Select Representative Sea States For Floating Offshore Wind Turbine Using Clustering Algorithms

Ding Peng Liu
Maseeh Department of Civil, Architectural and Environmental Engineering
The University of Texas at Austin
Email: dpl@utexas.edu

Kyungjung Cha
Maseeh Department of Civil, Architectural and Environmental Engineering
The University of Texas at Austin
Email: kjcha@utexas.edu

Fan Wang
Department of Communication Studies
The University of Texas at Austin
Email: fanwang@utexas.edu

April 28, 2024

Contents

1 Introduction and Previous Studies

Floating offshore wind turbines (FOWTs) have emerged as a promising technology to harness wind energy in deep-sea regions where fixed-bottom installations are not feasible. Unlike traditional offshore wind farms, FOWTs are subjected to complex dynamic forces influenced by wind, waves, and currents. Understanding the environmental conditions that FOWTs encounter is crucial for their design, operation, and maintenance.

One of the key challenges in the design and analysis of FOWTs is the selection of representative sea states for structural fatigue assessment. Sea states, characterized by wave height and period, are crucial environmental parameters that significantly impact the performance and reliability of these turbines. Long-term fatigue damage can be accumulated by the short-term fatigue damage from all possible sea states over the service life. However, considering the vast variability of sea conditions experienced in offshore environments, it becomes impractical to simulate all possible sea states. Therefore, a previous study [1] used principle analysis-based grids to select representative sea states based on the joint distribution of metocean conditions to reduce the number of simulations. If we have the observed data or Monte-Carlo samples from the joint distribution, clustering algorithms can also be useful to characterize the representative sea states [2]. For example, the maximum dissimilarity-based algorithm has been applied to characterize five-dimensional environmental conditions for FOWT [3], k-means algorithm has been used for selecting experimental conditions for wave energy converters [4].

In this research, we explore the use of clustering algorithms to identify and select a subset of representative sea states from a large dataset of environmental conditions. Clustering methods, such as k-means clustering or hierarchical clustering, offer a systematic approach to grouping similar sea states together based on their characteristics. These algorithms analyze patterns in the data to group sea states with similar characteristics together, which simplifies the number of simulations. By clustering sea states, we aim to reduce the complexity of the dataset while preserving the diversity of environmental conditions encountered by FOWTs. The primary objective of this study is twofold: first, to develop a methodology for identifying representative sea states that capture representative environmental conditions relevant to FOWTs, and second, to demonstrate the effectiveness of clustering algorithms in reducing the computational costs associated with simulating sea states for structural analysis using a coupled frequency domain simulation model [5] with an accurate estimation of long-term fatigue damage.

2 Data and Methods

Two types of data sets are used in this study (Figure 1): the environmental data used for representative sea state clustering and the simulation results of short-term fatigue damage for calculating long-term fatigue damage.

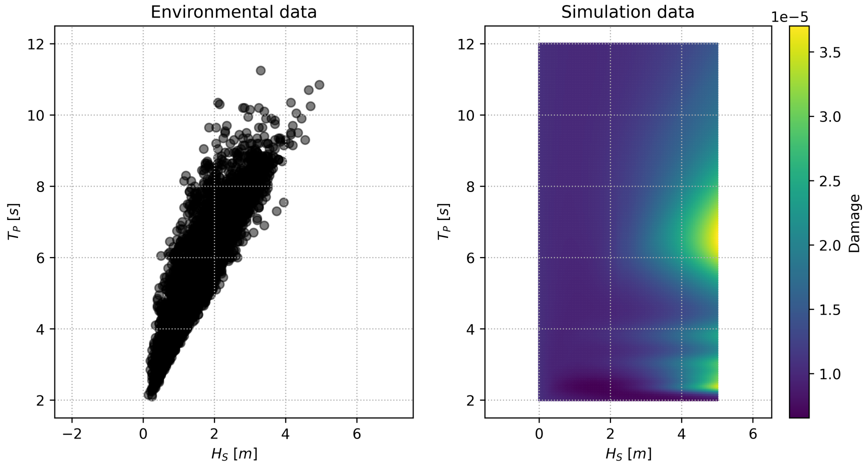


Figure 1: Environmental and simulation data sets

2.1 Environmental data

The environmental data used in this study is pre-processed from a 20-year dataset of wind and wave parameters collected every hour from port near Porto, Portugal. The dataset includes measurements of significant wave height (H_s), peak wave period (T_p), and mean wind speed (V), which are essential for assessing the environmental conditions for floating offshore wind turbine (FOWT) applications. These variables are defined as:

1. H_s: This parameter represents the average height of the highest one-third of waves in a given sea state. H_s is a critical factor in determining the structural loading and performance of FOWTs.

2. T_p: T_p denotes the period of the most energetic waves in a sea state. Understanding T_p is crucial for assessing wave dynamics and their impact on FOWT stability and response.

3. V: V over the measurement period provides insights into the prevailing wind conditions, which directly influence turbine power generation and structural design considerations.

For simplification, we only include the V between 10.4 and 12.4 m/s and set the representative V as 11.4 m/s (which is the rated wind speed of NREL 5 MW wind turbine) to demonstrate the proposed approach. Thus, the dimensions of the clustering are reduced to two-dimensional (H_s and T_p).

2.2 Simulation data

In addition to the environmental data preparation, this study involves the simulation of Floating Offshore Wind Turbine (FOWT) response using frequency domain analysis. The goal is to assess the fatigue damage of FOWTs under varying sea states represented by their Power Spectral Density (PSD) and subsequently derive the fatigue damage through post-processing. The response of FOWTs to environmental loads was simulated using frequency domain analysis. This method allows for efficient computation of structural responses across a range of sea states.

The PSD of the environmental loads was calculated based on the collected data of H_s, T_p, and V. The PSD provides a frequency-domain representation of the variance of these environmental loads over different wave frequencies and wind speeds. Using the calculated PSDs as input, the response stress PSD at the tower base of the FOWT system was simulated for each sea state. This involved modeling the dynamic behavior of the floating platform, mooring system, and turbine components. The selected FOWT system is NREL 5 MW reference wind turbine [6] supported by an OC4 semi-submersible platform [7]. Following the simulation, the Dirlik method [8] was applied to analyze the response data and derive the fatigue damage of every environmental data. The short-term fatigue damage is the accumulation of damage (unitless) within one hour for one sea state. The long-term fatigue damage is the average short-term fatigue damage from every sea state within one hour, which can help us estimate service life (in hours) before the accumulated damage reaches 1, indicating a fatigue failure.

2.3 Clustering algorithms

We employ several clustering algorithms to analyze the environmental data:

- **Principal Component Analysis (PCA):** PCA is employed to identify the principal components that follow the irregular shapes of wave data distribution. In the context of this study, PCA serves a specific purpose similar to that described in a previous study [1], where principle analysis-based grids were utilized to select representative sea states based on the joint distribution of metocean conditions. The grids are defined along the two principal components with same marginal probability. In this study, for the purpose of clustering, the grids are defined along two principal components to assure the size of each cluster is the same. This method effectively reduces the number of necessary simulations by identifying key features that represent the variance and dynamics of environmental conditions most relevant to our analysis.

- **k-means:** This algorithm partitions the data into k distinct clusters, each represented by the mean of the data points in the cluster. The number of clusters, k, is typically determined using methods such as the Elbow Method, which identifies a point where the rate of decrease in within-cluster variance significantly slows, indicating a natural grouping. The choice of k has significant implications, balancing between too few clusters, which can oversimplify the model, and too many, which may overfit the data and complicate the interpretation. Also, the randomness associated with the initialization of centroids in the k-means can affect the final clustering results in that k-means iteratively updates centroids to minimize the total within-cluster variance.

- **Mean-Shift:** Mean-shift clustering does not require the number of clusters to be specified in advance and works by updating candidates for centroids to be the mean of the points within a given region. These points are iteratively shifted towards the mode, determined by the kernel density estimate of the points. This method is adaptive and merges and finds the number of clusters based on the data itself, offering a more flexible approach to clustering. The choice of bandwidth parameter in this method influences the size of the kernel and thus impacts the convergence speed of mean-shift. Therefore selecting an appropriate bandwidth can substantially affect computation time.

- **Silhouette Score:** The Silhouette Score is used to assess the quality of the clusters formed by k-means and mean-shift algorithms. It provides a measure of how similar an object is to its own cluster compared to other clusters. A higher Silhouette Score indicates a more distinctly defined clustering. This metric is invaluable for validating the robustness of the clustering process across different algorithms, as it quantifies how well each object lies within its cluster. This helps in determining the appropriateness of the number of clusters and the effectiveness of the dimensionality reduction strategies employed.

2.4 Problem formulation

The long-term fatigue damage is calculated based on the short-term fatigue damage of all the possible sea states. In this study, we conducted a 1-hour coupled simulation for every observed environmental data, the reference solution of long-term fatigue damage in N hours can be derived as:

$$\boldsymbol{LTD} = \Sigma_i^N D(V, H_s, T_p)_i \tag{1}$$

where N is the total number of observed data and total time duration in hours. D is the short-term fatigue damage (unitless) and LTD is the long-term fatigue damage. When LTD reaches 1, the structure is considered failed and N is the expected service life. Nevertheless, the clustering algorithm summarizes n representative sea states with n clusters instead of all the observed data. LTD can be derived as:

$$\boldsymbol{LTD} = \Sigma_j^n D(V, H_s, T_p)_j \cdot t_j \tag{2}$$

where n represents the number of clusters or representative sea states, t_j is the total time duration of the observed data included in the cluster j. The computational costs can then be quantified by the number of simulations. By using clustering algorithms, it is reduced from N to n.

3 Results and Discussions

Initially, the environmental data underwent preprocessing, where the peak wave period (T_p) was modified to ensure equal spans for two variables (Figure 2). We believe this modification is beneficial for the clustering algorithms that generate spherical clusters, such as k-means algorithm. From a technical standpoint, note that after clustering with the modified data, reverting this scaling is necessary to accurately predict fatigue damage.

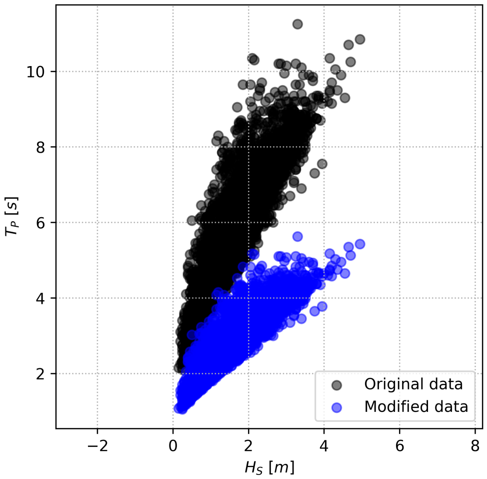
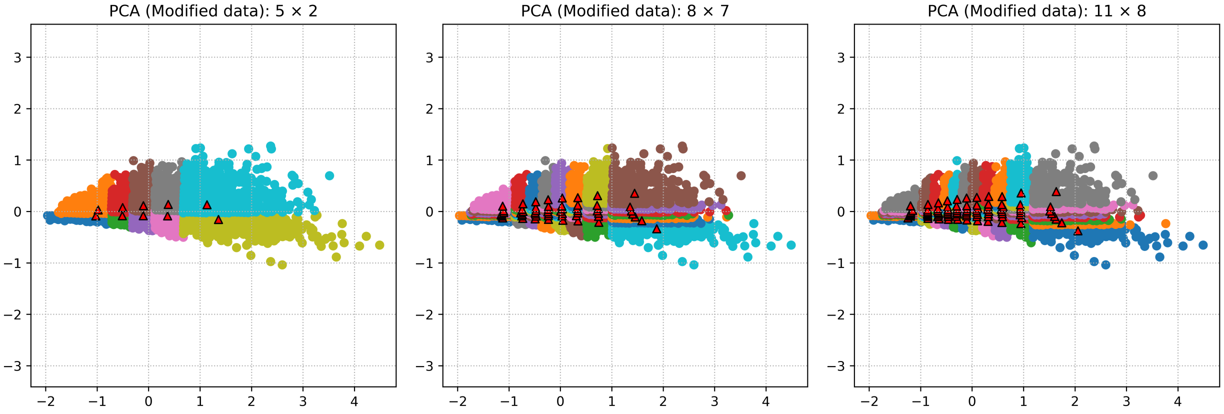


Figure 2: Data modification

3.1 Principal components grids

Three different configurations (5×2, 8×7, 11×8) of PCA were tested to observe how varying the grid resolution affected the data representation, as shown in Figure 3. Each configuration yielded unique insights into data structure, revealing how PCA can capture essential features across different scales.



(a)

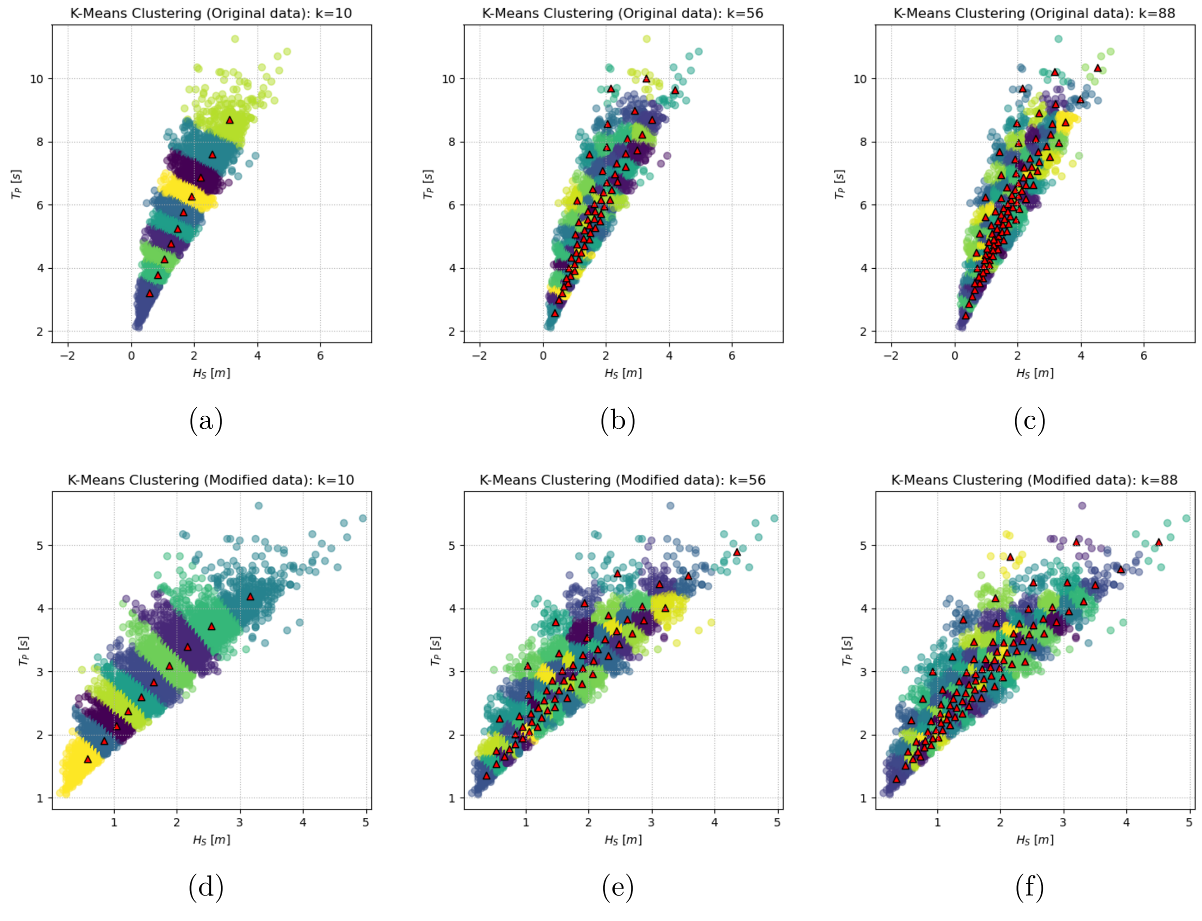
(b)

(c)

Figure 3: PCA (modified data set)

3.2 K-Means Clustering

K-means clustering was performed both on the original and the modified datasets with different numbers of clusters: 10, 56, and 88 (Figure 4). These experiments demonstrated how the choice of k influences the grouping and how modifications in the input data affect the clustering outcome. For each k, we generated visualizations to show how clusters adapted to the underlying data distribution. This helped in assessing the algorithm's sensitivity to the initial conditions and data transformations.



(a) (b) (c)

(d) (e) (f)

Figure 4: K-means clustering: original data (a), (b), (c); modified data (d), (e), (f)

3.3 Fatigue Damage Comparison

The effectiveness of the clustering approaches was assessed by comparing their ability to estimate the long-term fatigue damage against the exact calculated values (Figure 5). The analysis involved PCA with modified data, k-means with original data, and k-means with modified data. The exact damage calculations served as a benchmark to evaluate the accuracy of each method. This comparison highlighted the precision of each clustering method in capturing critical damage-inducing conditions and their potential to reduce simulation time while maintaining accuracy.

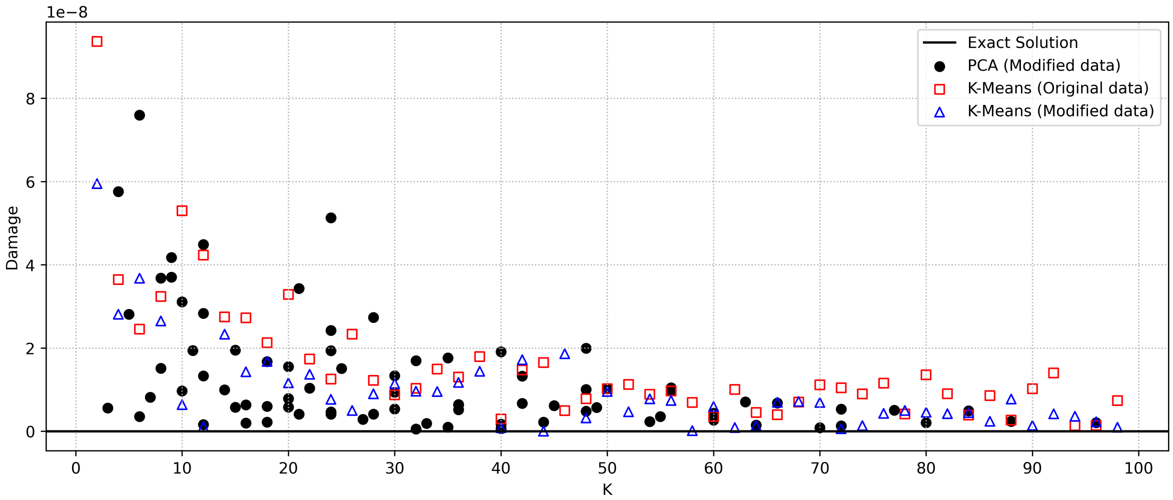


Figure 5: Fatigue damage vs. the number of clusters

3.4 K-Means vs Mean-Shift

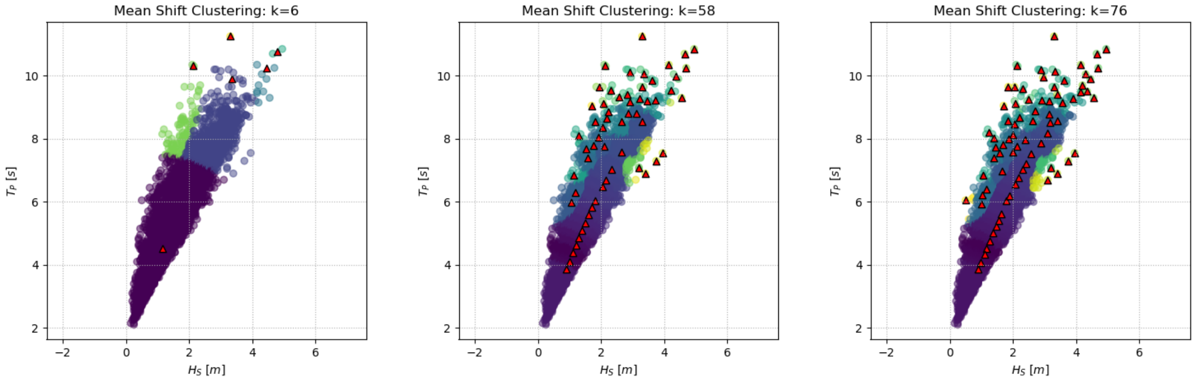
A comparative analysis of k-means and Mean-Shift clustering was conducted with the original dataset, focusing on different cluster counts. Figure 6 presents the cluster visualization generated by Mean-Shift algorithm for k values of 6, 58, 76. It is worth noting that calculating a scenario with 'k=6, 9, 10, 12, 14, 19, 21, 31, 42, 58, 76' clusters using k-means requires

8.45 seconds, whereas Mean-shift takes 3334.94 seconds.

Damage estimations for each method were plotted against the exact values to gauge their effectiveness, Figure 7. While both methods provided valuable insights into cluster-based damage estimation, k-means clusters were notably closer to the exact damage line across a broader range of k values. This suggests that k-means may be more reliable for predicting accurate fatigue damage in floating offshore wind turbines under various sea state conditions.

Additionally, Silhouette scores were calculated for each clustering setup to assess the quality of the clusters (Figure 7). The results indicated that for $k \leq 30$, Mean-Shift achieved higher silhouette scores, suggesting better cluster coherence and separation. This was particularly noticeable in scenarios with fewer clusters where the data's natural groupings were less fragmented. Conversely, for $k > 30$, k-means showed higher scores, reflecting its superior performance in managing larger cluster numbers and maintaining cluster quality even as the complexity of the clustering task increased.

This comparative analysis underscores the strengths of each clustering technique, with k-means providing a closer approximation to actual damage data, making it particularly useful for simulations that require high accuracy. Meanwhile, mean-shift's advantage in lower cluster configurations highlights its utility for exploratory data analysis where the exact number of clusters is unknown.

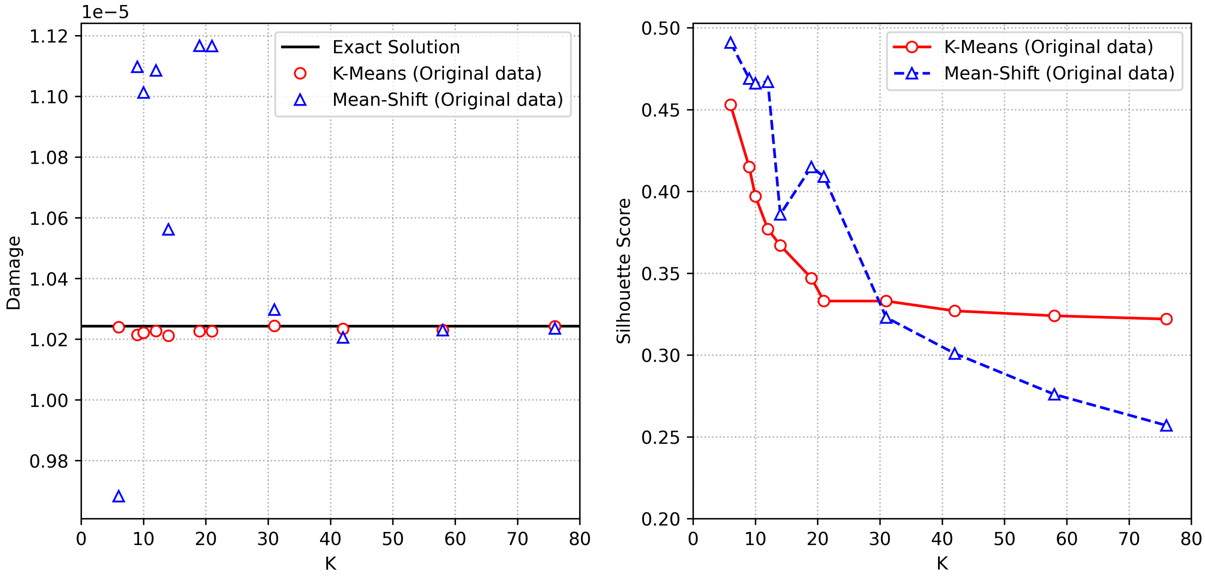


(a)

(b)

(c)

Figure 6: Mean-shift algorithm (original data)



(a)

(b)

Figure 7: K-means vs. Mean-shift; (a) Fatigue damage comparison, (b) Silhouette score

3.5 Discussion

The comprehensive analysis conducted in this study illuminates the applicability and efficiency of clustering algorithms in simplifying the environmental data crucial for FOWTs. While PCA and k-means have demonstrated robust capabilities in reducing data complexity and preserving essential information for structural analysis, other insights have emerged regarding the suitability and performance of different clustering techniques under various conditions.

K-means consistently provided results closest to the actual damage data, establishing its reliability for applications that demand high accuracy in fatigue damage prediction. On the other hand, mean-shift, while beneficial for its ability to autonomously determine the number of clusters, proved to be less efficient computationally. Notably, it took approximately one hour to run, which could be a significant drawback in scenarios where quick data processing is crucial.

Additionally, this study explored the use of Gaussian Mixture Models (GMM) and Bayesian Gaussian Mixture Models (BGMM). However, these methods did not perform as well in estimating the damage score compared to k-means and mean-shift. The likely reason for GMM and BGMM's underperformance is their sensitivity to the assumption of normally distributed components within the data. The real environmental conditions could be clustered based on multiple normal distributions; for example, different wind-wave directions may result in multiple normal distributions. However, due to the limitations of simulation model and simplifications made in this study, this kind of nature is not realized.

4 Conclusion

This study evaluated the effectiveness of various clustering algorithms for simplifying environmental data to predict long-term fatigue damage in floating offshore wind turbines. Our findings underscore the robust performance of k-means clustering, which aligns closely with exact damage calculations, making it a reliable choice for precise fatigue estimation. In contrast, mean-shift, while effective at determining natural clusters without predefined cluster counts, was hampered by its lengthy computational time. Gaussian Mixture Models and Bayesian Gaussian Mixture Models, despite their sophisticated approach, were less effective due to their sensitivity to non-normal data distributions and high computational demands. These insights highlight the importance of selecting appropriate clustering techniques that not only provide accurate damage estimations but also accelerate the whole design process. Future research should focus on enhancing the computational efficiency and adaptability of clustering methods to further support the more realistic and complex marine environmental data.

5 Acknowledgment

- **Ding Peng Liu:** Conceptualization, Methodology, Software, Writing
- **Kyungjung Cha:** Methodology, Software, Writing
- **Fan Wang:** Methodology, Software, Writing

References

[1] Taemin Heo et al. “Assessing fatigue damage in the reuse of a decommissioned offshore jacket platform to support a wind turbine”. In: *Journal of Offshore Mechanics and Arctic Engineering* 145.4 (2023), p. 042002.

[2] Paula Camus et al. “Analysis of clustering and selection algorithms for the study of multivariate wave climate”. In: *Coastal Engineering* 58.6 (2011), pp. 453–462.

[3] Samuel Kanner et al. “Maximum dissimilarity-based algorithm for discretization of metocean data into clusters of arbitrary size and dimension”. In: *International Conference on Offshore Mechanics and Arctic Engineering*. Vol. 51319. American Society of Mechanical Engineers. 2018, V010T09A062.

[4] Jarred Canning et al. *WEC Extreme Conditions Modeling Sandia Summer Intern 2017 Report*. Tech. rep. Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2017.

[5] Ding Peng Liu et al. “On long-term fatigue damage estimation for a floating offshore wind turbine using a surrogate model”. In: *Renewable Energy* (2024), p. 120238.

[6] Jason Jonkman et al. *Definition of a 5-MW reference wind turbine for offshore system development*. Tech. rep. National Renewable Energy Lab.(NREL), Golden, CO (United States), 2009.

[7] Amy Robertson and Jason Jonkman. *Definition of the Semisubmersible Floating System for Phase II of OC4*. Tech. rep. September. National Renewable Energy Laboratory Golden, CO, USA, 2014.

[8] Turan Dirlik. “Application of computers in fatigue analysis”. PhD thesis. PhD Thesis, University of Warwick, 1985.