

Research Article

Wind Pressure Coefficients Zoning Method Based on an Unsupervised Learning Algorithm

Danyu Li , Bin Liu , and Yongfeng Cheng

China Electric Power Research Institute, Beijing, China

Correspondence should be addressed to Danyu Li; danyuli@qq.com and Bin Liu; liubinliubin@yeah.net

Received 4 September 2019; Revised 5 January 2020; Accepted 17 January 2020; Published 14 February 2020

Academic Editor: Luca Chiapponi

Copyright © 2020 Danyu Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Damage of the cladding structures usually occurs from the wind-sensitive part, which can cause the damaged conditions to obviously vary from different areas especially on a large roof surface. It is necessary to design optimization due to the difference of wind loads by defining more accurate wind pressure coefficient (WPC) zones according to the wind vulnerability analysis. The existing wind pressure coefficient zoning methods (WPCZM) have successfully been used to characterize the simple roof shapes. But the solutions for the complex and irregular roof shapes generally rely on the empirical judgment which is defective to the wind loading analysis. In this study, a classification concept for WPC values on the roof surface is presented based on the unsupervised learning algorithm, which is not limited by the roof geometry and can realize the multitype WPC zoning more accurately. As a typical unsupervised learning algorithm, an improved K -means clustering is proposed to develop a new WPCZM to verify the above concept. And a method to determine the optimal K -value is presented by using the K -means clustering test and clustering validity indices to overcome the difficulty of obtaining the cluster number in the traditional methods. As an example, the most unfavorable pressure and suction WPC zones are studied on a flat roof structure with single wind direction and full wind direction based on the data obtained from the wind tunnel test. As another example, the mean pressure coefficient zones are studied on a saddle roof structure under 0- and 45-degree wind direction based on the data obtained by the wind tunnel test. And the proposed WPCZM is illustrated and verified.

1. Introduction

During recent storms with strong wind conditions [1, 2], cladding was found to suffer significantly more damage than the main structure. Therefore, the importance of cladding safety under wind loads has attracted increasing attention. At the same time, with economic development and improved construction quality, “beauty” has gradually gained importance in building construction. “New,” “odd,” and “special” shaped roofs have been created widely, and these bring new challenges to wind-resistant cladding design. The damage on cladding often begins in wind-sensitive areas such as the corners, edges, and other parts of large-scale roof structures. Therefore, the different parts of the roof surface have different wind vulnerability, and in particular, the roof corners and edges are more sensitive to wind. To evaluate the wind loading of the cladding more objectively and rationally,

different zones must be defined according to the degree of wind vulnerability to design cladding for different wind loads.

Studies have extensively investigated wind pressure coefficient (WPC) zoning of traditional low-rise building roofs such as flat roofs, double-sloping roofs, and four-sloping roofs. Most of their results have already been incorporated into several national load codes [3–5]. WPC zoning is generally divided into edges, corners, and central zones according to the distribution of WPCs on the roof surface. The edge zone size is defined as the distance required for the maximum negative peak or mean pressure coefficient at the leading edge to reach 70% of its total value or the point where the flow is being reattached [6–8]. This gives three types of WPC zones: edge, corner, and interior. The zoning method adopted in the ASCE/SEI7-10, NBCC-2005, and AIJ-2004 codes is mainly aimed at rectangular planar

buildings with universal representative roof shapes. However, it does not provide objective WPC zoning results for irregular and complex roof shapes. For the roof shapes not covered by the code, most studies considered subjective WPC zoning according to the geometric features of the roof and characteristics of WPC distribution on the roof surface. Some researchers also proposed WPC zoning methods from different angles. Dong and Ye [9] proposed a WPC zoning method based on the separation bubble theory that is only applicable to the average WPC zone. The WPC zoning method proposed by Sun et al. [10] was based on the WPC Gaussian and that proposed by Ke et al. [11] was based on WPC correlation. However, these WPC zoning methods are either for a specific building geometry or for a specific wind direction and a specific WPC type. Thus, an objective and widely applicable WPC zoning method remains lacking.

To solve the problems of subjectivity, practicability, and applicability of the current wind pressure coefficient zoning method (WPCZM), this study proposes classifying WPC values on a roof surface based on unsupervised learning algorithm to realize WPC zoning of the roof cladding. Because classification is performed according to the WPC magnitude alone, the zoning is not limited by the roof geometry and the objective multitype WPC zoning can be achieved. Based on the classification, the dynamic clustering theory is introduced for WPC zoning and a fast WPCZM based on K -means clustering is proposed. The K -means clustering method requires the user to provide K -value in advance. However, in the actual WPC zoning, it is often impossible to determine the number of zoning classes in advance. Thus, the current method has been improved by the search range of the K -value in advance along with the use of multiple cluster validity indicators to determine the optimal K -value. The WPCZM is thus illustrated and verified by considering the most unfavorable WPC zoning on a flat roof surface for single and full wind directions as examples.

2. Wind Pressure Coefficient Zoning Method Based on Unsupervised Learning Algorithm

As an advanced and potential data analysis technique, machine learning has presented the general opportunities on solving some problems in civil engineering [12]. Machine learning techniques are increasingly applied in civil engineering [13, 14]. For example, Sohn et al. [15] proposed a baseline-free damage classification on the carbon fiber-reinforced polymer using cluster analysis. The supervised and unsupervised learning procedures are the two main learning methods. The unsupervised learning is presented with the input data only, which is that the learning network organizes itself internally to create categories by correlating the information available in the input data. Dynamic clustering is the most common unsupervised learning method to achieve that the sets of multiple objects are divided into several classes. Cluster analysis divides data into multiple classes without using predetermined decision boundaries [16]. The dynamic clustering method is a novel development from the static classification paradigm in which the model remains unchanged during the recognition

phase. Actually, many tasks are dynamic and the training and recognition phases are interlinked. The dynamic clustering should consider the incrementality of the learning method to devise the clustering model and the self-adaptation of the learned model. As a consequence, dynamic aspects of the clustering model to be learned can be captured by adaptation of the current model.

K -means is a common algorithm to be applied for clustering, which uses the Euclidean distance and assigns a data point into the cluster with the shortest distance from its centroid [17]. K -means clustering, which was originally proposed by [18], has been a key technique in the field of clustering owing to its advantages of applicability to large-scale data, which is easy to understand the results, and low computational requirements. Therefore, by using K -means clustering, a clustering-based WPCZM is proposed to illustrate the concept of wind pressure zoning based on unsupervised learning algorithm.

For the designers, the number of WPC clusters that have been divided according to the characteristics of the WPC distribution on the roof surface cannot provide accurate data. It can only provide an approximate range based on the past experience or intuition. In this case, the K -means clustering algorithm requires users to provide K -value in advance and has some limitations. Considering the zoning requirements of WPCs for roof claddings, the construction cost of claddings should be minimized while ensuring cladding safety. Therefore, the current method is improved by limiting the search range of the K -value in advance and using cluster validity indicator (CVI) to determine the optimal K -value.

Actually, there are some other clustering algorithms including Gaussian mixture models, subspace clustering, self-organized mapping, etc., which have mainly differences on how to compute the distance between data samples and how to assign a data point to a specific cluster. But the comparison of different clustering techniques has not been involved in the scope of this study.

2.1. Search Range of Cluster Number. The K -means clustering algorithm performs in two different parts. The first is to select a K -value, where K is the number of clusters. The second part is to consider the distance between each data point to the nearest center. So to use the K -means algorithm requires the number of clusters (K -Value) in the data to be prespecified. In other words, the performance of a clustering algorithm is affected by the chosen K -value. But it is more difficult to find an appropriate k -value which is generally a trial and error process by deciding what constitutes “correct” clustering subjectively. In this study, a method to determine the optimal K -value is proposed by using the K -means clustering test and clustering validity indices to overcome the difficulty of obtaining the cluster number in the traditional methods.

The number of WPC zones affects the accuracy of WPC description on the roof surfaces of claddings and the convenience of design and construction of these claddings. Too few zones will lead to inaccurate description of the WPC on

TABLE 1: Common clustering validity indices.

No.	Name	Formula	Optimal clustering number	References
1	CH index	$CH(q) = \sqrt{(\text{trace}(B_q)/(q-1))/(\text{trace}(W_q)/(n-q))}$	Maximum value of the index	Calinski and Harabasz [22]
2	C Index	$C \text{ index} = (S_w - S_{\min})/(S_{\max} - S_{\min}), S_{\min} \neq S_{\max}, C \text{ index} \in (0, 1)$	Minimum value of the index	Hubert and Levin [23]
3	Gamma index	$\text{Gamma} = (S(+) - S(-))/(S(+) + S(-))$	Maximum value of the index	Baker and Hubert [24]
4	DB index	$DB(q) = 1/q \sum_{k=1}^q \max_{k \neq l} ((\delta_k + \delta_l)/d_{kl})$	Minimum value of the index	Davies and Bouldin [25]
5	Hartigan index	$\text{Hartigan} = ((\text{trace}(W_q))/(\text{trace}(W_{q+1}) - 1)(n - q - 1)$	Maximum difference between hierarchy levels of the index	Hartigan [26]
6	Tau index	$\text{Tau} = (s(+) - s(-))/((N_t(N_t - 1)/2 - t)(N_t(N_t - 1)/2))^{1/2}$	Maximum value of the index	
7	Scott index	$\text{Scott} = n \log((\det(T))/(\det(W_q)))$	Maximum difference between hierarchy levels of the index	Scott and Symons [27]
8	Marriot index	$\text{Marriot} = q^2 \det(W_q)$	Maximum value of second differences between levels of the index	Marriot [28]
9	Friedman index	$\text{Friedman} = \text{trace}(W_q^{-1} B_q)$	Maximum difference between hierarchy levels of the index	Friedman and Rubin [29]
10	Rubin index	$\text{Rubin} = ((\det(T))/\det(W_q))$	Minimum value of second differences between levels of the index	Friedman and Rubin [29]
11	KL index	$KL(q) = \text{DIFF}_q/\text{DIFF}_{q+1} $ $\text{DIFF}_q = (q-1)^{2/p} \text{trace}(W_{q-1}) - q^{2/p} \text{trace}(W_q)$	Maximum value of the index	Krzanowski and Lai [30]
12	Silhouette index	$\text{Silhouette} = \sum_{i=1}^n S(i)/n, \text{Silhouette} \in [-1, 1]$ $S(i) = (b(i) - a(i))/(\max\{a(i); b(i)\})$ $a(i) = \sum_{j \in \{C_r, \dots\}} d_{ij}/n_r - 1, b(i) = \min_{s \neq r} \{d_{iC_s}\}$ $d_{iC_s} = \sum_{j \in C_s} d_{ij}/n_s$	Maximum value of the index	Rousseeuw [31]
13	Dunn index	$\text{Dunn} = (\min_{1 \leq i < j \leq q} d(C_i, C_j))/(\max_{1 \leq k \leq q} \text{diam}(C_k))$ $d(C_i, C_j) = \min_{x \in C_i, y \in C_j} d(x, y)$ $\text{diam}(C_k) = \max_{x, y \in C_k} d(x, y)$ $\text{SD index}(q) = \alpha \text{Scat}(q) + \text{Dis}(q)$ $\text{Scat}(q) = (1/q \sum_{k=1}^q \ \sigma^{(k)}\)/(\ \sigma\)$ $\sigma = (\text{VAR}(V_1), \text{VAR}(V_2), \dots, \text{VAR}(V_p))$ $\sigma^{(k)} = (\text{VAR}(V_1^{(k)}), \text{VAR}(V_2^{(k)}), \dots, \text{VAR}(V_p^{(k)}))$	Maximum value of the index	Dunn [32]
14	SD index	$\text{Dis}(q) = D_{\max}/D_{\min} \sum_{k=1}^q (\sum_{z=1}^q \ c_k - c_z\)^{p-1}$ $D_{\max} = \max(\ c_k - c_z\) \forall k, z \in \{1, 2, 3, \dots, q\}$ $D_{\min} = \min(\ c_k - c_z\) \forall k, z \in \{1, 2, 3, \dots, q\}$ $\alpha = \text{Dis}(q_{\max})$	Minimum value of the index	Halkidi et al. [33]

the roof surface, and too many zones will result in inconvenient design and construction. The maximum number of zoning classes k_{\max} for WPC zoning of the roof is less than or equal to the root of the number of measured points n on the entire roof; that is, $k_{\max} \leq \sqrt{n}$ [19, 20]. Pal and Bezdek [19] noted that there are few theoretical deductions for the selection of k_{\max} ; however, many researchers use this rule in practice according to experiences. Yang et al. [21] proved the rationality of the abovementioned empirical rules from different perspectives.

2.2. Clustering Validity Indices (CVI). WPC zoning of a roof is aimed at creating a clear boundary between zones and fully separating zones with larger WPC values from those with smaller ones. The wind loads on these zones are assigned and designed separately to increase the convenience of design and construction and to ensure the safety. This requires maximizing the distance between different classes and

minimizing the distance within each class during clustering. Many clustering validity indices can satisfy these requirements. Table 1 summarizes 14 clustering validity indices that can reflect intra- and interclass information and also consider the geometric characteristics, statistical characteristics, similarities, and amount of data.

In this study, the selected CVI should include two parts. The first is to evaluate the differences between different zones. The second is to evaluate the internal consistency in each of the zones. At the same time, the number of zones should meet the engineering requirements. And a single representative value can fully represent the wind load in a single zone in order to meet the engineering requirements. With the higher zone number, the single value can represent the wind load more accurately for a single zone. The structural design will be safer with the increasing of zone number. And with the lower zone number, the discreteness of wind loading on a single zone is greater. Although the design and construction will be convenient,

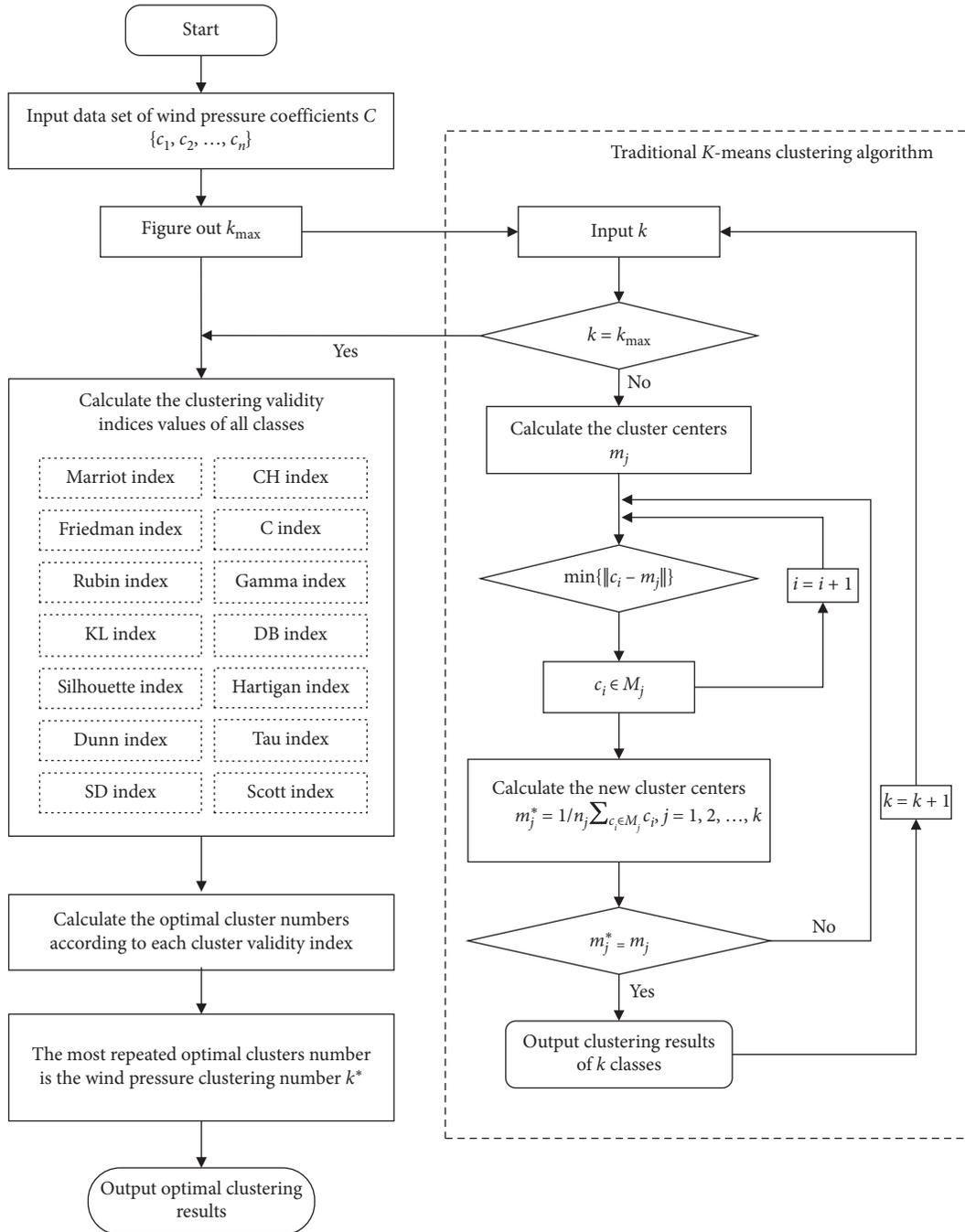


FIGURE 1: The flowchart of improved K-means clustering zoning algorithm for wind pressure coefficient.

the decreasing of zone number may make the designing unsafe due to the more conservative value of a single zone. As a result, an optimal wind pressure zoning should be obtained.

The optimal wind pressure zoning must present obvious differences between different zones and have a high consistency in each of the zones. And in the engineering application, it should reduce the difficulty of design and construction under the premise to fully ensure the structural safety. And a single representative value can fully represent the wind load in a single zone in order to meet the engineering requirements. The clustering validity

indices should be effective to determine the optimal wind pressure zoning.

The variables in Table 1 are described as follows.

n is the number of observations, p is the number of variables, and q is the number of clusters.

$X = \{x_{ij}\}$, $i = 1, 2, \dots, n$; $j = 1, 2, \dots, p$, is the $n \times p$ data matrix of p variables measured on n independent observations, \bar{X} is the $q \times p$ matrix of cluster means, and \bar{x} is the centroid of data matrix X .

n_k is the number of objects in cluster C_k , and c_k is the centroid of cluster C_k and x_i is the dimensional vector of observations of the i th object in cluster C_k .

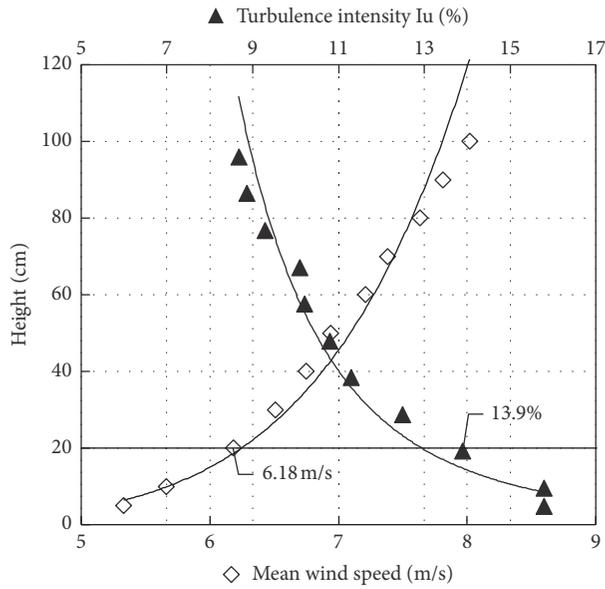


FIGURE 2: Wind field in wind tunnel test.

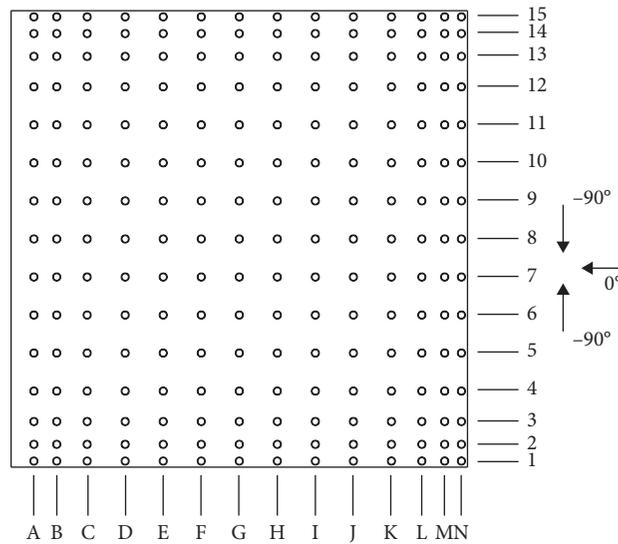


FIGURE 3: Pressure taps on flat roof.

TABLE 2: Optimal number of clusters.

Index name	0° pressure		0° suction		-45° pressure		-45° suction	
	k^*	Index value	k^*	Index value	k^*	Index value	k^*	Index value
CH	11	1121.14	14	1444.45	13	964.93	10	848.55
C	5	0.16	5	0.10	14	0.11	14	0.05
Gamma	2	0.97	14	0.97	14	0.96	8	0.98
DB	2	0.44	13	0.54	11	0.53	7	0.49
Hartigan	3	711.86	5	204.74	3	606.35	5	221.68
Tau	2	4847.49	3	4966.29	3	4240.27	3	5006.18
Scott	3	137.61	3	116.75	3	163.31	3	176.86
Marriot	13	3.37	4	42.97	8	5.84	5	106.16
Friedman	13	238.94	14	155.65	11	65.02	8	46.60
Rubin	13	-221.62	11	-31.08	11	-45.40	8	-27.08
KL	12	84.52	7	680.41	6	8.27	5	25.69
Silhouette	2	0.75	2	0.69	2	0.60	3	0.68
Dunn	10	0.06	7	0.02	13	0.02	5	0.04
SD	4	29.22	3	7.95	4	25.80	6	9.33

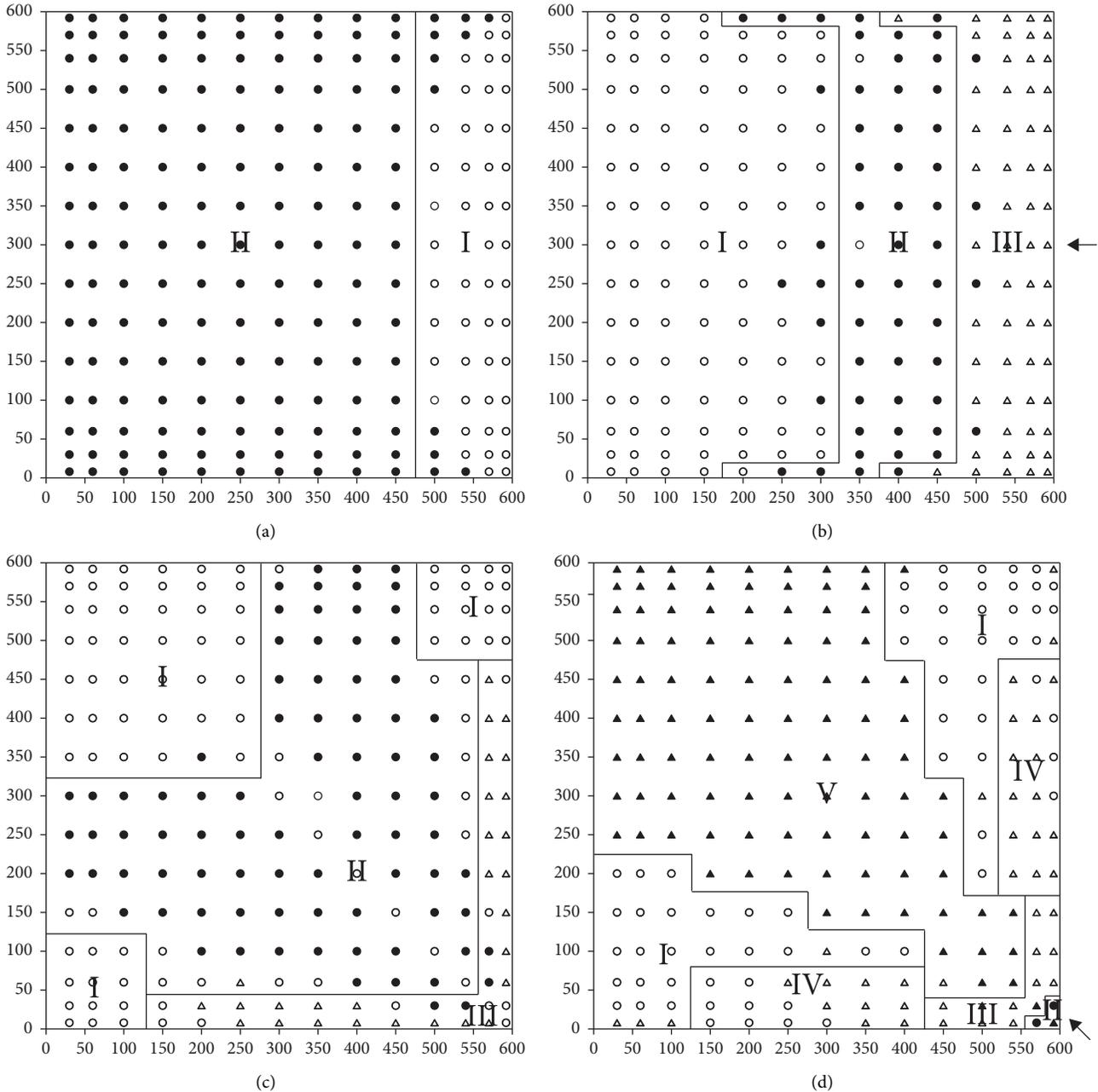


FIGURE 4: Optimal zoning of the most unfavorable WPC. (a) 0° pressure, (b) 0° suction, (c) -45° pressure, and (d) -45° suction.

$$\|x\| = (x^T x)^{1/2}.$$

$W_q = \sum_{k=1}^q \sum_{i \in C_k} (x_i - c_k)(x_i - c_k)^T$ is the within-group dispersion matrix for data clustered into q clusters.

$B_q = \sum_{k=1}^q n_k (c_k - \bar{x})(c_k - \bar{x})^T$ is the between-group dispersion matrix for data clustered into q clusters.

$N_t = (n(n-1))/2$ is the total number of pairs of observations in the data set.

$N_w = \sum_{k=1}^q (n_k(n_k-1))/2$ is the total number of pairs of observations belonging to the same cluster.

$N_b = N_t - N_w$ is the total number of pairs of observations belonging to different clusters.

$S_w = \sum_{k=1}^q \sum_{\substack{i, j \in C_k \\ i < j}} d(x_i, x_j)$ is the sum of the internal cluster distances.

$S_b = \sum_{k=1}^q \sum_{l=k+1}^q \sum_{\substack{i \in C_k \\ j \in C_l}} d(x_i, x_j)$ is the sum of the distances between different clusters.

S_{\min} is the sum of the N_w smallest distances between all the pairs of points in the entire data set (there are N_t such pairs), and S_{\max} is the sum of the N_w largest distances between all the pairs of points in the entire data set.

$s(+)$ is the number of concordant comparisons and $s(-)$ is the number of discordant comparisons.

2.3. Wind Pressure Coefficient Zoning Method Based on Improved K-Means Clustering Algorithm. A WPCZM based on an improved K-means clustering algorithm has been

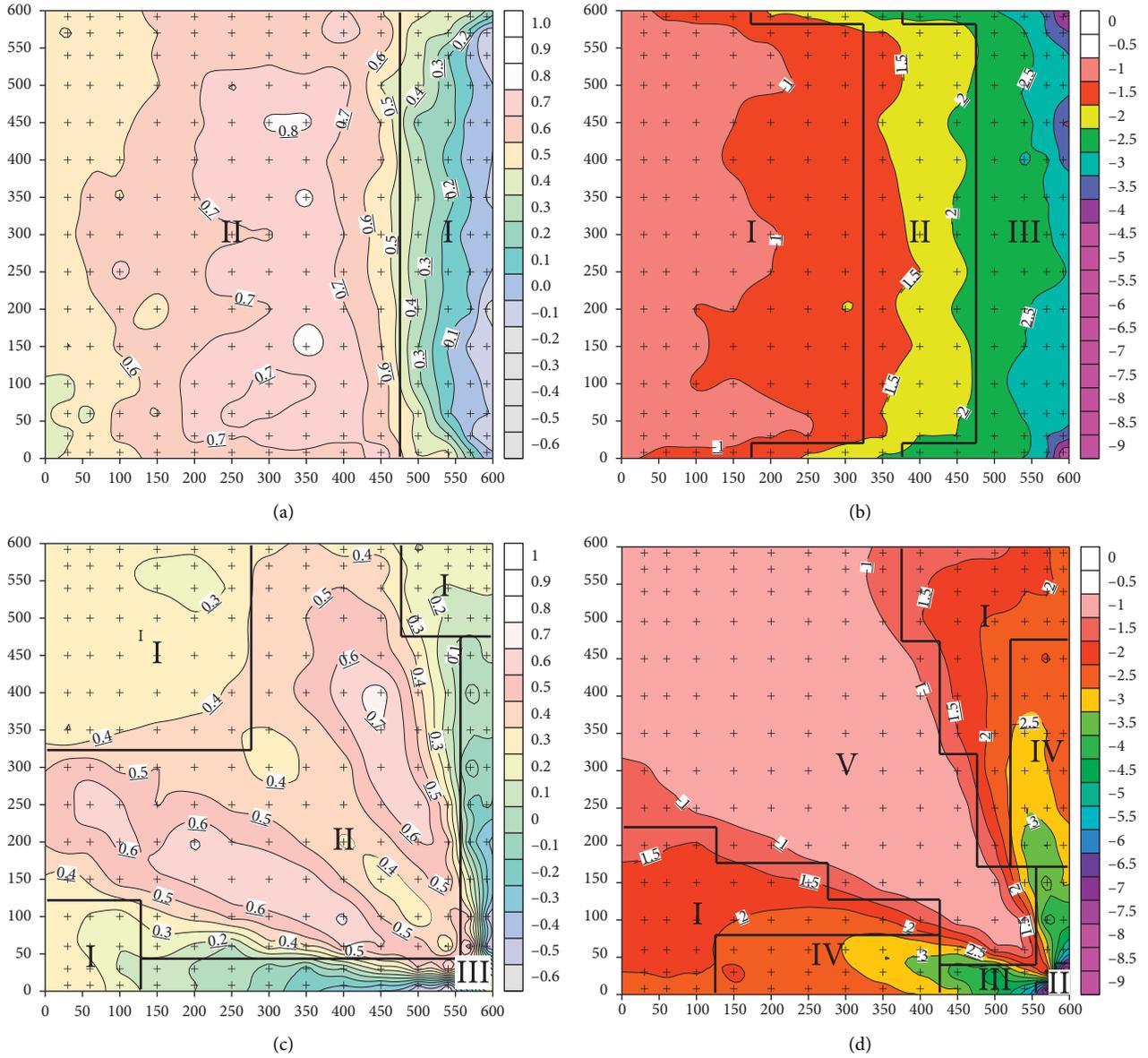


FIGURE 5: WPC distribution and zoning. (a) 0° pressure, (b) 0° suction, (c) -45° pressure, and (d) -45° suction.

proposed. The existing K -means clustering has limitations in that it requires a given K -value (cluster number) in advance. To reduce the difficulty faced by engineers in setting the K -value, a method to determine the optimal K -value is proposed in this study. First, the maximum K -value is limited, and a K -means clustering test is performed with all values in the range $[2, k_{\max}]$. Second, by using the clustering validity indices, the optimal K -value is obtained based on the test results in the first step.

The WPCZM based on the improved K -means clustering algorithm has the following steps.

Step 1. Input data set of WPCs for measuring points $C = \{c_i | i = 1, 2, \dots, n\}$, where c_i is the WPC at the i^{th} measuring point and n is the number of WPC sets.

Step 2. Calculate and determine the upper bound of the cluster numbers according to the empirical rule $k_{\max} \leq \sqrt{n}$.

Step 3. Clustering is performed for all cluster numbers using the K -means algorithm as follows.

- ① K objects are randomly selected from C as the initial cluster centers m_1, m_2, \dots, m_j .
- ② Calculate the distance of the WPC from each measuring point to each cluster center individually and classify the WPC of each measuring point according to the principle of closest distance, that is, $j, l = 1, 2, \dots, k, l \neq j$ for the corresponding categories.
- ③ Calculate the distance of the WPC from each measuring point to each cluster center individually. According to the principle of closest distance, which is given as $\|c_i - m_j\| < \|c_i - m_l\|, j, l = 1, 2, \dots, k, l \neq j$, the WPC of each measuring point falls into the corresponding cluster.
- ④ Calculate the central points of the newly formed clusters $m_1^*, m_2^*, \dots, m_j^*$.

TABLE 3: The most unfavorable pressure—the clustering validity indices in full wind direction.

k	CH	C	Gamma	DB	Hartigan	Tau	Scott	Marriot	Friedman	Rubin	KL	Silhouette	Dunn	SD
2	-42	0.250	0.659	0.755	172	1176	595	1.84	141	142	0.967	0.548	0.019	44
3	34	0.237	0.810	0.687	25	1343	675	2.12	277	278	0.198	0.532	0.022	36
4	36	0.216	0.859	0.633	10	1207	735	2.30	455	456	0.725	0.539	0.022	42
5	31	0.199	0.865	0.609	325	1133	761	2.89	565	566	0.488	0.528	0.019	45
6	159	0.284	0.927	0.578	122	1110	834	2.26	1041	1042	4.814	0.545	0.036	44
7	293	0.228	0.945	0.567	0	958	876	2.16	1483	1484	1.700	0.562	0.023	58
8	248	0.224	0.945	0.580	23	923	891	2.51	1673	1674	5.104	0.553	0.045	64
9	264	0.189	0.947	0.578	36	770	919	2.50	2119	2120	0.652	0.557	0.036	85
10	311	0.183	0.959	0.593	11	683	950	2.39	2740	2741	0.749	0.542	0.039	89

TABLE 4: The most unfavorable suction—the clustering validity indices in full wind direction.

k	CH	C	Gamma	DB	Hartigan	Tau	Scott	Marriot	Friedman	Rubin	KL	Silhouette	Dunn	SD index
2	-25	0.195	0.876	0.705	109	1195	358	287	18.7	19.7	4.3	0.688	0.112	13.5
3	30	0.129	0.789	0.698	614	1372	409	422	29.2	30.2	0.2	0.523	0.011	10.3
4	327	0.221	0.870	0.528	61	1393	554	224	100.3	101.3	113.0	0.586	0.021	4.6
5	387	0.149	0.950	0.427	46	1320	654	152	232.5	233.5	0.8	0.654	0.021	4.8
6	439	0.126	0.924	0.496	24	1047	684	170	299.0	300.0	0.2	0.597	0.021	8.1
7	443	0.110	0.921	0.523	45	891	706	193	358.0	359.0	1.6	0.566	0.013	11.6
8	533	0.089	0.941	0.515	7	792	738	193	467.8	468.8	0.4	0.590	0.013	14.4
9	493	0.085	0.944	0.478	10	763	748	226	506.9	507.9	1.2	0.598	0.013	14.7
10	476	0.071	0.951	0.464	26	648	765	241	586.6	587.6	5.0	0.628	0.011	18.7

TABLE 5: Optimal number of clusters.

	Pressure in full wind direction		Suction in full wind direction	
	k^*	Index value	k^*	Index value
CH	10	311.38	8	533.44
C index	10	0.18	10	0.07
Gamma	10	0.96	10	0.95
DB	7	0.57	5	0.43
Hartigan	5	314.54	4	552.76
Tau	3	1342.89	4	1393.11
Scott	3	80.39	4	145.15
Marriot	6	0.53	4	125.97
Friedman	10	621.00	5	132.20
Rubin	7	-251.46	8	-70.59
KL	8	5.10	4	112.97
Silhouette	7	0.56	2	0.69
Dunn	8	0.04	2	0.11
SD index	3	36.31	4	4.64

$$m_j^* = \frac{1}{n_j} \sum_{c_i \in M_j} c_i, \quad j = 1, 2, \dots, k. \quad (1)$$

⑤ According to the new central positions, recalculate the distances between the WPC of each measuring point and the new clustering central point and reclassify.

⑥ Repeat step ④. The algorithm converges and the program terminates when the centers of the newly formed classes are equal to the centers of the previous classes. To prevent the termination conditions from failing to satisfy the infinite cycle, the maximum number of iterations is set as 1,000,000 when the algorithm is executed.

⑦ Output the final clustering results, that is, K classes \bar{M}_j ($j = 1, 2, \dots, k$), and various clustering centers \bar{m}_j ($j = 1, 2, \dots, k$).

Step 4. Calculate the 14 clustering validity indices under different cluster numbers K .

Step 5. Calculate the optimal number of clusters determined by the 14 validity indicators. The most repeated optimal cluster number is the WPC classification number k^* . And the corresponding WPC classification results are recorded.

Step 6. Output classification number k^* , k^* classes M_j^* ($j = 1, 2, \dots, k$), and various centers m_j^* ($j = 1, 2, \dots, k$).

Figure 1 shows the flowchart of the above method.

3. Definition of Wind Pressure Coefficient (WPC)

The pressure coefficient is a dimensionless number which describes the relative pressures throughout a flow field in fluid dynamics. The pressure coefficient is used in aerodynamics. Every point in a fluid flow field has its own unique pressure coefficient.

In many situations in aerodynamics, the pressure coefficient at a point near a body is independent of body size. The pressure coefficient of an engineering model can be tested in a wind tunnel and determined at critical locations around the model. These pressure coefficients can be used to predict the fluid pressure at those critical locations around full-size structures.

The expression of pressure coefficient is obtained as follows:

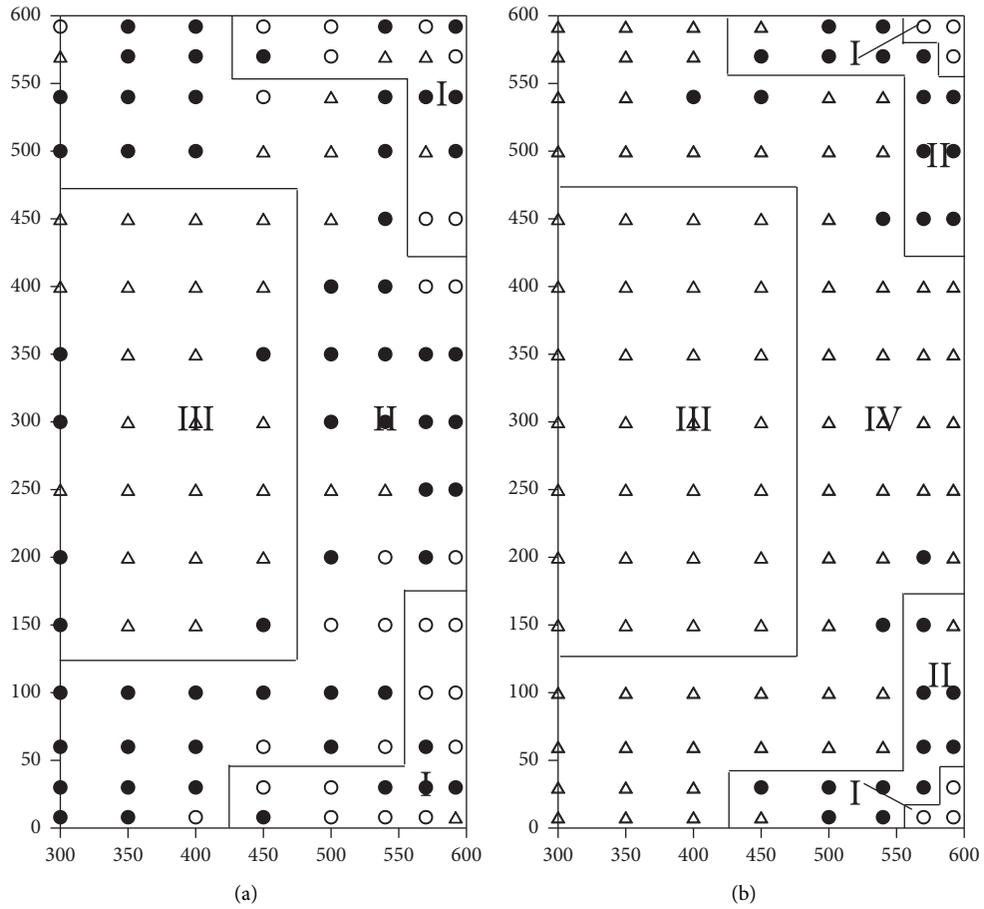


FIGURE 6: Optimal zoning of the most unfavorable pressure coefficient. (a) Pressure, (b) suction.

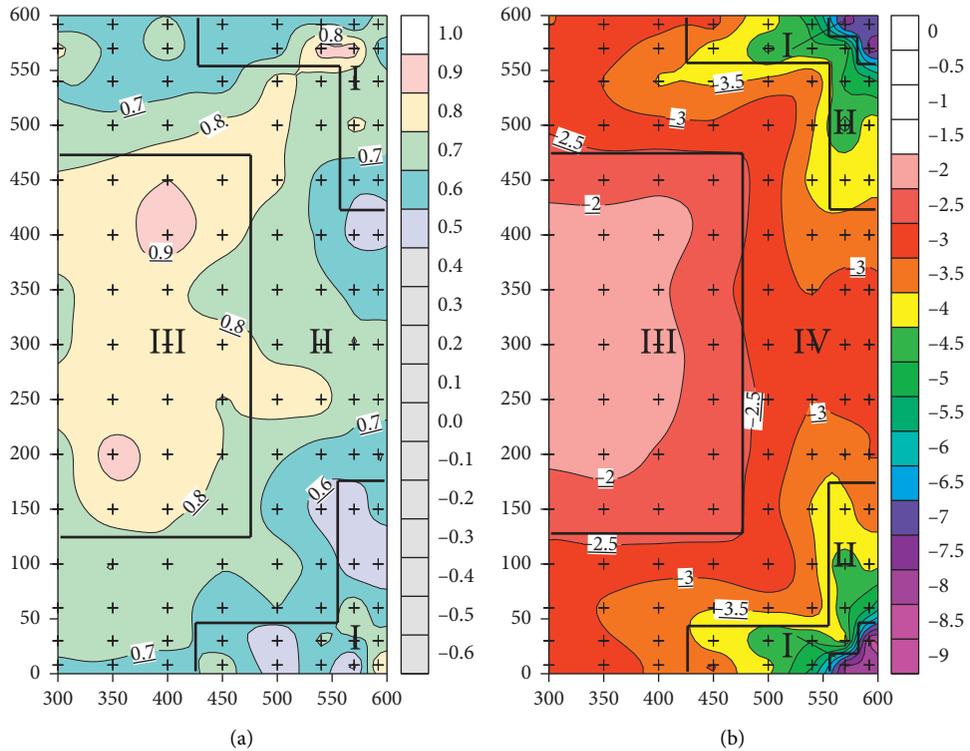


FIGURE 7: Wind pressure coefficient distribution and zoning. (a) Pressure. (b) Suction.

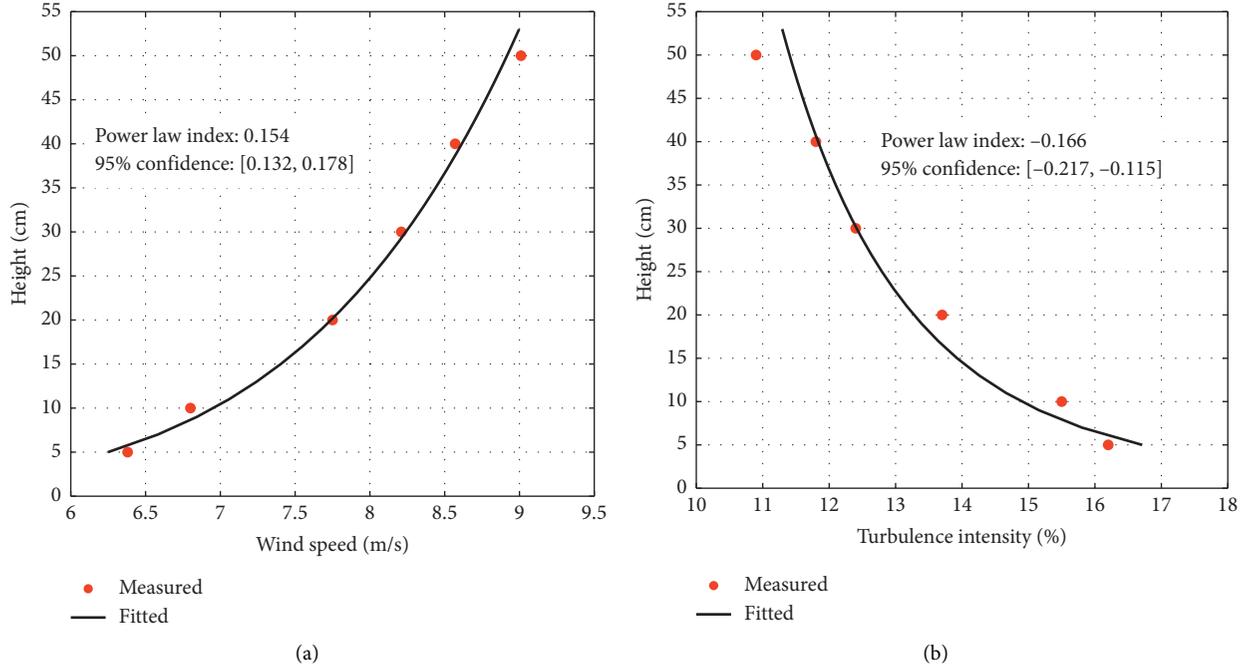


FIGURE 8: Wind field in wind tunnel test. (a) Mean wind speed profile. (b) Turbulence profile.

$$C_{pi} = \frac{(P_i(t) - P_{\infty})}{0.5\rho\bar{v}_{zT}^2}, \quad (2)$$

where $P_i(t)$ is the measured time series of WPC at the i^{th} measurement point, P_{∞} is the static pressure at the reference height, ρ is the air density, and \bar{v}_{zT} is the mean wind speed at the reference height.

The mean WPCs \bar{C}_p can be obtained by averaging one full sample cycle of WPCs (corresponding to 10 min length for the full scale). The peak of WPC can be calculated by averaging 10 peak values of 10 different samples.

$$\bar{C} = \frac{\sum_{i=1}^N C_{pi}}{N}, \quad (3)$$

$$\hat{C} = \sum_{i=1}^{10} \hat{C}_{pi},$$

where N is the size of the sample and \hat{C}_{pi} is the extreme pressure coefficient.

4. Examples of Simple Structures

A flat roof structure under two working conditions has been used as an example to illustrate and verify this zoning method. Two types of incoming flow conditions are considered in this study, single wind direction and full wind direction. Single wind direction inflow means that a building is only affected by a single wind direction. The angle between the building (rectangular plane) and the wind direction can be decomposed into the normal and the oblique directions. Full wind direction inflow means that a building is affected by wind from full directions (360°). The maximum or

minimum value is taken as the most unfavorable pressure or suction of the measuring point.

4.1. Wind Tunnel Tests. Wind tunnel tests for the flat roof structures were conducted in a wind tunnel at Beijing Jiaotong University by assuming that the structure is located in open rural terrain and the surface roughness coefficient is $\alpha = 0.15$. Figure 2 shows the mean wind speed and turbulence intensity profiles. The flat roof model has dimensions of $60 \text{ cm} \times 60 \text{ cm}$ on a plane with height (H) of 20 cm. There are 210 pressure taps on the top of the flat roof model, as shown in Figure 3. The mean wind speed at 20 cm height in this experiment is 4.36 m/s. The geometric scale, speed scale, and time scale of the wind test are 1/200, 3.125/20, and 1/31.25, respectively. The sampling frequency is 312.5 Hz, and the sampling time corresponding to 10 min in the full scale is 19.2 s.

4.2. WPC Zoning under Single Wind Direction Condition. The maximum clustering number k_{\max} is the root of the number of measuring points n on the flat roof to be zoned. In the present case, $n = 210$ gives $k_{\max} = \sqrt{n} = \sqrt{210} \approx 14.5$; that is, $k_{\max} = 14$. By using the traditional K-means algorithm to achieve all cluster numbers $K \in [2, k_{\max}]$ and clustering flat roofs with the most unfavorable pressure (maximum WPC, positive pressure) and suction (minimum WPC, negative pressure) at 0° and -45° , we calculate several clustering validity indices for different cluster numbers. Table 2 lists the optimal cluster numbers determined with each index.

The most unfavorable pressure is obtained at 0° . Four, three, and two clustering validity indices give optimal cluster numbers of 2, 13, and 3, respectively. The other five

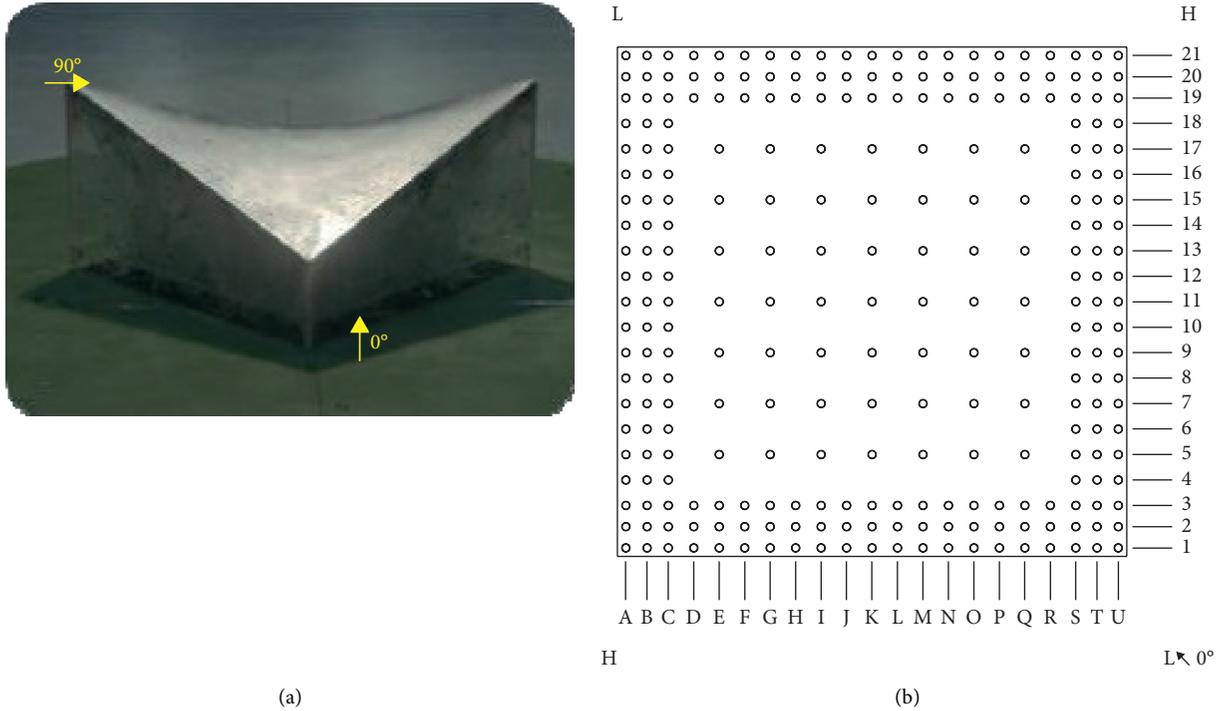


FIGURE 9: Pressure taps on saddle roof. (a) Wind tunnel test model. (b) Pressure taps.

TABLE 6: Mean WPC zoning—the clustering validity indices in full wind direction.

k	CH	C	Gamma	DB	Hartigan	Tau	Scott	Marriot	Friedman	Rubin	KL	Silhouette	Dunn	SD index
2	379.9	0.210	0.944	0.508	364.7	7304.0	732.4	54.4	14.9	15.9	5.844	0.735	0.017	35.6
3	633.2	0.208	0.960	0.507	93.3	8222.7	932.4	57.6	32.7	33.7	0.195	0.701	0.061	16.5
4	601.3	0.209	0.814	0.677	89.9	6255.4	1052.7	65.0	52.1	53.1	1.463	0.523	0.011	19.7
5	626.4	0.177	0.874	0.673	98.6	6124.9	1147.8	71.0	75.0	76.0	3.164	0.545	0.023	19.7
6	708.1	0.160	0.896	0.633	5.7	5448.4	1245.2	70.7	108.8	109.8	0.127	0.556	0.016	19.5
7	601.8	0.133	0.921	0.544	80.1	5160.2	1317.6	73.3	143.3	144.3	72.547	0.583	0.009	24.6
8	684.8	0.118	0.920	0.531	5.5	4445.5	1357.1	82.5	166.5	167.5	0.014	0.580	0.005	32.4
9	610.3	0.117	0.895	0.574	17.6	3704.2	1370.0	99.4	174.9	175.9	1.243	0.534	0.005	50.6
10	579.4	0.108	0.903	0.592	328.6	3402.9	1384.6	116.1	184.8	185.8	0.285	0.539	0.010	52.0
11	1221.5	0.128	0.934	0.566	10.9	3398.0	1568.6	70.2	371.2	372.2	6.320	0.551	0.015	46.8
12	1154.6	0.123	0.920	0.571	5.1	2770.2	1582.6	79.2	391.4	392.4	0.822	0.547	0.008	68.0
13	1075.9	0.121	0.921	0.566	135.1	2740.3	1590.4	90.3	403.0	404.0	0.534	0.544	0.008	70.7
14	1529.8	0.110	0.943	0.557	28.3	2658.8	1687.5	72.6	581.7	582.7	10.250	0.561	0.009	69.7
15	1576.4	0.101	0.951	0.564	13.9	2581.3	1714.6	75.2	644.5	645.5	0.395	0.563	0.008	71.0
16	1548.0	0.100	0.947	0.569	19.0	2399.4	1720.8	83.6	660.0	661.0	1.473	0.553	0.008	104.1

TABLE 7: Optimal number of clusters.

k^*	CH	C	Gamma	DB	Hartigan	Tau	Scott	Marriot	Friedman	Rubin	KL	Silhouette	Dunn	SD index
0°	15	16	3	3	11	3	3	11	11	11	7	2	3	3
45°	14	4	2	2	3	3	3	14	14	14	6	2	2	3

clustering validity indices have optimal cluster numbers of 4, 5, 10, 11, and 12, respectively. Specifically, four clustering validity indices give the optimal cluster number as 2, and this value is repeated most often. Therefore, the most unfavorable WPC zoning is obtained for optimal cluster number of 2 with wind direction of 0°, and the corresponding WPC zones are shown in Figure 4(a). Similarly, the proposed method determines the optimal cluster numbers of the most

unfavorable suction at 0°, most unfavorable pressure at -45°, and most unfavorable suction at -45° as 3, 3, and 5, respectively.

For validation, the WPC distribution on the flat roof surface is compared with the optimal zoning results. The results show that the zoning can match the WPC distribution on the surface of the cladding structure, as shown in Figure 5.

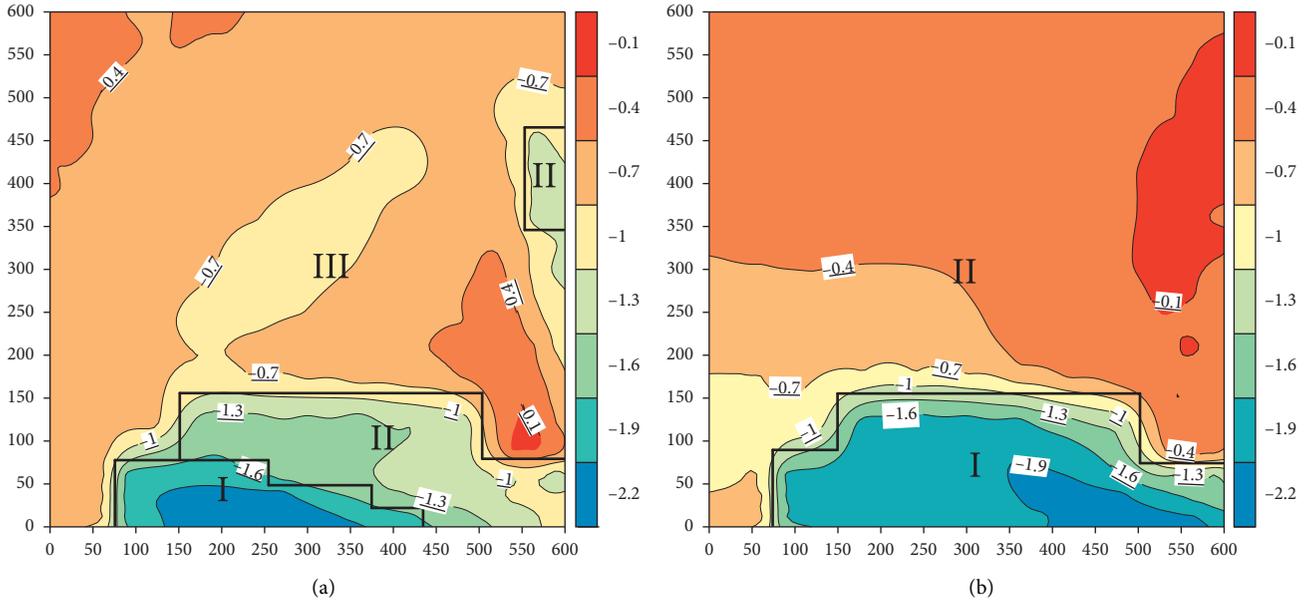


FIGURE 10: Optimal zoning of the mean pressure coefficient. (a) 0° . (b) 45° .

4.3. WPC Zoning under Full Wind Direction Condition.

Owing to the plane shape of the roof being biaxially symmetric, one-fourth of the roof can normally be used for analysis. However, for the purposes of verification and comparison, the right half of the roof surface is used for WPC zoning. For $n = 120$ measuring points on the right half surface (including the midline) of the flat roof, $k_{\max} = \sqrt{n} = \sqrt{120} \approx 10.95$; that is, $k_{\max} = 10$. By using the traditional K -means algorithm to achieve all cluster numbers $K \in [2, k_{\max}]$, clustering is performed with the most unfavorable pressure and suction for the full wind direction. Several clustering validity indices are calculated for different cluster numbers, as shown in Tables 3 and 4. According to Table 1, the optimal number of clusters determined by each index is shown in Table 5.

As shown in Table 3, the most unfavorable pressures in the full wind direction are as follows: four, three, and two clustering validity indices give optimal cluster numbers of 10, 3 and 7, and 8, respectively. The other two clustering validity indices give optimal cluster numbers of 5 and 6, respectively. For the most unfavorable pressure in the full wind direction, 10 optimal zoning classes are determined by four validity indices and the number of repetitions is the highest. However, considering that the distribution of the most unfavorable pressure has a gentle gradient, for convenience of design and construction, three optimal zoning classifications with the second highest number of repetitions are selected. Figure 6(a) shows the corresponding optimal zoning.

The proposed method determines the optimal cluster numbers for the most unfavorable suction in the full wind direction as 4, as shown in Figure 6(b). To verify the effectiveness of the WPC zoning, the results of the WPC distribution nephogram and the optimal zoning on the right side of the roof are compared. The results show that the zoning matches the WPC distribution, as shown in Figure 7.

5. Examples of Complex Structure

5.1. Wind Tunnel Tests. The wind tunnel test for a saddle roof structure was conducted in a wind tunnel at Beijing Jiaotong University. Figure 8 shows the mean wind speed and turbulence intensity profiles. The saddle roof model has dimensions of $60 \text{ cm} \times 60 \text{ cm}$ on a plane with lowest height (H) of 20 cm, and the rise span ratio is $1/8$. There are 265 pressure taps on the top of the saddle roof model, which is shown in Figure 9. The geometric scale, speed scale, and time scale of the wind test are $1/100$, $3.125/10$, and $1/31.25$, respectively. The sampling frequency is 312.5 Hz, and the sampling time corresponding to 10 min in the full scale is 19.2 s.

5.2. Mean WPC Zoning. The maximum clustering number k_{\max} is the root of the number of measuring points n on the saddle roof to be zoned. In the present case, $n = 265$ gives $k_{\max} \leq \sqrt{n} = \sqrt{265} \approx 16.3$; that is, $k_{\max} = 16$. By using the traditional K -means algorithm to achieve all cluster numbers $K \in [2, 16]$ and clustering saddle roofs with the mean pressure at 0° and 45° , we calculate several clustering validity indices for different cluster numbers. The clustering validity indices are calculated for different cluster numbers shown in Table 6. According to Table 1, the optimal number of clusters determined by each index is shown in Table 7.

According to the mean pressure obtained at 0° , six and four clustering validity indices have optimal cluster numbers of 3 and 11, respectively. The other four clustering validity indices have optimal cluster numbers of 2, 7, 15, and 16, respectively. From the results, the optimal cluster number 3 is repeated most often. Therefore, the mean pressure zoning is obtained for optimal cluster number of 3 with wind direction of 0° , and the corresponding WPC zones are shown in Figure 10(a).

Similarly, for the results of the mean pressure obtained at 45° , four clustering validity indices have optimal cluster numbers of 2, 3, and 14, respectively. The other two clustering validity indices have optimal cluster numbers of 4 and 6, respectively. Cluster number 2 is the minimum from the optimal cluster numbers of 2, 3, and 14 which have the same repetitions. So the mean pressure zoning is obtained for optimal cluster number of 2 with wind direction of 45° , and the corresponding WPC zones are shown in Figure 10(b).

For validation, the WPC distribution on the saddle roof surface is compared with the optimal zoning results. The results show that the zoning can match the WPC distribution on the surface of the saddle roof, which is shown in Figure 10.

6. Conclusions

This study presents a new concept to obtain the classifying WPC values on a roof surface based on unsupervised learning algorithm. The dynamic clustering theory is introduced for the WPC zoning and a fast WPCZM based on K -means clustering is proposed. Because this proposed classification is performed only based on the WPC magnitudes, the zoning process is not limited by the roof geometries and the objective multitype WPC zoning can be achieved. The proposed WPCZM is illustrated and verified. And the following conclusions can be drawn from this study.

- (1) The concept of classifying WPC values based on unsupervised learning algorithm to achieve the WPC zoning can provide an effective solution to overcome the problems of strong subjectivity, weak practicability, and limited scope caused by the application of the existing WPCZM.
- (2) Based on the improved K -means clustering method, the WPCZM can realize the multitype WPC zoning more accurately and adequately under the conditions with different geometric structures, different flow field characteristics, and different types of WPCs.
- (3) The layout of the wind tunnel test points may have some influence on the zoning results. It is suggested that the pressure measuring points should be laid out as evenly as possible when using this method.
- (4) Although this proposed WPCZM is originally based on the large-area flat roof and saddle roof, it is applicable to not only the roof claddings but also the other similar structures.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The supports provided in part by National Nature Science Foundation of China (51678548) and Science and Technology Project of State Grid Corporation of China (18-GW-07) which are greatly acknowledged.

References

- [1] N. J. Cook, "The designer's guide to wind loading of building structures, part 1: background, damage survey, wind data and structural classification," in *Building Research Establishment*, Garston and Butterworths, London, UK, 1985.
- [2] M. Kasperski, "Design wind loads for low-rise buildings: a critical review of wind load specifications for industrial buildings," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 61, no. 2-3, pp. 169–179, 1996.
- [3] American Society of Civil Engineers, *Minimum Design Loads for Buildings and Other Structures: ASCE/SEI 7-10*, American Society of Civil Engineers, Reston, VA, USA, 2013.
- [4] National Research Council of Canada, *User's Guide: NBC 2005 Structural Commentaries: Part 4 of Division B: NBC-2005*, National Research Council of Canada, Ottawa, Canada, 2005.
- [5] Architectural Institute of Japan, *Recommendations for Loads on Buildings: AIJ-2004*, Architectural Institute of Japan, Tokyo, Japan, 2004.
- [6] D. Surry and E. M. F. Stopar, "Wind loading of large low buildings," *Canadian Journal of Civil Engineering*, vol. 16, no. 4, pp. 526–542, 1989.
- [7] T. Stathopoulos and P. Saathoff, "Wind pressure on roofs of various geometries," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 38, no. 2-3, pp. 273–284, 1991.
- [8] H. Alrawashdeh and T. Stathopoulos, "Wind pressures on large roofs of low buildings and wind codes and standards," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 147, pp. 212–225, 2015.
- [9] X. Dong and J. Ye, "Study of shape factors for large-span flat roofs," *China Civil Engineering Journal*, vol. 45, no. 4, pp. 29–42, 2012, in Chinese.
- [10] Y. Sun, Y. Wu, Z. Lin, and S. Shen, "Non-gaussian features of fluctuating wind pressures on long span roofs," *China Civil Engineering Journal*, vol. 40, no. 4, pp. 1–5, 2007, in Chinese.
- [11] S. Ke, Y. Ge, and L. Zhao, "Research on features of fluctuating wind pressure on large hyperbolic cooling tower: features of non-gaussian," *Journal of Experiments in Fluid Mechanics*, vol. 24, no. 3, pp. 12–18, 2010, in Chinese.
- [12] Y. Reich, "Machine learning techniques for civil engineering problems," *Computer-Aided Civil and Infrastructure Engineering*, vol. 12, no. 4, pp. 295–310, 1997.
- [13] Y. Gao and K. M. Mosalam, "Deep transfer learning for image-based structural damage recognition," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 9, pp. 748–768, 2018.
- [14] D. J. Gunaratnam and J. S. Gero, "Effect of representation on the performance of neural networks in structural engineering applications," *Computer-Aided Civil and Infrastructure Engineering*, vol. 9, no. 2, pp. 97–108, 1994.
- [15] H. Sohn, S. D. Kim, and K. Harries, "Reference-free damage classification based on cluster Analysis," *Computer-Aided Civil and Infrastructure Engineering*, vol. 23, no. 5, pp. 324–338, 2008.
- [16] P. N. Tan, M. Steinback, and V. Kumar, *Introduction to Data Mining*, Addison-Wesley, New York, NY, USA, 2005.

- [17] S. I. Gallant, *Neural Network Learning and Expert Systems*, The MIT Press, London, UK, 1993.
- [18] J. B. MacQueen, "Some methods of classification and analysis of multivariate observations," in *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–197, University of California Press, Berkeley, CA, USA, 1967.
- [19] N. R. Pal and J. C. Bezdek, "On cluster validity for the fuzzy c-means model," *IEEE Transactions on Fuzzy Systems*, vol. 3, no. 3, pp. 370–379, 1995.
- [20] J. C. Bezdek and N. R. Pal, "Some new indexes of cluster validity," *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, vol. 28, no. 3, pp. 301–315, 1998.
- [21] H. L. Yang, Y. S. Li, X. X. Hu, and R. Y. Pan, "Optimization study on K value of K -means algorithm," *Systems Engineering-Theory & Practice*, vol. 26, no. 2, pp. 97–101, 2006, in Chinese.
- [22] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics*, vol. 3, no. 1, pp. 1–27, 1974.
- [23] L. J. Hubert and J. R. Levin, "A general statistical framework for assessing categorical clustering in free recall," *Psychological Bulletin*, vol. 83, no. 6, pp. 1072–1080, 1976.
- [24] F. B. Baker and L. J. Hubert, "Measuring the power of hierarchical cluster Analysis," *Journal of the American Statistical Association*, vol. 70, no. 349, pp. 31–38, 1975.
- [25] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, 1979.
- [26] J. A. Hartigan, *Clustering Algorithms*, John Wiley & Sons, New York, NY, USA, 1975.
- [27] A. J. Scott and M. J. Symons, "Clustering methods based on likelihood ratio criteria," *Biometrics*, vol. 27, no. 2, pp. 387–397, 1971.
- [28] F. H. C. Marriot, "Practical problems in a method of cluster analysis," *Biometrics*, vol. 27, no. 3, pp. 501–514, 1971.
- [29] H. P. Friedman and J. Rubin, "On some invariant criteria for grouping data," *Journal of the American Statistical Association*, vol. 62, no. 320, pp. 1159–1178, 1967.
- [30] W. J. Krzanowski and Y. T. Lai, "A criterion for determining the number of groups in a data set using sum-of-squares clustering," *Biometrics*, vol. 44, no. 1, pp. 23–34, 1988.
- [31] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.
- [32] J. C. Dunn, "Well-separated clusters and optimal fuzzy partitions," *Journal of Cybernetics*, vol. 4, no. 1, pp. 95–104, 1974.
- [33] M. Halkidi, M. Vazirgiannis, and Y. Batistakis, "Quality scheme assessment in the clustering process," *Principles of Data Mining and Knowledge Discovery*, vol. 1910, no. 1, pp. 265–276, 2000.