ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

An optimal sensor placement scheme for wind flow and pressure field monitoring

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:
Gao, H., Liu, J., Lin, P., Hu, G., Patruno, L., Xiao, Y., et al. (2023). An optimal sensor placement scheme for wind flow and pressure field monitoring. BUILDING AND ENVIRONMENT, 244, 1-19 [10.1016/j.buildenv.2023.110803].

Availability:
This version is available at: https://hdl.handle.net/11585/956751 since: 2024-07-16

Published:

DOI: http://doi.org/10.1016/j.buildenv.2023.110803

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/).
When citing, please refer to the published version.

(Article begins on next page)

08 November 2026

An Optimal Sensor Placement Scheme for Wind Flow and Pressure Field MonitoringAn Optimal Sensor Placement Scheme for Wind Flow and Pressure Field Monitoring

Huanxiang Gao[a], Junle Liu[a,b], Pengfei Lin[a,b], Gang Hu[a,c,*], Luca Patruno[d], Yiqing Xiao[a], K.T. Tse[b], K.C.S. Kwok[e]

[a]*School of Civil and Environmental Engineering, Harbin Institute of Technology, Shenzhen, 518055, China*
[b]*Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China*
[c]*Guangdong Provincial Key Laboratory of Intelligent and Resilient Structures for Civil Engineering, Harbin Institute of Technology, Shenzhen, 518055, China*
[d]*DICAM, University of Bologna, Viale Risorgimento 2, 40136, Bologna, Italy*
[e]*School of Civil Engineering, The University of Sydney, Sydney, NSW 2006, Australia*

Abstract

Sensing of wind pressures on wind-sensitive structures and wind speeds in densely populated cities is crucial to ensure structure safety and monitoring urban wind environment. However, these wind pressure or wind speed sensors are often sparsely distributed due to their high cost, and consequently the collected wind pressure or speed information is limited. Under these circumstances, it is particularly important to place these limited number of sensors in the optimal locations with the capacity of extracting the most information. In this study, an optimal sensor placement scheme based on multi-resolution dynamic mode decomposition (mrDMD), a space post-processing function: signed distance function (SDF) and particle swarm optimization-random forest (PSO-RF) is proposed, called mrDMD-SDF-PSO-RF (mrDSPR). The effectiveness of the proposed scheme is verified using two case studies of reconstruction based on sparse sensors: wind flow field of a city block and wind pressure field of a single building. The result indicates that the proposed scheme significantly improves the efficiency of the sensors. Therefore, based on the proposed optimal sensor placement scheme, the locations of placing sensors can be well-designed for various wind engineering problems, which builds a solid foundation of super-resolution reconstruction of wind pressure on buildings and wind flow field in urban environment.

Keywords: Multi-resolution Dynamic Mode Decomposition (mrDMD); Wind Environment; Sparse sensor; Sensor placement; PSO-RF

*Gang Hu, hugang@hit.edu.cn

1. Introduction

Three approaches including computational fluid dynamics (CFD) (Jiang et al., 2021; Wang et al., 2022), wind tunnel testing (Guo et al., 2015), and field measurement (Li et al., 2018a) have been often used to obtain wind flow and pressure field in wind engineering. Among them, field measurement is used for the real-time monitoring of wind pressures on wind-sensitive structures and wind flow field in densely populated cities. However, the number of sensors is often constrained by the high cost of sensors, complexity of terrain, and time consuming task for deploying sensors. How to place these limited number of sensors in the optimal locations in order to extract the most information of the wind pressure or speed field is a big challenge. Therefore, an optimal sensor placement scheme is desirable to guide sensor placement in field measurement in wind engineering. An optimized placement of sensors can achieve better performance for flow field reconstruction and prediction with lower cost and less effort compared to a random placement regime.

A number of studies have been devoted to find the optimal sensors placement locations, including the placement of sensors for flow field, temperature field or pollutant diffusion in the city. For instance, Kelp et al. (2022) introduced an optimal sensor placement method for air pollution estimation based on multi-resolution dynamic mode decomposition (mrDMD). It proposed an air quality sensor arrangement scheme for the entire U.S., which is more efficient than that determined by the Environmental Protection Agency (EPA). Liu & Li (2022a) and Liu & Li (2022b) conducted a series of studies on the monitoring of urban pollutants. They created an optimal sensor placement scheme based on the source term estimation (STE) of external pollution sources, which uses fewer sensors while considering a excellent STE performance. Malings et al. (2018) provided a sensor placement optimization scheme of monitoring urban temperature based on probabilistic model and vulnerability assessment. Their scheme is verified in the city of Pittsburgh, PA, and the result prove its efficiency. Du et al. (2015a) developed an optimal sensor placement scheme to measure the wind distribution over a large urban reservoir using a limited number of wind sensors. Their research is based on the measurement of wind that inherently exhibits strong non-Gaussian yearly distribution. By leveraging the local monsoon characteristics of wind, they segment a year into different monsoon seasons that follow a unique distribution respectively. Their results show that their scheme outperforms the state-of-the-art Gaussian model based on interpolation-based approaches. Papadopoulou et al. (2015) proposed a framework for optimal sensor placement in wind studies around buildings when knowledge is limited and no prior measurement data is available. The framework is built based on a multiple model falsification approach, and successfully used to improve simulation predictions of wind around buildings. Sousa et al. (2018) used wind data provided by sensors in a city to perform data assimilation and uncertainty analysis on CFD to improve the accuracy of CFD prediction of future wind fields. For maximizing the accuracy of CFD, they proposed the optimal arrangement of sensors on the roof of buildings under different wind directions. Based on POD and Linear Stochastic Estimation (LSE), Hu et al. (2022) selected the optimal sensor position and quickly estimated the high spatial resolution airflow distribution in the cubic building complex model, and explored in detail the impact of sensor number changes on the

airflow distribution estimation results. Fontanini et al. (2016) presented a method based on the finite dimensional Perron-Frobenius (PF) concept, which naturally considers sensor accuracy, location constraints, and sensing constraints. In addition, it can also provide the response times, sensor coverage maps, and the number of sensors needed. Besides, Sharma et al. (2019) added Kalman filter on the basis of PF method and achieved excellent results. Zhang et al. (2019) divided Xi'an into several areas based on the similarity of the urban layout of Xi'an, China, and arranged the appropriate amount of sensors in each area. However, these methods cannot handle problems with multiple wind attack angles, require a significant amount of computing resources or rely entirely on data-driven while ignoring the important physical characteristics of the wind itself, such as modes and implicit formulas. Therefore, although these methods perform well in their respective problems, they still suffer from insufficient generalization ability for wind engineering applications.

There are some studies focus on optimal flow field sensor placement for single object. For example, Semaan (2017) constructed a machine learning-based optimal sensors placement scheme for airwing, which is very easy to use and yields optimal results for multiple wind attack angles. Verma et al. (2020) studied the distribution of natural aquatic organisms used to sense the flow field in detail, and has proven the optimality of the distribution of these natural aquatic organisms, which has important guiding significance for the effective arrangement of sensors to monitor a single object in a flow field. Inoba et al. (2022) proposed a method for estimating the wind direction against a simple automobile model and reconstructing surface pressure distribution on it by the sparse pressure sensors, and the sparse pressure sensor placement was optimized by the three algorithms (Orthogonal matching pursuit, D-optimality-based greedy algorithm for vector-measurement sensors, and OMP-DG Hybrid algorithm) based on data-driven sparse sampling using the time-averaged surface pressure distributions as the training data. Their results indicate that the hybrid algorithm with four sensors is the most effective for both yaw angle estimation and pressure distribution reconstruction. Moreover, Sajedi & Liang (2021) proposed deep generative Bayesian optimization (DGBO) as a solution for a parallel optimization of black-box/expensive optimal sensor placement objective functions, which is able to effectively reduce the number of sensor on a single building for decreasing the measurement cost. However, models built for a single building can often use all positions on the building surface as candidate points, without considering its impact on the building function. It is challenging to apply these methods to attain optimal placement of urban sensors. It is not feasible to place sensors in certain areas, such as roads, to maintain the regular activities of the city. Additionally, the models created for a single building may not consider the obstacles in an urban setting.

There have been a lot of research on the optimal sensor arrangement of indoor temperature and pollutant detection. Chen & Gorlé (2022) used a polynomial chaos expansion method to quantify the effect of the variability in the boundary and initial conditions on the predicted buoyancy-driven flow and indoor temperature field. Based on the analysis results, they can determine the optimal indoor temperature sensor positions that provide robust monitoring of the volume-averaged time series temperature and the temperature range. Chen & Wen (2008) proposed a genetic algorithm-based sensor arrangement scheme for in-

door pollutant detection. A small commercial building was studied as a case. The research results show that this scheme can achieve the best system behavior and reduce the total system cost. Kyriacou et al. (2019) built a MATLAB toolbox called COMOB (contaminat monitoring in buildings) for positioning the contaminant sensors inside the building, such that the harmful effects due to a possible airborne contaminant event are minimized. Moreover, a system for detecting and isolating contaminant events in the building interior is also provided. Liu & Zhai (2009) designed a high-performance indoor air quality (IAQ) sensor network based on the adjoint probability method to identify potential contaminant source locations for an entire building based on limited sensor outputs, and the effectiveness of their method is verified by a case study. Sreedharan et al. (2011) combined various types of sensor information through a Bayesian model to improve the acquisition of indoor pollutant release characteristics. Real tracer gas data from a three-story building experiment were used to validate their method, demonstrating the benefit of multi-sensor information fusion. Eliades et al. (2013) proposed a method to determine the optimal sensor placement for monitoring a set of representative indoor contamination scenarios with minimum sensor costs. Two cases were used to prove the effectiveness of their method. Du et al. (2015b) integrated the building energy simulation (BES) and computational fluid dynamics (CFD) to optimize the indoor temperature sensor placement, which considers both of energy consumption and predicted mean vote (PMV). Cheng et al. (2022) proposed a method to optimize indoor temperature and CO2 sensor placement, which is supported by building information modeling (BIM) and considers different seasons, mechanical ventilation and air conditioning (MVAC) system settings and occupancy scenarios. These studies have played a very important role in the detection of indoor environments. However, compared with outdoor wind field, the indoor wind field is relatively simple. Specifically, the direction of wind flow is relatively fixed (wind can only enter from known windows), the wind speed is relatively small, and the location of indoor functional areas can be coordinated with the optimal sensor locations, thus almost all positions can be used as sensor candidate positions. Additionally, optimal indoor sensor placement becomes difficult to adapt for outdoor complex problems when the indoor geometry is simplified or requires excessive computing power, due to the utilization of overly intricate geometric models in indoor research. These characteristics make it difficult for indoor sensor optimization schemes to deal with complex outdoor problems, such as cities, with multiple wind attack angles, high-speed winds, and limited optional sensor placement locations.

There are also some techniques used for field measurement in some non-flow field research. Yi et al. (2011) constructed an improved genetic algorithm, generalized genetic algorithm (GGA), based on the criterion of the optimal sensor placement for the modal test. The dual-structure coding method instead of binary coding method is proposed to code the solution in GGA. The experiment results indicate that compared with exiting genetic algorithm, the GGA can improve the convergence of the algorithm in the case of the tallest building in the north of China. Wang et al. (2015) investigated temperature comfort issues inside airport terminals through questionnaires and indoor sensors. In the medical field, Kim et al. (2012) used the stereo digital image correlation technique for the full-field measurements of

aneurismal aortic tissues to investigate the behaviour of biomechanical failure. These full-field measurement methods have achieved excellent results in their respective fields, but not suitable for solving problems in the field of wind engineering. For example, the detection conditions are complex, the detection results are lagging behind, and the variables studied are difficult to obtain in wind engineering.

In summary, there are several obvious gaps in the existing research. First, studies on large spatial scales using unstructured sensor placement are not detailed enough to be implemented at the engineering level. Second, existing sensor optimization methods that consider building layout fail to deal with multiple wind attack angles. Third, methods designed for a single building only aim at solving specific problems and have poor generalization ability. Finally, methods from other research fields are difficult to transfer to wind engineering. Therefore, it is necessary to design a universal sensor optimization scheme for wind engineering applications. This study proposes an optimal sensor placement selection method based on mrDMD, Signed Distance Function (SDF) and Particle Swarm Optimization-Random Forest (PSO-RF). Specifically, the mrDMD method is applied to analyze the spatial-temporal data of the research variables, such as wind magnitude, pressure to identify the primary optimal sensors for each wind attack angle. Then, the SDF method is used to remove the sensor positions that are difficult or even infeasible for real engineering applications. Finally, the PSO-RF method is introduced for secondary selection based on the result generated by SDF. The objective of the secondary selection is to select the sensor locations, which are important when considering multiple wind angles. This scheme can be modified according to the practical needs of a particular project, and it is robust to be widely applied to a variety of working conditions and data types. After integrating these three methods, a feasible optimal sensor arrangement scheme considering multiple wind attack angles is established.

It is worth mentioning that, the method established in this article is a precursor to another study. After the optimal sensors are selected, another deep learning (DL) model will be built in the future work. The DL model is able to achieve real-time reconstruction data of wind speed or pressure field within a study area based on the sparse real-time data provided by these sensors. This paper is organized as follows. In section 2, the overall procedure of the optimal sensor position selection scheme is introduced. In section 3, the scheme of this study is validated based on two cases: an urban complex and a single building. Finally, the discussion and conclusion parts are presented in section 4.

2. Optimal sensor placement scheme

2.1. Methodology

The procedure of optimal sensor placement scheme is summarized in Fig.1. This scheme integrates mrDMD, SDF and PSO-RF to select the optimal sensor placement locations in practical engineering. At the beginning of this scheme, the mask data (1 for outer building region, and 0 for inner building region) that represents the basic geometry information is generated by using the flow field data of the target region. Based on the geometry information provided by mask data, the SDF data can be calculated according to Eqn.14. Then, the *Step 1*, which includes mrDMD analysis and QR pivoting selection is collectively referred to as

mrDMD Selection, is executed. With the help of mask data, the inner building data points can be eliminated from the uniform data matrix, thus preventing them from affecting the analysis of mrDMD on the flow field. Based on the analysis result of mrDMD, QR pivoting algorithm can find the spatial positions (optimal sensor positions) that can best represent the data feature. However, the locations chosen solely by mrDMD tends to be concentrated in some places where it is difficult to place the sensors (such as the middle of streets). For solving this problem, the *Step 2* is executed. Based on the initially calculated SDF value, the distance between any position in the space and the nearest building (that is, the SDF value) is known. From this, the SDF values of all existing sensors can be known, and the sensor positions that cannot be arranged in reality (such as sensors located in the middle of the road) are detected and eliminated. Nonetheless, the method described above is only valid under a specified circumstance (such as a given wind attack angle), because it costs extensive computational resources. For solving this problem, the *Step 3* is executed. Sensors selected by mrDMD at all wind attack angles are integrated. Their time series data are used as input into the PSO-RF algorithm, and the time-variation information of the wind field at each wind attack angle obtained by mrDMD analysis are concatenated together as output labels. After such a fitting relationship is established, the PSO-RF algorithm can judge the importance of each sensor under the comprehensive situation of all wind attack angles, and finally obtain the global optimal sensor arrangement scheme.

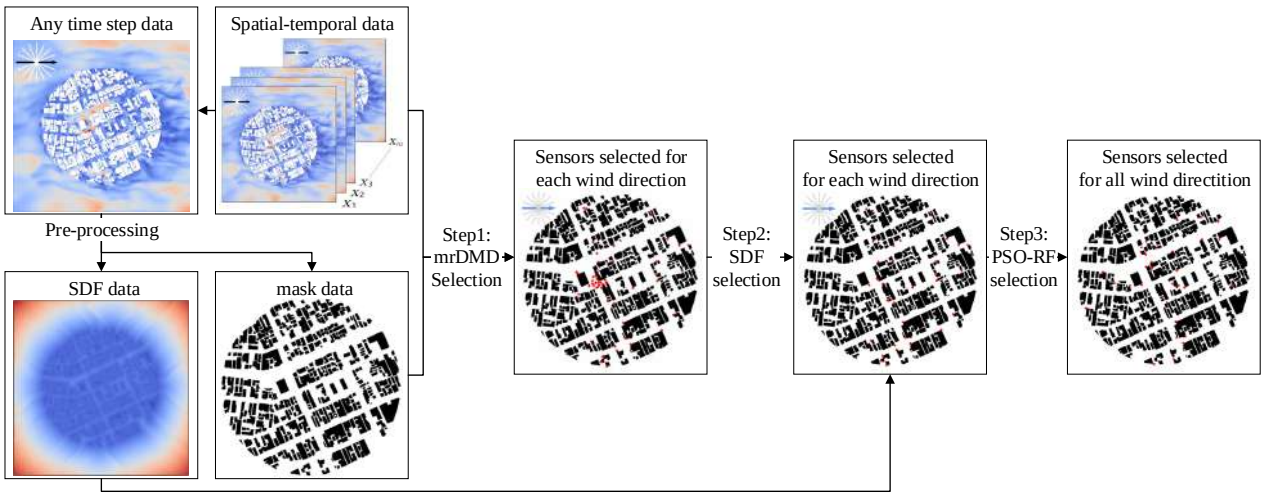


Figure 1: Overall procedure of the mrDMD-SDF-PSO-RF (mrDSPR) optimal sensor placement scheme

2.1.1. Dynamic Mode Decomposition (DMD)

To isolate the underlying multi-scale variability of data, a DMD modeling approach is used in this study. The DMD method provides a spatio-temporal decomposition of data into a set of dynamic modes derived from snapshots or measurements of a given system. The details of DMD can be obtained in the literature (Kutz et al., 2016). Here, for facilitating understanding of the following research, the procedure of DMD will also be simply described.

The DMD method is performed based on snapshots of data $x(t) \in R^n$ from a dynamical system (such as a flow field) at t time steps, where $t = 0, 1, 2, ..., m$. The DMD procedure builds a surrogate model that approximates locally linear dynamical system, as shown in Eqn. 1.

$$\frac{dx}{dt} = \frac{\Delta x}{\Delta t} = Hx \tag{1}$$

with initial condition $x(0)$ and well-known solution provided by Boyce et al. (2021), the $x(t)$ can be expressed as:

$$x(t) = \sum_{k=1}^{n} \phi_k \exp(\omega_k t) b_k = \vec{\phi} \exp(\vec{\Omega}\, t)\ \vec{b} \tag{2}$$

where ϕ_k and ω_k are the eigenvectors and eigenvalues of the matrix H, and the coefficients b_k are the coordinates of $x(0)$ in the eigenvector basis. For continuous dynamics as in Eqn.1, an analogous discrete-time system can always be sampled every Δt in time:

$$x(t+1) = Ax(t) \tag{3}$$

where

$$A = \exp(H\Delta t) \tag{4}$$

H refers to the matrix in the continuous-time dynamics from Eqn.1. The solution to this system can be expressed by the eigenvalues λ_k and eigenvectors ϕ_k of the discrete-time map A:

$$x(t) = \sum_{j=1}^{r} \phi_j \lambda_j^t b_j = \vec{\phi}\ \vec{\Lambda^t}\ \vec{b} \tag{5}$$

As before, $\vec{b}$ are the coefficients of the initial condition $x(1)$ in the eigenvector basis, thus $x(1) = \vec{\phi}\ \vec{b}$. Eqn.5 shows that, the DMD algorithm produces a low-rank eigendecomposition of A that optimally fits $x(t)$ for $t = 1, 2, ..., m$ in a least-square sense so that

$$\|x(t+1) - Ax(t)\|_2 \tag{6}$$

is minimized across all points for $t = 1, 2, ..., m-1$. The DMD reconstruction equation of all data is given as below:

$$\mathrm{X} = [x(1), x(2), ..., x(N-1)] = \vec{\phi}\cdot\vec{b}\cdot\vec{V}_{and} = [\phi_1, ..., \phi_r] \begin{bmatrix} b_1 & & \\ & \ddots & \\ & & b_r \end{bmatrix} \begin{bmatrix} 1 & \lambda_1^1 & \dots & \lambda_1^{N-2} \\ \vdots & \vdots & & \vdots \\ 1 & \lambda_r^1 & \cdots & \lambda_r^{N-2} \end{bmatrix} \tag{7}$$

where $x(t)$ $(t = 1, 2, ..., N-1)$ means the snapshot at time t, ϕ_k means the $k^{th}(k^{th} = 1^{st}, 2^{nd}, ..., r^{th})$ DMD mode, b_k means the amplitude of k^{th} mode, and $\vec{V}_{and}$ is the Vandermonde matrix. ϕ_k corresponds to the spatial information of the data, b_i represents the contribution of the k^{th} mode to the initial snapshot x_1. The Vandermonde matrix λ_k^{t-1} corresponds to the k^{th} eigenvalue of a matrix A, which represents the time-variance feature of the data. In this study, $\vec{b} \cdot \vec{V}_{and}$ is defined as the time-variation information of the data, and will be used in PSO-RF method as label data, as shown in Fig.10(d) and Fig.21(d).

2.1.2. Multi-resolution Dynamic Mode Decomposition (mrDMD)

The concept of multiple resolutions originated from video feeds. Specifically, the task of foreground/background subtraction can be achieved by separating the slow and fast modes (Kutz et al., 2015). The key of multi-resolution Dynamic Mode Decomposition (mrDMD) is to remove low-frequency modes from a given series of snapshots. The observed dynamics can be approximated by DMD modes with a chosen number of snapshots m $(m = 1, 2, ..., M)$. If all M snapshots are used for DMD analysis, which means the DMD decomposition is only performed once on M snapshots data, the obtained analysis results can well express the low-frequency characteristics of the data, and if a small part of M snapshots (e.g. $M/2, M/4...$) are used, the obtained analysis results can well express the high-frequency characteristics of the data. Therefore, the full utilization of M, which means the analysis result is based on the mrDMD decomposition results of all levels, is able to display high- and low-frequency content of the data. As shown in the bottom of Fig.2, during the procedure of mrDMD, modes will be divided into different levels according to their frequency (each row in the table represents a level). According to the code published in the study of Kelp et al. (2022), the cut-off frequency between the i^{th} layer and the $(i+1)^{th}$ layer is defined as $f_{threshold}(i) = max_cyc/T(i)$. Among them, $T(i)$ is the time span of the data analyzed by the i^{th} layer DMD. max_cyc is a parameter to adjust the overall frequency range for mrDMD analysis. This parameter has different optimal values for different data, and it is necessary to conduct trial parameter experiments to determine its value with the goal of obtaining the optimal reconstruction results.

From the initial DMD analysis of the M snapshots, k_1 modes with the lowest frequency are kept in the 1^{st} level, and the time domain is divided into two equal parts with $M/2$ snapshots. Then, the DMD analysis is performed on each $M/2$ snapshots again, and k_2 modes with the lowest frequency are kept in the 2^{nd} level. The algorithm keeps running the above iteration process until the time duration of the snapshot is small enough to satisfy the predetermined requirements. Mathematically, the DMD analysis in mrDMD based on M snapshots is described as follows:

$$\begin{aligned} X_{mrDMD}(t) &= \sum_{k=1}^{K} b_k^1 \psi_k^{(1,1)}(\xi) exp(\omega_k t) \\ &= \underbrace{\sum_{k=1}^{k_1} b_k^1 \psi_k^{(1,1)}(\xi) exp(\omega_k t)}_{(slow \quad modes)} + \underbrace{\sum_{k=k_1+1}^{K} b_k^1 \psi_k^{(1,1)}(\xi) exp(\omega_k t)}_{(fast \quad modes)} \end{aligned} \tag{8}$$

where K is the total mode number of the data, b_k^1 represents the initial amplitude of each

mode in 1^{st} level, $\psi^{(1,1)}(x)$ denotes the DMD modes based on all M snapshots, ξ is the spatial coordinate, $\omega_k = ln(\lambda_k)/\Delta t$ is the eigenvalue of each mode, and t is the collection time. The first term in the right hand side represents the low-frequency-mode dynamics, while the second term represents modes with higher frequency. Therefore, a matrix can be generated by computing the second term:

$$X_{M/2} = \sum_{k=k_1+1}^{K} b_k^1 \psi^{(1,1)}(\xi) exp(\omega_k(t)) \tag{9}$$

The DMD analysis procedure can now be exerted on the matrix $X_{M/2}$ once again. However, the matrix $X_{M/2}$ is divided into two matrices:

$$X_{M/2} = X_{M/2}^{(1)} + X_{M/2}^{(2)} \tag{10}$$

where $X_{M/2}^{(1)}$ and $X_{M/2}^{(2)}$ cover the first $M/2$ and the remaining $M/2$ snapshots, respectively. The k_2 low-frequency-DMD modes at this level are identified as $\psi_k^{(2,j)}$ $(j = 1, 2)$, where they are computed independently based on the first or second part of snapshots.

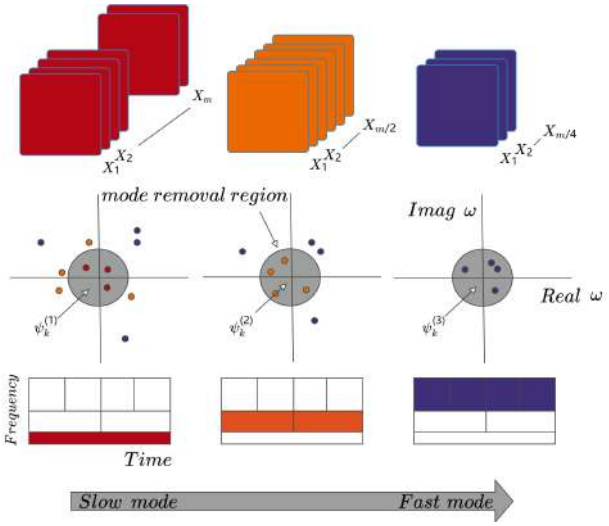


Figure 2: Schematic diagram of mrDMD theory (Reproduced from Kutz et al. (2015))

Based on the above procedure, an iteration process can be established by recurrently filtering low-frequency DMD modes and this iteration process also creates new matrices $X_{M/2}$, $X_{M/4}$, $X_{M/8}$, ... until an adequate multi-resolution decomposition has been achieved. The approximate DMD solution is given as follows:

$$X_{mrDMD}(t) = \sum_{k=1}^{k_1} b_k^{(1)} \psi_k^{(1,1)}(\xi) exp(\omega_k^{(1)}(t)) + \sum_{k=1}^{k_2} b_k^{(2)} \psi_k^{(2,j)}(\xi) exp(\omega_k^{(2)}(t)) + \sum_{k=1}^{k_3} b_k^{(3)} \psi_k^{(3,i)}(\xi) exp(\omega_k^{(3)}(t)) + ... \tag{11}$$

where the correct modes from the sampling window are selected at each level of the decomposition at the evaluation time t. $\omega_k^{(l)}$ represents the eigenvalues of DMD at l^{th} level of decomposition. $\psi_k^{(l,z)}$ represents the modes of DMD at l^{th} level and z^{th} part of decomposition. $b_k^{(l)}$ is the initial amplitude of each mode in l^{th} level, and k_i $(k_1, k_2, k_3...k_n)$ are the number of low-frequency modes retained at each level. Thus, the essential multi-resolution features can be represented by various spatio-temporal DMD modes through mrDMD. The process of mrDMD is shown in Fig.2, which describes the procedure of a three-level mode decomposition. In Fig.2, the lowest-frequency mode, mid-frequency mode and the highest-frequency mode are represented in order of red (snapshots and eigenvalues), orange and purple, respectively. As shown in Fig.2, the mrDMD method can analyze data with multiple time intervals. For large time interval, the periodic features of data are captured, and the features of sudden wind events are captured in small time interval data. They correspond to the overall ventilation capacity of the city and the extreme wind events that may cause danger, respectively. Therefore, mrDMD is more suitable for this study than DMD, which can only capture global periodic features.

2.1.3. QR pivoting

QR pivoting is one of the most commonly used optimized sensing greedy algorithms for finding near-optimal sensor locations in a computationally efficient way. Compared with compressed sensing, optimized sensing provides drastic reductions in the required number of sensors and improved reconstruction are observed in examples ranging from facial images to fluid vorticity fields. Among optimized sensing algorithms, QR pivoting is particularly favorable, as it is fast, simple to implement, and provides nearly optimal sensors tailored to a data-driven mode basis (Manohar et al., 2018a). For a very detailed description of QR decomposition, please refer to the works of Manohar et al. (2018b), and only a brief introduction is given in this article. In QR column pivoting, the rows in the modal library φ_r are identified with the highest 2-norm that represents the locations with the largest wind field frequencies and thus the greatest variability. The matrix $A \in \mathbb{R}^{m \times n}$ is decomposed into a column permutation matrix C^T, an upper-triangular matrix R, and a unitary matrix Q based on the reduced matrix QR column pivoting, such that $AC^T = QR$. Therefore, QR column pivoting generates r point sensors (pivots) which sample the r tailored φ_r best:

$$\varphi_r^T C^T = QR \tag{12}$$

It means that the spatial locations, corresponding to the most representative features in the mode library, are identified by each QR pivot (Kelp et al., 2022). Moreover, QR pivoting will sort all spatial points from important to unimportant. Sensors in these locations can monitor critical wind field variation above background variations. In this paper, the sensor positions selected by the combination of mrDMD method and QR pivoting method are referred to as the position selected by mrDMD.

2.1.4. Signed Distance Function (SDF)

Signed distance function (SDF) (Guo et al., 2016), which calculates based on the Cartesian grid, is employed to provide the universal information of different building shapes and

distances between buildings with only once calculation, and hence can help to determine the sensor locations in a real scale. The calculation of SDF needs to be executed based on mask data, because mask data can clearly provide the position of all obstacles in the research area, thus it is convenient for distance calculation. To calculate the SDF based on the discrete representation on a Cartesian grid of a 2D geometry (mask data), a zero level set is created first. The zero level set is a set of points (i, j) that describes the building geometry boundary (surface) in a region, $\Omega \subset R^2$:

$$Z = \{(i, j) \in R^2 : f(i, j) = 0\} \tag{13}$$

where f is the level set function, that is, $f(i, j) = 0$ when (i, j) is on the boundary of building geometry, $f(i, j) > 0$ when (i, j) is outside the building geometry, and $f(i, j) < 0$ when (i, j) is in the building geometry. A SDF $D(i, j)$ correlated with a level set function $f(i, j) > 0$ is determined as below:

$$D(i, j) = \min_{(i^{'}, j^{'}) \in Z} |(i, j) - (i^{'}, j^{'})| sign(f(i, j)) \tag{14}$$

$D(i, j)$ represents the shortest distance between a given point (i, j) and its nearest building boundary Z. The *sign* function is determined by the location of the point related to the building boundary. An example of SDF is shown in Fig.8(d), which describes the SDF value of the pedestrian height plane in Niigata, a city in Japan. For research with a dense building layout and only concerns about the outside space of the building, the calculation of the SDF of the inner building points can be skipped to save certain computing resources. For a complete display of the effect of SDF, the SDF values of inner building points are also calculated in this paper.

2.1.5. Particle Swarm Optimization-Random Forest (PSO-RF)

Particle swarm optimization (PSO) developed by Kennedy & Eberhart (1995) is one of the most commonly used evolutionary algorithms. PSO is a heuristic optimization algorithm developed based on the behavior of biological groups in nature. Compared with other bionic optimization algorithms, it has fewer codes and parameters, and has inherent parallelism, thus it is easy to use and has a faster calculation speed. The design concept of PSO is based on the predation behavior of a flock of birds. The first step of PSO is to initialize a group of random particles (random solution to the given question), and each particle is identified by two attributes: velocity and position. Then, the algorithm searches the optimal particles in an iterative process. Specifically, particles refresh themselves based on individual extreme (the optimal solution found by individual particle) and global extreme (the optimal solution determined by the entire particle swarm) during the iteration process. The iteration process is mathematically shown as below:

$$v_{i+1} = \omega \times v_i + c_1 \times rand_1 \times (pbest_i - L_i) + c_2 \times rand_2 \times (gbest_i - L_i) \tag{15}$$

$$L_{i+1} = L_i + v_{i+1} \times 1 \tag{16}$$

where c_1 and c_2 represent the learning factors, ω is the inertia factor, $pbest_i$ and $gbest_i$ respectively are the value in the i^{th} dimension of the best individual particle position and the best position of the entire population, $rand_1$ and $rand_2$ represent random numbers between $(0, 1)$, v_i and L_i are the velocity and position of the i^{th} dimensional particle value respectively.

The random forest (RF) bagging method developed by Breiman (1996) excels at handling high-dimensional problems in parallel and is able to avoid over-fitting issue. Compared with the serial decision tree GBRT, the parallel nature of RF can significantly speed up the program operation. The principle of random forest is to simultaneously train multiple weak learners, e.g. decision tree regressor, and the final result is generated by taking average value of outputs of all weak learners. This kind of procedure is called ensemble method (Zhou, 2012), which always performs better in prediction accuracy than a single learner, and reduces the impact of abnormal values and possibilities of overfitting. A very detailed illustration and mathematical explanation of random forests can refer to the research of Li et al. (2018b). In this section, only the random forest-based importance ranking function is supplemented. The pseudocode of importance ranking can be described as follows:

- Inputs: trained model m, dataset $D \in R^{N \times J}$, where N is the sample number, and J is the feature number.
- Compute the reference score s of the model m on dataset D (for instance the R^2 for a regressor).
- For each feature j (column of D):

  □ For each repetition k in $1, ..., K$:

  ▲ Randomly shuffle column j of dataset D to generate a corrupted version of dataset named $\widetilde{D_{k,j}}$.

  ▲ Compute the score $s_{k,j}$ of mdoel m on corrupted data $\widetilde{D_{k,j}}$.

  □ Compute the importance i_j for feature f_j defined as: $i_j = s - \frac{1}{K} \sum_{k=1}^{K} s_{k,j}$

In this section, PSO, whose ability in model optimization has been demonstrated (Hasanipanah et al., 2016a,b; Lin et al., 2021, 2022), is chosen to adjust the hyper-parameters of random forest model. The dominant hyper-parameters of RF include the maximum number of weak learners, the minimum number of samples to divide the nodes, the minimum number of samples at a node, the maximum number of features in RF partitioning, and the maximum depth of the tree.

2.2. Optimal Sensor Placement Procedure

2.2.1. Step 1: mrDMD selection

The first step of the optimal sensor placement scheme is operated on the matrix library ($\psi_r \in \mathbb{R}^{n \times r}$) including all mrDMD modes. Specifically, the locations are selected by employing QR pivoting to the mrDMD library (Heck et al., 1998; Manohar et al., 2018a). Optimal

sensor locations selected by mrDMD are the most representative position of the spatial-temporal wind field. However, those locations may not be feasible for placing sensors in the city block. For example, Fig.2.2.1 demonstrates that the optimal sensor positions are concentrated on the wake of the high-rise building as shown in the red circle under the WNW oncoming wind flow. The definition of wind attack angle is given in Fig.8(b). This result is reasonable from the viewpoint of wind flow field analysis, but not feasible for practical arrangement because it is hard to arrange a large number of sensors in the middle of the street.

Figure 3: mrDMD selected sensor and PSO-RF feature ranking

2.2.2. Step 2: SDF filter

To make the sensor arrangement feasible, the second step of the proposed scheme introduces SDF on the basis of mrDMD for adjusting the sensor position to satisfy the engineering requirement. Specifically, with the help of SDF, sensor locations can be restricted in the region that is feasible for practical arrangement. In this study, the region is set as 4m around the building, which is filled with sidewalks and green belts. In addition, considering the cost of sensor arrangement, after the SDF selection, the distance between the two sensors should be larger than a critical value that can be adjusted case by case. In this study, when the distance between two sensors is less than 10m, the less important one calculated by QR pivoting will be deleted. It is worth noting that the SDF method is not necessary for pressure sensors on a single building, where the sensors can be placed almost anywhere on the building surface.

2.2.3. Step 3: PSO-RF selection

The mrDMD-SDF method mentioned above can only deal with the situation of a certain wind attack angle. However, natural wind comes from all directions. In order to consider different wind attack angles, this study further optimizes the mrDMD-SDF method through the PSO-RF algorithm, because this algorithm not only require low computational resource, but also possesses the ability of ranking feature importance.

For PSO-RF in the third step of the proposed scheme, the input data is a set of time-history data of wind speed amplitudes, and they are monitoring results of the sensors jointly selected by mrDMD and SDF. Sensors designed for all considered wind attack angles are

included in the input data. Moreover, the label data of PSO-RF is a summation of time-variation information of all valid modes in every mrDMD level. The data structures of input and label data are shown in Fig.11 and Fig.10(d).

After training the PSO-RF algorithm to map the input data to the label data, the RF can rank the features (sensors in different locations) of the input data according to importance. Specifically, the reduction degree in model performance after randomly shuffling values of each feature is the main criterion for evaluating the importance of the feature. The procedure of feature importance ranking is described in detail in Sec.2.1.5.

Finally, the same number of sensors are selected by the mrDMD-SDF-PSO-RF (mrD-SPR) method. Based on the selected sensors, the wind flow field or wind pressure field is reconstructed by mrDMD, and the effectiveness of the sensor locations can be evaluated according to the reconstruction error. The reconstruction procedure for a snapshot can be explained by Eq.17.

$$x_{Recon} = Modes \times (Modes[sens, :]^{-1} \times x[sens]) \tag{17}$$

where x_{Recon} represents the reconstructed data (one snapshot); $\times$ means matrix multiplication; $sens$ stores the space index of the sensor selected by the algorithm in this paper; $Modes$ are all the modes reserved in each level of mrDMD correspond to the snapshots need to be reconstructed, each mode is a column of $Modes$, and $Modes[sens, :]$ implies spatially sparse sampling at the sensor locations of all mrDMD modals; x represents the original wind field/pressure data of the corresponding snapshot, and $x[sens]$ implies spatially sparse sampling at the sensor locations of the original data.

Based on Eq.17, it is necessary to explain that mrDMD reconstruction is not a process of reconstructing the wind or pressure field data only through the data of sparse sensors. In fact, besides the sparse sensors data, the mrDMD reconstruction also uses the mode data. Therefore, the mode reconstruction cannot be regarded as a real wind or pressure field reconstruction, but only shows the importance of the selected points to the wind or pressure field, as it reported in the literature (Cohen et al., 2004; Manohar et al., 2018a; Kelp et al., 2022; Willcox, 2006; Yildirim et al., 2009).

3. Case studies

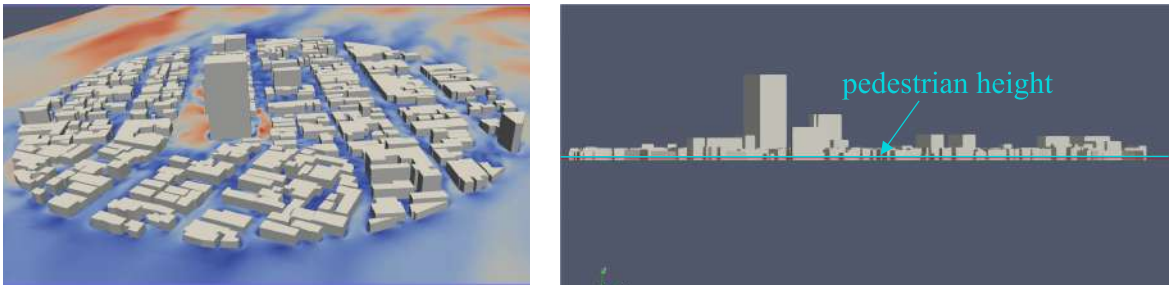
To demonstrate the effectiveness of the mrDSPR method, two different case studies in wind engineering are tested. The two cases are: (1) wind flow field of the Niigata city (a city of Japan) generated by large eddy simulation (LES); (2) wind pressure field of a tall building (located in Sanya, China) obtained by wind tunnel experiment. The mrDSPR method is applied to these two cases to compare the effectiveness of selected sensor positions based on reconstruction errors.

3.1. Case 1: wind flow field of city block

3.1.1. Case description

The city block model used in this section comes from AIJ benchmark, a part of Niigata, Japan, and the model scale is 1/250. As shown in Fig.4(a), the Niigata city block is composed

of complex building geometries, which often accompanies with complex wind flow field. It is challenging to design sensors placement locations to maximize the flow field information monitored by the sensors. Thus, the proposed mrDSPR method is employed to solve this problem.



(a) Geometry of Niigata with pedestrian level wind field

(b) Study plane, i.e. pedestrian height

Figure 4: Geometry model of Niigata and the pedestrian level wind flow field

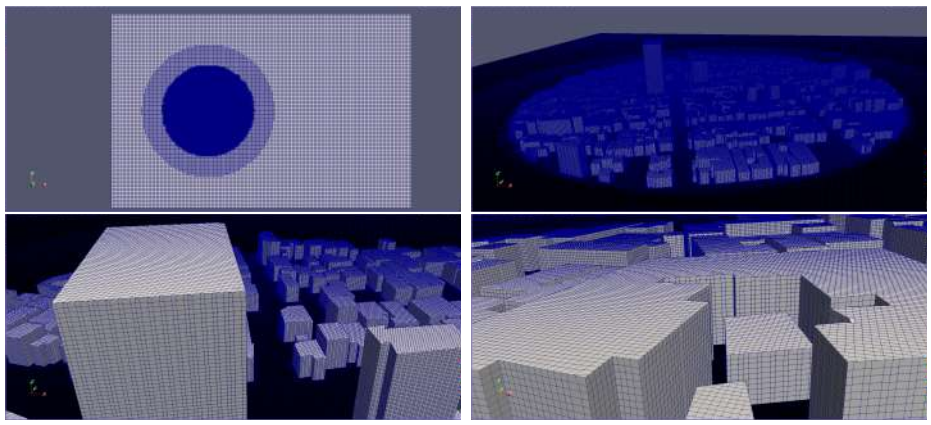


Figure 5: Grid distribution for the Niigata city model

In this case, the wind flow field of Niigata is simulated using OpenFOAM with a LES subgrid model. The numerical setup follows the research of Ricci et al. (2017); Weller (2012) and Bruno et al. (2010). In particular, velocity gradients are calculated relying on the *faceLimited Gauss linear* scheme, while for other quantities the *cellLimited Gauss linear* scheme is adopted. Notice that *faceLimited Gauss linear* and *cellLimited* schemes are OpenFOAM built-in limiters needed to enhance the scheme boundedness, thus, increase the simulation stability. The interpolation from the cell center to the face center is operated by the second-order central differentiation method. For laplacian terms, the *Gauss linear limited 0.5* scheme is used, in which *linear* is the interpolation scheme (central difference scheme, second-order), and *limited 0.5* is a flag that ensures that non-orthogonal contributions do not exceed the orthogonal part, which is still needed to ensure the simulation stability. As regards divergence terms, for velocity the *bounded Gauss LUST* scheme is used, which

is known to be a suitable compromise between accuracy and stability, and is obtained as a blend with 75% *linear* (second-order) plus 25% *linearUpwind* (second-order), which is numerically preferable compared to the pure linear scheme, which often leads to unsteady computations. The turbulence kinetic energy divergence is discretized by *bounded Gauss limitedLinear 1*. Pressure-velocity coupling is obtained using the well-known PISO algorithm. As regards solvers, the *GAMG* (multi-grid solver) is selected for the pressure equation, and *smoothSolver* (smoothing solver) is chosen for other variables. Time advancement is performed using the *Crank Nicolson* (Implict second-order discrete format) with a time step of 0.00002 s. We simulate 30 s for each wind attack angle, and disregard the first second to avoid initialization effects. The mesh is generated by OpenFOAM built-in functions blockMesh and snappyHexMesh. Considering the complexity of the city block model and the fact that we do not aim at measuring details of the flow in particularly delicate locations of the buildings, this study does not consider a yPlus value needed for integration through the boundary layer which is, by the way, often unnecessary for sharp edged bluff bodies. Van Driest damping is adopted at the walls. The mesh details are shown in Fig.5.

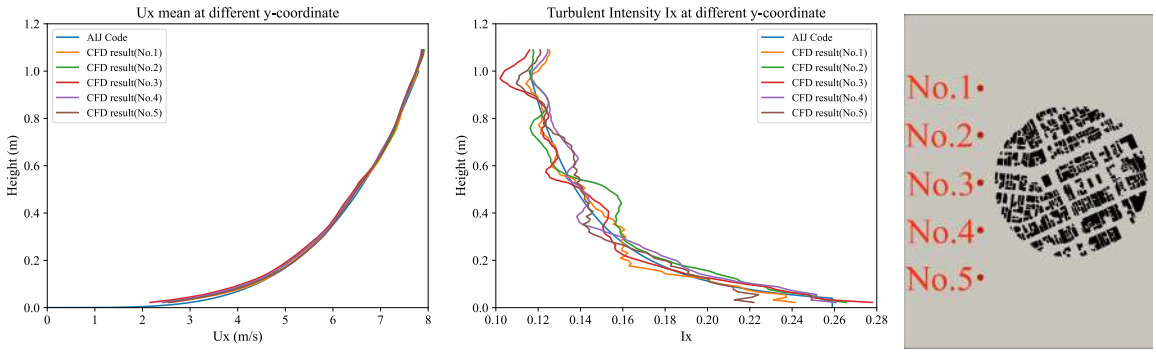


Figure 6: Mean wind speed, turbulence intensity profile and five positions of monitoring wind profiles

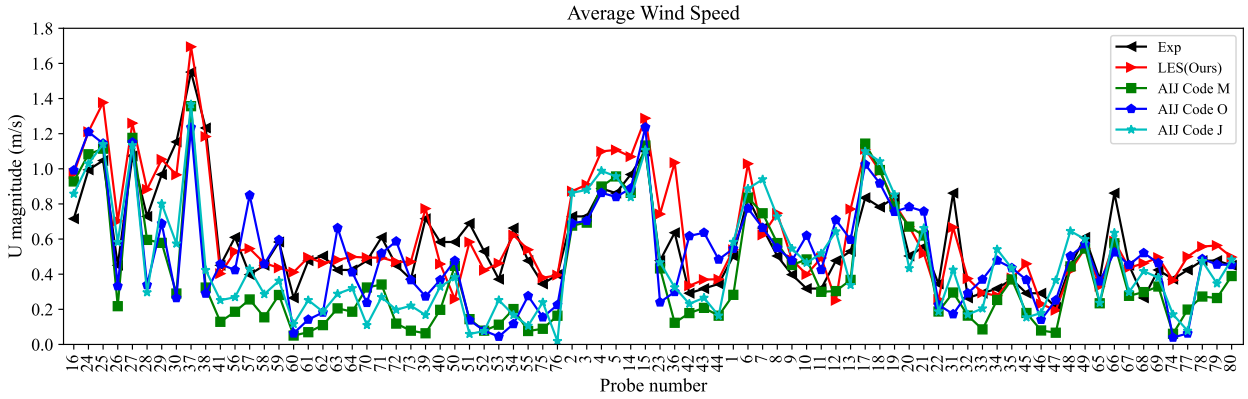
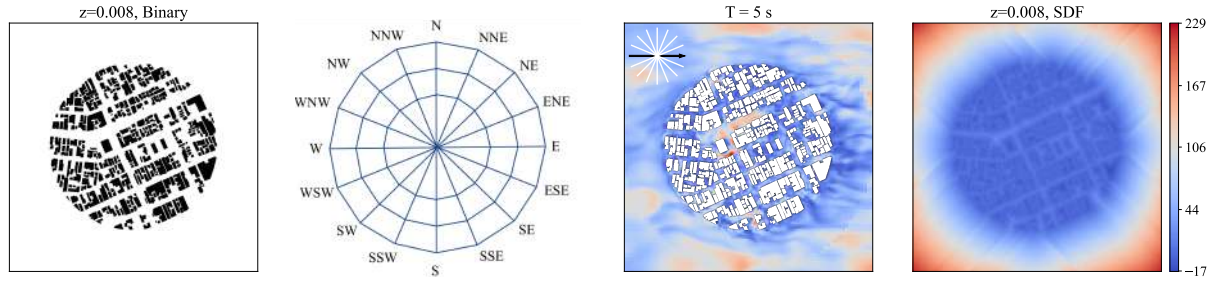


Figure 7: Comparison of LES and AIJ Results (The label 'AIJ Code M', 'AIJ Code O' and 'AIJ Code J' represent the three simulation schemes used by AIJ, described in (Tominaga et al., 2008).)

The inlet flow is generated by the method of Bervida et al. (2020) under the guidance of AIJ code. The mean wind speed and the turbulence intensity profiles at five locations are monitored and shown in Fig.6. It can be seen that the wind profiles agree well with those of the AIJ code. Furthermore, Fig.7 compares the wind speed data of 80 measuring points provided by AIJ wind tunnel experiment and the simulation data by LES. The simulation accuracy of the present LES simulation is better than the AIJ result in nearly half of detected points (point 1, 10, 21, 28-31, 33-35, 37-39, 42-44, 46-47, 51-56, 58, 60-64, 67, 70, 72, 74-77), and similar to AIJ results in most of other detected points. In addition, mesh independence is also tested based on the simulation results of these 80 points. This study produced a set of about 5.6 million grids with $yPlus = 6$ and a set of about 10 million grids with $yPlus = 2$. There is not obvious difference between the measured wind speeds of the two mesh sets, so this study uses 5.6 million mesh set for the LES simulation.



(a) Geometry defined by binary data (b) Wind directions (c) Study area (d) SDF

Figure 8: Geometry defined using binary data, definition of wind direction, study area, and SDF map for Niigata model

3.1.2. Data pre-processing

The main area studied in this study is the horizontal plane at the pedestrian-level height (2m above the ground, and 0.008m in model scale, as shown in Fig.4(b)). The diameter of the city block model is about 1.6m in model scale, and we select a 2.4m × 2.4m square centered on the model as the targeted area as shown in Fig.8(c). This region fully contains wind flow details in and around the city block. The wind flow information is extracted through OpenFOAM sensors. A sensor matrix with a resolution of 1024 × 1024 is set to cover the entire square area. In this way, the influence of uneven grid can be avoided, and an uniform matrix is obtained for subsequent analysis. It is worth mentioning that the sensor matrix method demands a high resolution, otherwise the information in the finest mesh area will be lost. Moreover, when the sensors are inside the building, the value of extracted data is infinite. Taking advantage of this property, we can convert the infinite values into 0, and normal values into 1. Then, a binarization map of the targeted area is obtained, as shown in Fig.8(a).

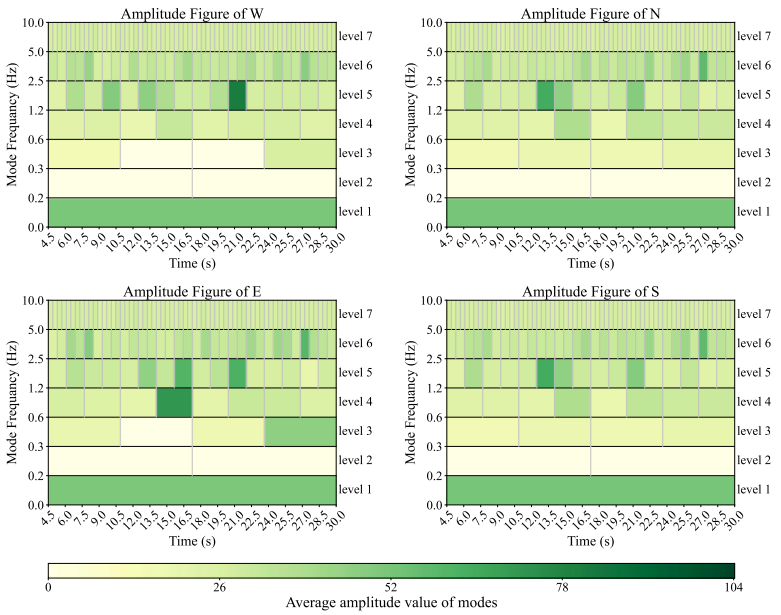


Figure 9: Maps of mrDMD modes for 4 wind attack angles: western (W), southern (S), eastern (E), and northern (N). When the mode obtained by DMD analysis is within the frequency range corresponding to its level, the amplitude figure will display a non-zero value, and these modes will be considered as valid modes by mrDMD. The mean values of valid mode amplitudes are displayed in the figure. Therefore, the more non-zero regions present in this figure, the more effective modes that mrDMD decomposes into different frequency ranges.

3.1.3. mrDMD procedure

As described in Section 2.1.2, the temporal intervals of different mrDMD layers are divided by 2, e.g. M, $M/2$, $M/4$.... Therefore, the raw data with 2^n time steps is required for mrDMD analyses (Kutz et al., 2016). In the present study, 256 snapshots with a time interval of 0.1s extracted from the last 25.6s (i.e. 4.5s-30s) simulation results are stored to perform mrDMD analyses for all different wind attack angle cases. There is some freedom in the choice of time interval. Specifically, this choice can be determined by the specific goals of the problem. When the higher frequency data feature are required, a smaller time interval can be selected, otherwise a larger time interval can be selected. But it should be noted that, a too low sampling frequency would cause heavy information lost and goes against the aim of mrDMD, which is to handle the transients in data. A too high sampling frequency that produces too many snapshots is unnecessary and inefficient because only the slow modes are of interest in decomposition of each level (Kutz et al., 2016). The main purpose of this study is to propose the complete flow of mrDMD-SDF-PSO-RF method, thus no special consideration is given to the sampling frequency.

The maps of mrDMD wind field data modes for four wind attack angles (N, W, S, E in Fig.8(b)) are shown in Fig.9. These maps illustrate the mode with different frequencies captured by the mrDMD analysis at four wind attack angles. From the bottom to the top of each subplot, the time durations of analyzed data correspond to each level are 25.6s, 12.8s,

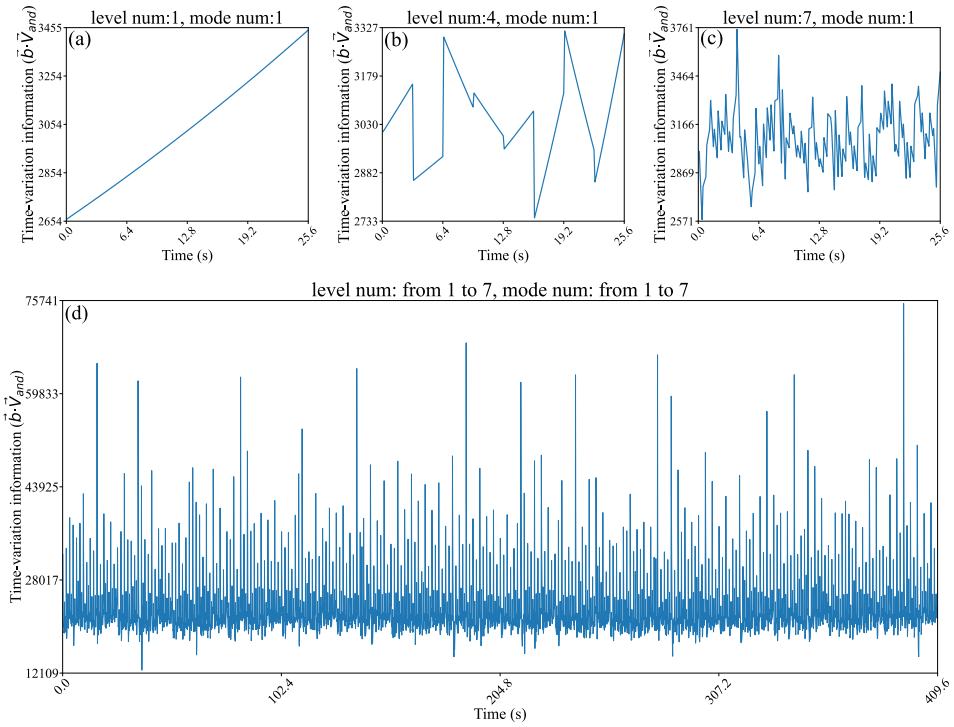


Figure 10: Time-variation information of different DMD level and order for Western wind direction: (a)Time-variation information for 1^{st} mode of level 1 (lowest level) mrDMD analysis; (b)Time-variation information for 1^{st} mode of level 4 (medium level) mrDMD analysis; (c)Time-variation information for 1^{st} mode of level 7 (highest level) mrDMD analysis; (d)Summation of time-variation information of every mrDMD layer and order. The y-axis of these figures represents the time-variation information of the data, as defined in Eq.7.

6.4s, 3.2s, 1.6s, 0.8s, and 0.4s respectively. It is worth mentioning that, a simply sharp cutoff of sampling window may cause information lost especially the transients (Kutz et al., 2016). Therefore, although this has not been done in most studies using mrDMD, an overlap between sampling windows is worthy trying in the future work. There is no significant mode when the time duration of analyzed data is less than 0.4s, and therefore the corresponding levels are not included, similar to Kelp et al. (2022).

The Vandermonde matrix is also calculated and plotted to demonstrate the time-variation information of the data. In detail, the corresponding value of an r^{th}-order mode at the t^{th} time step in the Vandermonde matrix is λ_r^{t-1}, where λ_r represents the DMD eigenvalue (plural) of the r^{th}-order mode, as shown in Eq.(1). The real part of λ_r^{t-1} represents the time-variance feature of the data. Fig.10(a-c) illustrate the Vandermonde matrix value of different levels and different orders of DMD modes for W wind direction. As shown in the figure, with the increase of layers and orders, the corresponding DMD time-variation information becomes more and more high-frequency and richer in variability. In this case, we use the summation of time-variation information of every layer and every order mode as the label value of PSO-RF. It can not only reflect the overall trend of wind field data in a large time span, but also reflect the fluctuation of a small time span. The summation result is shown

in Fig.10(d).

The matrix truncation order in the mrDMD analysis is obtained by selecting outliers for the diagonal matrix generated by the singular value decomposition. The data of each wind attack angle needs to be analyzed by the mrDMD and the QR pivoting. However, limited by space, only the data of W/N/E wind attack angles are shown in Fig.15(a). In order to keep the number of sensors identical for all wind attack angles after SDF filter, the number of sensors for each wind attack angle selected by the mrDMD results and the QR pivoting method is different, 50 for W wind attack angle, 56 for N wind attack angle, and 48 for E wind attack angle.

3.1.4. Signed distance function filter

As mentioned in Section 2.2, the signed distance function (SDF) method is used to constrain the sensor locations so that the placement scheme can be implemented in practical. In detail, the distance between the sensor and the building boundary is limited to be not more than 4m in full scale, and the distance between two sensors is limited to be not less than 10m. Certainly, these two values can be adjusted case by case. The filtered results of the mrDMD-selected sensor locations based on the Niigata SDF value matrix are shown in Fig.15(b). After the SDF filter, the sensor locations are no longer concentrated in the road or other locations that are difficult for actually placing sensors. Apparently, the SDF filter greatly improves feasibility of the sensor placement scheme.

		458 Sensors in total						
		Sensor 1(W)	Sensor 2(W)	Sensor …(W)	Sensor N(W)	Sensor 1(WNW)	Sensor …(…)	Sensor N(WSW)
4096 time steps in total	t = 4.5s(W)	Umag	Umag	…	…	…	…	Umag
	t = 4.6s(W)	…	…	…	…	…	…	…
	t = …(W)	…	…	…	…	…	…	…
	t =30s(W)	…	…	…	…	…	…	…
	t = 4.5s(WNW)	…	…	…	…	…	…	…
	t = …(…)	…	…	…	…	…	…	…
	t = 30s(WSW)	Umag	Umag	…	…	…	…	Umag

Figure 11: The input data structure for the PSO-RF algorithm (Umag means velocity magnitude of corresponding time, wind attack angle and sensor).

3.1.5. PSO-RF selection

The sensor placement locations for all wind attack angles are assembled in order to consider variability of natural wind direction, as shown in Fig.15(c). After removing duplicate points, 458 sensors remain in total. The wind speed monitored by these sensors are extracted, including all wind attack angles (256 time steps for each wind attack angle, thus 4096 time steps for 16 wind attack angles in total), to form the input matrix of the PSO-RF algorithm, as shown in Fig.11. Therefore, we use the aforementioned matrix as input, and the mrDMD time-variation information described in Fig.10(d) as output to train the PSO-RF model. The fitting result obtained after training is shown in Fig.12(b).

The performance of RF optimized by PSO is shown in Fig.12(a). It can be seen that the R^2 value of the fitting result is only 0.856, which means that the PSO-RF algorithm is

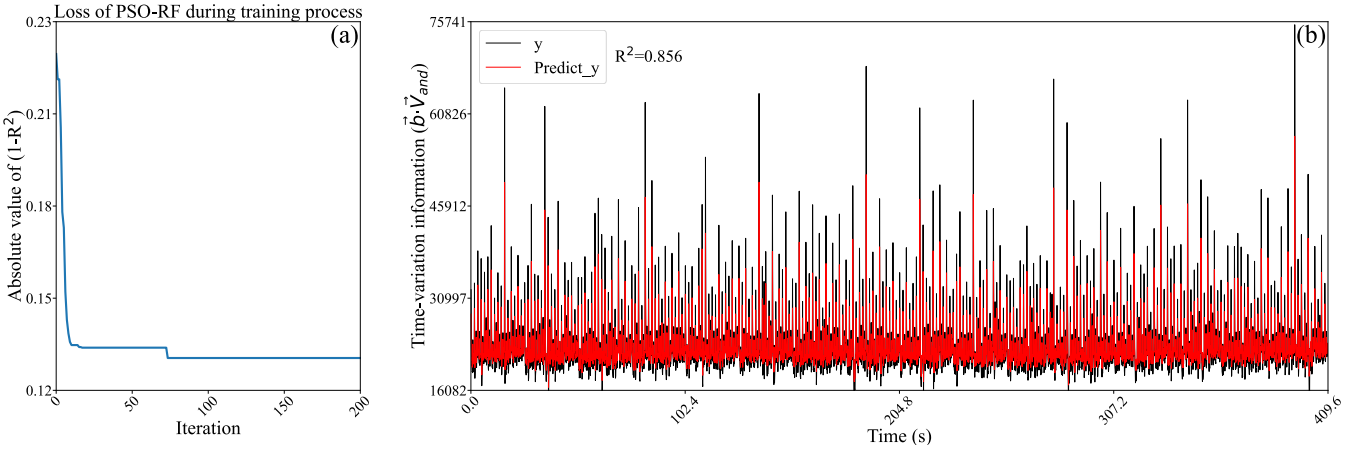


Figure 12: Performance of PSO-RF algorithm: (a) Performance of RF optimized by PSO; (b) PSO-RF fitting results for AIJ case. For (a), as a regression fitting problem, the closer the R^2 of the data is to 1, the better the fitting effect. Therefore, the lower the absolute value of $(1 - R^2)$, the better the fit of the algorithm. For (b), the y-axis of these figures means the time-variation information of the data, as defined in Eq.7.

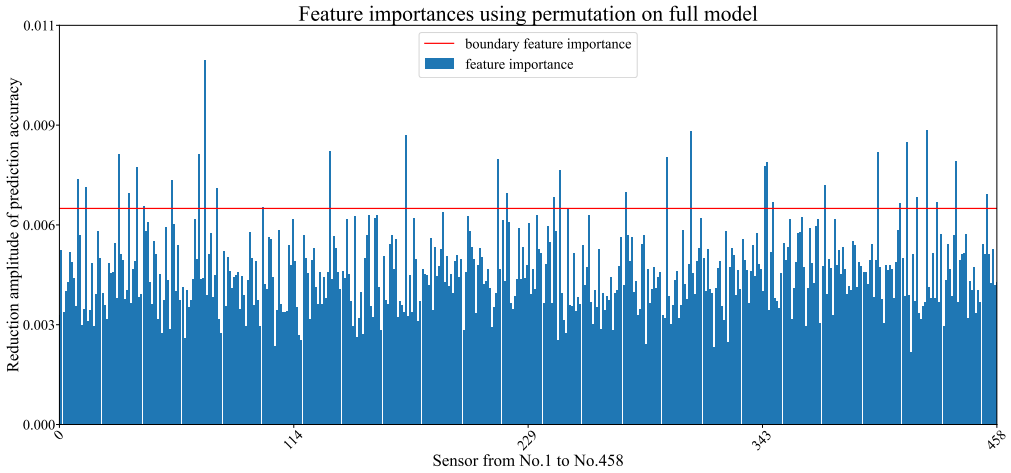


Figure 13: Sensor importance for AIJ case ranked using the PSO-RF algorithm. The principle of the importance ranking of the PSO-RF algorithm is that when the model finishes fitting the data, it stops modifying the internal parameters of the model. After that, the time-history data of each sensor is shuffled, while keeping the time-history data of other sensors unchanged. On the basis of such input data, time-variation information is predicted again. As the time-history data of one sensor is shuffled, it inevitably results in a drop in prediction accuracy. The more this accuracy drops, the more important the corresponding sensor is. The y-axis of figure corresponds to *importance* i_j described in Sec.2.1.5.

difficult to numerically and accurately fit the complex and variable mrDMD time-variation information. However, although PSO-RF cannot accurately fit the magnitude of each value, it captures almost the full trend of data variation, including regular fluctuation and surges at some time points, as can be seen from the details of Fig.12(b). The fitting of this trend

is particularly important for feature importance ranking.

Actually, the PSO-RF algorithm considers different sensor wind velocities as different features, then ranks the importance of these features. In other words, the algorithm tries to find the most important sensor locations considering all wind attack angles. Meanwhile, the RF algorithm can operate feature importance ranking. Hence, the importance of each sensor is figured out in Fig.13. The larger value in the figure means that after shuffling the time series of wind speed monitored by a sensor, the greater reduction on the prediction accuracy of the PSO-RF model will be achieved. We select 30 most important sensors, identical to the sensor number of mrDMD-SDF selection, to easily compare these two methods. The red line in Fig.13 represents the importance of the 30^{th} sensor. The selected 30 sensor locations are shown in Fig.15(d).

3.1.6. Result and Discussion

The quality of the selected sensor locations is evaluated by comparing the mrDMD reconstruction error of wind flow field. For evaluating the mrDMD-SDF method, we perform the mrDMD reconstruction for each wind attack angle based on 30 optimal sensors for this wind attack angle selected by the mrDMD-SDF method. For assessing the mrDSPR method, the mrDMD reconstructions for all wind attack angles are performed all simply based on the same set of 30 optimal sensors selected by the mrDSPR method. It should be noted that for the mrDMD-SDF method, the reconstruction is performed based on different sets of 30 sensors for different wind attack angles, while for the mrDSPR method, the reconstruction is conducted by sharing one set of 30 sensors for all wind attack angles. The convergence relationship between the mrDMD reconstruction error and the number of sensors is studied in the W wind attack angle, an shown in Fig.14. In the end, the superiority of the mrDSPR method to the mrDMD-SDF method can be demonstrated by comparing the reconstruction accuracy based on sensors selected by these two methods.

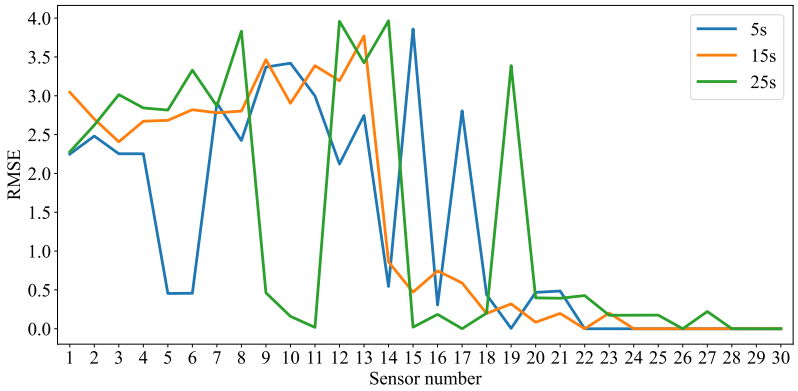


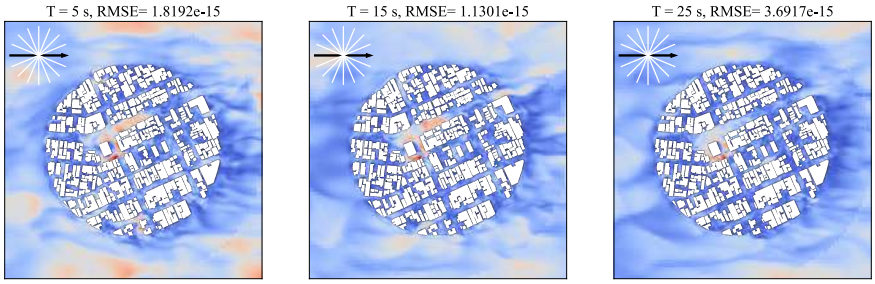
Figure 14: Convergence relationship between mrDMD reconstruction error and the number of sensors in W wind attack angle

Due to page limit, only the data of west wind direction is presented. The mrDMD-SDF reconstruction flow field, the mrDSPR reconstruction flow field, and the original wind flow

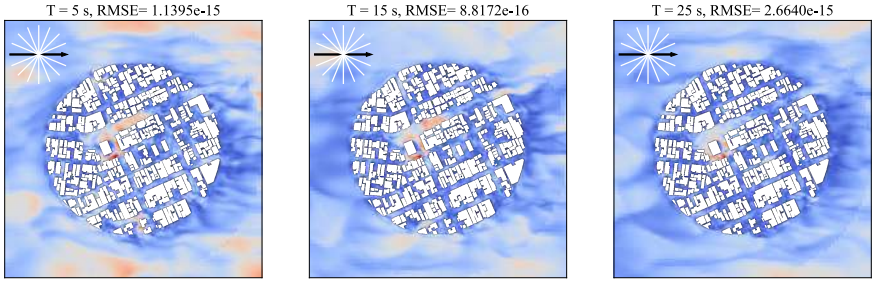
Figure 15: Optimal sensor selected procedure and results for the Niigata model case (Due to space constraints, only three wind attack angles data are shown).

field, at 5s, 15s and 25s instants are shown in Fig.16. Apparently, the reconstruction flow fields by these two methods agree well with the original wind flow field. It can be inferred that the sensor locations selected by both methods can capture dominant wind flow field information.

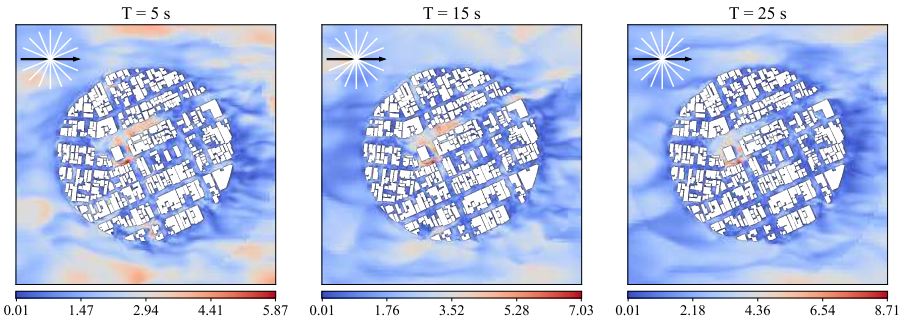
To further compare the reconstruction result of the two methods, Fig.17 presents the reconstruction errors of the two methods under every wind attack angle. Samples of 26 time instants (10% of the total data) are randomly selected and their average reconstruction errors at all wind attack angles are calculated. The result shows that for all wind attack angles, the mrDMD-SDF-PSO-RF method, which is the global optimal solution, has a little gap compared with the mrDMD-SDF method, which is the special solution for a single wind attack angle. But from the perspective of absolute value, the RMSE value near 0.04 is also



(a) mrDMD-SDF Reconstruction



(b) mrDSPR Reconstruction



(c) Ground Truth (wind velocity magnitude, m/s)

Figure 16: Comparison between real wind field data and different reconstructed data in western wind direction

a very small error, which is enough to prove the effectiveness of the method proposed in this study.

In summary, the mrDMD-SDF method selects the optimal sensor positions individually for each wind direction. In contrast, the mrDSPR method determines an invariable sensor position scheme for all wind directions. The low RMSE values of both methods indicate that they capture the flow field information very well, and also demonstrate the excellent generalization ability of the mrDSPR method for different wind directions.

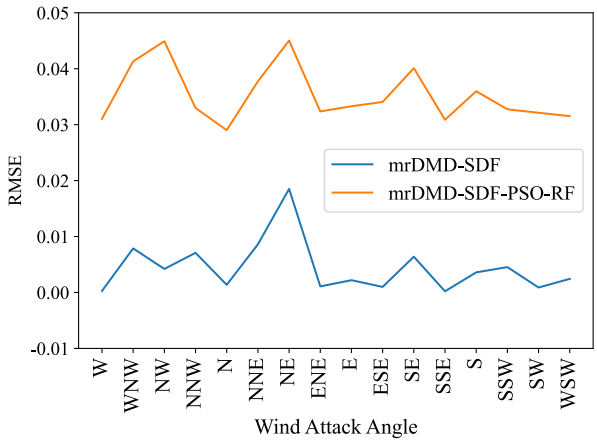


Figure 17: Comparison of mrDMD-SDF and mrDSPR reconstruction error

3.2. Case 2: wind pressure field of a tall building with experiment data

3.2.1. Case description

The mrDSPR method has been proven effective in selecting optimal sensor placement locations in an urban wind flow field. For wind-sensitive structures, real-time monitoring wind pressures on their surface are also important to ensure structural safety under action of strong wind. The effectiveness of the mrDSPR method on selecting optimal sensor placement locations on surface of wind-sensitive structures is tested on the Sanya Dayuecheng tower, the tallest building with a height of 210m in Sanya, China.

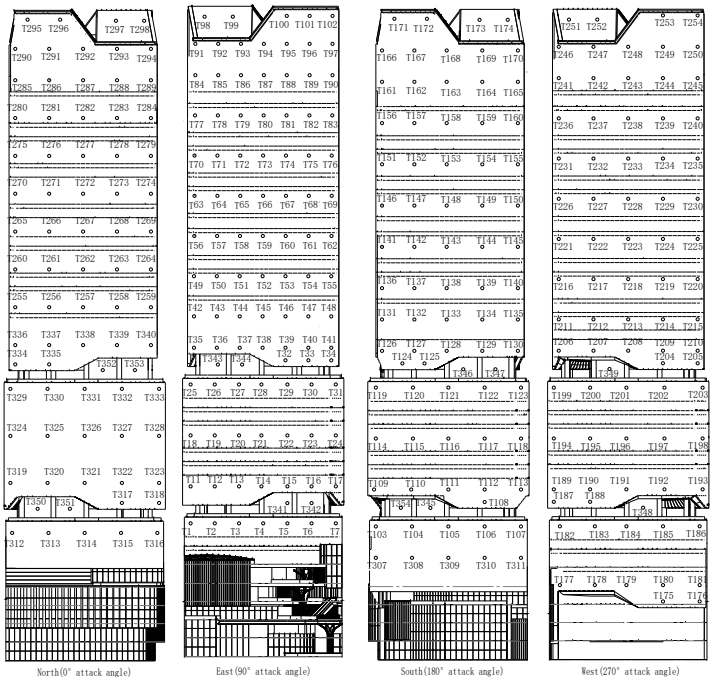


Figure 18: Pressure taps layout of the building model

(a) Layout diagram of Dayuecheng tower (b) Wind tunnel model and surrounding buildings (c) Wind attack angle

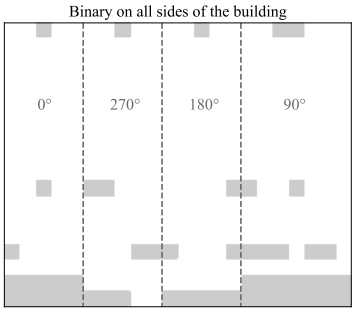
Figure 19: Dayuecheng building and the wind tunnel experiment setup

A wind tunnel pressure testing has been conducted on the Sanya Dayuecheng tower. The model length scale was 1:300. A total of 354 pressure taps were arranged on the model surface. 24 wind attack angles were tested from 0° to 345°. The overall wind tunnel test model is shown in Fig.19(b), and the layout of measuring points is shown in Fig.18. The maximum blocking ratio for all wind directions was 7.92%, not exceed the threshold of 8%. The sampling time of fluctuating pressure was 62 seconds, equivalent to the full-scale sampling time of 1 hour. The sampling frequency was 331Hz, which is a little higher than typical wind tunnel test sampling frequency range (100-300Hz) (Holmes, 1984) to ensure capture of fluctuating feature. Similarly, as mentioned in Sec.3.1.3, the main purpose of this paper is to propose a complete optimal sensors placement scheme, and the sampling frequency is not specially designed. The wind speed at the reference point was about 9.1m/s. The wind attack angles are defined in Fig.19(c).

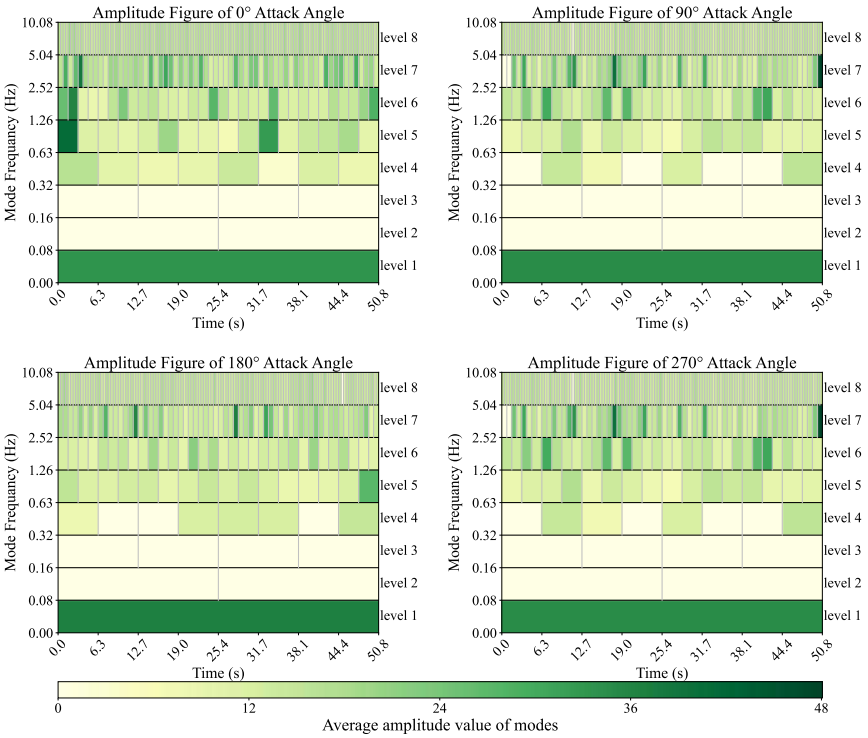
The surfaces of the building are irregular, resulting in unevenly distribution of pressure taps. Therefore, the same as removing invalid data inside the building in case 1, the mask data is used to find and remove invalid data at vacant positions on the building surface to avoid the influence of these data on the pressure field analysis. As shown in Fig.20(a), the grey part in the figure represents the region without sensors.

3.2.2. mrDMD procedure

As mentioned previously, the raw data with 2^n time steps is needed for the mrDMD analysis, and hence 16,384 time steps of data are selected for analyses in this section. To ensure the effectiveness of mrDMD analysis, a variety of frequency downsampling parameters were tested. The results show that the frequency range captured with 32 times downsampling is the widest. Therefore, the frequency downsampling factor is selected as 32, meaning that one step is taken every 32 time steps, and a total of 512 time instants are used for mrDMD analysis. The maps of the mrDMD modes for 4 wind attack angles of 0°, 90°, 180°, 270° are shown in Fig.20(b). Compared with Fig.9 (mrDMD modes map of case 1), mrDMD is still able to capture valid modes in different mode frequency ranges. The Vandermonde matrix values of different level and different order of DMD modes for zero wind direction are also



(a) Mask for single building data



(b) Maps of mrDMD modes for 4 wind attack angles

Figure 20: Mask data and amplitude map for the tall building case

shown in Fig.21(a-c), and the addition of the time-variation information of every layer and every order mode is shown in Fig.21(d).

Fig.24(a) shows the optimal sensors for wind attack angles of 0°, 90°, 180°, 270° selected by the mrDMD results and the QR pivoting method. Due to limited space, the results of other wind attack angles are not displayed graphically. The blue font represents the direction of the incoming flow. It is worth noting that the SDF method is unnecessary for this case, because the sensor can be placed almost anywhere on the building surface without causing a large impact on the surrounding environment. Therefore, the same number of 30 sensors are selected for each wind attack angle.

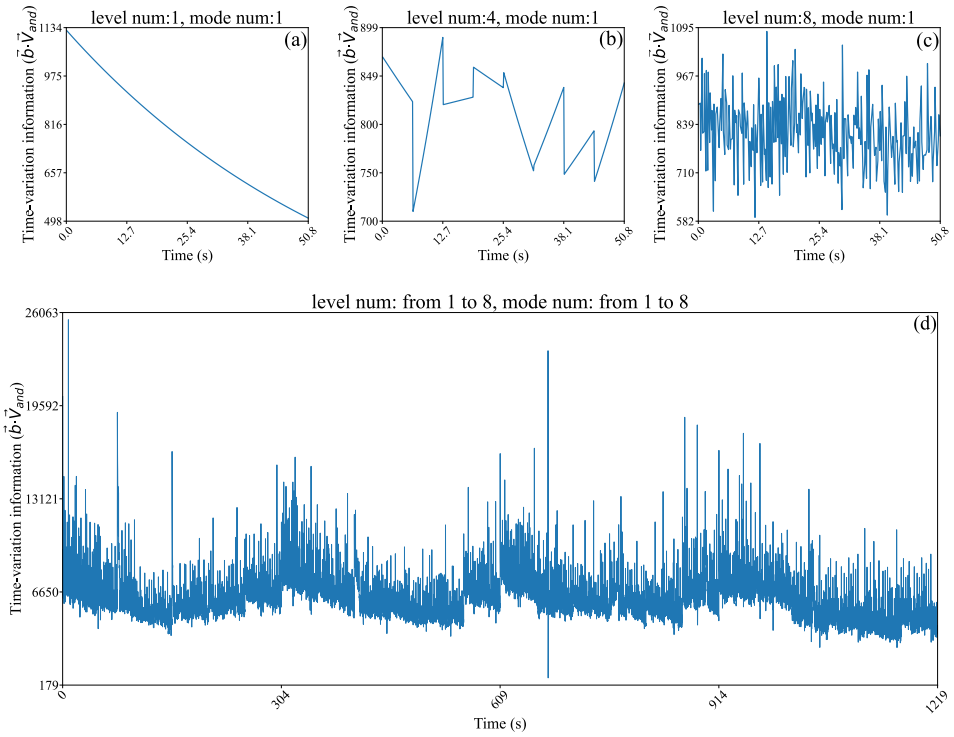


Figure 21: Time-variation information of different DMD level and order for single building with 0° wind attack angle: (a)Time-variation information for 1^{st} mode of level 1 (lowest level) mrDMD analysis; (b)Time-variation information for 1^{st} mode of level 4 (medium level) mrDMD analysis; (c)Time-variation information for 1^{st} mode of level 8 (highest level) mrDMD analysis; (d) Summation of time-variation information of every mrDMD layer and order.

3.2.3. PSO-RF selection

Although 30 wind pressure sensors are selected for each wind attack angle, it is impractical to integrate all 30 sensors from each wind attack angle like case 1. The integration of all sensors leads to the fact that almost all spatial points to be selected, which makes mrDMD analysis an unnecessary step. Therefore, the above 30 sensors will be used for the subsequent mrDMD reconstruction, while in the process of integrating the sensors from all wind attack angles, only the most important 10 sensor positions are selected under each wind attack angle. The result is shown in Fig.24(b).

After removing duplicate points, there are 107 sensor positions in total. The building surface pressure data of these sensors are extracted, including all wind attack angle cases (512 time steps for each wind attack angle case, thus 12,288 time steps for 24 wind attack angles in total), to form the input of the PSO-RF algorithm. The input data structure is similar to that of case 1. The mrDMD time evolution sequence demonstrated in Fig.21(d) is used as output to train the PSO-RF algorithm. The PSO-RF training process and fitting result of the time-variation information obtained after training are shown in Fig.22. The R^2 value of the fitting results is 0.915, better than the result shown in case 1, representing that the PSO-RF model is more competent for the feature importance ranking in this case. The

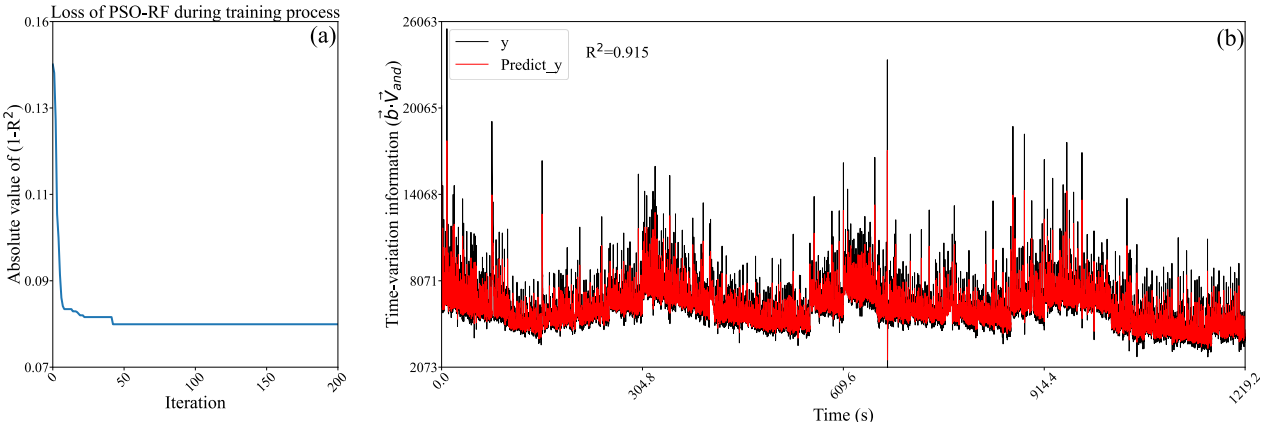


Figure 22: Performance of PSO-RF algorithm: (a) Performance of RF optimized by PSO; (b) PSO-RF fitting results for the tall building case.

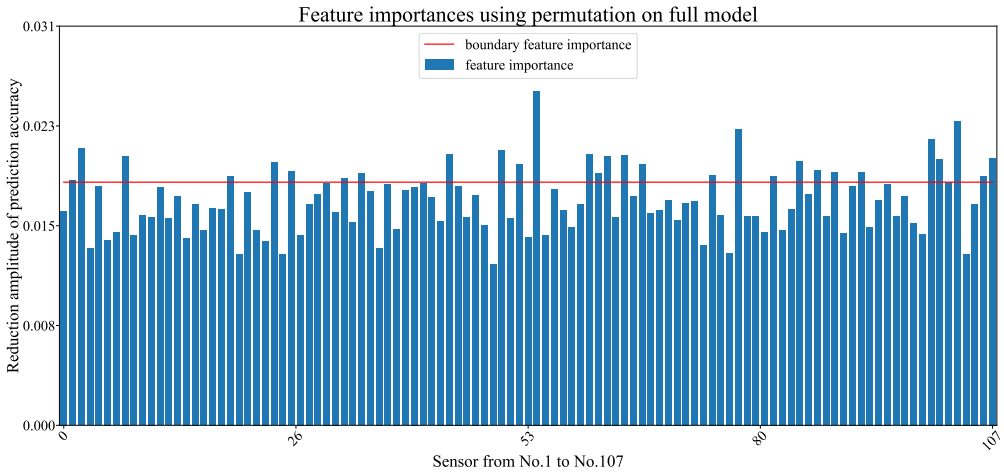


Figure 23: Sensor importance ranked by the PSO-RF algorithm. The y-axis of figure corresponds to *importance* i_j described in Sec.2.1.5.

feature importance ranking, i.e. the importance of each sensor ranked by PSO-RF algorithm is shown in Fig.23.

Compared with Fig.13 in case 1, it can be seen that the quantitative difference in sensor importance in this case is not significant, because the wind pressure distribution of a single building surface is much simpler than the wind flow field of a complex city block. Therefore, although the sensors arranged on the building surface can be ranked with different importance, the importance difference between sensors is not as large as that of sensors at different locations in a complex city block.

In this case, we also select top 30 important sensors, identical to the number of mrDMD-selected sensors for each wind attack angle, for easily comparing the two methods. The red line plotted in Fig.23 indicates the importance of the 30^{th} sensor, and it can also be seen that the number of 30 basically includes sensors of high importance. The selected 30 sensor

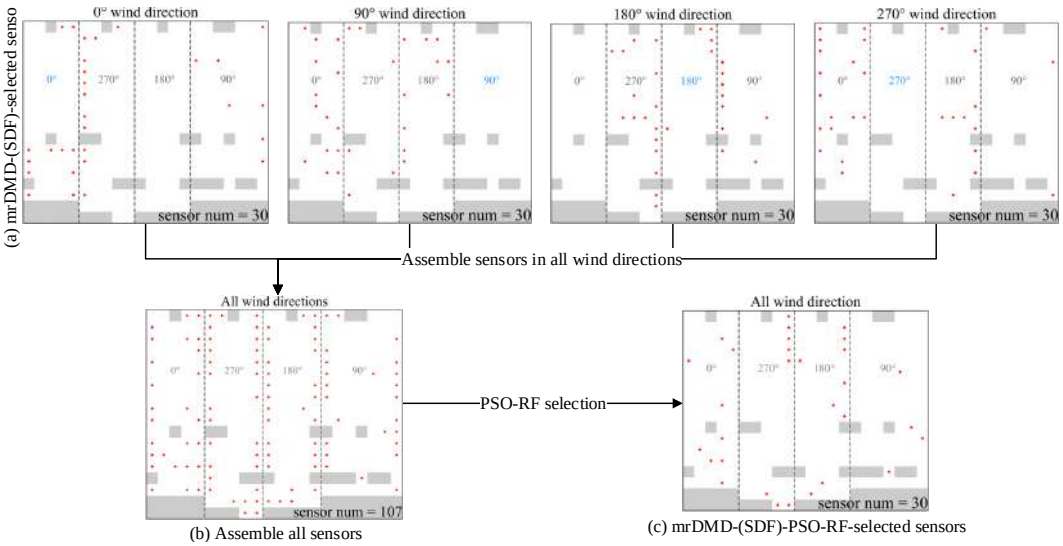


Figure 24: Optimal Sensor selected procedure and results for the tall building case (Due to space constraints, only four wind attack angles data are shown).

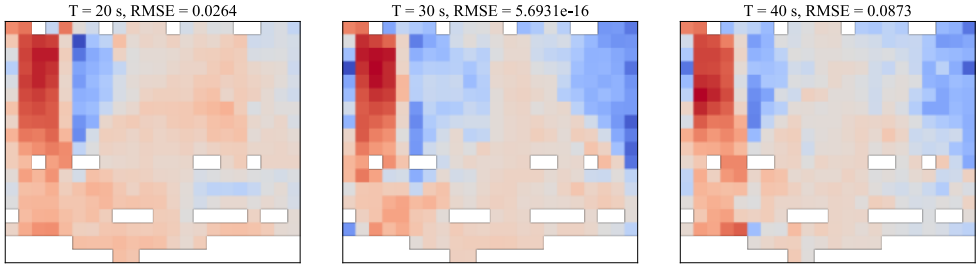
positions are shown in Fig.24(c).

3.2.4. Results and Discussions

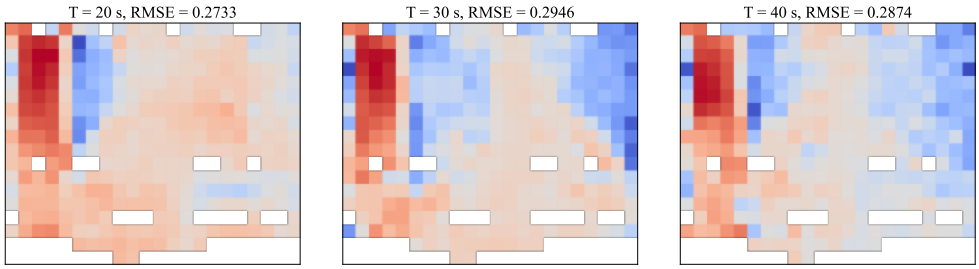
In this case, the mrDMD reconstruction error is also used to evaluate the effectiveness of the selected sensor positions. Similar to Section 3.1.6, 30 sensors are selected for each wind attack angle with the mrDMD-(SDF) method, and 30 sensors for all wind attack angles with the mrDMD-(SDF)-PSO-RF method. The convergence relationship between the mrDMD reconstruction error and the number of sensors is studied in the 0° wind attack angle, as shown in Fig.27. Afterwards, the mrDMD reconstruction is performed based on these sensor points.

Only the result of zero wind attack angle is presented due to page limit. The mrDMD-(SDF) reconstruction data (Fig.25(a)), the mrDMD-(SDF)-PSO-RF reconstruction data (Fig.25(b)), and the real wind field data (Fig.25(c)) at the 20s, 30s and 40s instants are given in Fig.25.

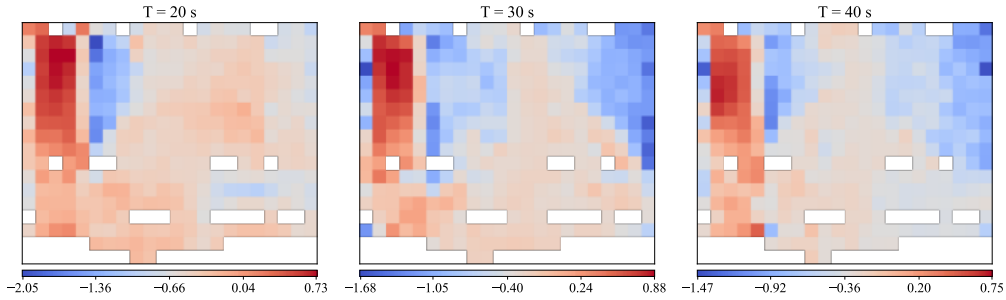
As shown in Fig.25, the sensor locations selected by both the mrDMD-(SDF) method and the mrDMD-(SDF)-PSO-RF method can provide sufficient wind pressure information for the mrDMD reconstruction. To further illustrate the reconstruction result of the two methods, Fig.26 presents the average reconstruction errors of the two methods at each wind attack angle using 50 randomly selected time instants (10% of the total data). When compared with the sensors selected by the mrDMD-(SDF) method, the reconstruction accuracy based on the sensors determined by the mrDMD-(SDF)-PSO-RF method is larger. When combined with the visualization results in Fig.25, it can be seen that the error range of 0.15-0.25 is still very small. Although such a reconstruction result is not as good as the optimal solution specifically designed for a single attack angle, it still captures the information of the flow field very effectively.



(a) mrDMD-(SDF) reconstruction at 0° wind direction



(b) mrDMD-(SDF)-PSO-RF reconstruction at 0° wind direction



(c) Wind tunnel pressure coefficient at 0° wind direction

Figure 25: Comparison between wind tunnel pressure data and reconstructed data at 0° wind direction for 3 time instants

4. Conclusions

In this study, we propose a framework for optimal sensor placement selection called mrDSPR. The framework is validated by using two case studies: wind flow field study around a city block, and wind pressure field study on a single building. A series of parameters including wind speed and direction, wind pressure, and high/low spatial resolution, are all taken into account. Hence, different situations that may occur in real engineering are studied. The mrDMD reconstruction error (in the form of RMSE) is set as the evaluation metric, which suggests how well the selected sensors represent the spatial-temporal characteristics of the data.

In case of wind flow field study around a city block, the experimental results show that using the mrDSPR method, 30 sensors are sufficient to capture the wind flow field characteristics of a complex urban area with size of $250m \times 250m$ in full scale. The reconstruction errors for different wind attack angles and different time steps are tested in this study. The

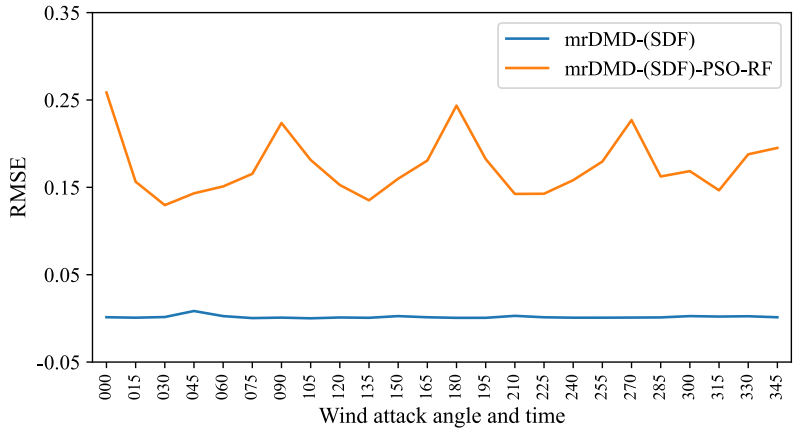


Figure 26: Comparing reconstruction error of mrDMD-(SDF) and mrDMD-(SDF)-PSO-RF

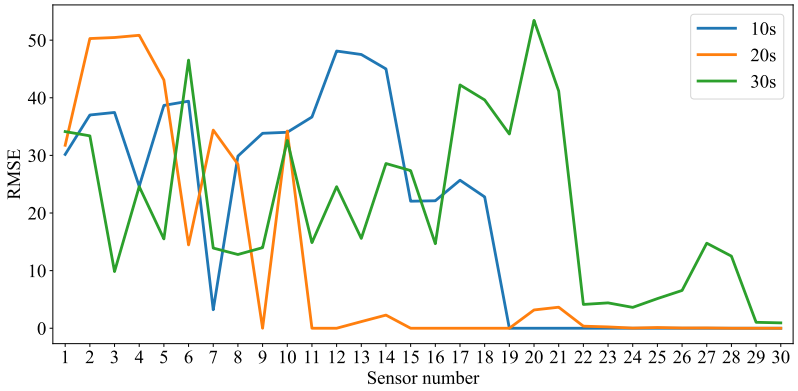


Figure 27: Convergence relationship between mrDMD reconstruction error and the number of sensors in 0° wind attack angle

average RMSE value of the specific optimal sensor position for each wind attack angle is between 0 and 0.02, and that of the global general optimal sensor position proposed in this paper is between 0.03 and 0.05. Despite the average error of the global solution being larger than that of the specific solution, 0.03 to 0.05 is a very small value in absolute terms, which is enough to prove that the global general solution can effectively capture the characteristics of the wind field.

In case of wind pressure field study on single building, the results show that 30 sensors are also sufficient to capture the wind pressure characteristics of a rectangular tall building surface. It is worth mentioning that in this case, although the reconstruction results are also very satisfactory, the main value range of the RMSE is in the order of 10^{-1}, which is larger than that in city block case. This result is mainly caused by the low spatial resolution of wind pressure data from wind tunnel testing, which makes mrDMD can only process decomposition on relatively rough data, and the obtained decomposition results naturally become rougher. On the other hand, the lower resolution only allows the sensor to be

arranged within a rough grid near the optimal spatial location, resulting in a long distance between the sensor and the true optimal location.

In summary, this study presents the strong capability of the mrDSPR framework to diverse conditions. Therefore, this novel framework provides a powerful tool to guide the optimal sensor placement scheme in wind engineering. For instance, it can provide suggestions for wind disaster detection in buildings. As a real building cannot be fully covered with sensors, a trade-off between detection effectiveness and economic cost must be made. Compared with the traditional sensor placement method, this novel framework can significantly reduce the number of sensors required in wind engineering, which not only reduces the monitoring cost, but also improves the efficiency of the sensors. Moreover, the method established in this article is a foundation for 3D optimal sensor arrangement scheme. At the same time, the method is a precursor to another our ongoing study. After the optimal sensors are selected, a deep learning model will be built. The deep learning model is able to achieve a real-time reconstruction for wind speed or pressure field within a study area (such as a city block or a building surface) based on the sparse real-time data acquired by these optimal-arranged sensors.

Acknowledgement

This study is supported by Artificial-intelligence-embedded wind flow field prediction for Guangdong-Hong Kong-Macao Bay Area (SGDX20210823103202018).

References

Bervida, M., Patruno, L., Stanič, S., & de Miranda, S. 2020, Journal of Wind Engineering and Industrial Aerodynamics, 196, 104040

Boyce, W. E., DiPrima, R. C., & Meade, D. B. 2021, Elementary differential equations and boundary value problems (John Wiley & Sons)

Breiman, L. 1996, Machine learning, 24, 123

Bruno, L., Fransos, D., Coste, N., & Bosco, A. 2010, Journal of Wind Engineering and Industrial Aerodynamics, 98, 263

Chen, C., & Gorlé, C. 2022, Building and Environment, 207, 108496

Chen, Y. L., & Wen, J. 2008, Building and Environment, 43, 1278

Cheng, J. C., Kwok, H. H., Li, A. T., Tong, J. C., & Lau, A. K. 2022, Building and Environment, 216, 108997

Cohen, K., Siegel, S., Wetlesen, D., Cameron, J., & Sick, A. 2004, Journal of Vibration and Control, 10, 1857

Du, W., Xing, Z., Li, M., et al. 2015a, ACM Transactions on Sensor Networks (TOSN), 11, 1

Du, Z., Xu, P., Jin, X., & Liu, Q. 2015b, Building and Environment, 85, 104

Eliades, D. G., Michaelides, M. P., Panayiotou, C. G., & Polycarpou, M. M. 2013, Building and environment, 63, 114

Fontanini, A. D., Vaidya, U., & Ganapathysubramanian, B. 2016, Building and Environment, 100, 145

Guo, A., Fang, Q., Bai, X., & Li, H. 2015, Journal of Bridge Engineering, 20, 04015012

Guo, X., Li, W., & Iorio, F. 2016, in Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 481–490

Hasanipanah, M., Armaghani, D. J., Amnieh, H. B., Majid, M. Z. A., & Tahir, M. M. D. 2016a, Neural Computing and Applications, 28, 1043

Hasanipanah, M., Shahnazar, A., Amnieh, H. B., & Armaghani, D. J. 2016b, Engineering with Computers, 33, 23

Heck, L. P., Olkin, J., & Naghshineh, K. 1998

Holmes, J. 1984, Journal of wind engineering and industrial aerodynamics, 17, 1

Hu, C., Jia, H., & Kikumoto, H. 2022, Building and Environment, 221, 109324

Inoba, R., Uchida, K., Iwasaki, Y., et al. 2022, in AIAA AVIATION 2022 Forum, 3606

Jiang, C., Vinuesa, R., Chen, R., et al. 2021, Physics of Fluids, 33, 055133

Kelp, M. M., Lin, S., Kutz, J. N., & Mickley, L. J. 2022, arXiv preprint arXiv:2201.01041

Kennedy, J., & Eberhart, R. 1995, in Proceedings of ICNN'95-international conference on neural networks, Vol. 4, IEEE, 1942–1948

Kim, J.-H., Avril, S., Duprey, A., & Favre, J.-P. 2012, Biomechanics and modeling in mechanobiology, 11, 841

Kutz, J. N., Brunton, S. L., Brunton, B. W., & Proctor, J. L. 2016, Dynamic mode decomposition: data-driven modeling of complex systems (SIAM)

Kutz, J. N., Fu, X., Brunton, S. L., & Erichson, N. B. 2015, in 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), IEEE, 921–929

Kyriacou, A., Michaelides, M. P., Eliades, D. G., Panayiotou, C. G., & Polycarpou, M. M. 2019, Building and Environment, 154, 348

Li, S., Laima, S., & Li, H. 2018a, Journal of Wind Engineering and Industrial Aerodynamics, 172, 196

Li, Y., Zou, C., Berecibar, M., et al. 2018b, Applied Energy, 232, 197

Lin, P., Ding, F., Hu, G., et al. 2022, Journal of Wind Engineering and Industrial Aerodynamics, 220, 104860

Lin, P., Hu, G., Li, C., et al. 2021, Journal of Wind Engineering and Industrial Aerodynamics, 211, 104549

Liu, X., & Zhai, Z. J. 2009, Building and Environment, 44, 2276

Liu, Z., & Li, X. 2022a, Building and Environment, 109276

—. 2022b, Building and Environment, 213, 108859

Malings, C., Pozzi, M., Klima, K., et al. 2018, Building and Environment, 141, 143

Manohar, K., Brunton, B. W., Kutz, J. N., & Brunton, S. L. 2018a, IEEE Control Systems Magazine, 38, 63

—. 2018b, IEEE Control Systems Magazine, 38, 63

Papadopoulou, M., Raphael, B., Smith, I. F., & Sekhar, C. 2015, in Proceedings of the 14th International Conference on Wind Engineering ICWE14, ICWE14

Ricci, M., Patruno, L., Miranda, S., & Ubertini, F. 2017, Computers & Fluids, 149

Sajedi, S., & Liang, X. 2021, Computer-Aided Civil and Infrastructure Engineering

Semaan, R. 2017, Computers & Fluids, 159, 167

Sharma, H., Vaidya, U., & Ganapathysubramanian, B. 2019, Building and Environment, 159, 106148

Sousa, J., García-Sánchez, C., & Gorlé, C. 2018, Building and Environment, 132, 282

Sreedharan, P., Sohn, M. D., Nazaroff, W. W., & Gadgil, A. J. 2011, Building and Environment, 46, 438

Tominaga, Y., Mochida, A., Yoshie, R., et al. 2008, Journal of wind engineering and industrial aerodynamics, 96, 1749

Verma, S., Papadimitriou, C., Lüthen, N., Arampatzis, G., & Koumoutsakos, P. 2020, Journal of Fluid Mechanics, 884

Wang, Q., Yan, L., Hu, G., et al. 2022, Physics of Fluids, 34, 081801

Wang, Z., Zhao, H., Lin, B., et al. 2015, Building and Environment, 94, 593

Weller, H. 2012, Monthly Weather Review, 140, 3220

Willcox, K. 2006, Computers & fluids, 35, 208

Yi, T.-H., Li, H.-N., & Gu, M. 2011, Mathematical Problems in Engineering, 2011

Yildirim, B., Chryssostomidis, C., & Karniadakis, G. 2009, Ocean Modelling, 27, 160

Zhang, Y., He, S., Gu, Z., et al. 2019, Indoor and Built Environment, 28, 1171

Zhou, Z.-H. 2012, Ensemble methods: foundations and algorithms (CRC press)