Vernay, D.G., Raphael, B. & Smith, I.F.C. "A model-based data-interpretation framework for improving wind predictions around buildings" Journal of Wind Engineering and Industrial Aerodynamics, 145, 2015 pp.219-228 http://dx.doi.org/10.1016/j.jweia.2015.06.016

A MODEL-BASED DATA-INTERPRETATION FRAMEWORK FOR IMPROVING WIND PREDICTIONS AROUND BUILDINGS

Didier G. Vernay [a,c,*], Benny Raphael [b] and Ian F.C. Smith [a,c]

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

[b)] Civil Engineering Department, Indian Institute of Technology, Madras, India.

[c)] Applied Computing and Mechanics Laboratory, School of Architecture, Civil and Environmental Engineering (ENAC), EPFL, Lausanne, Switzerland.

* Corresponding author: Didier G. Vernay, Future Cities Laboratory, Singapore-ETH Centre 1 CREATE Way #06-01 CREATE Tower Singapore 138602. Tel.: +65 9723 1127, E-mail: didier.vernay@epfl.ch

Abstract:

Although Computational Fluid Dynamics (CFD) simulations are often used to assess wind conditions around buildings, the accuracy of such simulations is often unknown. This paper proposes a data-interpretation framework that uses multiple simulations in combination with measurement data to improve the accuracy of wind predictions. Multiple simulations are generated through varying sets of parameter values. Sets of parameter values are falsified and thus not used for predictions if differences between measurement data and simulation predictions, for any measurement location, are larger than an estimate of uncertainty bounds. The bounds are defined by combining measurement and modelling uncertainties at sensor locations. The framework accounts for time-dependent and spatially-distributed modelling uncertainties that are present in CFD simulations of wind. The framework is applied to the case study of the CREATE Tower located at the National University of Singapore. Values for time-dependent inlet conditions, as well as values for the roughness of surrounding buildings, are identified with measurements carried out around the CREATE Tower. Results show that, on average, ranges of

horizontal wind-speed predictions at an unmeasured location have been decreased by 65% when measurement data are used.

Keywords: Computational Fluid Dynamics (CFD); wind modelling; field measurements; data interpretation; multi-model reasoning

1) Introduction

Wind around buildings affects the comfort and health of residents as well as the energy consumption of buildings, particularly in tropical climates. For example, the convective heat flux at the building façade, influencing energy consumption of buildings, depends on the surrounding wind [1]. Wind can also be harnessed for natural ventilation of buildings [2]. Computational Fluid Dynamics (CFD) simulations have been widely used to simulate wind around and through buildings [3, 4]. Although guidelines have been established to improve simulation predictions [5], large discrepancies remain when simulation predictions are compared to field measurements. Moreover, uncertainties in simulation predictions are usually not quantified [6].

The steady Reynolds-averaged Navier-Stokes (RANS) equations are usually employed in CFD simulations to describe the fluid-flow behavior. These equations are time-averaged or ensemble-averaged equations of the fluid flow motion. Large discrepancies have been observed in wakes of buildings when predictions of RANS-based simulations are compared with wind-tunnel experiments [7, 8]. Wind-tunnel experiments are usually employed to evaluate the performance of approximate equations of fluid-flow solved in CFD simulations because values of parameters are known (e.g. inlet conditions and surface roughness). Large Eddy Simulation (LES) is an alternative strategy for modeling fluid-flow behavior in which time-dependent predictions are computed. LES has been found to provide better agreement with wind-tunnel experiments than RANS-based simulations [7].

Thermal processes may affect the wind behavior around buildings, especially in street canyons which can be subject to combinations of low wind speeds and high differential heating between surfaces [9]. However, if thermal processes are implemented into the CFD model, modeling complexity is increased [3, 10] along with the number of parameters that cannot be easily estimated, such as the

thermal properties of surfaces. Therefore, thermal processes are not often included in CFD simulations. Effects of thermal processes have been evaluated by using field measurements [9]; by simulating thermal processes with CFD simulations [11]; or by using wind-tunnel experiments with a heated floor [12]. However, the effects have been estimated for standard building configurations (street canyons) and they are likely to vary for other topologies.

Model-based data interpretation has the potential to improve the accuracy of simulation predictions through the use of a population of CFD simulations and measurement data. In model-based data-interpretation approaches, many model instances (simulation instances) are generated through assigning sets of parameter values to a model class. In this work, the model class is a CFD model with un-assigned parameter values. Measurement data are used to estimate sets of parameter values by solving an inverse problem. The inverse problem involves estimating sets of parameter values by comparing measurement data with predictions of model instances. Several approaches are described in following chapters.

Model calibration, in which an "optimal" model is found by minimizing the sum of the squared difference between simulation predictions and measurement data, is not appropriate because there rarely is a single answer to the inverse problem. Many set(s) of parameter values within a model class might give same responses at sensor locations in complex systems [13]. Such ambiguities are amplified by measurement and modelling uncertainties. Modeling uncertainties refer to uncertainties (probability distributions of errors) in the model class (e.g uncertainties associated with RANS equations). Moreover, model calibration approaches provide values of parameters, which compensate modeling and measurement errors at sensor locations. Therefore, the "optimal" model is conditional on sensor locations (and modeling errors at those locations). Furthermore, calibration approaches do not provide information that can lead to estimates of uncertainties of subsequent predictions [14].

Bayesian inference identifies conditional probability distributions of parameter values given measurement data [15]. Probability distributions are required to represent measurement and modeling uncertainties at sensor locations. Uncertainties in CFD simulations are difficult, if not impossible to

determine precisely. If incorrect probability distributions are defined, it may lead to over-conditioning of parameter values [14]. Furthermore, modeling errors are often systematic and this introduces additional error correlations between measurement locations [16, 17]. Implementation of Bayesian inference requires a complete knowledge of all correlations in order to avoid biased predictions. In wind modeling, the values of such correlations are unknown.

An alternative is to use a model-falsification approach, such as error-domain model falsification [16, 18] and Generalized Likelihood Uncertainty Estimation (GLUE) [14], in which incorrect sets of parameter values are falsified using measurement data. Only bounds of measurement and modeling uncertainties are needed. Error-domain model falsification has been developed in the application of bridge diagnosis and leak detection in water networks. Error-domain model falsification involves falsification of model instances for which differences between measurement data and simulation predictions, for any measurement location, are larger than an estimate of uncertainty bounds; the bounds are defined by combining measurement uncertainties and modeling uncertainties at that location. When the entire set of model instances is falsified, the model class is incorrect. This could mean that either additional processes need to be included, boundary conditions are incorrect, etc. or modeling and measurement uncertainties have been underestimated. In this way, model falsification provides a way to test the validity of model classes.

The main objective of this paper is to present a model-based data-interpretation framework which is appropriate for the identification of parameter values of CFD simulations, and subsequent predictions at unmeasured locations. The framework is based on error-domain model falsification. Improvements have been made to error-domain model falsification in order to reproduce time variability (at the scale of 15 minutes) of wind through allowing identification of different sets of inlet conditions at different times. In this framework, time-dependent inlet conditions as well as the roughness of the surrounding buildings are identified using time series of measurement data.

Modelling and measurement uncertainties affect the information content of measurement data. A systematic methodology to evaluate modeling uncertainties is proposed that recognizes their time-

dependent and spatially-distributed characteristics. The final objective is to apply the methodology to the case study of the "CREATE Tower". The CREATE Tower is a 16-storey building located at the National University of Singapore.

The structure of the paper is as follows. In the next section, the model-based data-interpretation framework is described. Section 3 introduces the case study and the model class including the parameters requiring identification. The experimental setup is presented in Section 4. Section 5 presents a methodology to estimate modeling uncertainties that can be incorporated to the model-based data-interpretation framework. The model-based data-interpretation framework is applied in Section 6 using simulation predictions, measurement data and knowledge of measurement and modeling uncertainties. The paper ends with a discussion of the results and plans for future work.

2) Methodology

This section presents the model-based data-interpretation framework used to identify sets of parameter values of the CFD simulation and predict wind variables at unmeasured locations. This framework is based on error-domain model falsification which has been found to be useful in applications of bridge diagnosis and leak detection in water networks [16, 18]. In such systems, parameter values are identified using measurements carried out only at specific times. In the assessment of wind behavior around buildings, parameter values of CFD simulations need to be identified dynamically using time series of measurements.

2.1) Error-domain model falsification

Error-domain model falsification involves generating sets of model instances $M\left(\boldsymbol{\theta}_j\right)$ through assigning a combination of parameter values $\boldsymbol{\theta}_j = \left[\theta_1,\ldots,\theta_p\right]_j$ to a model class M with $j \in \{1,\ldots,n_m\}$. p is the number of parameters requiring identification and n_m is the number of model instances. When correct sets of parameter values $\boldsymbol{\theta}^*$ are assigned to the model class, the predicted value of an output variable of the model instance $M\left(\boldsymbol{\theta}^*\right)$ differs from the real value (Q)

by the modelling error $\grave{o}_{model}$. Modelling errors are errors associated with the model class. The real value is also equal to the measured value y plus a measurement error $\grave{o}_{measure}$. This is expressed in Equation (1).

$$M\left(\boldsymbol{\theta}^*\right)+\grave{o}_{model}=Q=y+\grave{o}_{measure} \tag{1}$$

Equation (2) is derived by rearranging the terms in Equation (1). The difference between the predicted and the measured value is equal to the difference between the measurement and the modelling error.

$$M\left(\boldsymbol{\theta}^*\right)-y=\grave{o}_{measure}-\grave{o}_{model} \tag{2}$$

However, errors are seldom known in environmental systems. Errors are represented with probability distributions (uncertainties), which are often assumed uniform in absence of more information [19]. Measurement and modelling uncertainties are combined using Monte Carlo technique [20]. Threshold bounds $\left[T_{low},T_{high}\right]$ are defined using the combined uncertainty and a confidence level of $\phi=95\%$ as presented in Figure 1.

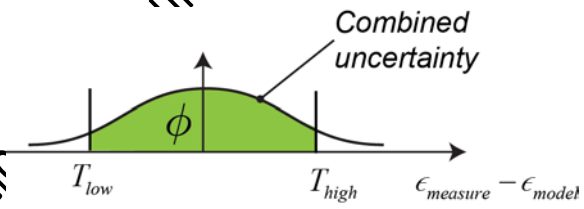

Figure 1: Threshold bounds determined with the combined uncertainty and a confidence level $\phi=95\%$

The model instance $M\left(\boldsymbol{\theta}_j\right)$ is candidate if Equation (3) is satisfied. Otherwise, the model instance is falsified ($\boldsymbol{\theta}_j\neq\boldsymbol{\theta}^*$).

$$T_{low}\le M\left(\boldsymbol{\theta}_j\right)-y\le T_{high} \tag{3}$$

Compared with typical Bayesian approaches (zero-mean Gaussian probability forms), this approach is more robust when error correlations are not known [17].

2.2) Application to wind modeling

In this work, the parameters requiring identification are the inlet wind speed $U_{inlet,16m}$ at 16m height, the inlet wind direction ϑ_{inlet} and the sand-grain roughness of the surrounding buildings ($\boldsymbol{\theta} = \left[U_{inlet,16m}, \vartheta_{inlet}, k_s \right]$).The output variables of interest are the horizontal wind speeds U and wind directions ϑ at sensor location $i \in \{1,\ldots,n_s\}$ and at unmeasured locations of interest, where n_s is the number of sensor locations.

Modeling uncertainties are time dependent and spatially distributed (Section 5). Thus, modelling uncertainties (and threshold bounds $\left[T_{low,i,j,k,t}, T_{high,i,j,k,t} \right]$) are estimated at each sensor location i, for each model instance j, for each compared variable $k \in \{U, \vartheta\}$ and at each time step t. At time step t , model instances are candidate models if, for each and every sensor location and for each and every compared variable, the difference between the measured and the predicted value falls inside the interval $\left[T_{low,i,j,k,t}, T_{high,i,j,k,t} \right]$. This corresponds to the situation when Equation (4) is satisfied.

$$\forall i \in \{1,\ldots,n_s\} \text{ and } \forall k \in \{u_h, \vartheta\} : T_{low,i,j,k,t} \le M\left(\boldsymbol{\theta}_j\right)_{i,k} - y_{i,k,t} \le T_{high,i,j,k,t} \quad (4)$$

Where $M\left(\boldsymbol{\theta}_j\right)_{i,k}$ is the predicted value of the variable k at sensor location i by the model instance j and $y_{i,k,t}$ is the measured value of the variable k at sensor location i at time step t .

Model instances are steady-state models. At each time step, specific set(s) of parameter values are identified in order to represent the dynamic behavior of wind.

2.3) Wind predictions at unmeasured locations

Each candidate model is assumed to be equally probable. A discrete distribution of values can be predicted with the set of candidate models for each variable k, at each unmeasured location of interest l and at each time step t. These distributions correspond to uncertainties associated with parameter values, which have been reduced through falsification of incorrect sets of parameter values.

In order to obtain unbiased predictions, modelling uncertainties need to be combined to these distributions as expressed in Equation (5). For each candidate-model prediction, 2000 samples are drawn from distributions of modelling-uncertainty sources in order to build predictive distributions at unmeasured locations.

$$P_{l,k,t} = M\left(\theta_{j^*}\right)_{l,k} + \grave{o}_{model,j^*,l,k,t} \tag{5}$$

Where $P_{l,k,t}$ is the predictive distribution of the variable k at the unmeasured location l and at the time step t. $M\left(\theta_{j^*}\right)_{l,k}$ is the prediction of the variable k by the candidate model j^* at the unmeasured location l. $\grave{o}_{model,j^*,l,k,t}$ is a random variable representing the modelling uncertainty of the candidate model j^* in the prediction of the variable k at the unmeasured location l and at the time step t.

Ranges of predictions corresponding to any confidence level (for example 50, 70 and 95%) can be computed using the predictive distributions.

3) Numerical simulations

CFD simulations have been executed using the commercial software FLUENT 14.5 [21] in order to simulate possible wind behavior around the CREATE Tower. The CREATE Tower is a 16-storey building located at the National University of Singapore.

The 3D models of the buildings have been built using photogrammetry techniques [22]. The resolution of the 3D model of the CREATE Tower is approximatively 5cm. A lower resolution is employed for the surrounding buildings. The surrounding buildings of the CREATE Tower have been

modelled following best practice guidelines [23]. If the distance from a building of height H to the CREATE Tower is lower than $6H$, the building is explicitly modelled. The size of the computational domain has also been determined using best practice guidelines [23], creating a computational domain with dimensions $L \times W \times H = 2233m \times 1144m \times 368m$.

CutCell Meshing has been employed to generate the grid. CutCell Meshing generates a high percentage of hexahedral cells which provide better iterative convergence than tetrahedral cells [23]. The expansion ratio is set to 1.1. The minimal size of the cells is set to 0.05m resulting in a grid with 13.1×10^6 cells. Results have been compared with a finer grid in which the minimal size of the cells has been reduced to 0.03m. In average over 187 locations close to the CREATE Tower (possible sensor locations mentioned in Section 4), the mean difference between horizontal wind speeds predicted with the two grids corresponds to 1.1% of the inlet wind speed at the same height. The finer grid does not improve significantly the results and, therefore, has not been selected.

The 3D steady RANS equations are used in the model class to describe the flow behavior around the CREATE Tower. The Realizable $k-\varepsilon$ model is employed to provide closure with two additional transport equations [24]. The first new transported variable is the turbulence kinetic energy k and the second is the turbulence dissipation rate ε. Isothermal conditions were imposed. More sophisticated model class can be employed to simulate the effect of thermal processes on the wind behavior. However, it would increase the modelling complexity and would require the definition of new parameter values such as the thermal properties of the surfaces as well as the sun radiation which vary with respect to time. Thus, isothermal conditions have been considered in the model class. Nevertheless, the effect of thermal processes on wind predictions is evaluated using statistical methods on measurement data taken at different times of the day (and night) in Section 5.3 in order to identify reliable sets of parameter values and predict reliable wind variables at unmeasured locations.

The Coupled algorithm is employed for pressure-velocity coupling. A second-order discretization scheme is used to interpolate pressure from values at cell centers. The convergence criteria, based on

the scaled residuals, are set to 10^{-4} for all variables. Before simulations were terminated, the predicted values of the variables of interest were constant. Thus, the solution was regarded as converged [25].

A user-defined function (UDF) in FLUENT is employed to define a vertical profile of wind speed U_{inlet}, turbulence kinetic energy k_{inlet} and turbulence dissipation rate ε_{inlet} at the inlet of the computational domain. For the $k-\varepsilon$ model, profiles have been proposed by Richards and Hoxey [26]. They are expressed in Equation(6), Equation (7) and Equation (8).

$$U_{inlet}(y)=\frac{\mathrm{u}^{*}_{\mathrm{ABL}}}{\kappa}\ln(\frac{y+y_0}{y_0}) \tag{6}$$

$$k_{inlet}(y)=\frac{{\mathrm{u}^{*}_{\mathrm{ABL}}}^{2}}{\sqrt{\mathrm{C}_{\mu}}} \tag{7}$$

$$\varepsilon_{inlet}(y)=\frac{{\mathrm{u}^{*}_{\mathrm{ABL}}}^{3}}{\kappa(y+y_0)} \tag{8}$$

where y is the height coordinate, y_0 is the roughness length of the terrain, κ is the von Karman constant, $\mathrm{u}^{*}_{\mathrm{ABL}}$ is the atmospheric boundary layer (ABL) friction velocity and C_{μ} is a model constant of the $k-\varepsilon$ model.

The flow behavior near walls is modeled with the standard-wall function [27]. The sand-grain roughness k_s needs to be defined at wall surfaces in FLUENT. A relationship between the sand-grain roughness k_s and the roughness length y_0 (commonly used in wind engineering) has been established for the standard-wall function in FLUENT by Blocken et al. [28]. This relationship is expressed in Equation (9).

$$k_s=\frac{9.793y_0}{C_s} \tag{9}$$

Where C_s is the roughness constant. The value of k_s cannot be larger than y_p, which is the distance between the wall and the centroid of the wall-adjacent cell. Therefore, C_s may need to be adjusted in FLUENT to satisfy Equation (9).

The terrain of the computational domain is decomposed into two surfaces. The first surface is the terrain where the surrounding buildings are explicitly modelled. The second surface is the upstream terrain where the buildings are implicitly modelled using an equivalent roughness length y_0. The roughness length used to model the upstream terrain is set to $y_0 = 0.8m$, which represents regularly-built large town [29]. By using Equation (9), the sand-grain roughness and the roughness constant have been set to k_s =1.33 and C_s =6. The roughness length imposed on the upstream terrain is also used for the definition of inlet profiles (Eqs. (6)-(8)) in order to avoid unintended streamwise gradient associated with roughness modification in the upstream part of the computational domain [28]. The terrain where the surrounding buildings are explicitly modelled is set to $y_0 = 0.24m$. Symmetry boundary conditions are imposed on the sides and on the top of the computational domain. Zero-pressure boundary condition is imposed at the outlet.

The parameters requiring identification are the inlet wind speed at 16m height $U_{inlet,16m}$, the inlet wind direction ϑ_{inlet} and the roughness of the surrounding buildings k_s. A population of model instances $M\left(\boldsymbol{\theta}_j\right)$ has been generated through assigning sets of parameter values $\boldsymbol{\theta}_j = \left[U_{inlet,16m}, \vartheta_{inlet}, k_s\right]_j$ to the model class M described in previous paragraphs. These parameters have been selected because they have the highest impact on wind predictions and are difficult to estimate.

Wind speeds and wind directions at the inlet of the CFD simulation are sensitive parameters and are difficult to estimate in urban areas [30]. Buildings and streets in the area of interest are modelled with a certain degree of geometrical simplification in CFD simulations. A roughness is imposed on these surfaces to implicitly model those simplifications. This roughness is difficult to estimate and may have a strong influence on predictions of wind speeds (variations up to 25% for mean wind speeds)

[31]. In the proposed framework, representative inlet conditions and roughness of the surrounding buildings are identified using measurement data from sensors located around the CREATE Tower.

Grid-based sampling is used to select sets of parameter values uniformly within the parameter space for the generation of model instances. Table 1 presents the maximal and minimal values of parameters as well as their discretization intervals. CFD simulations have been executed for each combination of inlet wind direction ϑ_{inlet} and roughness of the surrounding buildings k_s. A single value of inlet wind speed $U_{inlet,16m}$ is used in the CFD simulations. A total of 48 simulations have been executed in batch mode using 12 processes in parallel on a Windows Server 2012 containing four Hexa-Core Intel Xeon E54607 2.20GHz processors and 64 GB memory which requires approximatively 192 hours of execution time. When a new value of inlet wind direction is assigned to the model class, the outside box of the computational domain modifies its orientation, leading to the generation of a new grid.

Wind predictions for other inlet wind speeds are obtained using a linear relationship. Indeed, it was observed that same amplification factor of wind speeds and same wind directions are predicted at one location when the value of inlet wind speed is varied. In total, wind predictions of 768 model instances are obtained although only 48 CFD simulations have been executed.

Horizontal wind speeds and wind directions are predicted at sensor locations for each model instance. Figure 2 presents the horizontal wind speeds and wind directions at 30 meter height predicted by one model instance.

Table 1: Minimal and maximal values of parameters requiring identification as well as discretization intervals

Parameter requiring identification	Minimal value	Maximal value	Discretization intervals
Wind direction at the inlet ϑ_{inlet} form North [°]	0	337.5	22.5
Wind speed at the inlet $U_{inlet,16m}$ (at 16m height) [m/s]	0.5	8	0.5

Sand-grain roughness of surrounding buildings k_s [m]	0.02	0.32	0.15

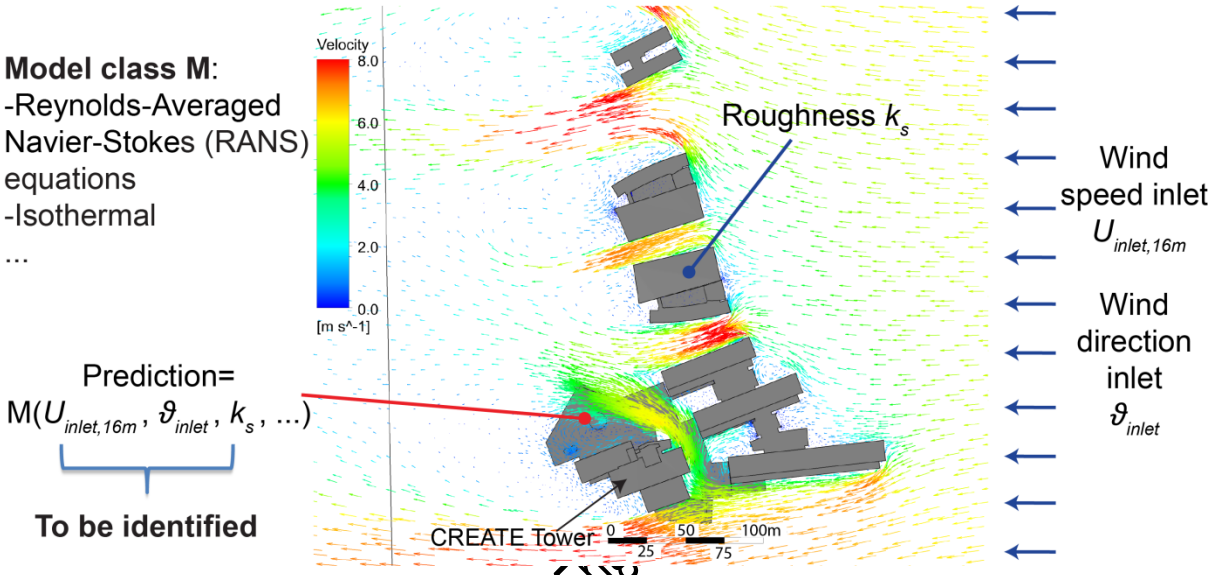

Figure 2: Wind velocities predicted by one model instance ($\vartheta_{inlet} = 90°, U_{inlet,16m} = 6m/s$, $k_s = 0.17m$) at 30 meter height determined through a CFD simulation (top view) and description of the system.

4) Experimental setup and field measurements

Measurements were carried out from February 14. 2014 to March 21. 2014 during the North-East monsoon season.

A sensor placement methodology was applied in order to define measurement locations that should bring the largest amount of information on the wind behavior around the CREATE Tower [32]. Possible locations were limited to the balconies and terraces of the CREATE Tower. From the 187 possible locations defined initially, the 8 best locations have been determined. Figure 3 presents the possible locations of weather stations as well as their actual locations. Seven weather stations are used to falsify incorrect sets of model instances (sensor S1 to S7) and the last weather station is used to test

the framework (sensor S8). The weather stations were deployed around the CREATE Tower at different floor levels (L3, L7 and L13).

Four Davis weather stations, each equipped with a wind-cup anemometer and a wind vane were used to measure horizontal wind speeds and wind directions (sensor S1 to S4) and four HOBO weather stations equipped with a wind-cup anemometer, a wind vane and a temperature sensor were used to measure horizontal wind speeds, wind directions and temperatures (sensor S5 to S8).

A data logger was installed on each weather station to store measurement data. Average values of measurement data were logged every minute. The resolutions of the Davis weather stations are 0.1m/s for horizontal wind-speed measurements and 22.5° for wind-direction measurements. The starting threshold is 0.4m/s. The resolutions of the HOBO weather stations are 0.19m/s for horizontal wind-speed measurements and 1.4° for wind-direction measurements. The starting threshold is 0.5m/s.

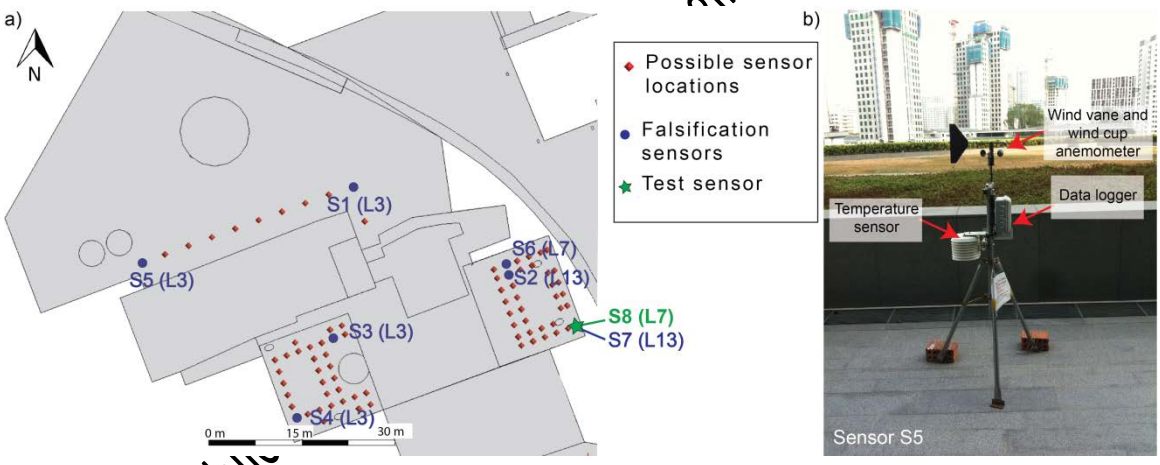

Figure 3: a) Measurement locations around the CREATE Tower (top view) and b) description of the HOBO weather station

5) Modeling uncertainties

Modelling uncertainties at sensor locations affect the information content of measurement data. This section proposes strategies to estimate main sources of uncertainties and to combine them in order to define threshold bounds used for falsification of incorrect model instances.

5.1) Uncertainties associated with RANS-based simulation in the predictions of mean wind variables

LES has been found to be more accurate than RANS-based simulation in regions of flow separation and recirculation when compared to wind-tunnel experiments because flow is highly unsteady in these regions [7]. However, LES takes two order of magnitude longer to execute than a RANS-based simulation [33] and, therefore, it has not been employed for the generation of model instances (Section 3). Uncertainties associated with RANS-based simulation are estimated by comparing predictions of a RANS-based simulation with those of a LES. Horizontal wind speeds and wind directions predicted at 10000 locations around the CREATE Tower with the RANS-based model are compared with mean values of LES predictions.

The geometry and grid settings described in Section 3 are used for both LES and the RANS-based simulation. The same wind conditions are defined at the inlet of both simulations. The turbulence kinetic energy imposed at the inlet of the RANS-based simulation is reproduced in LES by imposing a time-dependent wind-speed profile using the vortex method [34] in FLUENT 14.5. 190 vortices are used in the vortex method because it has provided good results in previous studies on the wind behavior around a wall-mounted cube [35]. In LES, the dynamic Smagorinsky model is employed to model the small eddies of the flow [36].

A zero sand-grain roughness is imposed on all surfaces of the RANS-based simulation in order to be consistent with LES. In order to avoid unintended streamwise gradient in the upstream part of the computational domain, a small roughness length ($y_0 = 0.001m$) is used to define the inlet profiles (Eqs. (6)-(8)), which is different from the roughness length used to model the upstream terrain in the generation of model instances (Section 3). In this section, the goal is to estimate the uncertainties associated with steady RANS equations by comparing RANS predictions with equivalent LES predictions. The goal is not to predict the wind velocity accurately for the selected case study. Since the uncertainties in the RANS predictions are defined as a function of the inlet wind speed, avoiding

unintended streamwise gradients in the upstream part of the computational domain leads to consistent estimations of uncertainties.

Responses of the RANS-based simulation are employed as initial conditions for LES. In LES, the time step size is set to 0.25s. After 1h of real time, mean values of LES do not vary significantly.

Figure 4 presents differences between mean responses of LES and responses of the RANS-based simulation with respect to the amplification factor of wind speeds predicted with the RANS-based simulation. The amplification factor of wind speeds is defined as the ratio between the predicted horizontal wind speed U and the inlet wind speed at the same height U_0. It is thus important to avoid streamwise gradients in the upstream part of the computational domain.

Differences in the predictions of horizontal wind speeds are higher in regions of low amplification factors of wind speeds (Figure 4a). Differences in the predictions of horizontal wind speeds are defined as a percentage of the inlet wind speed at the same height U_0. Indeed, in a previous paper, LES and RANS-based predictions around a single cubical building have been compared for two inlet wind speeds [37]. Differences in the predictions of horizontal wind speeds between LES and RANS-based simulations have been found to be proportional to U_0 [37].

Large variations in the predictions of wind directions are observed in regions of low amplification factors of wind speeds. This is in agreement with results found using a single cubical building [37]. This originates from the over-estimation of the region of reverse flow in wakes of buildings with RANS-based simulations. Therefore, RANS-based simulations may predict reverse flow at locations where LES do not. Model calibration mentioned in Section 1 is especially hazardous when sensors are located in regions of high uncertainties because the calibrated set of parameter values may compensate with possible high errors in the model class.

Two regions are defined in the model-based data-interpretation framework in order to acknowledge for the spatial variability of modeling uncertainties; the first region is defined by amplification factors of wind speeds that are lower than 0.33 and the second region is defined by amplification factors of

wind speeds that are higher than or equal to 0.33. Locations of those regions vary for each model instance. Based on the distributions of differences, this source of uncertainty is described with a normal distribution for each flow variable and for each region.

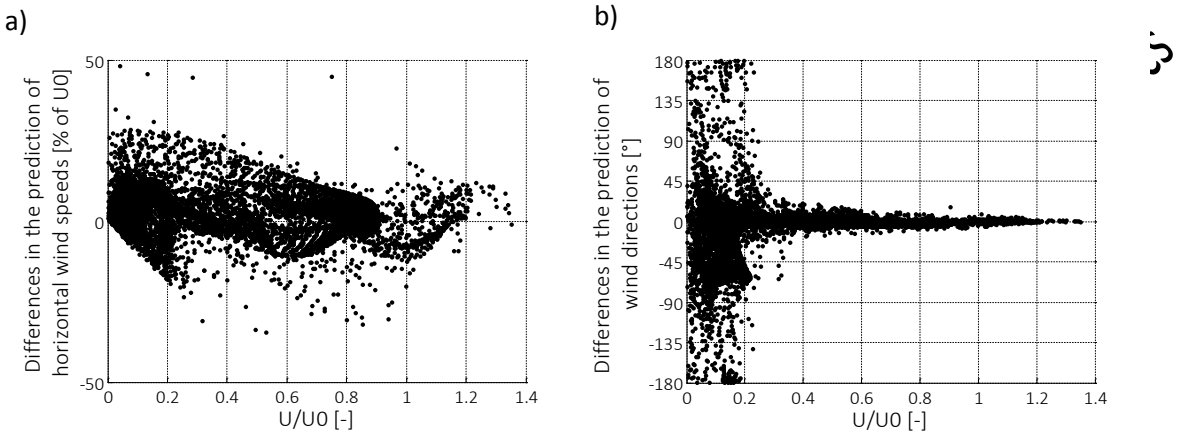

Figure 4: Differences between LES and a RANS-based simulation in the predictions of a) horizontal wind speeds and b) wind directions with respect to the amplification factor of wind speeds

5.2) Uncertainties associated with turbulence

In the model-based data-interpretation framework, an averaging window needs to be chosen to average measurement data. Defining the averaging window is a challenging task because of the presence of two conflicting objectives: 1) the averaging window should be short enough to be able to assume constant wind conditions at the inlet of the RANS-based simulations used in the model-based data-interpretation framework, 2) the averaging window should also be long enough in order to avoid fluctuations of wind variables associated with turbulence at locations of interest. Uncertainties associated with turbulence refer to uncertainties originating from turbulent fluctuations at locations of interest when the averaging window is not long enough.

Unlike RANS-based simulations, LES predicts time series of horizontal wind speeds and wind directions. Uncertainty associated with turbulence is estimated using time series predicted at sensor locations (sensor S1 to S7). Figure 5 presents maximal fluctuations of horizontal wind speeds and wind directions around the mean values of LES, averaged over all sensor locations, with respect to the

averaging window. Fluctuations decrease rapidly when the averaging window increases from 1 minute to 15 minutes. Smaller reductions of these fluctuations are observed if the averaging window is further increased. Therefore, the averaging window is set to 15min. For this averaging window, constant wind conditions are assumed at the inlet of the RANS-based simulations used in the model-based data-interpretation framework.

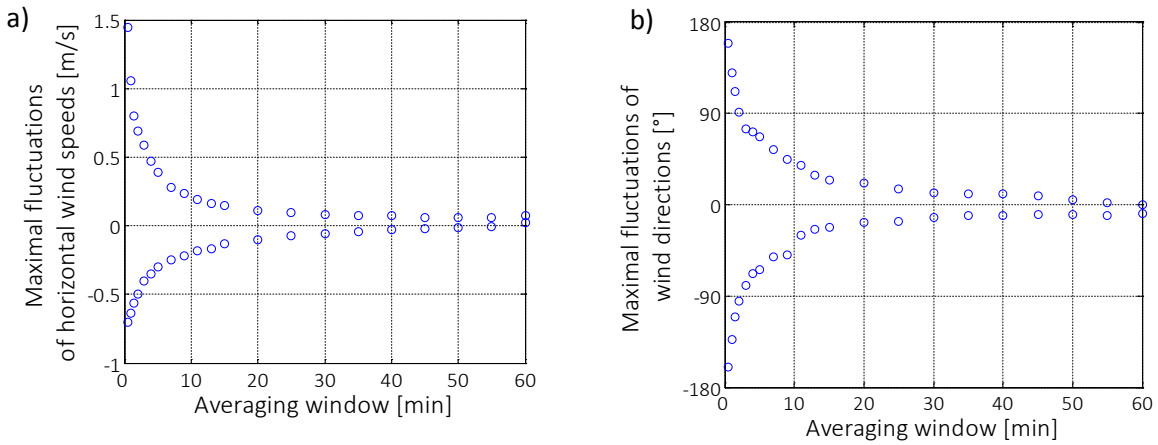

Figure 5: Maximal fluctuations of a) horizontal wind speeds and b) wind directions around the mean value of LES, averaged over all sensor locations, with respect to the averaging window

5.3) Uncertainties associated with thermal processes

Uncertainties associated with thermal processes are estimated with measurement data because modelling thermal processes in CFD simulations is complex (Section 1). It would require the definition of new parameter values such as the thermal properties of surfaces and the execution of a transient simulation for a period of several days [3].

A reference station should be placed on the roof of a building, close to the area of interest in order to measure the ambient wind conditions [61]. In the urban canopy layer, the wind can be significantly affected because of the combination of low wind speeds below roof level and high differential heating between surfaces. In this work, the reference station is located on a small ridge at the rooftop of the Engineering Faculty of the National University of Singapore. Measurement data at this station are provided online (https://inetapps.nus.edu.sg/fas/geog/stationInfo.aspx). Information from the

reference station is not used as inlet conditions in the CFD simulations because it might not be representative of the overall conditions at the inlet [30].

The steps followed to estimate the effect of thermal processes on the horizontal wind speed and the wind direction at sensors located around the CREATE Tower are described below.

1) Calculate the mean value of horizontal wind speeds and wind directions at the reference station.
2) Select measurement data that are taken when wind conditions at the reference station are similar to the mean wind conditions (mean value of horizontal wind speeds and mean value of wind directions).
3) Regression analyses are employed in order to separate out the effect of thermal processes from a data sample taken at different times of the day in which the wind conditions above roof level are similar.

   Determining the sample size is difficult because of two competing objectives. The sample size should be high enough in order to be able to separate out the effect of thermal processes from other sources of variability (e.g. turbulence, etc.). However, the sample size should be small enough in order to have same ambient wind conditions measured above roof level. Thus, different sample sizes are used in the methodology. The regression analyses are expressed in Equation (10) and Equation (11).

$$U_{i,s} = a_{i,s} \times T_{i,s} + b_{i,s} \times U_{ref,s} + c_{i,s} \tag{10}$$

$$\vartheta_{i,s} = d_{i,s} \times T_{i,s} + e_{i,s} \times \vartheta_{ref,s} + f_{i,s} \tag{11}$$

   Where $U_{i,s}$, $\vartheta_{i,s}$ and $T_{i,s}$ are vectors of horizontal wind speeds, wind directions and temperatures respectively measured at the sensor location i whose sizes correspond to the number of sample s. $U_{ref,s}$ and $\vartheta_{ref,s}$ are vectors of horizontal wind speeds and wind directions respectively measured at the reference station whose sizes correspond to the number of sample s. $a_{i,s}$, $b_{i,s}$, $c_{i,s}$, $d_{i,s}$, $e_{i,s}$ and $f_{i,s}$ are the regression coefficients.

4) Find out if the values of the regression coefficients, $a_{i,s}$ and $d_{i,s}$, are significantly different than 0 by performing a *t*-test [38]. If not, values of the regression coefficients are reported to be 0.

Figure 6 presents values of the regression coefficient, $a_{i,s}$, with respect to the number of samples at sensors S5 to S8 (where temperature sensors are installed). $a_{i,s}$ is the regression coefficient that determines the effect of thermal processes on the horizontal wind speed at sensor location i using a number of samples s. Values of the regression coefficient, $a_{i,s}$, are almost constant with respect to the number of samples used in the regression.

Uncertainties in the prediction of horizontal wind speed at sensor location i are represented by uniform distributions with bounds $\left[\Delta U_{low,th,i}, \Delta U_{high,th,i}\right]$ computed using Equation (12).

$$\left[\Delta U_{low,th,i}, \Delta U_{high,th,i}\right] = \left[\min_s(a_{i,s}) \times \left(T_i - T_{\min,i}\right), \max_s(a_{i,s}) \times \left(T_i - T_{\min,i}\right)\right] \quad (12)$$

Where T_i is the temperature measured at sensor location i. $T_{\min,i}$ is the minimal temperature measured at sensor location i. It is assumed that thermal processes have no-significant effects when the temperature measured is minimal.

Wind behavior is affected differently by thermal processes from one sensor to another as presented in Figure 6. Spatially distributed uncertainties are thus considered in the model-based data-interpretation framework in order to identify reliable set(s) of parameter values. Furthermore, bounds depend on the temperature measured at sensor locations. Therefore, time-dependent uncertainties are also considered in the model-based data-interpretation framework.

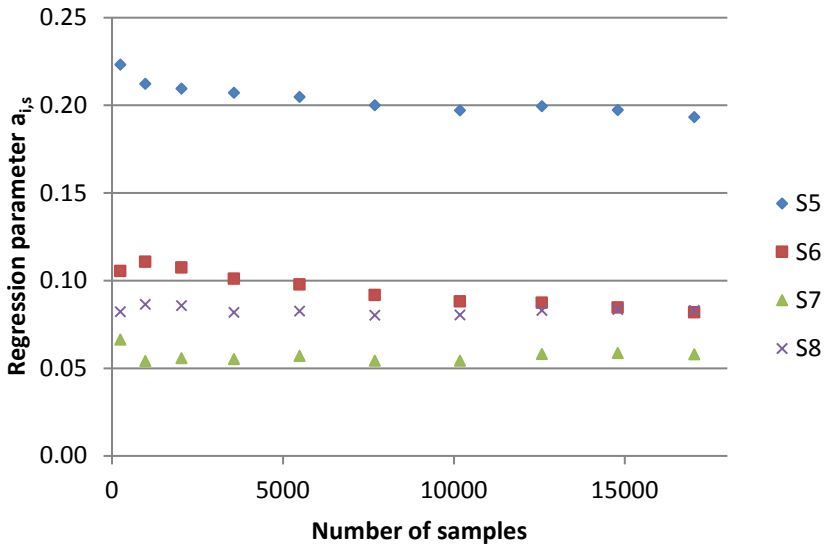

Figure 6: Regression coefficient, $a_{i,s}$, with respect to the number of samples at sensors S5 to S8 (averaging window = 15 minutes)

The maximal and minimal values of the regression coefficients $a_{i,s}$ at sensor locations are used to estimate the maximal and minimal effect of thermal processes on the horizontal wind speed at unmeasured locations. Therefore, uncertainties at unmeasured locations are represented by uniform distributions with bounds $\left[\Delta U_{low,th}, \Delta U_{high,th}\right]$ defined by Equation (13).

$$\left[\Delta U_{low,th}, \Delta U_{high,th}\right] = \left[\min_i\left(\min_s\left(a_{i,s}\right)\right)\times\left(\overline{T}_i - \overline{T}_{\min,i}\right), \max_i\left(\max_s\left(a_{i,s}\right)\right)\times\left(\overline{T}_i - \overline{T}_{\min,i}\right)\right] \quad (13)$$

Where $\overline{T}_i$ is the average value of temperatures measured at sensor locations. $\overline{T}_{\min,i}$ is the minimal value of $\overline{T}_i$.

The same procedure has been followed in order to estimate the effect of thermal processes on the wind direction. Table 2 summarizes the minimal and maximal values of the regression coefficients at sensor locations and at unmeasured locations. The maximal and minimal values of the regression coefficients at unmeasured locations are relatively low. A stronger effect of thermal processes would be expected if measurements would be carried out in street canyons [9]. In such situations, thermal processes might need to be incorporated into the model class.

Table 2: Minimal and maximal values of the regression coefficients at sensor locations and at unmeasured locations used to estimate uncertainty associated with thermal processes.

Wind variable		$U\,[m/s]$		$\vartheta[°]$	
		$\min_s(a_{sensor,i,s})$	$\max_s(a_{sensor,i,s})$	$\min_s(d_{sensor,i,s})$	$\max_s(d_{sensor,i,s})$
Sensor locations	S5	0.19	0.22	-1.65	-1.27
	S6	0.08	0.11	3.47	4.60
	S7	0.05	0.07	2.09	2.38
	S8	0.08	0.09	1.21	1.34
Unmeasured locations		$\min_i\left(\min_s\left(a_{i,s}\right)\right)$	$\max_i\left(\max_s\left(a_{i,s}\right)\right)$	$\min_i\left(\min_s\left(d_{i,s}\right)\right)$	$\max_i\left(\max_s\left(d_{i,s}\right)\right)$
		0.05	0.22	-1.65	4.60

5.4) Uncertainty combination

Measurement uncertainty, uncertainty associated with RANS-based simulation (Section 5.1), uncertainty associated with turbulence (Section 5.2) and uncertainty associated with thermal processes (Section 5.3) have been combined using Monte Carlo technique. Measurement uncertainty corresponds to the resolutions of the sensors mentioned in Section 4. The combination of uncertainties is illustrated in Figure 7.

Threshold bounds used to falsify incorrect sets of model instances are defined using the combined uncertainty and a confidence level of $\varphi = 95\%$. The Šidák [39] correction is employed to adjust the confidence level ($\varphi' = \varphi^{1/N}$) in order to ensure that the target reliability is respected when multiple measurements are employed simultaneously to falsify model instances [40]. N is the number of measurements.

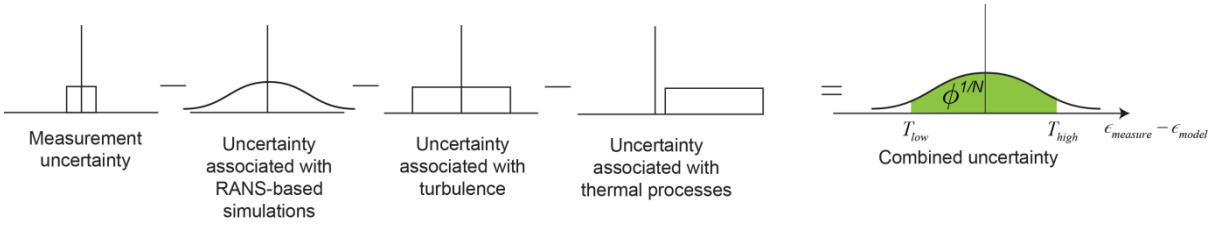

Figure 7: Uncertainty combination for the definition of threshold bounds

6) Results

6.1) Identification of parameter values of the CFD simulation

Falsification of model instances is performed every 15 minutes. Model instances are candidate models if the differences between the measured and the predicted values of horizontal wind speeds and wind directions fall within threshold bounds at each and every sensor location (Section 2.2). Figure 8 presents the falsification of model instances using the horizontal wind speed measured at sensor S1 at 12pm on March 11. 2014. The purple dashed line is the measured value; blue lines are threshold bounds; red crosses are falsified models and green points are candidate models. Falsified models that appear inside threshold bounds in Figure 8 have been falsified using measurement data at other sensor locations or using the measured wind direction at this sensor location.

Threshold bounds are not the same for all model instances because model instances are not defined with the same inlet wind speed and they don't predict the same amplification factor of wind speeds at sensor S1. Furthermore, systematic bias in modeling uncertainties has led to threshold bounds that are not centered on the measured value.

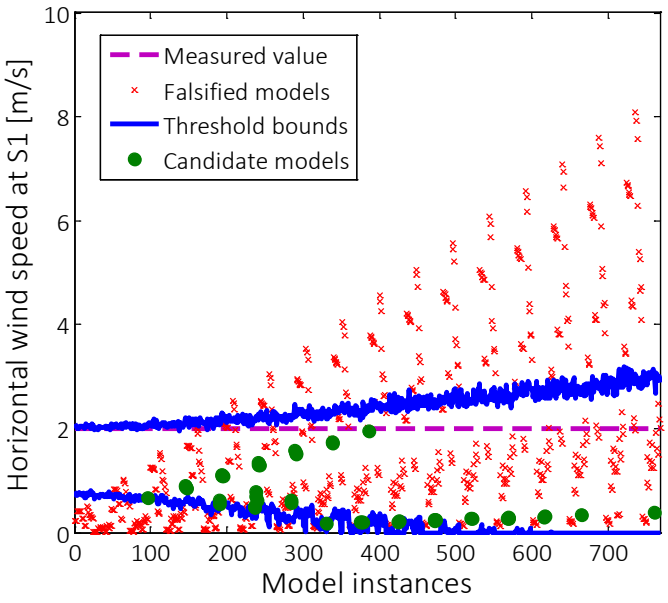

Figure 8: Falsification using the horizontal wind speed measured at sensor S1 at 12pm on March 11. 2014

6.2) Predictions at unmeasured locations

A distribution of predictions can be obtained at an unmeasured location using the final candidate model set (Section 2.3). The likelihood weight for each candidate model is assumed to be equal. This distribution, which corresponds to the uncertainty associated with parameter values propagated through the model, is combined with all sources of modeling uncertainty using Monte Carlo technique. The combination of uncertainties is illustrated in Figure 9. Prediction bounds $[P_{low}, P_{high}]$ are defined using the predictive distribution and a confidence level of $\varphi = 95\%$.

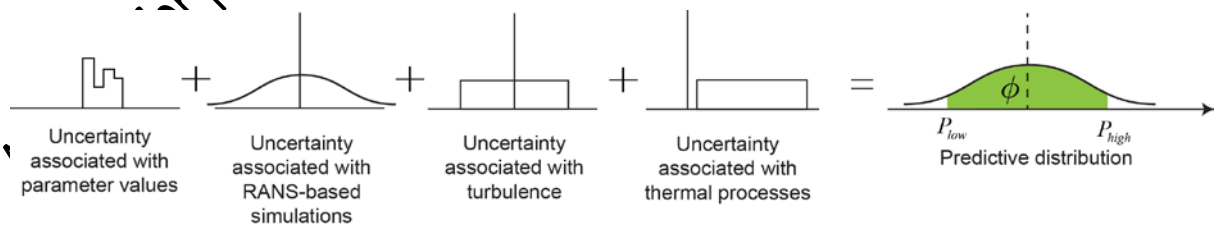

Figure 9: Uncertainty combination for predictions at unmeasured locations

Figure 10 presents predicted and measured values of horizontal wind speeds at the test sensor S8. Red points are horizontal wind speeds measured at the test sensor (sensor S8). Blue lines are prediction

ranges of horizontal wind speeds before measurements. These ranges are obtained using the methodology presented in Figure 9; in which uncertainty associated with parameter values corresponds to the distribution of predictions obtained with the initial set of model instances. Prediction ranges vary with respect to time because of time-dependent uncertainties (uncertainties associated with thermal processes change according to the measured temperatures).

Grey lines are prediction ranges of horizontal wind speeds after measurements. Measurement data have been employed to reduce the uncertainty associated with parameter values. On average, ranges of horizontal wind-speed predictions have been decreased by 65% when measurement data are used. Moreover, horizontal wind speeds measured at the test sensor fall within ranges of predictions 99% of the time after measurements.

The purple dashed line is the mean values of predictions. The root-mean-square difference between the measured values of horizontal wind speeds and the mean values of predictions at the test sensor is 0.39m/s.

Time variability (at the scale of 15 minutes) of measured horizontal wind speeds is observed at the test sensor S8. Time variability is also observed in predictions. This demonstrates that the proposed framework is able to identify time-dependent inlet conditions. At 2:15pm, all model instances have been falsified. This might originate from the underestimation of uncertainties associated with thermal processes, which effects are important at this time of the day.

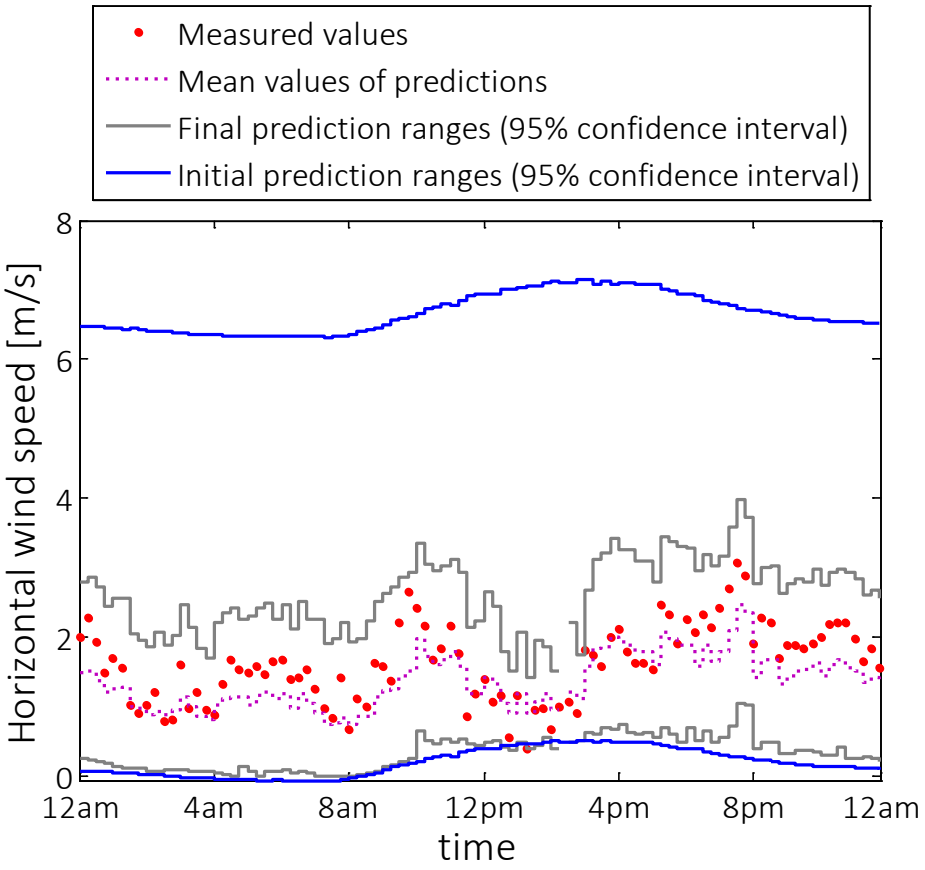

Figure 10: Predicted and measured values of horizontal wind speeds at the test sensor S8 every 15 minutes on March 11. 2014

7) Discussion

This paper proposed a model-based data-interpretation framework in order to identify time-dependent sets of parameter values and predict time-dependent ranges of wind variables at unmeasured locations. Time-dependent and spatially-distributed modeling uncertainties, which affect the information content of measurement data, have been considered.

Limitations are as follows: In the present study, uncertainties associated with RANS-based simulation have been estimated with LES. Although LES has been found to be in better agreement than RANS-based simulation when compared to wind-tunnel experiments, LES is not perfectly accurate.

Uncertainties associated with turbulence have also been estimated with LES. One simulation using LES has been executed under isothermal conditions. Otherwise definition of many new parameters such as thermal properties of surfaces would need to be defined. Furthermore, a transient simulation for a period of several days is needed to simulate such processes, which would take much longer computation time [3]. Fluctuations predicted with LES would have been larger if thermal processes were considered because of the additional turbulence generated. Moreover, a small roughness length is employed to define the wind profiles at the inlet of LES in order to avoid unintended streamwise gradients in the upstream part of the computational domain. This leads to small turbulence at the inlet. Thus, the turbulent fluctuations predicted by LES at sensor locations are mainly due to the surrounding buildings that are explicitly modelled.

Furthermore, the wind profiles used at the inlet of the CFD simulations assumed an equilibrium boundary layer and neutral conditions. These assumptions may not be valid in all urban environments.

Additional sources of uncertainties associated with these simplifications should be added in order to identify reliable sets of parameter values and predict reliable ranges of predictions at unmeasured locations. Even though uncertainties associated with these simplifications were not considered in this study, reliable ranges of predictions were still obtained at unmeasured locations (Section 6.2).

8) Conclusions

In this paper, a framework is proposed to integrate information obtained from measurements with simulation results. The information provided by measurements is used to estimate the parameter values of the simulation, including those for inlet wind conditions, through multiple solutions of the inverse problem. The information content of measurement data depends on levels of measurement and modelling uncertainties at sensor locations. Specific conclusions are:

- Differences between predictions of the RANS-based simulation and LES have been found to be large in regions of low amplification factors of wind speeds. This has led to the definition of modeling uncertainties that vary with respect to space.

- Thermal processes significantly influence horizontal wind speeds at sensors located around the CREATE Tower. This has led to a systematic bias in the modeling uncertainty of horizontal wind speed which depends on temperature measurements.
- Reliable prediction ranges of horizontal wind speeds at unmeasured locations are obtained dynamically with the proposed framework.
- Tighter prediction ranges of horizontal wind speeds are possible using the framework without compromising reliability.

Acknowledgements

This research has been conducted at the Singapore-ETH Centre for Global Environmental Sustainability (SEC), co-funded by the Singapore National Research Foundation (NRF) and ETH Zurich. The authors would like to gratefully acknowledge the support of Prof. C. Sekhar and Prof. M. Santamouris; Prof. N.H. Wong and Asst. Prof. C. Winston for providing additional measurement equipment; Prof. em. Armin Gruen and his FCL team for sharing their 3D models.

References

1. Defraeye, T., B. Blocken, and J. Carmeliet, *Convective heat transfer coefficients for exterior building surfaces: Existing correlations and CFD modelling.* Energy Conversion and Management, 2011. **52**(1): p. 512-522.
2. Ghiaus, C. and F. Allard, *Natural ventilation in the urban environment: assessment and design* 2005: Earthscan.
3. Van Hooff, T. and B. Blocken, *Coupled urban wind flow and indoor natural ventilation modelling on a high-resolution grid: A case study for the Amsterdam ArenA stadium.* Environmental Modelling & Software, 2010. **25**(1): p. 51-65.
4. Ramponi, R. and B. Blocken, *CFD simulation of cross-ventilation flow for different isolated building configurations: validation with wind tunnel measurements and analysis of physical and numerical diffusion effects.* Journal of Wind Engineering and Industrial Aerodynamics, 2012. **104**: p. 408-418.
5. Franke, J., et al. *Recommendations on the use of CFD in wind engineering*. in *Cost Action C*. 2004.
6. Blocken, B. and C. Gualtieri, *Ten iterative steps for model development and evaluation applied to Computational Fluid Dynamics for Environmental Fluid Mechanics.* Environmental Modelling & Software, 2012. **33**: p. 1-22.
7. Tominaga, Y., et al., *Comparison of various revised k–ε models and LES applied to flow around a high-rise building model with 1: 1: 2 shape placed within the surface boundary layer.* Journal of Wind Engineering and Industrial Aerodynamics, 2008. **96**(4): p. 389-411.

8. Yoshie, R., et al., *Cooperative project for CFD prediction of pedestrian wind environment in the Architectural Institute of Japan.* Journal of Wind Engineering and Industrial Aerodynamics, 2007. **95**(9): p. 1551-1578.
9. Niachou, K., I. Livada, and M. Santamouris, *Experimental study of temperature and airflow distribution inside an urban street canyon during hot summer weather conditions. Part II: airflow analysis.* Building and Environment, 2008. **43**(8): p. 1393-1403.
10. Assimakopoulos, V., C. Georgakis, and M. Santamouris, *Experimental validation of a computational fluid dynamics code to predict the wind speed in street canyons for passive cooling purposes.* Solar Energy, 2006. **80**(4): p. 423-434.
11. Xie, X., et al., *The impact of solar radiation and street layout on pollutant dispersion in street canyon.* Building and Environment, 2005. **40**(2): p. 201-212.
12. Allegrini, J., V. Dorer, and J. Carmeliet, *Wind tunnel measurements of buoyant flows in street canyons.* Building and Environment, 2013. **59**: p. 315-326.
13. Beven, K., *A manifesto for the equifinality thesis.* Journal of hydrology, 2006. **320**(1): p. 18-36.
14. Beven, K., *Environmental modelling: An uncertain future?* 2008: Taylor & Francis.
15. Box, G.E. and G.C. Tiao, *Bayesian inference in statistical analysis*. Vol. 40. 2011: John Wiley & Sons.
16. Goulet, J.-A., S. Coutu, and I.F.C. Smith, *Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks.* Advanced Engineering Informatics, 2013. **27**(2): p. 261-269.
17. Goulet, J.-A. and I.F. Smith, *Structural identification with systematic errors and unknown uncertainty dependencies.* Computers & Structures, 2013. **128**: p. 251-258.
18. Goulet, J.-A., C. Michel, and I.F.C. Smith, *Hybrid probabilities and error-domain structural identification using ambient vibration monitoring.* Mechanical Systems and Signal Processing, 2012. **37**(1): p. 199-212.
19. Goulet, J.-A. and I.F. Smith. *Extended uniform distribution accounting for uncertainty of uncertainty*. in *International Conference on Vulnerability and Risk Analysis and Management/Fifth International Symposium on Uncertainty Modeling and Analysis, Maryland, USA*. 2011.
20. Cox, M.G. and B.R. Siebert, *The use of a Monte Carlo method for evaluating uncertainty and expanded uncertainty.* Metrologia, 2006. **43**(4): p. S178.
21. Fluent, A., *14.5.* Theory Guide, 2012. **117**.
22. Gruen, A., et al., *Joint Processing of Uav Imagery and Terrestrial Mobile Mapping System Data for Very High Resolution City Modeling.* ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2013. **1**(2): p. 175-182.
23. Franke, J., et al., *The COST 732 Best Practice Guideline for CFD simulation of flows in the urban environment: a summary.* International Journal of Environment and Pollution, 2011. **44**(1): p. 419-427.
24. Shih, T.-H., et al., *A new k–ε eddy viscosity model for high reynolds number turbulent flows.* Computers & Fluids, 1995. **24**(3): p. 227-238.
25. Franke, J., et al., *Best practice guideline for the CFD simulation of flows in the urban environment 2007.* Hamburg: COST Action. **732**.
26. Richards, P. and R. Hoxey, *Appropriate boundary conditions for computational wind engineering models using the k-ϵ turbulence model.* Journal of Wind Engineering and Industrial Aerodynamics, 1993. **46**: p. 145-153.
27. Launder, B.E. and D. Spalding, *The numerical computation of turbulent flows.* Computer methods in applied mechanics and engineering, 1974. **3**(2): p. 269-289.
28. Blocken, B., T. Stathopoulos, and J. Carmeliet, *CFD simulation of the atmospheric boundary layer: wall function problems.* Atmospheric environment, 2007. **41**(2): p. 238-252.

29. Wieringa, J., *Updating the Davenport roughness classification.* Journal of Wind Engineering and Industrial Aerodynamics, 1992. **41**(1): p. 357-368.
30. 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.
31. Blocken, B. and J. Persoon, *Pedestrian wind comfort around a large football stadium in an urban environment: CFD simulation, validation and application of the new Dutch wind nuisance standard.* Journal of Wind Engineering and Industrial Aerodynamics, 2009. **97**(5): p. 255-270.
32. Papadopoulou, M., et al., *Optimal sensor placement for time-dependent systems: application to wind studies around buildings.* Journal of Computing in Civil Engineering, (under review).
33. Martilli, A., J.L. Santiago, and F. Martín, *Micrometeorological modelling in urban areas.* Física de la Tierra, 2008. **19**: p. 133-145.
34. Mathey, F., et al., *Assessment of the vortex method for large eddy simulation inlet conditions.* Progress in Computational Fluid Dynamics, An International Journal, 2006. **6**(1): p. 58-67.
35. Gousseau, P., et al., *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.
36. Germano, M., et al., *A dynamic subgrid-scale eddy viscosity model.* Physics of Fluids A: Fluid Dynamics, 1991. **3**: p. 1760.
37. Vernay, D.G., B. Raphael, and I.F. Smith, *Augmenting simulations of airflow around buildings using field measurements.* Advanced Engineering Informatics, 2014.
38. Zar, J.H., *Biostatistical analysis. 2nd.* Prentice Hall USA, 1984.
39. Šidák, Z., *Rectangular confidence regions for the means of multivariate normal distributions.* Journal of the American Statistical Association, 1967. **62**(318): p. 626-633.
40. Goulet, J.A. and I.F. Smith, *Performance-driven measurement system design for structural identification.* Journal of Computing in Civil Engineering, 2012. **27**(4): p. 427-436.