# Optimal Sensor Placement and Measurement of Wind for Water Quality Studies in Urban Reservoirs

Wan Du[†], Zikun Xing[§], Mo Li[†], Bingsheng He[†], Lloyd Hock Chye Chua[§] and Haiyan Miao[‡]
[†]School of Computer Engineering, Nanyang Technological University, Singapore
[§]School of Civil and Environmental Engineering, Nanyang Technological University, Singapore
[‡]Institute of High Performance Computing (IHPC), A*Star, Singapore

***Abstract*—We collaborate with environmental scientists to study the hydrodynamics and water quality in an urban district, where the surface wind distribution is an essential input but undergoes high spatial and temporal variations due to the complex urban landform created by surrounding buildings. In this work, we study an optimal sensor placement scheme to measure the wind distribution over a large urban reservoir with a limited number of wind sensors. Unlike existing sensor placement solutions that assume Gaussian process of target phenomena, this study measures the wind which inherently exhibits strong non-Gaussian yearly distribution. By leveraging the local monsoon characteristics of wind, we segment a year into different monsoon seasons which follow a unique distribution respectively. We also use computational fluid dynamics to learn the spatial correlation of wind in the presence of surrounding buildings. The output of sensor placement is a set of the most informative locations to deploy the wind sensors, based on the readings of which we can accurately predict the wind over the entire reservoir surface in real time. 10 wind sensors are finally deployed around or on the water surface of an urban reservoir. The in-field measurement results of more than 3 months suggest that the proposed sensor placement and spatial prediction approach provides accurate wind measurement which outperforms the state-of-the-art Gaussian model based or interpolation based approaches.**

***Keywords*—*Sensor placement; Spatial prediction; Wind measurements; Water Quality; Urban reservoir***

## I. Introduction

A healthy aquatic ecosystem and water quality monitoring is essential for good understanding of the water resources and social security, especially for countries with limited water resources like Singapore. Recent limnological studies [1, 2] reveal that the distribution of wind stress on the surface of a lake can significantly impact water hydrodynamics and affects water quality. Most existing limnological studies are conducted in rural lakes and based on simple assumptions of surface wind including uniform [1] or interpolated surface wind distribution [3]. In this work, we collaborate with environmental scientists to understand the effect of wind on the water quality of Marina Reservoir in Singapore, a typical urban water field. It is located in downtown of Singapore with a water surface of $2.2km^2$, as depicted in Fig. 1. Due to seasonal effects and the urban landform created by a variety of high-rise buildings surrounding two basins of the reservoir, the wind field is of high temporal and spatial variations.

To investigate the impact on water quality evolution numerically, the wind distribution above water surface as well as other environmental parameters (e.g., air temperature and precipitation) are used as inputs to a three-dimensional hydrodynamics-ecological model, Estuary Lake and Coastal Ocean Model - Computational Aquatic Ecosystem Dynamics Model (ELCOM-CAEDYM) [3]. In a previous study on sensitivity analysis [4], based on uniform wind distributions, we have found that in Marina Reservoir, wind forcing variability has significant impact on vertical and spatial variability of phytoplankton distribution which can cause substantial change to the water quality. In this study, we deploy a limited number of wind sensors to measure the wind direction and speed. Based on the limited sensor readings, we derive the wind distribution over the entire Marina reservoir. The accurate wind distribution is critical for studying and predicting the water quality in Marina reservoir. In order to maximize the accuracy of field measurements, we need to find the most informative locations to deploy the wind sensors, based on the observations from which we can accurately predict the wind at other unobserved locations. The optimal sensor placement together with the spatial prediction is therefore the key problem this paper will address.

The problem of optimal sensor placement has been studied in many applications that monitor spatial phenomenon, like temperature sensing [5] and field soil moisture estimation [6]. Techniques like spatial statistics [7] and subset selection [8] have been proposed in previous works. As commonly assumed in those studies, the underlying phenomenon at one location can be modeled by a Gaussian distribution and the phenomena over the target area is thus a Gaussian Process (GP), where the marginal and conditional distributions of a multi-variant Gaussian distribution are still Gaussian. The optimal sensor placement is then calculated as the most informative locations by information theory criteria like entropy [7] or mutual information [9]. Based on the sensor readings, spatial prediction is performed by estimating the posterior values of unobserved locations through Gaussian regression. In this paper, we also refer to wind distribution as the wind field over the target area at a given point in time.

Unfortunately, existing GP based approaches cannot be applied to wind measurement in this study mainly due to the following three challenges. First, as we will detail in Section II, the wind directions in the field do not follow Gaussian process over time. Blindly applying GP based approaches assuming Gaussian distribution of wind directions leads to sub-optimal sensor placement and incurs large errors in spatial prediction.


Fig. 1: Water surface and surrounding topography of Marina Reservoir in Singapore.

Second, existing approaches typically require sufficient prior knowledge on data distribution (usually collected from a denser pre-deployment) to train their GP model so as to capture pairwise correlations among different locations. Such prior knowledge is not available in our study. We do not possess historical wind distribution data of the field and it is cost prohibitive for us to pre-deploy numerous wind sensors to gain such knowledge. Third, in our study the water quality in the reservoir has varied sensitivity to the wind input at different locations due to diverse morphometrics and flow patterns, which calls for non-uniform measurement accuracy over the field. We need to optimize the sensor placement in a sense that the sensors are deployed at locations with higher sensitivity to wind variations.

In this paper we propose a novel sensor placement and measurement approach to address the above challenges. We propose a mixture model of wind as the sum of several Gaussian and uniform distributions. Inspired by the local monsoon characteristics of wind in Singapore, we do time series segmentation and divide one year into periods of different monsoon or intermonsoon seasons, during which the wind can be described or transformed to different Gaussian distributions and different prediction models can thus be trained. To derive the prediction model in each season, we obtain wind correlations among different locations in the field through Computational Fluid Dynamics (CFD) simulation instead of learning from pre-deployments. The optimal sensor placement is determined based on the information utility for all seasons and adjusted according to the sensitivity of water quality to wind in the field. When the sensor readings are collected at real-time, we use an online clustering algorithm to flexibly determine the boundaries of these seasons with instant wind measurement in different years, and thus perform proper spatial prediction accordingly. Finally, to further consider water quality prediction before deploying any wind sensors, we conduct a series of ELCOM-CAEDYM simulations for sensitivity analysis of water quality to the wind input and adjust accordingly the sensor placement scheme to factor the non-uniform accuracy requirement in wind measurement.

10 wind sensors are finally deployed around or on the water surface of Marina Reservoir according to the sensor placement scheme obtained from our analytical results. More than 3 month in-field measurement results suggest that the proposed approach provides accurate spatial prediction of wind in both time and space. Compared with previous GP or interpolation based approaches, our approach reduces average root-mean-squared error of measurement in wind direction by 81% and 26% respectively.

The rest of this paper is organized as follows. Section II gives the problem statement and presents the overview of the proposed approach. Section III presents the detailed design and analysis of the approach. Section IV describes the in-field deployment experience and presents the experimental evaluation results. Section V summarizes the lessons we learnt from this work and Section VI introduces related works. Section VII concludes this paper.

## II. Problem Statement and Overview

In this section, we formally formulate the sensor placement and spatial prediction problem. We present the unique challenges from our application and an overview of our approach.

### A. Problem Statement

In this wind measurement application, we divide Marina Reservoir into small grids of 20m*20m. We assume that each grid is a location with uniform wind field. Totally, we need to cover more than 5k locations. The set of all locations over Marina Reservoir is denoted as $\mathcal{V}$, where $|\mathcal{V}| = N$. The observations at each location $v_i \in \mathcal{V}$ can be modeled as a random variable $X_i$. All variables jointly form a random process. The objective of optimal sensor placement is to select a subset $\mathcal{A}$, $\mathcal{A} \subset \mathcal{V}$ and $|\mathcal{A}| = K << N$, from which we can predict the observations of the other locations, presented as $\mathcal{V} \backslash \mathcal{A}$, with minimum estimation errors.

Common approaches that have been applied to similar spatial prediction problems assume that the random variable $X_i$ at each location follows a Gaussian distribution and the joint distribution of the variables over all locations can be modeled as a Gaussian process [5, 10]. With such GP assumption, existing approaches benefit from the feature that the marginal and conditional distributions of a multi-variant Gaussian distribution are still Gaussian. Therefore, the most important sensor locations can be selected by some informative criteria like entropy [7] or mutual information [9]. The observations on the other unobserved locations can then be predicted as the mean of conditional distribution $X_{\mathcal{V}\backslash\mathcal{A}}|X_{\mathcal{A}}$ with an uncertainty $\sigma^2_{v|\mathcal{A}}$:

$$\mu_{v|\mathcal{A}} = \mu_v + \sum_{v\mathcal{A}} \sum_{\mathcal{A}\mathcal{A}}^{-1} (x_{\mathcal{A}} - \mu_{\mathcal{A}}) \tag{1}$$

$$\sigma^2_{v|\mathcal{A}} = \sum_{v,v} - \sum_{v\mathcal{A}} \sum_{\mathcal{A}\mathcal{A}}^{-1} \sum_{v\mathcal{A}}^{T} \tag{2}$$

where $\sum_{v\mathcal{A}}$ is a vector of covariance between $v$ and each element in $\mathcal{A}$, and $\sum_{\mathcal{A}\mathcal{A}}$ is the covariance matrix of $\mathcal{A}$.

The GP assumption, however, does not hold for wind directions over a large time period in our application. As depicted in Fig. 3a, the actual distribution of wind directions over one year is far from Gaussian. The data is collected by the meteorological station in Marina Channel over the year 2007. The inaccurate Gaussian fitting leads to large errors in understanding correlations within the wind field. As a result, it jeopardizes the results of sensor placement and

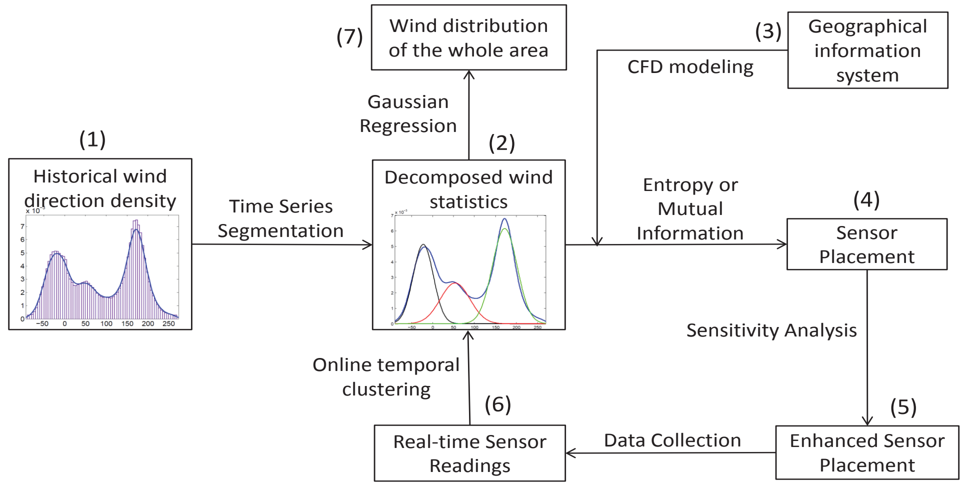



Fig. 2: Framework of the proposed approach. The design procedure is illustrated by the sequence number in brackets.

spatial prediction. As will be shown in Section IV, the average prediction error of wind direction will reach as high as 89° if we blindly apply such a biased and mis-modeled fitting.

In addition, to train the GP model, existing approaches require full prior-knowledge of data distribution over the entire field such that the pairwise correlations of all locations in the field can be captured. For instance, in [9], the training data for the GP model is collected with 54 temperature sensors pre-deployed for 5 days with a sampling interval of 0.5min. Another example is community sensing in [11] which provides the best route prediction based on a GP model trained by 2110 route planning requests obtained from volunteers during 2006 and 2007. Such full prior knowledge about the wind over Marina Reservoir, however, is not available. Intrusively gaining such knowledge through pre-deploying sensors is also not possible. First, it is cost prohibitive to deploy an adequate number of sensors (more than 5k) to precisely cover the underlying field. In our study, the cost of a land sensor is about 6000 USD and a floating sensor on the water surface costs about 8000 USD. We can only plan the most informative sensor placement beforehand and then deploy a limited number of sensors (10 in this study). Second, due to topography and regulatory requirements in such an iconic center of the city, we are not able to deploy sensors at all desired places for full data survey. As a matter of fact, it took us several months to get permits from Singapore government agencies for deploying the wind sensors in the allowed areas as shown in Fig. 10 of Section III).

Finally, the wind measurements are often not the final objective but used to infer some consequential phenomena such as energy distribution [12] and water circulation in a lake [3]. We need to consider the water quality modeling during designing the optimal sensor placement scheme since the winds of different locations impose variant impact on water quality of a whole lake due to diverse morphometries and flow patterns. Although much effort is made to reduce the spatial prediction errors as small as possible, the wind distributions obtained by estimating the observations through the readings of limited deployed sensors will inevitably contain some errors. Therefore, we intend to provide direct measurements by deploying sensors at locations with high impact on the final water quality studies and eliminate the prediction inaccuracy for the other unobserved locations maximally.

### B. Approach Overview

We propose a novel approach to address the sensor placement and spatial prediction problem by considering the unique features of wind measurement applications. Fig. 2 illustrates the main framework in steps. First, we possess the historical wind data from the two meteorological stations at Marina Bay (2007-2008) and Marina Channel (2007-2008 and 2011-2013). We find clear difference of the dominant wind directions between different time periods, which is consistent with the monsoon climate in Singapore[1]. We develop a time series segmentation method and divide the sensor data of the whole year into two monsoon seasons and two intermonsoon seasons. In each segment, the wind at one location can thus be modeled or transformed to a Gaussian distribution, and the optimal sensor locations are selected according to certain information criteria, e.g., entropy in this work. The results of all seasons are combined to calculate the optimal sensor placement scheme in a whole year.

We incorporate CFD modeling to simulate the wind distributions above Marina Reservoir based on 3D geographical information. CFD modeling can capture the detailed impact of surrounding high-rise buildings to the wind distribution by numerically solving the classic formulas of fluid mechanics [13]. In this study, we perform offline CFD simulations to generate coarse wind distributions at different conditions and learn the correlations in the field rather than obtaining the final wind distribution in real time since CFD modeling is computational complex and time-consuming.

To further consider the water quality sensitivity before deploying any wind sensors, we conduct a series of ELCOM-CAEDYM simulations to quantify the sensitivity of water

[1]Singapore has two monsoon seasons every year, Northeast (NE, roughly Dec.-Mar.) and Southwest (SW, roughly Jun.-Sep.). The name indicates their dominant wind direction. The monsoon seasons are separated by two intermonsoon periods, PreSW and PreNE, in which the wind is more evenly distributed.

quality to the wind input at different locations over Marina Reservoir. We then adjust accordingly the sensor placement scheme to factor the non-uniform accuracy requirement in wind measurement.

Once we have obtained the optimal sensor locations, we deploy a certain number of wind sensors at the most critical locations. Based on a wireless data collection system, the technical details of which are beyond the scope of this paper, we retrieve the real time sensor readings from our server. Finally, we use an online clustering algorithm to dynamically identify the transitional point between different monsoon and intermonsoon seasons with instant wind measurements. Different spatial prediction parameters are applied in the identified seasons with real sensor readings.

## III. Wind Measurement Approach

In this section, we present the detailed development procedure of our approach for wind measurements, including monsoon based time series segmentation, data set generation based on CFD modeling, optimized sensor placement and spatial prediction.

### A. Monsoon based Time Series Segmentation

Fig. 3a presents the histograms of wind direction and speed in the whole year of 2007 drawn by the data of meteorological station at Marina channel. From the historical data, we see that two obvious peaks in the density of wind directions corresponds to the two monsoon seasons in one year of Singapore, which are caused by the seasonal changes in global atmospheric circulation upon asymmetric heating of land and sea [14]. In each monsoon season, the wind is mainly from a dominant direction. It has been found from historical wind data of multiple years [15] that the wind directions of the two monsoon seasons are strongly Gaussian and the wind during the intermonsoon seasons is weak and more evenly distributed over all directions. The distribution of the whole year is the sum of all segments, exhibiting a mixture model. In this section, we introduce our monsoon based time series segmentation such that a whole year is segmented into different monsoon or intermonsoon seasons that follow different GP model.

*1) Time Series Segmentation Algorithm:* The traditional monsoon division scheme based on experience only provides month level granularity. The start and end of a monsoon season may largely vary at different years. We thus need an accurate segmentation scheme to find the critical changing time points for monsoon season transitions.

The objective is to find four critical change points to make the wind directions in the monsoon seasons the most follow a Gaussian distribution and the wind directions in the intermonsoon seasons the most follow a uniform distribution. We use Maximum Likelihood (ML) method to find the optimal time points that separate the one-year data from an meteorological station into four segments including $M$, $N$, $K$ and $J$ samples respectively, which maximize the likelihood function of the mixture model (two Gaussian and two uniform).

$$\mathcal{L}(\mu_1, \sigma_1, \theta_1, \mu_2, \sigma_2, \theta_2 | x_1, x_2, \ldots, x_{M+N+K+J})$$
$$= \prod_{i=1}^{M} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{1}{2\sigma_1^2}(x_i - \mu_1)^2\right] * \left[\frac{1}{\theta_1}\right]^N \quad (3)$$
$$* \prod_{i=1+M+N}^{K+M+N} \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left[-\frac{1}{2\sigma_2^2}(x_i - \mu_2)^2\right] * \left[\frac{1}{\theta_2}\right]^J$$

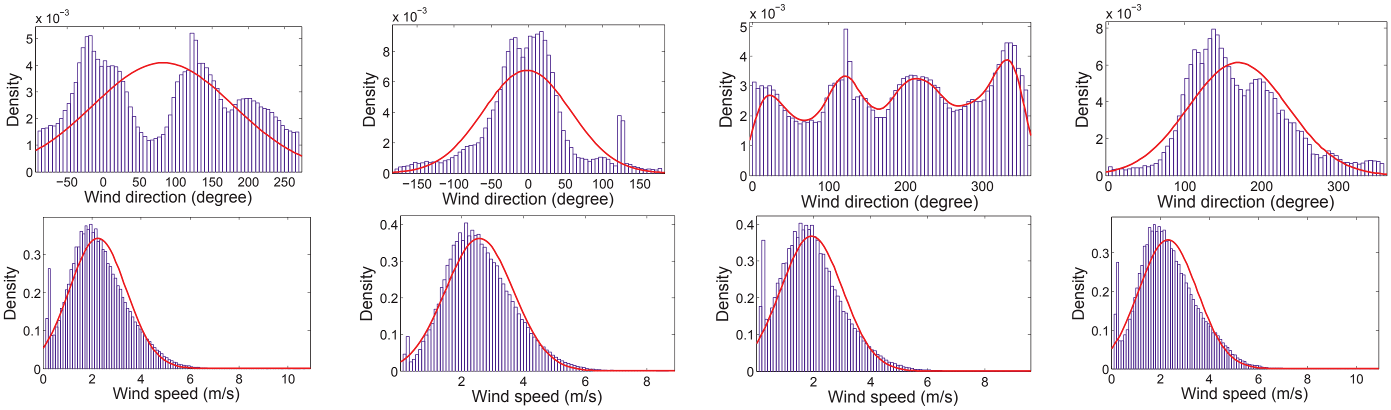
where $\mu_1 = (1/M)\sum_{i=1}^{M} x_i$ and $\sigma_1 = (1/(M-1))\sum_{i=1}^{M}$ $(x_i - \mu_1)^2$ are the unbiased estimation of parameters in the first Gaussian distribution including $M$ samples, and $(1+1/N)max(x_{M<i\leq N+M})$ and $(1+1/J)max($ $x_{M+N+K<i\leq N+M+K+J})$ are the unbiased estimation of parameters ($\theta_1$ and $\theta_2$) in the two uniform distributions.

The computation complexity to solve Eq. (3) is $\mathcal{O}(n^3)$ where $n$ is the search space for each time point. Since we know the approximate start and end of each monsoon season, we can restrict the search space. Algorithm 1 presents the ML-based time series segmentation algorithm searching the optimal time points heuristically. We can obtain the same results with the method searching in the whole data set exhaustively, but with much less computation. If we search in two months span centered at the experience based time point which starts with the first day of relative transitional month (e.g., April 1st for the transition from NE monsoon season to PreSW monsoon season), it takes less than 1 hour to converge.

**Algorithm 1** Heuristic ML-based time series segmentation

**Input**: One year wind data.
**Output**: Time points, $t_1$, $t_2$, $t_3$ and $t_4$.
**Initialization**: $t_{1,old} = Mar.15$; $t_{2,old} = t_2 = Jun.1$; $t_{3,old} = Oct.1$; $t_{4,old} = t_4 = Dec.1$;
**Step 0**: concatenate the start and end of data, so that the last NE part is merged to the first NE part;
**Step 1**: In ($t_4$, $t_2$), search $t_1$ in a Gaussian/Uniform mixture model by an equation similar to Eq. (3);
**Step 2**: In ($t_2$, $t_4$), search $t_3$ as **Step 1**;
**Step 3**: Based on the updated $t_1$ and $t_3$, search $t_2$ in ($t_1$, $t_3$) and search $t_4$ in ($t_3$, $t_1$);
**if** $(t_1 \neq t_{1,old} || t_2 \neq t_{2,old} || t_3 \neq t_{3,old} || t_4 \neq t_{4,old})$ **then**
    $t_{1,old} = t_1; t_{2,old} = t_2; t_{3,old} = t_3; t_{4,old} = t_4;$
    go to **Step 1**;
**else**
    **return** $t_1$, $t_2$, $t_3$ and $t_4$;
**end if**

Fig. 3 presents the decomposed monsoon seasons for year 2007. Two intermonsoon seasons are combined together since they present the same pattern. We see that the wind direction in each individual season is well fitted by a Gaussian or uniform model. Fig. 3 also shows that the wind speed of each season can be perfectly modeled as a Gaussian distribution. It is because the wind speed of the whole year is also Gaussian distributed. Therefore, we mainly focus on the segmentation of wind direction. The wind speed will automatically follow a Gaussian distribution processed according to the results of wind direction.




(a) The whole year (b) NE monsoon (c) Inter monsoon (d) SW monsoon

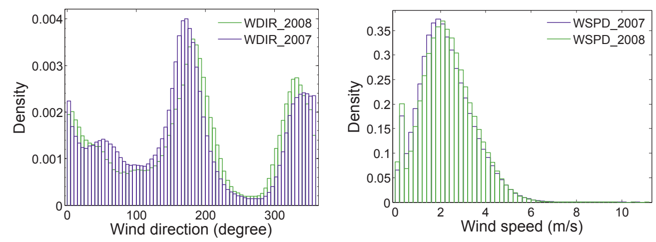
Fig. 3: Density of wind direction and speed over the year 2007 and its decomposed monsoon seasons. The red line is normal or uniform fitting curve for relative distributions.

*2) Segmentation Result Analysis:* Likelihood ratio, $D = 2(\ln \mathcal{L}_{new} - \ln \mathcal{L}_{null})$, is normally used to compare the fitness of two models. It expresses how many times more likely the data are under one model than the other. The likelihood ratio $D$ between the mixture model derived by the proposed segmentation algorithm ($\mathcal{L}_{new}$) and the uni-Gaussian model ($\mathcal{L}_{null}$) used by traditional sensor placement approach is 16.8k. The winds divided by the proposed segmentation algorithm are modeled obviously better than the uni-Gaussian model. We also divide the winds by the fixed division scheme based on experience. The likelihood ratio between this experience based mixture model and the uni-Gaussian model is 13.8k, which also shows that the division scheme derived by our segmentation algorithm can better fit the winds into proper statistical models.

*3) Application to Wind Measurements:* To apply the segmentation results generated with the historical data of one meteorological station over one year, we need to answer two questions. Do the other locations in the target area hold the same segmentation scheme? Can the segmentation scheme generated by one year historical data be applied to other years or even to the current year?

All locations in Marina Reservoir area share the same monsoon division scheme. The segmentation derived from the historical data of Marina Channel meteorological station can be applied to other locations, since it is based on a general environmental phenomenon which is consistent across the region. The monsoon climates are caused by the seasonal changes in global atmospheric circulation due to the asymmetric heating of land and sea [14]. Compared with the large scale atmospheric circulation, Marina Reservoir is small in size and therefore all locations are dominated by the same monsoon pattern. For example, we study the historical data (2007) from both observatory sites at Marina Bay and Marina Channel and find the exactly same segmentation results (Dec. 1 to Mar. 15 for NE monsoon season and Jun. 9 to Sep. 15 for SW).

It has also been proven based on historical wind data [15] that the main directions of winds in monsoon seasons are stable for different years and the winds in intermonsoon seasons are




(a) Direction (b) Speed

Fig. 4: Wind Density for two consecutive years (2007 and 2008) collected at Marina Bay meteorological station.

evenly distributed. Fig. 4 presents the wind statistics from Marina Bay meteorological station for two consecutive years (2007-2008) which suggest highly similar distributions. As we find later, the derived parameters of GP models based on the data from different years are very close to each other for the same seasons. The most likely variation from year to year is the variance of Gaussian distribution for monsoon seasons. Such small fluctuation is easily flattened by looking at multi-year wind data. For the wind data of multiple years (e.g., 2007-2008 data from Marina Channel meteorological station), the division is performed for each year respectively and the relative parts of different years are combined together to derive the best Gaussian fitting. Beside wind direction, according to Fig. 4, the wind speeds are stable for different years and always follow Gaussian.

### B. CFD Modeling

We apply CFD modeling to obtain simulated surface wind distributions above Marina Reservoir for different wind directions above the atmospheric boundary layer. CFD studies the physical aspects of fluid flows by algebraically solving the fundamental governing equations like continuity and momentum conservation. Numerical results are finally obtained at discrete points in time and space. The CFD modeling of wind distribution takes two inputs, atmospheric flow and topography

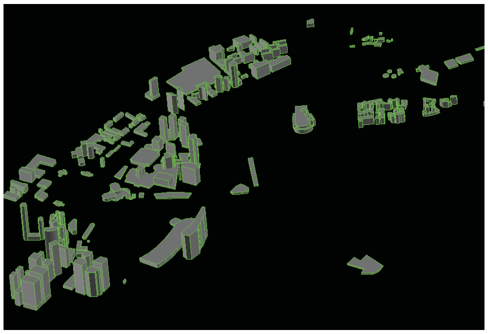


Fig. 5: 3D CFD model of geographical information around Marina Reservoir

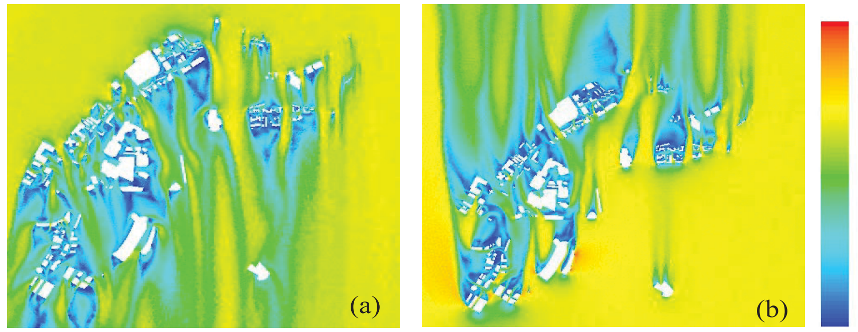



Fig. 6: CFD modeling results of wind velocity (m/s) distribution at 1.5 meter height with an incoming atmospheric flow from North (a) and South (b)

information of the land surface. On one hand, since Marina Reservoir is relatively small in size compared with the large scale atmospheric circulation, the atmospheric motion above this area can be treated as uniform. The atmospheric flow is therefore a vector comprising of a dominant wind direction and speed. On the other hand, a three dimensional CFD model of Marina Reservoir area is developed based on the geographical information that contains building locations, building shapes as well as heights. Fig. 5 depicts the built 3D CFD model of topography around Marina Reservoir, which models all buildings located within 3 blocks from the reservoir offshore. The computational domain is 3.5km long, 2.5km wide and 0.8km high. The number of computational cells used for each simulation is approximately 40 million.

The commercial CFD software FLUENT 13.0 is used to calculate the surface wind distribution over Marina Reservoir area. To capture the turbulent nature of the flow around buildings, the popular k-$\epsilon$ turbulence model is chosen because of its high computational efficiency [13]. Standard and second-order discretization schemes are adapted for pressure interpolation. It takes almost 2 days for one simulation case to be converged using a workstation of 12 cores (running 8 parallel-Fluent licenses) and 32GB memory.

The CFD modeling results cannot provide accurate instant wind distribution due to the following two limitations. First, CFD requires the real time and accurate atmospheric circulation data as input to derive instant wind distribution, which is difficult to obtain. Second, CFD simulation is computationally complex and time-consuming, which makes the instant CFD computation impossible. To capture the main characteristics of all possible wind distributions over the water surface, we run many simulations with different atmospheric flow inputs. 16-point compass rose is used to categorize the incoming atmospheric flows into 16 directions evenly spanning $0°$ to $360°$. For each direction, we run 10 gradually increasing speeds to explore all possible atmospheric motion velocities ($0 \sim 9$m/s) in Singapore. By doing this, we obtain a data set of 160 independent surface wind distributions for the underlying area. Two examples are given in Fig. 6, with an incoming atmospheric flow from north and south respectively. We can see that the surface wind distributions have distinctive patterns for different incoming flows due to the influence of surrounding architectures.

For all wind distribution results of CFD simulations, the wind direction and speed at the location of Marina Channel meteorological station is one-to-one mapped to the incoming atmospheric flow, because Marina Channel is in a relatively free space. In the data set, we extract a wind vector at the location of Marina Channel meteorological station from each of the 160 CFD wind distributions. At the same time, these vectors divide the historical wind data of Marina Channel meteorological station for the year 2007 into 160 segments. The occurrence frequency of each wind distribution in the whole year or in each monsoon or intermonsoon season can thus be computed. Based on this information and all 160 CFD wind distributions, we can derive the GP model over all locations of Marina Reservoir for each season. The key parameters of the GP models (e.g., mean vector and covariance matrix) used in the consequential sensor placement and spatial prediction are thus obtained. We will show in Section IV that the derived GP models are fine enough to provide high prediction accuracy.

### C. Sensor Placement

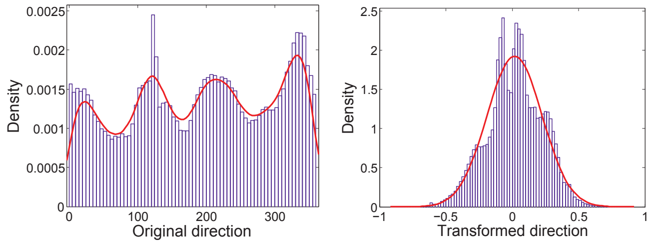
Once we divided one year into different monsoon seasons and obtained the prior knowledge on the wind distributions in each segment, we can learn the spatial correlation between any two locations over the target area and find the optimal sensor locations in each segment. For intermonsoon seasons, we need to transform the uniform wind direction distribution to a Gaussian distribution. Finally, a permanent sensor placement scheme can be obtained by combining the results of all segments and considering the water quality sensitivity.

*1) Sensor Placement for Single Monsoon Season:* With the data set of CFD modeling, we obtain a GP of wind for each season. It is NP-hard to select optimal sensor locations for predicting the mean, maximum or minimum of other locations [8]. Two widely used criteria to guide the sensor placement are entropy [7] and mutual information [9]. For entropy, the optimal sensor locations form a set which can provide the largest joint entropy.

$$\arg\max_{\mathcal{A}:|\mathcal{A}|=K} H(\mathcal{A}) \tag{4}$$

$$H(\mathcal{A}) = H(X_{a_k|a_{k-1},\ldots,a_1}) + \cdots + H(X_{a_2|a_1}) + H(X_{a_1})$$

Heuristic algorithms can be used to find the locations with largest entropy or conditional entropy iteratively. The selected locations provide the best prediction of observations




(a) Uniform (b) Gaussian

Fig. 7: Transformation of uniform distribution in intermonsoon season of year 2007 at Marina Channel meteorological station to Gaussian distribution.

at unobserved locations. For each location $v$, we treat the wind direction and speed as a random variable vector. Its entropy is calculated as $H(X_v) = \frac{1}{2}log|2\pi e\sum_{v,v}|$, where $\sum_{v,v}$ is the covariance matrix of direction and speed at $v$.

The entropy criterion finds the most informative locations which are located far away from each other. An alternative [9] searches for locations that most significantly reduce the uncertainty of rest space through maximizing the mutual information between the selected locations and the rest, presented as $MI(\mathcal{V}\backslash\mathcal{A}, \mathcal{A}) = H(\mathcal{V}\backslash\mathcal{A}) - H(\mathcal{V}\backslash\mathcal{A}|\mathcal{A})$.

*2) Transformation of Uniform Distribution:* The uniform distribution of winds in intermonsoon seasons can be transformed to a Gaussian distribution using Inverse Transform Sampling (ITS). If $X$ is uniformly distributed on $[0, 1]$ and $F(y_i) = x_i$, the random variable $Y$ is drawn from a normal distribution described by its cumulative function $F = 1/2 + 1/2erf(x/\sqrt{2})$. Fig. 7 depicts that the transformed data by ITS can be fitted by a Gaussian distribution. The advantage of ITS is that it supports bidirectional transformation. We can transform the wind data to Gaussian distribution to study the sensor placement and spatial prediction, and convert the estimated values of unobserved locations back to normal readings after processing.

*3) Sensor Placement for the Whole Year:* The sets of sensor locations found for different monsoon or intermonsoon seasons are not exactly the same. Ideally, we deploy sensors in one season according to its optimal placement scheme and move the sensors according to another optimal placement scheme when the next season starts. However, in reality, we cannot do that due to high reinstallation cost in terms of both finance and time. We resort to provide a suboptimal solution to find a best balance among different seasons.

We consider the above problem while calculating the entropy of each location. Assuming the entropy of location $v$ in $j$th time segment is $H(X_{v,j})$, the entropy of that location for all segments can be computed as:

$$H(X_v) = \textstyle\sum_{j=1}^{3} w_j * H(X_{v,j}) \tag{5}$$

where $w_j$ is the weight of $j$th time segment in the entire time series totally including two monsoon seasons and one combined intermonsoon season. From the viewpoint of information theory, by doing this, the information utility of each location is the sum of its entropies in each time segment weighted by the relative proportion. Once the location of the highest entropy is found, we search for the second and consequential locations by calculating the weighted conditional entropy until the maximum number of sensors we can deploy. The optimal sensor placement scheme for one year is also the best solution for multiple years, as the wind pattern simply repeats with negligible changes for different years.

*4) Sensitivity of water quality:* To consider water quality during the design of the optimal wind sensor placement scheme, a sensitivity analysis is conducted to find the relative influence of wind at each location on the water quality in Marina Reservoir. We first run the water quality simulation with uniform wind distribution of whole area and then repeat that simulation by doubling the wind speed at one location. We record the differences of all water quality parameters at each location between the two simulations. Fig. 8 depicts the obvious differences of chlorophyll distributions for two scenarios. The chlorophyll sensitivity to the wind at location $v$ is calculated as:

$$S_v = \sum_{j=1}^{N} \left| \frac{CHL_j^v - CHL_j}{CHL_j} \right| \tag{6}$$

where $N$ is the number of possible sensor locations and $CHL_j^v$ is the chlorophyll value of $j$th location when the wind speed at location $v$ is doubled. The sensitivity of water quality is the average of all water quality parameters including chlorophyll, temperature and dissolved oxygen. We find the water quality sensitivity to the wind at each location by repeating the experiments with doubled wind speed at that location. Fig. 9 shows that the sensitivities of water quality at different locations are significantly distinct.

We factor the sensitivity analysis in calculating sensor placement by adjusting the information utility of each location with its normalized sensitivity.

$$H'(X_v) = S_v * H(X_v) \tag{7}$$

We normalize the raw sensitivity of each location by the highest sensitivity over Marina Reservoir which is at location (1750, 2250). From the viewpoint of information theory, the information utility is the quantity one location can offer to eliminate the uncertainty of wind distribution of the whole reservoir. We reduce the information utility at one location if it has small impact on water quality. By doing this, the locations with high sensitivity will have more chance to be selected and wind sensors are deployed at the selected locations which provide direct measurements with minimal error. The final studies over water quality will be benefitted from these wind fields with intended error distribution.

After the final adjustment to the sensor placement scheme according to sensitivity analysis, we calculate the entropy for all locations and obtain the final sensor placement with locations of the highest entropy or conditional entropy. Due to topography and regulatory constraints, we cannot install wind sensors at all desired locations. The allowed area we may finally get permit to deploy sensors are depicted in Fig. 10. We therefore choose the first location only if it has the highest entropy and is available to deploy sensor. If the location of the highest entropy is not permitted for sensor deployment, we turn to the next location with the highest entropy. We repeat this

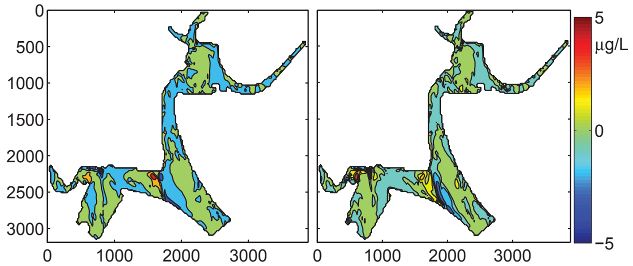



Fig. 8: The chlorophyll distribution differences between uniform wind and speed doubled at location (1750, 2250) and (1850, 2250)

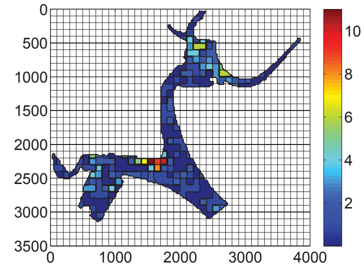


Fig. 9: Water quality sensitivity to wind at different locations in Marina Reservoir

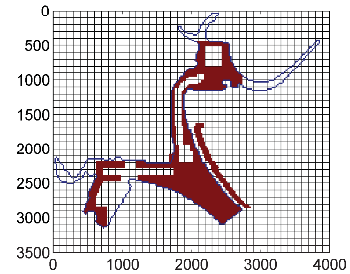


Fig. 10: The red region shows the allowed area for installing wind sensors.

procedure until enough sensor locations are found. The number of sensors to deploy is constrained to the project budget and the prediction accuracy. In this study, we finally deployed 10 wind sensors which can provide acceptable prediction accuracy. The deployment layout of wind sensors is given in Section IV.

### *D. Spatial Prediction*

We predict the observations at the unobserved locations as the mean of conditional distribution $X_{\mathcal{V}\backslash\mathcal{A}}|X_{\mathcal{A}}$ in Eq. (1), with the decomposed Gaussian models and the input of real-time sensor readings. One problem is to determine which GP model should be used to perform the prediction. Because the start and end of monsoon seasons are variable for different years, we cannot cluster the sensor readings according to a fixed division scheme derived using the time series segmentation algorithm in Section III-A1.

An online temporal clustering algorithm is developed to dynamically search for the critical change point of monsoon seasons. When a new sensor reading is received, the likelihood of last $N$ samples is calculated using the statistical model (Gaussian or Uniform distribution) of current monsoon season. When the likelihood decreases to a user-defined threshold, $\tau$, we infer that the transition of monsoon seasons occurs.

When a set of sensor readings measured by all deployed wind sensors at a given time point is categorized to a certain monsoon season, the relative GP model can then be applied to estimate the wind field on other unobserved locations using Eq. (1).

## IV. Deployment and Evaluation

In this section, we introduce the in-field deployment of wireless wind sensor network in Marina Reservoir area and evaluate the performance of the proposed sensor placement and spatial prediction approaches with real measurement results.

In summary, the measurement results show that compared with the traditional single Gaussian sensor placement method (presented as UniGau) and the linear interpolation, the proposed approach improves the wind direction prediction accuracy by 81% and 26% respectively. For wind speed, the performance of our approach and UniGau is comparable and they both improve the performance of Interpolcation by 26%.

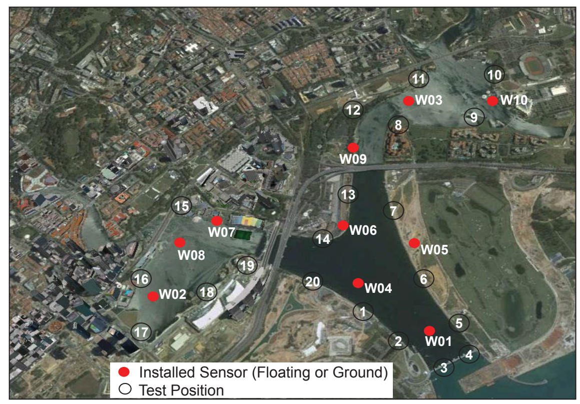



Fig. 11: Locations of deployed wind sensors and the test positions of the mobile sensor.

### *A. Deployment of Wireless Sensor Network*

The potential deployment area covers a water surface space of $2.2km^2$ plus the terrain space within 100m from the water's edge since some locations on land may provide more information than those on water surface to infer the wind observations on other locations. We divide the underlying area of Marina Reservoir into small grids of $20m * 20m$ which provides finest resolution. More than 5k locations need to be considered.

10 wind sensors are finally deployed, as marked by the red dots in Fig.11, including 5 land sensors installed on the ground around the water and 5 floating sensors on the water surface. The locations are selected according to the proposed approach based on the historical data of the meteorological station on Marina Channel over two years from 2007 to 2008. Due to high computational complexity of calculating mutual information over the large set of potential locations, entropy is used as the criterion for selecting the optimal sensor locations.

For the wind sensor, the wind monitor model 05305L of R.M. YOUNG is used. It provides an accuracy of 0.2m/s for speed and 3° for direction. With the current design, all wind sensors are equipped with a RTCU DX4 data logger and GSM communication module. The minutely measured data is first logged and then transmitted back to our backend server directly through cellular network. The real time data is then hosted in the server and can be accessed online. Accurate clock is

(a) Land (W10) (b) Floating (W2) (c) Mobile (6)

Fig. 12: Three types of wind sensors.

provided in the data logger and the readings of all wind sensors are instantly synchronized in-field. A solar panel is equipped on each wind sensor to provide continuous power to the wind anemometer and data logger.

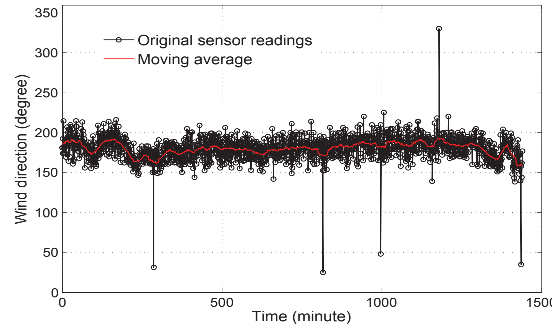
Fig. 12 depicts the wind sensors that we construct in this study. A land based sensor as depicted in Fig. 12a is fixed on the ground with an absolute reference direction. A floating sensor (Fig. 12b) is anchored to the bottom of the water but floats on the water surface. It has limited rotational freedom. We add a compass of high accuracy for each floating sensor to determine the instant reference direction which will be used to calculate the absolute wind direction by offsetting the raw measurement. We also build a mobile wind sensor, as depicted in Fig. 12c. It can be easily moved and set up temporarily at an arbitrary location. We use the mobile sensor to collect wind data for performance evaluation. Instead of solar panel, a portable battery is used to provide energy. All the other components are the same as other permanent wind sensors.

### *B. Experiment Setup*

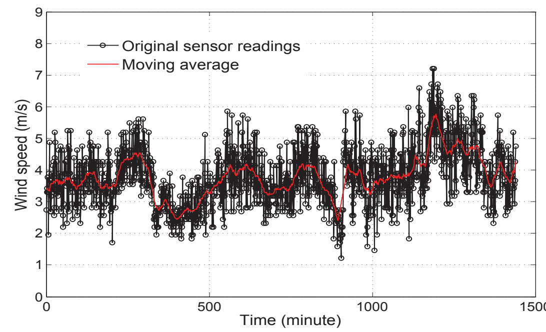
We evaluate the performance of the proposed approach by real measurements. With the deployed wireless wind sensor network, we have collected the wind data since July 2013. We study the accuracy of spatial prediction with reference of UniGau and linear interpolation. The latter method is widely used by current environmental analysis.

The performance gain of our proposed approach comes from two aspects: optimal sensor placement and accurate spatial prediction. Spatial prediction is based on sensor placement and they share the same system model. Since the advantages of Gaussian-based sensor placement over random deployment have completely proved in previous works [5, 6] and it is costly in terms of budget and time (more than 3 months) to reinstall the deployed wind sensors, we focus on evaluating the potential improvement of spatial prediction accuracy by comparing our approach (MIX) with UniGau and Interpolation based on the real wind measurements on the optimal sensor placement. The prediction error is measured by the average Root-Mean-Squared Error (RMSE) between the estimated values of unobserved locations $\hat{X}_{\mathcal{V}\setminus\mathcal{A}}$ and their actual values $X_{\mathcal{V}\setminus\mathcal{A}}$.

$$RMSE(X_{\mathcal{V}\setminus\mathcal{A}}|X_{\mathcal{A}}) = \frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{\sum_{i\in\mathcal{V}\setminus\mathcal{A}}^{N}(\hat{X}_i^t - X_i^t)^2}{N}} \quad (8)$$




(a) Direction




(b) Speed

Fig. 13: Time series of wind direction and speed from 00:00 to 23:29 on 1 September 2013

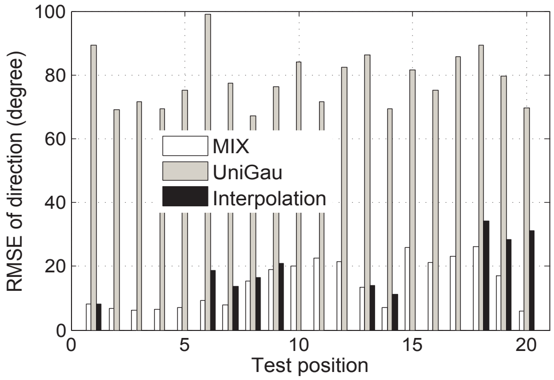
Assume we have $T$ sets of samples to conduct the evaluation and $N$ locations are included in $\mathcal{V}\setminus\mathcal{A}$. The result is the average error in both temporal and spatial aspects.

### *C. Measurement Results*

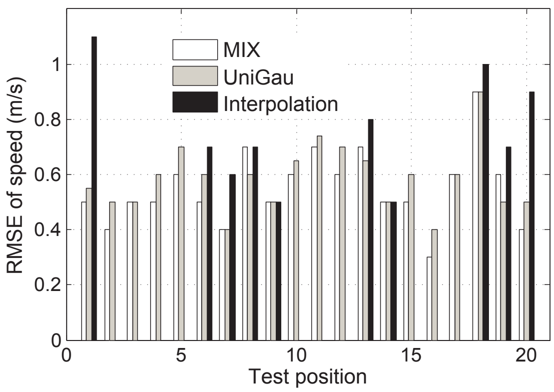
Fig. 13 presents a snapshot of the measured wind distribution and speed for 24 hours at wind sensor W01. The minutely measured data is plotted as the black line and the smoothed data with a 30 minute moving average window is depicted as the red line. The latter is required as a stable input to feed the ELCOM-CAEDYM model. We will evaluate the performance of the proposed approach and the benchmark methods using the moving average data.

**Overall performance in space.** To evaluate the spatial prediction accuracy at different locations, we measure wind direction and speed at 20 randomly selected locations along the water's edge of Marina Reservoir using the mobile wind sensor. The test positions are depicted in Fig. 11. At each location, minutely wind data is collected for 1$\sim$2 hours. Fig.14 presents the average RMSE of predicted direction and speed for each location. The default linear interpolation method implemented in matlab, *griddata*, is used. For some locations, we cannot perform the linear interpolation since they are located outside the effective region of this approach.

Compared with UniGau and interpolation, the proposed approach reduces average RMSE of wind direction prediction by 81% and 26% respectively. By the monsoon based time series segmentation, MIX can accurately model the wind and

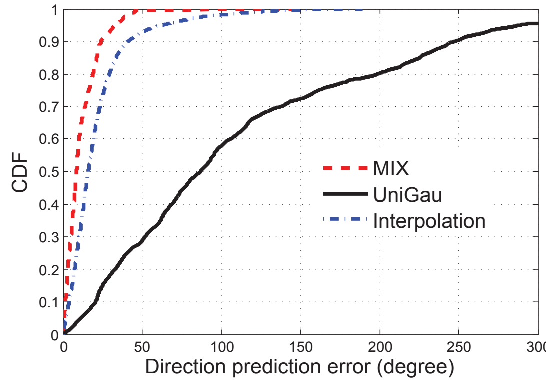



(a) Direction

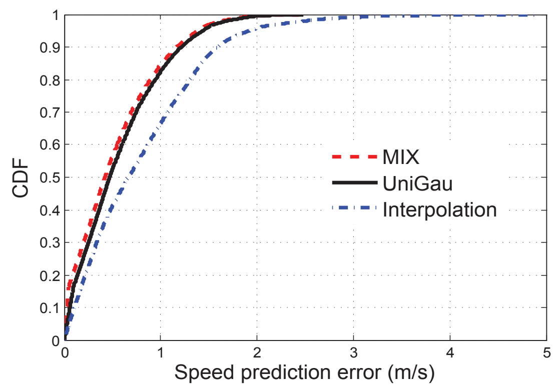



(b) Speed

Fig. 14: Average prediction RMSE of wind direction and speed for 20 test positions




(a) Direction




(b) Speed

Fig. 15: Prediction error of wind direction and speed validated by the installed wind sensors

provide high prediction accuracy. Because the wind direction distribution for the entire year is not Gaussian, UniGau gives large errors. The average RMSE of interpolation is relative large because it does not consider the effect of surrounding buildings to wind field and thus cannot accurately capture the spatial variation of wind distribution. For wind speed prediction, the performance of MIX and UniGau is comparable since the wind speed of whole year is still Gaussian. They reduce the average RMSE of Interpolation by 26%. From Fig.14, we can also see that the average RMSE of locations near installed sensors or in open space is relatively small, because the wind patterns can be better captured by the statistic models and CFD modeling.

**Overall performance in time.** To further investigate the performance of proposed approach for long term wind measurement, we use the measurement data of all 10 sensors for three months. At one time, we choose one sensor and use the measurement data from the rest 9 sensors to predict its wind direction and speed. We use its own measurement as the reference to calculate RMSE. We perform this evaluation in 10 rounds for all 10 sensors. We do not have data from W09 since it was missing a short time after installation. We are redeploying it to the opposite edge of the Kallang river which is more secure and provides the same level of information for the spatial prediction over the target area.

Fig.15 presents the cumulative distribution of the absolute difference between predicted observation and the measured value for each sample. In this case, $T$ and $N$ in Eq. (8) are both equal to one. Similar to the results of the mobile sensor testing, MIX improves the prediction accuracy of wind direction by 87% for UniGau and 27% for Interpolation, and the performances of MIX and UniGau for speed prediction are comparable and 21% higher than that of Interpolation. MIX offers an average spatial prediction accuracy of 24°. The error is larger than the mobile test since the data of one installed wind sensor is used as the reference for evaluation but not included in the calculation of spatial prediction.

**Sensitivity of water quality.** We take into account water quality while solving the sensor placement problem so as to obtain an intended error distribution in space. Fig.16 presents the average RMSE at each test position in the experiment with the mobile sensor corresponding to the sensitivity of water quality at that location. The results show that the average RMSE is relatively low at locations with high water quality sensitivity. The linear regression between the water quality sensitivities and the relative RMSEs of predicted direction for different locations reveals such an inverse trend.

**Online clustering algorithm.** We use historical data to evaluate the efficiency of our online clustering algorithm. The experiments are done using the historical wind data of

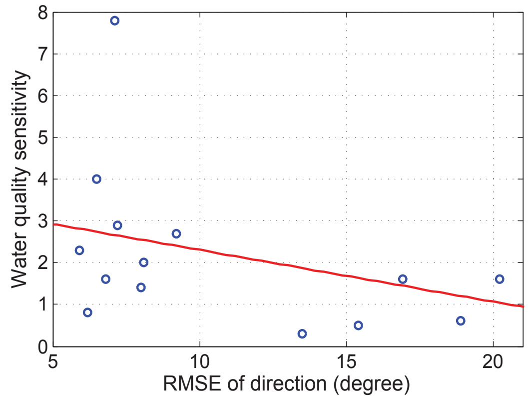



Fig. 16: The RMSE of direction prediction corresponding to the water quality sensitivity for each mobile test location. The red line is the linear fitting curve of the scatter points.

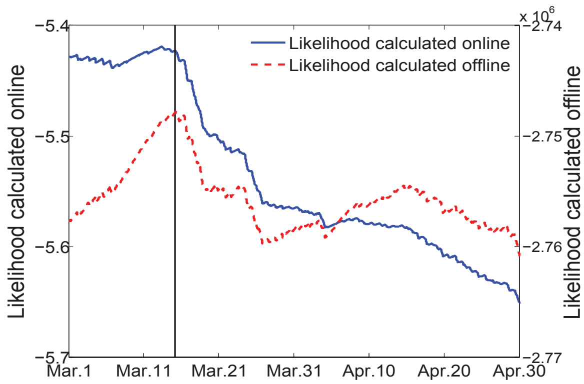



Fig. 17: Likelihood calculated online and offline.

Marina Channel meteorological station over the year 2008. The likelihood calculated by the online clustering algorithm is the average likelihood of all $N$ samples in the sliding window. Only the last $N$ samples before the time point under consideration are used in the calculation. We set the sliding window size $N$ to 1440 samples corresponding to one day and the likelihood threshold $\tau$ to 5.45. The likelihood calculated offline is obtained using the monsoon based time series segmentation algorithm introduced in Section III-A with the whole year data. It is the sum likelihood of all samples. Fig.17 shows that the likelihoods calculated online and offline peak almost at the same time point for the transition from NE monsoon season to PreSW intermonsoon season. The online temporal clustering algorithm can find the critical change point of time series with a small error of 3.6 days in this case.

## V. Lessons Learnt

The wind we measure in this study exhibits distinct non-Gaussian process from previously studied phenomena. Besides, no full prior-knowledge is possessed to model the spatial correlation in the field. To work with these challenges, we therefore developed a unique sensor placement approach including monsoon based time series segmentation and the data set generate using CFD modeling.

The proposed approach, on the other hand, demonstrates a complete procedure which extends the state-of-the-art sensor placement and spatial prediction methodology to solve a wider range of applications. There are many other phenomena which exhibit similar non-Gaussian process captured by a mixture model over time, e.g., city traffic flow [16], soil pollutant [17], etc. Generalizing the time series segmentation approach in such applications will significantly improve the accuracy of sensor placement and spatial prediction over clustered time periods.

Incorporating computational simulation, e.g., CFD in this work, to build a data set and study the field data correlation provides us a new thought to gain prior-knowledge when intrusive way of learning such knowledge is not preferred. We do not need to run the time-consuming computational simulations online, but just conduct enough simulations to cover most cases and capture the statistical features of the target phenomena. In many applications in which it is impractical to pre-deploy enough sensors due to various constraints, computational simulation procedure can provide coarse yet sufficient knowledge to statistically capture the spatial correlations that we need.

## VI. Related Works

The optimal sensor placement problem has been addressed in many previous works [18–20]. Among them, GP based approaches [8–10] have been used in many applications monitoring spatial phenomena like temperature [5] and soil moisture [6]. However, they cannot be directly applied to wind distribution measurement due to the temporal and spatial variations of wind.

The coverage problem of sensor networks has been extensively studied [21–23]. However, most of the existing theoretical works are based on the deterministic disc model. Data fusion is considered in [24, 25]. The coverage and connectivity in duty-cycled sensor network are analyzed in [26–28].

CFD modeling is a widely used tool to capture the fluid patterns in many applications, such as environmental engineering and aircraft design [13]. In [29], CFD has been successfully applied for geospatial risk assessment of wind channels in urban area with high accuracy. CFD modeling has also been used in sensor placement problem [30] and temperature forecasting [31] in data center environment. CFD models are built to capture extra hot spot scenarios. A thermal forecasting model is proposed in [32] to model and predict temperatures around servers in data center based on principles from thermodynamics and fluid mechanics.

The time series segmentation algorithms [33, 34] search for critical change points by iteratively dividing data into small segments with the same statistical model (e.g., Gaussian distribution). However, they cannot be applied for our mixture model of different statistic models, i.e., Gaussian and Uniform. Expectation Maximum algorithms [35] are utilized widely to divide a Gaussian mixture into individual Gaussian distributions. However, they cannot be used in the application of wind measurement either. First, the spatial correlation cannot be calculated since the samples of all locations at a given time point are not clustered in the same cluster. Second, the samples in the same cluster are not continuous in time. As a consequence, it is difficult to assign the online sensor readings to a proper cluster and apply the relative spatial prediction.

## VII. Conclusions

In this paper, we propose a novel sensor placement and spatial prediction approach for wind distribution measurements. It leverages the monsoon characteristics of wind to study its statistic properties. A data set is built using CFD modeling which captures the impact of surrounding buildings on wind distribution. Optimal sensor locations are selected through segmented wind statistical models and adjusted according to the sensitivity of water quality to wind at different locations. We deployed 10 wind sensors around or on the water surface of an urban reservoir. The observations of unobserved locations are predicted by the readings of deployed sensors clustered through an online algorithm. The in-field measurement results show that the proposed approach can significantly improve the accuracy of wind measurements.

We believe the proposed solution best exploits the underlying nature of wind measurement in our study and achieves optimal accuracy with a limited number of sensors. Such a procedure nevertheless can be generalized to handle other applications in observing phenomenon of similar nature.

## Acknowledgments

This work is supported by the Singapore National Research Foundation under its Environmental & Water Technologies Strategic Research Programmes and administered by the Environment & Water Industry Programme Office (EWI), under project 1002-IRIS-09, Singapore MOE AcRF Tier 2 MOE2012-T2-1-070, and the NTU Nanyang Assistant Professorship (NAP) grant M4080738.020.

## References

[1] R. Alexander and J. Imberger, "Spatial distribution of motile phytoplankton in a stratified reservoir: the physical controls on patch formation," *Journal of plankton research*, 2009.

[2] Z. Xing, D. A. Fong, E. Y.-M. Lo, and S. G. Monismith, "Thermal structure and variability of a shallow tropical reservoir," *Limnology and Oceanography*, 2014.

[3] B. Laval, J. Imberger, B. R. Hodges, and R. Stocker, "Modeling circulation in lakes: Spatial and temporal variations," *Limnology and Oceanography*, 2003.

[4] Z. Xing, C. Liu, L. H. C. Chua, B. He, and J. Imberger, "Impacts of variable wind forcing in urban reservoirs," in *the7th International Symposium on Environmental Hydraulics*, 2014.

[5] A. Krause, C. Guestrin, A. Gupta, and J. Kleinberg, "Near-optimal sensor placements: Maximizing information while minimizing communication cost," in *ACM/IEEE IPSN*, 2006.

[6] X. Wu, M. Liu, and Y. Wu, "In-situ soil moisture sensing: Optimal sensor placement and field estimation," *ACM Trans. Senor Networks*, 2012.

[7] N. Cressie, *Statistics for spatial data*. John Wiley and Sons, Inc, 1993.

[8] A. Das and D. Kempe, "Sensor selection for minimizing worst-case prediction error," in *ACM/IEEE IPSN*, 2008.

[9] C. Guestrin, A. Krause, and A. P. Singh, "Near-optimal sensor placements in gaussian processes," in *ICML*, 2005.

[10] M. A. Osborne, S. J. Roberts, A. Rogers, S. D. Ramchurn, and N. R. Jennings, "Towards real-time information processing of sensor network data using computationally efficient multi-output gaussian processes," in *ACM/IEEE IPSN*, 2008.

[11] A. Krause, E. Horvitz, A. Kansal, and F. Zhao, "Toward community sensing," in *ACM/IEEE IPSN*, 2008.

[12] T. Burton, N. Jenkins, D. Sharpe, and E. Bossanyi, *Wind energy handbook*. John Wiley & Sons, 2011.

[13] J. D. Anderson *et al.*, *Computational fluid dynamics: the basics with applications*. McGraw-Hill New York, 1995.

[14] K. E. Trenberth, D. P. Stepaniak, and J. M. Caron, "The global monsoon as seen through the divergent atmospheric circulation," *Journal of Climate*, 2000.

[15] L. S. Chia, A. Rahman, and D. B. H. Tay, *The biophysical environment of Singapore*. Singapore University Press, 1991.

[16] S. Sun, C. Zhang, and G. Yu, "A bayesian network approach to traffic flow forecasting," *IEEE Transactions on Intelligent Transportation Systems*, 2006.

[17] Y.-P. Lin, B.-Y. Cheng, G.-S. Shyu, and T.-K. Chang, "Combining a finite mixture distribution model with indicator kriging to delineate and map the spatial patterns of soil heavy metal pollution in chunghua county, central taiwan," *Elsevier Environmental Pollution*, 2010.

[18] S. S. Dhillon and K. Chakrabarty, "Sensor placement for effective coverage and surveillance in distributed sensor networks," in *IEEE WCNC*, 2003.

[19] F. Y. Lin and P.-L. Chiu, "A near-optimal sensor placement algorithm to achieve complete coverage-discrimination in sensor networks," *IEEE Communications Letters*, 2005.

[20] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Transactions on Signal Processing*, 2009.

[21] S. Kumar, T. H. Lai, and J. Balogh, "On k-coverage in a mostly sleeping sensor network," in *ACM MobiCom*, 2004.

[22] T. Yan, Y. Gu, T. He, and J. A. Stankovic, "Design and optimization of distributed sensing coverage in wireless sensor networks," *ACM Transactions on Embedded Computing Systems*, 2008.

[23] J. Chen, J. Li, S. He, T. He, Y. Gu, and Y. Sun, "On energy-efficient trap coverage in wireless sensor networks," *ACM Transactions on Sensor Networks*, 2013.

[24] G. Xing, C. Lu, R. Pless, and J. A. O'Sullivan, "Co-grid: an efficient coverage maintenance protocol for distributed sensor networks," in *ACM/IEEE IPSN*, 2004.

[25] G. Xing, R. Tan, B. Liu, J. Wang, X. Jia, and C.-W. Yi, "Data fusion improves the coverage of wireless sensor networks," in *ACM MobiCom*, 2009.

[26] X. Wang, G. Xing, Y. Zhang, C. Lu, R. Pless, and C. Gill, "Integrated coverage and connectivity configuration in wireless sensor networks," in *ACM SenSys*, 2003.

[27] Y. Gu, L. Cheng, J. Niu, T. He, and D. H. Du, "Achieving asymmetric sensing coverage for duty cycled wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, 2013.

[28] T. Yan, T. He, and J. A. Stankovic, "Differentiated surveillance for sensor networks," in *ACM SenSys*, 2003.

[29] T. K. Lim, H. Miao, C. Chew, K. K. Lee, and D. K. Raju, "Environmental modeling for geospatial risk assessment of wind channels in singapore," in *GSDI*, 2012.

[30] X. Wang, X. Wang, G. Xing, J. Chen, C.-X. Lin, and Y. Chen, "Towards optimal sensor placement for hot server detection in data centers," in *IEEE ICDCS*, 2011.

[31] J. Chen, R. Tan, Y. Wang, G. Xing, X. Wang, X. Wang, B. Punch, and D. Colbry, "A high-fidelity temperature distribution forecasting system for data centers," in *IEEE RTSS*, 2012.

[32] L. Li, C.-J. M. Liang, J. Liu, S. Nath, A. Terzis, and C. Faloutsos, "Thermocast: a cyber-physical forecasting model for data-centers." in *ACM KDD*, 2011.

[33] P. Bernaola-Galván, R. Román-Roldán, and J. L. Oliver, "Compositional segmentation and long-range fractal correlations in dna sequences," *Physical Review E*, 1996.

[34] V. Guralnik and J. Srivastava, "Event detection from time series data," in *ACM KDD*, 1999.

[35] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, 1977.