Article

Hierarchical Sensor Placement Using Joint Entropy and the Effect of Modeling Error

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

[1] Future Cities Laboratory, Singapore-ETH Centre, Singapore 138602, Singapore

[2] Department of Building, National University of Singapore, Singapore 117566, Singapore;
E-Mail: bdgscs@nus.edu.sg

[3] Civil Engineering Department, Indian Institute of Technology Madras, Chennai 600036, India;
E-Mail: benny@iitm.ac.in

[4] Applied Computing and Mechanics Laboratory, EPFL, Lausanne 1015, Switzerland;
E-Mail: ian.smith@epfl.ch

* Author to whom correspondence should be addressed; E-Mail: papmara@gmail.com;
Tel.: +65-8287-0922.

Abstract: Good prediction of the behavior of wind around buildings improves designs for natural ventilation in warm climates. However wind modeling is complex, predictions are often inaccurate due to the large uncertainties in parameter values. The goal of this work is to enhance wind prediction around buildings using measurements through implementing a multiple-model system-identification approach. The success of system-identification approaches depends directly upon the location and number of sensors. Therefore, this research proposes a methodology for optimal sensor configuration based on hierarchical sensor placement involving calculations of prediction-value joint entropy. Computational Fluid Dynamics (CFD) models are generated to create a discrete population of possible wind-flow predictions, which are then used to identify optimal sensor locations. Optimal sensor configurations are revealed using the proposed methodology and considering the effect of systematic and spatially distributed modeling errors, as well as the common information between sensor locations. The methodology is applied to a full-scale case study and optimum configurations are evaluated for their ability to falsify models and improve predictions at locations where no measurements have been taken. It is concluded that a sensor placement strategy using joint entropy is able to lead to predictions of wind

characteristics around buildings and capture short-term wind variability more effectively than sequential strategies, which maximize entropy.

1. Introduction

With more than half of the global population living in cities and with an estimated annual increase of urban dwellers reaching nearly 60 million [1], much recent research work has focused on urban-related aspects, including studying the wind environment where buildings are, or will be, placed. Common concerns of wind studies are pedestrian comfort [2], air quality [3,4], safety [5], energy use and natural ventilation [6]. The approach used in each study depends on the length scale of the study area. In small-scale studies, such as those around buildings (distances up to 1–2 km) [7], Computational Fluid Dynamics (CFD) modeling is commonly used to predict wind behavior.

Advantages of CFD modeling are that it allows treatment of a wide range of complicated geometries and it provides detailed information on airflow. Although CFD modeling may lead to reasonable predictions, results can be very different from field and laboratory experiments [8]. Even so, predictions from the same mathematical model applied by different modelers may differ, or more than one model may exist that generates the same predictions [9]. The application of CFD requires experienced users, and predictions are subject to challenges associated with precision, computational storage and execution time [10]. Guidelines are available in the literature on the application of CFD in wind studies around and through urban canyons [11,12]. Recommendations on appropriate boundary conditions have also been provided [13,14], while others have proposed methodologies for evaluating environmental models [15].

Uncertainties are inherent in wind modeling and they are associated with modeling assumptions as well as model application [8]. A difficulty is that models used in CFD are derived from experiments repeated under controlled laboratory conditions or from real-world data collected under specific contexts. In most cases they are able to capture only an approximation of the true conditions [16]. Consequently, models are not always applicable to a wide range of situations. In addition, wind studies around buildings involve open systems; a large number of reference variables influences climatic conditions and values of these variables are often unknown [2,17]. Furthermore, no information related to the accuracy of results is usually available.

Multiple-model approaches can be used to accommodate uncertainties involved in modeling and to account for many possible flow conditions with less risk of parametric-value compensation. In 1998, Raphael and Smith introduced a multiple-model approach for system identification of civil structures in order to account for uncertainties involved in modeling and measurement [18,19]. Mathematical models that describe system behavior are parameterized and populations of possible predictions are obtained. These model predictions are then compared with measurements of real systems and models whose predictions are incompatible with measured values are falsified. The remaining models comprise the candidate-model set. This task is known as system identification and among the

approaches proposed to date, model falsification is the most robust when values of correlations are not known [20].

Regardless of the approach used, good sensor placement is important for identifying candidate models. In system identification, measurement systems have been designed in order to place sensors at locations of high information value. Entropy, from information theory (also known as Shannon entropy or information entropy), has been used in earlier studies as a design criterion to identify sensor locations and number of sensors needed for identification; good locations were either positions of high entropy in values of model predictions [21,22] or positions that offered high entropy reduction [23]. Although these early studies used model predictions to identify optimal sensor locations, they did not explicitly incorporate modeling error into the measurement-system design process. Systematic modeling errors have not been considered (except in previous research by the authors) and the effect of spatial distribution of modeling error on sensor placement has not been studied.

Goulet and Smith [24] proposed a measurement-system-design methodology that includes error dependencies and their values. These error dependencies were described by correlation coefficients that were quantified using qualitative reasoning formulation (low, moderate, and high), given that little information was available to the authors. Comparing previous work [21,22], Goulet and Smith evaluated the usefulness of monitoring through the capability to reduce the number of candidate models and not through maximizing the entropy of model predictions. Papadimitriou and Lombaert have used an entropy-based sensor placement to stress the effect of spatial correlation of prediction errors [25]. However, optimum locations were selected as positions of minimum entropy in probabilities of model parameter values. Such an approach is difficult to apply to complex, time-dependent systems where multiple models and varying parameter values exist, such as wind studies around buildings.

Earlier studies in system identification have demonstrated that information entropy can be successfully used as a design criterion to optimize measurement systems in order to improve the accuracy of model predictions [22,23]. The two sensor-placement strategies that have been identified are essentially sequential; sensors are placed one at a time at locations that provide either the higher reduction or higher value in information entropy. Subsequent sensor placement does not change the location of sensors that have already been placed. Sequential strategies are preferred to global search strategies such as genetic algorithms due to computational cost [23]. Nevertheless, during sequential sensor placement, entropy is usually calculated at each location individually, disregarding the possibility of selecting locations with similar information content.

Field measurements are useful in wind studies in order to evaluate predictions and ensure that the modeling is sound; even when using modeling methods of high predictability, such as the large eddy simulation (LES) [26]. However, field measurements have been rare since they are difficult to perform, expensive, but also result in limited quantities of data with low repeatability. The challenge is that flow properties vary considerably with space and time, and so the location of sensors, their type, and the duration of measurements significantly affect the value of such information [27]. In addition, a limited number of feasible measurement locations remains a challenge [28].

Until now, measurement systems in wind studies around buildings have been designed using educated guesses and common sense. Some research has studied optimal sensor configurations in built environments, either in terms of pollutant dispersion to protect against nuclear, biological, and

chemical attacks (NBC) [29] or with the aim to reconstruct a close approximation of the flow field [30]. Recent work by Du *et al*. [31] proposed a methodology to identify optimal sensor locations for wind studies in an urban reservoir. In their study, an entropy-based sensor placement has been applied to wind predictions obtained from CFD simulations in order to identify optimum sensor locations. The objective was to use the readings from a limited number of sensors to predict the entire wind field over a reservoir surface. Although a limited number of two model parameters were selected to run CFD simulations, systematic modeling errors and the effect of modeling uncertainties on sensor placement have not been considered. Moreover, sensor placement has been performed iteratively and the possibility of selecting sensors having mutual information has not been considered. No rational and systematic methodology for sensor placement has been presented that includes modeling uncertainties and identifies configurations of sensors that could be used to improve the accuracy of wind predictions.
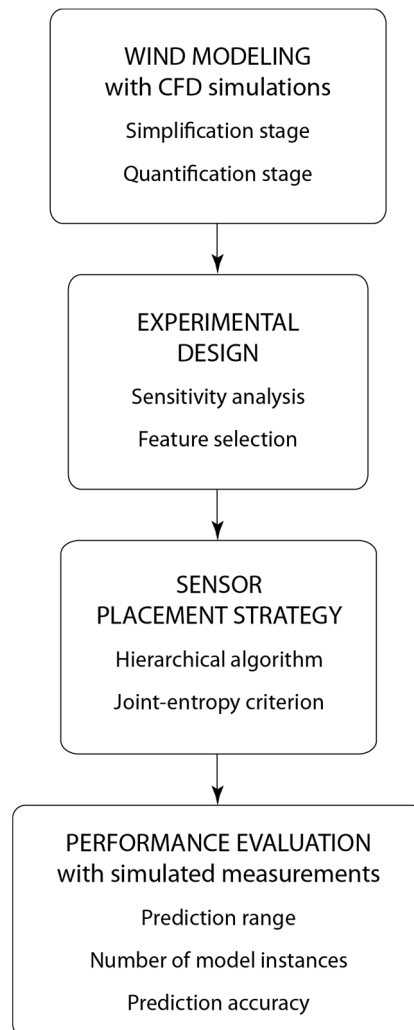
This paper proposes a hierarchical sensor placement strategy using joint entropy that explicitly incorporates spatial distribution of modeling errors and their values. The study also builds upon previous work in sensor placement where entropy was identified as a better design criterion than subset size [32]. Another aim of this paper is to evaluate the effect of modeling errors on optimal sensor configurations. In Section 2 the sensor placement strategy is explained. The hierarchical algorithm and the entropy design-criterion are further summarized. Results of applying the framework to a full-scale real-world building are presented in Section 3. The final two sections discuss research findings as well as limitations of the framework.

2. Sensor Placement Strategy

Measurements are performed to collect quantitative information about physical variables, by comparing them with a known standard. The aim is to enhance knowledge and provide a better understanding of the underlying processes which otherwise could only be estimated. In this work, measurements are used to improve wind predictions around buildings and capture short-term wind variability. A sensor placement strategy is developed to identify optimal sensor configurations prior to measuring, with limited knowledge of wind behavior.

The research design is comprised of four stages as illustrated in Figure 1. First, wind modeling is performed using CFD simulations to obtain possible wind predictions. Modeling focuses on two aspects in the simplification stage to include decisions related to geometrical simplifications and numerical methods, and the quantification stage, during which mathematical models, parameters, variables and constants that describe the system are identified and quantified. Due to the complexity of wind modeling, a significant degree of uncertainty is associated with mathematical models, parameters and boundary conditions.

During experimental design, sensitivity analysis is employed in order to evaluate the effect of variations in the values of model parameters on model predictions and feature selection to select a small number of parameters that have the highest impact on predictions. A multiple-model approach is adopted [18] and values of model parameters are varied within plausible ranges to create populations of model instances. These instances comprise the initial model set, with which multiple, steady-state CFD simulations are executed to obtain a discrete population of predictions at possible measurement locations.

Figure 1. Flowchart of sensor placement methodology.

```
┌─────────────────────────────┐
│      WIND MODELING          │
│    with CFD simulations     │
│                             │
│     Simplification stage    │
│                             │
│     Quantification stage    │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│      EXPERIMENTAL           │
│         DESIGN              │
│                             │
│     Sensitivity analysis    │
│                             │
│      Feature selection      │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│        SENSOR               │
│   PLACEMENT STRATEGY        │
│                             │
│    Hierarchical algorithm   │
│                             │
│    Joint-entropy criterion  │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│   PERFORMANCE EVALUATION    │
│   with simulated measurements│
│                             │
│      Prediction range       │
│                             │
│   Number of model instances │
│                             │
│     Prediction accuracy     │
└─────────────────────────────┘
```

Sensor placement is performed using the simulation predictions and a hierarchical algorithm is used to place sensors at locations that satisfy the desired design criterion, in this case maximum information content that corresponds to maximum joint-entropy. The objective is to design a measurement system that supports model falsification approaches, such as [20,33], and improves predictions. Further details are given in the Sections 2.2 and 2.3.
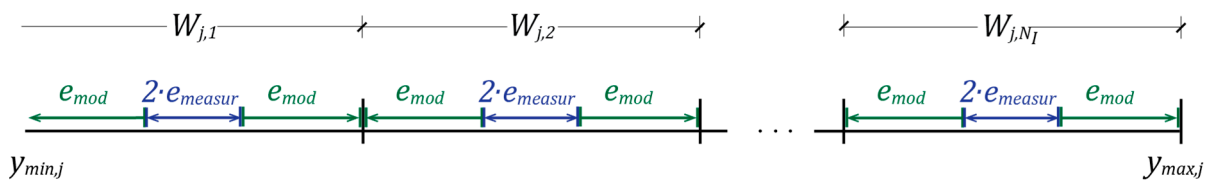
The performance of the sensor placement strategy is evaluated by demonstrating that optimal sensor configurations can improve wind predictions through the following three metrics: reducing the number of candidate models, minimizing the prediction range and increasing the accuracy of predictions. During performance evaluation, simulated measurements are created at optimum locations through combining predictions of the initial model set with modeling and measurement uncertainties. Historically, measured data from a full-scale study, available at other sensor locations, are used to create a more realistic distribution; it is assumed that the sample distribution at the same locations should follow the probability distribution of the measured data.

2.1. Errors in Wind-Speed and Wind-Direction

Measurements are essential for theory testing, yet in order to be useful they need to be accurate and precise. Collecting incorrect measurements results in misleading conclusions about the state of the system. The term accuracy is linked to how close the measurement is to the actual value. Precision is an indication of the consistency of a measurement. Obtaining precise measurements does not imply that they are accurate and accurate measurements are not necessarily precise. A good measurement system should perform well in terms of both these characteristics. In addition, for the purpose of this study, CFD simulations are used to obtain wind predictions and these predictions include modeling errors whose influence is taken into account.

In order to incorporate modeling and measurement uncertainties in the sensor placement strategy, a histogram of model predictions is built at potential sensor locations $j = 1, \ldots, n_s$, where n_s is a predetermined number of possible locations. The width of the histogram intervals is computed such that the frequency count in each interval is the number of model predictions that lie within the error threshold if the measured value is at the midpoint of the interval. This is done by dividing the maximum range of prediction values, $max\{ (y_{max,j} - y_{min,j}) : j = 1, \ldots, n_s\}$ of an output variable y into intervals $i = 1, \ldots, N_I$ (N_I the maximum number of intervals at the jth location) of width W equal to the sum of measurement errors e_{measur} and modeling errors e_{mod} (Figure 2). The intervals create subsets of model predictions that are then used to compute probabilities and entropy. Each subset represents model predictions that, given a potential measurement, will not be possible to separate further.

Figure 2. Constructing subsets of model predictions of width W for measurement location j using modeling and measurement errors.

2.2. Hierarchical Sensor Placement

A hierarchical sensor placement strategy has been used to identify optimum sensor locations that increase the entropy in wind predictions. High entropy at a sensor location represents uniform distribution of model predictions among the intervals, which means that the sensor location has high potential to separate model predictions. Given a measurement at this location a large number of models will be falsified and a small number of candidate models will remain. Locations are selected iteratively during sensor placement in order to maximize the entropy of the sensor configuration. The advantage of employing a hierarchical strategy is reduced computational cost through the use of an efficient data structure. The proposed algorithm is described in pseudo-code in Algorithm 1.

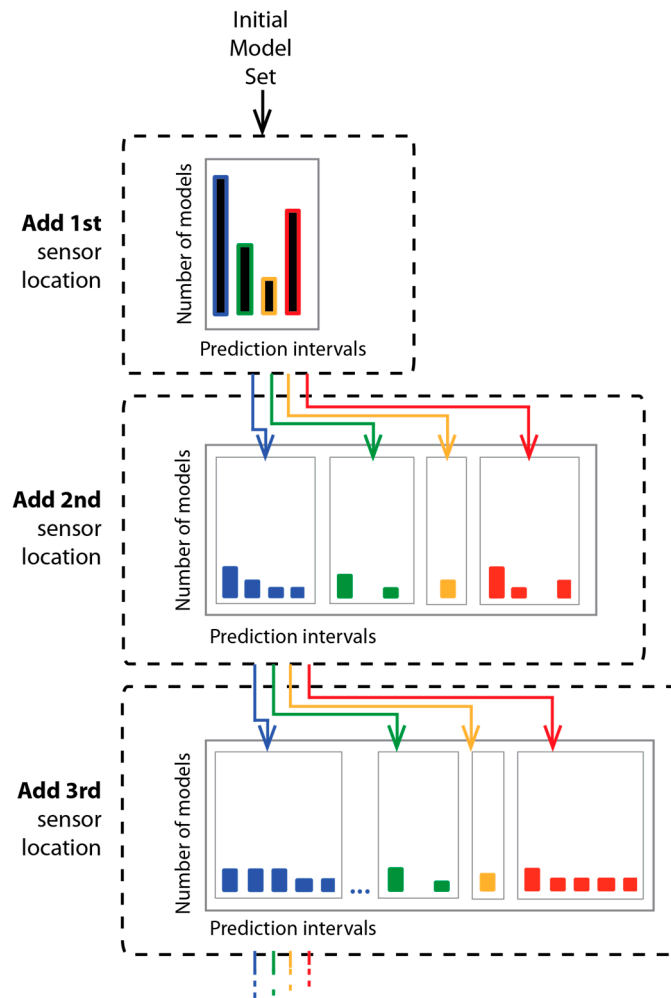**Algorithm 1.** Pseudo-code of the hierarchical algorithm.

1: Create a list locationList containing all possible locations.

2: Create a set sensorOptimum to store all possible sensors. The set is empty to start with.

3: Create a set modelSubsets to store subsets of models that cannot be separated using the current
sensor configuration. To start with this set contains a single element, which is the initialModelSet.

4: Add the first sensor location that corresponds to maximum entropy to sensorOptimum.

5: Create a list of subsets of models that cannot be separated by the first sensor location and add these
subsets to modelSubsets. Remove the initialModelSet from modelSubsets.

6: Repeat while locationList is not empty
{

7:     Select a sensor location from locationList, let it be currentLocation.

8:     Repeat for each set in modelSubsets
{

9:       Divide and distribute models in the current set into intervals of the currentLocation.
}

10:     Calculate the entropy of the distribution of the currentLocation.
}

11: Select the sensor location with maximum entropy. Add to sensorOptimum and remove from
locationList.

Model data is organized in a tree structure in which the initial model set (called initialModelSet in the algorithm description) is at the root, and branches contain subsets of model predictions (modelSubsets). Branches from a node in the tree represent division of the parent model set into smaller groups that can potentially be separated using measurements from the new sensor that is added to the configuration at each level in the tree. The number of model subsets that cannot be further separated with the sensor configuration at each stage (sensorOptimum) is stored for evaluating the performance of the configuration (more details are provided in Section 3.3).

Figure 3 provides a schematic of the hierarchical sensor placement strategy proposed in this work. At the top of the figure is shown the intervals of model predictions at the first sensor location. Each interval contains a subset of models, which is shown in a different color. When the second sensor is added to the configuration, the subset in each interval is further subdivided. The rectangular box in the middle of Figure 3 shows the intervals of model predictions of each subset at the second sensor location. This process is repeated to form a hierarchy of model subsets.

A tree data structure follows a hierarchical organization and takes advantage of an O(constant) computational complexity. At each stage of the sensor placement, a location is added to the configuration sensorOptimum that divides the existing subsets of model predictions into smaller subsets. The maximum number of divisions is restricted to the number of models within the prediction subsets.

Figure 3. Schematic of the hierarchical sensor placement strategy.

2.3. Joint Entropy as a Design Criterion

The information obtained from measured data is clearly a major criterion for selecting sensor locations and this can be evaluated using entropy from information theory (also known as Shannon's entropy or Information entropy). The importance of entropy is that it is a measure of uncertainty in parameter values, since it evaluates disorder in predictions. Here entropy is defined as:

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

where $H(y)_j$ is the entropy of a random variable y at a measurement location j, $p(y_i)_j$ is the probability of the ith interval of a variable's distribution with $i = 1, \ldots, N_I$ and N_I the maximum number of intervals at the jth location. In order to compute the entropy, the number of models that lie within each interval m_i is calculated and the probability of the interval is calculated as (m_i /N).

Equation (1) is used to calculate the entropy of a variable at one sensor location. However, during sensor placement more than one location is selected. Adding sensors to the sensor configuration requires evaluating the common information between multiple sensor locations; it avoids selecting locations that are redundant. For example, the next sensor location having the highest entropy might

contain substantially similar information as the previous sensor location. Therefore, selecting this sensor location does not improve the information obtained.

Joint entropy is a measure of uncertainty associated with multiple variables. It requires evaluating multiple sensor configurations while including the mutual information of data. In order to calculate joint entropy of two sensor locations j and $(j + 1)$, the models that lie within each interval of location j are further divided into sub-intervals using values of location $(j + 1)$ resulting in a rectangular grid (Figure 4).

When this process is repeated for more sensors, a multi-dimensional grid corresponding to each combination of intervals is obtained. Then, the probability of each sub-interval is calculated by dividing the number of models in the sub-interval by the total number of models. The entropy between two sensor locations j and $(j + 1)$, and the relation to mutual information, $I(y)_{j,j+1}$ is defined as:

$$H(y)_{j,j+1} = -\sum_{k=1}^{N_K}\sum_{i=1}^{N_I} p((y_i)_j, (y_i)_{j+1}) \log_2(p((y_i)_j, (y_i)_{j+1})) \tag{2}$$

$$H(y)_{j,j+1} = H(y)_j + H(y)_{j+1} - I(y)_{j,j+1} \tag{3}$$

where $i = 1, ..., N_K$ and N_K the maximum number of intervals at the location $(j + 1)$.

Figure 4. Example of a two-dimensional regular grid created using the intervals of two sensor locations j and $(j + 1)$.

From Equations (2) and (3) in order to calculate joint entropy all prediction subsets of the two sensor locations must be evaluated. In this work sensor placement is performed prior to measuring and no data is available, therefore all prediction subsets need to be evaluated. In forward sequential strategies, when the number of sensors increases, computation cost increases exponentially if a multi-dimensional

regular grid is used to organize model subsets. This is because probabilities have to be summed up over every combination of intervals corresponding to each variable [34]. In contrast, in a hierarchical strategy, when a new sensor is added to the configuration, the subsets of model predictions either remain the same or are further subdivided (Figure 3) causing the probability to be further divided. The hierarchical strategy analyses how the initial models are distributed within subsets, which allows calculations of joint entropy of the sensor configurations, thereby avoiding exponential complexity.

3. Results

The sensor placement strategy was applied to lab-type building called BubbleZERO, which is an experimental facility of rectangular geometry, with dimensions 4.88 m × 6.06 m × 2.9 m, located at NUS Campus in Singapore. Wind modeling and CFD simulations around BubbleZERO were performed with ANSYS Workbench 14.5, a platform that offers FLUENT as a solver for the equations of flow and design exploration tools for sensitivity analysis and feature selection.

In the first stage of simulations, simplifications were made according to recommendations [11,12] in the geometrical representations of the BubbleZERO and the surrounding obstacles, as well as in the numerical methods that control the solver. The simulation volume, called computational domain, represented the atmospheric boundary domain with dimensions 220 m × 140 m × 40 m (Figure 5). The entire domain was decomposed into finite elements, using the CutCell meshing method, within which an approximate solution was sought (Figure 6). CutCell meshing is as a discretization method that generates a high percentage of hexahedral elements with minimum user input; it results in a quicker solver run time and better convergence compared to tetrahedral meshes. The SIMPLE algorithm was employed in order to achieve pressure-velocity coupling and a second-order discretization scheme in order to interpolate pressure values from the elements center to the faces. A single-precision solver was selected as sufficiently accurate for this study.

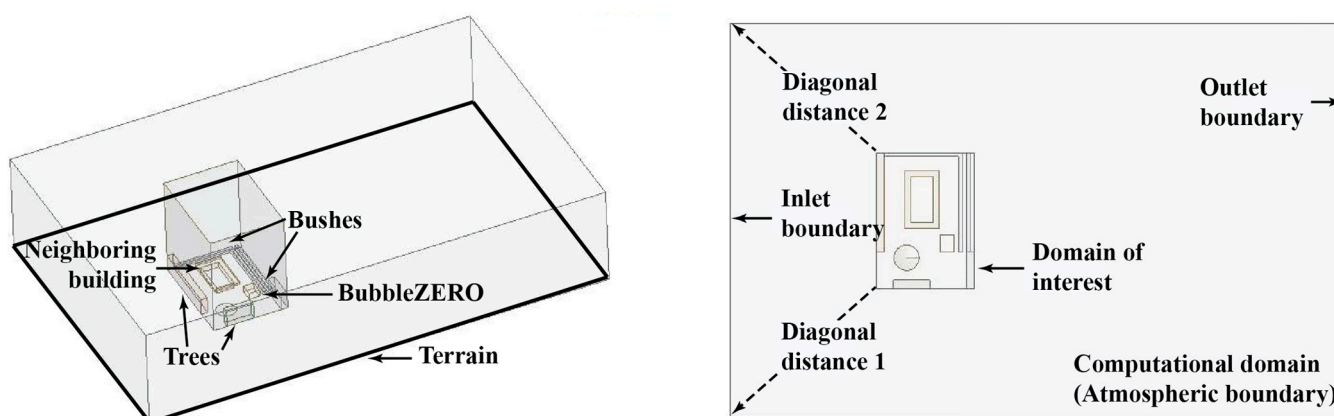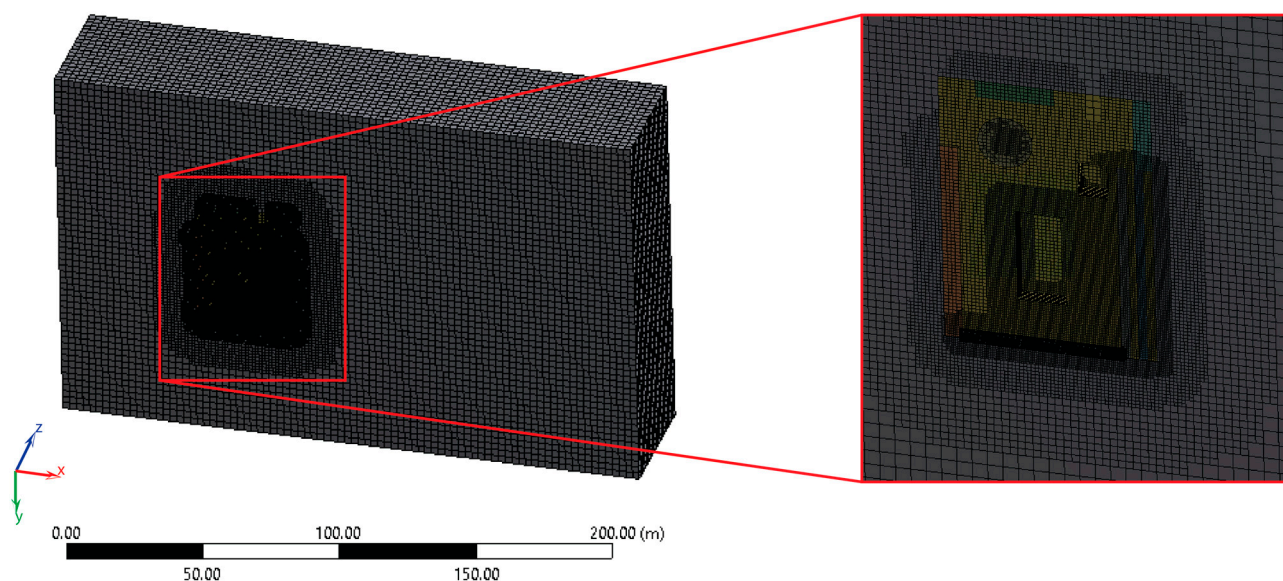**Figure 5.** 3D (left) and plan (right) views of the computational domain.

Figure 6. CutCell Cartesian meshing for the computational domain; bottom view (left) and the domain of interest magnified (right).

In the quantification stage, the behavior of the system was characterized by a set of mathematical models, parameters, variables and constants that describe flow motion. The mathematical models were selected in order to minimize computational cost and were: the steady RANS-equations, the realizable k-ε equations to represent turbulence and the standard wall-functions to treat near-wall turbulence. In total 15 parameters were selected related to the geometry, meshing as well as parameters for wall boundary conditions, such as terrain and surface roughness, for porous boundary conditions, such as the inertial resistance of vegetation and for atmospheric boundary conditions, such as wind speed, wind direction, turbulence kinetic energy (TKE) and turbulence eddy dissipation (TDE). Details of the parameters and their values are shown in Table 1 (in Fluent the term boundary conditions is a general term used to describe bounds between fluid and solid regions; for instance the terrain roughness is a boundary-condition parameter of the wall boundary: terrain [34]). Plausible ranges were specified for all parameter values based on engineering judgment and literature where available. For instance, the mesh growth rate, the rate at which the mesh grows away from the boundary, varied from 1.05 to 1.1 resulting in approximately 5.2×10^5 and 10.4×10^5 mesh elements.

The Equations (1)–(4) were used to describe boundary conditions [13,14,37]:

$$U(z) = \frac{u_* \, ln\left(\frac{z+z_0}{z_0}\right)}{\kappa} \tag{1}$$

where $U(z)$ 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:

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

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

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

where $\varepsilon(z)$ 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, $k_{s,ABL}$:

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

(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 was modeled as porous media, C, with inertial resistance set in the x- and y-direction as [38]:

$$C = C_d d_{SA}$$

(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 [39].

Table 1. The 15 parameters in the CFD simulations and their ranges of values.

Parameter in CFD simulations	Lower bound	Upper bound	Comments
Height of computational domain [m]	40	88	
Diagonal distance 1 from inlet boundary [m]	83	117	The lower and upper bounds were set according to ([11,12,35])
Diagonal distance 2 from inlet boundary [m]	83	117	
Mesh growth rate	1.05	1.1	
Terrain roughness of computational domain [m]	0.5	1	The lower and upper bounds were set according to ([36]), for suburbs (lower bound) to regularly-build large towns (upper bound).
Terrain roughness of area of interest [m]	3×10^{-3}	0.3	The lower and upper bounds were set according to ([36]), for concrete surfaces (lower bound) to long grass (upper bound).
Surface roughness of BubbleZERO [m]	0.03	1	A double and inflatable membrane from PTFE was installed on the outside of the BubbleZERO. The lower bound corresponded to the typical PTFE value and the upper bound to the height of the inflated PTFE.
BubbleZERO doors [m]	2×10^{-5}	0.16	The lower and upper bounds were typical thickness of glazing and wooden frames.
Inertial resistance of trees	0.1	3.5	The lower and upper bounds were set according to ([30])
Inertial resistance of bushes	0.1	5.2	
Wind direction at inlet boundary [deg]	1	360	The wind direction varied from 1 to 360 degrees in order to account for possible direction values.
Wind speed at inlet boundary [m/s]	0	7.2	The lower and upper bounds were set according to meteorological data obtained from the weather station Changi WMO in Singapore.
TKE at inlet boundary [J/kg]	0	7.2	The lower and upper bounds were set according to ([11,12]).
TDE at inlet boundary [m²/s³]	0	1.3	

Multiple, steady-state CFD simulations were run varying values of the identified parameters within the plausible ranges shown in Table 1. In order to reduce computational complexity, sensitivity analysis was performed using ANSYS DesignXplorer: an Optimal Space-Filling design [40] and CCD sampling [41] were selected that reduced that number of simulations to 283. The simulations output variables were wind speed and wind direction and their output distributions were built as full second-order polynomial response-surface, which are expressed as a function of the input parameters. Predictions of wind speed and direction were obtained at 63 possible sensor locations, fixed uniformly and in close distance to the BubbleZERO (Figure 7). These locations were selected in order to be in proximity to the BubbleZERO. The dimensions of the BubbleZERO, the measurement equipment characteristics and the orography were considered during the selection of the possible locations.

Figure 7. Possible measurement locations displayed in the simulation environment: 3D view on the left and plan view on the right.

Spearman's rho correlation coefficient ρ_j (Equation (6)) was calculated between the 15 parameters and the wind predictions over all measurement locations. The wind speed, the wind direction and the turbulence kinetic energy at the inlet boundary were identified as the features with the highest impact on wind predictions (with average coefficients of 0.35, 0.35 and 0.8, respectively, for both output variables and over all locations):

$$\rho_j = -\frac{\sum_k (x_{k,j} - \bar{x}_j)(y_{k,j} - \bar{y}_j)}{\sqrt{\sum_k (x_{k,j} - \bar{x}_j)^2 (y_{k,j} - \bar{y}_j)^2}} \tag{6}$$

where $x_{k,j}$, $y_{k,j}$ are 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.

A second set of multiple steady-state CFD simulations were performed varying values of these features within plausible ranges using a simple-grid sampling and selecting values uniformly within the ranges. A set of 1024 combinations of values was created that were used to run simulations and obtain a discrete population of wind predictions at the 63 sensor locations. This population of model instances of wind speed and wind direction formed the initial model set.

A hierarchical sensor placement strategy was employed using the initial model set in order to reveal optimal sensor configurations. Since model predictions were used in sensor placement, systematic

modeling errors, as well as spatial correlations between errors, were considered. Recent research in our group has demonstrated that the range of modeling errors can vary from location to location between [−0.6, +0.4] and [−1, +0.8] m/s for wind speed and [−30, +30] and [−180, +180] deg for wind direction, depending on boundary conditions and sensor locations [33]. Indeed, errors associated with wind direction can be the most that is possible—up to 180 degrees both ways—due to the RANS-based modeling used in this work, since time-averaged equations of flow motion are carried out. Although steady-state RANS is one of the most computational efficient approaches to approximate turbulent flows, thereby allowing multiple simulations to be run, it does not model small-scale local vortices that occur in reality due to local disturbances. The following systematic modeling errors with non-uniform spatial distribution were used for wind speed, $e_{mod,speed}$, and wind direction, $e_{mod,dir}$, following [33]:

$$e_{mod,speed} = \begin{cases} [(-0.33 \cdot U(z) - 0.12), (0.34 \cdot U(z) + 0.36)\,], if\ \dfrac{u_j(z)}{U(z)} < 1 \\[2em] [(-0.18 \cdot U(z) - 0.12), (0.12 \cdot U(z) + 0.34)\,], if\ \dfrac{u_j(z)}{U(z)} \geq 1 \end{cases} \tag{7}$$

$$e_{mod,dir} = \{(1.18 \cdot U(z) - 37.06), (2.09 \cdot U(z) + 26.44)\,\} : \dfrac{u_j(z)}{U(z)} \geq 0.33 \tag{8}$$

where $u_j(z)$ is the wind speed at height z at possible sensor locations $j \in \{1, ..., 63\}$. Wind direction with $\frac{u_j(z)}{U(z)} < 0.33$ were not considered since modeling errors were high (around ± 180 deg).

3.1. Effect of Modeling Error

The effect of modeling error on sensor placement is evaluated through comparing a sensor placement strategy that includes spatial variations in modeling errors against a strategy that assumes uniform values for errors at every location. Variations in error values are defined according to Equations (7) and (8), while uniform error-values are set constant and equal to the upper and lower bounds of the estimated ranges (Section 3). Measurement errors depend on the characteristics of the measurement equipment and for this case, error ranges are set to 0.1 m/s for wind speed and 22.5 deg for wind direction. At each stage of the sensor placement, a location was added to the optimum configuration and the joint entropy in wind predictions was calculated.

Figure 8 shows a comparison of the calculated joint entropy in wind-speed predictions of optimal sensor configurations, when either spatially uniform (± 0.4 and ± 1 m/s) or varying modeling errors (Equation (7)) are considered. The increase in joint entropy is higher when the spatial variation in modeling errors is considered during sensor placement. In addition, entropy values are found to be higher when compared with using uniform and constant values of modeling errors, particularly when modeling errors are large.

Figure 8. Comparison of the joint entropy in wind-speed predictions calculated during sensor placement; errors in predictions were considered either spatially uniform (±0.4 and ±1 m/s) or varying (only the first 15 optimum locations are displayed in the graph).
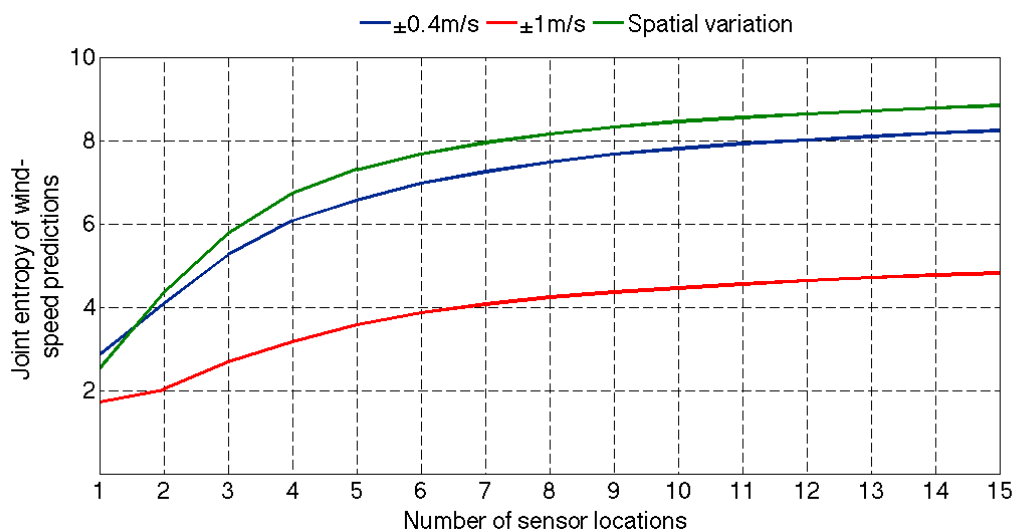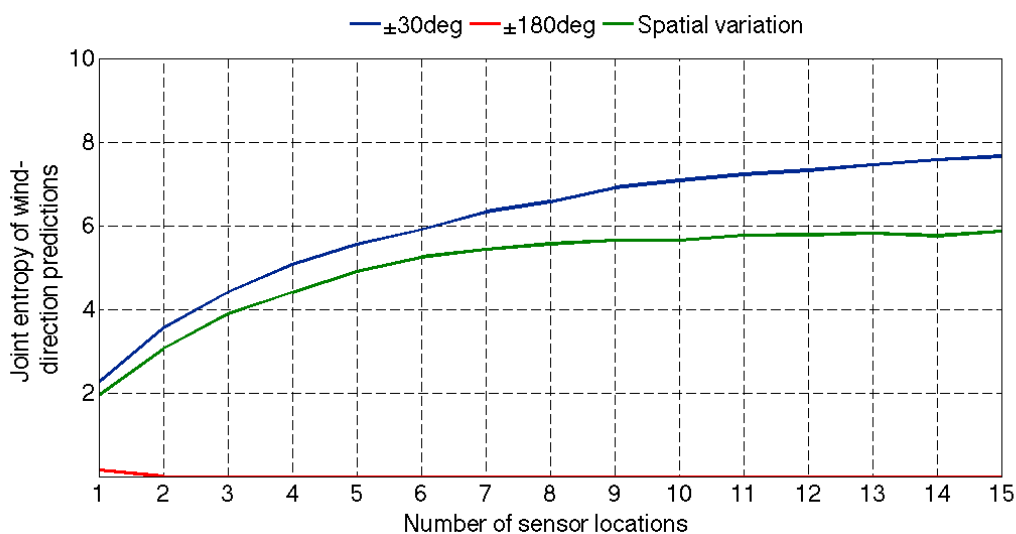

Figure 9 presents the joint entropy in wind-direction predictions of optimal sensor configurations, when either spatially uniform ($\pm$30 and $\pm$180 deg) or varying modeling errors are considered (Equation 8). Although using small and uniformly distributed errors ($\pm$30 deg) leads to a high increase in joint entropy, this increase stabilizes slower than using spatial variations in errors. Finally, assuming large and spatially uniform modeling errors ($\pm$180 deg), does not provide any optimum sensor location.

Figure 9. Comparison of the joint entropy in wind-direction predictions calculated during sensor placement; errors in predictions are taken to be either spatially uniform (±30 and ±180 deg) or varying (only the first 15 optimum locations are displayed in the graph).

The results demonstrated that during sensor placement, the joint entropy in wind predictions is influenced by the spatial distribution of modeling errors that is assumed. In contrast with wind speed, using small and uniform errors identifies locations of that provide high entropy increase in wind

direction. However the entropy increase stabilizes faster for both wind speed and wind direction when spatially variations in errors are included.

3.2. Optimum Sensor Configurations

Figure 10 provides a comparison of the calculated joint entropy in wind predictions using optimum configurations for wind speed and wind direction and including spatial variations in modeling errors. Overall, the entropy in wind speed is higher than that in wind direction, for the same number of sensor locations. For the purpose of this paper, an incremental change in joint entropy below half a unit is taken to be insignificant. This occurs after the 6th sensor is added to the configurations for wind-speed and wind-direction predictions.

Figure 11 shows the expected maximum number of candidate models of wind speed and wind direction during sensor placement; spatial variations in modeling errors are included. Less than 1/5 of the model instances of wind speed and wind direction are retained using a sensor configuration of two sensors. Similarly to Figure 10, the incremental reduction in the maximum number of candidate models stabilizes after the 4th sensor for both wind-speed and wind-direction predictions. The maximum number of candidate models of wind speed and of wind direction is retained to around 10% of the initial model set after the 4th sensor location was selected.

Figure 12 illustrates the optimum configurations of four sensors for wind-speed and wind-direction predictions in the simulation environment. The selected sensor locations for predicting wind speed are different from that for predicting wind direction: for wind speed locations are selected near all façades except the south, while for wind direction locations are selected near all façades except the east. One sensor location is commonly identified as optimal for predicting both wind speed and direction—location L16. Locations L39 and L37 have the same position, although L39 is at 2.7 m height and L37 at 0.6 m height (Table 2).

> **Figure 10**. A comparison of the joint entropy in wind-speed and wind-direction predictions calculated during sensor placement; errors in predictions vary spatially (only the first 15 optimum locations are displayed in the graph).
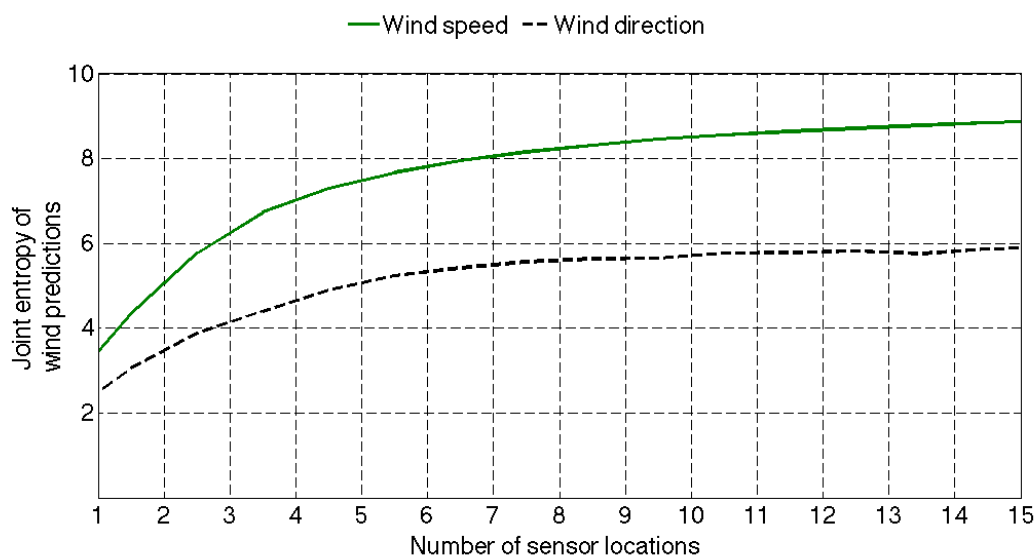
Figure 11. A comparison of the maximum number of candidate models of wind-speed and wind-direction that is expected during sensor placement; errors in predictions vary spatially (only the first 15 optimum locations are displayed in the graph).
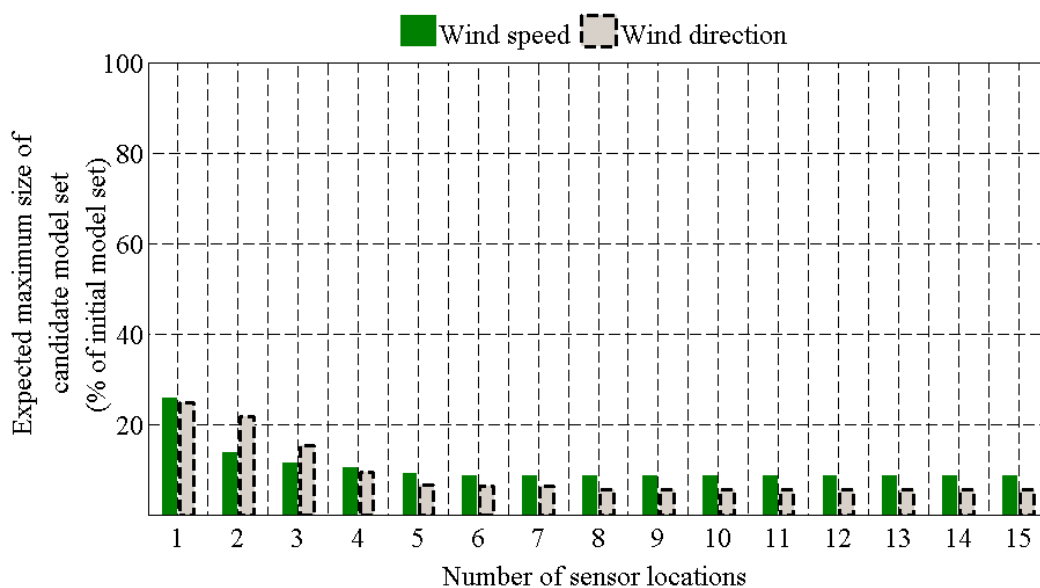

Figure 12. The optimum configurations of four sensors for wind speed (left) and wind direction (right) displayed in the simulation environment; the markers represent the selected sensor locations.
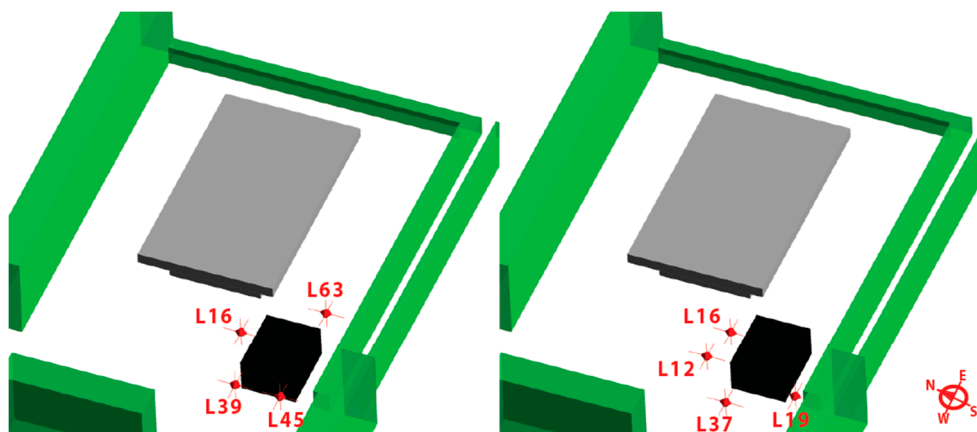

Table 2. The selection order of the optimum configurations of four sensors for predicting wind speed and wind direction.

Selection order	Sensor location	
	Wind speed	**Wind direction**
1st	L45	L12
2nd	L16	L16
3rd	L63	L19
4th	L39	L37

3.3. Performance Evaluation

The performance of the sensor placement strategy is evaluated for its ability to reduce the number of models, minimize the prediction range and increase the accuracy of predictions. Measurements taken with the optimum sensor configurations are compared with model predictions. Models whose predictions do not match measurements at any location are falsified in order to obtain candidate models at each point in time. Candidate models are used to predict wind speed and wind direction at an unseen location that has been randomly selected. The resulting prediction range is compared with the measurements at the unseen location. The objective is to show that when using the optimal sensor configuration the prediction ranges are narrow, while they still contain data.

Since no measurements were available at the optimum locations, simulated measurements are used. They are created for both wind speed and wind direction at the identified optimum locations through combining predictions from the initial model set with modeling and measurement errors. In order to create a more realistic distribution of simulated measurements, historically measured data from a full scale-study at other sensor locations were available and are used. The objective is to replicate trends observed in measured data in simulated measurements through ensuring that the distribution of simulated measurements follows the probability distribution of the measured data at the same locations. Therefore simulated measurements are not obtained through taking random values. Instead, they are created by sampling simulation predictions of models whose predictions at locations where sensors are available match past measurements.

3.3.1. Wind-Flow Predictions

Figure 13 presents a comparison of wind-speed prediction ranges at an unseen location obtained with the optimum configuration of four sensors and the simulated measurements at this location. Prediction ranges are shown as a grey area and simulated measurements at each time instant with points. A 15-minute period is taken from a 2-hour total prediction period. On average a 54% reduction in prediction ranges of wind-speed is achieved. The number of candidate models is 76, whereas the initial model set is 1024. Finally, good prediction accuracy is achieved since 88% of the simulated measurements are within the prediction range.

Figure 14 presents a comparison of wind-direction prediction ranges at an unseen location obtained with the optimum configuration of four sensors and the simulated measurements at this location. Similarly to Figure 13, prediction ranges are shown as a grey area and simulated measurements at each time instant with points. To account for wind-direction discontinuities, two prediction ranges are used; a 15-minute period is taken from a 2-hour total prediction period. In contrast with wind speed predictions, the reduction in prediction ranges of wind-direction is 36% and the identified number of candidate models is 88. Moreover, the prediction accuracy is lower, since 42% of the simulated measurements are within the prediction range. From the results shown in Figure 10, in order to reduce the entropy change below 0.5, a minimum of six sensors is required. Therefore, two additional locations are added to the optimal configuration. These are locations L22 and L7, selected near the south and north façades respectively.

Figure 13. Comparison of the wind-speed prediction ranges at an unseen location obtained with the optimum configuration of four sensors and the simulated measurements at this location; 15 min are taken from the 2 h measurement period.
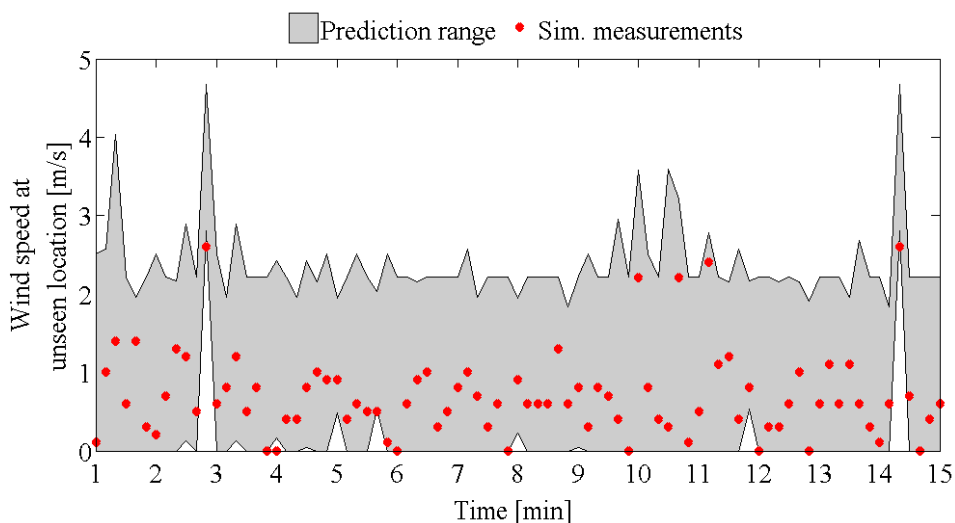

Figure 14. Comparison of the wind-direction prediction ranges at an unseen location obtained with the optimum configuration of four sensors and the simulated measurements at this location; 15 min are taken from the 2 h measurement period.
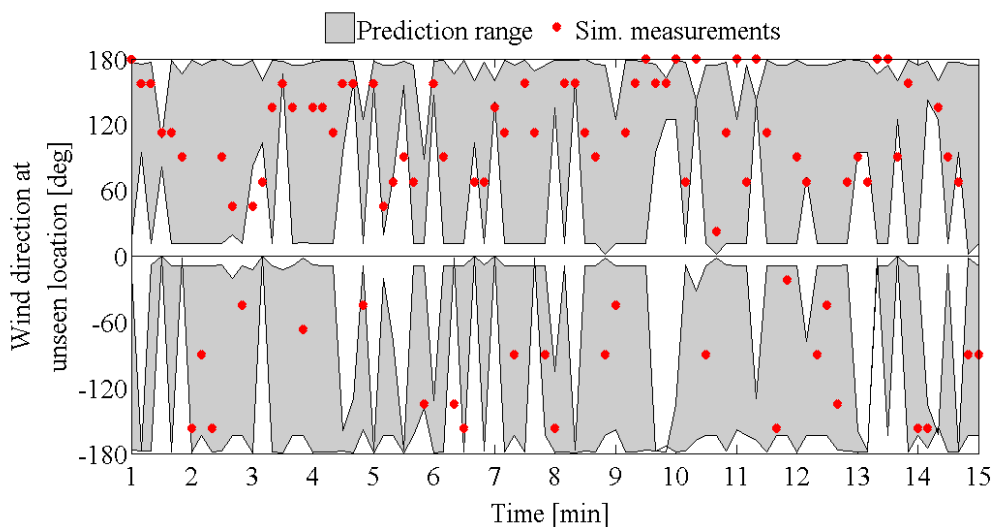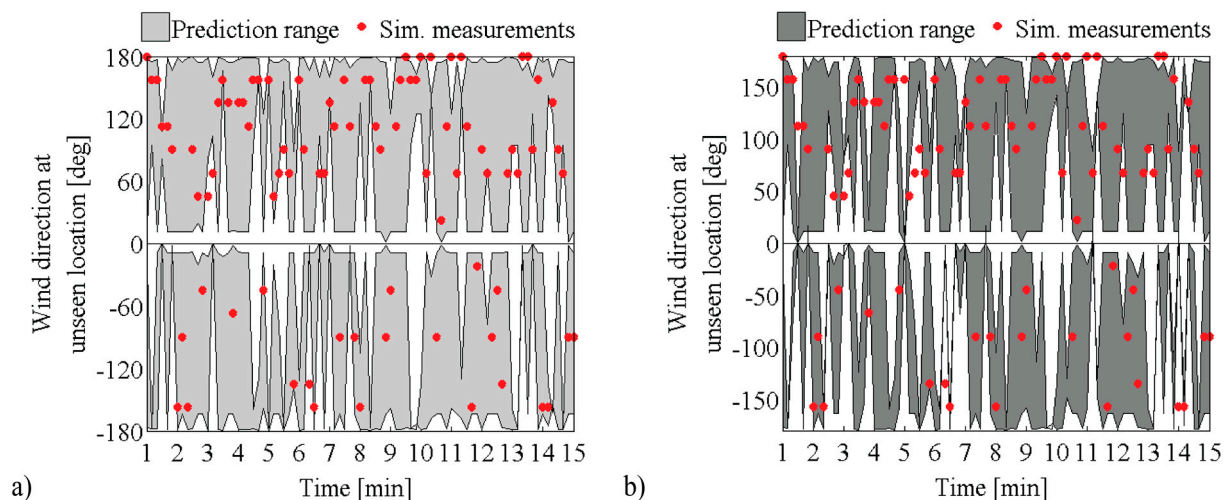

Figure 15 presents a comparison of wind-direction prediction ranges at an unseen location obtained with optimum configurations of (a) four and (b) six sensors and the simulated measurements at this location.

The wind-direction prediction ranges obtained using the optimum configuration of (a) four sensors and (b) six sensors are shown as light and dark grey areas respectively. A slight improvement is achieved in minimizing prediction ranges and reducing the number of candidate models with six sensors: prediction ranges are reduced on average by 47% and the candidate models are 69. However, the prediction accuracy is also reduced to 39%.
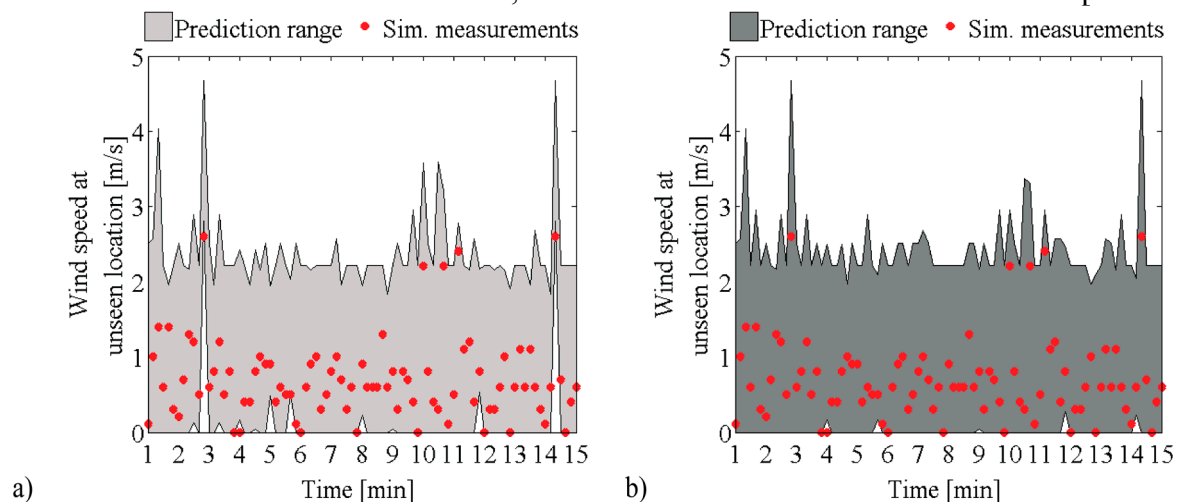
Figure 15. Comparison of the wind-direction prediction ranges at an unseen location obtained using the optimum configuration of (**a**) four sensors and (**b**) six sensors and the simulated measurements at this location; 15 min are taken from the 2 h measurement period.

a)   b)

3.3.2. Sequential *vs.* Hierarchical Sensor Placement

The performance of the hierarchical sensor placement strategy was compared against the sequential strategy proposed in [22] using the three metrics: prediction ranges, number of candidate models and prediction accuracy. Figure 16 shows the results of the comparison between the two strategies for predicting wind-speed at an unseen location, using the optimum configurations of four sensors. The wind-speed prediction ranges obtained using the hierarchical strategy are presented in (a), while the prediction ranges obtained using the sequential sensor placement strategy are presented in (b). The sequential strategy provides lower scores than the hierarchical strategy for two out of three metrics. The sequential strategy achieves a smaller reduction in the prediction range of 52%, although the candidate models are 136, which is almost two times higher than the results using the hierarchical strategy. The accuracy of predictions shows a marginal increase to 89%.

Figure 16. Comparison of the wind-speed prediction ranges at an unseen location obtained using (**a**) hierarchical and (**b**) sequential optimum configurations of four sensors and the simulated measurements at this location; 15 min are taken from the 2 h measurement period.

a)   b)

4. Discussion

Optimal sensor configurations for predicting wind behavior around buildings are identified using hierarchical sensor placement and calculating the joint entropy in predictions. The performance of the optimal sensor configurations is evaluated in terms of minimizing ranges of wind predictions, reducing the number of candidate models and increasing the accuracy of predictions.

The hierarchical sensor placement strategy proposed in this paper has been inspired by limitations found in previous studies. Sequential strategies, such as those proposed in [22,23] are advantageous when compared with global search strategies with regard to computational cost. However, joint-entropy calculations may be computationally prohibitive and as a result, locations with similar information content can be selected. Moreover, the above studies were evaluated for identification of structural systems. Sensor placement strategies have not yet been proposed for predicting behavior of time-dependent systems, such as wind studies around buildings.

An important contribution of this work is that the effect of systematic modeling errors and spatial variations in errors are evaluated and incorporated in the sensor placement strategy. Others have also stressed the effect of spatial correlation of prediction errors on entropy-based sensor placement [25]. However, in this approach entropy calculations have been based on probabilities of model parameter values, which are difficult to apply to wind studies where multiple models and parameter values are able to explain measurements.

A limitation of this work is that the same measurement error has been used for all sensors. In reality there are different types of sensors with different characteristics. Measurement errors associated with wind speed may also be different than those for wind direction. Moreover, during measurement campaigns, a single sensor configuration for measuring simultaneously wind speed and wind direction is common. Although in wind studies the number of available sensors is usually fixed as well as their locations, optimal sensor configurations for wind speed and wind direction differ not only in sensor locations but also in their number. A sensor placement strategy that combines these aspects in terms of information entropy is currently under study. Finally, determining the number of sensors should take into account the trade-off between the incremental increase in entropy and the cost of adding new sensors. This is a topic for future research.

5. Conclusions

A hierarchical sensor placement strategy using joint entropy as a design criterion is successfully employed to predict wind characteristics around buildings and capture short-term wind variability. Overall it is shown that a hierarchical placement strategy using a joint-entropy design criterion can better predict wind speed at un-measured locations compared with sequential algorithms that maximize entropy at each stage. Moreover, it has been demonstrated that by correctly modeling the spatial distribution of modeling errors higher values of joint entropy are obtained. The current methodology shows that in contrast with wind speed, wind direction cannot be predicted with reasonable accuracy for any number of this type of sensor. Finally, the optimal sensor configurations used to predict wind speed are different from those for wind direction and the latter are more sensitive to the magnitude of modeling errors.

Acknowledgments

Author Contributions

All authors contributed to the conception and design of the study, have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. *World Urbanization Prospects The 2011 Revision*; Department of Economic and Social Affairs, Population Division, United Nations: New York, NY, USA, 2012.
2. Mochida, A.; Lun, I.Y.F. Prediction of wind environment and thermal comfort at pedestrian level in urban area. *J. Wind Eng. Ind. Aerodyn.* **2008**, *96*, 1498–1527.
3. Balczó, M.; Gromke, C.; Ruck, B. Numerical modeling of flow and pollutant dispersion in street canyons with tree planting. *Meteorol. Z.* **2009**, *18*, 197–206.
4. Gousseau, P.; Blocken, B.; Stathopoulos, T.; van Heijst, G.J.F. 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. *Atmos. Environ.* **2011**, *45*, 428–438.
5. Blocken, B.; Janssen, W.D.; van Hooff, T. Cfd simulation for pedestrian wind comfort and wind safety in urban areas: General decision framework and case study for the Eindhoven university campus. *Environ. Model. Softw.* **2012**, *30*, 15–34.
6. Chen, Q. Ventilation performance prediction for buildings: A method overview and recent applications. *Build. Environ.* **2009**, *44*, 848–858.
7. Britter, R.; Hanna, S. Flow and dispersion in urban areas. *Annu. Rev. Fluid Mech.* **2003**, *35*, 469–496.
8. Schatzmann, M.; Leitl, B. Issues with validation of urban flow and dispersion cfd models. *J. Wind Eng. Ind. Aerodyn.* **2011**, *99*, 169–186.
9. *Can Theories be Refuted?: Essays on the Duhem-Quine Thesis*. Harding, S.G., Ed; D. Reidel Publishing Company: Dordrecht, The Netherlands, 1976.
10. Blocken, B.; Stathopoulos, T.; Carmeliet, J.; Hensen, J.L.M. Application of computational fluid dynamics in building performance simulation for the outdoor environment: An overview. *J. Build. Perform. Simul.* **2011**, *4*, 157–184.
11. Franke, J. *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 Institute: Oslo, Norway, 2007.

12. Tominaga, Y.; Mochida, A.; Yoshie, R.; Kataoka, H.; Nozu, T.; Yoshikawa, M.; Shirasawa, T. Aij guidelines for practical applications of cfd to pedestrian wind environment around buildings. *J. Wind Eng. Ind. Aerodyn.* **2008**, *96*, 1749–1761.

13. Richards, P.; Hoxey, R. Appropriate boundary conditions for computational wind engineering models using the k-ε turbulence model. *J. Wind Eng. Ind. Aerodyn.* **1993**, *46*, 145–153.

14. Richards, P.; Norris, S. Appropriate boundary conditions for computational wind engineering models revisited. *J. Wind Eng. Ind. Aerodyn.* **2011**, *99*, 257–266.

15. Jakeman, A.J.; Letcher, R.; Norton, J. Ten iterative steps in development and evaluation of environmental models. *Environ. Model. Softw.* **2006**, *21*, 602–614.

16. Ljung, L. *System Identification Toolbox: User's Guide*; MathWorks Incorporated: Natick, MA, USA, 1988.

17. Oreskes, N.; Shrader-Frechette, K.; Belitz, K. Verification, validation, and confirmation of numerical models in the earth sciences. *Science* **1994**, *263*, 641–646.

18. Raphael, B.; Smith, I.F.C. Finding the right model for bridge diagnosis. *Artif. Intell. Struct. Eng.* **1998**, *145*, 308–319.

19. Raphael, B.; Smith, I.F.C. *Fundamentals of Computer-Aided Engineering*; Wiley: New York, NY, USA, 2003.

20. Goulet, J.A.; Coutu, S.; Smith, I.F.C. Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks. *Adv. Eng. Inf.* **2013**, *27*, 261–269.

21. Kripakaran, P.; Smith, I.F.C. Configuring and enhancing measurement systems for damage identification. *Adv. Eng. Inf.* **2009**, *23*, 424–432.

22. Robert-Nicoud, Y.; Raphael, B.; Smith, I.F.C. Configuration of measurement systems using Shannon's entropy function. *Comput. Struct.* **2005**, *83*, 599–612.

23. Papadimitriou, C. Optimal sensor placement methodology for parametric identification of structural systems. *J. Sound Vib.* **2004**, *278*, 923–947.

24. Goulet, J.A.; Smith, I.F.C. Performance-driven measurement system design for structural identification. *J. Comput. Civ. Eng.* **2012**, *27*, 427–436.

25. Papadimitriou, C.; Lombaert, G. The effect of prediction error correlation on optimal sensor placement in structural dynamics. *Mech. Syst. Signal Proc.* **2012**, *28*, 105–127.

26. Tamura, T. Towards practical use of LES in wind engineering. *J. Wind Eng. Ind. Aerodyn.* **2008**, *96*, 1451–1471.

27. Pavageau, M.; Schatzmann, M. Wind tunnel measurements of concentration fluctuations in an urban street canyon. *Atmos. Environ.* **1999**, *33*, 3961–3971.

28. Van Hooff, T.; Blocken, B. Full-scale measurements of indoor environmental conditions and natural ventilation in a large semi-enclosed stadium: Possibilities and limitations for cfd validation. *J. Wind Eng. Ind. Aerodyn.* **2012**, *104*, 330–341.

29. Hamel, D.; Chwastek, M.; Farouk, B.; Dandekar, K.; Kam, M. A Computational Fluid Dynamics Approach for Optimization of a Sensor Network. In Proceedings of the 2006 IEEE International Workshop on Measurement Systems for Homeland Security, Contraband Detection and Personal Safety, Alexandria, VA, USA, 18–19 October 2006; pp. 38–42.

30. Mokhasi, P.; Rempfer, D. Optimized sensor placement for urban flow measurement. *Phys. Fluids* **2004**, *16*, 1758–1764.

31. Du, W.; Xing, Z.; Li, M.; He, B.; Chua, L.H.C.; Miao, H. Optimal sensor placement and measurement of wind for water quality studies in urban reservoirs. In Proceedings of the 13th international symposium on Information processing in sensor networks, Berlin, Germany, 15–17 April 2014; pp. 167–178.

32. Papadopoulou, M.; Raphael, B.; Smith, I.F.C.; Sekhar, C. Optimal sensor placement for time-dependent systems: Application to wind studies around buildings. *J. Comput. Civ. Eng.* **2014**, submitted.

33. Vernay, D.G.; Raphael, B.; Smith, I.F. Augmenting simulations of airflow around buildings using field measurements. *Adv. Eng. Inf.* **2014**, in press.

34. Cover, T.M.; Thomas, J. *Elements of Information Theory*; Wiley: New York, NY, USA, 1991.

35. *ANSYS Fluent User's Guide*; Ansys Inc.: Canonsburg, PA, USA, 2011.

36. Wieringa, J. Updating the davenport roughness classification. *J. Wind Eng. Ind. Aerodyn.* **1992**, *41*, 357–368.

37. Shih, T.-H.; Liou, W.W.; Shabbir, A.; Yang, Z.; Zhu, J. A new k-ϵ eddy viscosity model for high reynolds number turbulent flows. *Comput. Fluids* **1995**, *24*, 227–238.

38. Guo, L.; Maghirang, R.G. Numerical simulation of airflow and particle collection by vegetative barriers. *Eng. Appl. Comput. Fluid Mech.* **2012**, *6*, 110–122.

39. Tiwary, A.; Morvan, H.P.; Colls, J.J. Modelling the size-dependent collection efficiency of hedgerows for ambient aerosols. *J. Aerosol Sci.* **2006**, *37*, 990–1015.

40. *Design Exploration User Guide*; Ansys Inc.: Canonsburg, PA, USA, 2011.

41. Box, G.E.; Hunter, J.S. Multi-factor experimental designs for exploring response surfaces. *Ann. Math. Stat.* **1957**, *28*, 195–241.