Advanced Engineering Informatics 30 (2016) 127–142

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

Evaluating predictive performance of sensor configurations in wind studies around buildings

Maria Papadopoulou [a,*], Benny Raphael [c], Ian F.C. Smith [a,d], 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

ARTICLE INFO

Article history:
Received 26 August 2015
Received in revised form 14 December 2015
Accepted 12 February 2016
Available online 4 March 2016

Keywords:
Computational fluid dynamics (CFD)
System identification
Sensor configuration
Performance evaluation
Simulation prediction

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 configurations. Results show that the performance metrics successfully evaluate the robustness of sensor configurations with respect to reducing uncertainty of wind predictions at other unmeasured locations.

1. Introduction

The continuous growth of the global population living in cities has increased interest in outdoor thermal comfort [1], air quality [2,3], safety [4], and particularly in warm climates, building energy consumption and natural ventilation [5]. The wind environment has a primary role in mitigating these issues and, consequently, improving knowledge of wind behavior around buildings has been the focus of much recent research work (a detailed review can be found in [6]). The most common approach for wind prediction is based on computational models, such as those used in computational fluid dynamics (CFD) simulations.

Today, CFD simulations are used to overcome constraints of laboratory and field measurements [5], since they provide detailed information on wind flow and allow treatment of complex geometries with a high degree of repeatability. Although CFD simulations provide reasonable predictions, the accuracy is not superior to laboratory and field measurements [6]. Uncertainties are large in both modeling and measurements and should be taken into account [7].

Tamura [8] and Schatzmann et al. [9] suggested that measurements, both laboratory and field, should be used in a complementary manner to ensure that simulation results are sound, even when using modeling methods of high predictability, such as the large eddy simulation (LES) [8]. However, the use of simplified arrays of roughness elements in laboratory measurements results in idealized representations of the parameters affecting wind flow [10], and it is often unclear how sensitive wind predictions are to these parameter uncertainties [11]. Moreover, certain physical phenomena, such as buoyancy-driven natural ventilation, cannot be fully represented in reduced-scale laboratory experiments [5].

In these situations, field measurements are essential for ensuring that modeling is sound, especially in studies involving high-rise buildings [10,11]. Nevertheless, field measurements have been rare, and sensor placement still remains a challenging task [12]. Most of earlier studies have used historically measured data and

* Corresponding author at: Future Cities Laboratory, Singapore-ETH Centre, 1 CREATE Way, #06-01, CREATE Tower, Singapore 138602, Singapore. Tel.: +65 8287 0922.

E-mail address: papmara@gmail.com (M. Papadopoulou).

http://dx.doi.org/10.1016/j.aei.2016.02.004
1474-0346/

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

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

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

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

This paper proposes metrics to evaluate the predictive performance of sensor configurations and assess their usefulness in improving simulation predictions. The performance of the configurations is assessed for their capability to falsify multiple model instances whose predictions are inconsistent with the data (Section 2.1). Expected identifiability metrics found in literature [29] are adapted (Section 2.2) and then new metrics are developed (Section 2.3) to estimate the sensor configurations robustness-to-uncertainties associated with model predictions. In the end, a multi-criteria decision-making (MCDM) approach is proposed to evaluate the influence of the conflicting metrics on the choice of optimal sensor configuration (Section 2.4). In Section 3, the proposed metrics are applied to evaluate the performance of several hierarchically-constructed sensor configurations [31], in improving wind predictions around a full-scale building system in Singapore. A list of the main conclusions and a critical assessment on the results are provided in the final two sections (Sections 4 and 5).

2. Development of predictive performance metrics

Metrics are developed to estimate the expected predictive performance of sensor configurations and assess their usefulness in improving predictions. The study builds upon previous work on sensor placement performed by the authors [25,31], where the premise is that sensor data are best used for falsifying multiple model instances[1] whose predictions are inconsistent with the data. Therefore the performance of sensor configurations is assessed in terms of their capability to falsify multiple model instances (Section 2.1). First, expected identifiability metrics are adapted from literature to estimate the expected reduction in the number of candidate models (retained model instances) and in the prediction range associated with each sensor configuration (Section 2.2). Several optimal sensor configurations, which have been constructed using the hierarchical-sensor placement strategy, have been compared in [31]. Metrics are then developed to estimate the robustness of the sensor configurations to uncertainty associated with model predictions (Section 2.3). In the end, an MCDM approach is proposed to evaluate the influence of multiple metrics on the selection of optimal sensor configurations (Section 2.4).

2.1. Falsification based on multiple models

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

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

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

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

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

An illustration of the falsification process is shown in Fig. 1. The remaining model instances—called candidate models—are used to update the predictions at other, unmeasured, locations and

[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.*