Investigation of the Effects of the Classification of Building Stock Geometries Determined Using Clustering Techniques on the Vulnerability of Galvanized Iron Roof Covers Against Severe Wind Loading

Tan, Liezl Raissa E., Acosta, Timothy John S.*, Gumaro, Joshua Joseph C., Agar, Joshua C., Tingatinga, Eric Augustus J., Plamenco, Dean Ashton D., Ereno, Mary Nathalie C., Musico, John Kenneth B., Pacer, Jihan S., Baniqued, Julius Rey D., Hernandez, Jaime Jr. Y., Villalba, Imee Bren O.

1Institute of Civil Engineering, University of the Philippines Diliman, Philippines

E-mail: tsacosta@up.edu.ph

Abstract. In the risk assessment of buildings against severe wind loading, the vulnerability component of risk is highly affected by the geometry of the structure. Parameters such as the height, width, length, aspect ratio, eaves length, and roof slope, affect the pressure distribution around the structure, which in turn affects the response of galvanized iron (GI) roof covers to wind loadings. In developing countries, there is a large variation in the building geometric parameters which poses a challenge in determining the archetypes that would best represent the building population for risk assessment. This paper aims to develop and propose a method in determining the building archetypes based on its geometry. The hierarchy for grouping of geometries started with the roof type. These were gable type roofs, mono-slope type roofs and hip type roofs. The building datasets per roofing type were then clustered using a two-stage approach involving Hierarchical and K-means clustering which were based on the aforementioned geometric parameters. These algorithms will aggregate buildings having similar sets of geometric parameters but the number of clusters must be specified. In order to determine the optimal number of clusters, this study employs various validation tools or measures namely – dendrograms, variation of the variance ratio criterion (VRC) across number of clusters, validity indices such as, Davies Bouldin, Silhouette and Calinski Harabasz, and the elbow method. Although guided by these validation methods, the final selection of the number of clusters were determined considering computational time and resources. To define an archetype, the mean values of each parameter per cluster were selected. Resulting to 5, 3, and 3, archetypes for gable, hip, and mono-slope roof buildings, respectively. The selection of the archetype was further evaluated by investigating its effects on the vulnerability of GI roof covers in order to see how distinct each archetype would behave. A Kruskal Wallis test on the vulnerability curves of the different building archetypes showed that there is a significant difference between vulnerability curves under a roof shape category, which reinforces the distinction between the selected building archetypes.
1. Introduction

The determination of the building stock is a key parameter in the field of risk analysis. Albayrak et al. (2015) were able to classify buildings into existing building stock for rapid seismic risk assessment [1]. The classification of building stock helps most especially in the classification of local housing construction wherein these are non-engineered buildings. Parameters such as the main structural frame, load-bearing systems, connections between structural elements among others, are used to identify basic typologies for certain countries [2]. One classification system was developed for the Prompt Assessment of Global Earthquakes for Response (PAGER) system. This system is used to estimate an earthquake’s impact in order to guide governments, insurance agencies and relief organizations in post-disaster relief operations and strategies [3]. This system made use of a global building inventory that was developed through the use of different inventory data sources worldwide [4]. Databases such as the UN statistical database on global housing [4], the United Nations Human Settlement Program (UN-HABITAT) database [4], World Housing Encyclopedia (WHE) database [5] have information on the housing dwelling type, construction materials for wall, roof or floors, structural systems and other parameters that contribute to the vulnerability of building to seismic activity [4]. Aside from these databases, FEMA was able to develop HAZUS standard model building classification schemes. These building stock classifications were not only used for seismic risk assessment but were used for multi-hazard loss estimations.

One of the hazards which HAZUS assesses is severe wind events. For their methodology, two building stock classification methods were used. The first method was through the inspection of aerial photography samples of Dade, Broward and Palm Beach of Southeast Florida. It was concluded that, for commercial building stocks, simple rectangular-shaped building models were adequate for damage simulation which came from random sampling from the aerial photographs wherein they observed that 80% of the samples accounted for rectangular or L shaped plan areas. For the residential building stock, they used three simple default classifications which were the single-story gable, single story hip and two-story gable. It was also noted that no attempt was made to quantify the plan shapes. The other method made use of contractor surveys to determine the distribution of GI roof cover types. The method was able to come up with 13 building stock classifications [6]. Each building stock classification had multiple combinations of building envelope components that contributed to the vulnerability against wind loadings. For the building geometries used to determine the wind loads were based on expert opinion for each building stock classification [6,7]. The different building geometries which were derived using expert opinion may also affect the determination of the wind loads which in turn will affect the resulting vulnerability of the building stock.

With the advancement of computational fluid dynamics (CFD), many researches have been done to explore the effects of geometric parameters on the wind loads of a building [8,13,14,15]. In the application of risk assessment, the use of CFD to individually analyze the configuration of the population of buildings is very complicated and computationally heavy. Thus, there is a need to explore the use of a building geometry that would statistically best represent the population of buildings being assessed and will also improve the risk assessment of the population of buildings being considered. One method is to use clustering techniques on diverse data to come up with a representative ensemble for the population of buildings. Since the wind pressures on the building will be computed using computational fluid dynamics, the objective of determining the building geometry archetypes, instead of individual buildings of the population, will significantly reduce the computational cost.

The use of clustering techniques in the classification of building stock for different applications has been prominent in the past decade. The technique has been applied to cluster buildings; based on their sensitivity to retrofitting measures [16], in the context of air conditioning electricity demand flexibility.
[17], performance in energy savings [18], energy efficiency and emission reduction [19], and many other areas. With this in mind, this paper seeks to investigate the use of clustering techniques to develop building stock geometries for risk assessment of low-rise buildings against severe wind loadings.

This proposed methodology was implemented in a pilot study to assess the risk of the province of Cebu, Philippines against severe wind loadings. The pilot study is part of the Philippine government’s national effort to produce scientifically based vulnerability profiles which would help in disaster risk and reduction strategies. The greater metro manila area risk analysis project (GMMA-RAP) [20] is a precedent of the current pilot study wherein key government agencies and institutions such as the Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAG-ASA), Philippine Institute of Volcanology and Seismology (PHIVOLCS) and Institute of Civil Engineering, University of the Philippines Diliman (UPD-ICE) were involved. And given a lack of a high-resolution exposure database, this study was performed to improve the risk analysis methodology.

2. Methodology
2.1. Parameter Identification.
Various researches have shown that the roof shape, slope, length, width, height [9,10,11,12] have significant effect on the pressures for low-rise structures. The hierarchy of the parameters was patterned after the aerodynamic database developed by Tokyo Polytechnic University (TPU) [21]. The dataset was then classified by the roof shapes.

2.2. Clustering Analysis.
Clustering techniques group data points into clusters in a way that the data points in the same cluster are similar to one another compared to data points in another cluster [22]. This paper employed a two-stage approach involving Hierarchical and K-means clustering algorithms employed in MatLAB. The two algorithms complemented each other in a way that k-means clustering, known for its efficiency in large datasets, works best when a non-random starting point is specified, which is provided by the cluster centroids obtained from the hierarchical clustering. A series of the number of clusters were run using the k-means approach where each building data point was assigned to the nearest cluster. The cluster centroids were recomputed and data points reassigned to minimize the variability within clusters whilst maximizing the variability between clusters. This approach is similar to the ones employed in the study on electricity usage for consumer archetypes [22].

2.3. Evaluation of the results.
In order to assess the optimum number of clusters, various validation methods were utilized. Visual methods such as the use of dendrograms and the elbow method, were used to roughly assess the distinction across clusters. The elbow method plots the within-cluster-sum-of-squared error (WSS), for every number of clusters considered. The number of clusters corresponding to the elbow of the graph is regarded as the optimum number of clusters. WSS is computed as follows,

\[ WSS = \sum_k \sum_{i \in C_k} (x_i - \mu_k)^2 \]  

Where \( x_i \) is an element within the \( k^{th} \) cluster \( C_k \) with mean value equal to \( \mu_k \). Thus \( WSS \) takes the summation of the summation of the sum-of-square error (SSE) within the cluster which is then summed across all clusters \( C \).

Internal validity indices were also used to evaluate the clustering performance. These are the Silhouette index, Davies Bouldin Index and the Calinski Harabasz Index. The silhouette width indicates the belongingness of a datapoint in a cluster it was assigned. Values ranges from -1 to 1 wherein a value close to 1 means that the “within” dissimilarity is much smaller than the “between” dissimilarity, hence
it is said to be well-clustered. Davies Bouldin index is a ratio of the total scatter within the cluster to the separation between clusters, and hence a smaller value is regarded as good clustering. Calinski Harabasz is the ratio of the overall between-cluster variance to the overall within-cluster variance multiplied by the factor ratio between the difference of the total data points and the number of clusters to the number of clusters minus 1. There is no acceptable cut-off value, but the higher the index the better.

Another validation method makes use of the variance ratio criterion (VRC) [24,25]. F value were determined for each parameter, \( p \), per number of clusters, \( C \). The F value, \( F_p \), measures the ratio of variability of each parameter, \( p \), between clusters, \( k \), given by the Between-Group-Sum-of-Square Error (BGSS) and the variability within a cluster or WGSS. [23].

\[
F_p = \frac{BGSS}{\frac{c-1}{WGSS}}
\]

\[
BGSS = \sum_k n_k ||\mu_k - \mu||^2
\]

\[
WGSS = \sum_k \sum_{i \in k} ||x_i - \mu_k||^2
\]

Where \( \mu \), represents the centroid of all the \( n \) number of datapoints, \( x_i \), while \( \mu_k \) represents the centroid within clusters of \( n_k \) datapoints.

The F value of each parameter can be used to identify which parameter greatly affects the distinction between clusters. So, for a number of clusters considered, the higher the F value of a parameter the higher its influence to the clustering process, and therefore can be regarded as a critical clustering parameter. In relation to determining the optimum number of clusters, the F values were totaled to get the VRC for every number of clusters. The higher the value of the VRC implies that either clusters are highly distinct from each other or that the data points within each cluster is highly similar with each other, and thus a high value of VRC is preferable in the selection of the number of clusters. However, the VRC value generally increases as the number of clusters approaches the total number of datapoints. To aid in selecting an efficient number of clusters, the VRC values were then compared to its neighboring number of clusters using the equation for \( \omega_k \), which quantifies the difference between the deviations of VRC from the current cluster to the succeeding and preceding clusters.

\[
\omega_k = (VRC_{k+1} - VRC_k) - (VRC_k - VRC_{k-1})
\]

A positive value for the first term indicates that the current cluster performs worse than the succeeding cluster while a positive value for the second term indicates that the current cluster performs better than the previous cluster. Therefore, the more negative the value of the \( \omega_k \) indicates a more efficient selection of the number of clusters.

2.4. Roofing Vulnerability Analysis.

In order to illustrate the use of the above geometries, the vulnerability curves for the roof sheeting were determined in this paper. The framework was patterned after the GMMA-RAP simulation methodology [20]. The archetypes developed in the clustering analysis served as the representative geometries that would be subjected to computational fluid dynamics. The damage ratio was finally derived by determining the percent area of the roofing sheets that exceeded the threshold capacity and dividing this by the total area. The vulnerability curve for each building archetype was then derived by getting the mean of the models accounting for different wind directions and plotting by the corresponding wind speed.
3. Results: Case Study of Cebu, Philippines

3.1. Clustering Analysis.

For the pilot study area, which is Cebu, Province, the parameters of interest were surveyed by a team coming from the Institute of Civil Engineering, University of the Philippines Diliman. The target areas in Cebu were chosen based on the analysis of the construction material typology per municipality wherein the data on the construction materials came from the housing tables of the Philippine Statistical Authority (PSA) [26]. A total of 179 buildings from 15 municipalities of the province were gathered. The prominent roof shapes were gable roofs, mono-slope roofs, and hip roofs which comprised 50.8%, 24.58%, and 12.89% of the data, respectively. Hence these were selected as the three main groups used for the cluster analysis.

The figure below shows the dendrograms produced by the hierarchical clustering algorithm. The centroids of these results were used as the input cluster centroids for the k-means clustering algorithm. Each number in the horizontal axis represents the building ID number and hence represents one building data point. While the vertical axis represents the mean values of the parameters or also termed as cluster centroid. The height of the links in a dendrogram is indicative of the distinction between clusters. Each color represents a cluster in a group of a particular roof shape, and consequently shows how big of the building population a cluster represents.

**Figure. 1** Dendrograms from Hierarchical Clustering of different roof shapes using Ward’s Method.

**Figure. 2** Plots of the a) elbow method, b) Davies-Bouldin Indices, c) Silhouette indices, and d) Calinski-Harabasz indices against the number of clusters for gable, mono-slope and hip roof types.
The validation indices from various methods provided the team a range of number of clusters to select from and also gives a measure of how distinct these clusters are from each other yet contains data points that are similar within each cluster. Consideration of the computational time and resources were also a major influence in the selection of the final number of clusters.

The $\omega_k$ values were obtained for 3 to 10 clusters for all roof shapes and are shown in the following figure.

![Figure 3](image)

Figure 3 Variation in VRC between neighboring clusters ($\omega_k$) for every number of clusters.

A selection between 2 to 7, 4 to 8 and 3 to 10 number of clusters were considered for gable, mono-slope and hip roof types, respectively. Considering the roof type composition in the building and computational time and resources the author’s settled at 5, 3 and 3 number of clusters for gable, mono-slope and hip roof types, respectively. In order to understand the characteristics of the population of buildings, the relative contribution of each clustering parameter to the development of the clusters was investigated. The F values of each parameter were ranked to see the importance of each parameter to the clustering algorithm.

| Parameter | Gable Clusters $k = 5$ | Hip Clusters $k = 3$ | Mono-slope Clusters $k = 3$ |
|-----------|------------------------|-----------------------|-----------------------------|
| Height    | Rank: 3, F: 12.79       | Rank: 3, F: 6.66      | Rank: 3, F: 14.32           |
| Length    | Rank: 1, F: 263.88      | Rank: 1, F: 108.93    | Rank: 1, F: 71.38           |
| Width     | Rank: 2, F: 142.69      | Rank: 2, F: 50.65     | Rank: 2, F: 45.51           |
| Slope     | Rank: 6, F: 0.77        | Rank: 6, F: 2.84      | Rank: 4, F: 4.44            |
| Eaves-L   | Rank: 5, F: 1.16        | Rank: 5, F: 4.36      | Rank: 5, F: 1.48            |
| Eaves-S   | Rank: 4, F: 5.91        | Rank: 4, F: 4.76      | Rank: 6, F: 1.31            |
| VRC       | Rank: 4, F: 427.22      | Rank: - F: 178.19     | Rank: - F: 138.43           |
| wk        | Rank: - F: 139.094      | Rank: - F: -83.80     | Rank: - F: -276.87          |

The table above shows us that the primary parameters that affected the formation of clusters are the length (ranked 1st) and width (ranked 2nd) of the building geometries. This implies that the building population can be categorized more by their plan area rather than their roof shape parameters. This also
tells us that the construction practice for the roof slope, eaves and length has a low variability in the pilot study area.

The final geometry clusters produced through the clustering techniques are now considered as the archetypes which will be used to determine the wind loads on the roofing sheets. The building archetypes for the different roof shapes can be found in table 2 to 4. The illustration in figure 4 gives a reference on how the different parameters are measured.

![Figure 4](image)

**Figure 4** Illustration of parameter measurements for gable roof archetypes

### Table 2. Centroids for Gable K-means clusters

| ARCHETYPE LABEL | HEIGHT | LENGTH | WIDTH | ROOF SLOPE | EAVES-L | EAVES-S |
|-----------------|--------|--------|-------|------------|---------|---------|
| G1              | 4.08   | 10.50  | 5.07  | 15.33      | 1.01    | 0.50    |
| G2              | 8.50   | 41.67  | 27.00 | 19.00      | 1.10    | 1.90    |
| G3              | 4.26   | 5.97   | 4.62  | 19.00      | 0.85    | 0.58    |
| G4              | 8.75   | 22.00  | 16.35 | 19.00      | 1.22    | 0.25    |
| G5              | 5.84   | 11.6   | 8.76  | 19.5       | 0.68    | 0.58    |

### Table 3. Centroids for Hip K-means clusters

| ARCHETYPE LABEL | HEIGHT | LENGTH | WIDTH | ROOF SLOPE | EAVES-L | EAVES-S |
|-----------------|--------|--------|-------|------------|---------|---------|
| H1              | 3.95   | 7.37   | 5.70  | 24.20      | 0.83    | 0.70    |
| H2              | 8.90   | 24.95  | 17.45 | 38.31      | 1.45    | 1.45    |
| H3              | 5.90   | 13.42  | 8.75  | 21.03      | 1.03    | 0.875   |

### Table 4. Centroids for Mono-slope K-means clusters

| ARCHETYPE LABEL | HEIGHT | LENGTH | WIDTH | ROOF SLOPE | EAVES-L | EAVES-S |
|-----------------|--------|--------|-------|------------|---------|---------|
| M1              | 5.42   | 6.92   | 4.78  | 8.83       | 0.70    | 0.20    |
| M2              | 9.00   | 18.23  | 11.00 | 2.50       | 0.50    | 0.10    |
| M3              | 20.00  | 50.00  | 30.00 | 0.00       | 0.00    | 0.00    |
3.2. Vulnerability Results.
The simulations were done for wind speeds from 10 m/s to 100 m/s and considered different orientations of the wind relative to the building surfaces. Due to the symmetry of the archetypes, only angles between 0 degrees to 90 degrees at increments of 15 degrees were considered. The various combinations gave a total of 770 models that were simulated with 350 models for the 5 gable roof archetypes, 210 models for the 3 hip roof archetypes and 210 for the 3 mono-slope roof archetypes. All CFD simulations were implemented through the software ANSYS CFX. The GI roof cover failure which was primarily characterized by pullout strength of 4.3 kPa [27] was compared to the positive and negative wind pressures developed on the building archetypes. The simulated data points were then fitted into a lognormal cumulative distribution function (CDF) using a nonlinear regression analysis that minimized the sum of squared estimate of errors (SSE) to represent the final vulnerability curves. The final vulnerability curves are then shown in figure 5.

| Building Archetypes | CDF Parameters | G1 | G2 | G3 | G4 | G5 | H1 | H2 | H3 | M1 | M2 | M3 |
|---------------------|----------------|----|----|----|----|----|----|----|----|----|----|----|
| Mean                |                | 4.92| 4.86| 4.87| 4.67| 4.69| 4.92| 4.89| 4.83| 4.65| 4.54| 4.38|
| Std                 |                | 0.31| 0.37| 0.30| 0.23| 0.21| 0.29| 0.29| 0.32| 0.22| 0.21| 0.36|

Figure 5 Roof Sheeting Vulnerability for different building archetypes

A Kruskal-Wallis test was used to determine if the vulnerability curves for each building archetype were significantly different from one another. For gable, hip and mono-slope roof building archetypes, it was observed that for wind speeds from 60 m/s – 100 m/s, there was a significant difference ($p = .001$) for the vulnerability curves between the 5 building archetypes at a level of significance of 5%. For hip roof building archetypes, the same conclusion of a significant difference ($p = .001$) was also observed between the three building archetypes. And for the mono-slope building archetypes, the vulnerability curves were also significantly ($p < .0001$) different from each other. These results show that the use of building archetypes developed from the two-stage clustering algorithm has a significant effect on the vulnerability analysis of roof sheeting against severe wind loadings. The use of multiple building
archetypes that best represent the population of buildings considers the variability of the geometric parameters of the area being assessed. The analysis can be extended to investigate the vulnerability curves that also consider other building envelope components aside from the roofing sheets that are also vulnerable to severe wind loadings.

4. Conclusion
This paper presents the application of clustering techniques in determining the building geometry archetypes which will be used for vulnerability and risk assessments. Specifically, the vulnerability assessment for galvanized iron roof covers against severe wind loading was evaluated. The paper implemented a two-stage clustering approach wherein Hierarchical and K-means clustering algorithms were used to develop clusters for the population of buildings in Cebu, Philippines. A total of 179 buildings were considered and subdivided into 3 main categories based on roof shape namely the gable, mono-slope and hip. Cluster analysis was employed on buildings in each roof shape category considering the parameters, length, width, height, roof slope, and eaves lengths. The analysis considers various numbers of clusters where cluster centroids and cluster assignments were obtained. The clusters were then evaluated using validation scores for cluster analysis namely the elbow method, silhouette index, Davies-Bouldin, Calinski Harabasz and the variation in VRC between neighboring clusters ($\omega_k$).

The validation scores were used to serve as a guide in the selection of the number of clusters to define the archetypes per roof type. Considering the composition of building population and the computational time and resources, the final number of clusters were selected for each roof shape. These are 5, 3, and 3 for gable, mono-slope and hip roof type, respectively. The cluster centroids for the selected number of clusters per roof type were then used as the dimensions for the building archetypes to be subjected to vulnerability analysis. For each building archetype, the simulated damage ratios for every wind speed data were fitted with lognormal cumulative distribution functions in obtaining the vulnerability curves. The vulnerability curves for the gable, hip and mono-slope building archetypes were observed to have a significant difference within each respective roof category at a level of significance of 5%. This shows that the consideration of the building archetypes from a sample population will significantly affect the results of the vulnerability analysis for wind speeds greater than 60 m/s. Further studies can be done to investigate the effect on vulnerability curves that also consider different building components that are vulnerable to severe wind loadings.

Acknowledgements
This work is part of the project entitled “Enhanced Vulnerability Curves for Different Building Types in the Philippines” under the program “Severe Wind Risk Assessment of Cebu City” wherein this project received funding from the Department of Science and Technology (DOST) through the Grants-In-Aid (GIA) program. The program was implemented in collaboration with the Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA) and Philippine Institute of Volcanology and Seismology (PHIVOLCS).

References
[1] Albayrak, Uğur, Mehmet Canbaz, and Gülçağ Albayrak. "A rapid seismic risk assessment method for existing building stock in urban areas." Procedia engineering 118 (2015): 1242-1249.

[2] Ngoma, Ignasio, Innocent Kafodya, Panos Kloukinas, Viviana Novelli, John Macdonald, and Katsuichiro Goda. "Building classification and seismic vulnerability of current housing construction in Malawi." Malawi Journal of Science and Technology 11, no. 1 (2019): 57-72.

[3] Earle, Paul S., David J. Wald, Kishor S. Jaiswal, Trevor I. Allen, Michael G. Hearne, Kristin D. Marano, Alicia J. Hotovec, and Jeremy M. Fee. "Prompt Assessment of Global Earthquakes for Response (PAGER): A system for rapidly determining the impact of earthquakes worldwide." US Geological Survey Open-File Report 1131 (2009): 15.
[4] Jaiswal, Kishor, and David Jay Wald. Creating a global building inventory for earthquake loss assessment and risk management. US Geological Survey open-file report-1160. Reston, VA: US Geological Survey, 2008.

[5] EERI, 2007, World Housing Encyclopedia (http://www.world-housing.net/)

[6] FEMA. "HAZUS-MH 2.1 hurricane model technical manual." (2012).

[7] Vickery, Peter J., Peter F. Skerlj, Jason Lin, Lawrence A. Twisdale Jr, Michael A. Young, and Francis M. Lavelle. "HAZUS-MH hurricane model methodology. II: Damage and loss estimation." Natural Hazards Review 7, no. 2 (2006): 94-103.

[8] Singh, Jagbir, and Amrit Kumar Roy. "Effects of roof slope and wind direction on wind pressure distribution on the roof of a square plan pyramidal low-rise building using CFD simulation." International Journal of Advanced Structural Engineering 11, no. 2 (2019): 231-254.

[9] Alrawashdeh, Hatem, and Ted Stathopoulos. "Wind pressures on large roofs of low buildings and wind codes and standards." Journal of Wind Engineering and Industrial Aerodynamics 147 (2015): 212-225.

[10] Kasperski, Michael. "Design wind loads for low-rise buildings: a critical review of wind load specifications for industrial buildings." Journal of wind engineering and industrial aerodynamics 61, no. 2-3 (1996): 169-179.

[11] Holmes, John D., Yukio Tamura, and Prem Krishna. "Wind loads on low, medium and high-rise buildings by Asia-Pacific codes." In The 4th International Conference on Advances in Wind and Structures, Jeju, Korea. 2008.

[12] Prasad, Deepak, Tuputa Uliate, and M. Rafiuddin Ahmed. "Wind loads on low-rise building models with different roof configurations." International Journal of Fluid Mechanics Research 36, no. 3 (2009).

[13] Janajreh, Isam. "CFD Analysis of Wind Loads on Permeable Low-Rise Structures." In Applied Mechanics and Materials, vol. 152, pp. 1814-1820. Trans Tech Publications Ltd, 2012.

[14] Tominaga, Yoshihide, Shin-ichi Akabayashi, Takuya Kitahara, and Yuki Arinami. "Air flow around isolated gable-roof buildings with different roof pitches: Wind tunnel experiments and CFD simulations." Building and Environment 84 (2015): 204-213.

[15] Atmaca, M. U. S. T. A. F. A. "Wind tunnel experiments and CFD simulations for gable-roof buildings with different roof slopes." Acta Physica Polonica A, vol 135 (2019): 690-3.

[16] Geyer, Philipp, Arno Schlüter, and Sasha Cisar. "Application of clustering for the development of retrofit strategies for large building stocks." Advanced Engineering Informatics 31 (2017): 32-47.

[17] Patteeuw, Dieter, Gregor P. Henze, Alessia Arteconi, Charles D. Corbin, and Lieve Helsen. "Clustering a building stock towards representative buildings in the context of air-conditioning electricity demand flexibility." Journal of Building Performance Simulation 12, no. 1 (2019): 56-67.

[18] Lee, Yunseok, Pierre Boisson, Mathieu Rivallain, and Olivier Baverel. "Application of clustering technique for the segmentation of building stocks to renovate." 2016.

[19] Geyer, Philipp, and Arno Schlueeter. "Performance-based clustering for building stock management at regional level." In Proceedings of the International Conference ‘Smart Energy Regions’, pp. 230-241. Cardiff University, 2016.

[20] UP Diliman Institute of Civil Engineering, Development of Vulnerability Curves of Key Building Types in the Greater Metro Manila Area, Philippines (2014).

[21] TPU Aerodynamic database (http://www.wind.arch.t-kougei.ac.jp/system/eng/)
[22] Hughes, Mark., Moreno, Jorge Garcia. “Further Analysis of Data from the HEUS Consumer Archetypes.” Final report for the Department of Energy and Climate Change and Department for the Environment Food and Rural Affairs. (2013).

[23] Nikolaou, Triantafyllia G., Dionysia S. Kolokotsa, George S. Stavrakakis, and Ioannis D. Skias. "On the application of clustering techniques for office buildings' energy and thermal comfort classification." IEEE Transactions on Smart Grid 3, no. 4 (2012): 2196-2210.

[24] Caliński, Tadeusz, and Jerzy Harabasz. "A dendrite method for cluster analysis." Communications in Statistics-theory and Methods 3, no. 1 (1974): 1-27.

[25] Milligan, Glenn W., and Martha C. Cooper. "An examination of procedures for determining the number of clusters in a data set." Psychometrika 50, no. 2 (1985): 159-179.

[26] Philippine Statistics Authority. 2015 FIES Housing Tables (https://psa.gov.ph/content/2015-fies-housing-tables)

[27] Baskaran, Bas A., Steven Kee Ping Ko, and Suda Molleti. "A novel approach to estimate the wind uplift resistance of roofing systems." Building and environment 44, no. 4 (2009): 723-735.