1

Evaluating predictive performance of sensor configurations in wind studies around buildings

Maria Papadopoulou[a,*]; Benny Raphael[c]; Ian F.C. Smith[a,d] and Chandra Sekhar[b]

a) Future Cities Laboratory, ETH Zurich, Zurich, Switzerland.

b) Department of Building, National University of Singapore, Singapore

c) Civil Engineering Department, Indian Institute of Technology, Madras, India.

d) Applied Computing and Mechanics Laboratory, EPFL, Lausanne, Switzerland.

* Corresponding author: Maria Papadopoulou, Future Cities Laboratory, Singapore-ETH

Centre 1 CREATE Way #06-01, CREATE Tower Singapore 138602. Tel.: +65 8287 0922, E-

mail: papmara@gmail.com

ABSTRACT

1. Abstract

A great challenge associated with urban growth is to design for energy efficient and healthy built environments. Exploiting the potential for natural ventilation in buildings might improve pedestrian comfort and lower cooling loads, particularly in warm and tropical climates. As a result, predicting wind behavior around naturally ventilated buildings has become important and one of the most common prediction approaches is computational fluid dynamics (CFD) simulation. While accurate wind prediction is essential, simulation is complex and predictions are often inconsistent with field measurements. Discrepancies are due to the large uncertainties associated with modeling assumptions, as well as the high spatial and temporal climatic variability that influences sensor data. This paper proposes metrics to estimate the expected predictive performance of sensor configurations and assesses their usefulness in improving simulation predictions. The evaluations are based on the premise that measurement data are best used for falsifying model instances whose predictions are inconsistent with the data. The potential of the predictive performance metrics is demonstrated using full-scale high-rise buildings in Singapore. The metrics are applied to assess previously proposed sensor

1   configurations. Results show that the performance metrics successfully evaluate the

2   robustness of sensor configurations with respect to reducing uncertainty of wind predictions

3   at other unmeasured locations.

4   *Keywords*: Computational fluid dynamics (CFD); System identification; Sensor

5   configuration; Performance evaluation; Simulation prediction

6   # 2. Introduction

7   The continuous growth of the global population living in cities has increased interest in

8   outdoor thermal comfort [1], air quality [2, 3], safety [4], and particularly in warm climates,

9   building energy consumption and natural ventilation [5]. The wind environment has a primary

10   role in mitigating these issues and, consequently, improving knowledge of wind behavior

11   around buildings has been the focus of much recent research work (a detailed review can be

12   found in [6]). The most common approach for wind prediction is based on computational

13   models, such as those used in computational fluid dynamics (CFD) simulations.

14   Today, CFD simulations are used to overcome constraints of laboratory and field

15   measurements [5], since they provide detailed information on wind flow and allow treatment

16   of complex geometries with a high degree of repeatability. Although CFD simulations

17   provide reasonable predictions, the accuracy is not superior to laboratory and field

18   measurements [6]. Uncertainties are large in both modeling and measurements and should be

19   taken into account [7].

20   Tamura [8] and Schatzmann et al. [9] suggested that measurements, both laboratory and field,

21   should be used in a complementary manner to ensure that simulation results are sound, even

22   when using modeling methods of high predictability, such as the large eddy simulation (LES)

23   [8]. However, the use of simplified arrays of roughness elements in laboratory measurements

24   results in idealized representations of the parameters affecting wind flow [10], and it is often

25   unclear how sensitive wind predictions are to these parameter uncertainties [11]. Moreover,

26   certain physical phenomena, such as buoyancy-driven natural ventilation, cannot be fully

27   represented in reduced-scale laboratory experiments [5].

28   In these situations, field measurements are essential for ensuring that modeling is sound,

29   especially in studies involving high-rise buildings [10, 11]. Nevertheless, field measurements

30   have been rare, and sensor placement still remains a challenging task [12]. Most of earlier

1   studies have used historically measured data and only a few configured full-scale
2   measurement campaigns [13-16]. In most cases, sensors locations have been selected based
3   on educated guesses, although some researchers have investigated optimal sensor
4   configurations using systematic and data-driven strategies [17-20]. Such strategies require
5   prior knowledge of data distributions and spatial correlations, obtained from a denser pre-
6   deployment of sensors, a task that is however both costly and time-intensive. A recent study
7   in Singapore has used the concept of maximum entropy for sensor selection, in a similar way
8   to [21], yet modeling and measurement data were assumed to be free of errors and sensor
9   placement was performed iteratively, disregarding the mutual information content between
10   sensor locations.

11   All of these studies employed data from sensor locations placed outside the urban canopy, or
12   on building rooftops, although reports on methods to obtain representative data suggest that,
13   unless interested in topographic-generated climate patterns, sensor locations subject to local
14   and mesoscale effects should be avoided and target measurement areas of acceptable
15   homogeneity at screen-level (~1.5 m) or high-level (about the roughness sub-layer) should be
16   selected [22]. Wind flow varies considerably over space and time and measurements within
17   the urban canopy depend on the location of sensors and sampling frequency [23, 24]. In
18   addition, it has been shown [7] that even under steady ambient conditions, large discrepancies
19   occur between measured and predicted values that are caused by low frequency variations of
20   the wind flow.

21   A recent study done by the authors explored systematic sensor placement strategies that are
22   applicable to time-dependent wind prediction within the urban canopy, involving buildings of
23   varying size and use [25]. The study adapted and compared sequential strategies and criteria
24   used in the field of infrastructure diagnosis, which can achieve high levels of accuracy with
25   low computational cost compared to global search strategies [26] and genetic algorithms [27].
26   The typical information-based criteria found in literature for optimal sensor configuration
27   were information entropy [26-28], cost and expected identifiability [29], while some studies
28   incorporated uncertainty correlations and their values [29, 30]. Based on the conclusions of
29   the authors' initial study [25], a novel hierarchical sensor placement strategy has been
30   developed that uses the concept of joint-entropy to account for the mutual information
31   between sensor locations [31]. The strategy also explicitly incorporated the spatial
32   distributions of modeling errors and their values, which has been shown to affect optimal
33   sensor configuration [30-32].

1   In conclusion, typical information-based criteria used for optimal sensor configuration
2   focused primarily on increasing information value, measured either with entropy from
3   information theory or using identifiability metrics to reduce parameter-value uncertainty.
4   Nonetheless, earlier studies [33] have suggested that the performance of sensor configurations
5   in reducing uncertainty of model predictions should be assessed by additional criteria, such as
6   robustness-to-uncertainty and "prediction-looseness"—equivalent to the range of
7   predictions—which are often conflicting.

8   This paper proposes metrics to evaluate the predictive performance of sensor configurations
9   and assess their usefulness in improving simulation predictions. The performance of the
10 configurations is assessed for their capability to falsify multiple model instances whose
11 predictions are inconsistent with the data (Section 3.1). Expected identifiability metrics found
12 in literature [29] are adapted (Section 3.2) and then new metrics are developed (Section 3.3)
13 to estimate the sensor configurations robustness-to-uncertainties associated with model
14 predictions. In the end, a multi-criteria decision-making (MCDM) approach is proposed to
15 evaluate the influence of the conflicting metrics on the choice of optimal sensor configuration
16 (Section 3.4). In Section 4, the proposed metrics are applied to evaluate the performance of
17 several hierarchically-constructed sensor configurations [31], in improving wind predictions
18 around a full-scale building system in Singapore. A list of the main conclusions and a critical
19 assessment on the results are provided in the final two sections (section 5 and 6).

3. Development of predictive performance metrics

21 Metrics are developed to estimate the expected predictive performance of sensor
22 configurations and assess their usefulness in improving predictions. The study builds upon
23 previous work on sensor placement performed by the authors [25, 31], where the premise is
24 that sensor data are best used for falsifying multiple model instances[1] whose predictions are
25 inconsistent with the data. Therefore the performance of sensor configurations is assessed in
26 terms of their capability to falsify multiple model instances (§3.1). First, expected
27 identifiability metrics are adapted from literature to estimate the expected reduction in the
28 number of candidate models (retained model instances) and in the prediction range associated
29 with each sensor configuration (§3.2). Several optimal sensor configurations, which have
30 been constructed using the hierarchical-sensor placement strategy, have been compared in
31 [31]. Metrics are then developed to estimate the robustness of the sensor configurations to

---

[1] A specific combination of values for the input parameters in a CFD simulation and the corresponding wind predictions at all potential locations is one *model instance*.

1 uncertainty associated with model predictions (§3.3). In the end, an MCDM approach is
2 proposed to evaluate the influence of multiple metrics on the selection of optimal sensor
3 configurations (§3.4).

3.1. Falsification based on multiple models

5 Falsification of multiple model instances is performed at each sensor location of the
6 hierarchically-constructed sensor configurations using simulated measurements generated
7 following the procedure described in [25]. As justified in these earlier studies done by the
8 authors ([31], [25]), the reason for employing simulated measurements is that the evaluations
9 of sensor locations are done prior to measuring and therefore data at these sensor locations are
10 not yet available. In addition, the performance of several sensor configurations needs to be
11 evaluated and compared at the same time instant, which would require a costly pre-
12 deployment of a large number of sensors.

13 During falsification, multiple model instances are rejected if the difference between their
14 predicted values and the measurements falls outside defined threshold bounds. These bounds
15 correspond to confidence intervals that include plausible model instances. The interval width
16 at each sensor location is equal to the prediction range, obtained from the model instances,
17 and estimates of modeling and measurement errors.

18 The term model instance refers to a computational model in which input parameters are
19 assigned a definite combination of values and the corresponding values of output variables
20 are predicted using simulation. The core of the methodology is the multiple-model approach
21 introduced by Raphael and Smith [34].

22 $$P\left(\cap_{i=1}^{n_m} T_{low,i} \leq U_{c,i} \leq T_{high,i}\right) \tag{1}$$

23 where φ is the confidence level, $U_{c,i}$ is obtained from by subtracting the modeling $U_{mod,i}$ and
24 measurement $U_{meas,i}$ uncertainty, $T_{low,i}$ and $T_{high,i}$ are the computed threshold bounds equal
25 to the minimum and maximum values of the combined modeling and measurement residuals,
26 with n_m the number of measurements used.

27 An illustration of the falsification process is shown in **Fig. 1**. The remaining model
28 instances—called candidate models—are used to update the predictions at other, unmeasured,
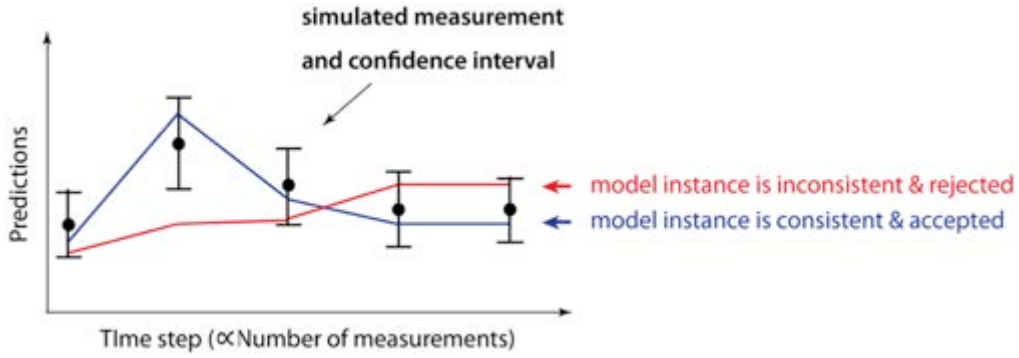29 locations and enhance inference (more information on this process is available in [25, 31]).

1

2 **Fig. 1.** Model falsification using the simulated measurements from one sensor location and predictions
3 from two model instances. One model instance is rejected because its prediction at one time step falls
4 outside the confidence interval of the measured value.

5 Since optimal sensor configurations often include more than one sensor locations, model

6 instances are evaluated multiple times. However, a complication arises when confidence

7 intervals result from multiple testing, which is known as the inflation of the alpha level [35]

8 The Šidák correction (Equation (1)) is therefore employed to correct the alpha values, $\alpha =$

9 $1 - \varphi$, for the threshold bounds in order to maintain a constant confidence level $\varphi = 0.95$.

10 The Šidák correction has been successfully applied in the field of infrastructure diagnosis [29]

11 to counteract the effect of the inflation of the alpha level during multiple hypothesis testing.

12 This adjustment will affect the number of candidate models estimated after the sensor

13 addition, since it corrects for the models instances that have been falsely rejected.

14 $\quad \alpha' = 1 - (1 - \alpha)^{1/N}$ (2)

15 where N is the number of sensors.

16 The expected number of candidate models obtained during model falsification is a trade-off

17 between the number of sensor locations, the confidence level (related to α) and the threshold-

18 bound width, which depends on measurement and modeling errors. This trade-off is

19 quantified with the expected identifiability metrics described below.

3.2. Expected identifiability

21 Expected identifiability metrics have been used in literature [29] to predict probabilistically to

22 what extent a sensor configuration is useful in falsifying model instances. These metrics have

23 been employed to select optimal sensor locations, whereas in this work they are adapted for

1 evaluating the performance of existing sensor configurations in improving simulations
2 predictions.

3 The expected identifiability of configurations is calculated with respect to the expected
4 number of candidate models and their range of predictions at unmeasured locations.
5 Empirical cumulative distribution functions of the number of candidate models and prediction
6 range are built for each sensor configuration using a sample set of simulated measurements.
7 Two quantities are then extracted from the cumulative distributions, the maximum number of
8 candidate models and the maximum prediction range. These quantities express the expected
9 identifiability of the current sensor configuration and depend on the confidence level, φ
10 (often fixed at 95%), the number of sensors n_s, and modeling e_{mod} and measurement errors
11 e_{meas}, since these define the width of the falsification thresholds.

12 **Fig. 2** shows example cumulative distributions that are used to estimate the maximum number
13 of candidate models, as a percentage of the initial-model-set size. **Fig. 2** (a) represents the
14 trade-off between the number of sensors, n_s, and the level of confidence. **Fig. 2** (b) represents
15 the effect that variations in modeling error, e_{mod}, have on the expected maximum number of
16 candidate models of a fixed configuration. The same example can be provided by displaying
17 on the horizontal axis the prediction range as a percentage of the initial possible range. For a
18 fixed confidence level of 95%, the expected maximum number of candidate models (or
19 prediction range) decreases with an increase in the number of sensors ($n_{s2} > n_{s1}$) and with a
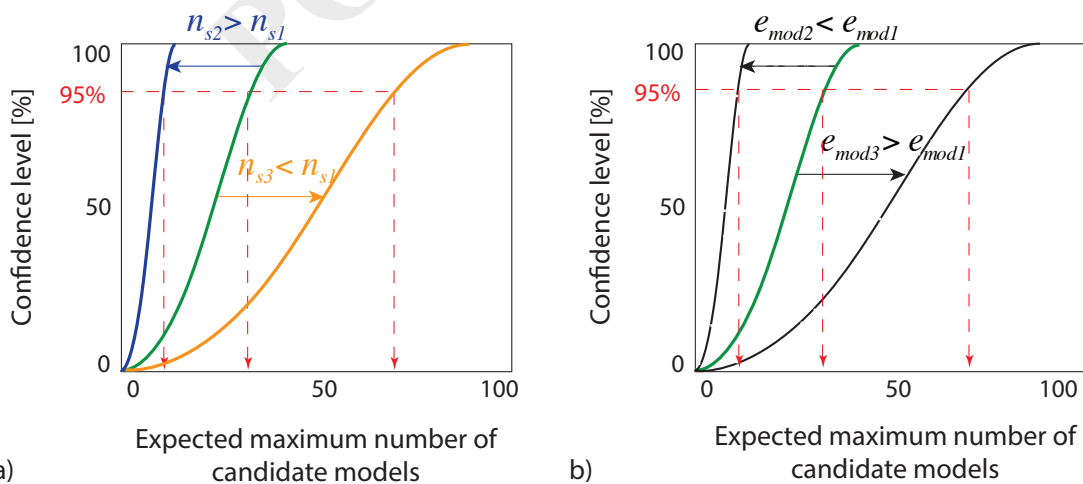20 decrease in modeling errors ($e_{mod2} < e_{mod1}$).

21 a)   b)

22 **Fig. 2.** Examples of cumulative distribution function used to compute the expected identifiability of the
23 candidate-model-set size for a target confidence level depending on (a) the number of sensors and (b)
24 modeling uncertainty (modified from [29]).

7

1   Although expected identifiability metrics are useful for providing information on the

2   performance of sensor configurations with respect to reducing the number or candidate

3   models and the prediction ranges, they do not provide information on the success of the

4   identification. This represents the number of correct predictions that are made at an

5   unmeasured location using the information obtained from the current configuration. Such a

6   metric evaluates the robustness of the configuration with respect to the uncertainty associated

7   with predictions and is discussed in the following section.

8   3.3. Robustness to prediction uncertainties

9   Model falsification is performed using threshold bounds that are equal to modeling and

10   measurement uncertainties. Therefore, if modeling errors are misevaluated, there is a

11   probability that a model is falsely rejected, a Type I error. This probability is estimated by

12   counting the number of simulated measurements that fall outside the predicted range at an

13   unmeasured location, signifying that a correct model instance has been wrongly rejected. On

14   the other hand, the probability of falsely accepting a model instance, a Type II error, and is

15   equal to the size of the candidate model set minus one—the correct model instance.

16   The probability of a Type I and Type II error is calculated at an unmeasured location for each

17   sensor configuration, in a similar manner to Section 3.2. The empirical cumulative

18   distribution functions are used to extract the probability values of Type I and Type II error, as

19   well as the prediction range (% of initial plausible range), for a 95% certainty. **Fig. 3** (a),

20   illustrates an example of the conflicting relationship between the probability of prediction

21   error with the number of sensors in a configuration: more sensors decrease the number of

22   models that are falsified (Type II), yet the chance of making false predictions (Type I errors)

23   increases. There is however a number of sensors that provide an equal probability of

24   committing either error. Similarly **Fig. 3** (b) illustrates the expected maximum prediction

25   range (as a percentage of the initial plausible range), with the number of sensors: the larger

26   the number of sensors in the configuration the lower the range of predictions. However, the

27   minimum range of predictions depends on the range of measurement data and is restricted by

28   the range of modeling uncertainty. Overall, such metrics reveal the conflicting relationships

29   between the probabilities of committing prediction errors and assess the robustness of sensor

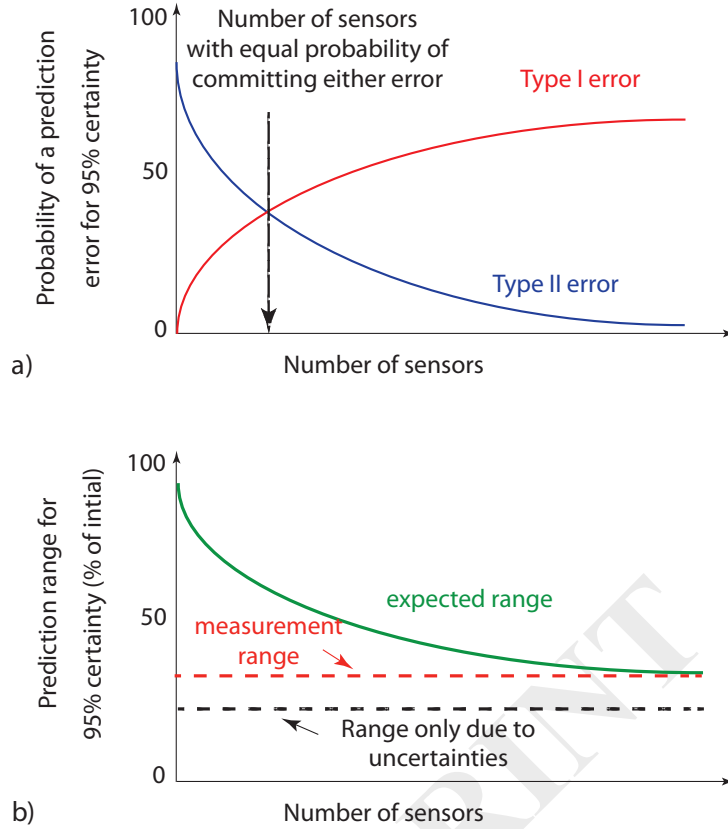30   configurations to misevaluation of modeling uncertainties.

1

2　**Fig. 3. (a)** The relationship between the number of sensors and the probability of a Type I and Type II
3　error; **(b)** the expected range of predictions, the measurement range and the uncertainty range, with the
4　number of sensors; both are displayed for a 95% confidence level.

5　In comparison with the metrics described in Section 3.2, the above robustness-to-uncertainty

6　metrics are conflicting criteria to consider in an optimal sensor configuration problem. The

7　following section proposes a multi-criteria decision making approach (MCDM) that allows

8　assessment of the influence of these criteria on the selection of optimal sensor configurations.

9　3.4. Multi-criteria decision making

10　MCDM approaches are concerned with solving decision problems that typically involve

11　multiple conflicting criteria. Therefore an optimal sensor placement problem can be treated

12　with an MCDM approach to take into account the multiple, conflicting criteria that influence

13　optimal sensor configuration.

14　**Table 1** lists a set of 5 criteria that influence an optimal sensor placement problem based on

15　the earlier studies. These criteria are ranked according to a decision maker's preference on

16　measurement objectives: minimize cost, reduce the range of predictions and the prediction

17　errors (Type I and Type II), and maximize joint-entropy of predictions. Among these, cost

1 conflicts with joint-entropy and prediction range, and Type I errors conflicts with Type II

2 errors.

3 **Table 1** Criteria evaluated during MCDM for selecting optimal sensor configurations.

Order	Criterion	Description
1	Cost	Proportional to the number of sensors in the configuration
2	Prediction range	Expected range of simulation predictions obtained using the current sensor configuration
3	Type I error	Probability of falsely rejecting a model instance ($\propto$ number of measurements that fall outside the prediction range)
4	Type II error	Probability of falsely accepting a model instance ($\propto$ number of candidate models minus one)
5	1/Joint-entropy	Inverted joint-entropy of predicted values of the sensor configuration

4 The RR-Pareto3 algorithm (in improved version of RR-Pareto algorithm [36]) is used to

5 identify a good compromise solution by first ordering the criteria according to their

6 importance, normalizing the criteria-values and finally filtering through bisecting the range of

7 values, one at a time and according to their order, until a unique solution is found. **Fig. 4**

8 illustrates an example where the RR-Pareto3 algorithm is applied to assess the 5 criteria of

9 **Table 1** for 10 sensor configurations, in order to obtain a compromise configuration of 4
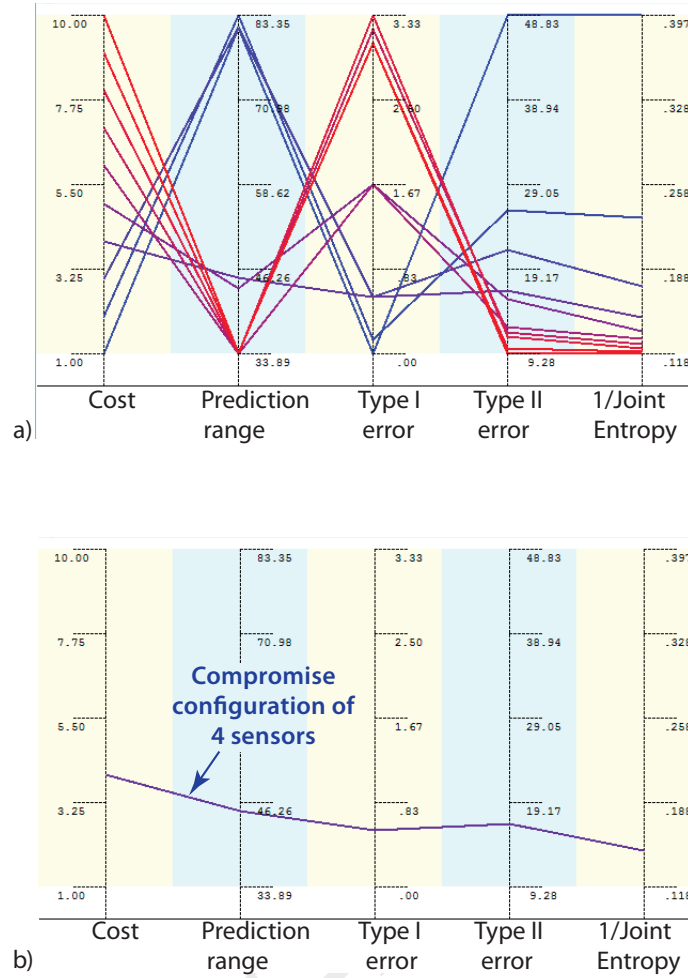
10 sensors.

Fig. 4. Example of RR-Pareto3 algorithm used to evaluate (a) 5 sensor selection criteria for 10 optimal sensor configurations and (b) the best compromise sensor configuration of 4 sensors.

Typically in an MCDM problem there is no single optimal solution and a set of best compromise solutions is provided. This is illustrated with an example in **Fig. 5**, where the RR-Pareto3 algorithm recognized a set of 10 compromise optimal configurations, of 7 to 16 sensors, from the initial set of 60 configurations, with respect to the 5 criteria of **Table 1**. The decision maker can then decide using additional information on a single optimal configuration that meets the measurement needs.
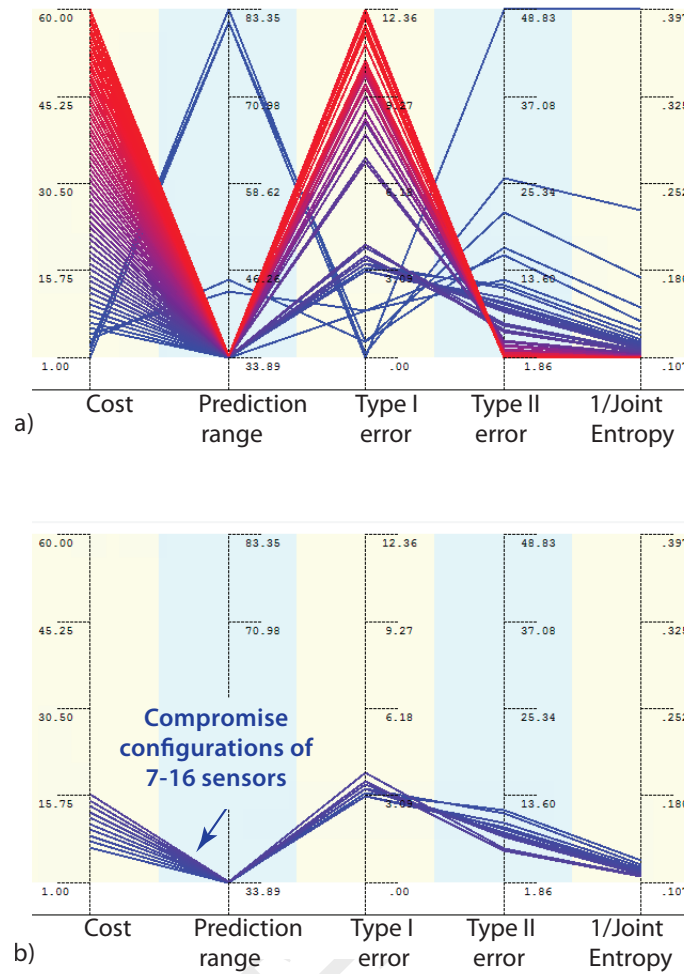
1

Fig. 5. Example of RR-Pareto3 algorithm used to evaluate (a) 5 sensor selection criteria for 60 optimal sensor configurations and (b) the set of 10 best compromise sensor configurations of 7 to 16 sensors.

Often in wind studies several variables are measured with the same sensor configuration, such as wind speed and wind direction, which necessitates that a common configuration is found between these variables. A solution involves calculating criteria values for all possible sensor configurations and variables, and employing the RR-Pareto3 algorithm to find the best compromise solution among them. Such calculations however, have high computational costs that are related to the number of variables and criteria.

In the presence of a decision maker's preferences, the above MCDM problem can be solved in a hierarchical manner with low computational cost: first, variables are ranked in order of importance; then the RR-Pareto3 algorithm is employed to obtain a reduced set of compromise sensor configurations, from the initial set of optimal configurations, using the criteria values of the variable of highest importance (i.e. wind speed); this set of compromise configurations is used to recalculate the criteria values for the second variable in order (i.e.

1   wind direction); the RR-Pareto3 algorithm is rerun to recognize a reduced set of compromise

2   configurations that satisfy the measurement objectives for both variables. When applicable,

3   additional variables or information are used to further reduce the number of compromise

4   sensor configurations until a single optimal configuration is obtained.

5 # 4. Application to wind prediction around buildings

6   The predictive performance metrics for sensor configurations were applied to a full-scale

7   residential estate, called Treelodge, which is located in Punggol, Singapore (**Fig. 6** (a)). The

8   estate consists of seven 16-story building blocks with an average height of 63 m. Treelodge is

9   Singapore's first experimental ecofriendly public housing project [37].
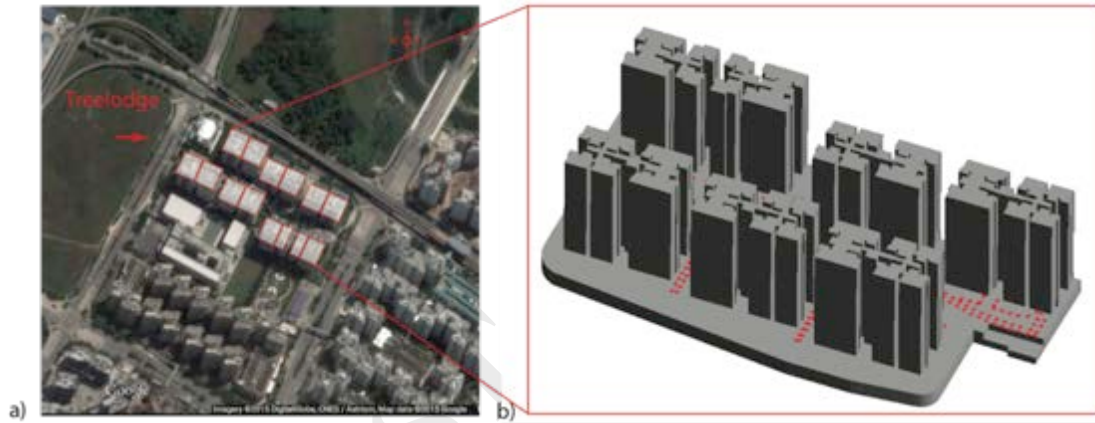
10

11   **Fig. 6. (a)** The buildings of the Treelodge estate at Punggol considered in modeling and **(b)** their 3D
12   view in the simulation environment.

13 ## 4.1. Multiple-model CFD simulation

14   The experimental design for multiple CFD simulations is organized in a similar way to [25,

15   31]: FLUENT 14.5 is used as a solver for the equations of flow behavior and the Design

16   Exploration tool for sensitivity analysis and parameter selection.

17   The geometric simplifications and assumptions related to the numerical methods that control

18   the solver are made as follows:

19   1. The geometries of the domain of interest consist of 7 building blocks with average

20   dimensions 26 x 53 x 63 m (**Fig. 6** (b)). The orography of the area is assumed to be

21   uniform and surface details of buildings are omitted.

1   2. The entire size of the computational domain (or ABL domain) is 955 m x 1598 m x
2   504 m (**Fig. 7** (a)), according to recommendations available in literature [38, 39].

3   3. The CutCell Cartesian meshing is used as a discretization method to generate a
4   predominantly hexahedral mesh with minimum user input (**Fig. 7** (b)). After grid
5   sensitivity analysis the minimal element size to 0.05 m and the mesh growth rate to
6   1.175, generating 13.4 x 10e6 elements.

7   4. The SIMPLE algorithm is employed to achieve pressure-velocity coupling.

8   5. Second-order discretization is used as a pressure interpolation scheme.

9   6. A single-precision solver is assumed to be sufficiently accurate for this study.

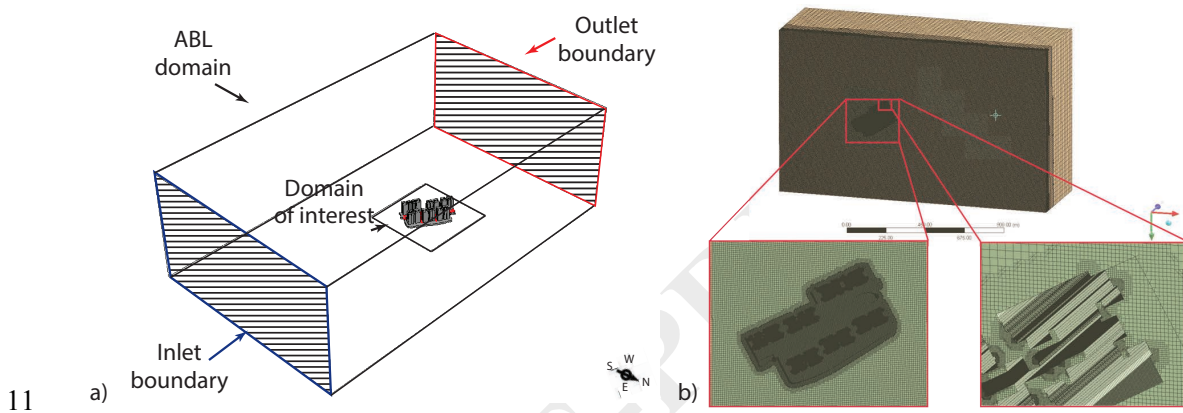10   7. The convergence criteria of the scaled residuals for all variables are set to 10^{-4}.

11 a)

12 **Fig. 7.** View of the ABL domain of the Treelodge@Punggol (a) and bottom view of the CutCell
13 Cartesian meshing with two magnified views (b) in the simulation environment.

14 The multiple-model CFD simulation approach is applied to predict possible wind behavior
15 around the Treelodge estate according to the following assumptions:

16   1. RANS-equations are used to characterize wind behavior around buildings, the
17   realizable k-ε equations to represent turbulence and the standard wall-functions to
18   treat near-wall turbulence, since it is one of the most computationally efficient
19   approaches to approximate turbulent flows.

20   2. The inlet boundary conditions of wind speed are described in Equation (2) and
21   Equations (3) and (4) are used to calculate the TKE and TDE as functions of the
22   varying wind speed profile at the inlet. The sand-grain roughness of the ABL domain
23   and domain of interest are calculated using Equation (5). Finally, the pressure at the
24   outlet boundary is set to zero Gauge pressure.

25 $$U(z) = \frac{u_* \ln\left(\frac{z+z_0}{z_0}\right)}{\kappa} \qquad\qquad (3)$$

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 κ is the von Kármán constant.

$$k = \frac{u_*{}^2}{\sqrt{C_\mu}}, \tag{4}$$

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

$$\varepsilon(z) = \frac{u_*{}^3}{\kappa(z+z_0)} \tag{5}$$

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

$$k_{s,ABL} = \frac{9.793 z_0}{C_s} \tag{6}$$

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).

In total 9 uncertain parameters are found related to the assumptions made in geometric simplifications, meshing and boundary conditions. Ranges of values for these parameters are found in the literature and are based on engineering judgment.

The output variables of the simulations are the wind speed, referring to the magnitude of the horizontal component of the velocity vector, and the horizontal direction. Simulation predictions of wind speed and horizontal direction are obtained at 290 potential sensor locations, which are fixed at 3 m height near the pedestrian eco-deck, which is the canyon separating the building blocks) (**Fig. 8**). Potential locations are selected remote from the buildings and distances are set taking into account the eco-deck length, the size of the measurement equipment and the orography of the domain of interest.

Sensitivity analysis is then carried out in ANSYS Workbench 14.5 with design exploration tools in order to evaluate the parameter uncertainties on predictions and reduce computational cost. An Optimal Space-Filling design [40] with CCD sampling [41] is applied and Spearman's rho correlation coefficient, ρ_j, is calculated from Equation (6) for the 9 uncertain parameters and the output variables, wind speed and horizontal direction, at each potential sensor location shown in **Fig. 8**. Computational cost is reduced through selecting three of the uncertain parameters with the highest $\overline{\rho_j}$ and Z-scores, averaged over all locations. These are the inlet wind speed, the inlet horizontal direction and the surface roughness of buildings, with coefficients 0.76, 0.4 and 0.1 and Z-scores 2.4, 1.3 and 0.23, respectively.

$$\rho_j = -\frac{\sum_k (x_{k,j}-\overline{x_j})(y_{k,j}-\overline{y_j})}{\sqrt{\sum_k (x_{k,j}-\overline{x_j})^2 (y_{k,j}-\overline{y_j})^2}} \tag{7}$$

1   where $x_{k,j}$, $y_{k,j}$ are the ranks of the input parameters and output variables respectively at each

2   location $j \in \{1, \dots, 63\}$, with $k = 1, \dots, n$ the size of the sample and $\overline{x}_j$, $\overline{y}_j$ the mean values.

3   Multiple simulations are the performed taking discrete values for the three parameters within

4   plausible ranges shown in **Table 2**. The reduced number of parameters allows a simple-grid

5   sampling with values selected uniformly within the ranges, with discretization intervals of 0.1

6   m/s, 22.5 deg and 0.25 m. A set of 3648 combinations of parameter values is created and used

7   to carry out CFD simulations. A discrete population of predictions of wind-speed and

8   horizontal-direction at the 290 potential sensor locations (**Fig. 8**) is the output of the

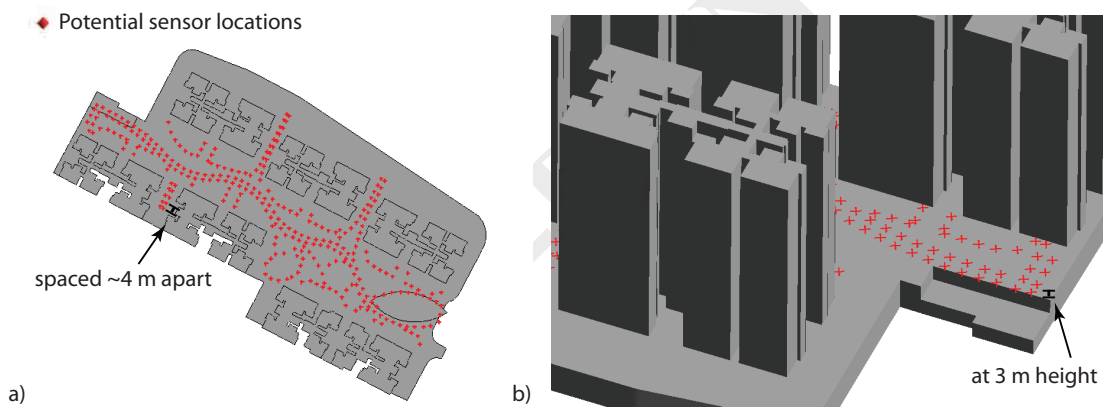9   simulations that comprises the initial model set.

10

11   **Fig. 8.** Potential sensor locations L1-L290 near the Treelodge estate shown in the simulation
12   environment: (a) top view and (b) front, magnified view.

13   **Table 2** The selected parameters and their ranges of values used to generate the initial model set for the
14   Punggol case study.

Parameter in CFD simulations	Lower bound	Upper bound	Comments
Boundary conditions			
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	8.7	The lower and upper bounds were set according to meteorological data obtained from the weather station Changi WMO in Singapore.
Surface roughness of buildings [m]	8E-3	1.12	The lower and upper bounds were set according to ([38, 39]).

16

1   ## 4.2. Hierarchically-constructed sensor configurations

2   The hierarchical sensor placement strategy using joint entropy [31] is applied to the initial

3   model set in order to reveal optimal sensor configurations. Modeling and measurement errors

4   are explicitly incorporated in the strategy, with measurement errors set to ±0.05 m/s for wind

5   speed and ±1 deg for wind direction, while Equations (7) and (8) are used to define modeling

6   errors with non-uniform spatial distribution for wind speed, $e_{mod,speed}$ and wind

7   direction, $e_{mod,dir}$. These equations are based on recent findings of the studies conducted by

8   the Simulation-Platform research group at the Future Cities Laboratory. Simplifications are

9   made similar to [42]:

10   $$e_{mod,speed} = \begin{cases} [(-0.33U(z) - 0.62), (0.34U(z) + 0.86)], & \frac{u_j(z)}{U(z)} < 1 \\ [(-0.18U(z) - 0.62), (0.12U(z) + 0.84)], & \frac{u_j(z)}{U(z)} \geq 1 \end{cases} \quad (8)$$

11   $$e_{mod,dir} = \{(1.18U(z) - 65.06), (2.09U(z) + 30.44)\} : \frac{u_j(z)}{U(z)} \geq 0.33 \quad (9)$$

12   where $u_j(z)$ is the local wind speed at height z at possible sensor locations $j \in \{1, ..., 290\}$

13   and $U(z)$ is the wind speed that would occur without the presence of buildings. Wind-

14   direction predictions with amplification factor $u_j(z)/U(z) < 0.33$ are not considered since

15   modeling errors are high (around ± 180 deg).

16   **Fig. 9** shows a comparison of the calculated (a) joint-entropy in predictions and (b) maximum

17   number of candidate models, of wind speed and horizontal direction during hierarchical

18   sensor placement. Only the first 15 optimal locations are displayed in the graph, yet

19   calculations over all 290 possible sensor locations are considered. An effect similar to [31] is

20   observed, since joint entropy of wind-speed predictions is higher than that of wind-direction

21   predictions, which have relatively low values. **Fig. 10** illustrates the maximum number of

22   candidate models of wind speed and horizontal direction that is estimated during hierarchical

23   sensor placement. The incremental reduction in the maximum number of candidate models

24   flattens after selecting the 5th sensor for wind speed and the 3rd sensor for wind direction. This

25   is equivalent to an incremental change in the number of models of less than 1%. At maximum

26   1/5 of the initial-model-set size of both wind speed and direction is retained using a

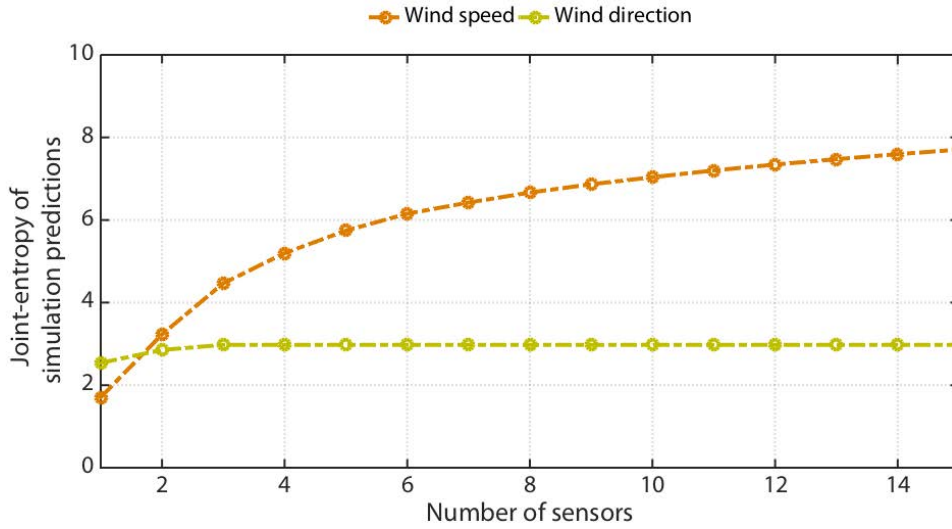27   configuration of five sensors.

Fig. 9. A comparison of the joint entropy of wind-speed and wind-direction predictions calculated for various sensor configurations near Treelodge@Punggol, using the hierarchical algorithm with joint-entropy criterion; spatial variation of predictions errors is considered (only the first 15 optimum locations are displayed in the graph).
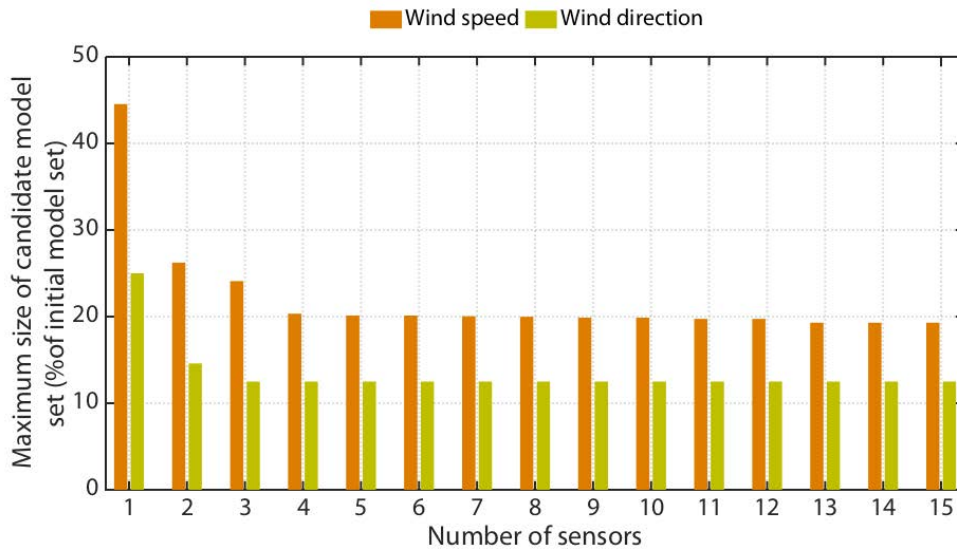

Fig. 10. A comparison of the maximum number of candidate models of wind-speed and wind-direction that is expected for various sensor configurations near Treelodge@Punggol, using the hierarchical algorithm; spatial variation of predictions errors is considered (only the first 15 optimum locations are displayed in the graph).

Fig. 11 illustrates the optimal sensor configurations of the first 5 sensors for predicting (a) wind-speed and (b) wind-direction. As observed in earlier studies done in Singapore ([25, 31]), the optimal locations for the two variables differ: wind-speed locations are selected uniformly within the canopy of buildings, while wind-direction locations are selected near the center and northeast part of the canopy.
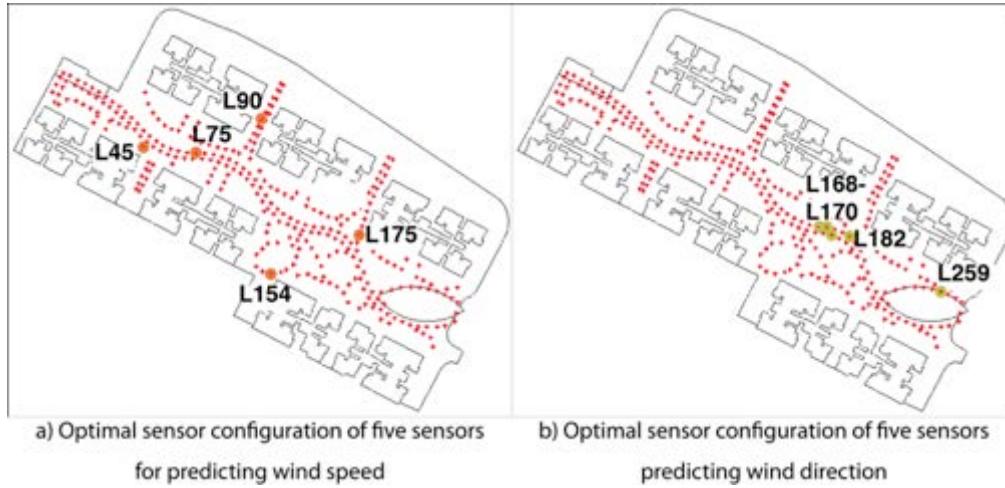
a) Optimal sensor configuration of five sensors for predicting wind speed

b) Optimal sensor configuration of five sensors predicting wind direction

Fig. 11. Optimal configurations of first five sensors for predicting (a) wind speed and (b) wind direction (right) near Treelodge@Punggol, using the hierarchical algorithm with joint-entropy criterion; the circled markers represent the selected locations.

4.3. Evaluation of predictive performance

A measurement campaign was carried out by Shanshan Pan near Treelodge at Punggol [43], as part of research conducted by Nanyang Technological University and the Singapore-MIT Alliance for Research and Technology (SMART) Center for Environmental Sensing and Modeling (CENSAM), with support from the National Research Foundation and the Ministry of Education. Measurement data were recorded between February 24 to March 23, 2012, and August 28 to September 9, 2012 with nine sets of Vaisala WXT520 Weather transmitters and data loggers employed at ground level.

Measurements at Punggol are used here to evaluate and compare the predictive performance of various sensor configurations. Out of the nine weather transmitters, six were used, since one was not in operation and two were in locations not modeled in CFD. One of the highest records of wind speed was observed on February 28, 2012 and the measured data (24 hours) were selected for evaluation in order to capture short-term variations in atmospheric boundary conditions and support the premise of negligible convective effects and isothermal conditions during modeling (Section 4.1). Sample distributions of simulated measurements were generated at all potential sensor locations shown in **Fig. 8** based on field measurements, following the procedure described in [25, 31].

1

Fig. 12. Selected sensor locations during field measurement near Punggol@Treelodge displayed in the simulation environment (center) and actual sensor installations (bottom left and top right, modified from [43]).

5   The sensor configurations constructed with the hierarchical strategy (Section 4.2) are

6   evaluated for their predictive performance at a random unmeasured location near the

7   northwest part of the canopy. The evaluation of each sensor configuration is done using the

8   metrics described in Sections 3.2-3.3. The aim is to support a model-falsification approach to

9   data interpretation (Section 3.1) and demonstrate the usefulness of the optimal configurations

10 in improving simulation predictions.

11 **Fig. 13** shows the relationship between the number of sensors and the probability of (a) the

12 Type I and Type II prediction errors and (b) the expected prediction range of wind speed, for

13 a 95% certainty. The probabilities of committing either error are equal when a configuration

14 of twelve sensors is deployed. Type II errors are increased using fewer sensors, while with

15 more sensors Type I errors are increased. When more than 12 sensors are deployed, a 55%

16 refinement from the initial plausible range is expected and the predicted range is nearly 10%

17 higher than the range of the measurements. Around 20% of this range is owed to modeling

18 uncertainty and the remaining to parameter uncertainties. During sensor placement, a large

19 reduction in prediction range occurs when the 5th, 8th, 9th and 13th sensors are deployed, yet

20 none exceeds 10%. After the 13th sensor, the reduction rate drops steadily with the addition of

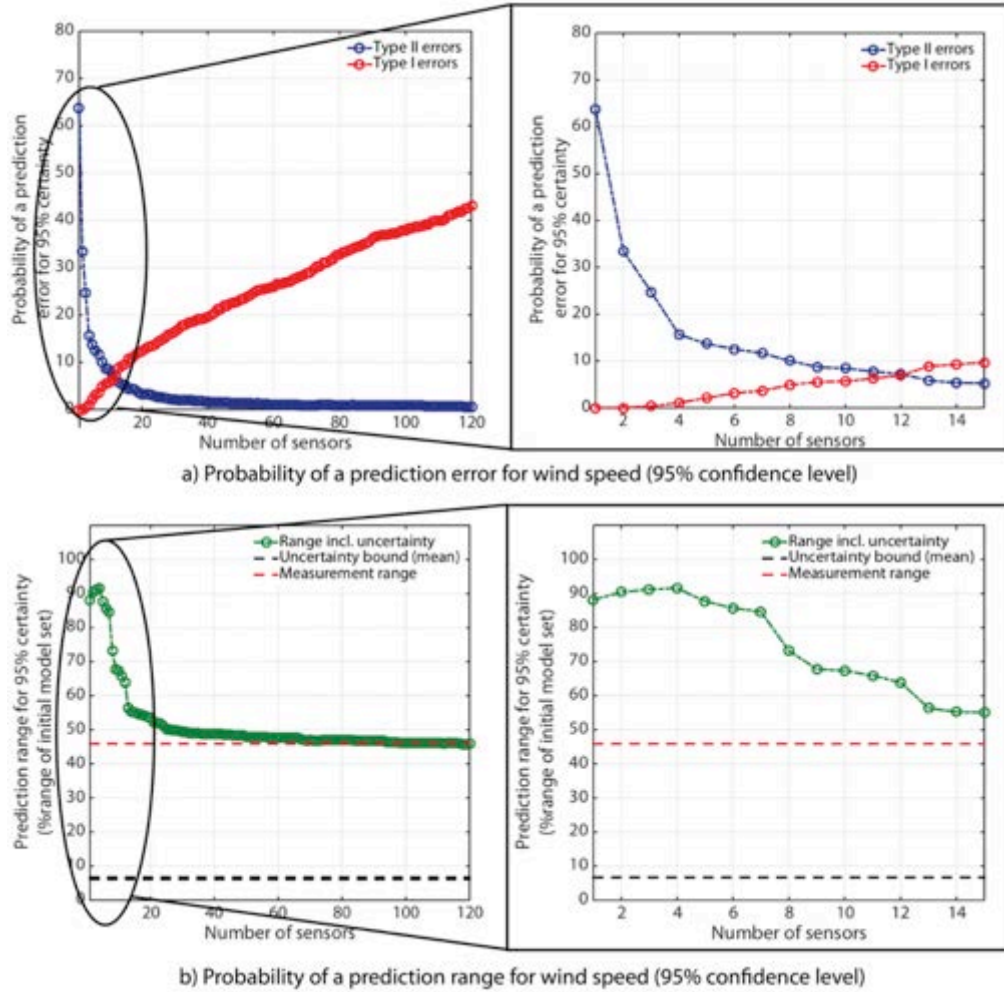21 sensors and converges towards the measurement range.

a) Probability of a prediction error for wind speed (95% confidence level)

b) Probability of a prediction range for wind speed (95% confidence level)

1

2 **Fig. 13.** The relationship between the number of sensors and (a) the probability of a prediction error
3 and (b) the expected range predictions for wind speed, for a 95% confidence level using the
4 hierarchical algorithm; all possible optimal configurations are displayed on the left side and only the
5 first fifteen configurations on the right side.

6 **Fig. 14** (a) shows the relationship between the number of sensors and the probability of Type
7 I and Type II errors in predictions of wind direction, for 95% certainty. In contrast to wind
8 speed predictions, the expected prediction range is not displayed since the range of
9 measurements is almost 360 deg due to the long duration of the measurement campaign (24
10 hours). Here, the probabilities of committing either error are diverging for all sensor
11 configurations, signifying no sensor configuration is useful.

12 Since Type II errors are considerably low and Type I errors large, even from the first sensor
13 selected, it is evident that a large number of model instances are falsified and retained
14 candidate models are not able to improve predictions at the unmeasured location. This
15 location was randomly selected near the northwest part of the canopy, where values of wind
16 speed predictions are low and modeling errors associated with wind direction are large. From

1   Equation (9), corresponding model instances of wind direction are not considered in the
2   calculations.

3   To cross-validate this finding, the procedure is repeated by randomly selecting another
4   (unmeasured) location near the center-east side of the canopy, where values of wind-speed are
5   high. **Fig. 14** (b), shows the relationship between the number of sensors deployed and the
6   probability of Type I and Type II errors in predictions of wind direction, for 95% certainty.
7   Compared to the NW location, the predictive performance is better at the center-east location:
8   the probabilities of committing either error are equal to almost 20% when five sensors are
9   deployed. Type II errors are increased with fewer sensors, while adding sensors increases
10 Type I errors. These results show that unlike wind-speed predictions, for the specific model-
11 set, no sensor configuration is able to improve predictions of wind direction at low wind-
12 speed regions where modeling errors are large.

13

a) @NW unmeasured location   a) @Center-east unmeasured location
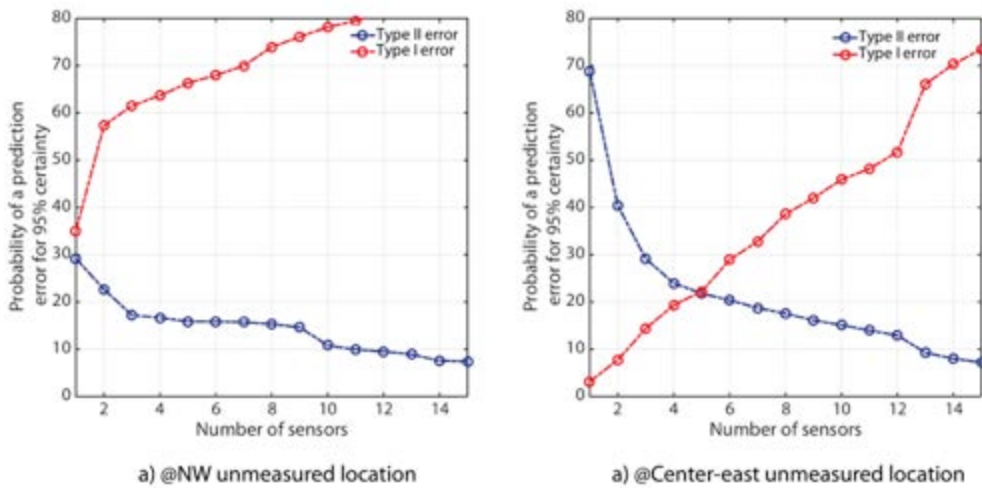
14 **Fig. 14.** The relationship between the number of sensors and the probability of prediction error for
15 wind direction at (a) a NW and (b) a center-east unmeasured location, for a 95% confidence level using
16 the hierarchical algorithm; only the first fifteen optimal configurations are displayed.

17 A comparison of the performance of the configuration of six sensors constructed with the
18 hierarchical strategy is performed against the historical sensor configuration selected by
19 Shanshan Pan (**Fig. 12**). **Table 3** shows the results of this comparison with respect to
20 improving predictions of wind speed, at the randomly selected NW location. The three
21 metrics used for comparison are Type I errors, Type II errors and prediction range. Both
22 configurations achieved an equal refinement in the range of predictions, which is well above
23 the measurement range of 46% (**Fig. 13**). However, the probabilities of prediction errors are
24 larger using the historical configuration as compared to the configuration constructed with the

hierarchical strategy: Type I errors are nearly 8 times larger, while Type II errors are increased around 30%.

Table 3 Comparison of the predictive performance with respect to wind speed at the randomly selected NW location: for the configuration of six sensors constructed with the hierarchical strategy and the initially fixed configuration during the measurement campaign.

Configuration of 6 sensors	Type I Prediction Error (%)	Type II Prediction Error (%)	Prediction range (%of initial plausible range)
Sensor configuration using the hierarchical strategy	3.1	12.5	85.7
Sensor configuration based on historical measurements	23.5	16.4	85.7

4.4. Multi-criteria decision making

A multi-criteria decision-making (MCDM) approach is adopted to evaluate the influence of complementary and conflicting criteria on the choice of sensor configuration. The MCDM approach described in Section 3.4 is applied to the optimal sensor configurations constructed with the hierarchical strategy and the criteria evaluated are listed in **Table 3**.

Fig. 15 illustrates the evaluation of the five criteria using the RR-Pareto3 algorithm for wind-speed sensor configurations: (a) all sensor configurations and (b) the set of 31 compromise configurations of 12 to 42 sensors with respect to these criteria. Although no single optimal solution is found, this set of solutions is refined when the decision maker provides additional information. Such an example is shown in **Fig. 16** (a) where the 31 compromise configurations are evaluated using only three criteria: range, cost and Type I errors. A reduced set of 6 compromise configurations, of 15 to 20 sensors is recognized with respect to these three criteria (**Fig. 16** (b)). When the decision maker provides additional information, a single compromise configuration can be obtained that meets the measurement needs.
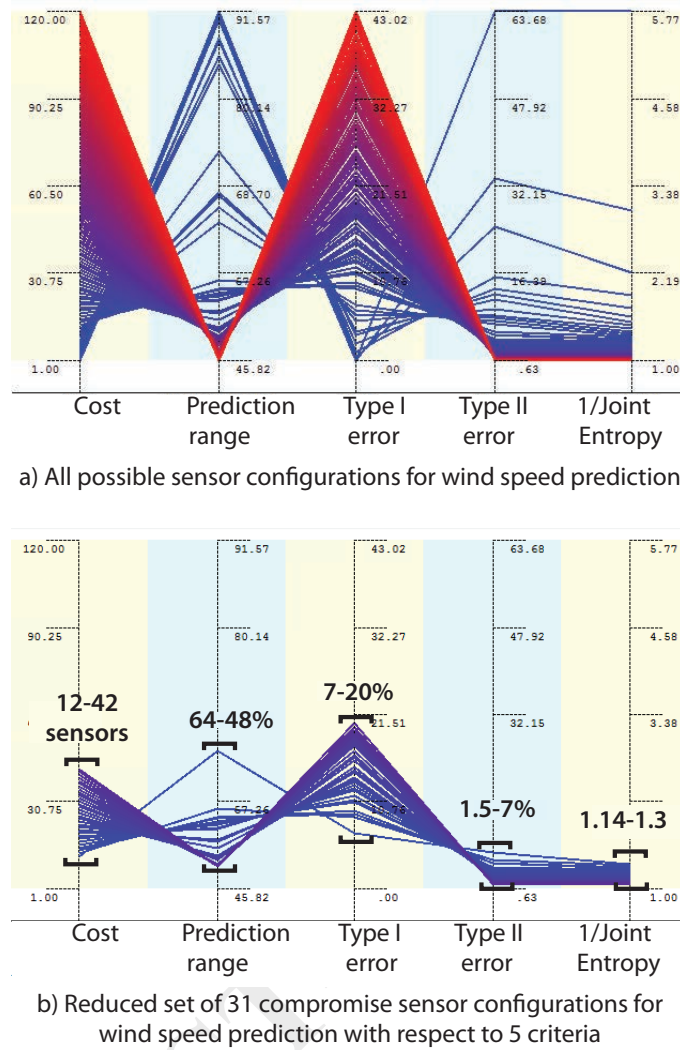
a) All possible sensor configurations for wind speed prediction

b) Reduced set of 31 compromise sensor configurations for
wind speed prediction with respect to 5 criteria

Fig. 15. Evaluation of 5 sensor selection criteria for predicting wind speed using the RR-Pareto3 algorithm: (top) all possible sensor configurations and (bottom) the set of 31 compromise configurations of 12 to 42 sensors with respect to all criteria.

In wind studies, however, it is often the case that more than one variable is measured with the same sensor configuration, such as wind speed and wind direction. As demonstrated in Section 4.2, the optimal sensor configurations can differ between these variables, which necessitates that a common configuration is found. In the presence of a decision maker's preference, this problem can be solved using an MCDM approach in a hierarchical manner with less computational cost (Section 3.4).

Firstly, the two variables, wind speed and wind direction are ordered according to their significance in measurement. Based on the earlier analysis (Section 4.3), it is not possible to predict wind direction with reasonable reliability and therefore it is assessed as a secondary variable of lesser significance, while wind speed is considered the variable of highest importance. The set of 6 compromise configurations for wind-speed predictions (**Fig. 16** (b)),

24

1   are employed to recalculate criteria values for the second variable, wind direction. The RR-

2   Pareto algorithm is rerun to recognize the single compromise configuration that satisfies the

3   measurement criteria for both wind speed and direction. Results are displayed in **Fig. 17**, with

4   the criteria values (a) for wind-direction predictions and (b) for wind-speed predictions using
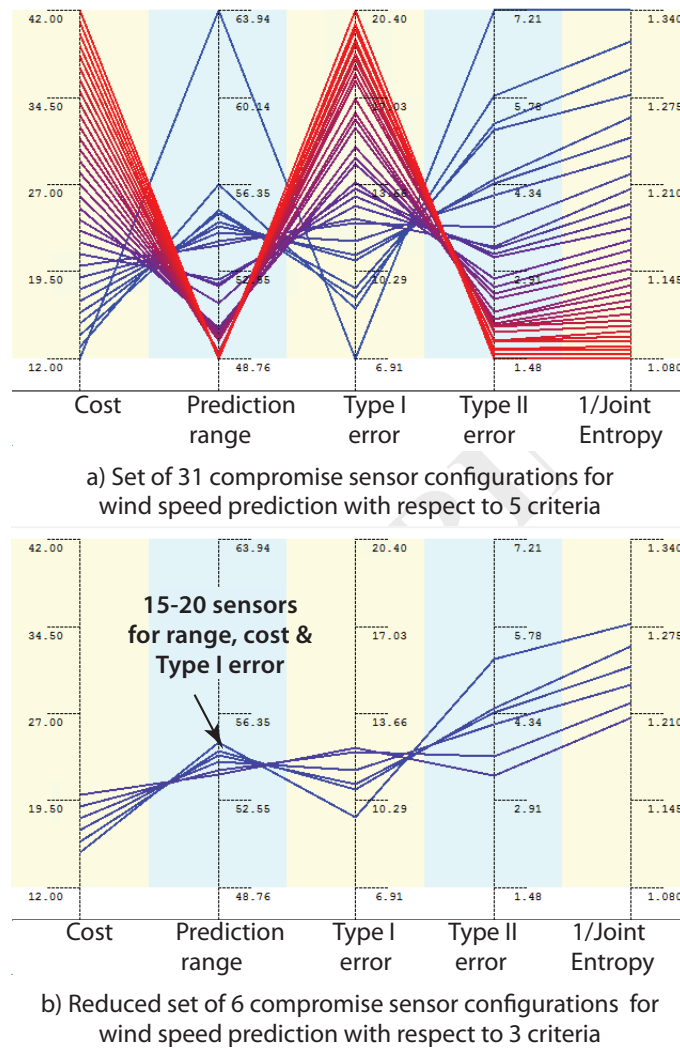
5   the single compromise configuration of 17 sensors.

a) Set of 31 compromise sensor configurations for
wind speed prediction with respect to 5 criteria

b) Reduced set of 6 compromise sensor configurations for
wind speed prediction with respect to 3 criteria

6

7   **Fig. 16.** Evaluation of three sensor selection criteria using the RR-Pareto3 algorithm: (a) the
8   compromise configurations of 12 to 42 sensors for wind speed and (b) the reduced set of 6 compromise
9   sensor configurations with respect to cost, prediction range and Type I prediction error.

10

a) Best compromise sensor configuration for
wind direction prediction with respect to 3 criteria

b) Criteria values for wind speed for the best compromise con-
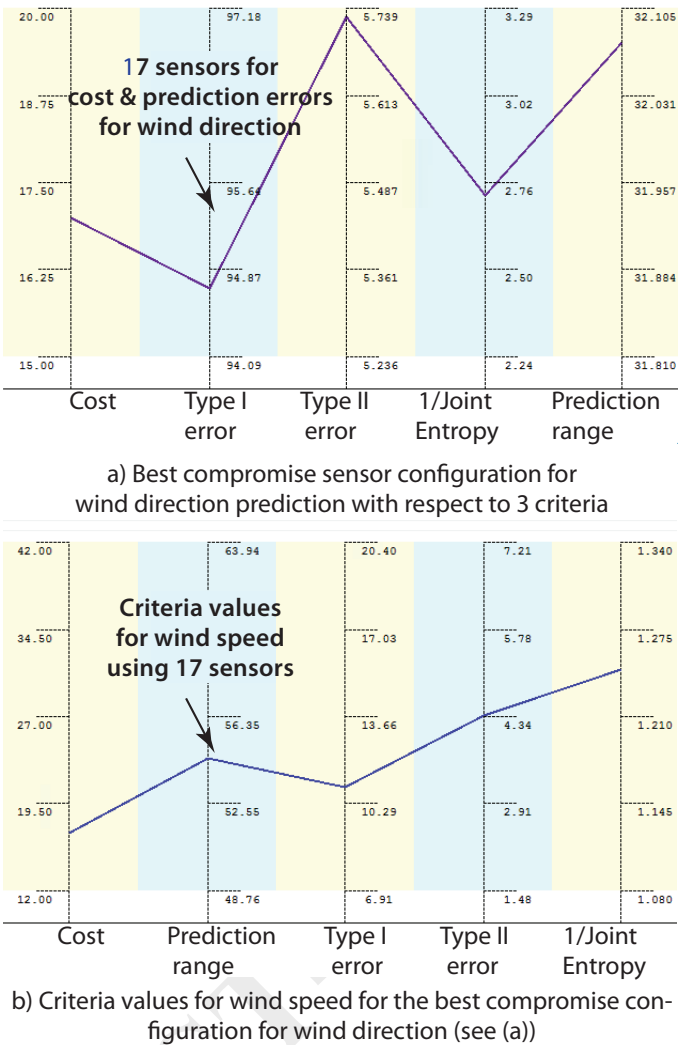figuration for wind direction (see (a))

Fig. 17. The best compromise configuration provided by the RR-Pareto3 algorithm (a) for wind direction predictions with respect to cost, range and Type I prediction error and (b) the criteria values for wind speed for the same configuration.

5. Discussion

Metrics have been developed to provide information on the expected predictive performance of sensor configurations, in terms of reducing the number or candidate models and the ranges of predictions at unmeasured locations. Moreover, an MCDM approach has been proposed for assessing the multiple conflicting criteria and variables that are involved in an optimal sensor configuration problem. Specifically:

1) Expected identifiability metrics are adapted from literature to evaluate the performance of sensor configurations in improving simulations predictions, whereas previously they have been employed to iteratively select optimal sensor locations.

2) Metrics are developed to assess the robustness of sensor configurations to prediction uncertainties. These include metrics to calculate the probability of committing Type I prediction error— reflected by the number of candidate models—and Type II prediction error—represented by the number of measurements that fall outside the predicted range.

3) An MCDM approach is proposed to evaluate the conflicting relationships between the performance metrics as well as to include additional criteria during optimal sensor configuration. A set of compromise sensor configurations is then identified that is subject to the preferences of the decision maker. When measuring several variables with the same configuration, the MCDM problem is solved in a hierarchical manner to identify common optimal configurations, while reducing computational cost.

Although this paper has proposed metrics to evaluate the predictive performance of sensor configurations in wind studies, there are a number of limitations that can be addressed for further development and application. Multiple CFD simulations were performed based on steady-RANS analysis and the realizable k-ε equations, since it is one of the most computationally efficient approaches to approximate turbulent flows. In addition, to manage computational cost no thermal affects were considered and were assumed not to significantly influence the results. Field measurements were therefore taken during the monsoon seasons and following rainy conditions. Approximate estimates of modeling errors associated with these premises are available in literature and were explicitly incorporated in the sensor placement methodology.

Another limitation is that potential measurement locations were subject to orographic constraints and sensor coverage was restricted by the size of mounting arrangement. Potential sensor locations on roofs were particularly avoided to minimize the impact of convection on flow when wind speed is low. In addition, resources were limited in this research. Only a few measurement points were available and used during the experiments to study short-term wind variation around buildings while seasonal variations were not considered.

In this work, the performance of selected sensor configurations has been evaluated for their ability to reduce prediction uncertainty at single point locations. Additional metrics can be developed to assess the robustness of the configurations in updating predictions at multiple locations and with varying area coverage. This is a particularly important procedure when the objective of measurement is to use limited measurement data to update model predictions at multiple locations that have not been measured. In addition, in wind studies, the number of available sensors and their locations are usually fixed, yet optimal sensor configurations for

wind speed and wind direction differ in both sensor locations and in their number. This work proposed an approach to identify a good compromise configuration between optimal configurations for several variables. However, accounting for multiple variables has not been explicitly incorporated in the sensor placement strategies. Further research may contribute to adapt sensor placement strategies, including the hierarchical strategy, in order to account for multiple variables during sensor placement.

Finally, the application of these metrics needs to be further investigated in real-time monitoring and prediction. The integration in monitoring and prediction of outdoor wind conditions can involve several building systems taking into account information gained from neighboring studies. Such studies may contribute additional information to understanding wind behavior and support decision making for the design of new buildings.

6. Conclusions

The main results of this study point to the following conclusions:

1. Sensor locations configured using a hierarchical sensor placement strategy enhance wind predictions compared to a situation with no measurement.
2. Predictive performance metrics evaluate the usefulness of sensor configurations prior to field measurements through calculating the expected accuracy and range of predictions at unmeasured locations.
3. The usefulness of sensor configurations in improving predictions is influenced by modeling errors and depends on the variable and location of interest to predict.
4. An MCDM approach to sensor placement identifies set of compromise sensor configurations that accounts for multiple and conflicting criteria, as well as the decision maker's preferences.
5. Solving an MCDM sensor-placement problem in a hierarchical manner identifies common optimal configurations between several variables of different significance.

7. Acknowledgements

8. References

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

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

[3] Gousseau, P., B. Blocken, T. Stathopoulos, and G.J.F. van Heijst, *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, 2011. **45**(2): p. 428-438.

[4] Blocken, B., W.D. Janssen, and T. van Hooff, *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, 2012. **30**(0): p. 15-34.

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

[6] Blocken, B., T. Stathopoulos, J. Carmeliet, and J.L.M. Hensen, *Application of computational fluid dynamics in building performance simulation for the outdoor environment: an overview.* Journal of Building Performance Simulation, 2011. **4**(2): p. 157-184.

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

[8] Tamura, T., *Towards practical use of LES in wind engineering.* Journal of Wind Engineering and Industrial Aerodynamics, 2008. **96**(10): p. 1451-1471.

[9] Schatzmann, M., H.R. Olesen, and J. Franke, *COST 732 model evaluation case studies: approach and results*, 2010, COST Office.

[10] Grimmond, C. and T.R. Oke, *Aerodynamic properties of urban areas derived from analysis of surface form.* Journal of Applied Meteorology, 1999. **38**(9): p. 1262-1292.

[11]   Britter, R. and S. Hanna, *Flow and dispersion in urban areas.* Annual Review of Fluid Mechanics, 2003. **35**(1): p. 469-496.

[12]   van Hooff, T. and B. Blocken, *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, 2012. **104-106**(May-July 2012): p. 330-341.

[13]   Rotach, M., *Profiles of turbulence statistics in and above an urban street canyon.* Atmospheric Environment, 1995. **29**(13): p. 1473-1486.

[14]   Brown, M.J., D. Boswell, G. Streit, M. Nelson, T. McPherson, T. Hilton, E.R. Pardyjak, S. Pol, P. Ramamurthy, and B. Hansen. *Joint urban 2003 street canyon experiment.* in *Symposium on "Planning, Nowcasting, and Forecasting in the Urban Zone.* 2004.

[15]   Eliasson, I., B. Offerle, C. Grimmond, and S. Lindqvist, *Wind fields and turbulence statistics in an urban street canyon.* Atmospheric Environment, 2006. **40**(1): p. 1-16.

[16]   Balázs, B., T. Hall, M. Roth, and L.K. Norford. *Microclimate in a high-rise residential development in Singapore.* in *The seventh Intl Conf on Urban Climate.* 2009.

[17]   Krause, A., A. Singh, and C. Guestrin, *Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies.* The Journal of Machine Learning Research, 2008. **9**(6 January 2008): p. 235-284.

[18]   Osborne, M.A., S.J. Roberts, A. Rogers, S.D. Ramchurn, and N.R. Jennings. *Towards real-time information processing of sensor network data using computationally efficient multi-output Gaussian processes.* in *Information processing in sensor networks (ISPN 2008).* 2008. ACM/IEEE, NY, USA.

[19]   Das, A. and D. Kempe. *Sensor selection for minimizing worst-case prediction error.* in *Information Processing in Sensor Networks (IPSN 2008).* 2008. ACM/IEEE, NY, USA.

[20]   Wu, X., M. Liu, and Y. Wu, *In-situ soil moisture sensing: Optimal sensor placement and field estimation.* ACM Transactions on Sensor Networks (TOSN), 2012. **8**(4): p. 33:1-30.

[21]   Papadopoulou, M., B. Raphael, C. Sekhar, and I.F.C. Smith. *Sensor placement for predicting airflow around buildings to enhance natural ventilation.* in *ASHRAE IAQ 2013: Environmental Health in Low Energy Buildings.* 2013. ASHRAE, Atlanta, GA, USA.

[22]   Oke, T.R., *Siting and exposure of meteorological instruments at urban sites*, in *Air Pollution Modeling and its Application XVII*2007, Springer. p. 615-631.

[23] Schatzmann, M., S. Rafailidis, and M. Pavageau, *Some remarks on the validation of small-scale dispersion models with field and laboratory data.* Journal of Wind Engineering and Industrial Aerodynamics, 1997. **67**: p. 885-893.

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

[25] Papadopoulou, M., B. Raphael, I.F.C. Smith, and C. Sekhar, *Optimal sensor placement for time-dependent systems: Application to wind studies around buildings.* Journal of Computing in Civil Engineering, 2015: p. 04015024.

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

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

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

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

[30] Papadimitriou, C. and G. Lombaert, *The effect of prediction error correlation on optimal sensor placement in structural dynamics.* Mechanical Systems and Signal Processing, 2012. **28**(0): p. 105-127.

[31] Papadopoulou, M., B. Raphael, I.F.C. Smith, and C. Sekhar, *Hierarchical sensor placement using joint entropy and the effect of modeling error.* Entropy, 2014. **16**(9): p. 5078-5101.

[32] Goulet, J.-A. and I.F.C. Smith, *Structural identification with systematic errors and unknown uncertainty dependencies.* Computers & structures, 2013. **128**(November 2013): p. 251-258.

[33] Ben-Haim, Y. and F.M. Hemez, *Robustness, fidelity and prediction-looseness of models.* Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science, 2012. **468**(2137): p. 227-244.

[34] Raphael, B. and I.F.C. Smith, *Engineering Informatics: Fundamentals of computer-aided engineering.* Second Edition ed2013, Hoboken, NJ, USA: Wiley. 356.

[35] Adbi, H. and D. Valentin, *Encyclopedia of Measurement and Statistics.* Thousand Oaks, CA: Sage. Agresti, A. (1990) Categorical data analysis., New York: Wiley.

Agresti, A.(1992) A survey of exact inference for contingency tables. Statist Sci, 2007. **7**: p. 131-153.

[36] Raphael, B., *Integrated control of indoor environmental quality*, in *R-296-000-102-112: Final Report,* 2010, Department of Building, National University of Singapore.

[37] Authority, B.a.C., *BCA Green Mark Awards 2007*, 2008.

[38] Franke, J., A. Hellsten, H. Schlünzen, and B. Carissimo, *Best practice guideline for the CFD simulation of flows in the urban environment: COST action 732 quality assurance and improvement of microscale Meteorological models,* 2007, Brussels, Belgium: COST Office.

[39] Tominaga, Y., A. Mochida, R. Yoshie, H. Kataoka, T. Nozu, M. Yoshikawa, and T. Shirasawa, *AIJ guidelines for practical applications of CFD to pedestrian wind environment around buildings.* Journal of Wind Engineering and Industrial Aerodynamics, 2008. **96**(10-11): p. 1749-1761.

[40] Ansys, *Design Exploration User Guide*, 2011, ANSYS, Inc.: Canonsburg, PA.

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

[42] Vernay, D.G., B. Raphael, and I.F.C. Smith, *Augmenting simulations of airflow around buildings using field measurements.* Advanced Engineering Informatics, 2014. **28**(October 2014): p. 412–424.

[43] Shanshan, P., *Urban canopy flow and thermal comfort: A case study in Singapore*, in *CENSAM Seminar presentation*2014: Singapore.