

1 Vernay, D.G., Raphael, B. and Smith, I.F.C. "Improving simulation predictions of wind around buildings using  
2 measurements through system identification techniques" Building and Environment, Volume 94, Part 2, 2015,  
3 pp 620-631. <http://dx.doi.org/10.1016/j.buildenv.2015.10.018>  
4  
5

6 IMPROVING SIMULATION PREDICTIONS OF WIND AROUND BUILDINGS USING MEASUREMENTS  
7 THROUGH SYSTEM IDENTIFICATION TECHNIQUES

8 Didier G. Vernay <sup>a,c,\*</sup>, Benny Raphael <sup>b</sup> and Ian F.C. Smith <sup>a,c</sup>

9 <sup>a)</sup> Future Cities Laboratory, ETH Zurich, Zurich, Switzerland.

10 <sup>b)</sup> Civil Engineering Department, Indian Institute of Technology, Madras, India.

11 <sup>c)</sup> Applied Computing and Mechanics Laboratory, School of Architecture, Civil and Environmental Engineering (ENAC),  
12 EPFL, Lausanne, Switzerland.

13 \* Corresponding author: Didier G. Vernay, Future Cities Laboratory, Singapore-ETH Centre 1 CREATE Way #06-01  
14 CREATE Tower Singapore 138602. Tel.: +41 79 784 62 03, E-mail: didiervrn@gmail.com

15 **Abstract**

16 Wind behavior in urban areas is receiving increasing interest from city planners and architects.  
17 Computational fluid dynamics (CFD) simulations are often employed to assess wind behavior around  
18 buildings. However, the accuracy of CFD simulations is often unknown. Measurements can be used to  
19 help understand wind behavior around buildings more accurately.

20 In this paper, a model-based data interpretation framework is presented to integrate information  
21 obtained from measurements with simulation results. Multiple model instances are generated from a  
22 model class through assigning values to parameters that are not known precisely, including those for  
23 inlet wind conditions. The information provided by measurements is used to falsify model instances  
24 whose predictions do not match measurements and to estimate the parameter values of the simulation.  
25 The information content of measurement data depends on levels of measurement and modelling  
26 uncertainties at sensor locations. Modelling uncertainties are those associated with the model class  
27 such as effects associated with turbulent fluctuations or thermal processes.

1 The model-based data interpretation framework is applied to the study of the wind behavior around  
2 the buildings of the Treelodge@Punggol estate, located in Singapore. The framework incorporates  
3 modelling and measurement uncertainties and provides probability-based predictions at unmeasured  
4 locations. This paper illustrates the possibility to improve approximations of modelling uncertainties  
5 through avoiding falsification of the entire set of model instances. It is concluded that the framework  
6 has the potential to infer time-dependent sets of parameter values and to predict time-dependent  
7 responses at unmeasured locations.

8 **Keywords:** data interpretation; uncertainty quantification; wind monitoring; computational fluid  
9 dynamics (CFD); wind modelling

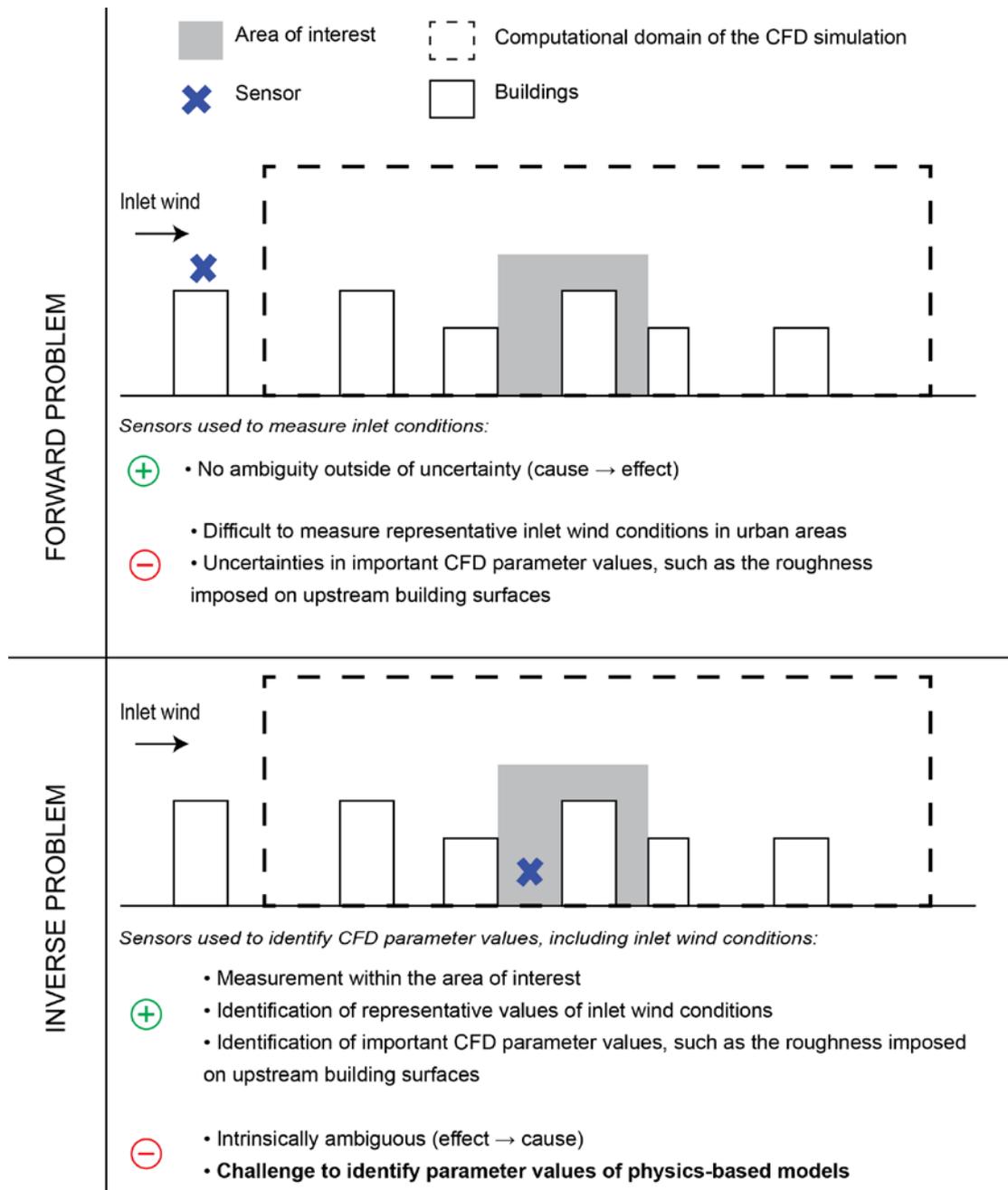
## 10 **1. Introduction**

11 Understanding wind behavior in cities has received increasing interest by city planners and architects.  
12 The assessment of wind behavior around buildings is sought for a wide range of applications, such as  
13 pedestrian wind comfort [1-3], pollutant dispersion [4-6], convective heat transfer at exterior building  
14 surfaces [7-9], natural ventilation [10-13], wind loading on buildings [14, 15], etc. Computational  
15 fluid dynamics (CFD) simulations are often employed to assess wind behavior around buildings.  
16 However, CFD simulations might not provide accurate predictions because of uncertainties in the  
17 values of parameters and the physical phenomena that are not modeled. Monitoring data can be used  
18 to enhance the knowledge of the wind behavior obtained with CFD simulations.

19 Two ways of using measurements in combination with CFD simulations are presented in Figure 1.  
20 The first way is to use sensors to directly measure the inlet wind conditions used as input for the  
21 simulations. Then, simulations are executed to deduce the wind conditions in the area of interest  
22 (forward problem). One measured inlet wind speed and inlet wind direction (cause) correspond to one  
23 model response (effect). Therefore, there is no ambiguity other than the uncertainty associated with  
24 the model. However, values of inlet wind conditions are difficult to measure in urban areas [16].  
25 Furthermore, other important CFD parameter values need to be estimated by engineering judgment,

1 such as the roughness imposed on upstream building surfaces, which can significantly impact the  
2 wind conditions in the area of interest [2].

3 The second way is to use measurements to infer inlet wind conditions and other important CFD  
4 parameter values by solving an inverse problem. The inverse problem consists of estimating the set(s)  
5 of parameter values by comparing simulation predictions of multiple CFD simulations (generated  
6 through assigning different sets of parameter values to the model) with measurement data. This  
7 technique is generally known as system identification. Measurements are carried out within the area  
8 of interest. This allows inference of representative inlet wind conditions and other important CFD  
9 parameter values. This also allows estimations of uncertainties associated with thermal processes  
10 using data measured at different times of day. This aspect is explained in Section 5.2.



1

2 Figure 1: Two ways to use sensors in combination with CFD simulations. This paper focuses on the  
 3 inverse problem.

4 The inverse problem is the focus of this paper. This way of using measurements to estimate parameter  
 5 values is intrinsically ambiguous because there might not be a single answer to the inverse problem  
 6 [17]. Many sets of parameter values might give the same responses at measurement locations in  
 7 complex systems [18, 19]. Such ambiguities are amplified by modelling and measurement  
 8 uncertainties, which reduce the information content of measurement data. Modelling uncertainties are

1 uncertainties associated with the model that cannot be accounted for when sets of parameter values  
2 are varied. Thus residual minimization approaches, which provide a single set of parameter values, are  
3 not appropriate for the inference of parameter values of the CFD simulation.

4 It is a challenge to infer the set(s) of parameter values of physics-based models (such as CFD models)  
5 using measurement data because of measurement and modelling uncertainties. Several approaches can  
6 be used to infer the set(s) of parameter values of physics-based models from measurement data. Their  
7 potential depends on the knowledge of uncertainties (measurement and modelling uncertainties) at  
8 measurement locations and correlations between uncertainties at different measurement locations.

9 Bayesian inference is a statistical method that updates the prior probability of a hypothesis (e.g. a set  
10 of parameter values) using evidence (e.g. measurement data). Bayesian inference has been developed  
11 in the fields of statistics, signal processing and control engineering. Bayesian inference has also been  
12 used in environmental applications such as groundwater modelling [20, 21], rainfall-runoff modelling  
13 [22, 23], climate change predictions [24, 25], etc. In Bayesian inference, prior information about  
14 parameter values as well as a likelihood function that corresponds to the information content of the  
15 measurement data can be defined [17]. The error structure, if known, can be employed to define the  
16 likelihood function [17]. The error structure refers to the probability density functions of errors  
17 (uncertainties) at measurement locations as well as the error correlations from one location to another.

18 In environmental systems, the error structure is seldom known because of modelling uncertainties [17,  
19 18, 26]. Assumptions related to the error structure might not be valid and, therefore, might provide  
20 biased estimates of sets of parameter values and subsequent predictions [26].

21 Approaches have been developed in the fields of civil and environmental engineering in which  
22 modelling uncertainties are seldom known. An example is the error-domain model falsification  
23 methodology developed by Goulet et al. in applications of bridge diagnosis [27, 28], leak detection in  
24 pipe networks [29], wherein only uncertainty bounds need to be defined. In a previous paper of the  
25 authors, this methodology has been adapted to time-variant situations, such as wind around buildings  
26 [30]. In error-domain model falsification, the likelihood function is assumed to be uniform. Each

1 candidate model is assumed to be equally probable. This is justified because there is often not enough  
2 information about error probability distributions between bounds at measurement locations or error  
3 correlations between measurement locations to justify any other likelihood function. Compared with  
4 typical Bayesian inference approaches, error-domain model falsification is more robust when error  
5 correlations between measurement locations are not known [31].

6 The main objective of this paper is to present and evaluate a model-based data-interpretation  
7 framework for improving simulation predictions of wind at unmeasured locations using measurement  
8 data. The framework is applied for the assessment of the wind behavior around the buildings of the  
9 Treelodge@Punggol estate, located in Singapore. This paper illustrates the possibility to improve  
10 approximations of modelling uncertainties through avoiding falsification of the entire set of model  
11 instances.

12 The structure of the paper is as follows. The model-based data-interpretation framework is presented  
13 in the next section. In Section 3, the CFD simulations of the selected case study are described. Field  
14 measurements are presented in Section 4. Strategies to estimate modelling uncertainties are presented  
15 in Section 5. Model falsification is performed in Section 6. Predictions are presented in Section 7 as  
16 well as the evaluation of the framework using test data. The paper ends with a discussion of the results  
17 and plans for future work.

## 18 **2. Methodology**

### 19 **2.1. Model falsification**

20 Error-domain model falsification involves generating model instances  $M(\theta_j)$  through assigning sets  
21 of parameter values  $\theta_j = [\theta_1, \dots, \theta_p]_j$  to a model class  $M$  with  $j \in \{1, \dots, n_m\}$ .  $p$  is the number of  
22 parameters that needs to be inferred and  $n_m$  is the number of model instances. If many parameter  
23 values need to be defined in the model class, a sensitivity analysis can be performed to find out which  
24 parameters have the highest impacts on predictions and use them for the generation of model  
25 instances. This allows reducing the complexity of sampling in high-dimensional parameter spaces.

1 When the correct set of parameter values,  $\theta^*$ , is assigned to the model class, the predicted value of  
 2 this model instance,  $M(\theta^*)$ , differs from the real value,  $r$ , by the modelling error,  $\dot{\theta}_{model}$ . The  
 3 modelling error is the error associated with the model class. The real value is also equal to the  
 4 measured value,  $y$ , plus a measurement error,  $\dot{\theta}_{measure}$ . This is expressed in Equation (1).

$$5 \quad M(\theta^*) + \dot{\theta}_{model} = r = y + \dot{\theta}_{measure} \quad (1)$$

6 The modelling error is the sum of all sources,  $s$ , of modelling errors, as expressed in Equation (2).

$$7 \quad \dot{\theta}_{model} = \sum_s \dot{\theta}_{model,s} \quad (2)$$

8 Equation (3) is derived by rearranging the terms in Equation (1). The difference between the predicted  
 9 and the measured values is equal to the difference between the measurement and the modelling errors.

$$10 \quad M(\theta^*) - y = \dot{\theta}_{measure} - \dot{\theta}_{model} \quad (3)$$

11 However, errors are seldom known in environmental systems. In error-domain model falsification,  
 12 errors are represented with probability distributions (uncertainties), which are often assumed to be  
 13 uniform in the absence of more information. Measurement and modelling uncertainties are combined  
 14 using the Monte Carlo method. 2000 samples are drawn from uncertainty sources in order to build the  
 15 combined uncertainty. Threshold bounds  $[T_{low}, T_{high}]$  are defined using the combined uncertainty and  
 16 a confidence level of 95%. The values of the threshold bounds don't change significantly if the  
 17 number of samples used to combine uncertainties is further increased. Lower and upper threshold  
 18 bounds are found by minimizing the distance between bounds while satisfying a confidence level of  
 19  $\phi = 95\%$ . The Šidák correction is employed when multiple measurements are used to test model  
 20 instances in order to maintain a confidence level of 95% [32]. The model instance  $M(\theta_j)$  is a  
 21 candidate model if Equation (4) is satisfied. Candidate models are used to make predictions at  
 22 unmeasured locations (Section 2.2). Otherwise, the model instance is falsified ( $\theta_j \neq \theta^*$ ).

$$1 \quad T_{low} \leq M(\theta_j) - y \leq T_{high} \quad (4)$$

2 The error-domain model-falsification methodology has been adapted for time-variant situations such  
 3 as wind around buildings [30]. In this adaptation, measurement data are represented as time series.  
 4 The user usually specifies an averaging window and a time step size in order to compute moving-  
 5 average time series of the measurement data. Moving-average time series are computed by replacing  
 6 the measured value at each time step with the average value of its neighboring steps.

7 Each model instance (set of parameter values) is tested at each time step by comparing its predictions  
 8 with measurements. Tests performed at time step  $t$  are assumed to be independent from those  
 9 performed at previous steps. Model instances are Reynolds-averaged Navier–Stokes (RANS)-based  
 10 simulations, which are steady-state simulations. At each time step, specific set(s) of parameter values  
 11 (e.g. inlet wind conditions) are inferred in order to represent the dynamic behavior of wind.

12 Modelling uncertainties can be spatially distributed (e.g. uncertainties associated with RANS-based  
 13 simulations) and time dependent (e.g. uncertainties associated with thermal processes). The  
 14 quantification of modelling uncertainties, including their spatial and temporal variations, is presented  
 15 in Section 5. Therefore, modelling uncertainty depends on the location of the prediction and the time  
 16 step. Each model instance provides an approximation of the wind behavior around buildings with a  
 17 corresponding spatial distribution of modelling uncertainties. Thus, modelling uncertainty at one  
 18 location also depends on the model instance.

19 Modelling uncertainties are estimated at each measurement location,  $i \in \{1, \dots, n_c\}$ , for each compared  
 20 variable,  $k \in \{1, \dots, n_v\}$ , each model instance,  $j$ , and at each time step,  $t$ .  $n_c$  is the number of  
 21 measurement locations and  $n_v$  is the number of compared variables. Thus, threshold bounds

22  $[T_{low,i,j,k,t}, T_{high,i,j,k,t}]$ , which are estimated by combining measurement and modelling uncertainties,  
 23 are estimated at each measurement location,  $i$ , for each compared variable,  $k$ , for each model  
 24 instance,  $j$ , and at each time step,  $t$ .

1 At time step  $t$ , the model instance  $j$  is a candidate model if, for each and every measurement location  
 2 and for each and every compared variable, the difference between the measured and the predicted  
 3 values 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 (5) is satisfied.

$$5 \quad \forall i \in \{1, \dots, n_c\}, \forall k \in \{1, \dots, n_v\} : T_{low,i,j,k,t} \leq M(\theta_j)_{i,k} - y_{i,k,t} \leq T_{high,i,j,k,t} \quad (5)$$

6 where  $M(\theta_j)_{i,k}$  is the predicted value of the compared variable  $k$  at measurement location  $i$  by  
 7 model instance  $j$  and  $y_{i,k,t}$  is the measured value of the compared variable  $k$  at measurement  
 8 location  $i$  and time step  $t$ .

9 When the entire set of model instances is falsified, the model class is not consistent with the error  
 10 structure. This could lead to changes such as adding physical processes to the model class, modifying  
 11 boundary conditions and revising uncertainties.

## 12 **2.2. Predictions at unmeasured locations**

13 Error-domain model falsification has been developed for falsification of wrong sets of parameter  
 14 values and, therefore, for reducing the uncertainty associated with parameter values. The next step is  
 15 to make predictions at unmeasured locations. The likelihood function between bounds is assumed to  
 16 be uniform. There is not enough information regarding error probability distributions between bounds  
 17 at measurement locations or error correlations between measurement locations to justify any other  
 18 likelihood function.

19 A discrete distribution of predictions is obtained using the candidate-model set for each flow variable,  
 20 at each unmeasured location of interest  $i^* = \{1, \dots, n_c^*\}$  and at each time step.  $n_c^*$  is the number of  
 21 unmeasured locations of interest. Candidate models are model instances defined with plausible sets of  
 22 parameter values. Thus, this distribution corresponds to the uncertainty associated with parameter  
 23 values propagated through the model.

1 Modelling uncertainties need to be combined with this distribution in order to predict unbiased wind  
 2 conditions at unmeasured locations, as expressed in Equation (6). For each candidate-model  
 3 prediction at an unmeasured location of interest, samples are drawn from modelling uncertainties at  
 4 this location using the Monte Carlo method in order to build the posterior predictive distribution.  
 5 Errors at measurement locations are assumed to be independent from errors at unmeasured locations  
 6 because error correlations are usually unknown. This results in conservative ranges of predictions at  
 7 unmeasured locations.

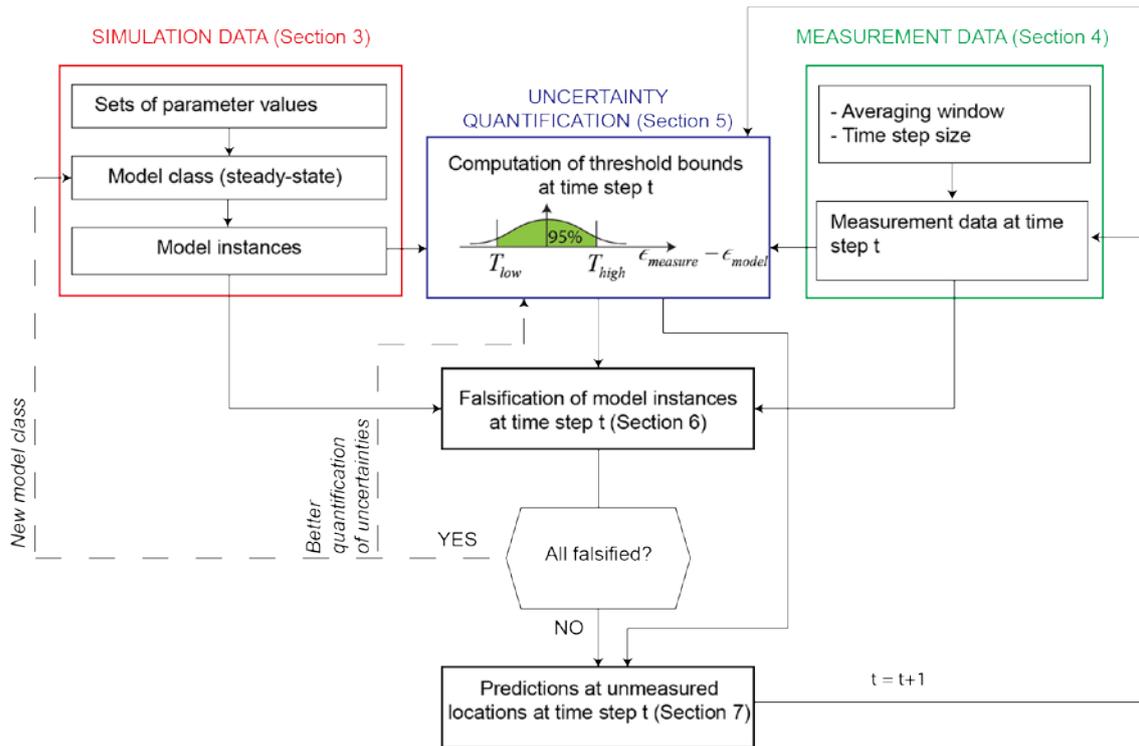
$$8 \quad R_{i^*,k,t} = M\left(\theta_j^*\right)_{i^*,k} + \dot{\theta}_{model,i^*,j^*,k,t} \quad (6)$$

9 where  $R_{i^*,k,t}$  is the posterior predictive distribution of the flow variable  $k$  at the unmeasured location  
 10  $i^*$  and at time step  $t$ .  $M\left(\theta_j^*\right)_{i^*,k}$  is the prediction of the flow variable  $k$  by candidate model  $j^*$  at  
 11 unmeasured location  $i^*$ .  $\dot{\theta}_{model,i^*,j^*,k,t}$  is a random variable representing the modelling uncertainty  
 12 associated with candidate model  $j^*$  in the prediction of the flow variable  $k$  at unmeasured location  $i^*$   
 13 and at time step  $t$ .

14 Figure 2 summarizes the steps leading to probability-based predictions at unmeasured locations and  
 15 indicates the sections of the case study in which those steps are illustrated. The principal goal is to  
 16 integrate information from measurements with information from simulations given prior knowledge  
 17 of measurement and modelling uncertainties.

18 It is emphasized that in this work, model parameters are not "tuned" like in traditional model  
 19 calibration and updating procedures. Through the process of model falsification, the ranges of values  
 20 of uncertain parameters are reduced. The prediction ranges at the locations of falsification sensors are  
 21 reduced according to the estimated uncertainties at these locations. There is a high probability that  
 22 prediction ranges at other locations are also reduced, since some of the model instances are  
 23 eliminated. How much the prediction ranges will reduce at locations far away from the selected  
 24 sensors is difficult to estimate a-priori. If the values at the prediction point have similar sensitivity to

1 those at the falsification sensor locations when input parameters are varied, the prediction ranges can  
 2 be expected to reduce by about the same ratio. However, where the responses are highly non-linear, it  
 3 is possible that the reduced set of candidate models has the same variations and the prediction ranges  
 4 do not get reduced. The strength of the model falsification approach is that an indication of the  
 5 accuracy of predictions is obtained at each point, which is represented by the prediction range.



6

7 Figure 2: Flowchart describing the model-based data interpretation framework

8 **3. Simulation data**

9 CFD simulations have been performed in order to simulate possible wind behavior around the  
 10 buildings of the Treelodge@Punggol estate using FLUENT 14.5. Figure 3 presents the  
 11 Treelodge@Punggol estate and a satellite image of this estate and its surrounding environment. The  
 12 estate is composed of seven buildings of 16 stories each.



1 Figure 3: The Treelodge@Punggol estate (left) and a satellite image of this estate and its surrounding  
2 environment (right).

3 The buildings footprints have been provided by the Urban Redevelopment Authority (URA) of  
4 Singapore. The 3D models of the buildings have been constructed by extruding building footprints to  
5 building heights.

6 The size of the computational domain has been determined using best practice guidelines [33], based  
7 on the height of the tallest building, creating a computation domain with dimensions  
8  $length \times width \times height = 1598m \times 955m \times 504m$ . CutCell Meshing has been used in order to  
9 generate a grid with a high percentage of hexahedral cells because hexahedral cells provide a better  
10 iterative convergence than tetrahedral cells [33]. A grid-sensitivity analysis has been performed in  
11 order to define the grid settings. The expansion ratio is set to 1.1. The minimal size of the cells is set  
12 to 0.05m resulting in a grid with  $13.4 \times 10^6$  cells. Results have been compared with a finer grid in  
13 which the minimal size of the cells has been reduced to 0.03m, resulting in a grid with  $15.9 \times 10^6$ . On  
14 average, over all sensor locations, the difference between horizontal wind speeds predicted with the  
15 two grids corresponds to 3.2% of the horizontal wind speed predicted with the finer grid. The finer  
16 grid does not significantly improve simulation predictions, and, thus has not been selected.

17 The steady RANS equations are used to describe the flow behavior. The Realizable  $k - \varepsilon$  model is  
18 used as turbulence model to provide closure of the RANS equations [34]. Isothermal conditions are  
19 employed. Thus, the effect of thermal processes on the wind behavior is not considered in the CFD  
20 simulations. This simplification has been made because modelling thermal processes in CFD

1 simulations is complex and requires the definition of a large number of new parameters such as the  
 2 thermal properties of the surfaces. Despite this simplification, the effect of thermal processes on the  
 3 wind behavior is estimated at measurement locations and at unmeasured locations in Section 5.2 in  
 4 order to infer reliable sets of parameter values and predict reliable wind conditions at unmeasured  
 5 locations.

6 The Coupled algorithm is used to include the pressure-velocity coupling. The Coupled algorithm is  
 7 preferable to the SIMPLE algorithm [35] as it reduces the number of iterations needed to reach  
 8 convergence for a large computational domain [36]. A second-order discretization scheme is  
 9 employed to interpolate pressure from values obtained at cell centers (finite volume method). The  
 10 convergence criterion is based on the scaled residuals which are set to  $10^{-4}$  for all variables. Before  
 11 simulations were terminated, the predicted values of the variables of interest at sensor locations were  
 12 constant. Thus, the solution was assumed to have converged [33]. A user-defined function is used in  
 13 FLUENT in order to impose vertical profiles of wind speed,  $U_{inlet}$ , turbulence kinetic energy,  $k_{inlet}$ ,  
 14 and the turbulence dissipation rate,  $\varepsilon_{inlet}$ , at the inlet of the computational domain. These profiles are  
 15 expressed in Equation (7), Equation (8) and Equation (9).

$$16 \quad U_{inlet}(y) = \frac{u_{ABL}^*}{\kappa} \ln\left(\frac{y + y_0}{y_0}\right) \quad (7)$$

$$17 \quad k_{inlet}(y) = \frac{u_{ABL}^{*2}}{\sqrt{C_\mu}} \quad (8)$$

$$18 \quad \varepsilon_{inlet}(y) = \frac{u_{ABL}^{*3}}{\kappa(y + y_0)} \quad (9)$$

19 where  $y$  is the height coordinate,  $y_0$  is the roughness length of the terrain,  $\kappa$  is the von Karman  
 20 constant,  $u_{ABL}^*$  is the ABL friction velocity and  $C_\mu$  is a model constant of the  $k - \varepsilon$  model.

21 The wind behavior close to the walls is modeled using the standard wall function [37]. The terrain of  
 22 the computational domain is decomposed into two surfaces. The first surface is the terrain where the

1 buildings are explicitly modelled. The second surface is the upstream terrain where the urban  
2 environment is implicitly modelled by imposing an equivalent roughness length,  $y_0$ , on the surface.  
3 The roughness length imposed on the surface of the upstream terrain is set to  $y_0 = 0.45m$ . In  
4 common CFD software packages, such as FLUENT, the sand-grain roughness height,  $k_s$ , is used to  
5 model the surface roughness instead of the roughness length,  $y_0$ . A relationship between the sand-  
6 grain roughness height,  $k_s$ , and the roughness length,  $y_0$  (commonly used in wind engineering), has  
7 been established by Blocken et al. for the standard-wall function implemented in FLUENT [38]. This  
8 relationship is expressed in Equation (10).

$$9 \quad k_s = \frac{9.793y_0}{C_s} \quad (10)$$

10 where  $C_s$  is the roughness constant.

11 By using Equation (10), the sand-grain roughness imposed on the surface of the upstream terrain has  
12 been set to  $k_{s,1} = 0.7m$  and the roughness constant has been set to  $C_{s,1} = 6$  in the CFD simulations.  
13 The same value of surface roughness is employed in order to define wind profiles at the inlet of the  
14 computational domain (Equation (7), Equation (8) and Equation (9)) in order to avoid unintended  
15 streamwise gradients in the upstream part of the computational domain. The roughness length  
16 imposed on the terrain surface where buildings are explicitly modelled is set to  $y_0 = 0.24m$ , which  
17 corresponds to a sand-grain roughness of  $k_{s,2} = 0.59m$  and a roughness constant of  $C_{s,2} = 4$ .  
18 Symmetrical boundary conditions are imposed on the sides and on the top of the computational  
19 domain. A zero-pressure boundary condition is imposed at the outlet of the computational domain.

20 A sensitivity analysis has been carried out in order to identify parameters with the highest impact on  
21 horizontal wind speeds and wind directions given their plausible ranges of values. These parameters  
22 were related to the size of the computational domain, the roughness of the terrain surfaces and the  
23 surrounding-building surfaces as well as the inlet boundary conditions (treated as parameters of the

1 CFD model). The inlet wind speed at 16m height,  $U_{inlet,16m}$ , the inlet wind direction,  $\mathcal{G}_{inlet}$ , and the  
 2 roughness of the surrounding-building surfaces,  $k_s$ , have the highest impact on wind predictions  
 3 given their prior ranges of values. Thus, these three parameters have been selected in the model-based  
 4 data interpretation framework. This work has been done by another researcher and is described in  
 5 [39].

6 Grid-based sampling is used to generate sets of parameter values that are assigned to the model class.

7 Table 1 presents plausible ranges of parameter values as well as discretization intervals defined by  
 8 engineering judgment and literature.

9 Table 1: Minimal and maximal values of parameters that need to be inferred as well as discretization  
 10 intervals

Parameter that needs to be inferred	Minimal value	Maximal value	Discretization intervals
Wind direction at the inlet form North, $\mathcal{G}_{inlet}$ [°]	0	337.5	22.5
Sand-grain roughness of surrounding-building surfaces, $k_s$ [m]	0.001	0.5	0.2495
Wind speed at the inlet (at 16m height), $U_{inlet,16m}$ [m/s]	0.5	8	0.1

11

12 CFD simulations have been executed for each combination of inlet wind direction,  $\mathcal{G}_{inlet}$ , and  
 13 roughness of the surrounding-building surfaces,  $k_s$ . A single value of inlet wind speed,  $U_{inlet,16m}$ , is  
 14 used in the CFD simulations. A total of 48 simulations have been executed in batch mode using 12  
 15 processes in parallel on a Windows Server 2012 containing four Hexa-Core Intel Xeon E54607  
 16 2.20GHz processors and 64 GB memory, which requires approximatively 200 hours of execution  
 17 time. When a new value of inlet wind direction is assigned to the model class in batch mode, the  
 18 outside box of the computational domain modifies its orientation, leading to the automatic generation

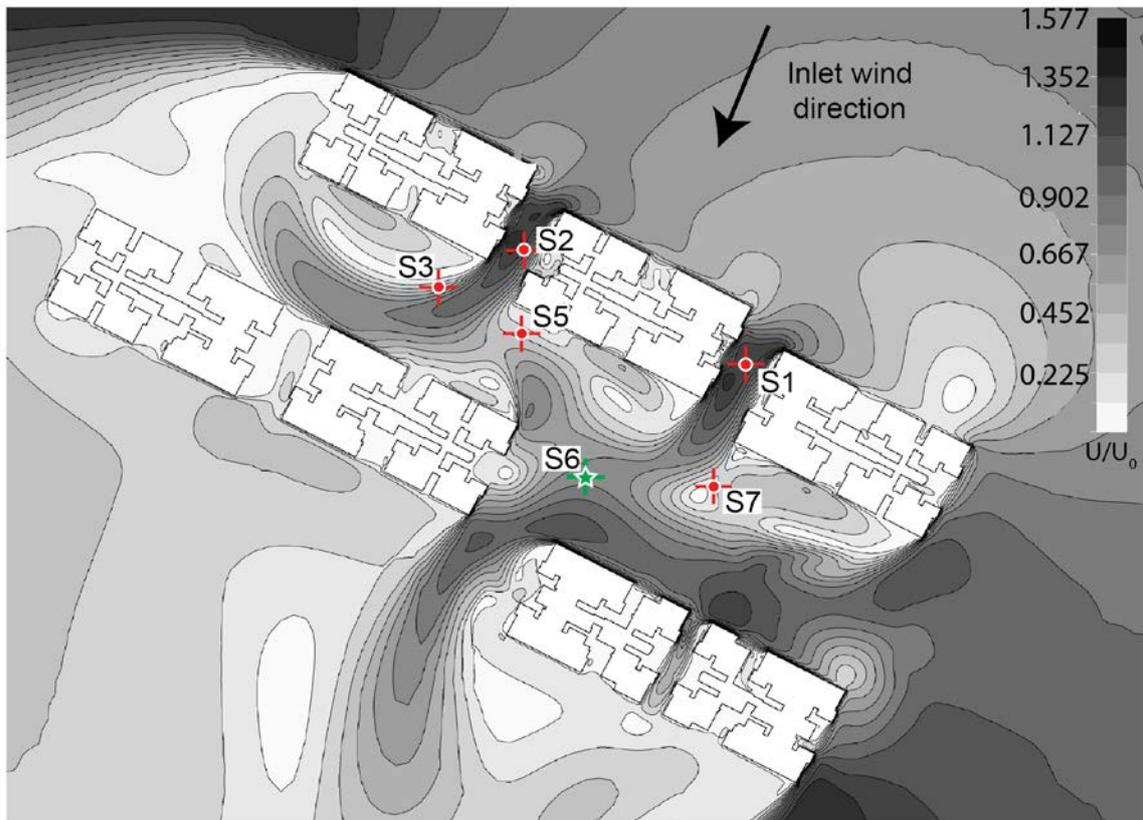
1 of a new grid. It is assumed that the grid settings obtained with the grid-sensitivity analysis provide  
2 good results for other inlet wind directions.

3 Wind predictions for other inlet wind speeds are obtained using a linear relationship. It was found that  
4 the same amplification factor of wind speeds,  $U/U_0$ , and the same wind directions are predicted  
5 when the value of inlet wind speed is varied. This was found by executing three CFD simulations with  
6 the same inlet wind direction and surface roughness and with a new value of inlet wind speed for each  
7 simulation.

8 48 CFD simulations have been executed through varying the inlet wind direction and the roughness of  
9 the surrounding-buildings. Wind predictions for 75 other inlet wind speeds are provided by  
10 considering that the amplification factor of wind speeds and the predicted wind direction are constant  
11 when the inlet wind speed is varied. A fine granularity of sampling in the inlet wind speed axis of the  
12 parameter space is needed in order to have a precise identification of the inlet wind speed. Moreover,  
13 if the granularity of sampling is too coarse, all model instances might be falsified using  
14 measurements. Thus, in total, wind predictions of 3648 model instances are obtained.

15 Horizontal wind speeds and wind directions are predicted at each location of interest and for each  
16 model instance. Figure 4 presents amplification factors of wind speeds predicted by one CFD  
17 simulation ( $\mathcal{G}_{inlet} = 22.5^\circ$  and  $k_s = 0.5m$ ) in a horizontal plane at sensor height (3m above the eco-  
18 deck). High spatial variability of wind speeds is observed around the buildings.

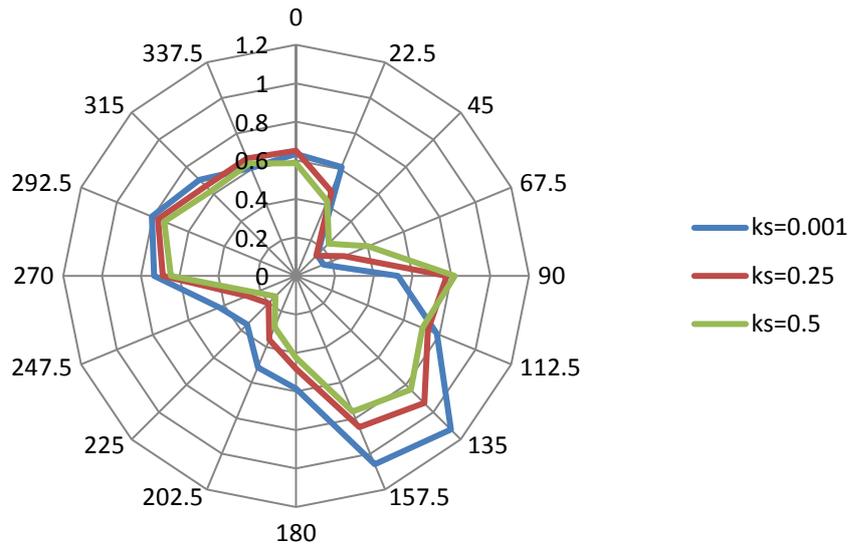
- Falsification sensor
- ★ Test sensor



1

2 Figure 4: Amplification factor of wind speeds predicted by a CFD simulation ( $\vartheta_{inlet} = 22.5^\circ$  and  
 3  $k_s = 0.5m$ ) in a horizontal plane at sensor height (3m above the eco-deck).

4 Figure 5 presents the impact of the inlet wind direction and the roughness of the surrounding-building  
 5 surfaces on the amplification factor of wind speeds predicted at sensor S3. The amplification factor of  
 6 wind speeds predicted at sensor S3 strongly depends on the wind direction at the inlet. Furthermore,  
 7 the impact of the surface roughness on the predicted amplification factor of wind speeds can go up to  
 8 0.30 when the wind direction at the inlet is  $135^\circ$  (South-East).



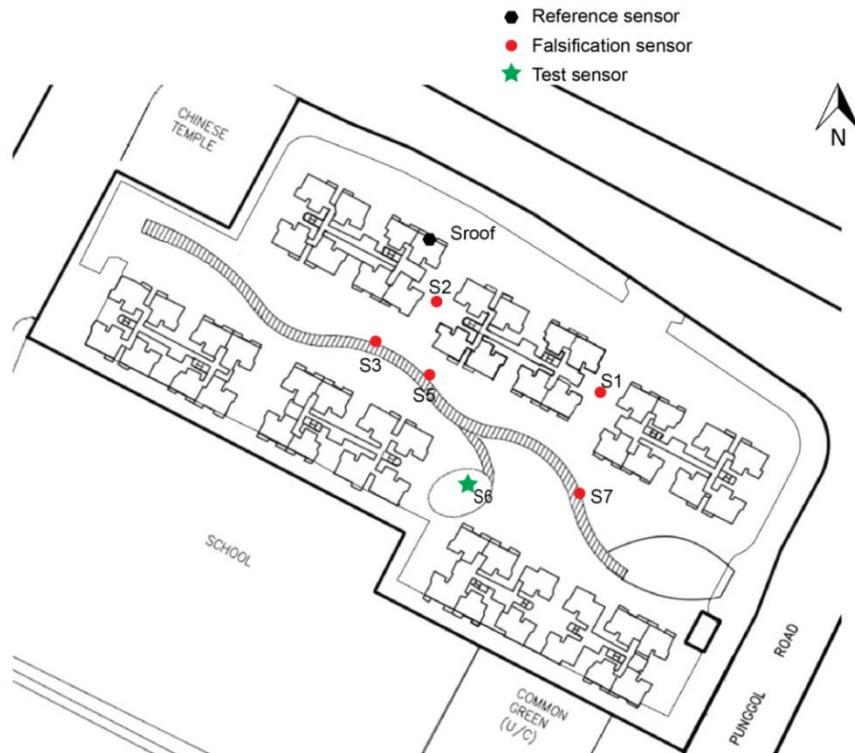
1

2 Figure 5: Impact of the inlet wind direction and the roughness of the surrounding building surfaces,  $k_s$ ,  
 3 , on the prediction of the amplification factor of wind speeds at sensor S3. The angular axis represents  
 4 the inlet wind direction and the radial axis represents the amplification factor of wind speeds predicted  
 5 at sensor S3.

6 **4. Measurement data**

7 Measurements were carried out during the North-East monsoon from February 24 to March 5, 2012.  
 8 The data was collected by Singapore-MIT Alliance for Research and Technology (SMART). No  
 9 sensor placement methodology has been used to define the locations of the weather stations. Six  
 10 weather stations are located on the eco-deck at the 2<sup>nd</sup> level ( $S_1$ - $S_2$ - $S_3$ - $S_5$ - $S_6$ - $S_7$ ) and one weather  
 11 station is located on the roof ( $S_{roof}$ ). Sensors  $S_1$ ,  $S_2$ ,  $S_3$ ,  $S_4$ ,  $S_5$  and  $S_7$  are used to falsify model  
 12 instances and the data from sensor  $S_6$  is used as test data to evaluate the predictions obtained by the  
 13 framework at unmeasured locations. The weather station located on the roof is used as reference  
 14 station to evaluate the effect of thermal processes on the wind behavior. Figure 6 presents a plan view  
 15 of the buildings as well as the locations of the weather stations.

16



1  
 2 Figure 6: Plan view showing the building footprints and the locations of the weather stations at  
 3 Treelodge@Punggol (modified from Pan ShanShan<sup>1</sup>).

4 Vaisala Weather Transmitters are employed to measure horizontal wind speed, wind direction and  
 5 temperature. A data logger was installed on each weather station in order to store measurement data.  
 6 The resolutions of the weather stations are 0.1m/s for horizontal wind-speed measurements and 1° for  
 7 wind-direction measurements. The installation of the weather stations is displayed in Figure 7.

---

<sup>1</sup> Internal report, Singapore-MIT Alliance for Research and Technology (SMART), 2014



1  
2 Figure 7: Installation of the weather stations (left: on roof; middle: on eco-deck trellis; right: on eco-  
3 deck railing) (modified from Pan ShanShan<sup>2</sup>).

#### 4 **5. Uncertainty quantification**

##### 5 **5.1. Uncertainty associated with RANS-based simulation in predictions of mean flow** 6 **variables and uncertainty associated with turbulent fluctuations**

7 Uncertainties associated with RANS-based simulation and turbulent fluctuations are estimated using a  
8 large eddy simulation (LES). LES is another approach for modelling turbulent flow, in which time-  
9 dependent predictions of flow variables are computed. Mean values of LES have been found to be  
10 much more accurate than values predicted with RANS-based simulations in regions of flow separation  
11 and recirculation when compared to wind-tunnel experiments [40]. RANS-based simulations are not  
12 able to provide accurate predictions in regions of flow separation and recirculation because flow is  
13 highly unsteady in those regions [40]. Moreover, LES is a transient approach and, thus, provides  
14 information about turbulent fluctuations in the area of interest. However, a LES takes two orders of  
15 magnitude ( $100\times$ ) longer to execute than a RANS-based simulation [41] and, therefore, it cannot be  
16 executed many times within model-based data interpretation frameworks.

---

<sup>2</sup> Internal report, Singapore-MIT Alliance for Research and Technology (SMART), 2014

1 Thus, LES is executed only once with the most likely wind conditions at the inlet. The goal is to  
2 identify spatial regions where RANS-based simulations are accurate (low differences between the  
3 predictions of the RANS-based simulation and the mean predictions of the LES) and regions where  
4 they are not (high differences between the predictions of the RANS-based simulation and the mean  
5 predictions of the LES). The advantage of using LES over wind tunnel experiments to estimate  
6 uncertainties is the possibility to compare and analyze high spatial-resolution data and the fact that  
7 LES does not suffer from experimental scale challenges.

8 In a previous study, mean predictions of a LES were compared with predictions of a RANS-based  
9 simulation in order to estimate uncertainties associated with RANS-based simulation. The results of  
10 this previous study have been used in this case study because the size of the buildings in these two  
11 case studies are similar (16-storey buildings) and the same turbulence model is used in both studies  
12 for the generation of model instances [42]. It was found that uncertainties associated with RANS-  
13 based simulation for the prediction of mean flow variable at a location of interest depend on the  
14 amplification factor of wind speeds predicted at that location. The amplification factor of wind  
15 speeds,  $U/U_0$ , is defined as the ratio between the predicted horizontal wind speed,  $U$ , and the inlet  
16 wind speed at the same height as the location of the prediction,  $U_0$ . Thus, spatially distributed  
17 modelling uncertainties are considered in the model-based data interpretation framework. Moreover,  
18 turbulent fluctuations were estimated using the LES. It was found that uncertainties associated with  
19 turbulent fluctuations depend on the averaging window used to average the measurements. It was  
20 found that fluctuations decrease rapidly when the averaging window increases from 1 minute to 15  
21 minutes and smaller reductions of these fluctuations were observed if the averaging window was  
22 further increased. Thus, an averaging window of 15 minutes was recommended. Since LES results are  
23 not entirely correct, levels of uncertainties estimated with this methodology are lower-bound  
24 estimates. Summary of results are presented in Table 3.

## 5.2. Uncertainty associated with thermal processes

The weather station on the roof ( $S_{\text{roof}}$ ) is used as a reference station in order to measure the ambient wind conditions (horizontal wind speed and wind direction). In the urban canopy layer (below the roof of the tallest building), the wind can be significantly affected by thermal processes because of the combination of low wind speeds below roof level and high differential heating between surfaces [43]. Regression analysis is employed in order to separate out the effect of thermal processes on the wind conditions below the roof level from other sources of variability. This is done using a data sample taken at different times of day, in which the ambient wind conditions are similar to mean ambient wind conditions. The mean ambient conditions are the mean values of the measured wind speeds and wind directions with respect to time.

Determining the sample size is difficult because of two competing objectives. On one hand, the sample size should be large enough in order to separate out the effect of thermal processes on the wind conditions below the roof level from other sources of variability. However, the sample size should be small enough in order to have similar ambient wind conditions. Thus, different sample sizes are used in the methodology. The largest sample size (2039 samples) corresponds to variations of  $28^\circ$  for wind direction and 3.8 m/s for horizontal wind speed while the smallest sample size (133 samples) corresponds to variations of  $5.6^\circ$  for wind direction and 0.76 m/s for horizontal wind speed. The regression analyses are expressed in Equation (11) and Equation (12).

$$U_{i,s} = a_{i,s} \times T_{i,s} + b_{i,s} \times U_{\text{roof},s} + c_{i,s} \quad (11)$$

$$\mathcal{G}_{i,s} = d_{i,s} \times T_{i,s} + e_{i,s} \times \mathcal{G}_{\text{roof},s} + f_{i,s} \quad (12)$$

where  $U_{i,s}$ ,  $\mathcal{G}_{i,s}$  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 samples  $s$ .  $U_{\text{roof},s}$  and  $\mathcal{G}_{\text{roof},s}$  are vectors of horizontal wind speeds and wind directions respectively measured at the reference station whose sizes correspond to the number of samples  $s$ .  $a_{i,s}$ ,  $b_{i,s}$ ,  $c_{i,s}$ ,  $d_{i,s}$ ,  $e_{i,s}$  and

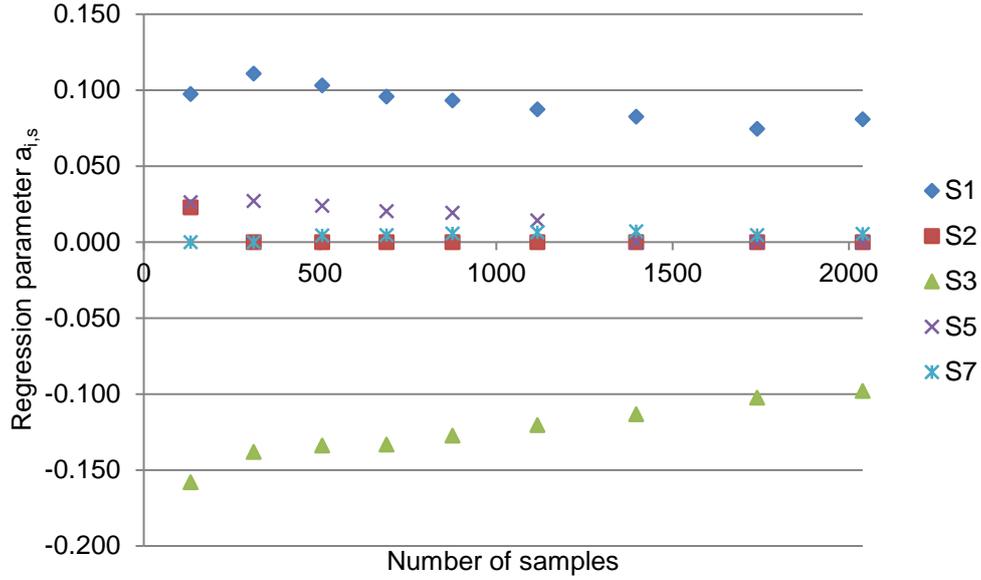
1  $f_{i,s}$  are the regression coefficients. A  $t$ -test is performed in order to find out if the regression  
 2 coefficients are significantly different from 0 [44]. If not, values of the regression coefficients are  
 3 reported to be 0.

4 Figure 8 presents values of the regression coefficients  $a_{i,s}$  with respect to the number of samples used  
 5 in the regression analyses at sensor location  $i$ .  $a_{i,s}$  is the regression coefficient that determines the  
 6 effect of thermal processes on the horizontal wind speed at sensor location  $i$  using a number of  
 7 samples  $s$ .

8 Values of the regression coefficient  $a_{i,s}$  are not constant with respect to the number of samples. The  
 9 maximal and minimal values of the coefficient  $a_{i,s}$  are used to estimate uncertainty bounds at sensor  
 10 locations. Bounds of uncertainty in the prediction of horizontal wind speed  $[\Delta U_{low,th,i}, \Delta U_{high,th,i}]$  at  
 11 sensor location  $i$  are computed using Equation (13).

$$12 \quad [\Delta U_{low,th,i}, \Delta U_{high,th,i}] = \left[ \min_s(a_{i,s}) \times (T_i - T_{min,i}), \max_s(a_{i,s}) \times (T_i - T_{min,i}) \right] \quad (13)$$

13 where  $T_i$  is the temperature measured at the sensor location  $i$  at the time when predictions at  
 14 unmeasured locations are sought.  $T_{min,i}$  is the minimal temperature measured over time at sensor  
 15 location  $i$ . It is thus assumed that the wind behavior is not affected by thermal processes when the  
 16 temperature is minimal (before sunrise). The horizontal wind speed is affected differently by thermal  
 17 processes from one sensor to another as presented in Figure 8. The regression coefficient  $a_{i,s}$  doesn't  
 18 change significantly with respect to the number of samples. The effect of thermal processes depends  
 19 also on the temperatures measured at sensor locations. Thus, time-dependent and spatially distributed  
 20 uncertainties are considered in the model-based data interpretation framework.



1

2 Figure 8: Regression coefficient  $a_{i,s}$  with respect to the number of samples

3 Uncertainties at unmeasured locations are represented by a uniform distribution with bounds

4  $[\Delta U_{low,th}, \Delta U_{high,th}]$  defined by Equation (14).

$$5 \quad [\Delta U_{low,th}, \Delta U_{high,th}] = \left[ \min_i \left( \min_s (a_{i,s}) \right) \times (\bar{T}_i - \bar{T}_{min,i}), \max_i \left( \max_s (a_{i,s}) \right) \times (\bar{T}_i - \bar{T}_{min,i}) \right] \quad (14)$$

6 where  $\bar{T}_i$  is the average value of temperatures measured at sensor locations.  $\bar{T}_{min,i}$  is the minimal  
7 value of  $\bar{T}_i$ .

8 The same procedure has been followed in order to estimate the effect of thermal processes on wind  
9 direction. Table 2 summarizes the minimal and maximal values of regression coefficients at sensor  
10 locations and at unmeasured locations.

11

12

13

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

Wind variable		$U [m / s]$		$\vartheta [^\circ]$	
		$\min_s(a_{sensor,i,s})$	$\max_s(a_{sensor,i,s})$	$\min_s(d_{sensor,i,s})$	$\max_s(d_{sensor,i,s})$
Sensor locations	S1	0.07	0.11	-1.62	-0.44
	S2	0	0.02	-0.44	-0.11
	S3	-0.16	-0.10	-3.65	-2.23
	S5	0	0.03	-4.37	-0.91
	S7	0	0.01	-7.29	-4.72
Unmeasured location		$\min_i(\min_s(a_{i,s}))$	$\max_i(\max_s(a_{i,s}))$	$\min_i(\min_s(d_{i,s}))$	$\max_i(\max_s(d_{i,s}))$
		-0.16	0.11	-7.29	-0.11

3

### 4 5.3. Uncertainty combination

5 Table 3 summarizes all sources of modelling uncertainty considered in this case study. Sources of  
 6 modelling uncertainty have been combined with measurement uncertainty using the Monte Carlo  
 7 method. Threshold bounds are defined using the combined uncertainty and a confidence level of 95%.  
 8 Threshold bounds are then used to falsify model instances that are defined with wrong sets of  
 9 parameter values (Section 2).

10

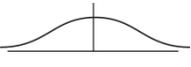
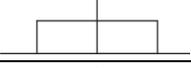
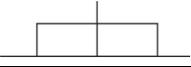
11

12

13

14

1 Table 3: Summary of modeling uncertainties used for the definition of threshold bounds

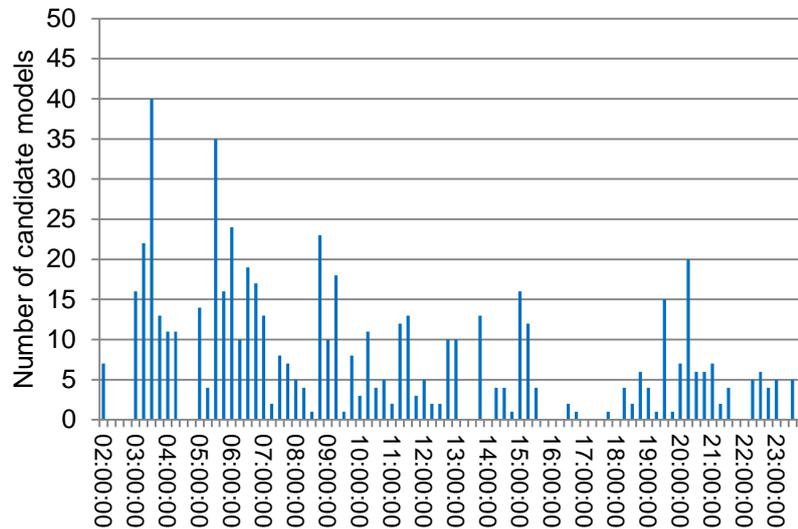
Source of uncertainty	Wind variable	Distribution	$U / U_0 < 0.33$	$U / U_0 \geq 0.33$
RANS-based simulation	$U$ [m/s]		$\mu = \frac{4.9}{100} \times U_0$ $\sigma = \frac{7.4}{100} \times U_0$	$\mu = \frac{1.0}{100} \times U_0$ $\sigma = \frac{6.1}{100} \times U_0$
	$\vartheta$ [°]		$\mu = 0; \sigma = 180$	$\mu = -0.8;$ $\sigma = 3.2$
turbulent fluctuations (averaging window=15min)	$U$ [m/s]		min = -0.13 ; max = 0.15	
	$\vartheta$ [°]		min = -22 ; max = 24	
thermal processes	$U$ [m/s]		<p>At sensor location <math>i</math> :</p> $\min = \min_s (a_{sensor,i,s}) \times (T_i - T_{min,i})$ $\max = \max_s (a_{sensor,i,s}) \times (T_i - T_{min,i})$ <p>At unmeasured locations:</p> $\min = \min_i \left( \min_s (a_{i,s}) \right) \times (\bar{T}_i - \bar{T}_{min,i})$ $\max = \max_i \left( \max_s (a_{i,s}) \right) \times (\bar{T}_i - \bar{T}_{min,i})$	
	$\vartheta$ [°]		<p>At sensor location <math>i</math> :</p> $\min = \min_s (d_{sensor,i,s}) \times (T_i - T_{min,i})$ $\max = \max_s (d_{sensor,i,s}) \times (T_i - T_{min,i})$ <p>At unmeasured locations:</p> $\min = \min_i \left( \min_s (d_{i,s}) \right) \times (\bar{T}_i - \bar{T}_{min,i})$ $\max = \max_i \left( \max_s (d_{i,s}) \right) \times (\bar{T}_i - \bar{T}_{min,i})$	

2

3 **6. Model falsification**

4 Falsification of model instances is performed every 15 min. Model instances are deemed candidate  
5 models if the differences between measured and predicted horizontal wind speeds as well as the  
6 differences between measured and predicted wind directions fall within threshold bounds at every  
7 sensor location. Figure 9 presents the number of candidate models at each time step on Feb. 28, 2012.  
8 The entire set of model instances is falsified 18 times out of 96 time steps (18.75% of the time) using

1 the proposed definition of threshold bounds. This suggests that the model class is not consistent with  
2 the error structure (Figure 2).



3  
4 Figure 9: Number of candidate models at each time step on Feb. 28, 2012 (3648 model instances  
5 generated initially)

6 In this case study, significant simplifications have been made to the geometry of the urban  
7 environment. For example, building voids and small-scale obstacles such as trees have been omitted.  
8 In order to account for these simplifications, an additional source of uncertainty in the prediction of  
9 horizontal wind speed has been added for the computation of threshold bounds. Table 4 presents  
10 plausible assumptions about bounds of uncertainties associated with geometric simplifications  
11 evaluated using engineering judgment. A uniform distribution is assumed between bounds. In this  
12 study, uncertainties associated with geometric simplifications are assumed to be constant with respect  
13 to space.

14  
15

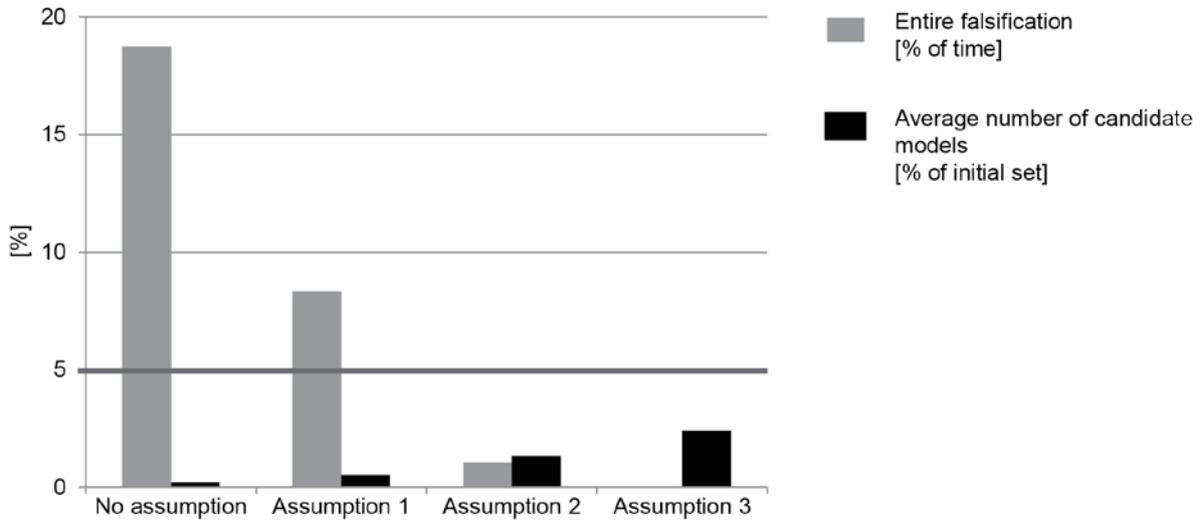
1 Table 4: Plausible assumptions about bounds of uncertainties associated with geometric  
 2 simplifications

	<b>Lower bound</b>	<b>Upper bound</b>
Assumption 1	-0.25m/s	0.25m/s
Assumption 2	-0.5m/s	0.5m/s
Assumption 3	-0.75m/s	0.75m/s

3  
 4 Figure 10 presents the percentage of time that the entire set of model instances is falsified as well as  
 5 the average number of candidate models over time on Feb. 28, 2012 (as a percentage of the number of  
 6 model instances initially generated) for the different assumptions related to uncertainties mentioned in  
 7 Table 4. The following methodology is used to choose the most representative assumption:

- 8 1) Assumptions of levels of modelling uncertainties that lead to the falsification of the entire set  
 9 of model instances less than 5% of the time are retained. This choice is justified because  
 10 threshold bounds have been defined using a confidence level of 95% (Section 2). Thus, there  
 11 is a probability of 5% to falsify a correct set of parameter values (Type-I error) and a  
 12 probability of falsifying the entire set of model instances of at most 5%.
- 13 2) Among the retained assumptions, the assumption related to modelling uncertainties that  
 14 provides the smallest amount of candidate models is chosen. This means that, among the  
 15 retained assumptions, the assumption that provides the smallest probability of accepting a  
 16 wrong set of parameter values is chosen (Type-II error).

17 Assumption 2 provides the smallest amount of candidate models among the assumptions that lead to  
 18 the falsification of the entire model set less than 5% of the time. Thus, this assumption is used to  
 19 represent uncertainties associated with geometric simplifications.



1

2 Figure 10: Percentage of time that the entire model set is falsified as well as the average number of  
 3 candidate models over time (as a percentage of the number of model instances initially generated) on  
 4 Feb. 28, 2012 for different assumptions about uncertainties associated with geometric simplifications

5 The proposed framework leads to modifying approximations of uncertainties through falsification of  
 6 the entire model set before making predictions at unmeasured locations.

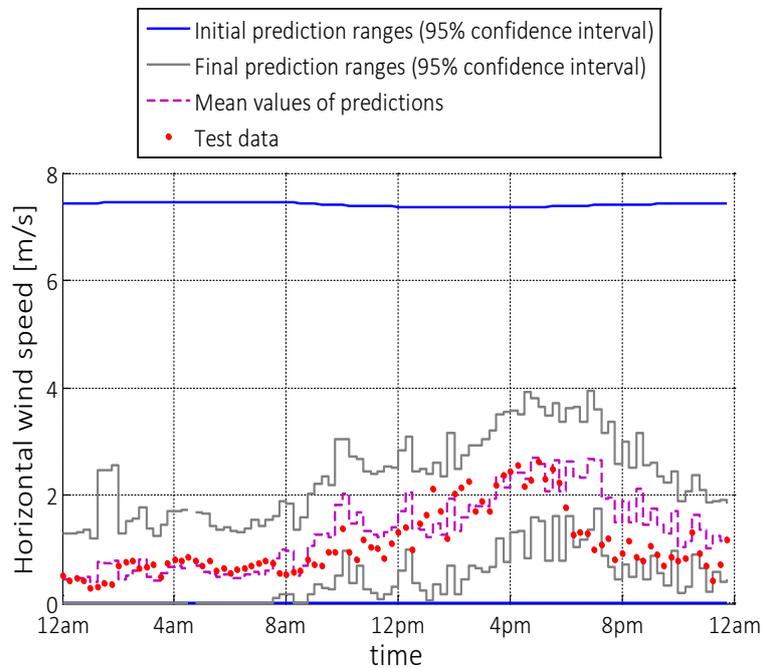
7 **7. Predictions at unmeasured locations and evaluation with test data**

8 Predictions of the candidate model set at an unmeasured location correspond to the uncertainty  
 9 associated with parameter values propagated through the model, which has been reduced through  
 10 measurements (Section 2). For each candidate-model prediction at an unmeasured location of interest,  
 11 samples are drawn from modelling uncertainty sources at that location using the Monte Carlo method  
 12 in order to build the posterior predictive distribution. Ranges of predictions corresponding to a  
 13 confidence level of 95% are computed using the posterior predictive distribution.

14 Figure 11 presents the predictions ranges of horizontal wind speeds before and after measurements,  
 15 the mean values of predicted horizontal wind speeds as well as the measured values at the test sensor  
 16 on Feb. 28, 2012.

1 On average, over all time steps, the prediction range of horizontal wind speeds has been reduced by  
2 74% after measurements. This corresponds to a mean prediction range of 1.9 m/s. The reliability  
3 index in the prediction of horizontal wind speed is 0.95. The reliability index is defined as the ratio of  
4 the number of time steps when the test data fall within the prediction range to the total number of time  
5 steps. Although prediction ranges have been reduced after measurements, the reliability index remains  
6 high.

7 The root-mean-square difference between the measured values of horizontal wind speeds and the  
8 mean values of predictions at test sensor S6 is 0.51m/s. The model-based data interpretation  
9 framework is able to predict reliable time-dependent prediction ranges of horizontal wind speeds at  
10 unmeasured locations. Between 7pm to 9pm, predictions are biased (high differences between mean  
11 values of predictions and the test data). These biases are due to the biased estimates of uncertainties at  
12 this time of day (e.g. effect of thermal processes). Improvement in the estimation of uncertainties  
13 would provide better predictions.

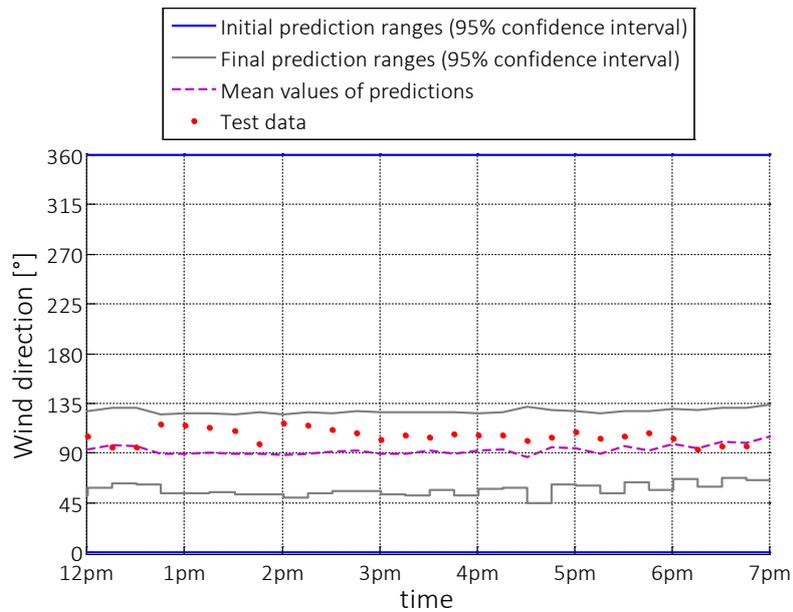


14  
15 Figure 11: Predicted and measured values of horizontal wind speeds at test sensor S6 on Feb. 28, 2012

1 Figure 12 presents the predictions ranges of wind directions before and after measurements, the mean  
 2 values of predicted wind directions as well as the measured values at the test sensor from 12pm to  
 3 7pm. In this figure, only time steps between 12pm to 7pm are presented. Indeed, from 12pm to 7pm,  
 4 all candidate models obtained after measurements predict high amplification factors of wind speeds (  
 5  $U / U_0 \geq 0.33$ ) at test sensor S6. Therefore, small uncertainties associated with RANS-based  
 6 simulation are added to the candidate-model predictions (Table 3). During this period, the mean  
 7 prediction range of wind direction is  $69^\circ$ . This corresponds to a reduction of prediction range of  
 8 approximately 81%.

9 At the other time steps (not presented in Figure 12), no reduction of prediction range is obtained after  
 10 measurements because several candidate models predict a low amplification factor of wind speeds at  
 11 the test sensor and, thus, provides predictions with high uncertainties associated with RANS-based  
 12 simulation (Table 3).

13 From 12pm to 7pm, the root-mean-square difference between the measured values of wind directions  
 14 and the mean values of predictions at test sensor S6 is  $18.1^\circ$ .



15

1 Figure 12: Predicted and measured values of wind directions at test sensor S6 on Feb. 28 2012 from  
2 12pm to 7pm

### 3 **8. Discussion**

4 In this work, a model-based data interpretation framework is proposed to improve wind predictions at  
5 unmeasured locations using measurements. This framework only requires the estimation of  
6 uncertainty bounds in order to falsify incorrect set(s) of parameter values and, thus, refine predictions  
7 at unmeasured locations. Therefore, this framework is suitable when knowledge of uncertainties,  
8 including error correlations between measurement locations, is incomplete. In contrast to the  
9 traditional CFD simulation methodology using a single model, prediction ranges are available and  
10 therefore, information about possible variations in responses due to modelling uncertainties is  
11 obtained.

12 Limitations are as follows: In the present study, three parameters have been selected for inference  
13 because: a) sensitivity analyses revealed that these three parameters dominated the variability of  
14 predictions, b) CFD simulations require high computation time and c) sampling in the parameter  
15 space using a grid-based approach is an exponentially complex task with respect to the number of  
16 parameters. Even with only three parameters selected for inference, the granularity of sampling was  
17 still coarse (e.g. inlet wind direction discretized into intervals of  $22.5^\circ$ ). More efficient sampling  
18 techniques, such as the use of surrogate models, can be used to reduce the time needed to generate  
19 model instances and, thus, allow denser sampling and/or sampling in higher parameter spaces.

20 In this study, it is assumed that the uncertainties associated with geometric simplifications are  
21 constant with respect to space. This might not be a valid assumption. In future work, strategies should  
22 be proposed to provide better estimates of uncertainties associated with geometric simplifications,  
23 including their spatial variability. An alternative strategy is to improve the model class without  
24 changing the error structure used to define threshold bounds. This requires more detailed information  
25 related to aspects such as building geometries.

1 Simple linear regression analyses were used to separate out the effects of thermal processes on the  
2 wind conditions below roof level. In future work, other statistical methods should be employed and  
3 their performance should be compared.

4 Finally, the scale of the study is small (seven buildings). In future work, the framework should be  
5 evaluated for an extended area, such as a district.

## 6 **9. Conclusions**

7 This study has led to the following conclusions:

- 8 • Reliable prediction ranges of horizontal wind speeds and wind directions at an unmeasured  
9 location can be obtained at different time steps using the proposed model-based data  
10 interpretation framework.
- 11 • Tighter prediction ranges of horizontal wind speeds and wind directions are possible using the  
12 framework without compromising reliability.
- 13 • The proposed model-based data interpretation framework leads to new approximations of  
14 uncertainties through falsification of the entire model set.

## 15 **Acknowledgements**

16 This research has been conducted at the Singapore-ETH Centre for Global Environmental  
17 Sustainability (SEC), co-funded by the Singapore National Research Foundation (NRF) and ETH  
18 Zurich. The authors would like to thank Pan ShanShan for the data recorded at the Treelodge in  
19 Punggol as part of research conducted by Nanyang Technological University and the Singapore-MIT  
20 Alliance for Research and Technology (SMART)'s Center for Environmental Sensing and Modeling  
21 (CENSAM), with support from the National Research Foundation and the Ministry of Education.

## 22 **References**

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

- 1 2. Blocken, B. and J. Persoon, *Pedestrian wind comfort around a large football stadium in an*  
2 *urban environment: CFD simulation, validation and application of the new Dutch wind*  
3 *nuisance standard*. Journal of Wind Engineering and Industrial Aerodynamics, 2009. **97**(5):  
4 p. 255-270.
- 5 3. Blocken, B. and J. Carmeliet, *Pedestrian wind conditions at outdoor platforms in a high-rise*  
6 *apartment building: generic sub-configuration validation, wind comfort assessment and*  
7 *uncertainty issues*. Wind and Structures, 2008. **11**(1): p. 51-70.
- 8 4. Tominaga, Y., S. Murakami, and A. Mochida, *CFD prediction of gaseous diffusion around a*  
9 *cubic model using a dynamic mixed SGS model based on composite grid technique*. Journal of  
10 Wind Engineering and Industrial Aerodynamics, 1997. **67-68**: p. 827-841.
- 11 5. Leitzl, B.M., et al., *Concentration and flow distributions in the vicinity of U-shaped buildings:*  
12 *wind-tunnel and computational data*. Journal of Wind Engineering and Industrial  
13 Aerodynamics, 1997. **67-68**: p. 745-755.
- 14 6. Blocken, B., et al., *Numerical evaluation of pollutant dispersion in the built environment:*  
15 *comparisons between models and experiments*. Journal of Wind Engineering and Industrial  
16 Aerodynamics, 2008. **96**(10): p. 1817-1831.
- 17 7. Defraeye, T., B. Blocken, and J. Carmeliet, *Convective heat transfer coefficients for exterior*  
18 *building surfaces: Existing correlations and CFD modelling*. Energy Conversion and  
19 Management, 2011. **52**(1): p. 512-522.
- 20 8. Defraeye, T. and J. Carmeliet, *A methodology to assess the influence of local wind conditions*  
21 *and building orientation on the convective heat transfer at building surfaces*. Environmental  
22 Modelling & Software, 2010. **25**(12): p. 1813-1824.
- 23 9. Blocken, B., et al., *High-resolution CFD simulations for forced convective heat transfer*  
24 *coefficients at the facade of a low-rise building*. Building and Environment, 2009. **44**(12): p.  
25 2396-2412.
- 26 10. Van Hooff, T. and B. Blocken, *Coupled urban wind flow and indoor natural ventilation*  
27 *modelling on a high-resolution grid: A case study for the Amsterdam ArenA stadium*.  
28 Environmental Modelling & Software, 2010. **25**(1): p. 51-65.
- 29 11. Jiang, Y. and Q. Chen, *Effect of fluctuating wind direction on cross natural ventilation in*  
30 *buildings from large eddy simulation*. Building and Environment, 2002. **37**(4): p. 379-386.
- 31 12. Ramponi, R. and B. Blocken, *CFD simulation of cross-ventilation flow for different isolated*  
32 *building configurations: validation with wind tunnel measurements and analysis of physical*  
33 *and numerical diffusion effects*. Journal of Wind Engineering and Industrial Aerodynamics,  
34 2012. **104-106**: p. 408-418.
- 35 13. Ramponi, R. and B. Blocken, *CFD simulation of cross-ventilation for a generic isolated*  
36 *building: impact of computational parameters*. Building and Environment, 2012. **53**: p. 34-  
37 48.
- 38 14. Tamura, T., et al., *Numerical prediction of wind loading on buildings and structures—*  
39 *Activities of AIJ cooperative project on CFD*. Journal of Wind Engineering and Industrial  
40 Aerodynamics, 1997. **67-68**: p. 671-685.
- 41 15. Nozu, T., et al., *LES of the flow and building wall pressures in the center of Tokyo*. Journal of  
42 Wind Engineering and Industrial Aerodynamics, 2008. **96**(10): p. 1762-1773.
- 43 16. Schatzmann, M. and B. Leitzl, *Issues with validation of urban flow and dispersion CFD*  
44 *models*. Journal of Wind Engineering and Industrial Aerodynamics, 2011. **99**(4): p. 169-186.
- 45 17. Beven, K., *Environmental modelling: An uncertain future?* 2008: Taylor & Francis.
- 46 18. Beven, K., *A manifesto for the equifinality thesis*. Journal of hydrology, 2006. **320**(1): p. 18-  
47 36.
- 48 19. Robert-Nicoud, Y., B. Raphael, and I. Smith, *Configuration of measurement systems using*  
49 *Shannon's entropy function*. Computers & Structures, 2005. **83**(8): p. 599-612.
- 50 20. Newman, S. and P.J. Wierenga, *Comprehensive Strategy of Hydrogeologic Modeling and*  
51 *Uncertainty Analysis for Nuclear Facilities and Sites*. 2003.
- 52 21. Feyen, L., et al., *Bayesian methodology for stochastic capture zone delineation incorporating*  
53 *transmissivity measurements and hydraulic head observations*. Journal of hydrology, 2003.  
54 **271**(1): p. 156-170.

- 1 22. Marshall, L., D. Nott, and A. Sharma, *A comparative study of Markov chain Monte Carlo*  
2 *methods for conceptual rainfall-runoff modeling*. Water Resources Research, 2004. **40**(2).
- 3 23. Yang, J., et al., *Hydrological modelling of the Chaohe Basin in China: Statistical model*  
4 *formulation and Bayesian inference*. Journal of hydrology, 2007. **340**(3): p. 167-182.
- 5 24. Tebaldi, C., et al., *Quantifying uncertainty in projections of regional climate change: A*  
6 *Bayesian approach to the analysis of multimodel ensembles*. Journal of Climate, 2005.  
7 **18**(10): p. 1524-1540.
- 8 25. Min, S.-K., D. Simonis, and A. Hense, *Probabilistic climate change predictions applying*  
9 *Bayesian model averaging*. Philosophical Transactions of the Royal Society A: Mathematical,  
10 Physical and Engineering Sciences, 2007. **365**(1857): p. 2103-2116.
- 11 26. Beven, K.J., P.J. Smith, and J.E. Freer, *So just why would a modeller choose to be*  
12 *incoherent?* Journal of hydrology, 2008. **354**(1): p. 15-32.
- 13 27. Goulet, J.-A., C. Michel, and I.F.C. Smith, *Hybrid probabilities and error-domain structural*  
14 *identification using ambient vibration monitoring*. Mechanical Systems and Signal  
15 Processing, 2012. **37**(1): p. 199-212.
- 16 28. Goulet, J.A. and I.F.C. Smith, *Performance-driven measurement system design for structural*  
17 *identification*. Journal of Computing in Civil Engineering, 2012. **27**(4): p. 427-436.
- 18 29. Goulet, J.-A., S. Coutu, and I.F.C. Smith, *Model falsification diagnosis and sensor placement*  
19 *for leak detection in pressurized pipe networks*. Advanced Engineering Informatics, 2013.  
20 **27**(2): p. 261-269.
- 21 30. Vernay, D.G., B. Raphael, and I.F.C. Smith, *Augmenting simulations of airflow around*  
22 *buildings using field measurements*. Advanced Engineering Informatics, 2014. **28**(4): p. 412-  
23 424.
- 24 31. Goulet, J.-A. and I.F.C. Smith, *Structural identification with systematic errors and unknown*  
25 *uncertainty dependencies*. Computers & Structures, 2013. **128**: p. 251-258.
- 26 32. Šidák, Z., Rectangular confidence regions for the means of multivariate normal distributions.  
27 Journal of the American Statistical Association, 1967. 62(318): p. 626-633.
- 28 33. Franke, J., et al., *Best practice guideline for the CFD simulation of flows in the urban*  
29 *environment*. COST Action 732: Quality assurance and improvement of microscale  
30 meteorological models, 2007.
- 31 34. Shih, T.-H., et al., A new  $k-\epsilon$  eddy viscosity model for high reynolds number turbulent flows.  
32 Computers & Fluids, 1995. 24(3): p. 227-238.
- 33 35. Patankar, S.V. and D.B. Spalding, *A calculation procedure for heat, mass and momentum*  
34 *transfer in three-dimensional parabolic flows*. International Journal of Heat and Mass  
35 Transfer, 1972. **15**(10): p. 1787-1806.
- 36 36. Keating, M., *Accelerating CFD solutions*. ANSYS Advantage, 2011. **5**(1): p. 48-49.
- 37 37. Launder, B.E. and D. Spalding, *The numerical computation of turbulent flows*. Computer  
38 methods in applied mechanics and engineering, 1974. **3**(2): p. 269-289.
- 39 38. Blocken, B., T. Stathopoulos, and J. Carmeliet, *CFD simulation of the atmospheric boundary*  
40 *layer: wall function problems*. Atmospheric environment, 2007. **41**(2): p. 238-252.
- 41 39. Papadopoulou, M., *Optimal sensor placement for prediction of wind environment around*  
42 *buildings*, NUS Ph.D. thesis, 2015
- 43 40. Tominaga, Y., et al., *Comparison of various revised  $k-\epsilon$  models and LES applied to flow*  
44 *around a high-rise building model with 1: 1: 2 shape placed within the surface boundary*  
45 *layer*. Journal of Wind Engineering and Industrial Aerodynamics, 2008. **96**(4): p. 389-411.
- 46 41. Martilli, A., J.L. Santiago, and F. Martín, *Micrometeorological modelling in urban areas*.  
47 Física de la Tierra, 2008. **19**: p. 133-145.
- 48 42. Vernay, D. G., Raphael, B., & Smith, I. F.C. A model-based data-interpretation framework  
49 for improving wind predictions around buildings. *Journal of Wind Engineering and Industrial*  
50 *Aerodynamics*, 2015, **145**: p. 219-228.
- 51 43. Niachou, K., I. Livada, and M. Santamouris, *Experimental study of temperature and airflow*  
52 *distribution inside an urban street canyon during hot summer weather conditions. Part II:*  
53 *airflow analysis*. Building and Environment, 2008. **43**(8): p. 1393-1403.
- 54 44. Zar, J.H., *Biostatistical analysis. 2nd*. Prentice Hall USA, 1984.
- 55

- 1 This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0
- 2 International License



3