *Mechanical Systems and Signal Processing*, 27, 461-470.

Tan, W. C. K. (2011). *Practical research methods*, Pearson Education South Asia Pte Ltd.

Tiwary, A., Morvan, H. P., and Colls, J. J. (2006). "Modelling the size-dependent collection efficiency of hedgerows for ambient aerosols." *Journal of aerosol science*, 37(8), 990-1015.

Tominaga, Y., Mochida, A., Yoshie, R., Kataoka, H., Nozu, T., Yoshikawa, M., and Shirasawa, T. (2008). "AIJ guidelines for practical applications of CFD to pedestrian wind environment around buildings." *Journal of Wind Engineering and Industrial Aerodynamics*, 96(10), 1749-1761.

van Hooff, T., and Blocken, B. (2012). "Full-scale measurements of indoor environmental conditions and natural ventilation in a large semi-enclosed stadium: possibilities and limitations for CFD validation." *Journal of Wind Engineering and Industrial Aerodynamics*, 104, 330-341.

van Hooff, T., Blocken, B., and van Harten, M. (2011). "3D CFD simulations of wind flow and wind-driven rain shelter in sports stadia: Influence of stadium geometry." *Building and Environment*, 46(1), 22-37.

Vernay, D. G., Raphael, B., and Smith, I. F. (2014). "Augmenting simulations of airflow around buildings using field measurements." *Advanced Engineering Informatics*.

Wu, X., Liu, M., and Wu, Y. (2012). "In-situ soil moisture sensing: Optimal sensor placement and field estimation." *ACM Transactions on Sensor Networks (TOSN)*, 8(4), 33.