ERASMUS +: ERASMUS MUNDUS MOBILITY PROGRAMME

Master of Science in

COASTAL AND MARINE ENGINEERING AND MANAGEMENT

CoMEM

INPUT REDUCTION ANALYSIS FOR LONG-TERM MORPHODYNAMIC SIMULATIONS

Delft University of Technology
July 3, 2017

Bruna de Queiroz

The Erasmus+: Erasmus Mundus MSc in Coastal and Marine Engineering and Management is an integrated programme including mobility organized by five European partner institutions, coordinated by Norwegian University of Science and Technology (NTNU).

The joint study programme of 120 ECTS credits (two years full-time) has been obtained at two or three of the five CoMEM partner institutions:

- Norges Teknisk- Naturvitenskapelige Universitet (NTNU) Trondheim, Norway
- Technische Universiteit (TU) Delft, The Netherlands
- Universitat Politècnica de Catalunya (UPC). BarcelonaTech. Barcelona, Spain
- University of Southampton, Southampton, Great Britain
- City, University London, London, Great Britain

During the first three semesters of the programme, students study at two or three different universities depending on their track of study. In the fourth and final semester an MSc project and thesis has to be completed. The two-year CoMEM programme leads to a multiple set of officially recognized MSc diploma certificates. These will be issued by the universities that have been attended by the student. The transcripts issued with the MSc Diploma Certificate of each university include grades/marks and credits for each subject.

Information regarding the CoMEM programme can be obtained from the programme coordinator:

Øivind A. Arntsen, Dr.ing.
Associate professor in Marine Civil Engineering
Department of Civil and Transport Engineering
NTNU Norway
Telephone: +4773594625 Cell: +4792650455 Fax: + 4773597021
Email: oivind.arntsen@ntnu.no

CoMEM URL: https://www.ntnu.edu/studies/mscomem

CoMEM Master Thesis

This thesis was completed by:
Bruna de Queiroz

Under supervision of:
Prof. dr. ir. A.J.H.M. Reniers, TU Deflt
Dr. ir. D. Walstra, TU Delft/Deltares
Dr. ir. S. Caires, Deltares
Ir. F. Scheel, Deltares
Ir. W. de Boer, TU Delft/Deltares

As a requirement to attend the degree of
Erasmus+: Erasmus Mundus Master in Coastal and Marine Engineering and Management
(CoMEM)

Taught at the following educational institutions:

Norges Teknisk- Naturvitenskapelige Universitet (NTNU)
Trondheim, Norway

Technische Universiteit (TU) Delft
Delft, The Netherlands

University of Southampton,
Southampton, Great Britain

At which the student has studied from August 2015 to July 2017.

Cover picture from http://www.noordwijk.com (Access in Jun 15, 2015)

SUMMARY

Morphodynamics is relevant for the coastal engineering field for many reasons, among them: safety of hinterland, shoreface and beach nourishment, generation of rip currents and nature development. Therefore, predictions of coastal morphology evolution are necessary to assess engineering solutions as well as understand coastal systems behaviors. Among the tools used to predict morphological evolution are the process-based models that make use of physical laws and empirical knowledge. Such models account for a considerable range of coastal processes and are rather complex, hence demanding substantial computational time. Usually, when using complex process-based models, reducing the size of the input parameters, named input reduction, is made necessary in order to reduce the computational effort.
The scope of the present study is to understand the influence of reduced wave climate on simulated morphological evolution. This is approached by focusing on four main research questions:

1. Which input reduction method presents the best performance in terms of morphology?

2. What is the influence of sequencing of the representative wave conditions on the performance of the input reduction methods?

3. What is the influence of the number of wave conditions on the performance of the input reduction methods?

4. How does the best input reduction method, sequencing and number of wave conditions performs in a validation case?

In this research, input reduction algorithms, sequencing methods, number of cases and duration of the reduced wave climate were investigated and evaluated with a 1D (cross-shore) brute force simulation of 3.3 years in Noordwijk, Netherlands. Noordwijk is a wave-dominated sandy beach characterized by a double sandbar system that has an inter-annual net offshore migration.
The assessment of the methods was carried out through cumulative skill scores, temporal evolution of profile perturbations (bars and troughs) and profiles at the end of the simulation. Furthermore, the findings on the Dutch coast were validated with a 1 year-long, 1D (cross-shore) brute force simulation in Anmok beach, South Korea. Anmok is a wave-dominated sandy beach with crescentic bars in the nearshore.
Firstly, 10 input reduction methods were analyzed with, initially, 12 representative wave conditions ($k = 12$).

- **Method 1: Conditions with the Largest Transport Contribution** - Initially, the wave observations are binned into wave height and wave direction bins. Then, the sediment transport contribution of each bin is determined and the "k" bins with the highest sediment transport contribution are selected.

- **Method 2: Grouping with Equal Sediment Influence** - The observations are aggregated in clusters with approximately the same sediment transport contribution.

- **Method 3: The Representative Wave Approach** - The observations are divided into sections over time (seasons) and each section has its averaged representative wave condition.

- **Method 4: Fixed Bins** - The observations are divided in pre-defined wave height and wave direction bins and the representative wave conditions are determined as the average wave conditions within the bins.

- **Method 5: Crisp k-means** - The centroids are selected based on an iterative process with the distance of the observations in relation to previous centroids. The method converges towards a stable solution which would be equal centroids from the last iteration and the previous one. The centroids are defined as the average of the observations within the cluster.

- **Method 6: Fuzzy k-means** - The principle of this method is similar to method 5, however the centroids are defined as the weighted average of all observations with the membership function as the weight. The membership function gives higher weights for observations that are closer to the pre-defined centroids.
- **Method 7: K-harmonic means** - Similar to method 6, but the weight in the definition of the centroids consists of the membership function and the dynamic weighting function. The latter gives more weight to observations that are farther from the pre-defined centroids.
- **Method 8: Maximum Dissimilarity** - The centroids are the observations with maximum distance among themselves.
- **Method 9: Energy Flux Method** - The observations are divided in pre-defined wave height and wave direction bins with equal amount of energy flux. The definition of the representative wave conditions is made by using the inverse function of the energy flux.
- **Method 10: Sediment Transport Bins Method** - The observations are divided in pre-defined wave height and wave direction bins with equal amount of sediment transport and the representative wave cases are defined as the average conditions of each bin.

Additionally, intra-methods variations regarding different input variables, definition of centroids and cluster initiation were also investigated.
Another variable analyzed was the time-scale that the reduced wave climate is applied. Each input reduction method was simulated with the following duration of the reduced wave climate: $T_{wc} = 1205$ days ($N_R = 1$); $T_{wc} = 602$ days ($N_R = 2$); $T_{wc} = 401$ days ($N_R = 3$); $T_{wc} = 301$ days ($N_R = 4$); $T_{wc} = 241$ days ($N_R = 5$) and $T_{wc} = 134$ days ($N_R = 9$), where T_{wc} is the duration of reduced wave climate and N_R is the number of repetitions of the reduced wave climate.
From the input reduction methods analysis, the sediment transport bins method stood out as the best input reduction method and it was selected for further sequencing and number of cases analysis. Four sequencing methods were used to determine the order of the wave conditions in the reduced wave climate. Among them are: random sequencing with 5 replicates, Markov Chain (MC) sequencing where the wave cases are ordered according to the highest Markov Chain transition probabilities, Monte Carlo Markov Chain (MCMC) sequencing with 5 replicates that orders the wave cases randomly with Markov Chain transition probabilities as weight, and finally, Monte Carlo Markov Chain with repetition (MCMCR) sequencing with 5 replicates which is similar to MCMC but allows wave cases to occur consecutively according to the number of repetition of the reduced wave climate (N_R).
Lastly, the influence of the number of representative wave conditions was analyzed by performing the sediment transport bins method with $k = 8$ (4 directional bins vs 2 wave height bins), $k = 10$ (2 directional bins vs 5 wave height bins), $k = 16$ (4 directional bins vs 4 wave height bins), $k = 24A$ (6 directional bins vs 4 wave height bins), $k = 24B$ (4 directional bins vs 6 wave height bins), $k = 24C$ (8 directional bins vs 3 wave height bins), $k = 32A$ (8 directional bins vs 4 wave height bins) and $k = 32B$ (4 directional bins vs 8 wave height bins).
The conclusions and recommendations obtained from this study are:

- Input reduction methods that have more control on the definition of bins such as pre-definition of wave height and wave direction bins perform better than methods that are based fully on the statistical properties of the wave dataset. Thus, the Sediment Transport Bins method, followed by the Energy Flux method are the ones recommended for input reduction applications.
- The order of the wave cases in the reduced wave climate must resemble at its best the natural variability of the full wave climate. The random sequencing is the best way of imposing variability in the reduced wave climate, however it should made aware of its possibility to aggregate energetic wave conditions.
- A less robust reduced wave climate (with less representative wave conditions) applied in a smaller time-scale performs better in terms of morphology than a more robust reduced wave climate (with more representative wave conditions) applied in longer time-scales. Therefore, it is recommended the application of input reduction methods with at least $k = 12$ and $T_{wc} = 100 - 150$ days.

PREFACE

This thesis completes the Master of Science in Coastal and Marine Engineering and Management (CoMEM) program carried out at Delft University of Technology, the Norwegian University of Science and Technology (NTNU) and the University of Southampton. The past two years have been a challenging but ultimately very rewarding experience. During this time, I have learned a lot and I have received a lot of support for which I'm grateful.

I would first like to thank my supervisory committee for all the guidance along the development of my research. I would like to thank Dr. Dirk-Jan Walstra for giving me the opportunity of carrying out an internship at Deltares that turned out to be the starting point of my Msc thesis. Thank you for the enlightening discussions about my research topic, for keeping me enthusiastic about it and for all the support on my academic carrier. To Prof. Ad Reniers, thank you for the help with the starting up of my internship and for chairing the committee, keeping me on track and for the insightful questions. To Freek Scheel, for developing the Input reduction tool, the helpful feedback and guidance on thinking critically about the methods. To Sofia Caires, for her applied research perspective that helped me in focusing on the practical aspects of my research. Thank you also to Wiebe de Boer for the guidance on handling the data of the CoMIDAS project. Also I would like to thank Derrick Olij, Lindino Benedet and João Dobrochinski for providing the initial scripts, they were of great help.

I would like to acknowledge KIOST that through the CoMIDAS project financed the Input reduction tool that was of great value for this research. I also would like to acknowledge EACEA/EU and the CoMEM Board for the funding and organization that made it possible for me to study in highly qualified institutions over Europe.

To the friends I made along the way, thank you for the great fun and memorable moments over this period. You made this master's degree a pleasant journey, despite dark and cold days. To Ben, for always making me smile. To my family and friends in Brazil for the effort on keeping in touch even with the distance and always rooting for me wherever I go.

B. de Queiroz
Delft, June 2017

CONTENTS

1

INTRODUCTION

1.1. MORPHODYNAMIC MODELS AND INPUT REDUCTION

Predictions of the evolution of coastal morphology are an important aspect for the coastal engineering sector because of its implications for safety, environment and financial aspects. For instance, coastal morphology influences nature and tourism through beach width. Besides that, coastal morphology also influences the occurrence of rip currents as well as protects the hinterland from storm events. Additionally, bar morphology, *i.e.* the net offshore bar migration phenomenon, has major implications to beach and dune morphological evolution (Guillen et al. 1999) and to the efficacy of shoreface nourishments (Grunnet & Ruessink 2005, Ojeda et al. 2008, van Duin et al. 2004).

There are many tools to predict the coastal morphological evolution, among them are the process-based morphodynamic models, which are very often used. Such models make use of physical principles (*e.g.* balance of forces, energy or transport) and empirical knowledge in such way that the equilibrium is consequence of the balance of forces and transport contributions. Usually, process-based morphodynamic models account for a considerable range of coastal processes such as wave generation and transformation, current generation and propagation, sediment transport and morphological changes that result in a high level of complexity and consequently on extensive computational effort. As complexity and, thus, computational effort increases, a process reduction or an input reduction is necessary in order to make morphodynamic modeling feasible for prediction of coastal morphology evolution.

Input reduction in morphology can be defined as the derivation of a reduced set of representative conditions that accurately approximates the long-term morphological evolution (Walstra et al. 2013). Input reduction can take place on wave, wind or tidal forcing. The present study focus on wave input reduction, hereafter referred as input reduction. Input reduction methods usually are based on a specific target such as longshore sediment transport or on statistical properties of the full dataset. A robust input reduction method consists of a reduced set of conditions that preserve some natural variability of the environment and hence, is able to represent accurately the full set of conditions.

1.2. MOTIVATION FOR RESEARCH

There are many techniques to perform input reduction and even though they have the same principle, they can vary greatly in relation to the selection of the representative wave conditions. Not only the selection of wave conditions can affect the morphological results, but also the sequencing of the cases and the number of cases influences the model results (Benedet et al. 2016). Although input reduction is a common practice in morphodynamic modeling, there is not a well-defined guideline on the application of input reduction methods around the world and there is a lack of understanding the consequences of input reduction on morphology. The effect of input reduction on the predicted morphological evolution is often not taken into account even though it influences beach, dune and shoreface nourishment development affecting the predictability of shoreline evolution (Walstra et al. 2011).

1.3. STUDY OBJECTIVES AND RESEARCH QUESTIONS

The main objective of this study is to understand the influence of a reduced wave climate on the simulated morphological evolution. In order to achieve that, the following aims of the study are established:

1. Which input reduction method presents the best performance in terms of morphology?
2. What is the influence of sequencing of the representative wave conditions on the performance of the input reduction methods?
3. What is the influence of the number of wave conditions on the performance of the input reduction methods?
4. How does the best input reduction method, sequencing and number of wave conditions performs in a validation case?

1.4. OUTLINE AND APPROACH

In order to achieve the study goals, primarily, 10 input reduction methods were analyzed in terms of the wave-driven morphological evolution of nearshore sandbars on the time scale of their quasi-cyclic-offshore-directed behavior in Noordwijk, Netherlands. The area is characterized by a double sandbar system that propagates offshore on the time scale of years (Wijnberg & Terwindt 1995). The bar dynamics in Noordwijk is classified as inter-annual net offshore migration (Ruessink et al. 2009, Walstra et al. 2013) where the cycle from bar inception in the swash zone to bar decay in the outer surf zone takes around 3 to 4 years and it is not really affected by individual wave events and chronology (Walstra et al. 2013).
Besides the input reduction methods (*i.e.* how to select representative wave cases), the sequencing of the selected wave cases and the number of representative wave cases are analyzed. The input reduction methods, sequencing and number of representative wave cases are evaluated with calibrated long-term one-dimensional (cross-shore) predictions performed with Unibest-TC forced with the full wave climate (brute force) for Noordwijk, Netherlands (Walstra et al. 2011).
Finally, the findings of the analysis on the Dutch coast are validated on Anmok beach, East coast of South Korea. In Figure 1.1 is a diagram that illustrates the framework performed in this study.

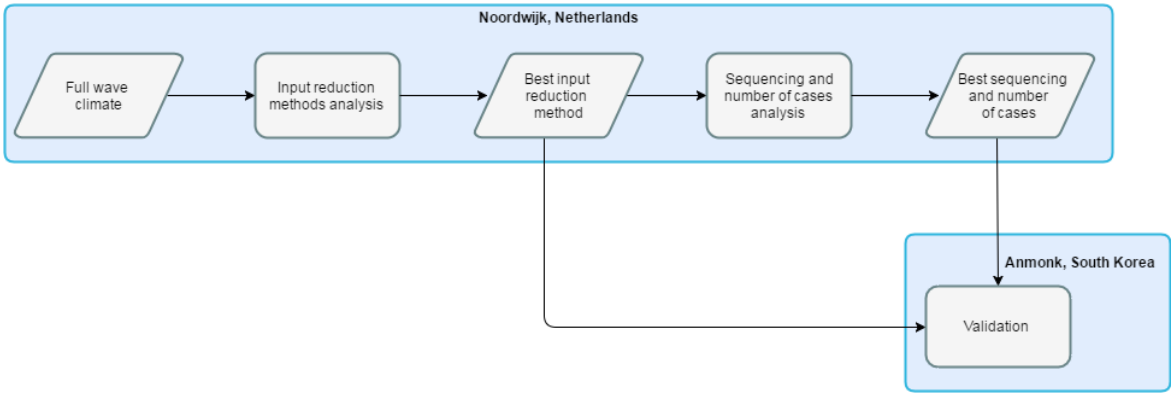


Figure 1.1: Flowchart of framework applied in the present study.

This report is structured as follows: first, the methodology applied in this study is described in chapter 2. In this chapter, the input reduction framework, the input reduction methods, sequences and number of representative cases are described as well as its evaluation and validation. Next, the results of the evaluation of the methods, sequences and number of representative cases as well as the validation are analyzed and reported in chapter 3 followed by a discussion and reasoning of the results of this study in chapter 4. The main conclusions and recommendations are presented in Chapter 5 and Chapter 6, respectively.

2

METHODOLOGY

The methodology applied in this study follows the input reduction framework proposed by Walstra et al. (2013):

1. Selection of the input reduction period;
2. Selection of the representative wave conditions – input reduction methods;
3. Sequencing of the selected conditions;
4. Determination of the wave climate duration.

2.1. SELECTION OF THE INPUT REDUCTION PERIOD

The input reduction period is the time-length of the original time-series that is used to perform the input reduction. The upper limit of the reduction period should be the time-scale related to the inherent morphological variation while the lower limit should be that it covers at least one year in order to account for the seasonal variations of the wave climate (Walstra et al. 2013). In this study the input reduction period is 3.3 years.

2.2. INPUT REDUCTION METHODS

There are mainly 2 categories of input reduction methods. Methods where a sediment proxy is used as a reference for the performance of the input reduction and methods without sediment proxy where the basis of the input reduction is on retaining the dataset's statistical properties. All methods aim to find a number k of wave conditions that best represent the full observed dataset.
A wave climate is composed by observations that describe the sea state over time. In a multivariate approach, the sea state is described with multiple variables such as significant wave height, wave peak period, dominant wave direction, wind direction, wind velocity and duration of storm events. The combination of these variables, not necessarily all of them, at a certain time is called a wave condition. In this study the variables used to describe a wave condition are: root-mean-square wave height (H_{rms}), peak wave period (T_p) and mean wave direction (θ).

2.2.1. NOTATION, NORMALIZATION AND DISTANCE MEASURES

NOTATION

In the description of the reduction methods analyzed in this study, a standard notation is used according to Olij (2015). The wave condition is described by a vector, x_i, with multiple variables, $x_i = x_{i,1}, x_{i,2}, \ldots, x_{i,z}$, where z is the number of variables used to describe a wave condition. In this study, a wave condition is described by the root-mean-square wave height (H_{rms}), peak wave period $\left(T_p\right)$ and mean wave direction with respect to the shore-normal (θ). So, in this case $z = 3$ and $x_i = H_{rms,i}, T_{p,i}, \theta_i$. All the wave conditions available are collected in the database $X = x_1, x_2, \ldots, x_N$ where N is the total number of wave conditions. All reduction methods describe the original dataset X with k representative wave conditions, called centroids. Centroid is defined as a vector that describes a representative wave condition, $v_j = v_{j,1}, v_{j,2}, \ldots, v_{j,}$. In this

study, $z = 3$, $v_j = H^{rep}_{rms,j}, T^{rep}_{p,j}, \theta^{rep}_j$, where the superscript "*rep*" means representative wave condition. The centroids are collected in the database $V = v_1, v_2, \ldots, v_k$.

DISTANCE MEASURES

Some reduction methods use distance measures, which is a technique to determine the distance between 2 points in a multivariate space. The distance is use to determine how similar the observations are to each other: the smaller the distance between 2 observations, the more similar they are. In this study, the Euclidean distance is used once the observations are time independent and can be freely located in the complete observation space. A distance measure between an observation x_i and a centroid v_j is indicated by: $||x_i, v_j||$. The Euclidean distance can be generalized as:

$$||x_i, v_j|| = \sqrt{\left(H'_i - H^{rep'}_j\right)^2 + \left(T'_i - T^{rep'}_j\right)^2 + \left(min\left(\left|\theta'_i - \theta^{rep'}_j\right|, 2 - \left|\theta'_i - \theta^{rep'}_j\right|\right)\right)^2} \tag{2.1}$$

where the accent indicates that the input variables are normalized.

NORMALIZATION

The distance between observations is determined in a space that is determined by multiple variables. Therefore, all reduction methods that use a distance function also use normalized input data. Non-circular variables such as wave height and wave period as scaled to values from 0 to 1 by a linear transformation:

$$H' = \frac{H - min(H)}{max(H) - min(H)} \tag{2.2}$$

$$T' = \frac{T - min(T)}{max(T) - min(T)} \tag{2.3}$$

Circular variables such as wave direction are transformed differently. This variable describes a circle where the distance between two values can be measure clockwise or counter clockwise. Thus, the maximum difference between 2 radial angles is π. In this way the distance between 2 angles are normalized between 0 and 1 while the absolute values of the angle are normalized between 0 and 2:

$$\theta' = \frac{\theta}{2} \tag{2.4}$$

2.2.2. METHOD 1: CONDITIONS WITH THE LARGEST TRANSPORT CONTRIBUTION

This method uses the sediment transport as target. First, it distributes the observed wave conditions in user-predefined wave height and direction classes. Each class has a representative wave condition. Then is calculated the sediment transport for each class by summing the sediment transport of the waves in the class and the total sediment transport of all wave conditions. After that, it is estimated how much each class contributes to the total sediment transport in order to obtain the relative influence of a certain class in the total transport. The relative influence of the waves in a class i (W_i) is calculated by:

$$W_i = \frac{P(H = H_{rms,i}, \theta = \theta_i) S(H_{rms,i}, \theta = \theta_i)}{S_{tot}} \tag{2.5}$$

$$S_{tot} = \sum_{i=1}^{N} P(H = H_{rms,i}, \theta = \theta_i) S(H_{rms,i}, \theta = \theta_i) \tag{2.6}$$

where $P(H = H_{rms,i}, \theta = \theta_i)$ is the probability of occurrence of a wave class, $S(H_{rms,i}, \theta = \theta_i)$ is the sediment transport of the class (generated by the representative wave condition) and S_{tot} is the total sediment transport generated by all wave conditions.

Next, the k classes with the highest contribution to the total sediment transport are selected. The reduced wave climate consists of the representative wave conditions of the wave classes. In order to obtain a well-balanced wave climate the probability of occurrence of the reduced wave conditions has to be scaled up in order to guarantee that the total sediment transport caused by the reduced wave climate is the same as the sediment transport caused by the full wave climate:

$$P_{up,i} = P(H_{rms,i}, \theta_i) \frac{S_{tot}}{\sum_{i=1}^{k} P(H_{rms,i}, \theta_i) S(H_{rms,i}, \theta_i)} \tag{2.7}$$

Where $P_{up,i}$ is the new probability of occurrence of the waves in class i.

The sediment transport adopted in this study is the longshore component of the sediment transport obtained from the simulation of the full wave climate, namely brute force, without bed changes computation.

2.2.3. METHOD 2: GROUPING WITH EQUAL SEDIMENT INFLUENCE

This method aggregates the observations in clusters with approximately the same sediment transport contribution. It starts by selecting k initial wave conditions as individual clusters, then in every iteration each cluster picks the closest observation that will become part of it until a sediment transport threshold is reached. The threshold is defined by the total sediment transport divided by the number of representative cases. In the end, the result is k clusters with approximately the same amount of sediment transport. The following steps for the algorithm are described:

1. Determine the sediment transport for each wave condition observed ($S(x_i)$), and then the total sediment transport as a sum of each individual sediment transport ($S_{abs,tot} = \sum_{i=1}^{N} |S(x_i)|$) . Assuming that the clusters will represent an equal amount of sediment transport, the sum of the absolute sediment transport of a certain cluster will be approximately equal to the total sediment transport divided by the amount of clusters ($S_{clu,j} = S_{tot}/k$). In this way, the algorithm will generate k clusters that all represent a certain number of observations with equal amount of sediment transport.

2. Select the farthest k observations as the initial clusters centroids V. This step makes use of Maximum Dissimilarity Method, further describe in section 2.2.9.

3. Start of the iterative process. In every iteration, every cluster selects an observation that is not in a cluster yet and has the smallest distance to the cluster centroid keeping the amount of sediment transport of the cluster equal to $S_{clu,j} < S_{abs,tot}/k$. After every iteration, the cluster centroids V are updated by taking the average values of the wave conditions that constitutes the cluster.

4. When no observations can join a cluster anymore (because $S_{clu,j} \geq S_{abs,tot}/k, j = 1,...k$) the remaining observations join the cluster to which they have the smallest distance to.

Again, the sediment transport adopted as input for this method is the longshore component of the sediment transport obtained from the brute force simulation without bed changes computation.

2.2.4. METHOD 3: THE REPRESENTATIVE WAVE APPROACH

This method was introduced by Brown & Davies (2009). In this method, the dataset is divided in sections over time, *e.g.* season. Thus, the chronology of the original dataset is retained. In each section, a representative wave height, wave direction and wave period is determined by taking the weighted average value of wave height and unweighted average of wave direction and wave period. The representative wave height of a section is determined as follows:

$$v_{j,1} = H_{rms,j}^{rep} = \left(\frac{\sum_{i:x_i \in C_j} \left(f_i H_{rms,i}^{p} \right)}{\sum_{g:x_g \in C_p} f_i} \right)^{1/p} \tag{2.8}$$

Where f_i is the frequency of occurrence of the root-mean-square wave height $H_{rms,i}$ of an observed wave condition x_i. C_j is the database of observations of section j, and p is the coefficient that represents the non-linearity between waves and sediment transport, also known as the power of the sediment transport that varies from $p = 2$ to $p = 3$. In this study, $p = 2.5$.

2.2.5. METHOD 4: FIXED BINS

This method uses bins with predefined boundaries. The user defines the number of directional and wave height bins. The algorithm first divides the observations within the number of directional bins. Next, each directional bin is divided according to the number of wave height bins and the maximum wave height of the observations of each directional bin. This results in equidistant directional bins but not necessarily equidistant wave height bins once the resolution of the wave height bins can vary among the directional classes. The representative wave conditions are determined as the average wave conditions within the bins.

2.2.6. METHOD 5: CRISP K-MEANS

This method is based on the statistical properties of the dataset. The method starts with k randomly selected centroids, then all observations X are assigned to the cluster that they have the smallest distance to. Next, the centroids are updated by averaging the wave conditions that constitutes the clusters. The crisp k-means algorithm has a 'hard' membership function which means that observations can only be a member of one

cluster. The hard membership function can lead to poor clustering and makes the method very dependent on the initial centroids. There are some measures to bypass these problems. One solution is to use a soft membership function which is adopted in the next methods. Another solution is increase to the accuracy and independence of initiation by choosing in a smart way the initial centroids leading the algorithm in the right direction. A third solution is to repeat the algorithm several times, so that the different outcomes can be compared and the outcome with the lowest difference will be then selected as the final reduced wave climate. In this study the k-means++ algorithm (Arthur & Vassilvitskii 2007) is used to selected the initial centroids values. In the following, there is a description of the algorithm:

1. Select the initial centroids using k-means++ algorithm (Arthur & Vassilvitskii 2007):

   Draw uniformly an observation from X. This will be the first centroid v_1;

   Compute the distance between v_1 and all observations of X, $D = ||x_i, v_1||, i = 1, ...N$

   Draw the second centroid v_2 from X with probability:

$$P(x_i = v_2) = \frac{||x_i, v_1||^2}{\sum_{j=1}^{N} ||x_j, v_1||^2}, i = 1, \ldots, N \tag{2.9}$$

   Assign the observations to the cluster for which the distance between the observation and the clusters centroid is the smallest. Draw the remaining centroids iteratively with the probability:

$$P(xi = v_g) = \frac{||x_i, v_p||^2}{\sum_{g:x_g \in C_p} ||x_i, v_p||^2,} g = 1, \ldots, k-2, i = 1, \ldots, N, p = 1, \ldots., k-g \tag{2.10}$$

   Where C_p is the database of observations that are in the cluster p.

2. Compute the distance between all observations and all centroids. Assign every observation to the cluster which centroid has the smallest distance to the observation;

3. Determine the new representative wave conditions V by averaging the wave conditions of the observations in the cluster:

$$v_g = \frac{\sum_{g:x_g \in C_p} x_g}{N_p}, p = 1, \ldots, k \tag{2.11}$$

   Where N_p = number of observations in the cluster p.

4. Determine the internal error for each iteration (ϵ_{it})

$$\epsilon_{it} = \frac{\sum_{p=1}^{k} \sum_{g:x_g \in C_p} ||x_g, v_p||}{N} \tag{2.12}$$

   And by that the improvement of V over the iteration:

$$\Delta\epsilon = \epsilon_{it-1} - \epsilon_{it} \tag{2.13}$$

5. Repeat steps 2, 3 and 4 until the value of improvement of V over the iteration ($\Delta\epsilon$) is smaller than a minimal value predefined by the user (ϵ_{min}):

$$\Delta\epsilon \leq \epsilon_{min} \tag{2.14}$$

   Sensitivity analysis has led to $\epsilon_{min} = 10^{-5}$. More details of the sensitivity analysis and parameter choice can be found in Appendix A.

2.2.7. METHOD 6: FUZZY K-MEANS

This method is also based on the statistical properties of the dataset and it is a variant of the Crisp k-means, however it has a soft membership function which means that this algorithm allows observations to be assigned to more than one cluster. In other words, all observations have some influence on the definition of the centroids determined by the fuzzy membership function. The description of the algorithm according to Chang et al. (2011) is as follows:

1. Select the initial centroids using k-means++ algorithm (Arthur & Vassilvitskii 2007):

   Draw uniformly an observation from X. This will be the first centroid v_1;

   Compute the distance between v_1 and all observations of X, $D = ||x_i, v_1||, i = 1,...N$

   Draw the second centroid v_2 from X with probability:

$$P(x_i = v_2) = \frac{||x_i, v_1||^2}{\sum_{j=1}^{N} ||x_j, v_1||^2}, i = 1,\ldots,N \tag{2.15}$$

   Assign the observations to the cluster for which the distance between the observation and the clusters centroid is the smallest. Draw the remaining centroids iteratively with the probability:

$$P(xi = v_g) = \frac{||x_i, v_p||^2}{\sum_{g:x_g \in C_p} ||x_i, v_p||^2,} g = 1,\ldots,k-2, i = 1,\ldots,N, p = 1,\ldots.,k-g \tag{2.16}$$

2. Determine how much each observation belongs to every cluster trough the membership function, M. M is a Nxk matrix that quantifies these memberships. The weights in this matrix are restricted by: $0 \leq M_{i,g} \geq 1$. And the total contribution of one observation sums up to one: $M_i = \sum_{j=1}^{k} M_{i,g} = 1$.

$$M_{i,g} = \frac{\left(1/||x_i, v_j||^2\right)^{1/(o-1)}}{\sum_{j=1}^{k} \left(1/||x_i, v_j||^2\right)^{1/(o-1)}} \tag{2.17}$$

   where o = fuzzy parameter.

3. Determine the new representative wave conditions V:

$$v_g = \frac{\sum_{i=1}^{N} \left(M_{i,g}^{o}\right) x_i}{\sum_{i=1}^{N} \left(M_{i,g}\right)^{o}} \tag{2.18}$$

4. Determine the internal error for every iteration (ϵ_{it}):

$$\epsilon_{it} = \frac{\sum_{i=1}^{N} \sum_{g=1}^{k} \left(M_{i,g}\right)^{o} ||x_i, v_j||}{\sum_{i=1}^{N} \sum_{g=1}^{k} \left(M_{i,g}\right)^{o}} \tag{2.19}$$

   From the internal error, the improvement of V over the iteration can be determined by:

$$\Delta\epsilon = \epsilon_{it-1} - \epsilon_{it} \tag{2.20}$$

5. Repeat the steps 2, 3, and 4 until $\Delta\epsilon \leq \epsilon_{min}$, where ϵ_{min} is the minimal internal error value predefined by the user.
   Sensitivity analysis has led to $\epsilon_{min} = 10^{-5}$ and $o = 1.5$. More details of the sensitivity analysis and parameters choice can be found in Appendix A.

2.2.8. METHOD 7: K-HARMONIC MEANS

The K-harmonic means method is another variation of the Crisp k-means method with an adaptation to decrease the dependency on initiation. Therefore is also based on the statistical properties of the dataset. This method can be seen as an extension of the fuzzy k-means method where all observations have some influence (weight) on the definition of centroids. The total weight of each observation is defined as a balance between the membership function and a dynamic weighting function. The dynamic weight accounts for the larger influence that outliers should have on determining the centroids rather than observations that are closer to the centroids. The dynamic weighting of the observations is done by a method called boosting (Zhang 2000, Zhang et al. 1999) where more weight is given to data points that are not well-clustered yet, in other words, that are far from the centroids. The k-harmonic algorithm has been indicated as accurate results while is almost independent of its initiation (Zhang 2000, Zhang et al. 1999).

1. Select the centroids using the k-means++ algorithm (Arthur & Vassilvitskii 2007):

   Draw uniformly an observation from X. This will be the first centroid v_1;

   Compute the distance between v_1 and all observations of X, $D = ||x_i, v_1||, i = 1,...N$

   Draw the second centroid v_2 from X with probability:

$$P(x_i = v_2) = \frac{||x_i, v_1||^2}{\sum_{j=1}^{N} ||x_j, v_1||^2}, i = 1, \ldots, N \tag{2.21}$$

   Assign the observations to the cluster for which the distance between the observation and the clusters centroid is the smallest. Draw the remaining centroids iteratively with the probability:

$$P(xi = v_g) = \frac{||x_i, v_p||^2}{\sum_{g:x_g \in C_p} ||x_i, v_p||^2,} g = 1, \ldots, k-2, i = 1, \ldots, N, p = 1, \ldots., k-g \tag{2.22}$$

2. Determine how much each observation belongs to every cluster trough the membership function, M. M is a Nxk matrix that quantifies these memberships. The weights in this matrix are restricted by: $0 \leq M_{i,g} \geq 1$. And the total contribution of one observation sums up to one: $M_i = \sum_{j=1}^{k} M_{i,g} = 1$.

$$M_{i,g} = \frac{\left(||x_i, v_j||\right)^{-o-2}}{\sum_{j=1}^{k} \left(||x_i, v_j||\right)^{-o-2}}, i = 1, ..., N, g = 1, ..., k \tag{2.23}$$

   where o is the fuzzy parameter.

3. Determine the dynamic weight of the observations. Some observations will have more influence on the determination of the centroids than others (boosting). K is a $Nx1$ matrix that contains these weights.

$$K_i = \frac{\sum_{j=1}^{k} ||x_i, v_j||^{-o-2}}{\left(\sum_{j=1}^{k} ||x_i, v_j||^{-o}\right)^2} \tag{2.24}$$

4. Determine the new representative wave conditions V:

$$v_g = \frac{\sum_{i=1}^{N} M_{i,g} K_i x_i}{\sum_{i=1}^{N} M_{i,g} K_i}, g = 1, ..., k \tag{2.25}$$

5. Determine the internal error for every iteration ϵ_{it}:

$$\epsilon_{it} = \sum_{i=1}^{N} \frac{k}{\sum_{j=1}^{k} \frac{1}{||x_i, v_j||^o}} \tag{2.26}$$

   From the internal error, the improvement of V over the iteration can be determined by:

$$\Delta\epsilon = \epsilon_{it-1} - \epsilon_{it} \tag{2.27}$$

6. Repeat the steps 2, 3, and 4 until $\Delta\epsilon \leq \epsilon_{min}$, where ϵ_{min} is the minimal internal error value predefined by the user.
   Sensitivity analysis has led to $\epsilon_{min} = 10^{-5}$ and $o = 4.2$. More details of the sensitivity analysis and parameters choice can be found in Appendix A.

2.2.9. METHOD 8: MAXIMUM DISSIMILARITY – MAXMIN

This method was introduced by Kennard & Stone (1969) and its principle is to create a subset V of k centroids that represents the full diversity of the observations. Moreover, this is done by maximizing the dissimilarity between the vectors in the subset. There are many methods of measuring the dissimilarity between vectors. One of the most accurate and efficient variant of the Maximum Dissimilarity Algorithm (MDA) is the MaxMin Algorithm (Willett 1996):

1. Determine the distances D of all observations in the full data sample X with each other:

$$D_{i,m} = ||x_i, x_m||^2, i = 1..., N \tag{2.28}$$

This distance can be interpreted as the dissimilarity between the observations. The initial centroid v_1 is the observation with the largest total distance in relation to the other observations.

2. Determine the dissimilarity of the remaining observations and the subset of centroids v_1:

$$D_{i,v_1} = ||x_i, v_1||^2, i = 1, ..., N \tag{2.29}$$

The observation with the largest dissimilarity is chosen as v_2.

3. To determine the values of the next subset of centroids $v_3, v_j, ..., v_k$ the efficient algorithm of Polinsky et al. (1996) is used. The observation with the largest D^{min}_{i,v_j} is chosen at the next centroid:

$$D^{min}_{i,v_j} = min\left[D_{i,v_j}, D^{min}_{i,v_{j-1}}\right], i = 1, ..., N-j, j = 3, ..., k \tag{2.30}$$

To start this iterative procedure $D^{min}_{i,v_2} = D_{i,v_1}$.

2.2.10. METHOD 9: ENERGY FLUX METHOD

In this method, the wave conditions are divided in bins of equal energy flux (Dobrochinski 2009). The wave energy flux is calculated for each wave condition by the formula:

$$E_f = \left(\frac{\rho g H^2}{8}\right) C_g \tag{2.31}$$

where:
E_f = energy flux;
ρ = water density;
H = wave height;
C_g = group wave celerity;
With this method, the wave bins present high resolution and small intervals in conditions with more wave energy and low resolution and large intervals in conditions with less wave energy. The wave conditions that represents each bin are calculated by the mean wave energy flux concept. The period is determined by the average of the bin while the wave height and wave direction are determined according to the inverse function of the bin's mean wave energy flux.

2.2.11. METHOD 10: SEDIMENT TRANSPORT BINS METHOD

This method aggregates the observations in groups of equal sediment contribution likewise method 2 (section 2.2.3), however the clustering process is different, being similar to the Energy flux method (section 2.2.10). First the observations are divided in directional bins of equal amount of sediment transport. In this step negative and positive sediment transport rates are treat separately in order to avoid averaging out wave conditions that generate opposite directed sediment transport. In other words, the directional bins are divided in such a way that any bin crosses the shore-normal angle. Next, each directional bin is divided in equal sediment transport bins. The representative wave conditions are determined by the average of the wave conditions of each bin.

2.2.12. INTRA-METHODS VARIATIONS

Some of the above methods present different possible combinations regarding input variables, definition of centroids and cluster initiation methods. All methods that use H_{rms},T_p,θ as input variables can also be set-up with H^p_{rms}, T_p, θ and S_y, T_p, θ as input variables, where S_y is the longshore sediment transport obtained from the brute force simulation and p is the power that represents the non-linearity between wave height and sediment transport, here assumed $p = 2.5$.
When using H^p_{rms}, T_p, θ as input variables, the definition of the centroids within a bin is defined by a non-linear weighting formula for the wave height and average for the wave period and wave direction:

$$v_{j,1} = H^{rep}_{rms,j} = \left(\frac{\sum_{i:x_i \in C_j}\left(f_i H^p_{rms,i}\right)}{\sum_{g:x_g \in C_p} f_i}\right)^{1/p} \tag{2.32}$$

To obtain the centroids as H_{rms}, T_p, θ values, the inverse function $H_{rms} = \left(H_{rms}^{p}\right)^{1/p}$ is applied. When using S_y, T_p, θ as input variables, the centroids are defined by the average of the observations within the bin and to obtain the values of H_{rms}, T_p, θ the nearest observation of the centroid with S_y value is used.
Additionally, the methods Crisp k-means (Method 5), Fuzzy k-means (Method 6) and K-harmonic means (Method 7) are initiated by the k-means++ algorithm that contain randomness in the selection of the initial centroids and thus, is not possible to reproduce the same selection of centroids. This is called inconsistency and in order to compared consistent methods, 2 different cluster initiations are analyzed: Fixed bins and Maximum dissimilarity (van Arkel 2016).
Figure 2.1 presents all the combinations of the methods analyzed in this study.

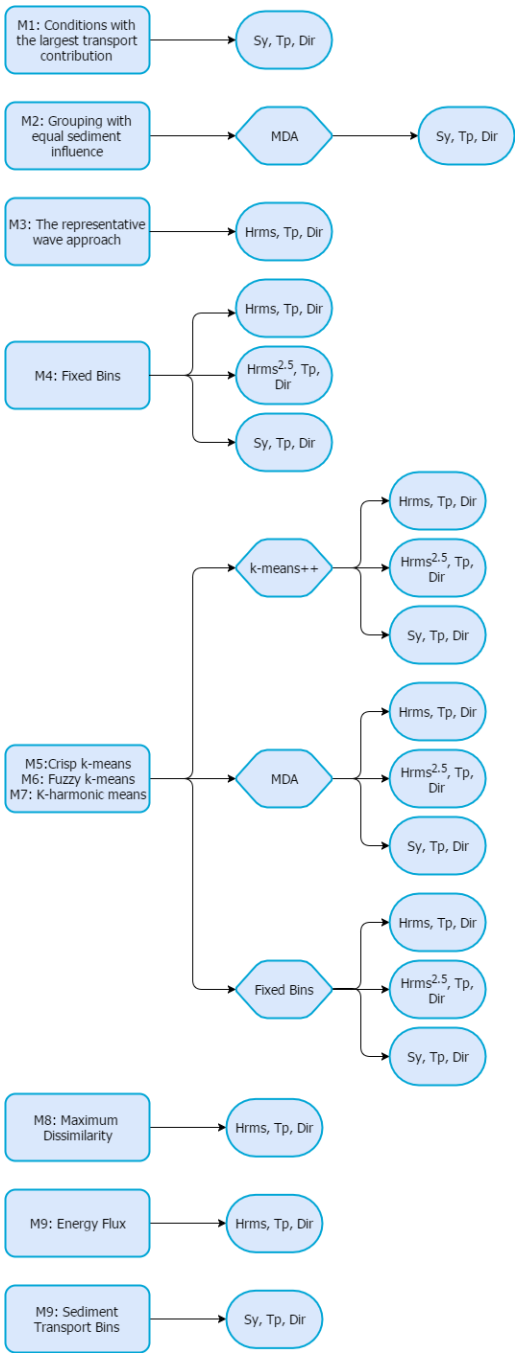


Figure 2.1: Combinations of possible variations of the methods analyzed.

2.3. WAVE CLIMATE DURATION

The wave climate duration indicates the time-scale in which the reduced wave climate is applied. It is determined by the sum of the duration of each representative wave case:

$$T_{wc} = \sum_{j=1}^{k} T_{wc,j} = \sum_{j=1}^{k} f_{rep,j} \frac{T_R}{N_R} \tag{2.33}$$

where T_R = input reduction period and N_R = number of repetitions of the reduced wave climate. $T_R = 3.3$ years in the present study. The number of repetitions is important once affects the duration of each wave condition. For instance, one wave condition with duration of 5 days presents different morphological response than the same wave condition forced for 2.5 days, but repeated two times, not consecutively, especially if one consider energetic wave conditions. The duration of a reduced energetic wave case should be similar as possible to the reality in order to avoid irreversible morphological changes. On the other hand, the reduced wave climate should be repeated the least possible in order to avoid a high number of transitions between conditions, NoT defined as $NoT = NoC * N_R - 1$, where NoC = number of reduced wave conditions. The higher the number of transitions, higher the computational demand.
In this study, the duration of the reduced wave climate analyzed are: $T_{wc} = 1205$ days ($N_R = 1$); $T_{wc} = 602$ days ($N_R = 2$); $T_{wc} = 401$ days ($N_R = 3$); $T_{wc} = 301$ days ($N_R = 4$); $T_{wc} = 241$ days ($N_R = 5$) and $T_{wc} = 134$ days ($N_R = 9$) according to Walstra et al. (2013).

2.4. NUMBER OF CASES

Initially, the number of representative wave cases considered is $k = 12$ (4 directional bins vs 3 wave height bins for methods that classify the observations in wave height and wave direction bins). After the methods analysis, the best method(s) and the different sequencing methods are simulated with number of cases $k = 8$ (when applicable, 4 directional bins vs 2 wave height bins), $k = 10$ (when applicable, 2 directional bins vs 5 wave height bins), $k = 16$ (when applicable, 4 directional bins vs 4 wave height bins), $k = 24A$ (when applicable, 6 directional bins vs 4 wave height bins), $k = 24B$ (when applicable, 4 directional bins vs 6 wave height bins), $k = 24C$ (when applicable, 8 directional bins vs 3 wave height bins), $k = 32A$ (when applicable, 8 directional bins vs 4 wave height bins) and $k = 32B$ (when applicable, 4 directional bins vs 8 wave height bins).

2.5. SEQUENCING REDUCED WAVE CLIMATES

The order of the wave conditions in the input reduction can influence the morphological results, specially for locations where the chronology of wave conditions is relevant for the morphological evolution. In this study, randomly sequenced reduced wave climates are initially considered with 5 replicates for each method. After a pre-selection of the most promising methods, the Markov Chain (MC), Monte Carlo Markov Chain (MCMC) and Monte Carlo Markov Chain with repetition (MCMCR) sequencing are analyzed.

2.5.1. MARKOV CHAIN SEQUENCING

The Markov Chain sequencing (MC) orders the representative wave conditions in the way they most likely would occur in the full dataset. The procedure is described as follows:

1. Number the representative wave conditions stored in the database V from 1:k;

2. Determine for every observation of the database X which of the representative wave conditions in V is most similar to it. In this step, a new vector F is created with size $Nx1$ where the number of the wave conditions that is most similar to each observation is stored:

$$F_i = \sum_{j=1}^{k} jI\left[||x_i, v_j|| = min(||x_i, V||)\right], j = 1,...,k, i = 1,...,N \tag{2.34}$$

   Where $I[.]$ is a true-false indicator that is 1 when the equation between brackets is true and 0 when it is false.

3. Determine the Markov transitions for the wave conditions in F. The Markov transitions are stored in a Markov transition matrix M of size kxk:

$$M(m,n) = P\left(F_{t+1} = n | f_t = m\right) = \sum_{t=1}^{N-1} \frac{I\left[F_t = m\right] I\left[F_{t+1} = n\right]}{N-1}, m = 1,...,k, n = 1,...,k \tag{2.35}$$

4. Define two time-series matrices: A_S and A_{NS}. The first matrix, A_S starts empty and will contain the numbers that are assigned to the wave conditions in step 1 in the sequence determined by the algorithm. The second matrix, A_{NS} contains the numbers assigned to the wave conditions in step 1 at the start of the algorithm. When a wave condition is selected by the algorithm its number will be deleted from matrix A_{NS} and added to the matrix A_S.

5. Define the first wave condition ($A_{S,1}$) as the most similar to the initial wave condition in the observation dataset (X_1):

$$A_{S,1} = \sum_{j=1}^{k} jI\left[||x_1, v_j|| = min(||x_1, V||)\right], j = 1,...,k \tag{2.36}$$

   The number assigned to the initial wave condition ($A_{S,1}$) will now be deleted from the matrix A_{NS} that thus reduces its size to $(k-1)x1$

6. The next wave condition to be selected for the reduced time-series is the one with the highest Markov transition probability, conditional on the previous selected wave condition and available in the matrix A_{NS}:

$$A_{S,t} = \sum_{q=1}^{k-t+1} A_{NS,q} I\left[M(A_{S,t-1}, A_{NS,q}) = max\left(M\left(A_{S,t-1}, A_{NS}\right)\right)\right] \tag{2.37}$$

7. Reorder the wave conditions in the database V according to their assigned numbers in matrix A_S.

2.5.2. MONTE CARLO MARKOV CHAIN SEQUENCING

The Monte Carlo Markov Chain sequencing (MCMC) orders the representative wave conditions randomly with weights in which the weights are the Markov transition probabilities. In other words, this sequencing method order the wave conditions in a random manner, but still pointing to the direction of the most likely wave case to occur based on the full dataset. The Monte Carlo Markov Chain sequencing was performed with 5 replicates. The procedure of this sequencing method is described as follows:

1. Follow steps 1 to 4 from section 2.5.1

2. Define the first wave condition ($A_{S,1}$) as the most similar to the initial wave condition in the observation dataset (X_1):

$$A_{S,1} = \sum_{j=1}^{k} jI\left[||x_1, v_j|| = min(||x_1, V||)\right], j = 1,...,k \tag{2.38}$$

   The Markov transition probability of the initial wave condition ($A_{S,1}$) will be reduced from the cumulative Markov transition matrix M and the remaining cumulative probabilities are normalized. Moreover, the number assigned to the initial wave condition ($A_{S,1}$) will now be deleted from the matrix A_{NS}, hence its size reduces to $(k-1)x1$;

3. Determine a random number between 0 and 1. The next wave condition to be selected is the one which the Markov transitions probability contains the random number previously defined:

$$A_{S,t} = \sum_{q=1}^{k-t+1} A_{NS,q} I\left[M(A_{S,t-1}, A_{NS,q}) \geq R_t\right], R_t = rand[0,1] \tag{2.39}$$

4. Subtract the Markov transition probability of the selected wave condition $A_{S,t}$ from the cumulative Markov transition matrix M and normalize the remaining probabilities:

$$M(m,n) = \frac{M(m, n \geq A_{S,t}) - M(m, A_{S,t})}{max\left(M(m,n)\right)}, m = 1,...,k, n = 1,...,k \tag{2.40}$$

5. Exclude the selected wave condition $A_{S,t}$ from the matrix A_{NS}

6. Reorder the wave conditions in the database V according to their assigned numbers in matrix A_S.

2.5.3. MONTE CARLO MARKOV CHAIN WITH REPETITION SEQUENCING

The Monte Carlo Markov Chain with repetition sequencing (MCMCR) has the same principle as the MCMC sequencing. It orders the representative wave conditions randomly with the Markov transition probabilities as weights. However, instead of excluding immediately the selected wave condition $A_{S,t}$, it allows the wave case to be repeated N_R times, where N_R is the number of repetition of the reduced wave climate (see section 2.3). Hence, in this sequencing method the reduced wave climate is not entirely repeated but instead the wave cases are allowed to persist N_R times. The Monte Carlo Markov Chain with repetition sequencing was performed with 5 replicates. The procedure of this sequencing method is described as follows:

1. Follow steps 1 to 3 from section 2.5.1
2. Define two time-series matrices: A_S with size $(kxN_R)x1$ and A_{NS} with size $kx1$. The first matrix, A_S starts empty and will contain the numbers that are assigned to the wave conditions in step 1 in the sequence determined by the algorithm. The second matrix, A_{NS} contains the numbers assigned to the wave conditions in step 1 at the start of the algorithm. When a wave condition is selected by the algorithm after N_R times its number will be deleted from matrix A_{NS} and added to the matrix A_S.
3. Define the first wave condition ($A_{S,1}$) as the most similar to the initial wave condition in the observation dataset (X_1):
$$A_{S,1} = \sum_{j=1}^{k} jI\left[||x_1, v_j|| = min(||x_1, V||)\right], j = 1, ..., k \tag{2.41}$$
4. Determine a random number between 0 and 1. The next wave condition to be selected is the one which the Markov transitions probability contains the random number previously defined:
$$A_{S,t} = \sum_{q=1}^{k-t+1} A_{NS,q} I\left[M(A_{S,t-1}, A_{NS,q}) \geq R_t\right], R_t = rand[0,1] \tag{2.42}$$
5. When the same wave condition is selected from A_{NS} N_R times, its Markov transition probability is reduced from the cumulative Markov transition matrix M and the remaining cumulative probabilities are normalized. Moreover, the number assigned to the wave condition ($A_{S,1}$) will now be deleted from the matrix A_{NS}:
$$M(m,n) = \frac{M(m, n \geq A_{S,t}) - M(m, A_{S,t})}{max\,(M(m,n))}, m = 1, ..., k, n = 1, ..., k \tag{2.43}$$
6. Reorder the wave conditions in the database V according to their assigned numbers in matrix A_S.

2.6. EVALUATION

The evaluation of the input reduction methods is performed by comparing the model predictions with reduced wave climate with the calibrated brute force model predictions for Noordwijk, Netherlands (Walstra et al. 2011). The performance of the reduced wave climate model predictions (z_{red}) is determined by a cumulative skill score R using the brute force predictions (z_{full}) as reference (Ruessink et al. 2007):

$$R(t) = 1 - \frac{\epsilon^2_{red}(t)}{\epsilon^2_{full}(t)} \tag{2.44}$$

where

$$\epsilon^2_{red}(t) = \sum_{x=x_1}^{x=x_{max}} \sum_{t=0}^{t=t} \left(z_{red}(x,t) - z_{full}(x,t)\right)^2 \tag{2.45}$$

$$\epsilon^2_{full}(t) = \sum_{x=x_1}^{x=x_{max}} \sum_{t=0}^{t=t} \left(z_{full}(x,t) - z_{full}(x,t=0)\right)^2 \tag{2.46}$$

Besides, the cumulative skill score, the instantaneous skill score BSS was also used as an auxiliary measure in the evaluation of the methods (Sutherland et al. 2004) :

$$BSS(t) = 1 - \frac{MSE_{red}}{MSE_{full}} \tag{2.47}$$

where

$$MSE_{red}(t) = \frac{1}{X} \sum_{x=x_1}^{x=x_{max}} \left(z_{red}(x,t) - z_{full}(x,t)\right)^2 \tag{2.48}$$

$$MSE_{full}(t) = \frac{1}{X} \sum_{x=x_1}^{x=x_{max}} \left(z_{full}(x,t) - z_{full}(x,t=0)\right)^2 \tag{2.49}$$

x = cross-shore distance and $x_1 - x_{end}$ is the range where the bars occur in the profile.
When $R = 1$ or $BSS = 1$ means a perfect match, no error, between the morphological prediction with the reduced wave climate and the brute force model. When R or BSS is smaller than one, indicates the difference between the performance of the models, thus, the error on the morphological prediction with the reduced wave conditions.

2.6.1. UNIBEST-TC

The Unibest-TC model is a 1D coupled model for cross-shore bed level evolution. It solves wave-averaged equations of hydrodynamics, sediment transport and bed level changes and computes the cross-shore distribution of hydrodynamics and sediment transport. The coupling between bed changes and hydrodynamic is made by the update of the changed bathymetry due divergence on sediment transport in the hydrodynamic model at the subsequent time-step. The equations of the model are described in (Delft Hydraulics 2000).

The Unibest-TC model consists of 5 sub-models:

1. Wave propagation model that computes the energy decay along the cross-shore including effects such as shoaling, refraction and energy dissipation;

2. Mean current profile model that computes the vertical distribution of the wave-averaged mean current considering wind shear stress, wave breaking, wave boundary layer bottom dissipation, and slope of free surface;

3. Wave orbital velocity model that calculates the near-bed wave orbital velocity accounting for wave asymmetry, wave group amplitude modulation and bound long waves;

4. Sediment transport model that consists of a suspended load transport module and the bed load transport module. The suspended load transport is dominated by the mean current and concentration profiles obtained from a time-averaged advection-diffusion equation. The bed load transport is dominated by the instantaneous bed shear stress which is computed from the near bed wave orbital and mean current velocities;

5. Bed level change model that computes the bed changes from the transport rates and the depth-integrated mass balance:
$$\frac{\partial(z)}{\partial(t)} + \frac{\partial(q_{bot+sus})}{\partial(x)} = 0 \tag{2.50}$$

2.6.2. STUDY AREA - NOORDWIJK, NETHERLANDS

The study area consists of Noordwijk, Netherlands (Figure 2.2). The area is characterized by a double sandbar system that propagates offshore on an yearly time scale (Walstra et al. 2013). In Noordwijk, the bar cycle (inception of bar in the swash zone followed by bar migration offshore until bar decay in the outer region of surf zone) lasts around 3-4 years. Furthermore, it seems that there is no relation between specific wave conditions and the bar cycle duration. Storms events cause an offshore migration, however the magnitude of the response is small in relation to the width of the bar system in the cross-shore direction. Thus, the bar dynamics in Noordwijk are classified as inter-annual net offshore migration (Ruessink et al. 2009). The tide at Noordwijk is semi-diurnal with a range of 1 m and 1.8 m during neap and spring tide, respectively. Storm surges eventually raised the water level by more than 1 m above the astronomical tide during the reduction period (Walstra et al. 2011). However, storm surge levels are not taken into account in the reduction models once the frequency of these events were considered not relevant for the overall morphological response of the system.

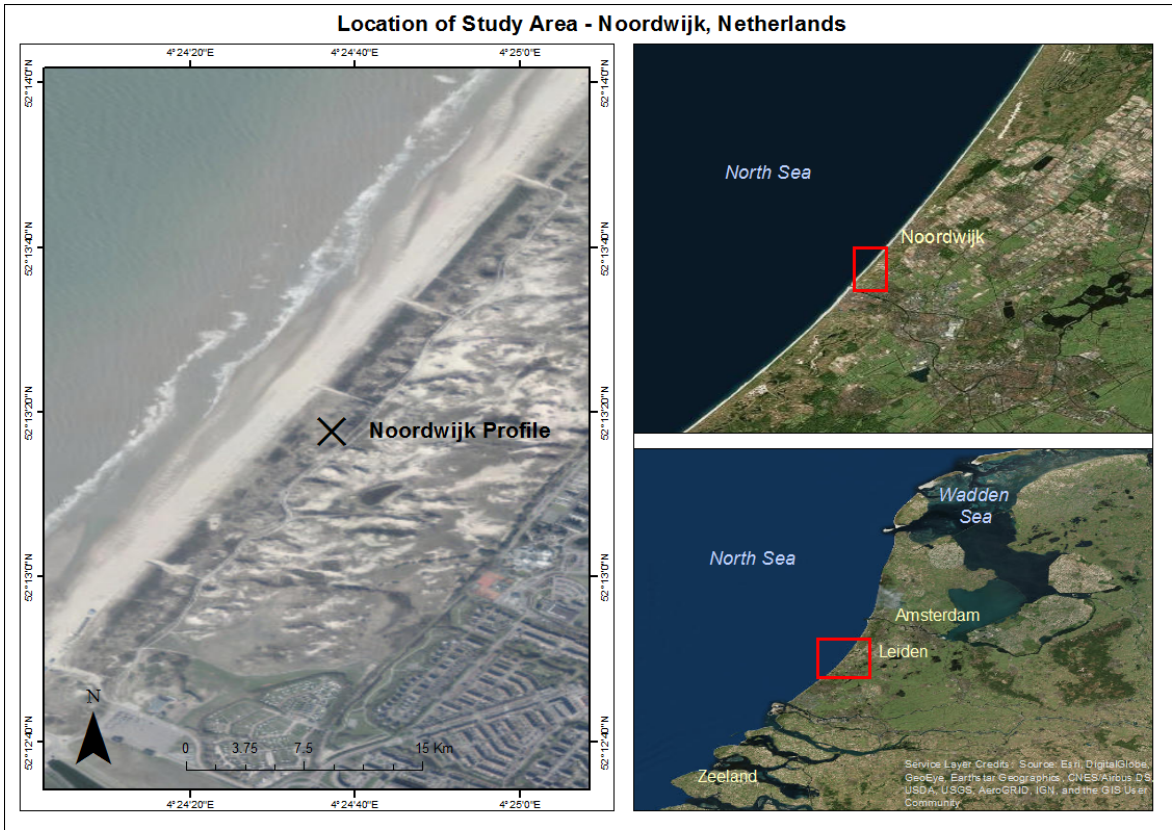


Figure 2.2: Location of study area, Noordwijk, Netherlands.

MODEL SETUP

The brute force model was used to simulate a complete bar cycle at Noordwijk from 1984 to 1987. The initial profile is based on a single transect measured in 1984. The computational grid has a resolution of 200 offshore and gradually decreases to 2 m across the active part of the profile (above 10 m water depth) in which the initial profile is interpolated. The model is forced with wave time-series measured 5 km offshore (around 18 m water depth) and water level time-series for the complete period with temporal resolution of 3 hours. The input reduction models utilized the same set-up of the brute force model except by the forcings, which were the reduced wave climate and astronomical tide predicted with only the M2 component.

2.7. VALIDATION

The validation of the findings on the Dutch coast is effectuated in Anmok beach, on the East coast of South Korea (Figure 2.3). The validation has as aim verify whether the best methodology of input reduction defined in the Dutch coast can also be applied in other locations with distinct characteristics and what are the implications of that. To do so, the best input reduction method, sequencing and number of cases defined in the study area are performed for a wave dataset measured at 128.95°E, 37.78°N (W1 point in Figure 2.3) for the period: 06/16/2015-06/18/2016. A Unibest-TC model is set-up with the resultant reduce wave climate and the results are compared with a full wave climate Unibest-TC model (brute force model).
Anmok is a sandy beach with crescentic bars in the nearshore. It is situated in a micro-tidal environment, dominated by waves coming mainly from the North-Northeast (Swinkels et al. 2016). The morphological evolution of Anmok is not yet fully understood which points the necessity of input reduction in this location.

MODEL SETUP

The brute force model of Anmok simulates a period of approximately one year, from 06/16/2015 to 06/18/2016. The initial profile is based on a survey campaign performed on 06/16/2015. The computational grid in which the initial profile is interpolated has a varying resolution of 100 offshore that gradually decreases to 10 m from 20 m water depth towards the shore, including the active part of the profile. The model is forced with

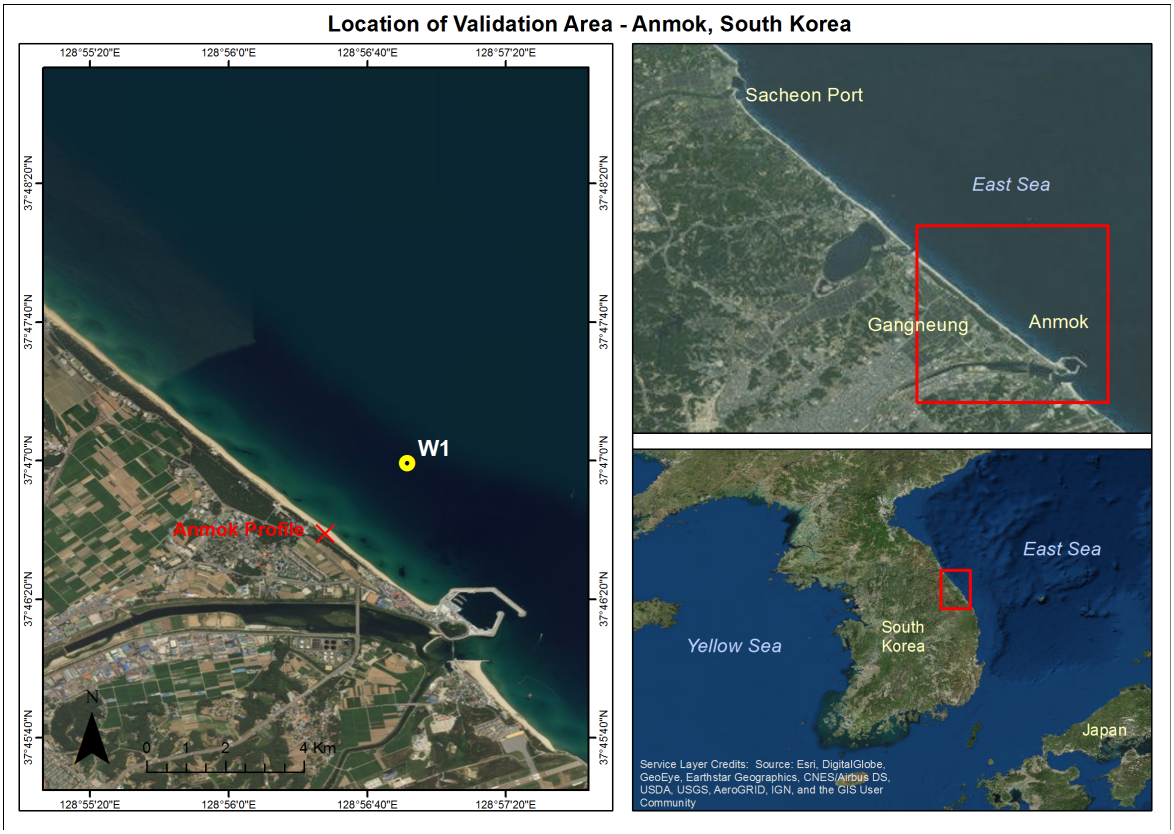


Figure 2.3: Location of validation area, Anmok, South Korea.

wave time-series measured 850 m offshore (around 20 m water depth) for the complete period with temporal resolution of 1 hour. The brute force model was roughly calibrated with a profile extracted from a survey campaign performed on 07/27/2016. More details on the calibration of Anmok's brute force model can be found in Appendix B. The input reduction model utilized the same set-up of the brute force model except by the wave climate that was reduced.

3

RESULTS

In this section the results of the brute force model are described, followed by the results and evaluation of the input reduction methods, sequencing, number of cases and validation.

3.1. BRUTE FORCE SIMULATION

Figure 3.1 and Figure 3.2 present the temporal evolution of the profile perturbations, the bars and troughs, and the initial and final profiles of the brute force simulation of Noordwijk. The profile perturbations are defined as the deviations from the time mean profile of the brute force simulation. It is possible to observe that in general, the behavior of the 2 bar system at Noordwijk is an offshore migration with alternating periods of growth and decay of the profile perturbations. Initially, the second bar (seawards) migrates offshore and decays until it is completely extinguished (around time = 500 days), while the second trough (seawards) continuously migrates offshore but with alternating periods of growth and decay. The first bar and trough (landwards) propagates offshore with periods of growth and decays.

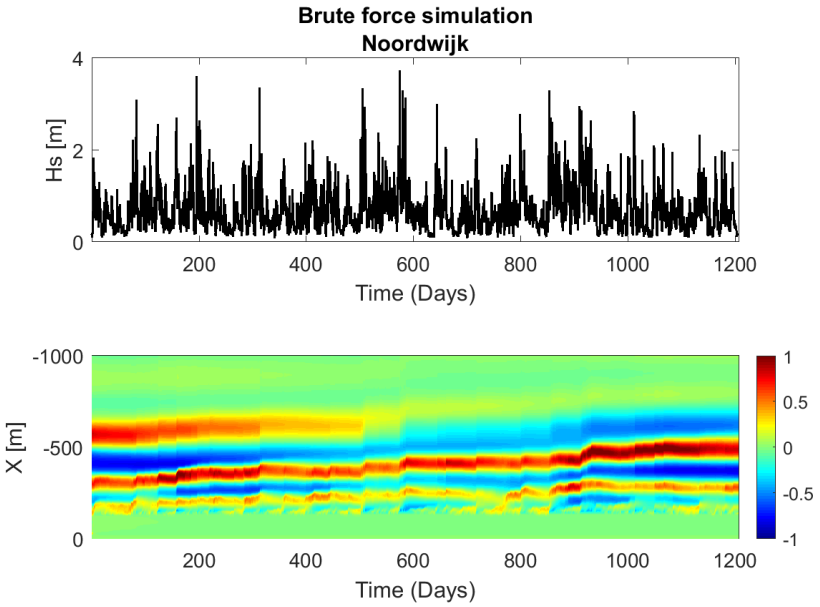


Figure 3.1: Time-series of offshore root-mean-square wave height H_{rms} (upper) and time stacks of profile perturbation (bottom) of the brute force simulation. Warm colors correspond to bars while cold colors correspond to troughs.

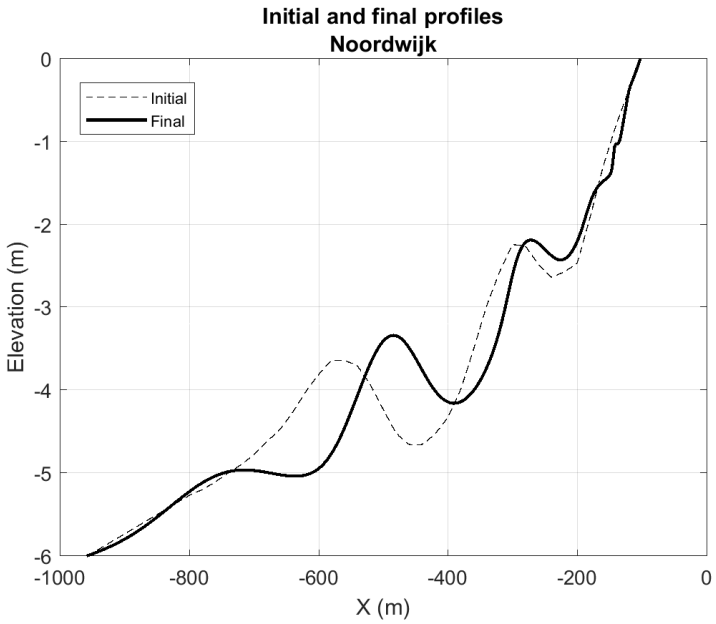
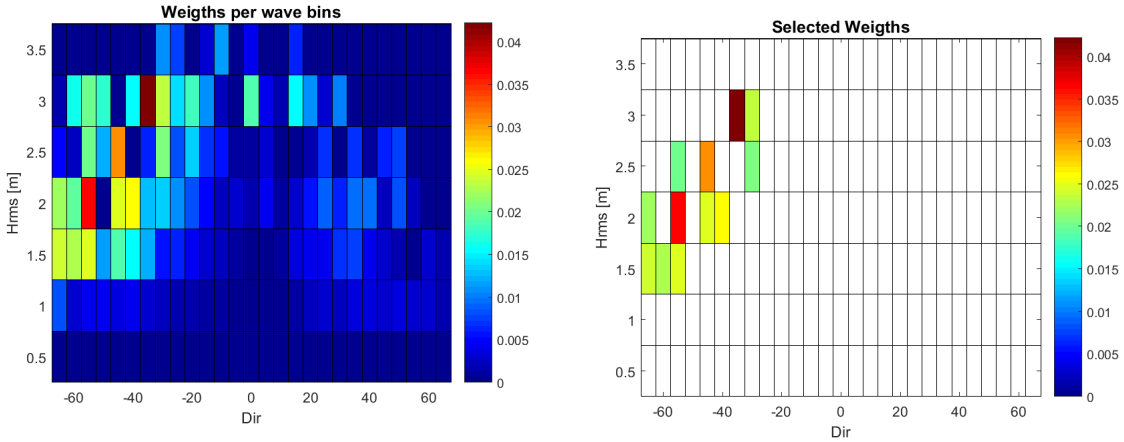


Figure 3.2: Initial (dashed line) and final (solid line) profiles of the brute force simulations. The second bar and trough are defined as the features most seawards while the first are the bar and trough most landwards.

3.2. METHOD 1: CONDITIONS WITH THE LARGEST TRANSPORT CONTRIBUTION

Method 1 initially, divides the observations into predefined wave height and direction bins, here defined as 28 directional bins and 8 wave height bins. Then, the contribution to the sediment transport of each bin (weight) is determined by summing the sediment transport of the wave conditions within the bins and dividing by the total sediment transport (Figure 3.3a). Finally, the method selects the highest weights as the representative wave cases (Figure 3.3b). This method selects mostly high wave height conditions predominantly from S and SW direction (negative direction values), overestimating the transport rates resulting in unrealistic results because it lacks mild conditions and directional variation. The morphological results of this method are unrealistic and are not presented in this report. Moreover, this method is not further investigated in this study.



(a) Sediment transport contribution (weights) of initial bins.

(b) Selected weights by method 1.

Figure 3.3: Selected wave cases - Method 1: Conditions with the Largest Transport Contribution.

3.3. METHOD 2: GROUPING WITH EQUAL SEDIMENT INFLUENCE

Method 2 initially selects the k most distant observations as initial centroids. Next, it groups the closest observations of each centroid until the cluster reaches a threshold of sediment transport. This threshold is defined as the total sediment transport of all observations divided by the amount of clusters (*e.g.* 12) . In this way, the method selects clusters of equal sediment contribution. The centroid of each cluster is defined as the average of the observations in the cluster and it is updated every time the cluster selects an observation. Figure 3.4 presents the selection of method 2. One drawback of this method is that it does not aggregate the observations into clusters very well, because once one cluster reaches the limit of sediment transport it closes and the observations closer to this cluster will be aggregated into another one that might be relatively far from the observations. This leads to a poor selection of representative wave cases once they are defined as the average of the observations in the cluster.

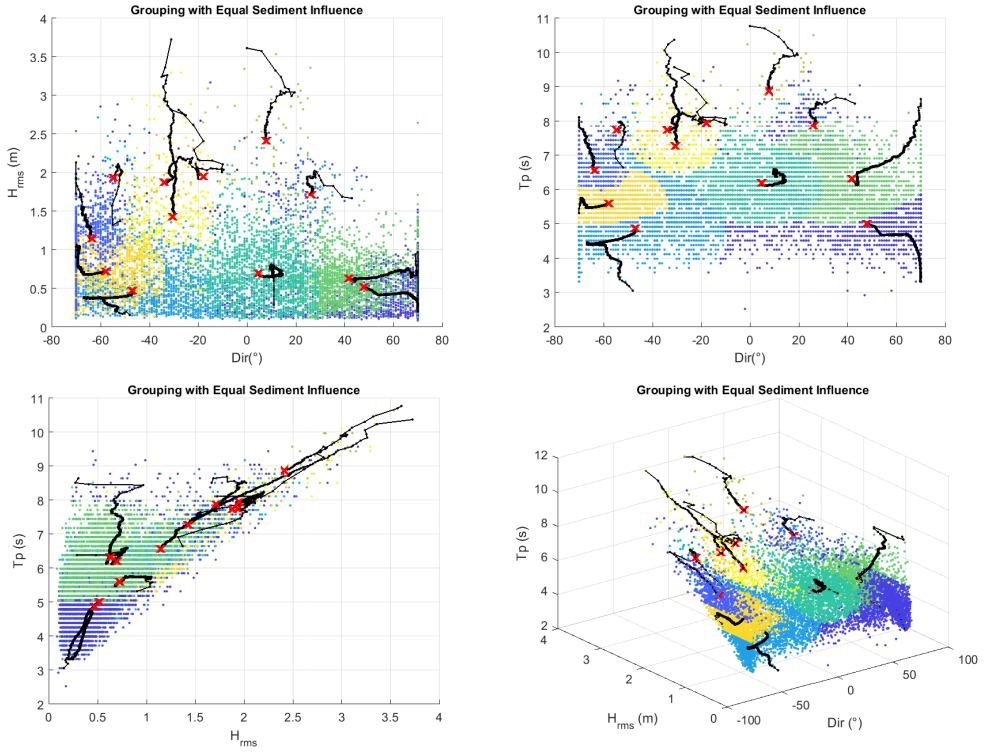


Figure 3.4: Selection of method 2: Grouping with Equal Sediment Influence. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

3.4. METHOD 3: THE REPRESENTATIVE WAVE APPROACH

Method 3 divides the observations into sections which are determined by seasons. For each season, a representative wave case is defined by the non-linear weighted function for wave height (see Eq. 2.8) and normal average for wave period and wave direction. This method presents the advantage of maintaining the chronology on the sequencing of the wave conditions, however it makes a poor selection of the representative wave conditions once it aggregates the observations through time. Even though usually in winter the frequency of extreme cases is higher, in summer and other seasons storms can occur as well. Besides, during winter mild wave conditions also occur, therefore the average conditions of the seasons end up being similar. Figure 3.5 presents the selection of method 3 and it shows that the representative wave conditions are very similar, especially regarding wave height and wave period. Additionally, even though the directions are not similar between the sections, the representative wave direction does not represent the range of wave directions of each section.

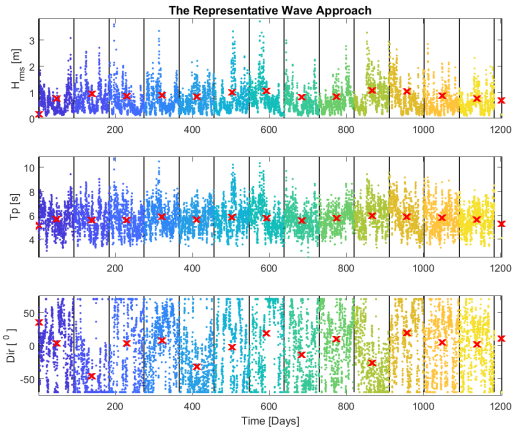


Figure 3.5: Selection of method 3: The Representative Wave Approach. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the sections.

3.5. METHOD 4: FIXED BINS

Method 4 is a method that requires some input from the user such as the amount of directional and wave height bins which results in a better guidance on the selection of the representative cases. Method 4 was performed with 3 different input variables. Table 3.1 lists the simulated methods. Figures 3.6 to 3.8 present the selection of wave cases from the different methods 4. In general, Method 4 presented a better spreading of the representative wave cases over direction and wave height, however M4B lacks on selection of mild conditions and M4C lacks on spreading over the wave height space.

Table 3.1: Simulated Methods 4 - Fixed Bins

Method	Input variables
M4A	H_{rms}, T_p, θ
M4B	$H_{rms}^{2.5}$, T_p, θ
M4C	S_y, T_p, θ

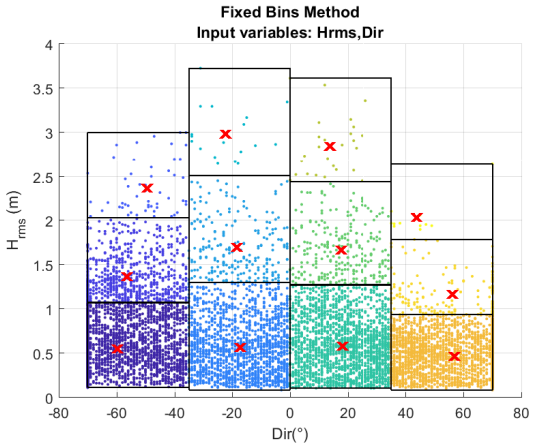


Figure 3.6: Selection of method 4: Fixed Bins - M4A. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the bins.

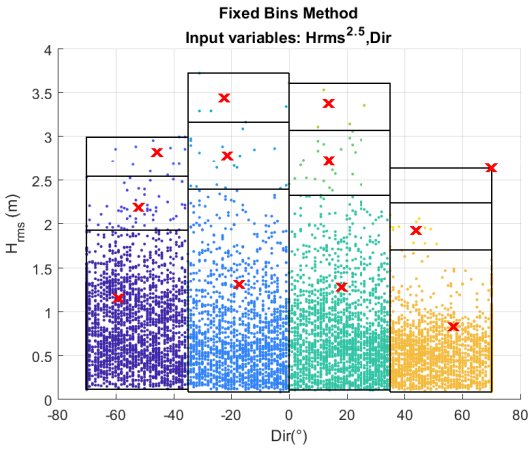


Figure 3.7: Selection of method 4: Fixed Bins- M4B. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the bins.

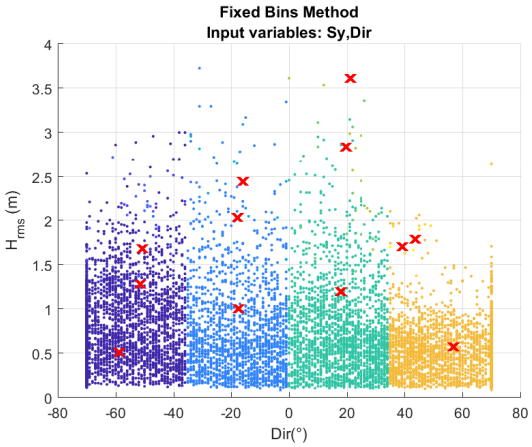


Figure 3.8: Selection of method 4: Fixed Bins- M4C. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the bins.

3.6. METHOD 5: CRISP K-MEANS

Method 5 selects the centroids based on the distance of the observations in relation to the previous centroids which can be initially defined as randomly with weights (K-means++), Fixed Bins or MDA. The method is based on an iterative process that leads towards a stable solution which would be equal centroids from the last iteration and the previous one. Besides the different cluster initiations, this method also allows for different input variables. Table 3.2 describes the variations simulated in method 5. Figures 3.9 to 3.17 present the selection of wave cases of the different methods 5.

This method has a high dependency on the frequency of occurrence of the observations. Low wave cases have higher occurrence than high wave cases in the observations and therefore have bigger influence on the definition of the centroids. However, by giving higher weights for the observations such as input variables $H_{rms}^{2.5}$, T_p, θ and S_y, T_p, θ results in centroids with higher wave height. Moreover, it is possible to notice that the cluster initiation influences the selection of the centroids, even though is difficult to state which cluster initiation is better just by looking at the selection once the centroids follow different paths depending on its cluster initiation.

Table 3.2: Simulated Methods 5 - Crisp k-means

Method	Cluster initiation	Input variables
M5A	K-means++	H_{rms}, T_p, θ
M5B	K-means++	$H_{rms}^{2.5}$, T_p, θ
M5C	K-means++	S_y, T_p, θ
M5D	MDA	H_{rms}, T_p, θ
M5E	MDA	$H_{rms}^{2.5}$, T_p, θ
M5F	MDA	S_y, T_p, θ
M5G	Fixed Bins	H_{rms}, T_p, θ
M5H	Fixed Bins	$H_{rms}^{2.5}$, T_p, θ
M5I	Fixed Bins	S_y, T_p, θ

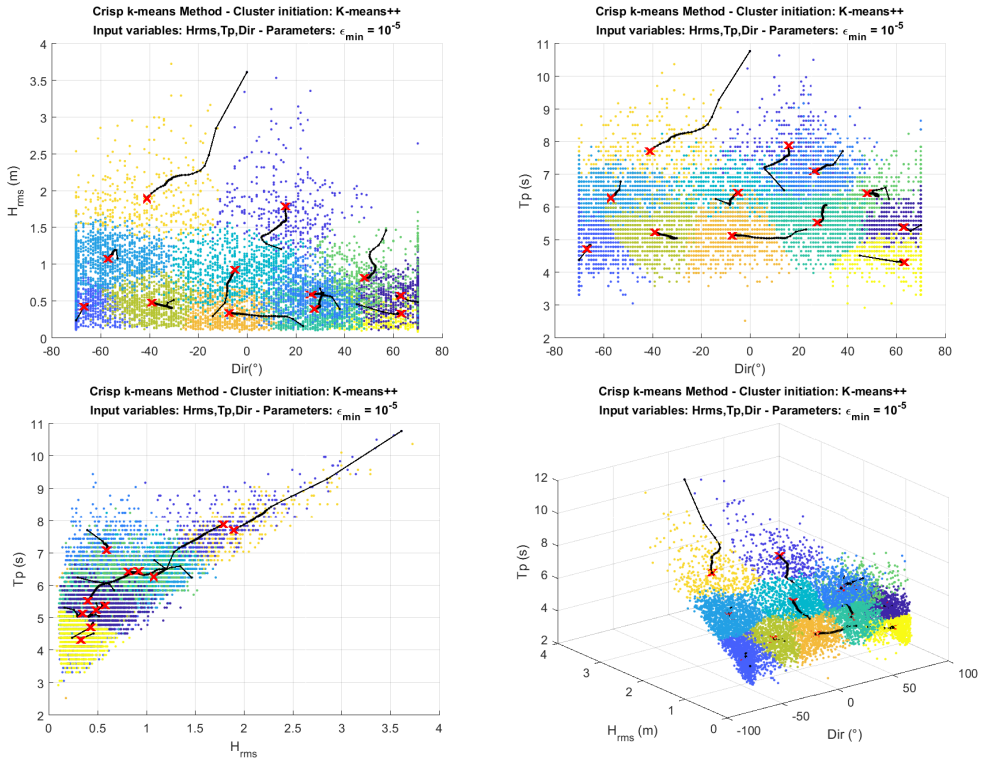


Figure 3.9: Selection of method 5: Crisp k-means - M5A. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

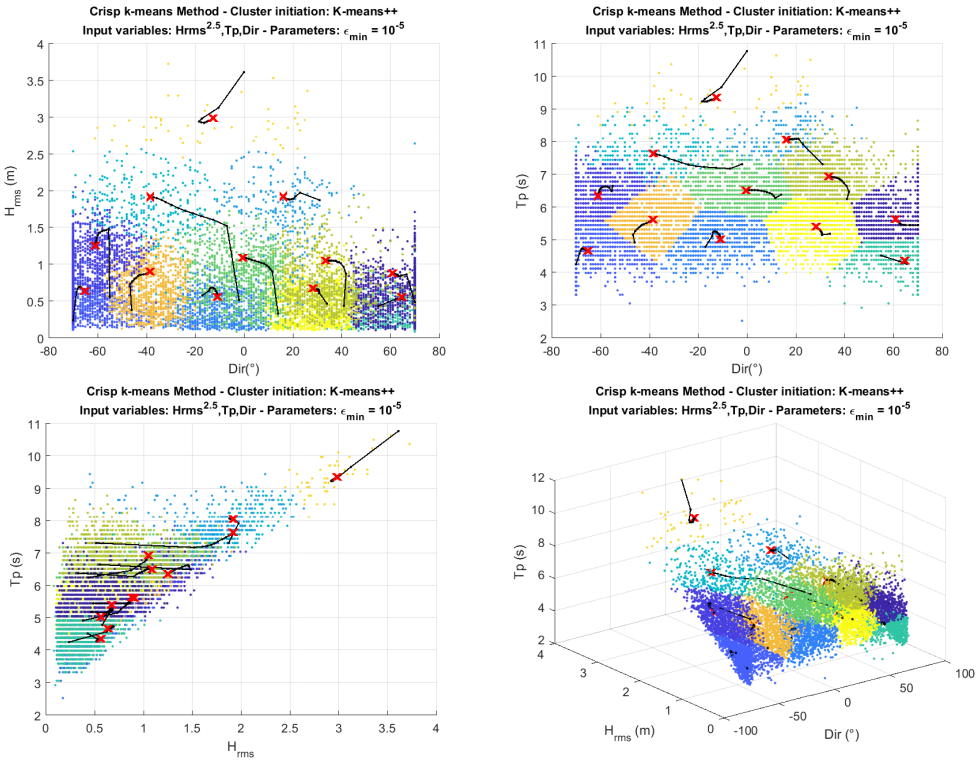


Figure 3.10: Selection of method 5: Crisp k-means - M5B. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

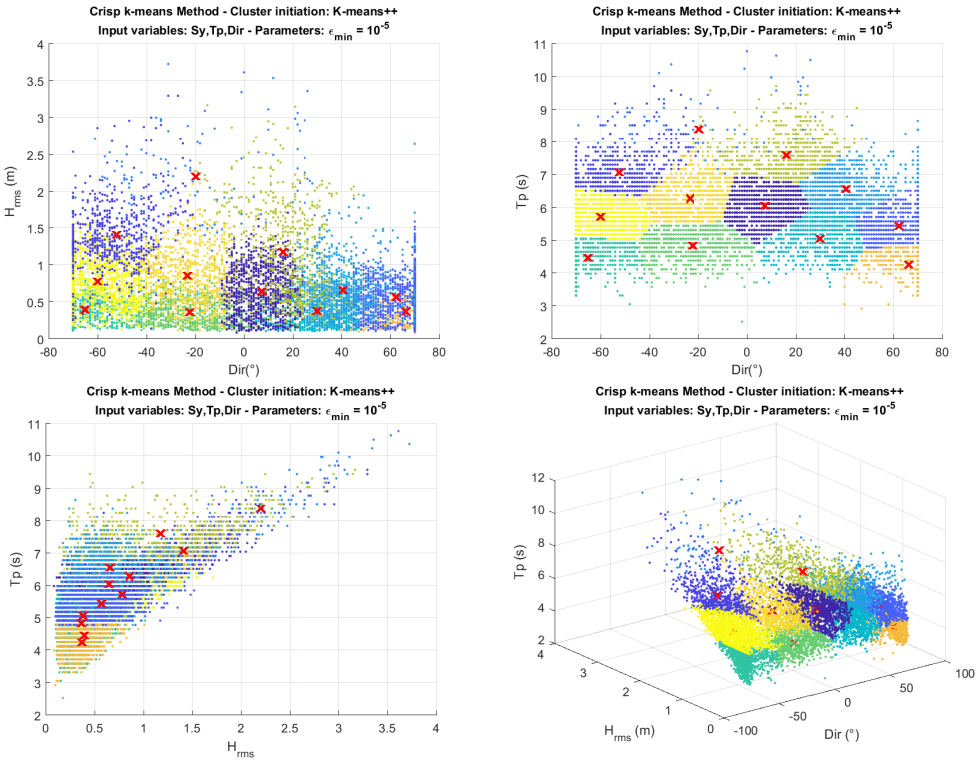


Figure 3.11: Selection of method 5: Crisp k-means - M5C. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

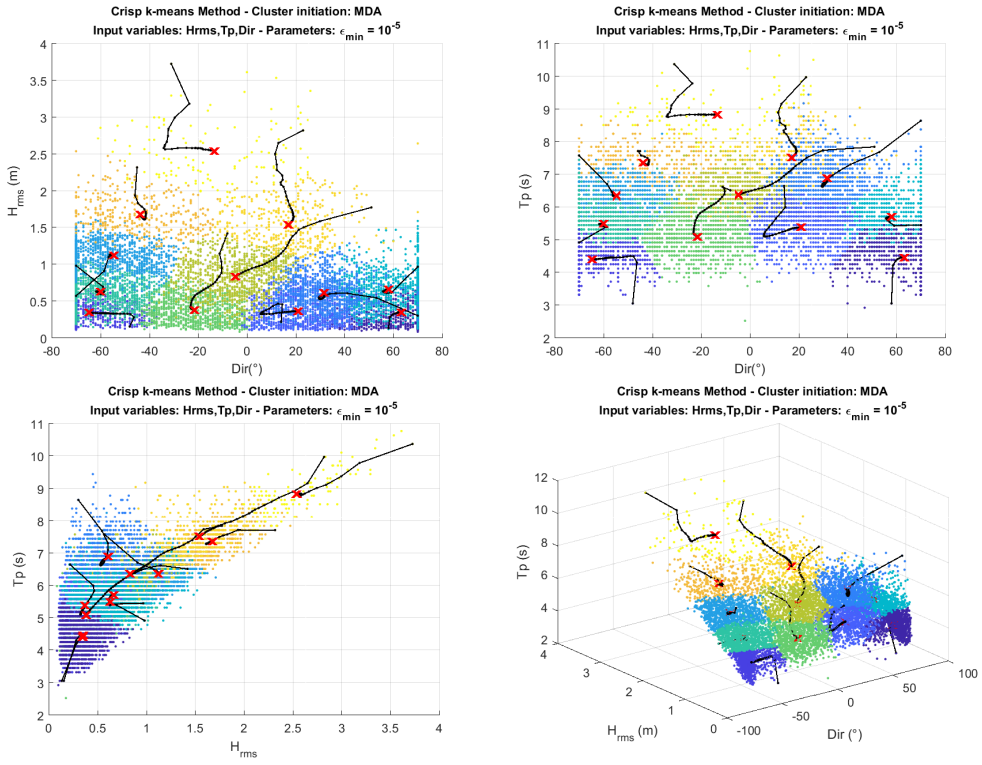


Figure 3.12: Selection of method 5: Crisp k-means - M5D. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

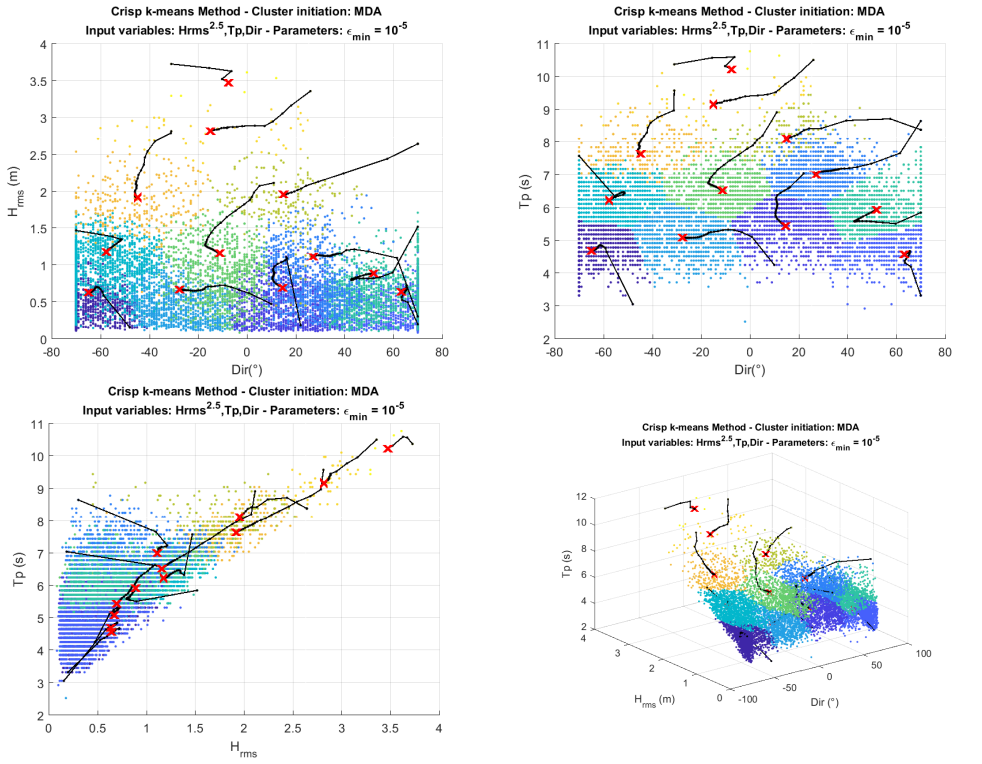


Figure 3.13: Selection of method 5: Crisp k-means - M5E. The red cross are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

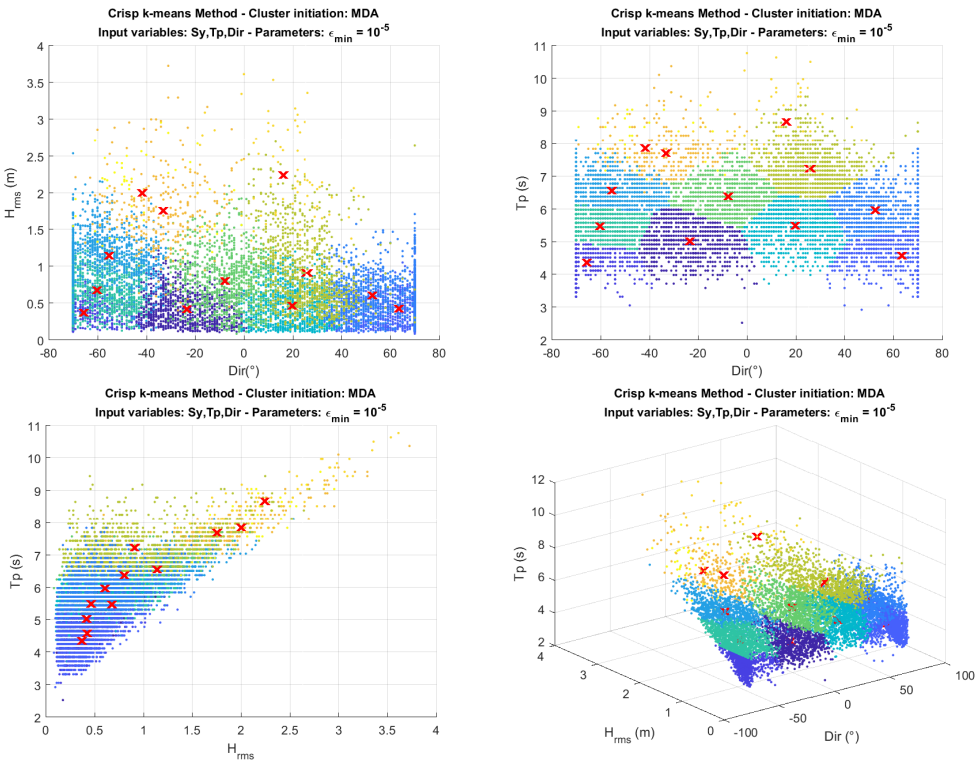


Figure 3.14: Selection of method 5: Crisp k-means - M5F. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

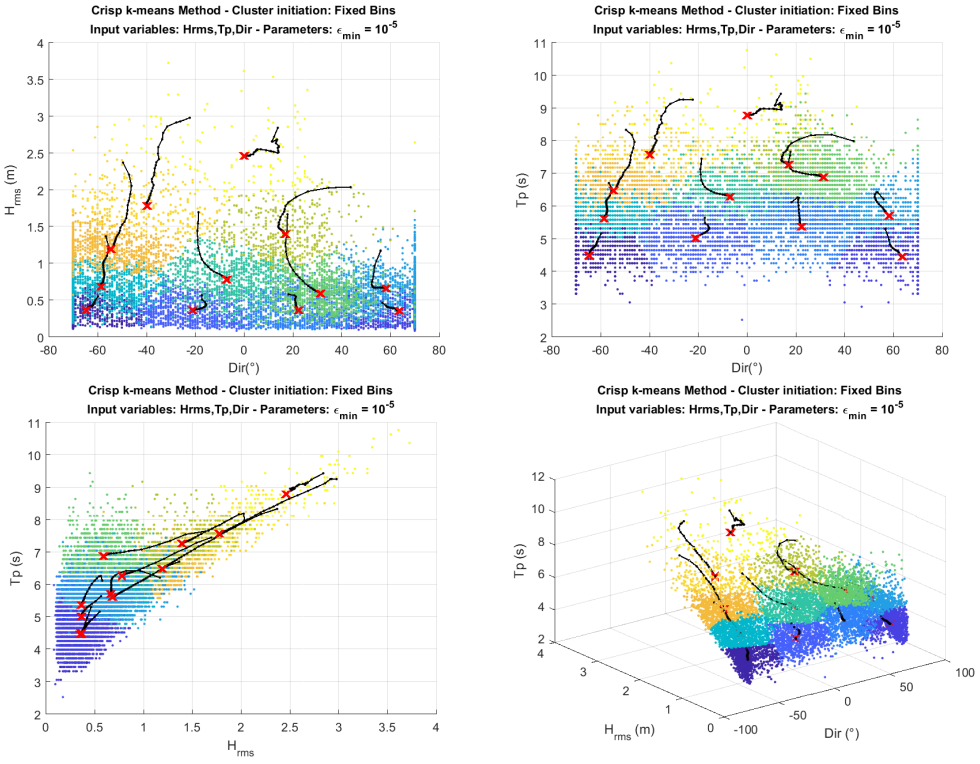


Figure 3.15: Selection of method 5: Crisp k-means - M5G. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

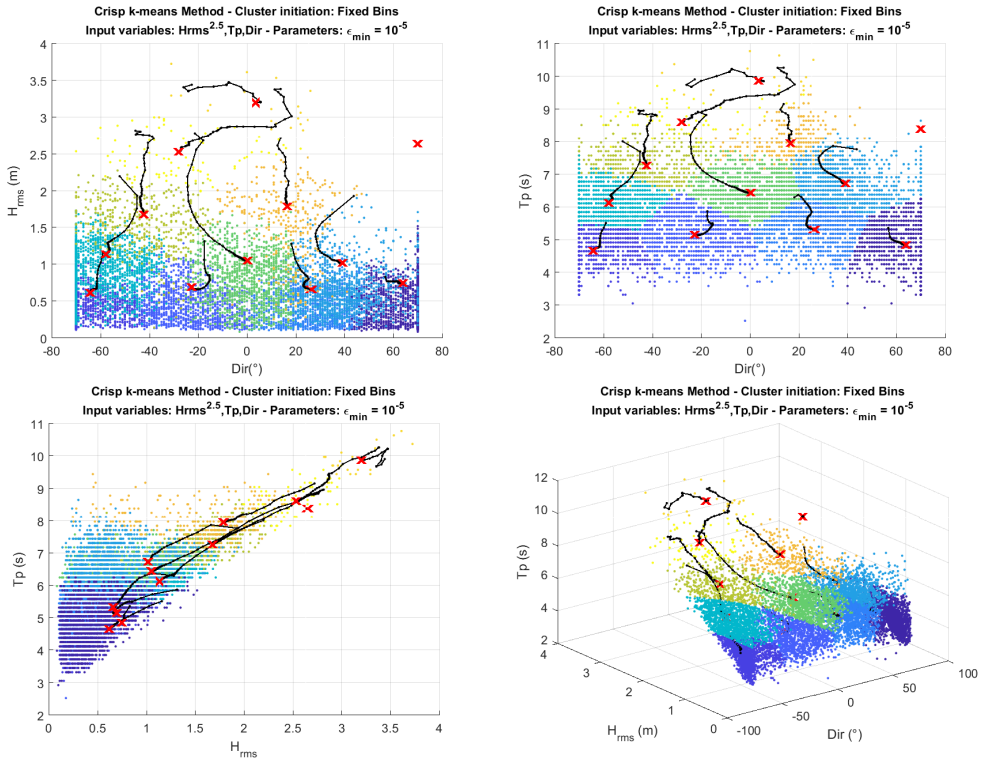


Figure 3.16: Selection of method 5: Crisp k-means - M5H. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

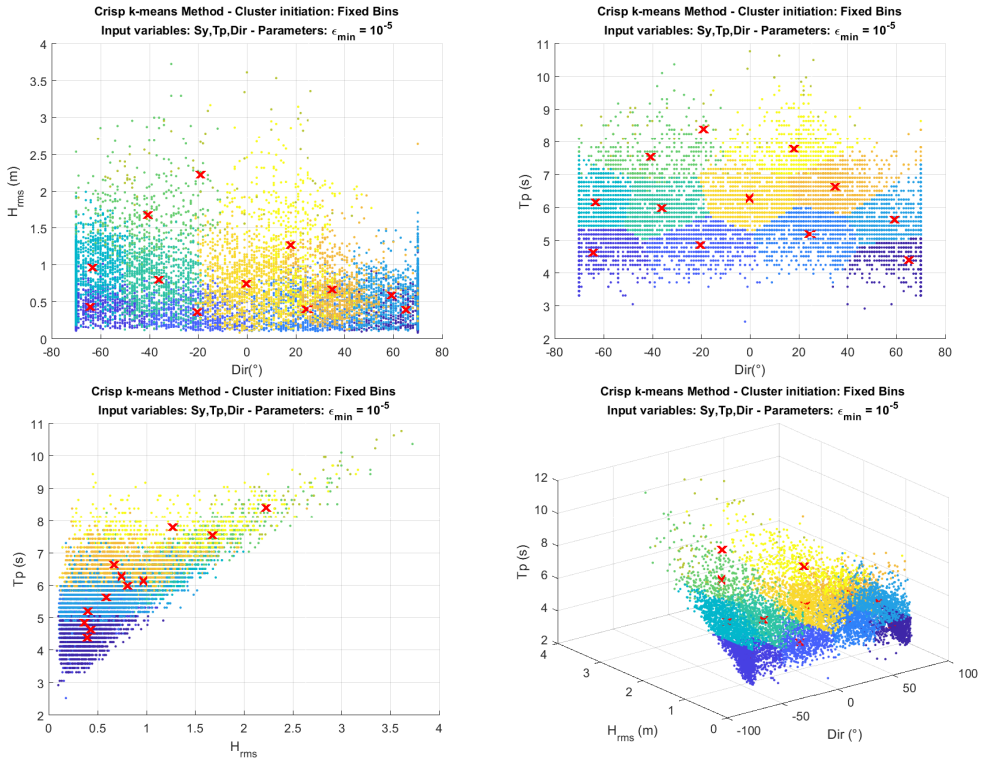


Figure 3.17: Selection of method 5: Crisp k-means - M5I. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

3.7. METHOD 6: FUZZY K-MEANS

Method 6 is a variant of method 5, it is also an iterative process where the centroids are defined based on the distance of the observations in relation to previous centroids. However, method 6 has a soft membership function which means that every observation has influence on the definition of the centroids. How much influence each observation will have is defined by the membership function which is a function that identifies how much each observation belongs to a cluster. The membership function gives higher weights for observations that are closer to the pre-defined centroids. In the end, the centroids are defined as the weighted average of all observations with the membership function as the weight.

The initial centroids can be defined randomly with weights (K-means++), Fixed Bins or MDA. Besides the different cluster initiations, this method also allows for different input variables. Table 3.3 describes the variations simulated in method 6. Figures 3.18 to 3.26 present the selection of wave cases from the different methods 6.

This method presents even a higher dependency on the low wave cases that have high frequency of occurrence. Even when using a weighting function such as input variables $H_{rms}^{2.5}$, T_p, θ and S_y, T_p, θ, the upper centroids are brought down (see Figures 3.19, 3.22 and 3.25). This occurs because even though lower observations have less weight on defining the upper centroids, their high frequency of occurrence overrules their low weight. Again, it is possible to notice that the cluster initiation influences the selection of the centroids, even though it is difficult to state which cluster initiation is better just by looking at the selection once the centroids follow different paths depending on its cluster initiation.

Table 3.3: Simulated Methods 6 - Fuzzy k-means

Method	**Cluster initiation**	**Input variables**
M6A	K-means++	H_{rms}, T_p, θ
M6B	K-means++	$H_{rms}^{2.5}$, T_p, θ
M6C	K-means++	S_y, T_p, θ
M6D	MDA	H_{rms}, T_p, θ
M6E	MDA	$H_{rms}^{2.5}$, T_p, θ
M6F	MDA	S_y, T_p, θ
M6G	Fixed Bins	H_{rms}, T_p, θ
M6H	Fixed Bins	$H_{rms}^{2.5}$, T_p, θ
M6I	Fixed Bins	S_y, T_p, θ

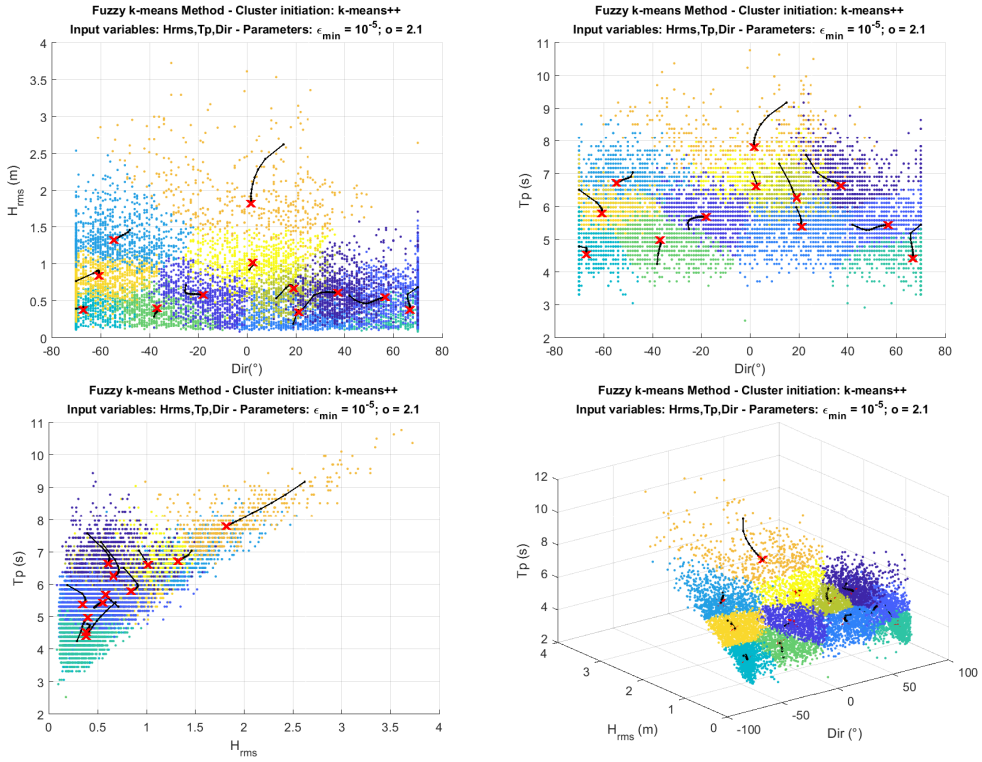


Figure 3.18: Selection of method 6: Fuzzy k-means - M6A. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

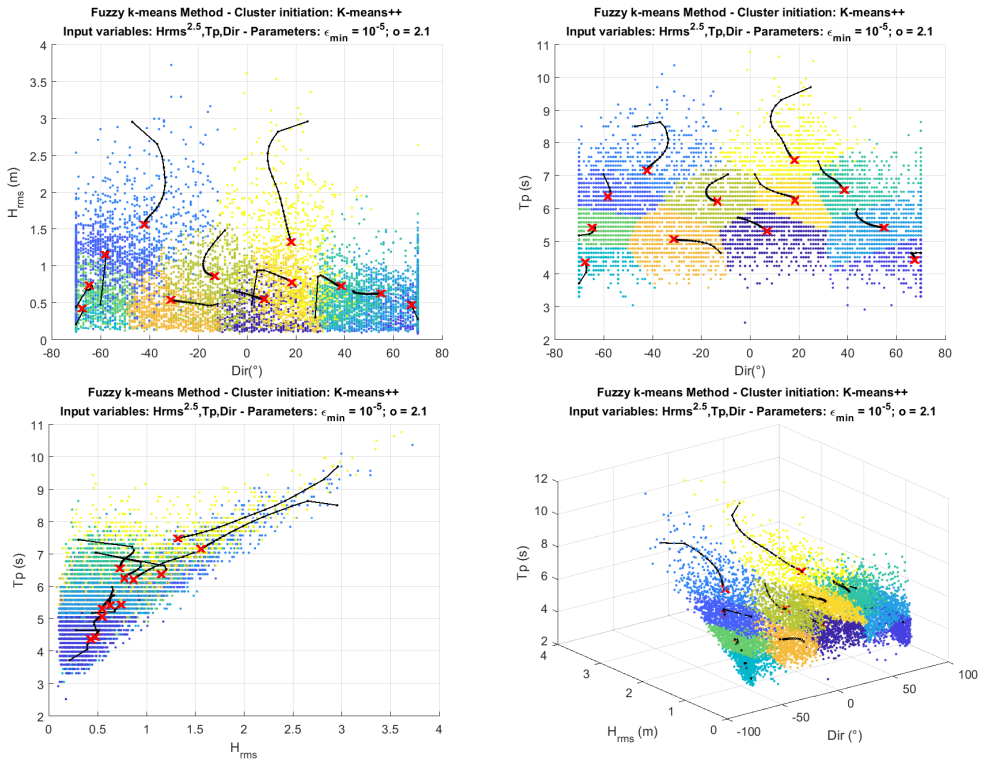


Figure 3.19: Selection of method 6: Fuzzy k-means - M6B. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

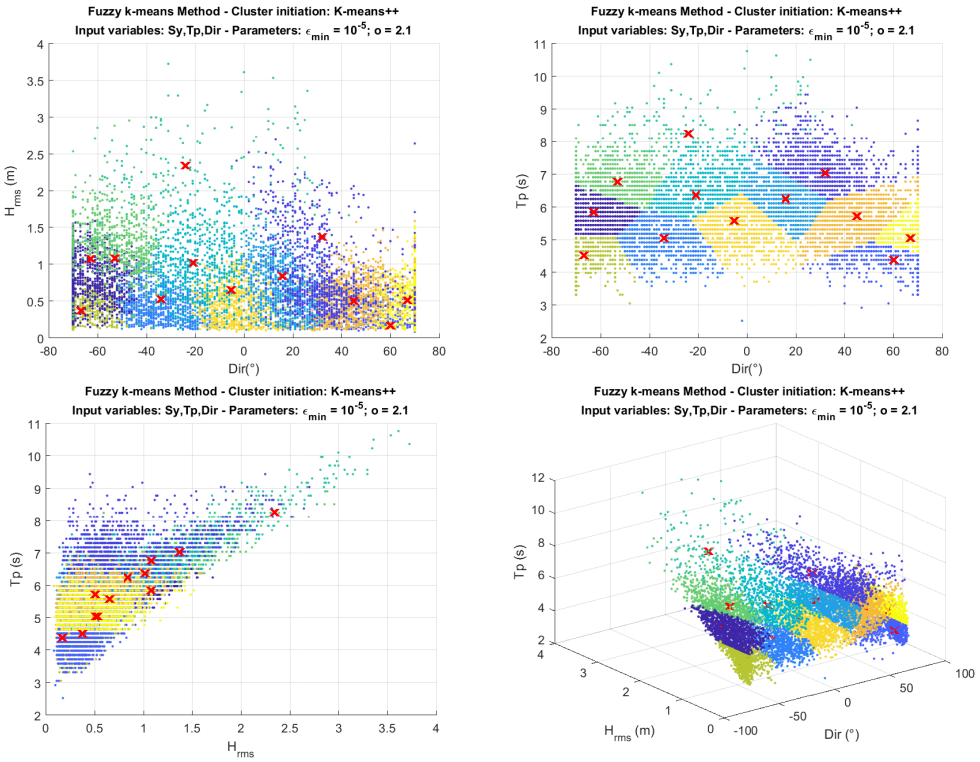


Figure 3.20: Selection of method 6: Fuzzy k-means - M6C. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

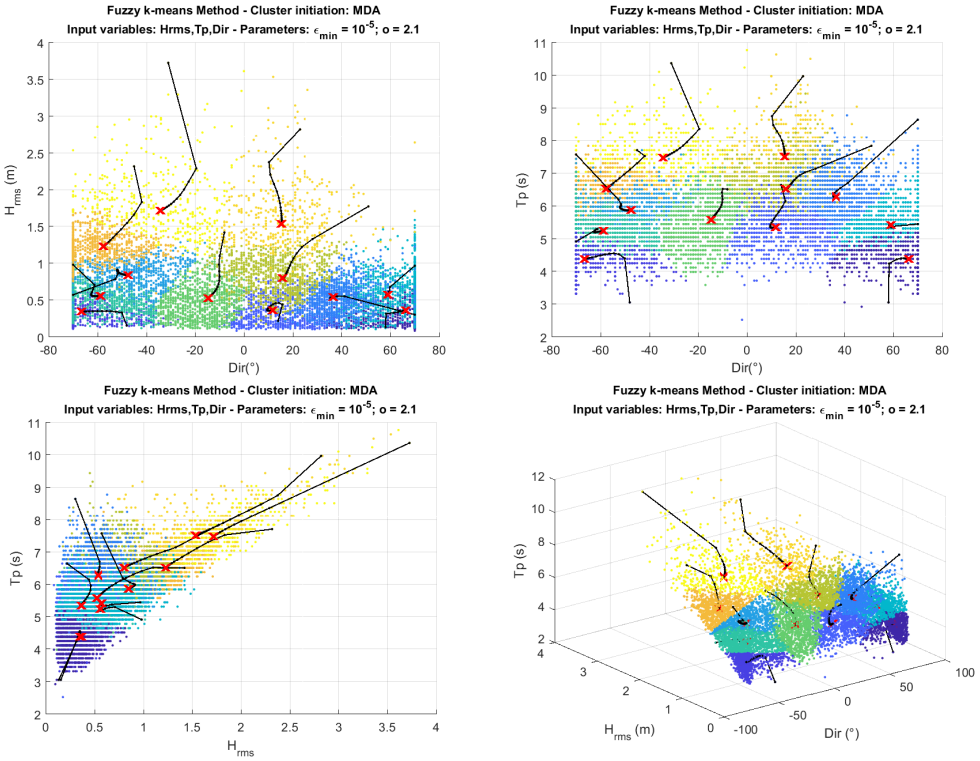


Figure 3.21: Selection of method 6: Fuzzy k-means - M6D. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

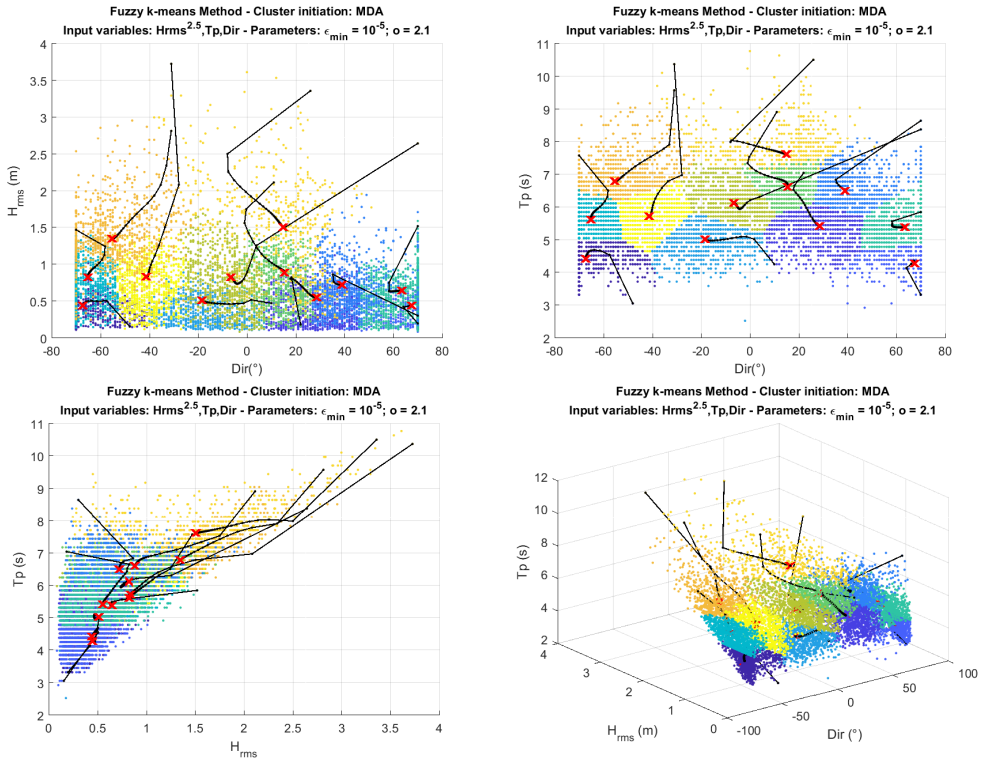


Figure 3.22: Selection of method 6: Fuzzy k-means - M6E. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

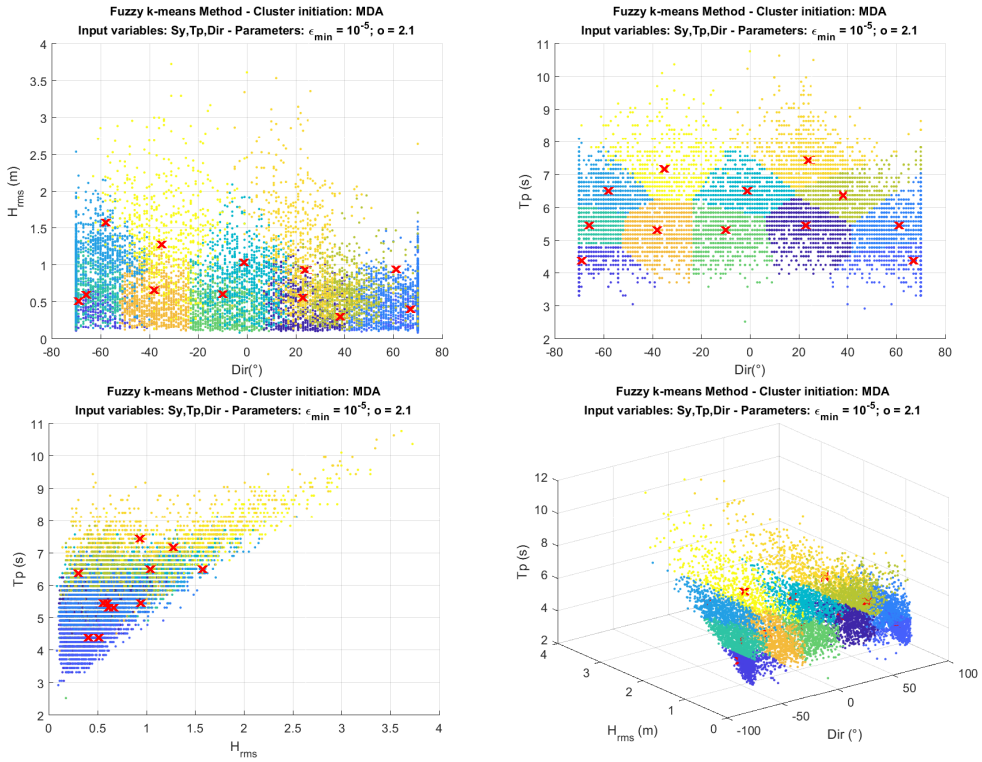


Figure 3.23: Selection of method 6: Fuzzy k-means - M6F. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

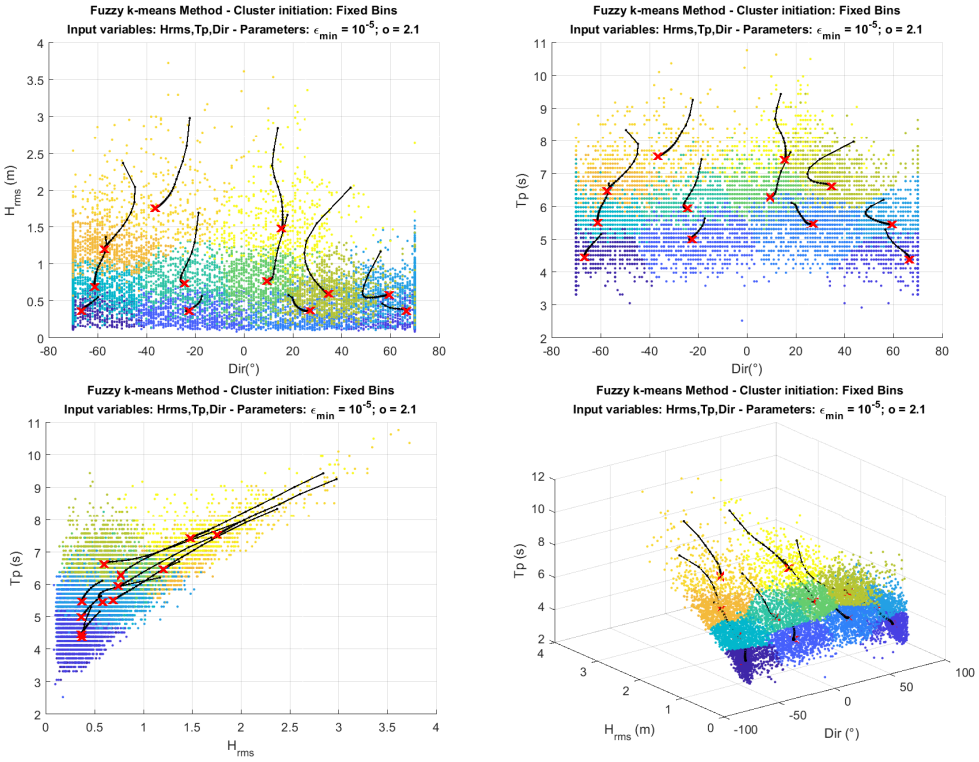


Figure 3.24: Selection of method 6: Fuzzy k-means - M6G. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

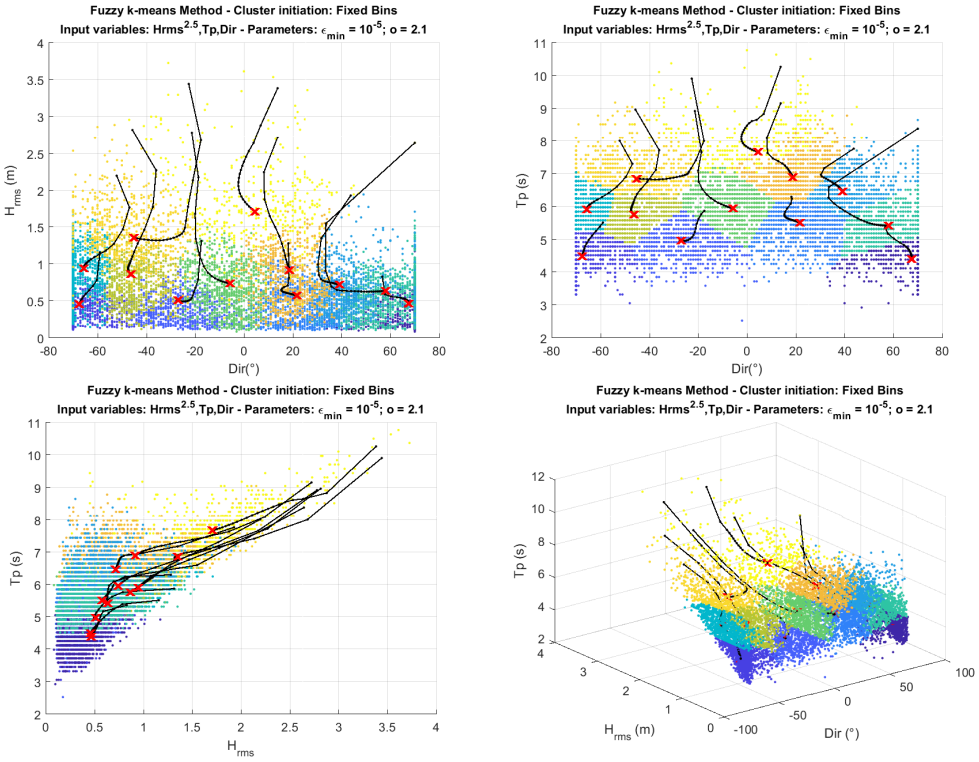


Figure 3.25: Selection of method 6: Fuzzy k-means - M6H. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

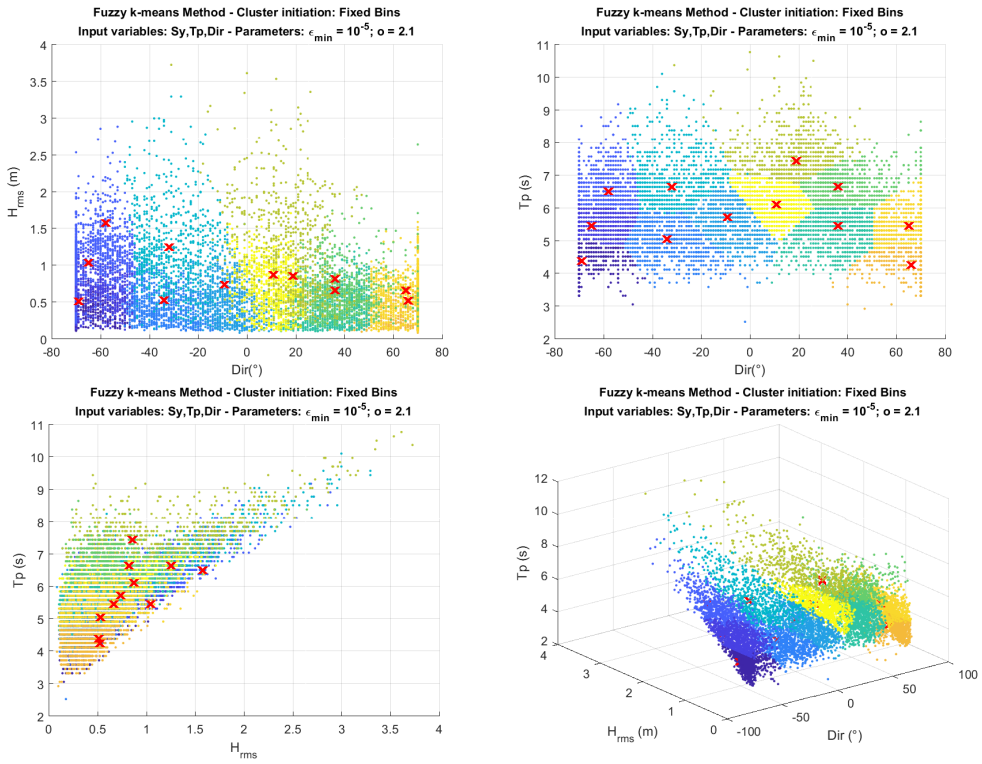


Figure 3.26: Selection of method 6: Fuzzy k-means - M6I. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

3.8. METHOD 7: K-HARMONIC MEANS

Method 7 is another variant of method 5 and it also accounts for the membership function as method 6 but in a different manner (see Eq. 2.23). Moreover, this method contains the dynamic weighting function which gives more weight to observations that are farther from the pre-defined centroids when defining the new centroids. As methods 5 and 6, method 7 allows for different cluster initiation as well as input variables. Table 3.4 lists the different variations simulated in method 7. Figures 3.27 to 3.35 present the selection of wave cases from the different methods 7.

The selection of method 7 also presents the same issue of dependency on the observation's frequency of occurrence as methods 5 and 6, although in a smaller scale due to its dynamic weighting function.

Table 3.4: Simulated Methods 7 - K-harmonic means

Method	Cluster initiation	Input variables
M7A	K-means++	H_{rms}, T_p, θ
M7B	K-means++	$H_{rms}^{2.5}$, T_p, θ
M7C	K-means++	S_y, T_p, θ
M7D	MDA	H_{rms}, T_p, θ
M7E	MDA	$H_{rms}^{2.5}$, T_p, θ
M7F	MDA	S_y, T_p, θ
M7G	Fixed Bins	H_{rms}, T_p, θ
M7H	Fixed Bins	$H_{rms}^{2.5}$, T_p, θ
M7I	Fixed Bins	S_y, T_p, θ

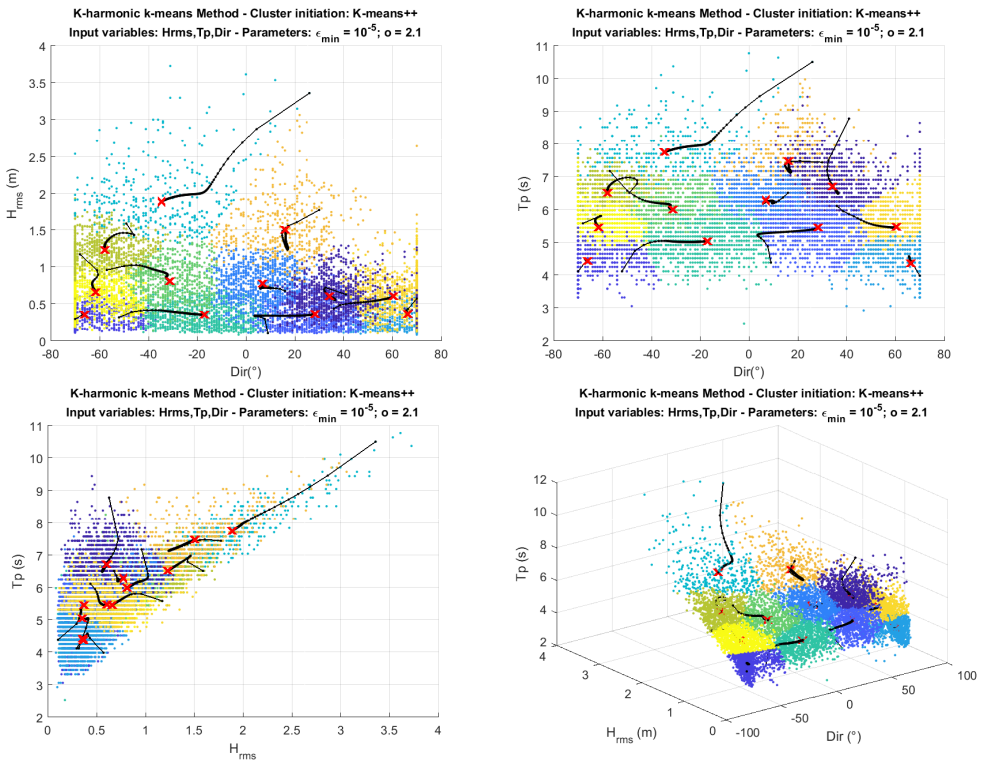


Figure 3.27: Selection of method 7: Fuzzy k-means - M7A. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

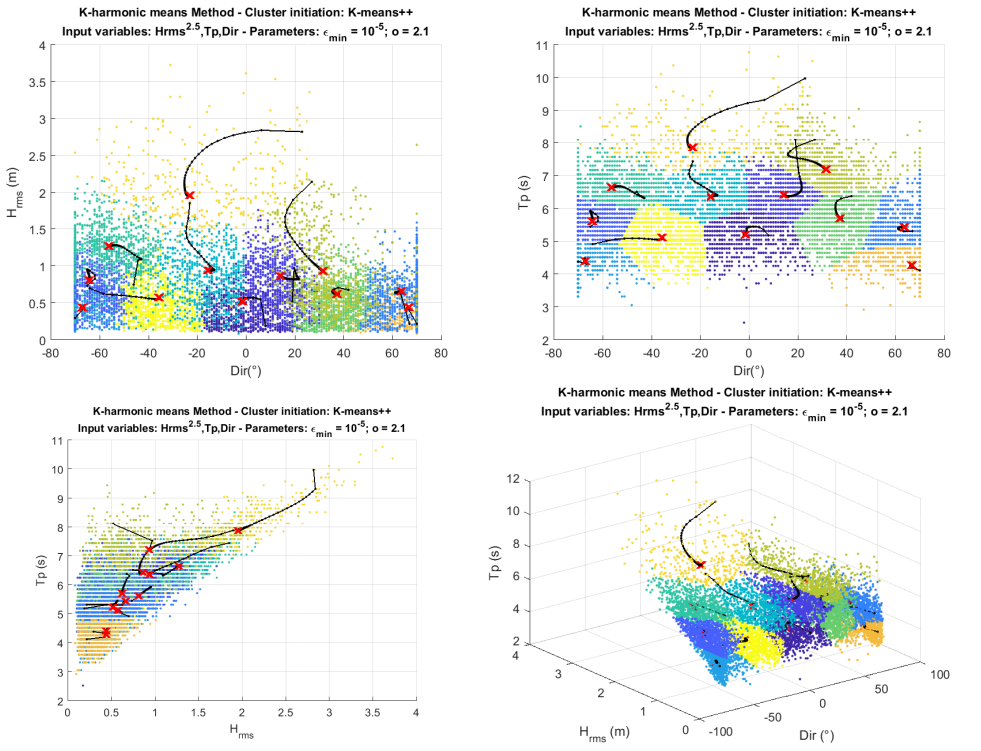


Figure 3.28: Selection of method 7: K-harmonic means - M7B. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

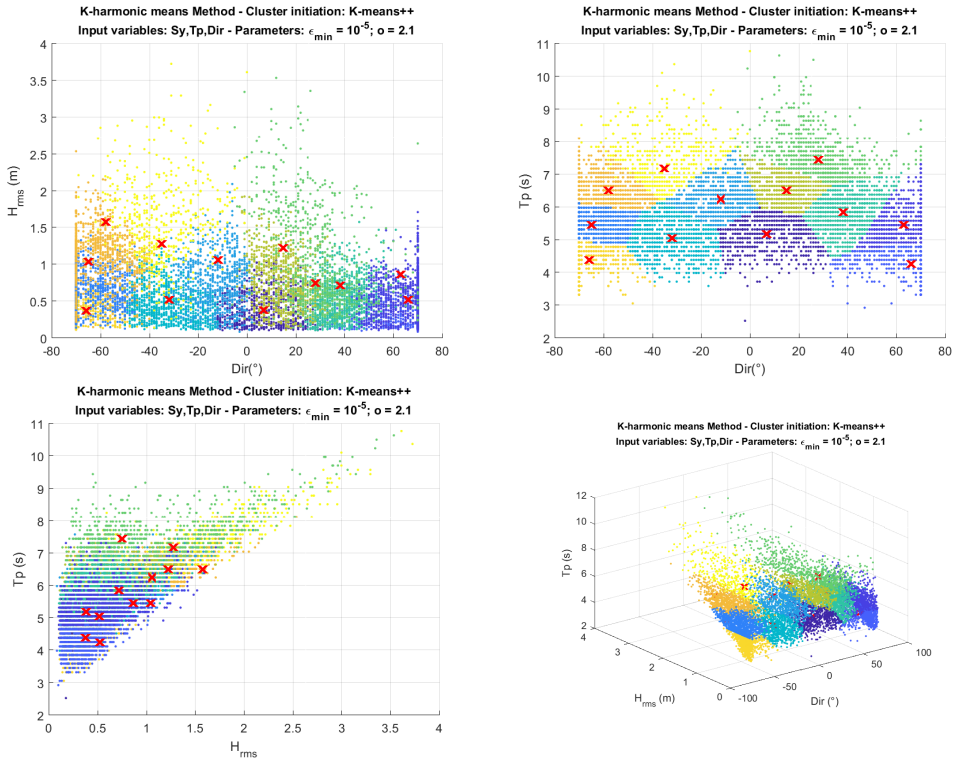


Figure 3.29: Selection of method 7: K-harmonic means - M7C. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

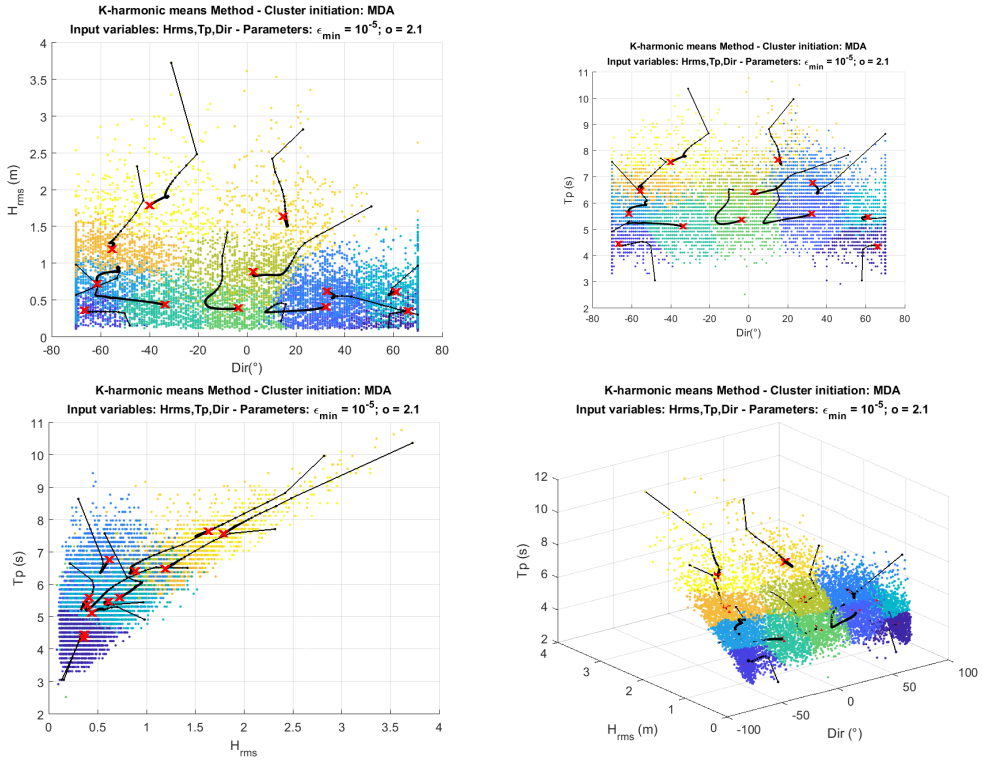


Figure 3.30: Selection of method 7: K-harmonic means - M7D. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

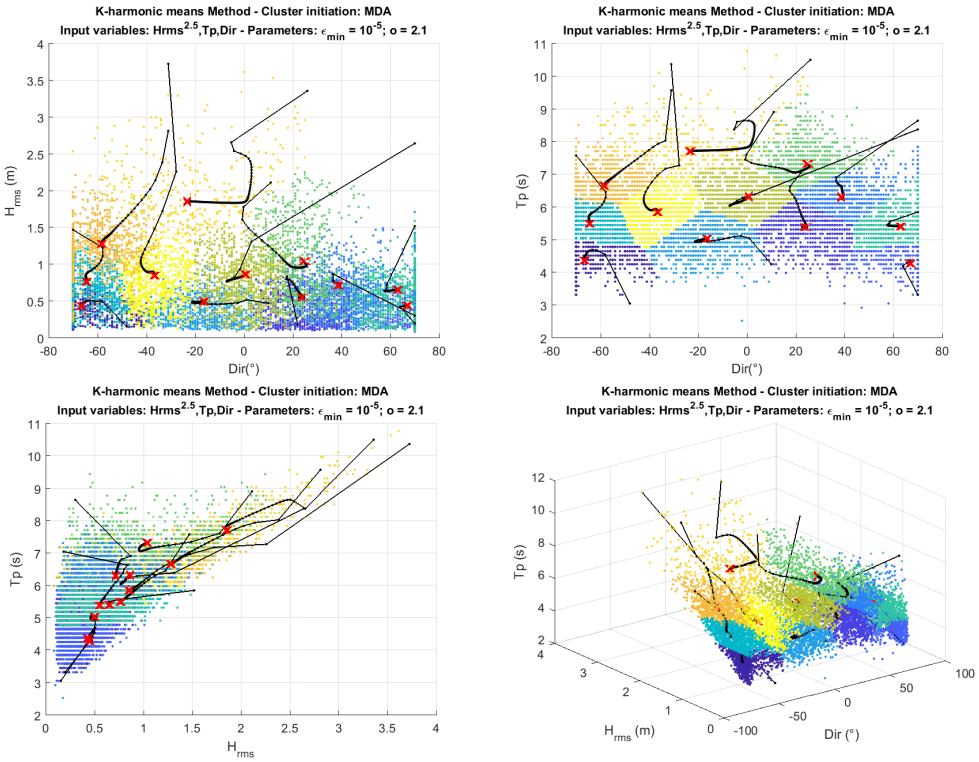


Figure 3.31: Selection of method 7: K-harmonic means - M7E. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

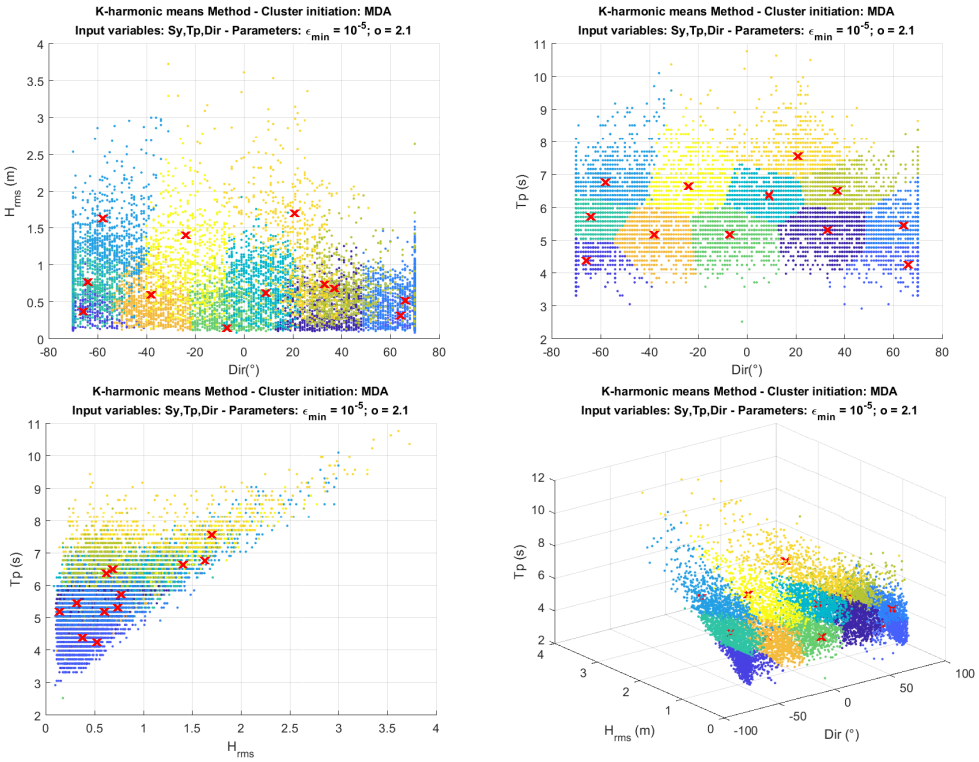


Figure 3.32: Selection of method 7: K-harmonic means - M7F. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

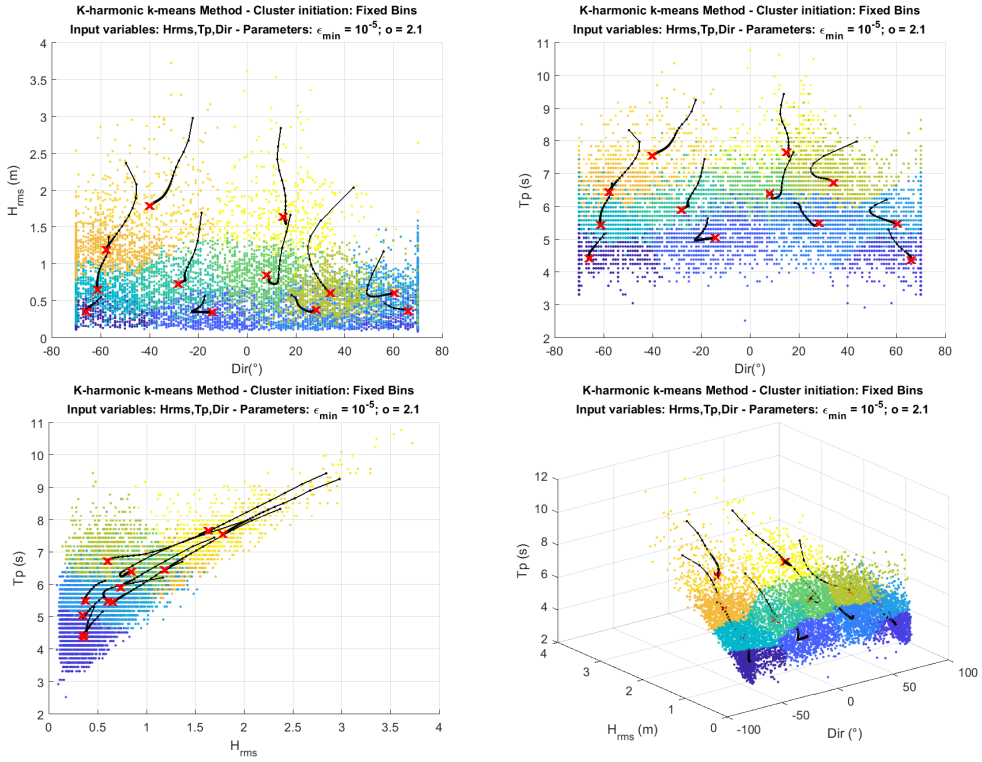


Figure 3.33: Selection of method 7: K-harmonic means - M7G. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

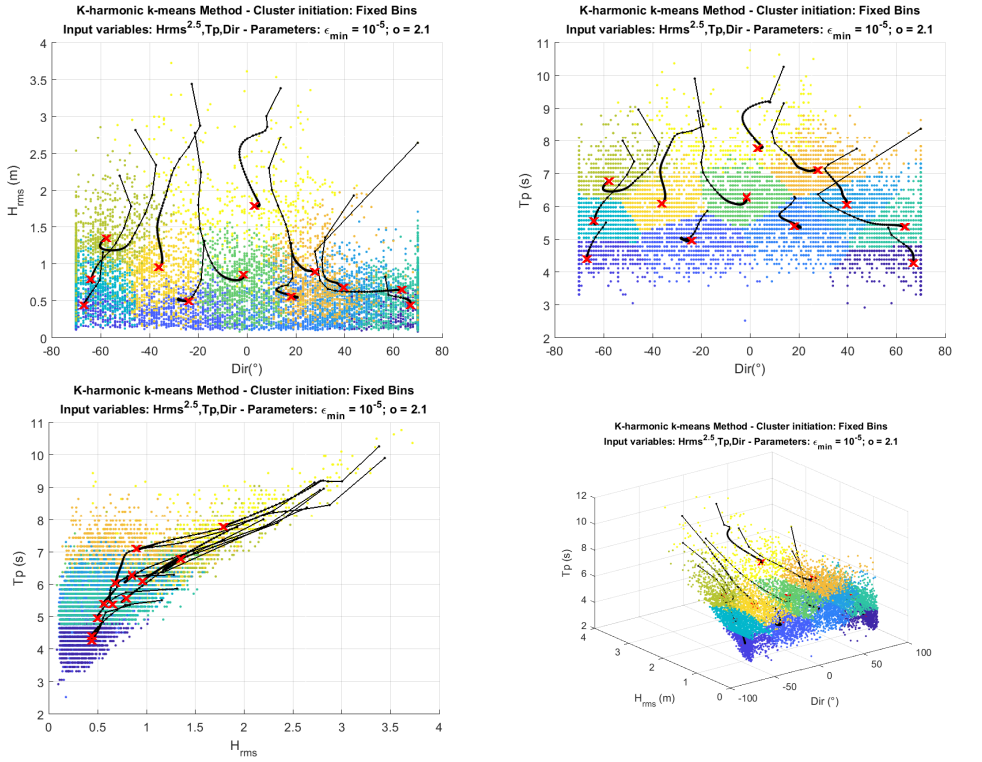


Figure 3.34: Selection of method 7: K-harmonic means - M7H. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

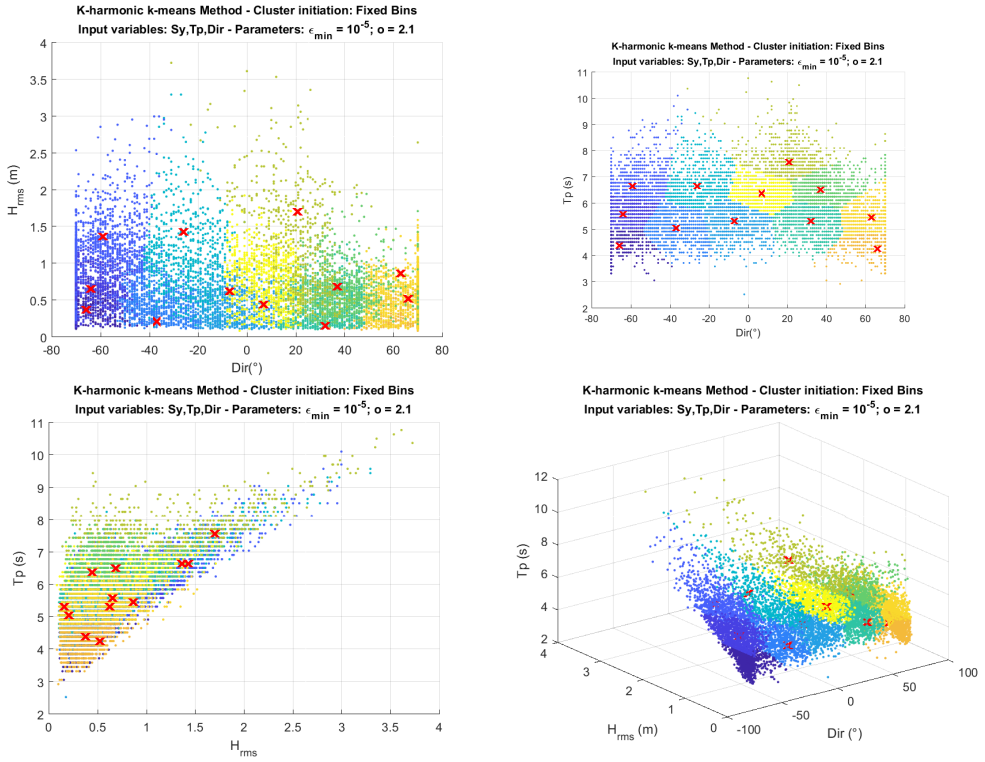


Figure 3.35: Selection of method 7: K-harmonic means - M7I. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

3.9. METHOD 8: MAXIMUM DISSIMILARITY – MAXMIN (MDA)

Method 8 selects as centroids the observations with maximum distance among themselves. Initially, the observation with maximum distance in relation to all other observations in the dataset is selected as the first centroid. Then, the first centroid is excluded from the observations. The second centroid is defined as the observation with the maximum distance to the first centroid. The third and next centroids are defined as the maximum distance among the minimum distance of the remaining observations to the previous centroids. Figure 3.36 presents the selection of wave cases by method 8. It is observed that method 8 only selects extreme conditions: either the highest wave cases or the smallest wave conditions. In order to select average wave conditions using this method a considerable increase of the number of centroids is needed. With k=12, method 8 is poor selecting representative cases. The morphological evolution obtained from method 8 is not realistic and is not presented in this report. Moreover, this method is not further investigated in this study.

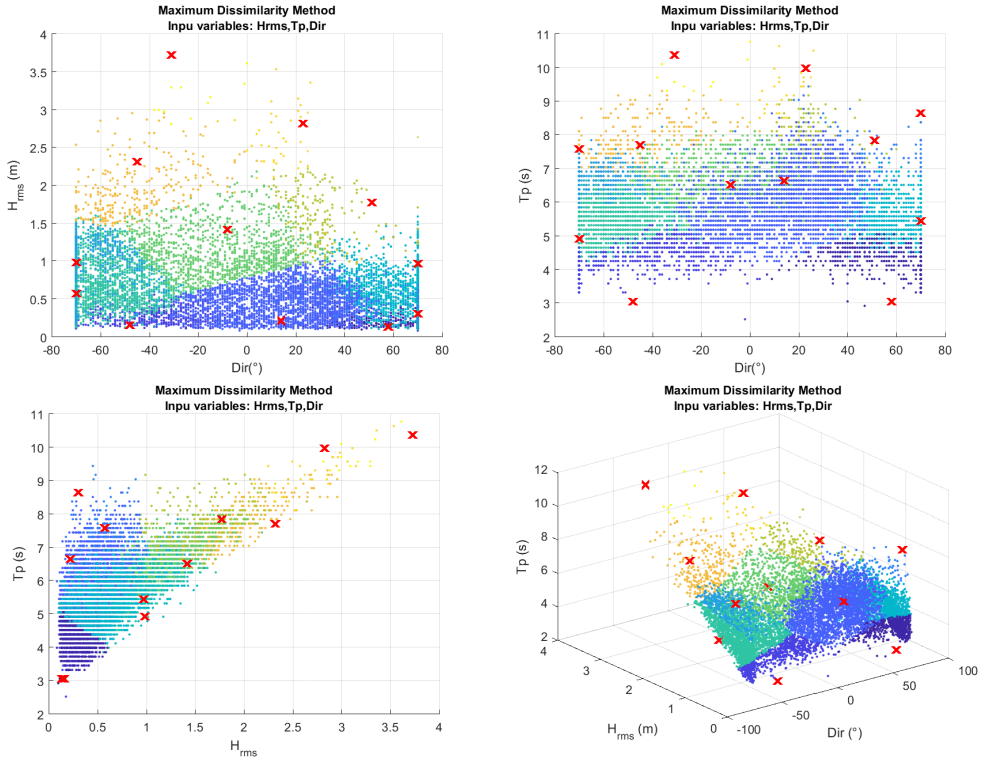


Figure 3.36: Selection of method 8: Maximum Dissimilarity. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

3.10. METHOD 9: ENERGY FLUX METHOD

Method 9 divides the observations in pre-defined directional and wave height bins with equal amount of energy flux and defines the centroids using the energy flux concept as well. The wave direction is defined as the mean of the energy flux direction of the bin, the period is the mean period of the bin and the wave height is obtained from the inverse of the equation of the energy flux within the bins. Figure 3.37 presents the selection resultant from method 9. This method selects representative wave conditions that better represent the overall observations distribution when compared to the previous methods described.

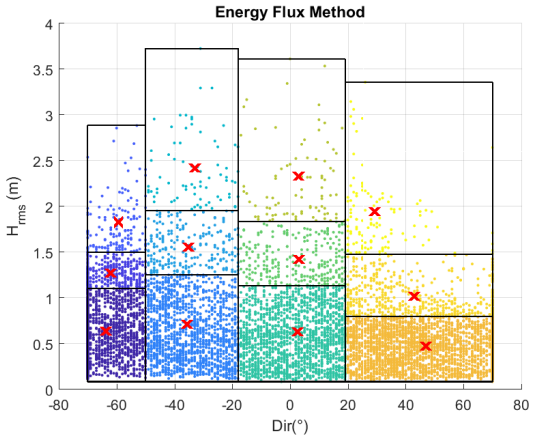


Figure 3.37: Selection of method 9: Energy Flux Method. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

3.11. METHOD 10: SEDIMENT TRANSPORT BINS METHOD

Method 10 has the same principle of method 9, but instead of dividing the observations in equal energy flux bins, it uses the sediment transport as weight to divide the observations into bins. The sediment transport used as input was obtained from the brute force simulations. The representative wave cases are defined as the average conditions of each bin, *e.g.* mean wave height, period and direction. Figure 3.38 presents the selection of wave cases of method 10. This method also selects representative wave conditions that better represent the overall observations distribution, (*i.e.* with higher spreading over direction and wave height spaces) when compared to the previous methods described.

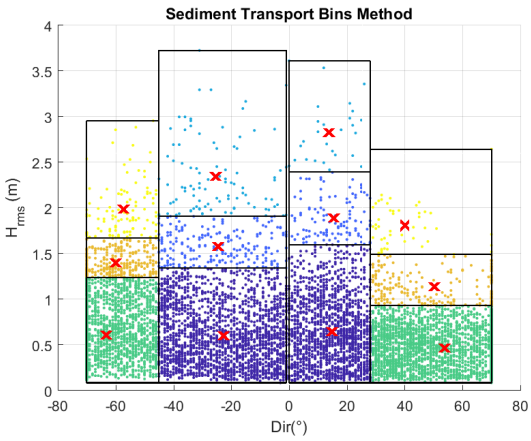


Figure 3.38: Selection of method 10: Sediment Transport Bins method. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

3.12. EVALUATION OF THE METHODS

Initially, the methods were evaluated with the cumulative skill score R based on the squared difference of the reduce wave climate profile and brute force profile relative to the squared difference of the brute force profile to its initial condition (see section 2.6). Table 3.5 lists the methods simulated. Figure 3.39 depicts the cumulative skill score of the methods at the end of the simulation of 5 random sequences and different wave climate durations simulated. It shows how the sequence of the wave conditions influences on the performance of the input reduction method. Figure 3.40 presents the mean skill score of the 5 random sequences for every wave climate duration simulated. Both figures demonstrate that besides the sequence, the duration of the wave climate also influences the performance of the methods. When the reduced wave climate is applied on the total time scale of the brute force model (3.3 years), the model presents poor skill scores once the duration of each wave condition is too long, especially for extreme conditions. Moreover, when decreasing the reduced wave climate time-scale, the performance of the model increases. This occurs due to the reduced duration as well as higher repetitions of individual wave conditions which resembles the brute force time-series better. However, since an optimal reduced wave climate should have a balance between wave climate duration and number of transitions, the best method(s) were selected based on $T_{wc} = 301$ days which means applying the reduced wave climate on an approximately yearly time-scale.

Among the Fixed Bins methods, M4A has relatively good results, but only when the reduced wave climate was repeated considerable ($T_{wc} = 134 - NR = 9$). M4B presents very poor performance once due to the weighting function it does not selects low wave height cases. M4C shows good performance even though its selection was not representative of the variability of wave cases for morphology. This method was further investigate in terms of final profiles and instantaneous *BSS*.

When looking at the Crisp k-means methods, the cluster initiation MDA presents slightly improved results compared to K-means++ or Fixed Bins. Again, when using the weighting function $H_{rms}^{2.5}$ the method performed poorly due to the absence of low wave height cases. Additionally, when using Sy as input variable the model performance improves when compared to using H_{rms} as input variable with $T_{wc} = 301$ days. This occurs due to sequencing effects, once when the duration of the wave climate is decreased there is little improvement in the skill score of the method 5 with S_y as an input variable.

For the Fuzzy k-means and K-harmonic means methods, the cluster initiation of Fixed Bins presented slightly

improved performance compared to K-means++ and MDA. These methods performed better with $H_{rms}^{2.5}$ as input variable because there is a balance between the weighting function that leads to a selection of centroids with higher wave height, and the high dependency on the dense cloud of observations that leads to a selection of centroids with low wave height. When using S_y as input variable, these methods had a lower skill score. This could be due to the absence of the inverse function for the sediment transport used in this study that resulted in a poor selection of wave cases.
Methods 9 and 10, presented promising results because predefining bins guarantees the desirable spreading over the direction and wave height for the selection of representative wave conditions. M10 performed better than M9 because its bins definition accounts for wave height and wave direction through the sediment transport weight while M9 only accounts for wave height through the energy flux weight.
Additionally, for methods that have a poor selection such as M4B, M5E, M5H, M6C, M6F, M6I, M7C, M7F and M7I, decreasing the duration of the reduced wave climate does not result in improvement of the skill scores whereas for methods that have a reasonable or good selection, decreasing the reduced wave climate duration results in an improvement of the input reduction performance.

Table 3.5: Simulated Methods

Name	Method	Cluster initiation	Input variables
M2	Grouping with Equal Sediment Contribution	-	S_y, T_p, θ
M3	The Representative wave approach	-	H_rms, T_p, θ
M4A	Fixed Bins	-	H_rms, T_p, θ
M4B	Fixed Bins	-	$H_{rms}^{2.5}$, T_p, θ
M4C	Fixed Bins	-	S_y, T_p, θ
M5A	Crisp k-means	K-means++	H_rms, T_p, θ
M5B	Crisp k-means	K-means++	$H_{rms}^{2.5}$, T_p, θ
M5C	Crisp k-means	K-means++	S_y, T_p, θ
M5D	Crisp k-means	MDA	H_rms, T_p, θ
M5E	Crisp k-means	MDA	$H_{rms}^{2.5}$, T_p, θ
M5F	Crisp k-means	MDA	S_y, T_p, θ
M5G	Crisp k-means	Fixed Bins	H_rms, T_p, θ
M5H	Crisp k-means	Fixed Bins	$H_{rms}^{2.5}$, T_p, θ
M5I	Crisp k-means	Fixed Bins	S_y, T_p, θ
M6A	Fuzzy k-means	K-means++	H_rms, T_p, θ
M6B	Fuzzy k-means	K-means++	$H_{rms}^{2.5}$, T_p, θ
M6C	Fuzzy k-means	K-means++	S_y, T_p, θ
M6D	Fuzzy k-means	MDA	H_rms, T_p, θ
M6E	Fuzzy k-means	MDA	$H_{rms}^{2.5}$, T_p, θ
M6F	Fuzzy k-means	MDA	S_y, T_p, θ
M6G	Fuzzy k-means	Fixed Bins	H_rms, T_p, θ
M6H	Fuzzy k-means	Fixed Bins	$H_{rms}^{2.5}$, T_p, θ
M6I	Fuzzy k-means	Fixed Bins	S_y, T_p, θ
M7A	K-harmonic means	K-means++	H_rms, T_p, θ
M7B	K-harmonic means	K-means++	$H_{rms}^{2.5}$, T_p, θ
M7C	K-harmonic means	K-means++	S_y, T_p, θ
M7D	K-harmonic means	MDA	H_rms, T_p, θ
M7E	K-harmonic means	MDA	$H_{rms}^{2.5}$, T_p, θ
M7F	K-harmonic means	MDA	S_y, T_p, θ
M7G	K-harmonic means	Fixed Bins	H_rms, T_p, θ
M7H	K-harmonic means	Fixed Bins	$H_{rms}^{2.5}$, T_p, θ
M7I	K-harmonic means	Fixed Bins	S_y, T_p, θ
M9	Energy Flux Method	-	H_rms, T_p, θ
M10	Sediment Transport Bins Method	-	S_y, T_p, θ

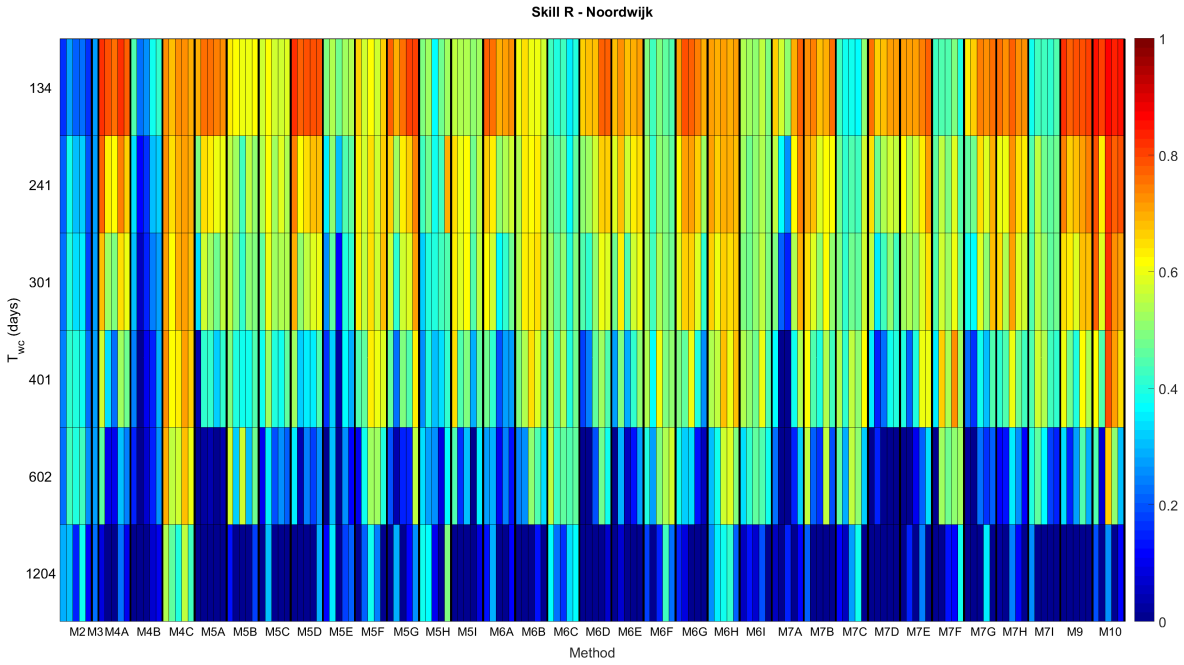


Figure 3.39: Cumulative Skill Score of the 5 random sequences of the simulated methods.

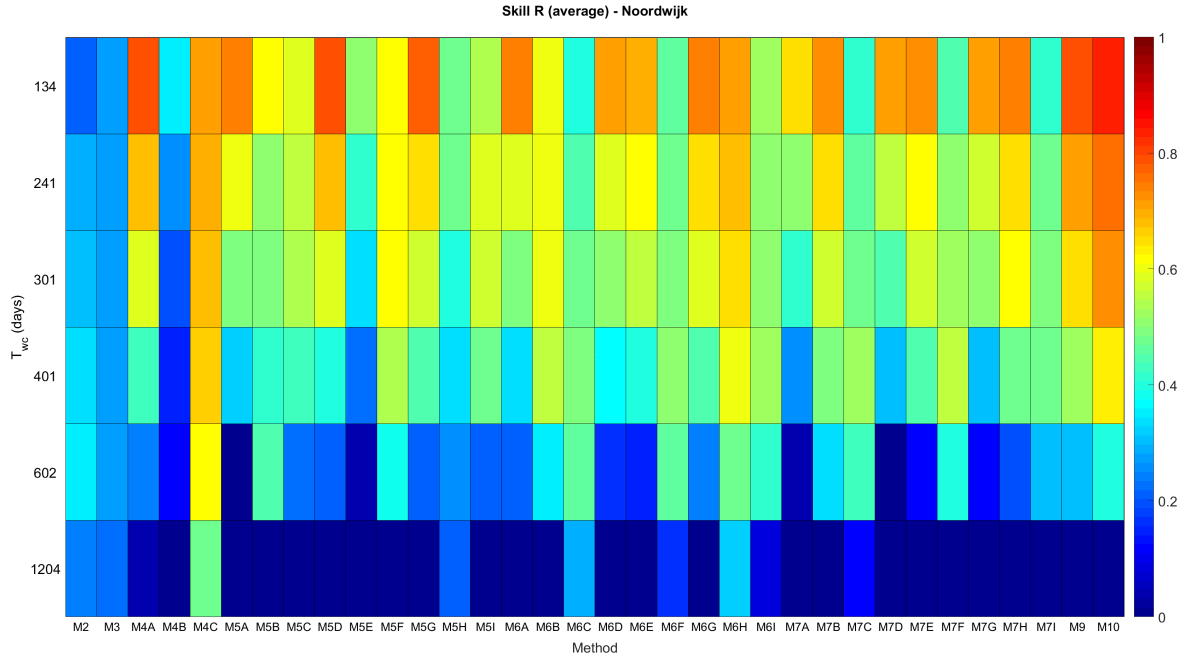


Figure 3.40: Average cumulative Skill Score of the 5 random sequences of the simulated methods.

According to the cumulative skill scores, methods M4C, M6H, M9 and M10 stand out as promising methods and were further investigated in terms of morphology as well as statistics through the instantaneous *BSS*. Figure 3.41 and Figures 3.42 to 3.46 present the profiles at the end of the simulation and the morphological evolution of the methods M4C, M6H, M9 and M10 for a reduce wave climate duration of T_{wc} = 301 days, respectively.

Even though M4C presented relatively high skill score its morphological response is very poor. Furthermore, methods M6H and M9 presented similar skill scores but the morphological response of M6H is evidently worse than M9. This could be due to the tendency of *BSS* (both cumulative and instantaneous) to reward predictions that underestimate the overall magnitude of bed changes (Bosboom et al. 2014). M9 presented a

fairly good morphological response. Nonetheless, M10 presented the highest skill scores as well as the best morphological response.

Figure 3.47 presents the temporal evolution of the instantaneous *BSS* of the methods M4C, M6H, M9 and M10. It is possible to observe that all the methods initially (time ≤ 200 days) have low *BSS*, due to the small variability of the full wave climate model (see Eq. 2.47). Moreover, methods M4C and M6H present a decay of the *BSS* towards the end of the simulation (time ≥ 900 days). M9 and M10 present overall high *BSS* with specific periods of very low scores, however in those specific periods of low *BSS*, M9 has lower scores than M10.

According to the skill scores and morphological evolution, method 10: Sediment Transport Bins was selected for further analysis regarding sequencing and number of representative cases.

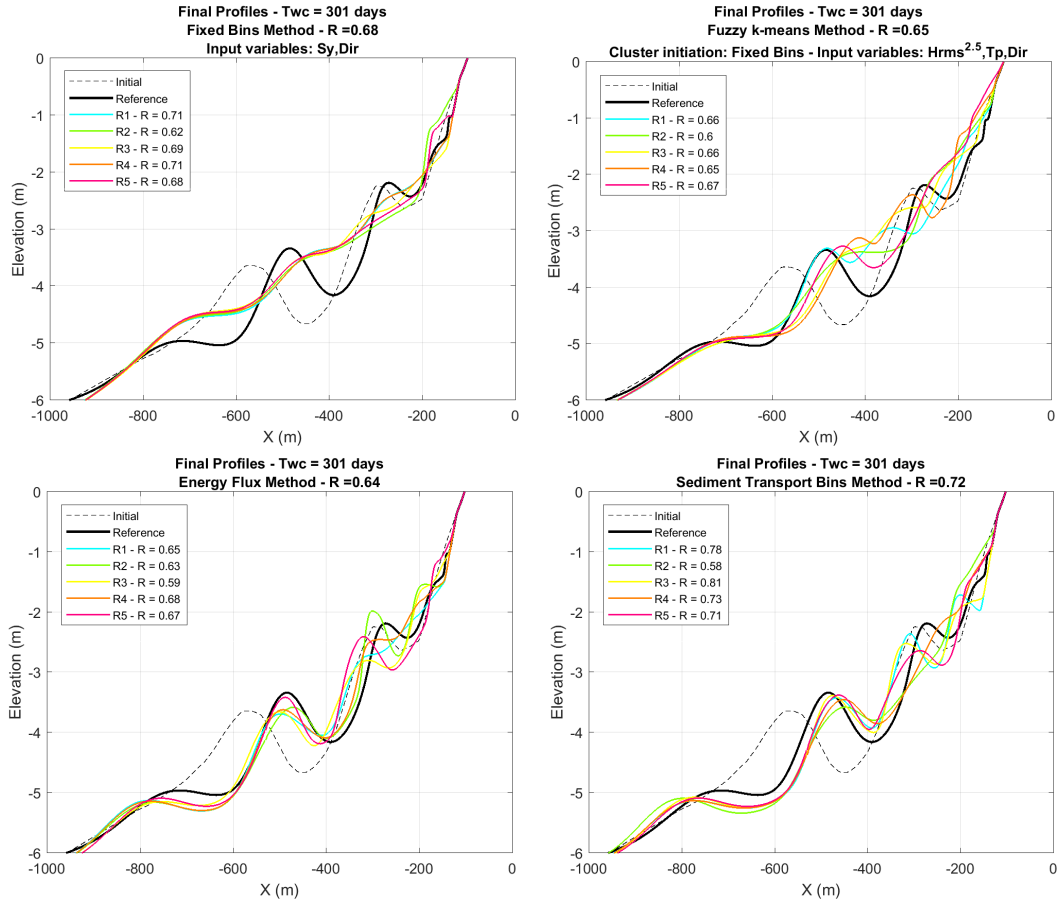


Figure 3.41: Profiles at the end of the simulation with the reduced wave climate of methods M4C (top-left), M6H (top-right), M9 (bottom-left) and M10 (bottom-right). The black lines are the initial (dashed) and final (solid) profiles from the brute force model. The colored lines are the final profiles of each random sequence.

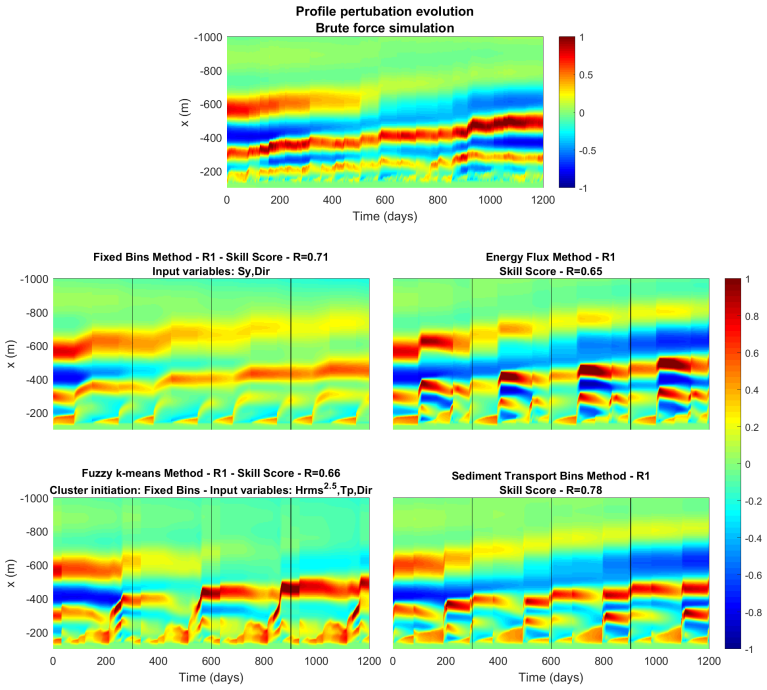


Figure 3.42: Time stacks of profile perturbation of the brute force simulation (top) and reduced wave climates of methods M4C (top-left), M6H (top-right), M9 (bottom-left) and M10 (bottom-right) with random sequence R1. Warm colors correspond to bars while cold colors correspond to troughs.

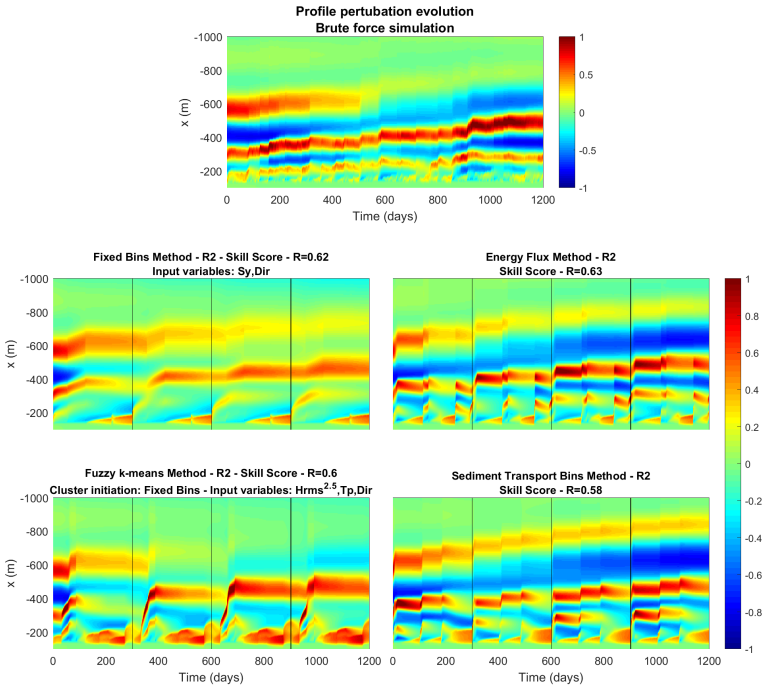


Figure 3.43: Time stacks of profile perturbation of the brute force simulation (top) and reduced wave climates of methods M4C (top-left), M6H (top-right), M9 (bottom-left) and M10 (bottom-right) with random sequence R2. Warm colors correspond to bars while cold colors correspond to troughs.

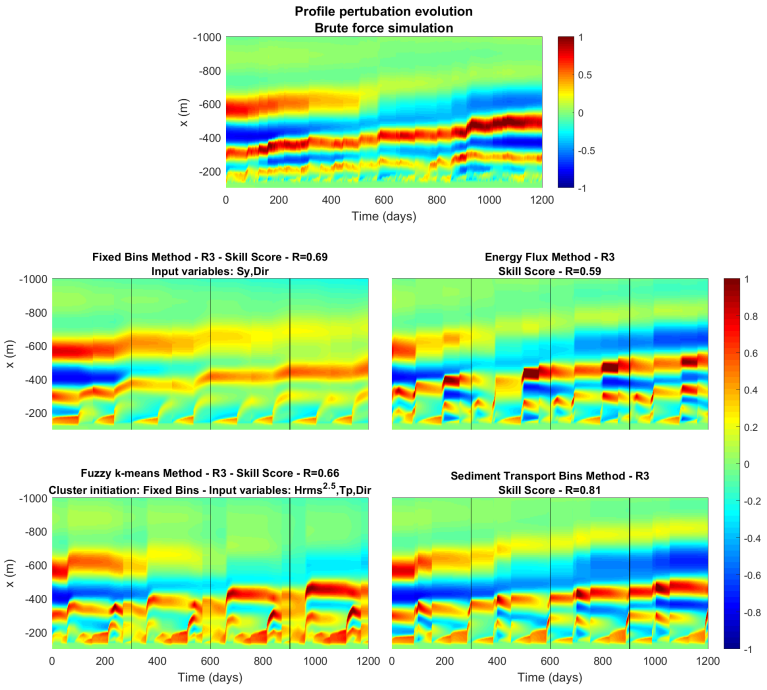


Figure 3.44: Time stacks of profile perturbation of the brute force simulation (top) and reduced wave climates of methods M4C (top-left), M6H (top-right), M9 (bottom-left) and M10 (bottom-right) with random sequence R3. Warm colors correspond to bars while cold colors correspond to troughs.

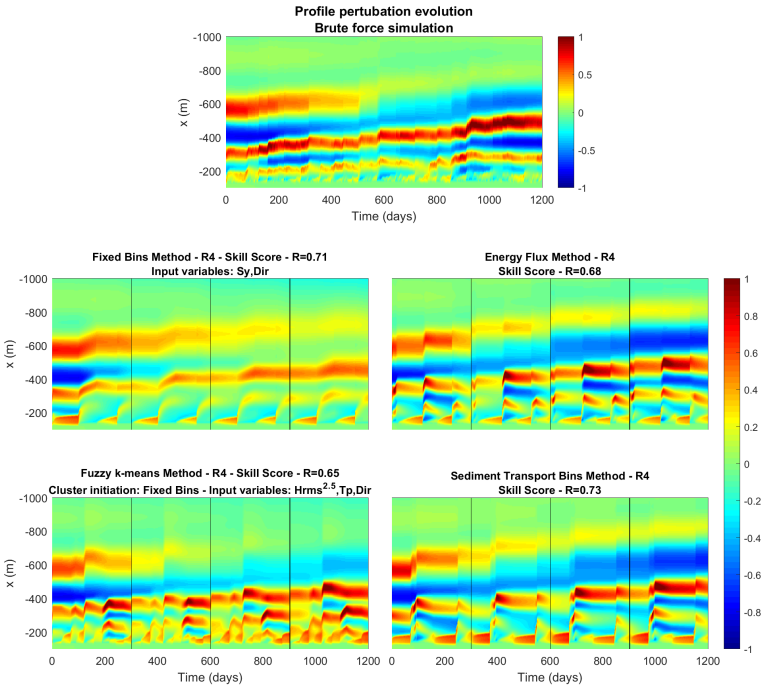


Figure 3.45: Time stacks of profile perturbation of the brute force simulation (top) and reduced wave climates of methods M4C (top-left), M6H (top-right), M9 (bottom-left) and M10 (bottom-right) with random sequence R4. Warm colors correspond to bars while cold colors correspond to troughs.

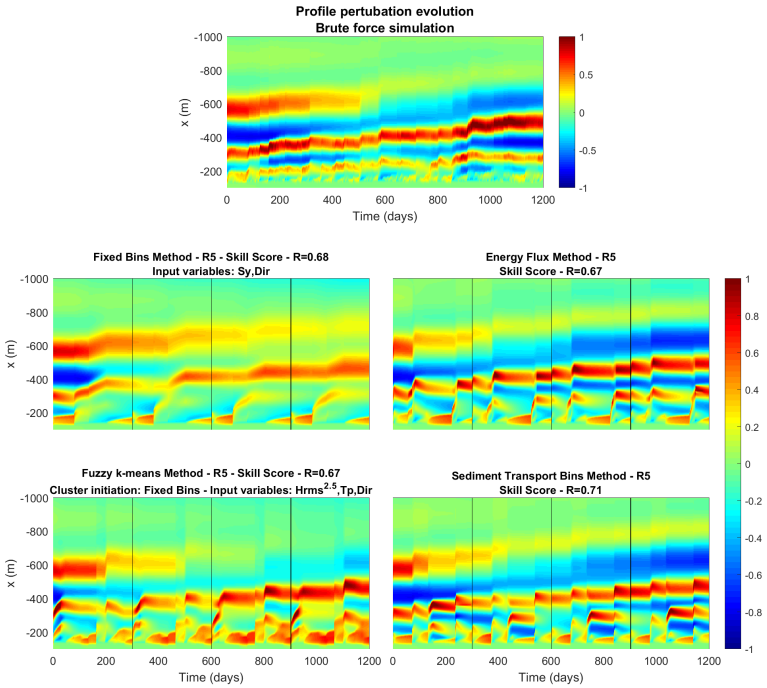


Figure 3.46: Time stacks of profile perturbation of the brute force simulation (top) and reduced wave climates of methods M4C (top-left), M6H (top-right), M9 (bottom-left) and M10 (bottom-right) with random sequence R5. Warm colors correspond to bars while cold colors correspond to troughs.

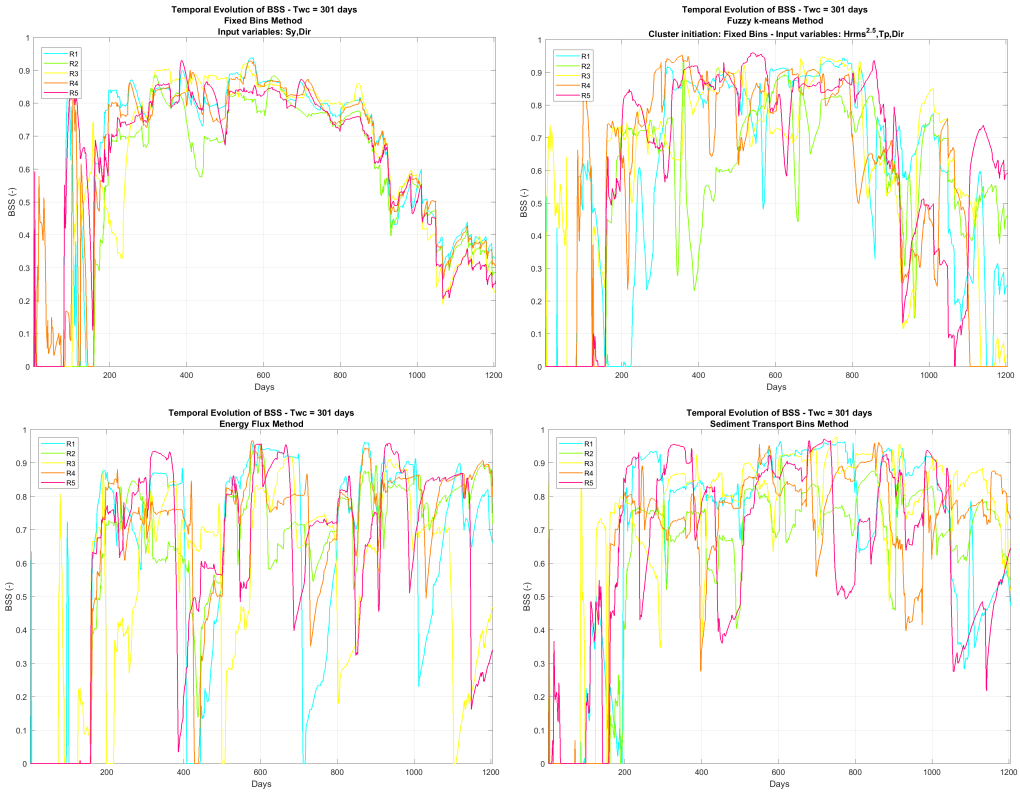


Figure 3.47: Time evolution of instantaneous BSS of methods M4C (top-left), M6H (top-right), M9 (bottom-left) and M10 (bottom-right).

3.13. NUMBER OF CASES

Figure 3.48 presents the selection of the wave cases by method 10: Sediment Transport Bins for the different number of wave cases (k).

3.14. SEQUENCING

Figures 3.49 to 3.57 present the wave height time-series of the reduced wave climate by method 10 with duration of 301 days ($N_R = 3$) sequenced by the random, Markov Chain, Monte Carlo Markov Chain and Monte Carlo Markov Chain with repetition methods with $k = 8$ (4 directional bins vs 2 wave height bins), $k = 10$ (2 directional bins vs 5 wave height bins), $k = 12$ (4 directional bins vs 3 wave height bins), $k = 16$ (4 directional bins vs 4 wave height bins), $k = 24A$ (6 directional bins vs 4 wave height bins), $k = 24B$ (4 directional bins vs 6 wave height bins), $k = 24C$ (8 directional bins vs 3 wave height bins), $k = 32A$ (8 directional bins vs 4 wave height bins) and $k = 32B$ (4 directional bins vs 8 wave height bins).

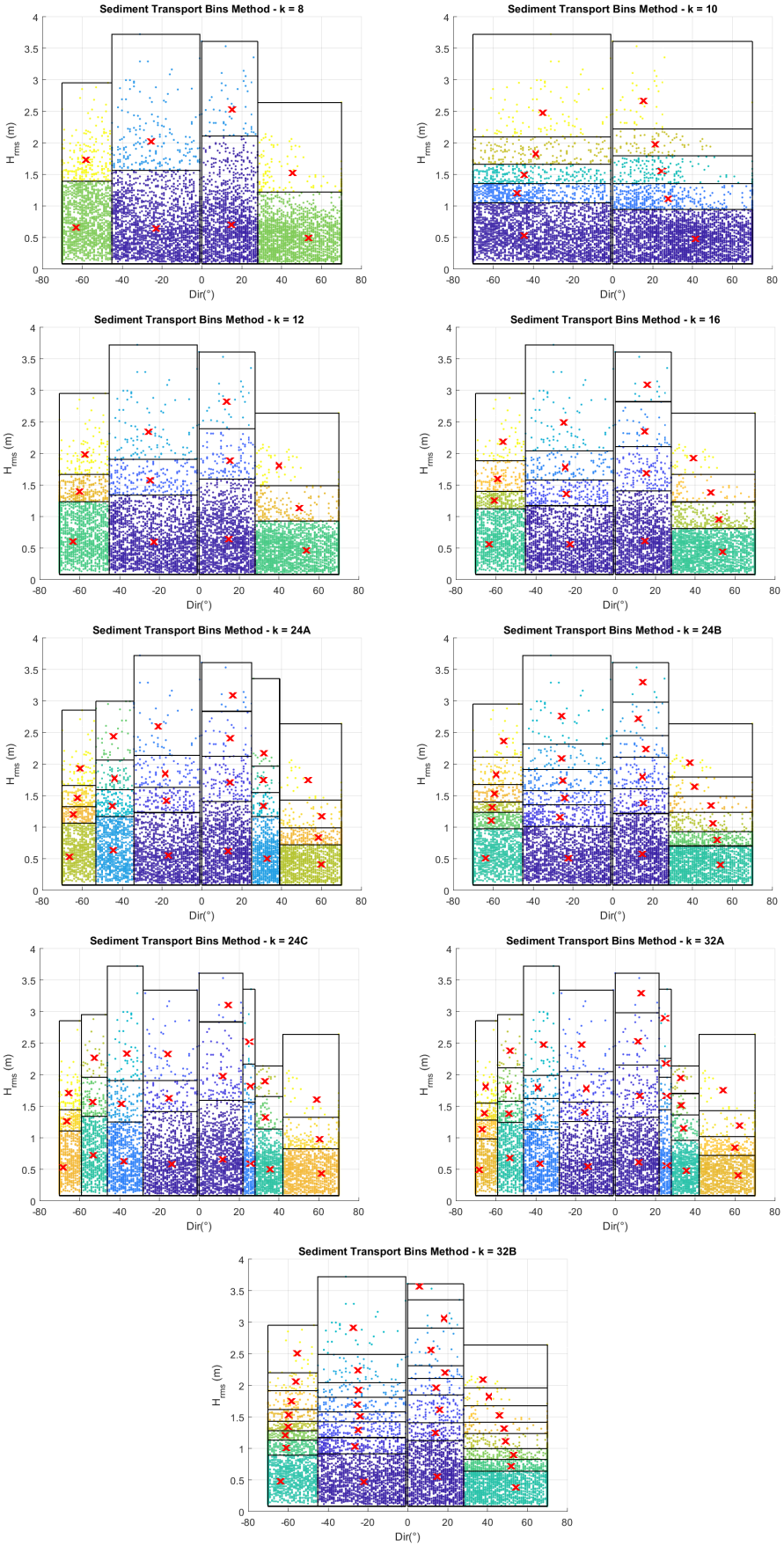


Figure 3.48: Selection of method 10: Sediment Transport Bins with different number of wave cases, *k*. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

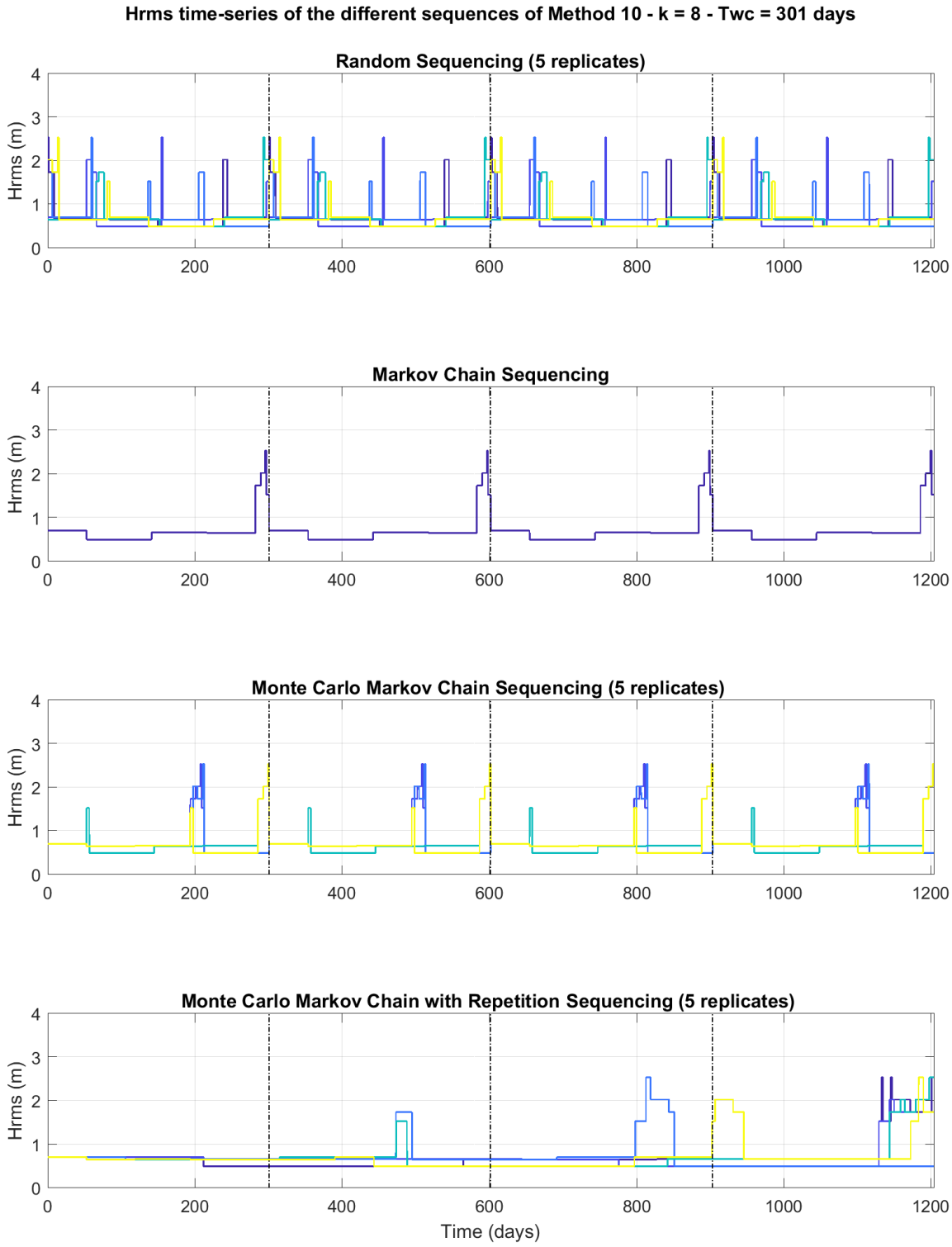


Figure 3.49: Wave height time-series of the reduced wave climate with duration of 301 days sequenced by random (top), Markov Chain (top-down), Monte Carlo Markov Chain (bottom-up) and Monte Carlo Markov Chain with repetition (bottom) methods with $k = 8$.

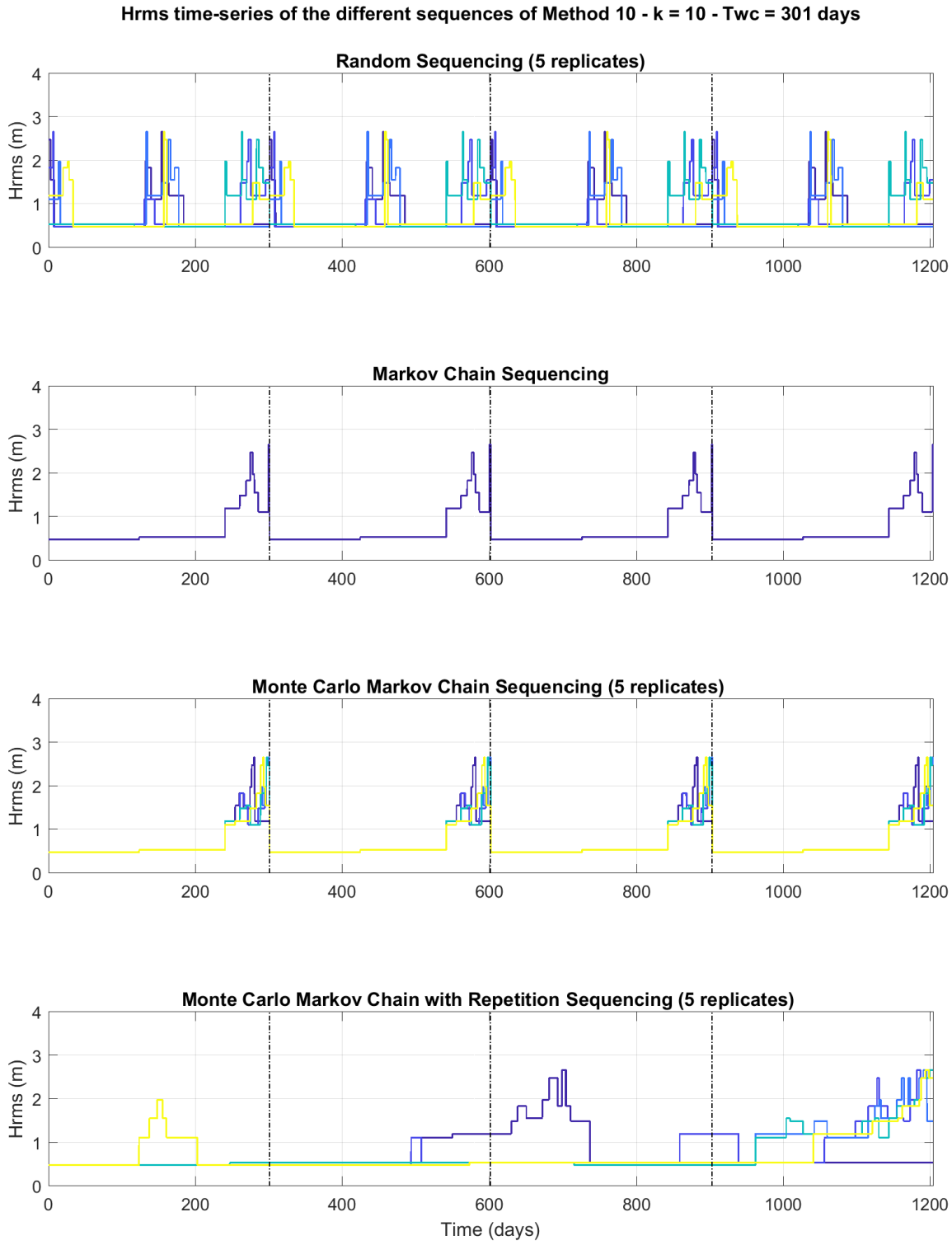


Figure 3.50: Wave height time-series of the reduced wave climate with duration of 301 days sequenced by random (top), Markov Chain (top-down), Monte Carlo Markov Chain (bottom-up) and Monte Carlo Markov Chain with repetition (bottom) methods with $k = 10$.

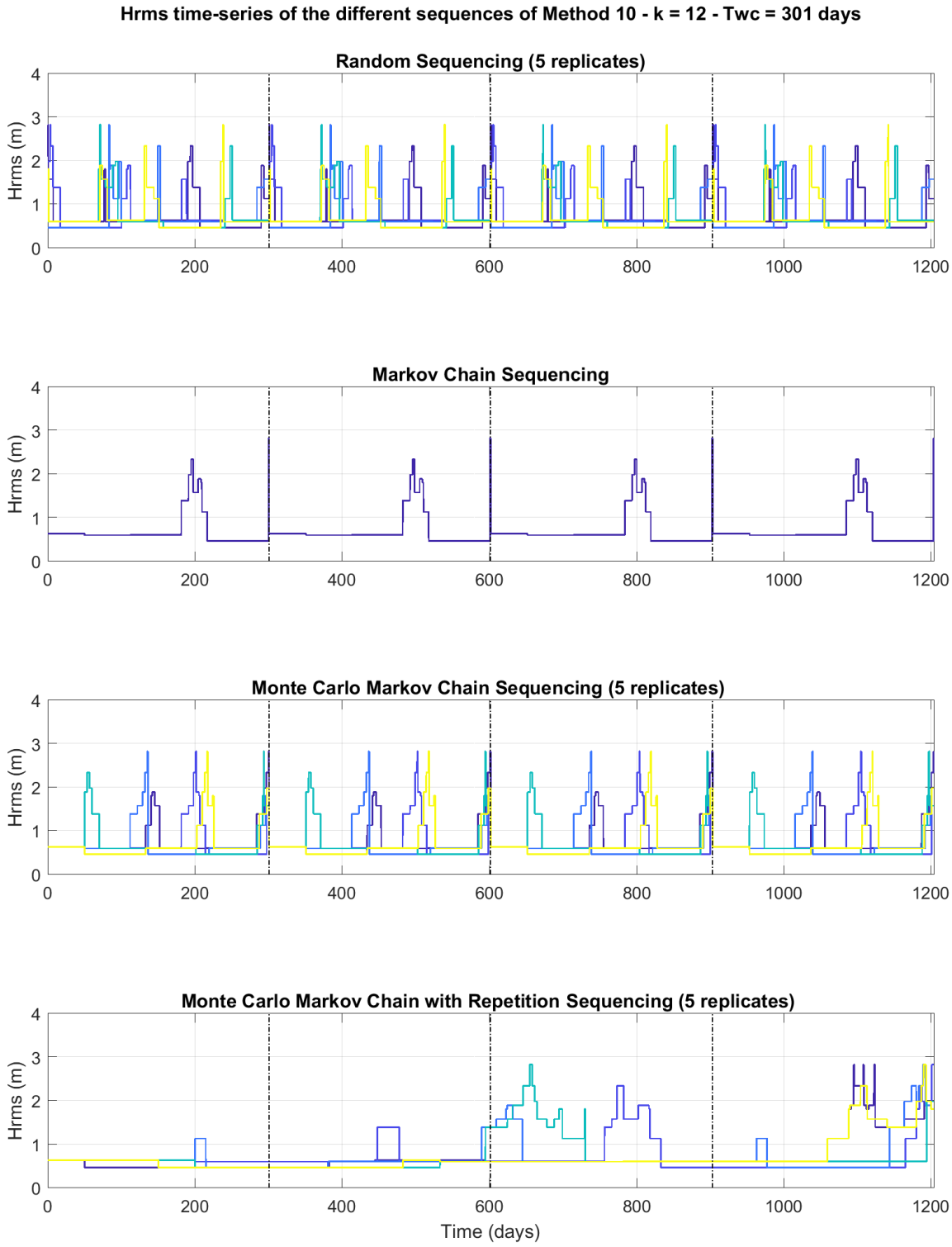


Figure 3.51: Wave height time-series of the reduced wave climate with duration of 301 days sequenced by random (top), Markov Chain (top-down), Monte Carlo Markov Chain (bottom-up) and Monte Carlo Markov Chain with repetition (bottom) methods with $k = 12$.

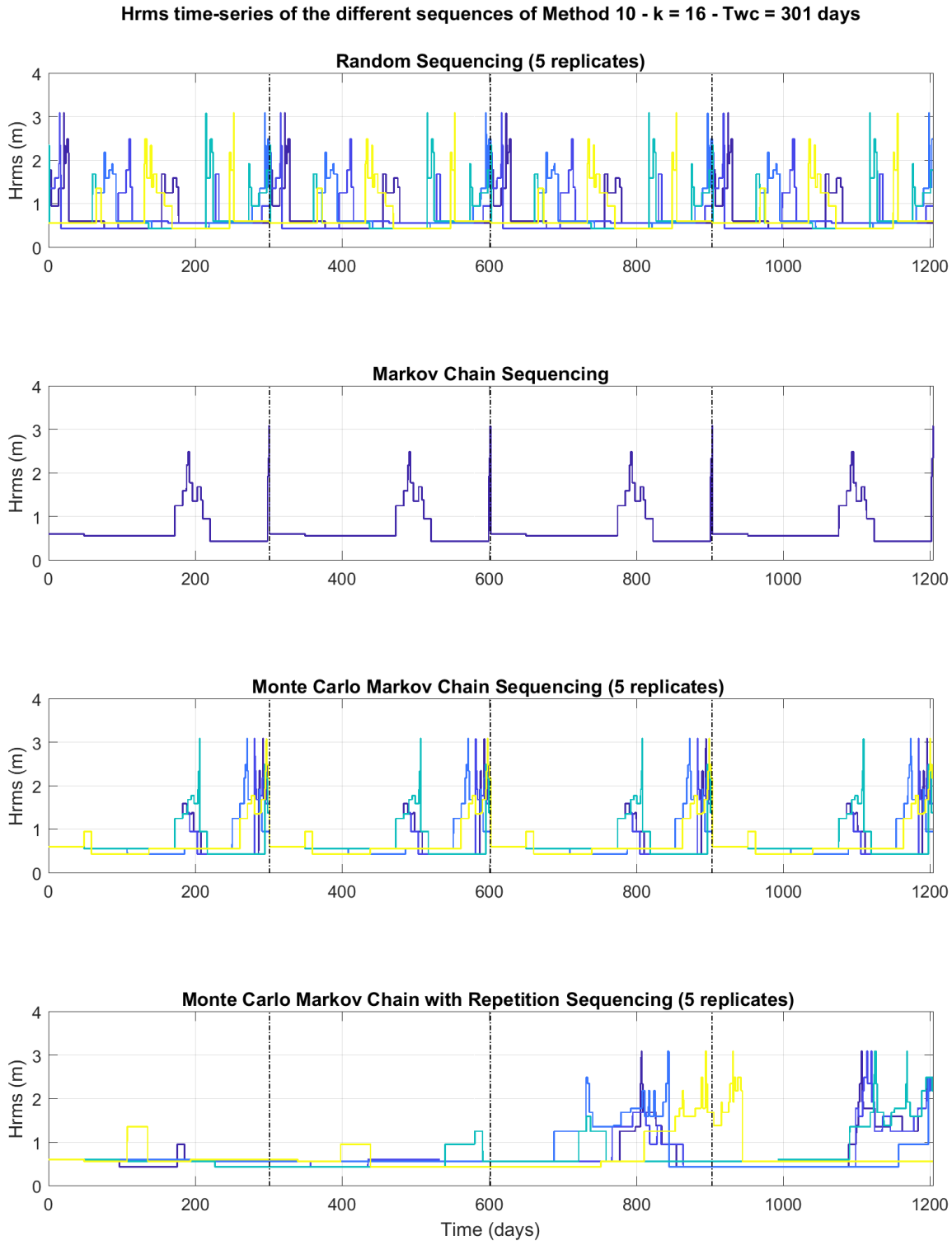


Figure 3.52: Wave height time-series of the reduced wave climate with duration of 301 days sequenced by random (top), Markov Chain (top-down), Monte Carlo Markov Chain (bottom-up) and Monte Carlo Markov Chain with repetition (bottom) methods with $k = 16$.

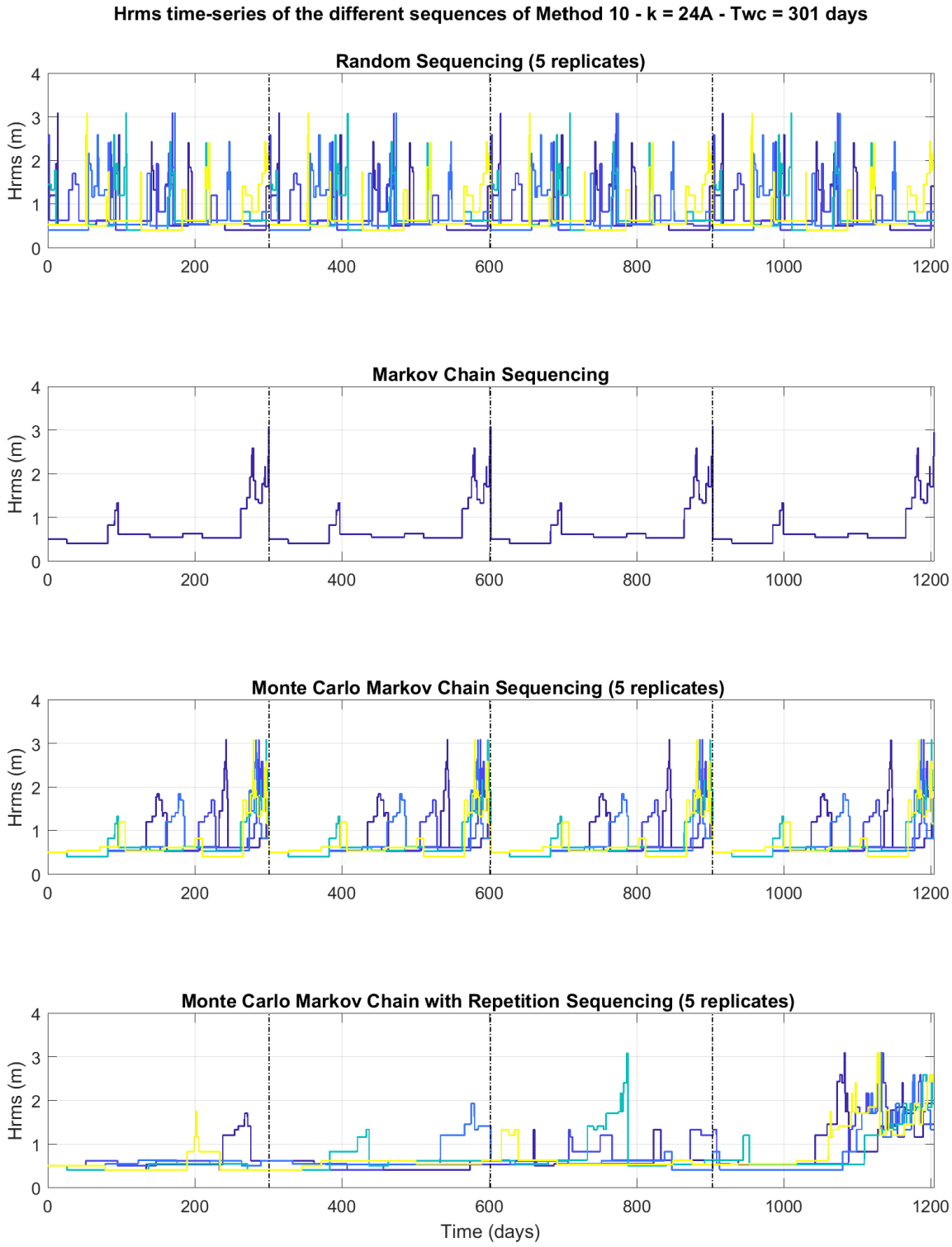


Figure 3.53: Wave height time-series of the reduced wave climate with duration of 301 days sequenced by random (top), Markov Chain (top-down), Monte Carlo Markov Chain (bottom-up) and Monte Carlo Markov Chain with repetition (bottom) methods with $k = 24A$.

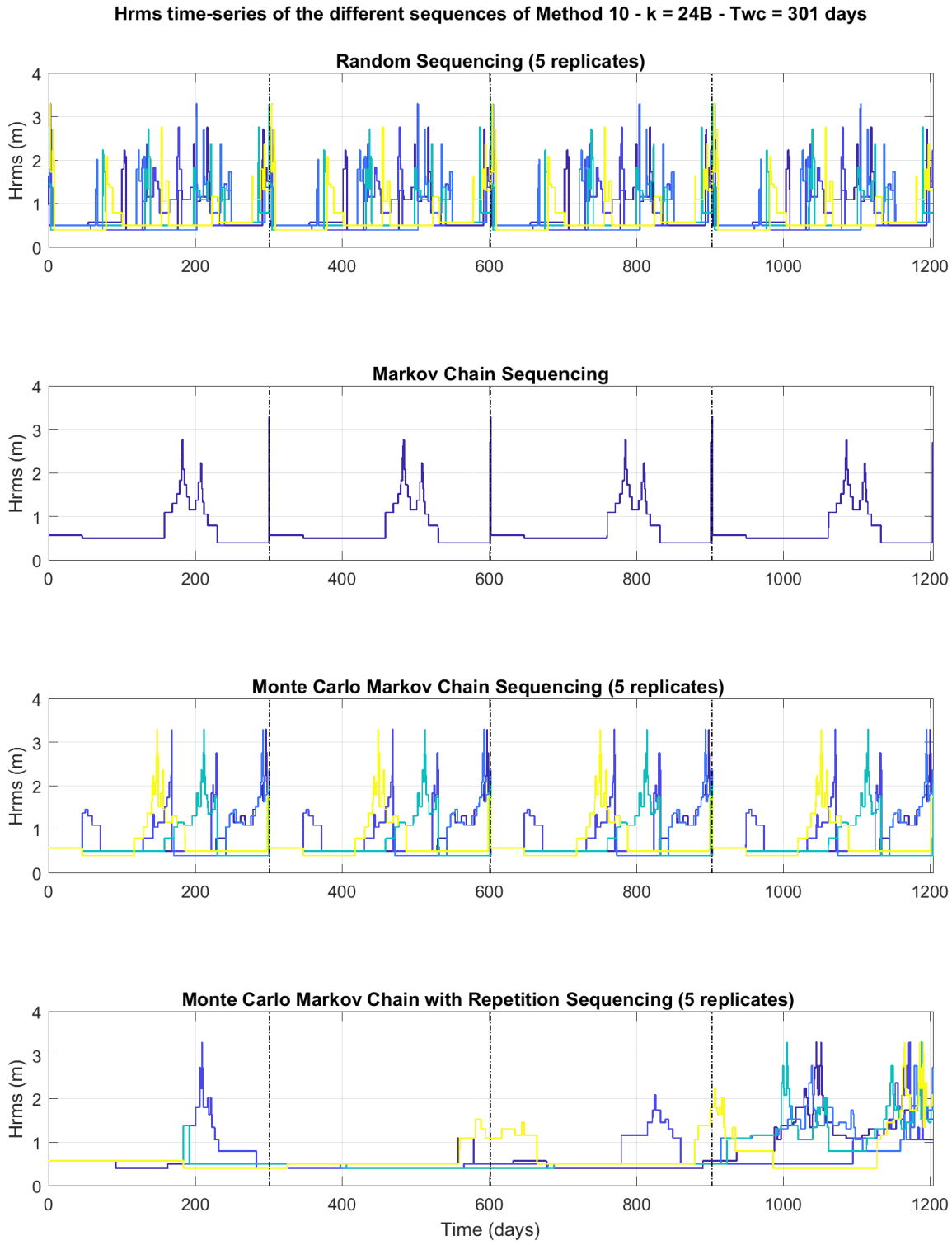


Figure 3.54: Wave height time-series of the reduced wave climate with duration of 301 days sequenced by random (top), Markov Chain (top-down), Monte Carlo Markov Chain (bottom-up) and Monte Carlo Markov Chain with repetition (bottom) methods with $k = 24B$.

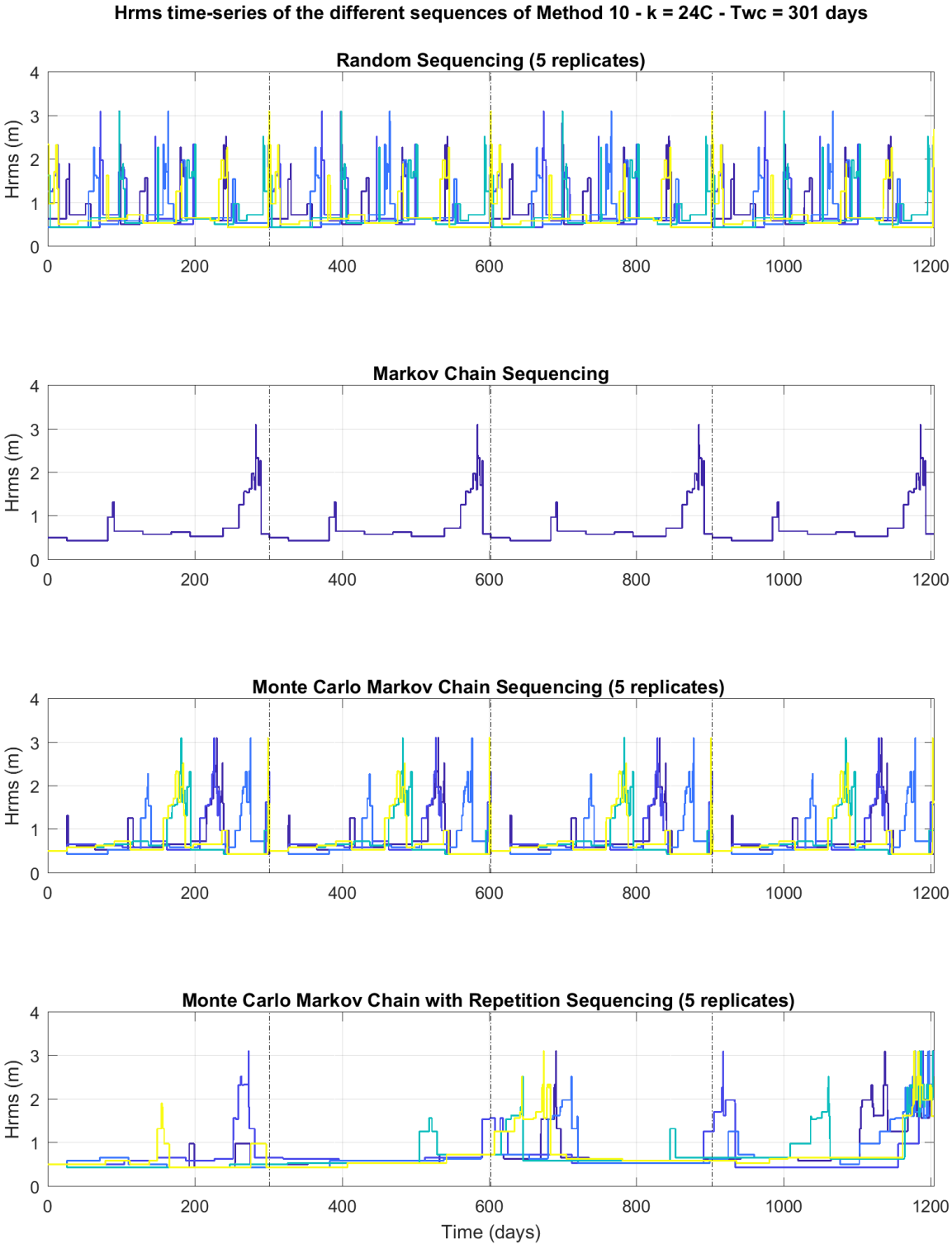


Figure 3.55: Wave height time-series of the reduced wave climate with duration of 301 days sequenced by random (top), Markov Chain (top-down), Monte Carlo Markov Chain (bottom-up) and Monte Carlo Markov Chain with repetition (bottom) methods with $k = 24C$.

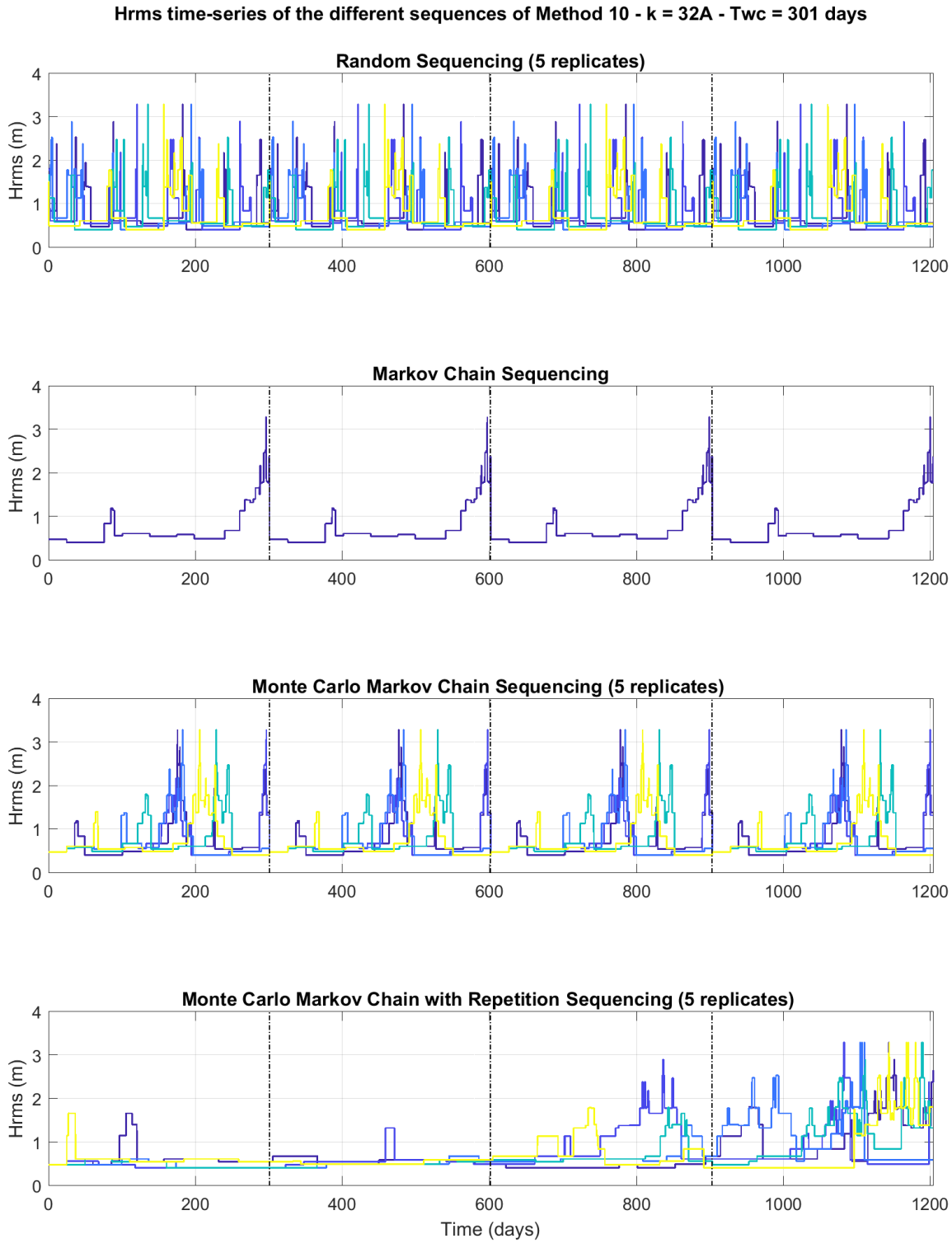


Figure 3.56: Wave height time-series of the reduced wave climate with duration of 301 days sequenced by random (top), Markov Chain (top-down), Monte Carlo Markov Chain (bottom-up) and Monte Carlo Markov Chain with repetition (bottom) methods with $k = 32A$.

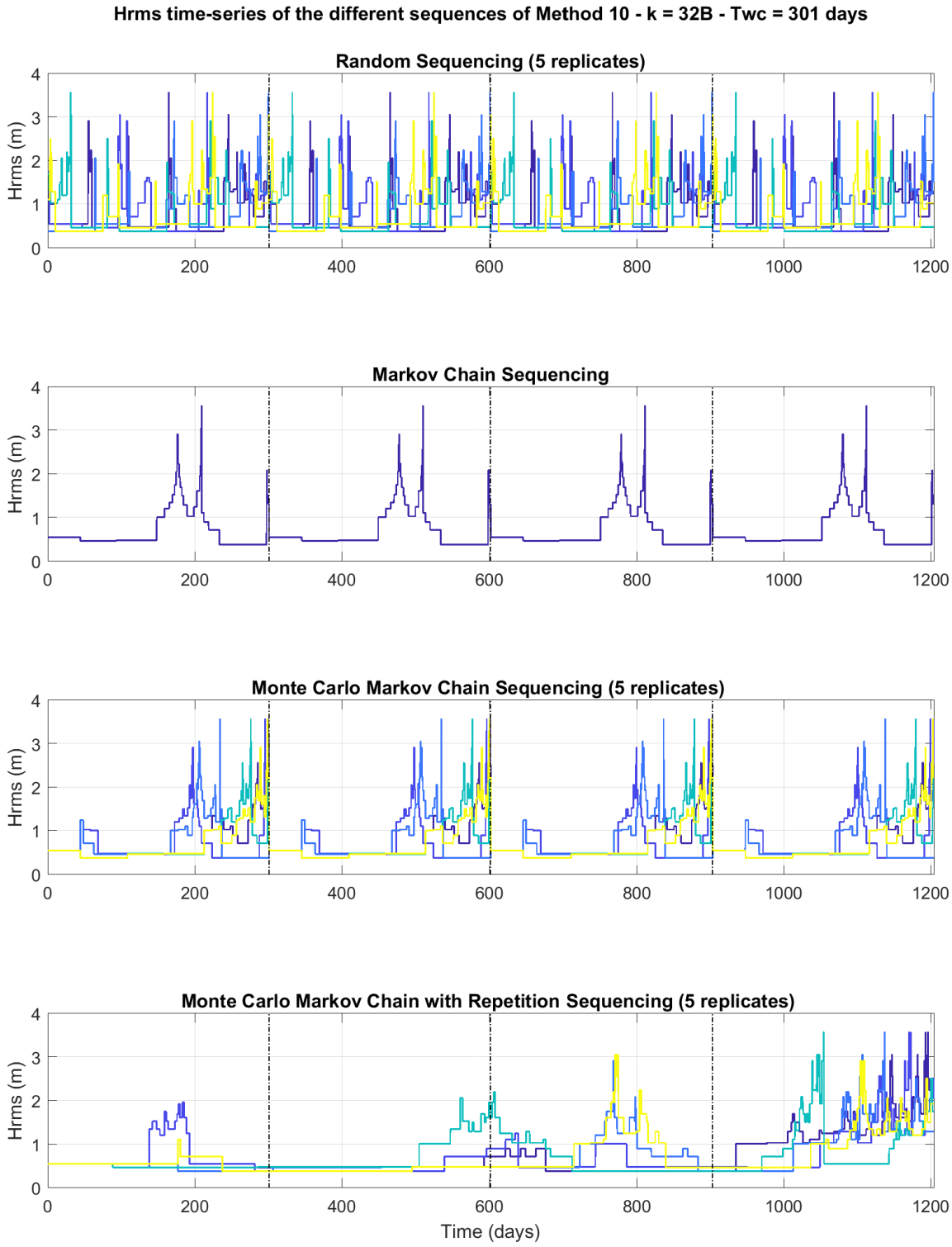


Figure 3.57: Wave height time-series of the reduced wave climate with duration of 301 days sequenced by random (top), Markov Chain (top-down), Monte Carlo Markov Chain (bottom-up) and Monte Carlo Markov Chain with repetition (bottom) methods with $k = 32B$.

3.15. EVALUATION OF NUMBER OF CASES AND SEQUENCING

The sequencing and number of cases were, initially, evaluated with the cumulative skill score R (see section 2.6). Table 3.6 lists the different sequencing methods simulated of method 10: Sediment Transport Bins.

Table 3.6: Simulated sequencing methods

Name	Sequencing Method
S1	Random (5 replicates)
S2	Markov Chain
S3	Monte Carlo Markov Chain (5 replicates)
S4	Monte Carlo Markov Chain with repetition (5 replicates)

Figure 3.58 presents the cumulative skill score at the end of the simulation of the 5 replicates of the sequencing methods and the number of cases for each wave climate duration simulated of method 10. Figure 3.59 presents the mean skill score of the 5 replicates of the sequencing methods and the number of cases for every wave climate duration simulated of method 10. Note that Markov Chain Sequencing (S2) does not contains randomness, thus does not have replicates. Also, the Monte Carlo Markov Chain Sequencing with repetition (S4) does not apply for $T_{wc} = 1205$ days, once the reduced wave climate is applied to the full timescale of the reduction period (1205 days), therefore repetition of wave conditions is not possible.

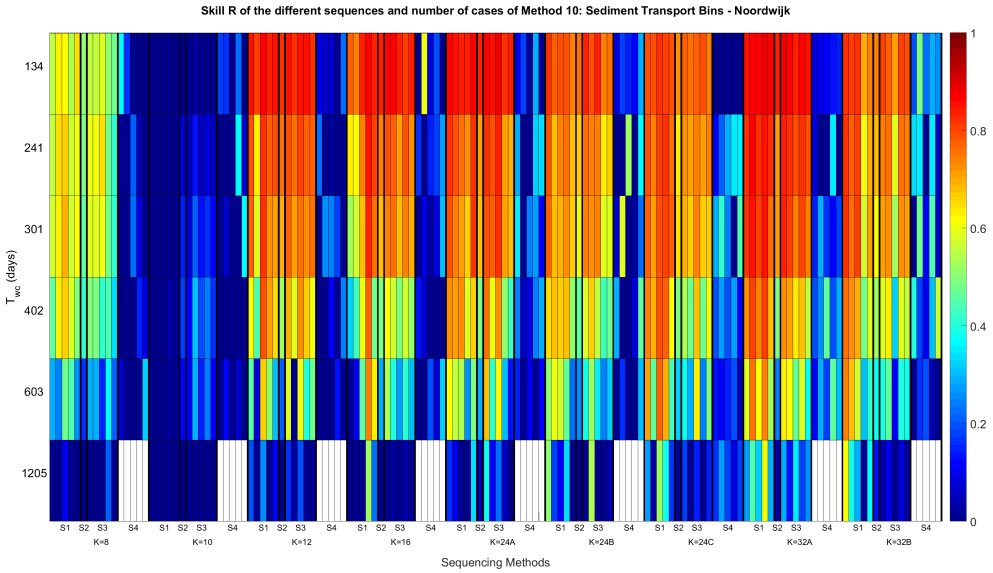


Figure 3.58: Cumulative Skill Score of the 5 replicates of the different sequencing and number of cases of method 10: Sediment Transport Bins.

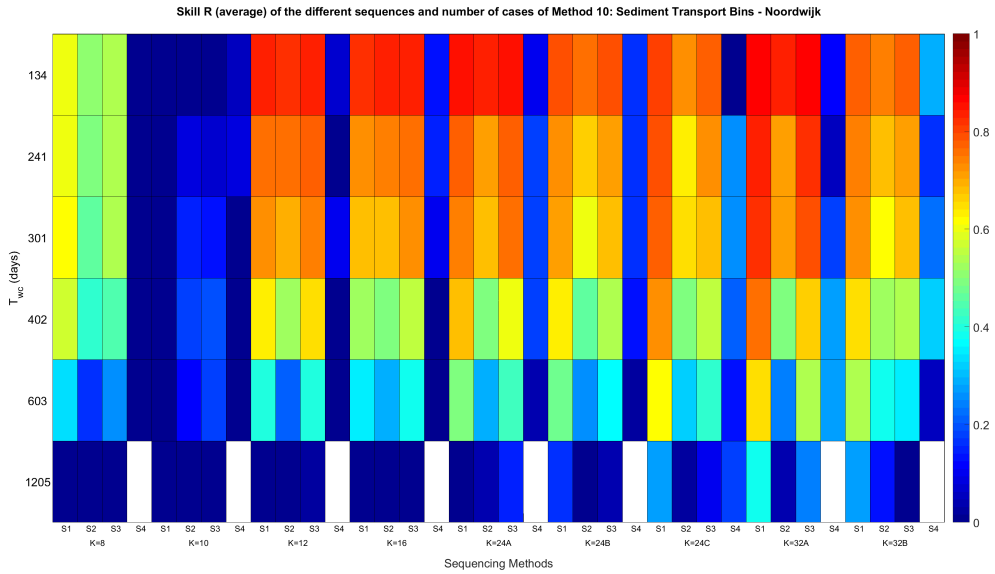


Figure 3.59: Average cumulative Skill Score of the 5 replicates of the different sequencing and number of cases of method 10: Sediment Transport Bins.

Once again, the analysis focused on the wave climate duration of 301 days due to its applicability. It is observed that $k = 8$ (4 directional bins vs 2 wave height bins) and $k = 10$ (2 directional bins vs 5 wave height bins) present very low skill scores relative to other number of cases. These results indicate that $k = 8$ and $k = 10$ are poor quantities of representative wave conditions for morphology. Moreover, $k = 10$ presents even lower skill scores than $k = 8$. This is because $k = 10$ has only 2 directional bins whereas $k = 8$ has 4. The resolution of the directional bins is a relevant aspect on reducing wave climate for morphology since the influence of the wave angle on the longshore transport and consequently morphology has a different behavior than the wave height. The influence of the wave angle can be approximated to a S-ϕ curve where the maximum transport is around $\phi = 45°$ relative to the shore-normal if one consider a straight stretch of coast (Figure 3.60). So wave angles relative to the shore-normal from 0 to 45° present increasing transport rates while wave angles relative to the shore-normal from 45° to 90° present decreasing transport rates (Bosboom & Stive 2011). Therefore, having low resolution on the directional bins implied in averaging out transport contributions resulting in poor morphological evolution. The influence of the wave height in the longshore transport and morphology even though non-linear, is always proportional. So higher the wave height, much higher the longshore transport rates due to its non-linearity that is usually represented by a power in the range of 2-3. Furthermore, increasing the number of cases does not imply in substantial improvement on the skill scores, except when the number of directional bins is increased considerably such as $k = 24C$ and $k = 32A$ with 8 directional bins.

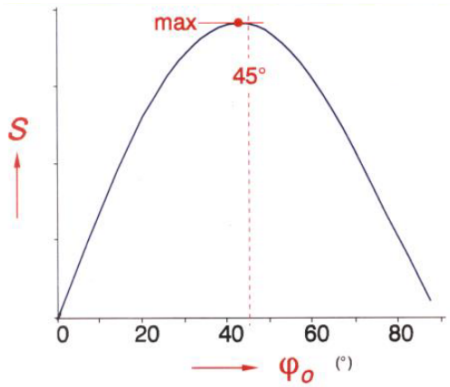


Figure 3.60: Theoretical S-ϕ curve. Adapted from Bosboom & Stive (2011)

Regarding sequencing, the Monte Carlo Markov Chain with repetition (MCMCR - S4) presents very low skill scores and that it is due its lack of variability on the reduced wave climate caused by the allowance of repetition of wave conditions. Once the highest probabilities of the Markov Chain transitions are to remain on the same state, the wave conditions tend to repeat themselves creating time-series with long periods of calm wave conditions followed by short and aggregated periods of energetic wave conditions. The reduced wave climate requires a certain variability that resembles the full dataset and the reduced time-series originated by the MCMCR sequecing (S4) does not provide this. Nevertheless, this is a good demonstration of how much the sequencing of the reduced wave climate can affect morphological evolution.
Additionally, the Markov Chain (MC - S2) sequencing presents lower skill scores than random (S1) and Monte Carlo Markov Chain (MCMC - S1). This occurs because the MC sequencing method tends to select primarily low wave cases which have the highest Markov Chain transition probabilities, thus leaving to the end the remaining high wave height conditions. Hence, the MC sequencing also generates reduced time-series with very low variability, similarly to ordinating the wave conditions from low to high wave height. Moreover, when decreasing the wave climate duration, that is increasing the number of repetition of the reduced wave climate, the skill score of the MC sequencing becomes equivalent to the random and MCMC sequencing. This occurs because the repetition of the reduced wave climate overcomes the low variability of the MC sequencing method. However, it does not improve at the same rate the sequencing methods that have already high variability such as random and MCMC. The random and MCMC sequencing methods present very similar performances once these sequencing methods contain randomness implying in more variability of the reduced time-series.
The random, Markov Chain and Monte Carlo Markov Chain sequencing methods and the number of cases $k = 12$, $k = 16$, $k = 24A$, $k = 24B$, $k = 24C$, $k = 32A$, $k = 32B$ were further investigated in terms of morphological evolution. Figures 3.62 to 3.70 present the profiles at the end of the simulation and the morphological evolution of method 10: Sediment Transport Bins with wave climate duration of 301 days for $k = 12$, $k = 16$, $k = 24A$, $k = 24B$, $k = 24C$, $k = 32A$, $k = 32B$ and random, MCMC and MC sequencing methods.

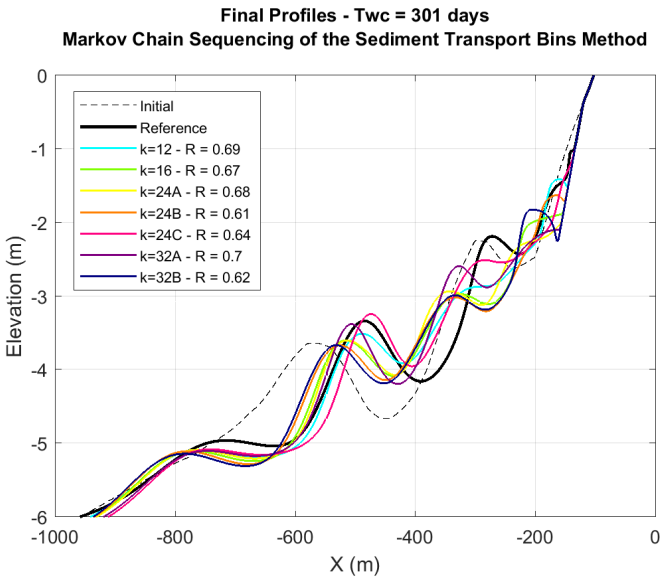


Figure 3.61: Profiles at the end of the simulation with the reduced wave climate of method 10: Sediment Transport Bins with $k = 12$, $k = 16$, $k = 24A$, $k = 24B$, $k = 24C$, $k = 32A$, $k = 32B$ and Markov Chain sequencing. The black lines are the initial (dashed) and final (solid) profiles from the brute force model. The colored lines are the final profiles of each number of wave cases.

For $k = 12$, the final profile with MCMC sequencing presents slightly better results than the random sequencing, specially on the most seawards bar. However, the morphological evolution points the random sequencing as slightly better than the MCMC sequencing. So, for $k = 12$, random and MCMC sequencing can be consider equivalent. This occurs because the variability on the wave height time-series of the random and MCMC sequencing are very similar (see Figure 3.51)
For $k = 16$, $k = 24A$, $k = 24B$, $k = 24C$, $k = 32A$ and $k = 32B$, the random sequencing presents both final profiles and morphological evolution better than the MCMC sequencing. This is because the variability of the

time-series with random sequencing is still higher than with MCMC sequencing (see Figures 3.52 to 3.57). Additionally, increasing the number of cases from 12 (4 directional bins vs 3 wave height bins) to 16 (4 directional bins vs 4 wave height bins) does not result in any significant improvement on the performance of the model. This is because $k = 16$ only increases the resolution of the wave height bins that does not have such influence on the resultant morphology as the directional bins that remained the same as $k = 12$. The same occurs with $k = 24A$ (6 directional bins vs 4 wave height bins), $k = 24B$ (4 directional bins vs 6 wave height bins), $k = 24C$ (8 directional bins vs 3 wave height bins), $k = 32A$ (8 directional bins vs 4 wave height bins) and $k = 32B$ (4 directional bins vs 8 wave height bins) in which $k = 24A$, $k = 24C$ and $k = 32A$ present better morphological results than $k = 24B$ and $k = 32B$, respectively, due to the increase on the directional bins resolution. Furthermore, it is important to note that increasing the number of cases from $k = 16$ to $k = 24$ resulted in some improvement in the morphology but not as much as increasing the number of wave cases to $k = 32$.

The MC sequencing presented considerably worse morphological representation in relation to random and MCMC sequencing even when increasing the number of wave cases which demonstrates that the effect of sequencing can overcomes the influence of the number of directional bins.

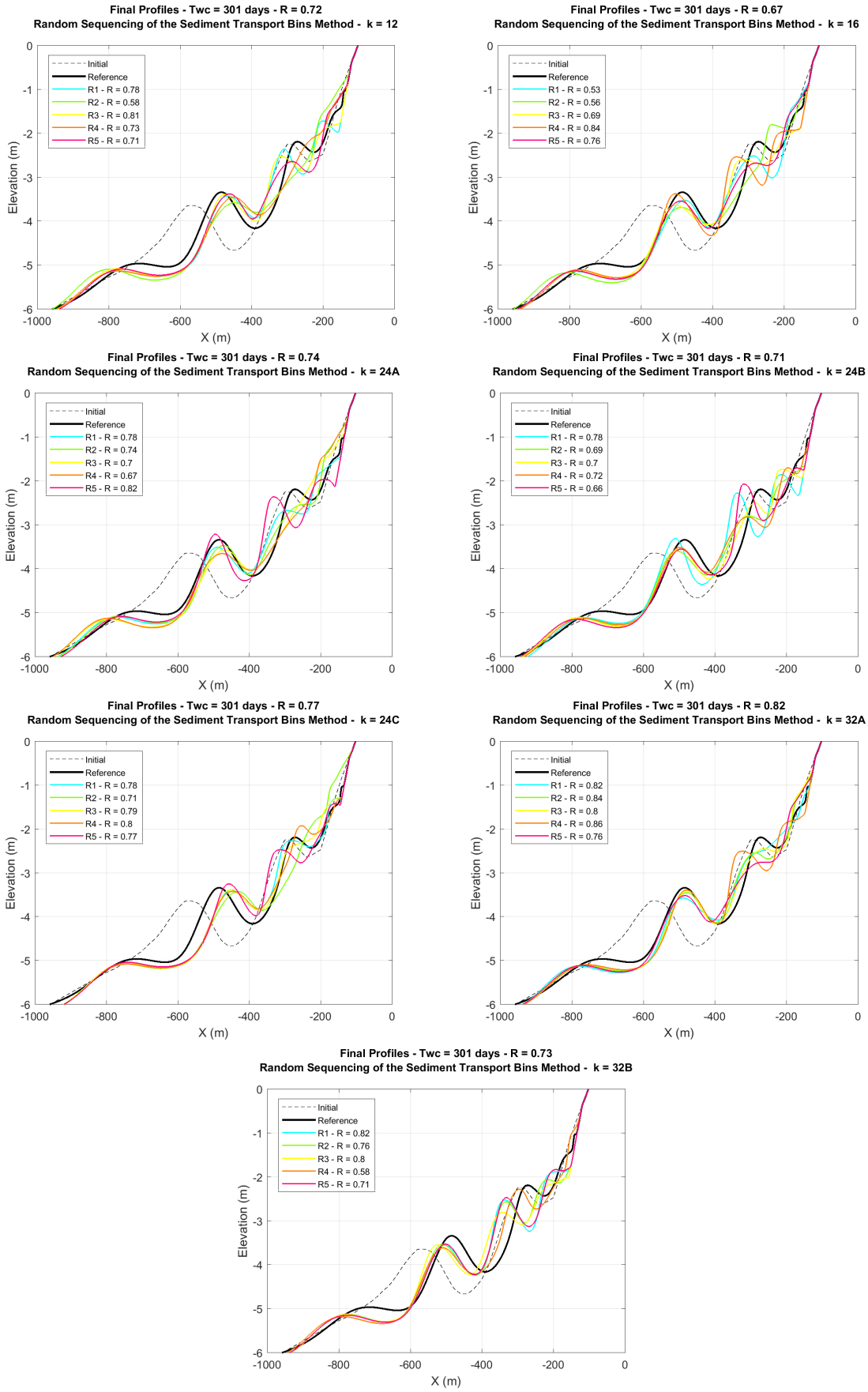


Figure 3.62: Profiles at the end of the simulation with the reduced wave climate of method 10: Sediment Transport Bins with $k = 12$, $k = 16$, $k = 24A$, $k = 24B$, $k = 24C$, $k = 32A$, $k = 32B$ and random sequencing. The black lines are the initial (dashed) and final (solid) profiles from the brute force model. The colored lines are the final profiles of each random sequence.

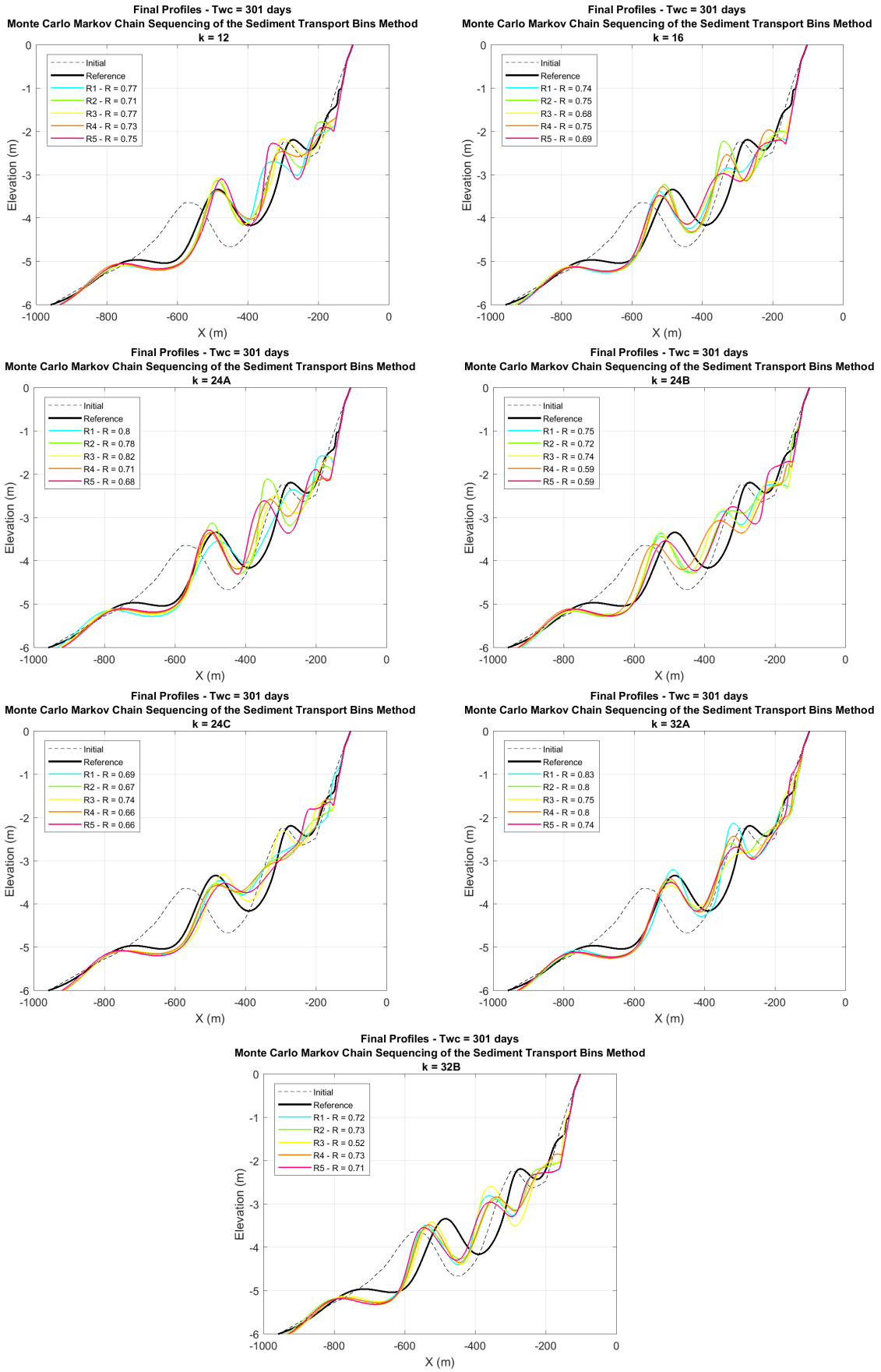


Figure 3.63: Profiles at the end of the simulation with the reduced wave climate of method 10: Sediment Transport Bins with $k = 12$, $k = 16$, $k = 24A$, $k = 24B$, $k = 24C$, $k = 32A$, $k = 32B$ and Monte Carlo Markov Chain sequencing. The black lines are the initial (dashed) and final (solid) profiles from the brute force model. The colored lines are the final profiles of each replicate sequence.

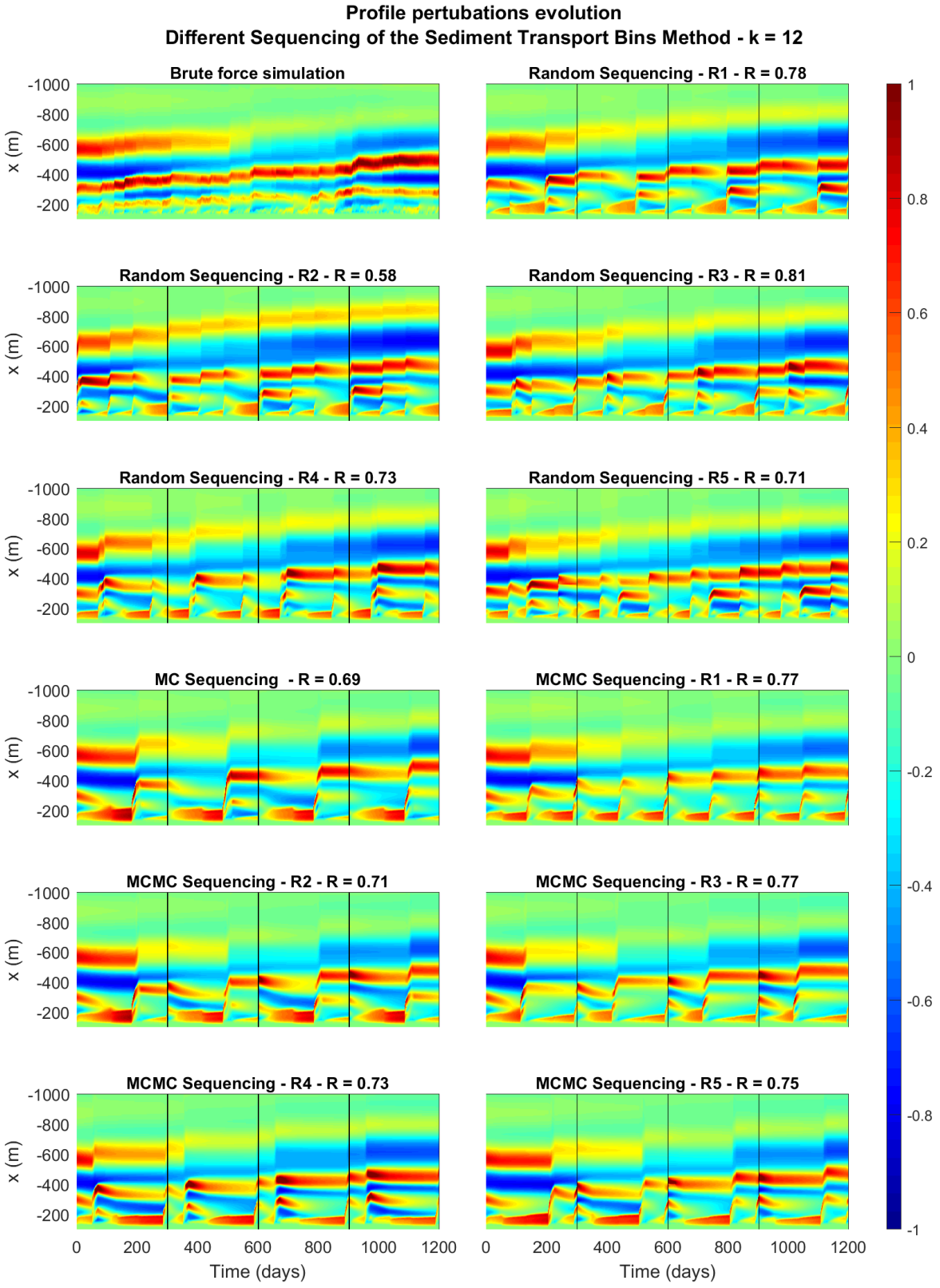


Figure 3.64: Time stacks of profile perturbation of the brute force simulation (top) and reduced wave climates of method 10: Sediment Transport Bins with $k = 12$ for the random, Markov Chain and Monte Carlo Markov Chain sequencing. Warm colors correspond to bars while cold colors correspond to troughs.

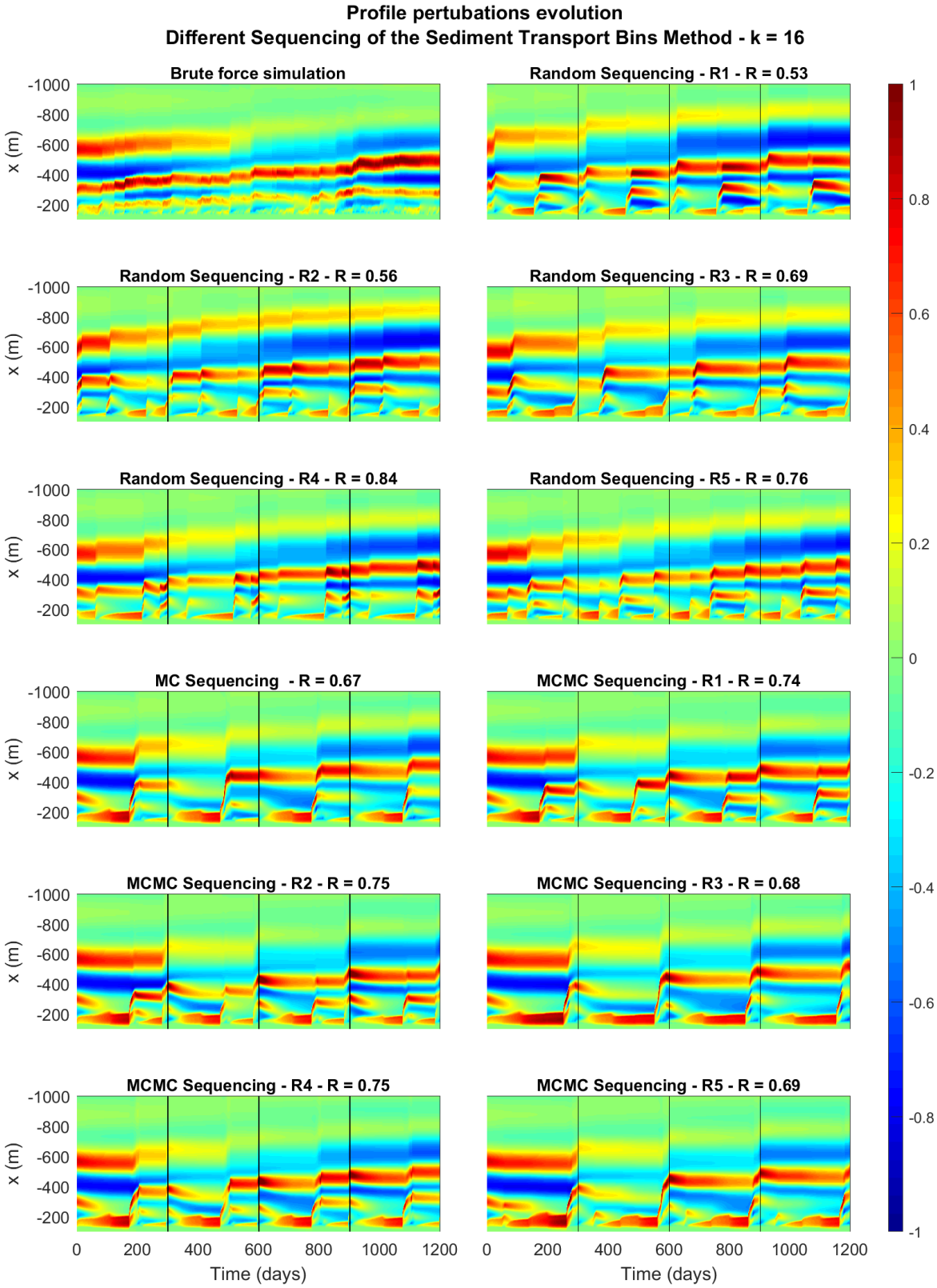


Figure 3.65: Time stacks of profile perturbation of the brute force simulation (top) and reduced wave climates of method 10: Sediment Transport Bins with $k = 16$ for the random, Markov Chain and Monte Carlo Markov Chain sequencing. Warm colors correspond to bars while cold colors correspond to troughs.

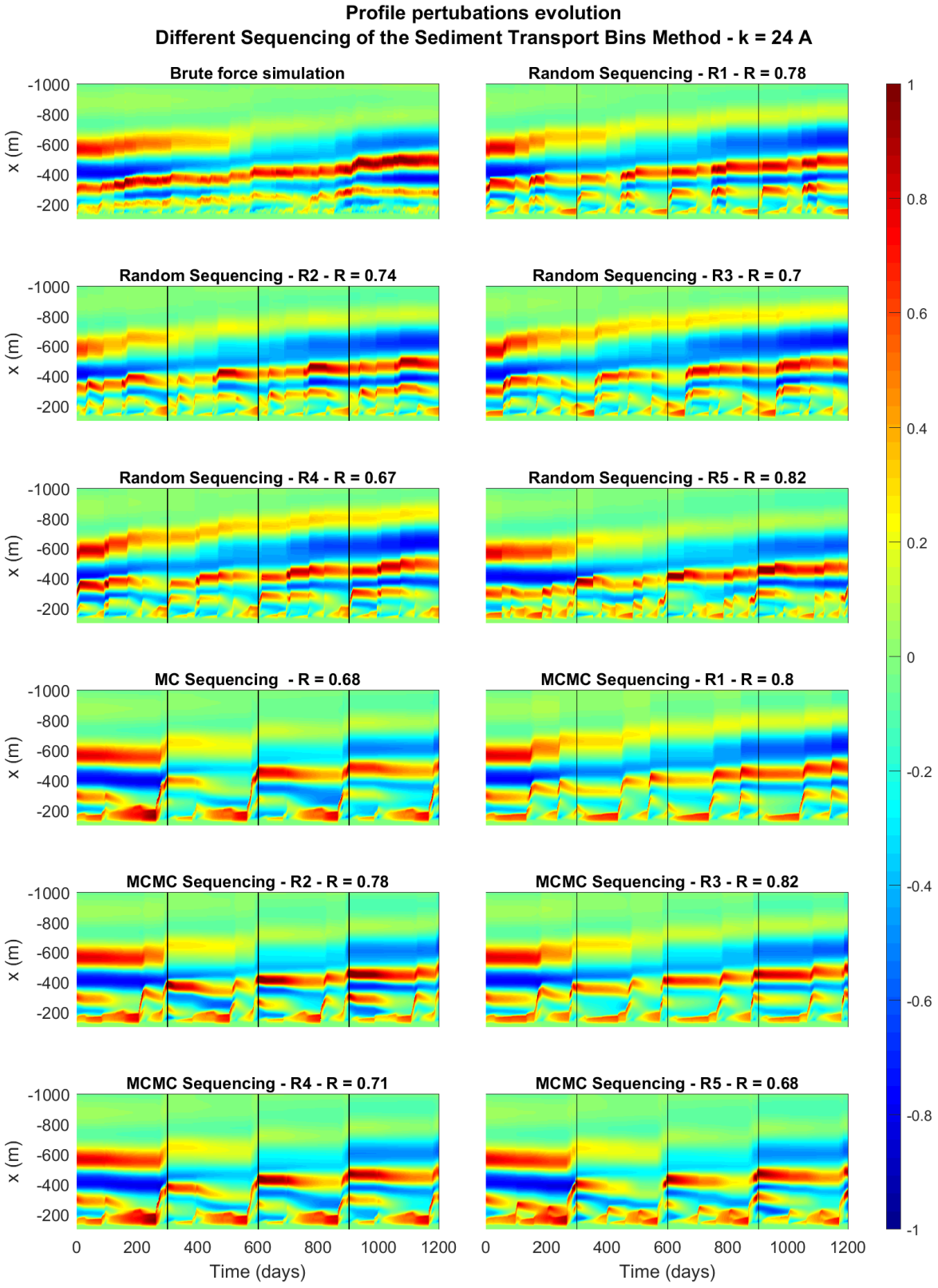


Figure 3.66: Time stacks of profile perturbation of the brute force simulation (top) and reduced wave climates of method 10: Sediment Transport Bins with $k = 24A$ for the random, Markov Chain and Monte Carlo Markov Chain sequencing. Warm colors correspond to bars while cold colors correspond to troughs.

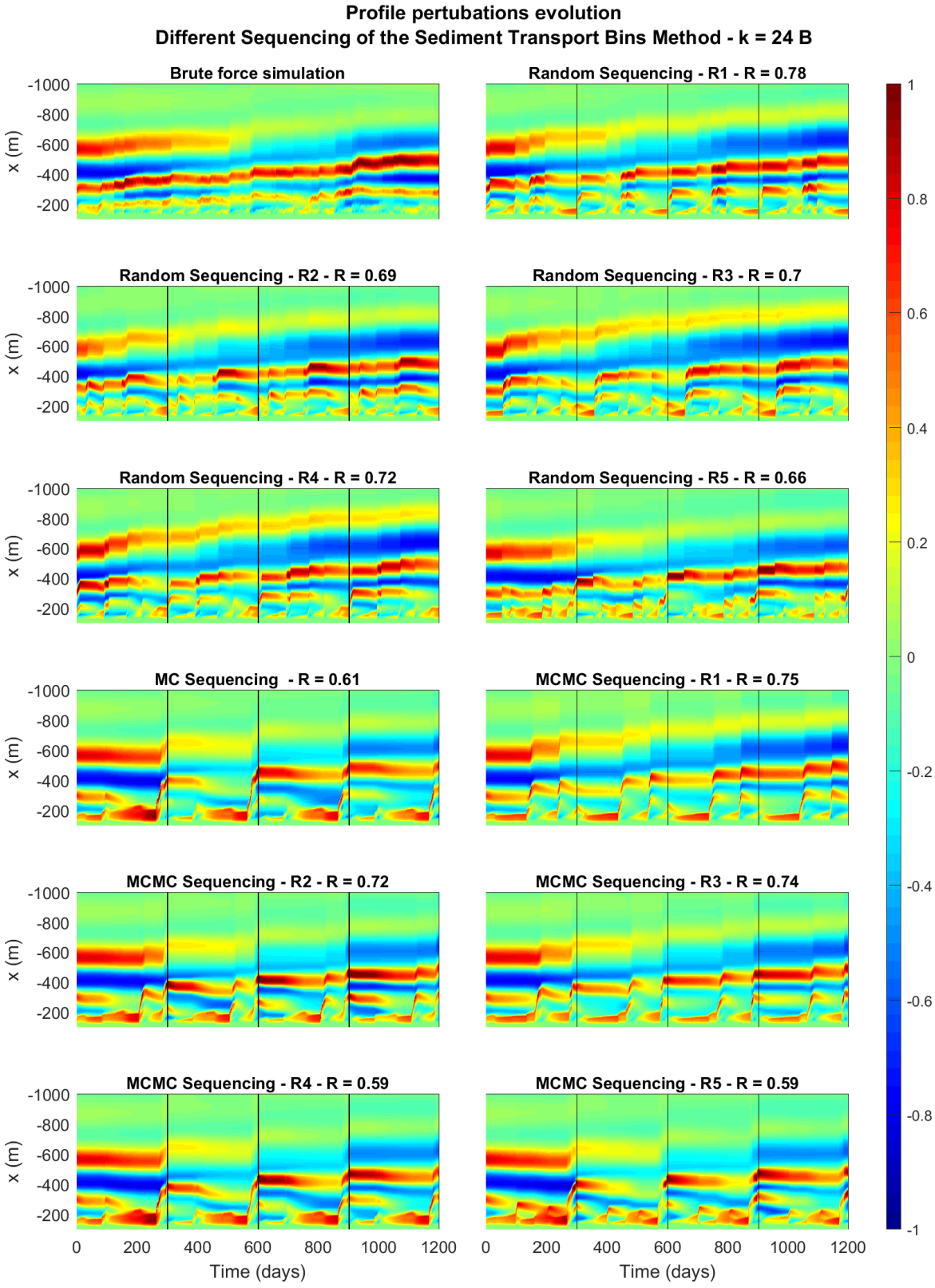


Figure 3.67: Time stacks of profile perturbation of the brute force simulation (top) and reduced wave climates of method 10: Sediment Transport Bins with $k = 24B$ for the random, Markov Chain and Monte Carlo Markov Chain sequencing. Warm colors correspond to bars while cold colors correspond to troughs.

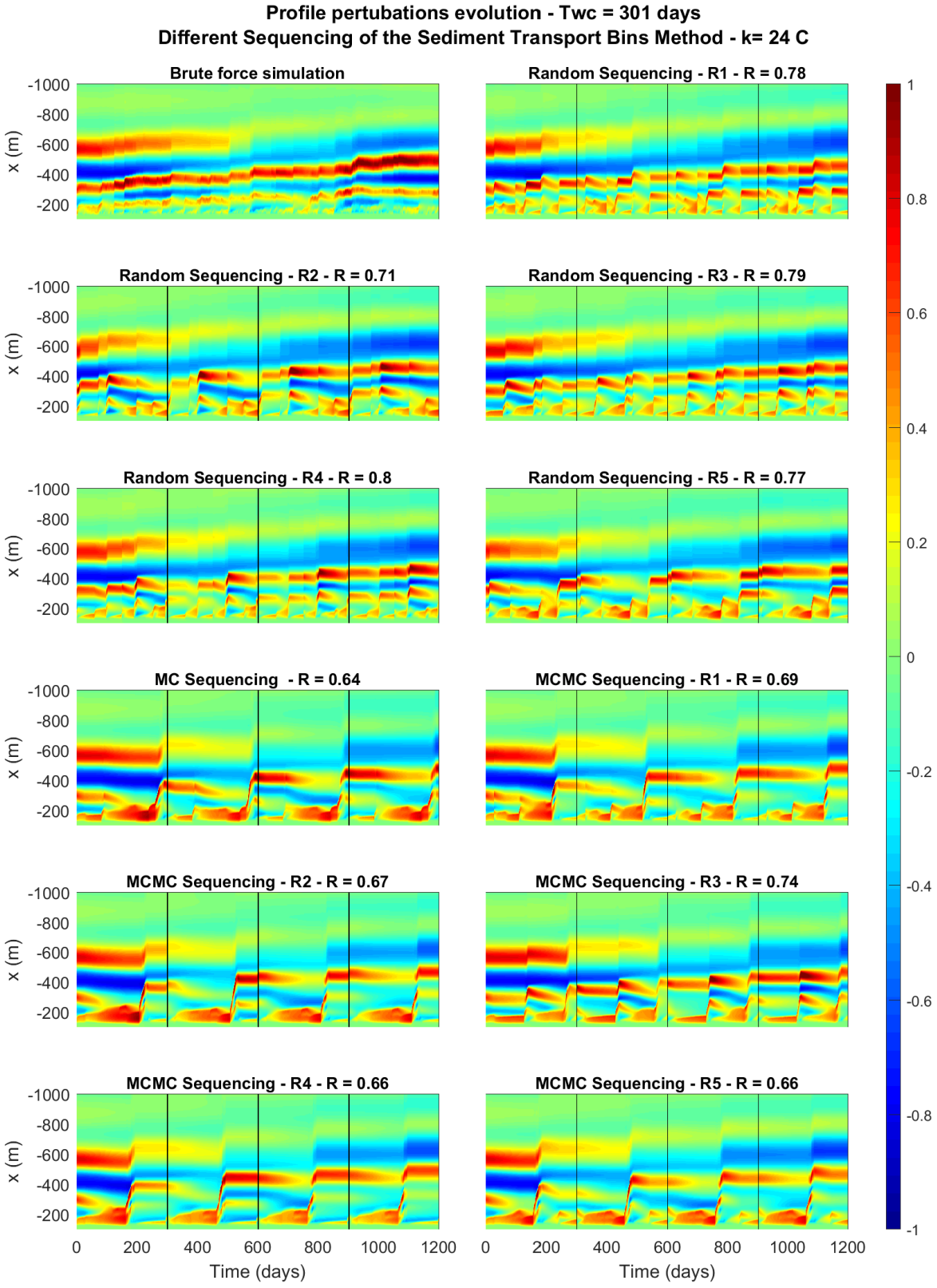


Figure 3.68: Time stacks of profile perturbation of the brute force simulation (top) and reduced wave climates of method 10: Sediment Transport Bins with $k = 24C$ for the random, Markov Chain and Monte Carlo Markov Chain sequencing. Warm colors correspond to bars while cold colors correspond to troughs.

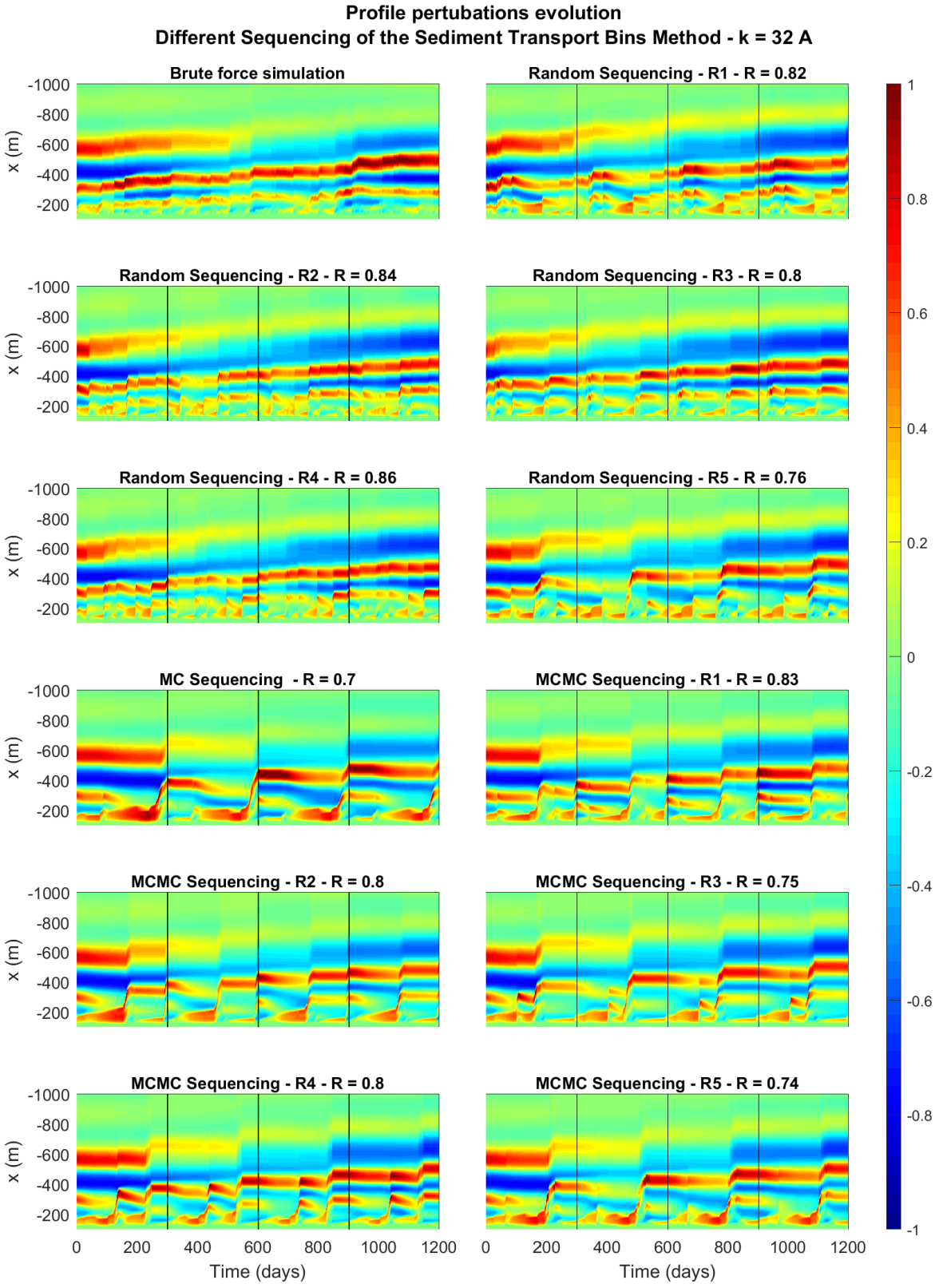


Figure 3.69: Time stacks of profile perturbation of the brute force simulation (top) and reduced wave climates ofmethod 10: Sediment Transport Bins with $k = 32A$ for the random, Markov Chain and Monte Carlo Markov Chain sequencing. Warm colors correspond to bars while cold colors correspond to troughs.

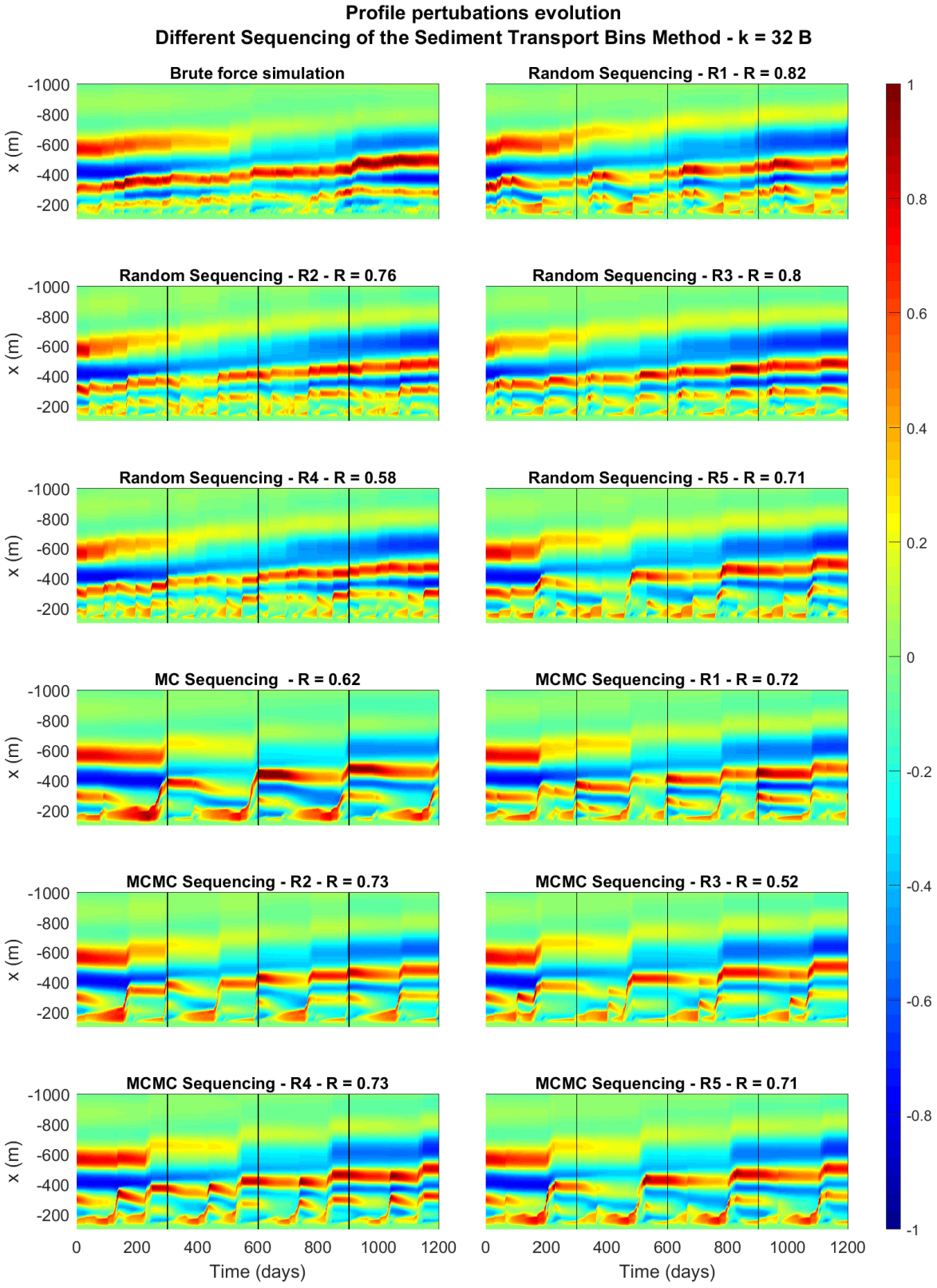


Figure 3.70: Time stacks of profile perturbation of the brute force simulation (top) and reduced wave climates of method 10: Sediment Transport Bins with $k = 32B$ for the random, Markov Chain and Monte Carlo Markov Chain sequencing. Warm colors correspond to bars while cold colors correspond to troughs.

The random sequencing and $k = 32A$ is the best set-up of the sediment transport bins method when applying the reduced wave climate in an yearly time-scale, $T_{wc} = 301$ days ($N_R = 4$). The influence of the duration of the reduced wave climate was further investigated with $k = 12$ and $k = 32A$. Figures 3.71 to 3.73 present the profiles at the end of the simulation and the morphological evolution of method 10: Sediment Transport Bins with wave climate duration of 134 days, random sequencing for $k = 12$ and $k = 32A$, respectively.
When applying the reduce wave climate in a smaller time-scale such as 134 days increasing the number of representative cases does not result in much improvement of morphology. Indeed, $k = 32A$ presents morphological evolution more similar to the full wave climate than $k = 12$. On the other hand, $k = 32A$ with $T_{wc} = 134$ days ($N_R = 9$) has $NoT = 287$, while $k = 12$ with $T_{wc} = 134$ days ($N_R = 9$) has $NoT = 107$ which is considerably less than $k = 32A$, even when $T_{wc} = 301$ days (Table 3.7). Moreover, the morphological evolution and final profile of $k = 12$ with $T_{wc} = 134$ days are better than $k = 32A$ with $T_{wc} = 301$ days (see Figures 3.62 and 3.69), specially on the most shorewards bar. Overall, $k = 12$ with $T_{wc} = 134$ days can be consider as good as $k = 32A$ with $T_{wc} = 134$ days.

Table 3.7: Number of cases, duration of wave climate and number of transitions

Number of cases (k)	**Duration of wave climate (T_{wc})**	**Number of transitions (NoT)**
$k = 12$	$T_{wc} = 301$ ($N_R = 4$)	$NoT = 47$
$k = 32A$	$T_{wc} = 301$ ($N_R = 4$)	$NoT = 127$
$k = 12$	$T_{wc} = 134$ ($N_R = 9$)	$NoT = 107$
$k = 32A$	$T_{wc} = 134$ ($N_R = 9$)	$NoT = 287$

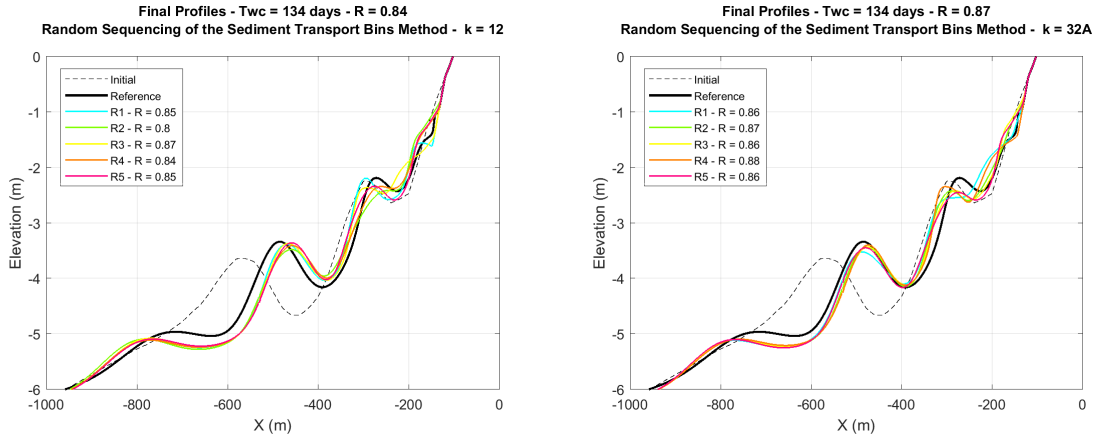


Figure 3.71: Profiles at the end of the simulation with the reduced wave climate of method 10: Sediment Transport Bins with $k = 12$ and $k = 32A$, random sequencing and $T_{wc} = 134$ days. The black lines are the initial (dashed) and final (solid) profiles from the brute force model. The colored lines are the final profiles of each replicate sequence.

Therefore, the best set-up of the Sediment Transport Bins method is $k = 32A$ with random sequencing and $T_{wc} = 134$ days and the optimal set-up is $k = 12$ with random sequencing and $T_{wc} = 134$ days. In Figure 3.74 is presented the comparison between the full wave time-series and the reduced wave time-series generated by the Sediment Transport Bins method, with random sequencing and $T_{wc} = 134$ days. It is possible to observe that even though with certain limitations, the reduced wave time-series represent the variability of the full wave time-series.

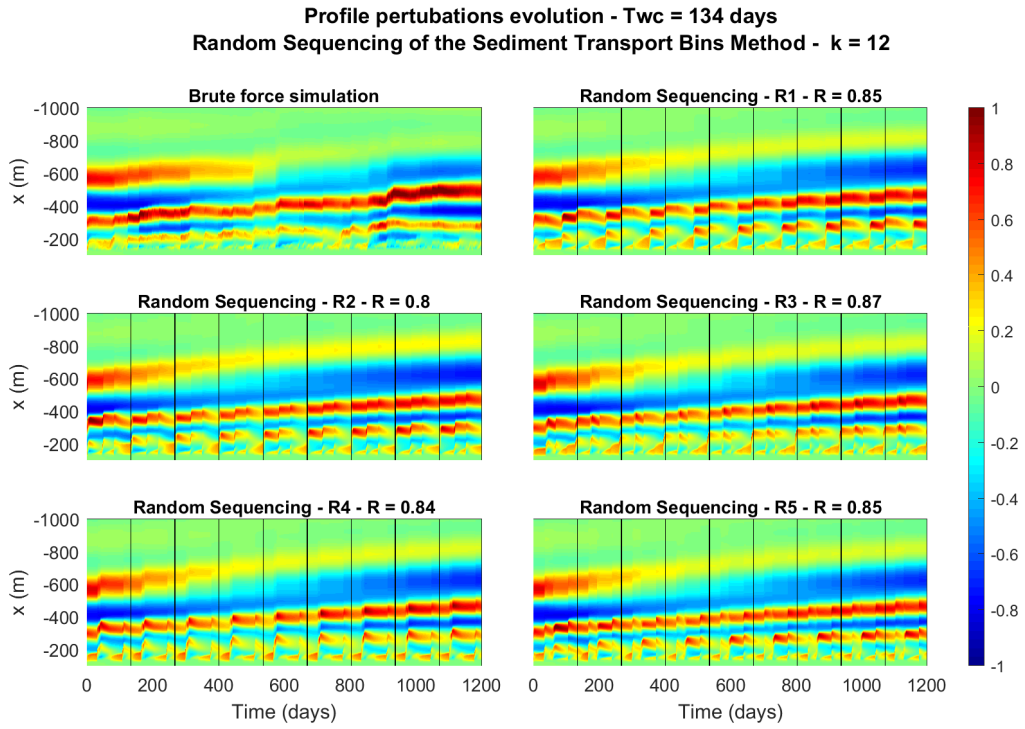


Figure 3.72: Time stacks of profile perturbation of the brute force simulation (top) and reduced wave climates of method 10: Sediment Transport Bins with $k = 12$, random sequencing and $T_{wc} = 134$ days. Warm colors correspond to bars while cold colors correspond to troughs.

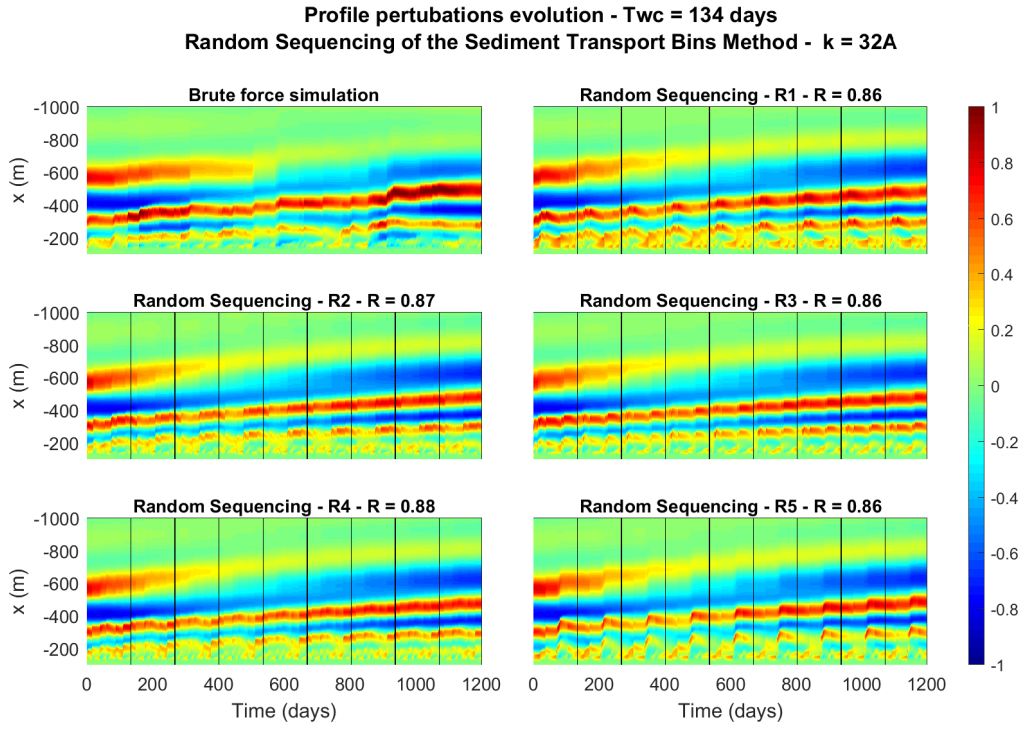


Figure 3.73: Time stacks of profile perturbation of the brute force simulation (top) and reduced wave climates of method 10: Sediment Transport Bins with $k = 32A$, random sequencing and $T_{wc} = 134$ days. Warm colors correspond to bars while cold colors correspond to troughs.

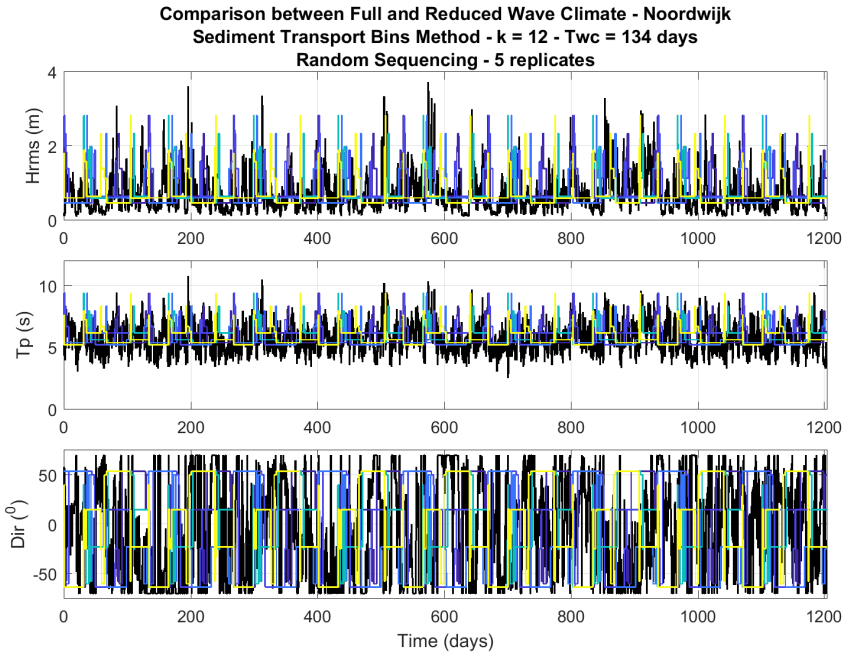


Figure 3.74: Comparison of full wave time-series and (black lines) and reduced wave time-series generated by method 10: Sediment Transport Bins with $k = 12$, random sequencing and $T_{wc} = 134$ days. The colored lines represent the 5 random replicates of the reduced wave climate.

3.16. VALIDATION - APPLICATION AT ANMOK BEACH

The best input reduction method, sequencing and number of cases previously defined are validated in Anmok beach, on the East coast of South Korea (see section 2.7). Thus, the Sediment Transport Bins method was used to perform the reduction in Anmok's wave dataset together with random sequencing and $T_{wc} = 123$ days. The difference on the time-scale of the reduced wave climate is due to rounding of the number of repetition of the wave climate. Nonetheless, the duration of the reduced wave climates of Noordwijk and Anmok are still comparable.

Figures 3.75 to 3.77 present the comparison of the reduced and full wave time-series, profiles at the end of the simulation and morphological evolution, respectively. The comparison of the reduced and full wave time-series looks promising, the reduced wave time-series seems to represent the variability of the full wave time-series. The profile perturbations temporal evolution as well as the final profiles are very similar to the brute force. It is possible to note that R2 and R3 random sequences present small instabilities in the profile.These instabilities occur due to the coarse sediment characteristic of the profile ($D_{50} = 400\mu m$ - see Appendix B) and they can be eliminated with improvement in the calibration of the model. Despite these instabilities, the overall performance of the input reduction in Anmok is not impaired by them. The average skill score is 0.8, with maximum and minimum skill score of 0.83 and 0.76, respectively (see Figure 3.76) and the morphological evolution of the brute force model is represented reasonably accurately by the reduced wave climate. Hence, the input reduction set-up: Sediment transport method, random sequencing, $k = 12$ and $T_{wc} = 123$ days is considered validated when applied in a different location.

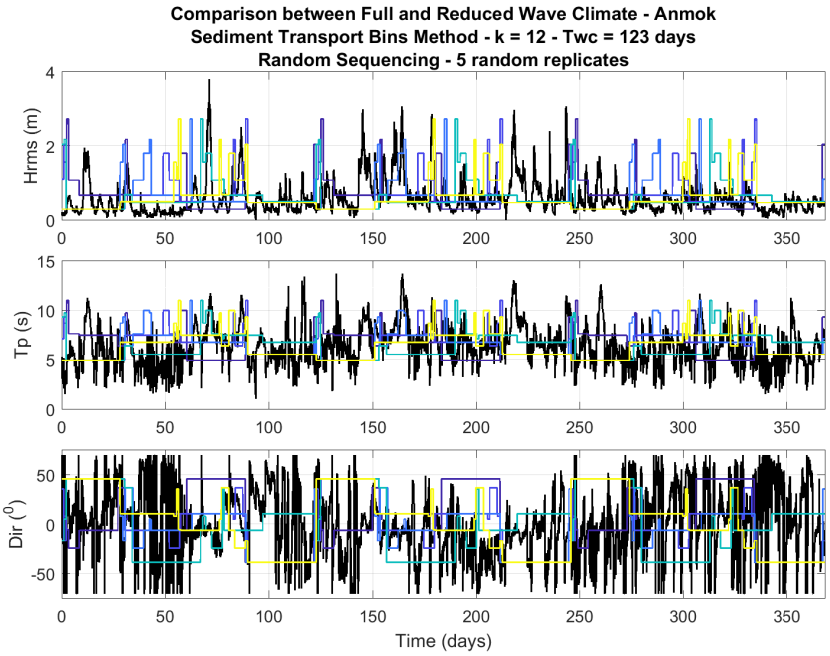


Figure 3.75: Comparison of full wave time-series and (black lines) and reduced wave time-series generated by method 10: Sediment Transport Bins with $k = 12$, random sequencing and $T_{wc} = 123$ days at Anmok beach, East coast fo South Korea. The colored lines represent the 5 random replicates of the reduced wave climate.

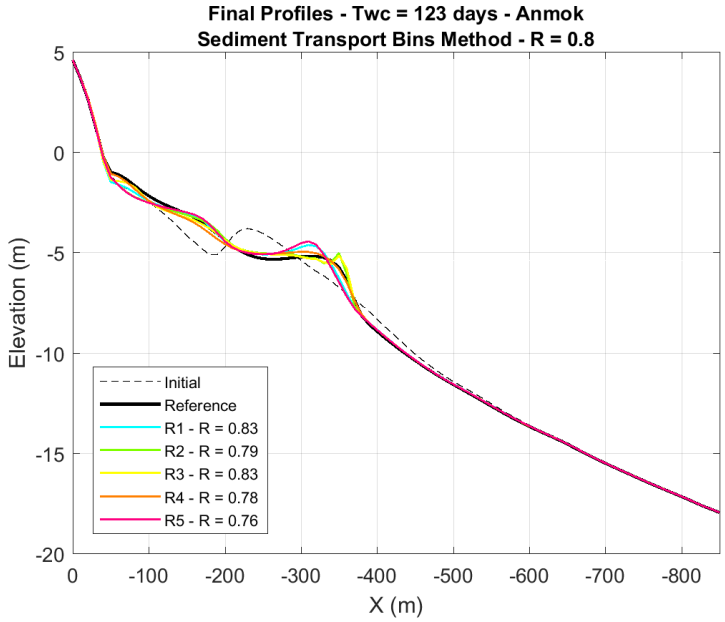


Figure 3.76: Profiles at the end of the simulation with the reduced wave climate of method 10: Sediment Transport Bins with $k = 12$, random sequencing and $T_{wc} = 123$ days at Anmok beach, East coast fo South Korea. The black lines are the initial (dashed) and final (solid) profiles from the brute force model. The colored lines are the final profiles of each replicate sequence.

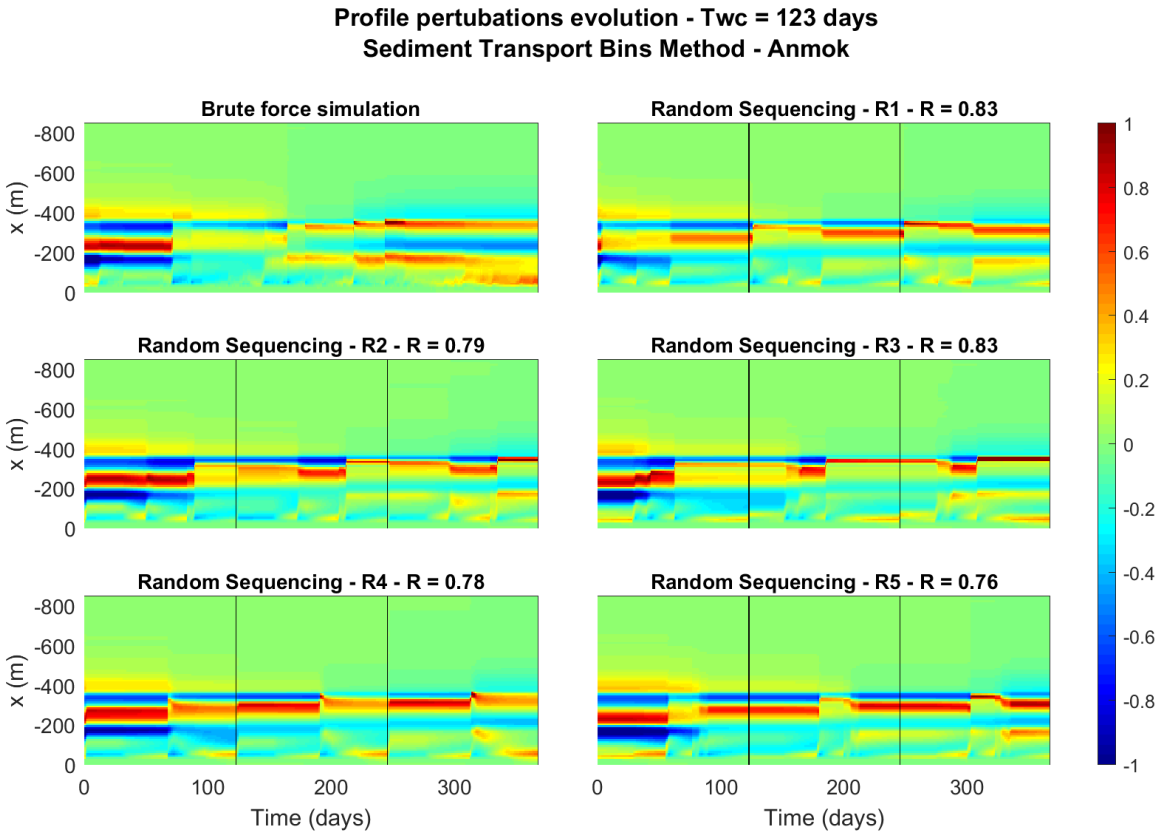


Figure 3.77: Time stacks of profile perturbation of the brute force simulation (top) and reduced wave climates of method 10: Sediment Transport Bins with $k = 12$, random sequencing and $T_{wc} = 123$ days at Anmok beach, East coast fo South Korea. Warm colors correspond to bars while cold colors correspond to troughs.

4

DISCUSSION

In this section a discussion is presented regarding the methods, their advantages and disadvantages, the sequencing and number of cases analysis as well as the validation. Additionally, the limitations and applicability of the present research are discussed as well as its comparison with similar studies in literature.

4.1. INPUT REDUCTION METHODS

Method 1: Conditions with the Largest Transport Contribution is not considered a good method for reproducing morphological changes via input reduction. This method selects wave conditions with the highest sediment transport contribution which leads to an overestimation of the sediment transport in the reduced model and therefore irreversible and not realistic morphological changes.
Method 2: Grouping with Equal Sediment Influence has a good physical background, selecting wave conditions that represent an equal amount of sediment transport contribution is a good way of representing morphological variations. However, the way it performs the clustering allows for superposition of the clusters which misleads the definition of the centroids, therefore resulting in not a good selection of representative wave conditions for morphological evolution.
Method 3: The Representative Wave Approach is not considered a good method because it aggregates the observation through time. The aggregation in time preserves the chronology, however in one section, defined as the seasons of the year, all kind of conditions occur and the resultant representative wave conditions defined as the average of the section are all very similar. In this method, the extreme wave conditions and the directional spreading of the bins are average out.
Method 4: Fixed Bins has the advantage of predefining the bins and in this way guarantees a desirable spreading over direction and wave height. However, the same advantage can also be seen as a disadvantage once an arbitrary definition of the bins is likely to lead for a poorer morphological representation when compared to other methods that make use of physical concepts when dividing the observations into bins such as equal sediment contribution or energy flux.
Method 5: Crisp k-means, method 6: Fuzzy k-means and method 7: K-harmonic means present a high dependency on the observation's frequency of occurrence when defining the centroids. High wave height cases are mostly not selected due to their low frequency of occurrence in the dataset. The Fuzzy k-means method is the one that has the highest dependency on the observation's frequency of occurrence. This is because the method has a soft membership function which means that all observations have some influence on the definition of the centroids. How much each observation will influence on the definition of the centroids is defined by the membership function that gives higher weight for observations closer to the pre-defined centroids. So, for instance when a centroid is defined as a high wave height condition, even though the observations closer to that centroid have higher weight, the excess of lower wave height observations with smaller weight pulls down the centroid. The same occurs with the k-harmonic means, although in a smaller scale due to its dynamic weighting function that gives higher weight to observations far from the centroids. The Crisp k-means method has a hard membership function which means that each observation belongs to only one cluster. The way that this method depends on the observation's frequency of occurrence is by the definition of the clusters that when averaging the wave conditions tends to the most recurrent observations, characteristic of calm wave conditions.

Besides that, the randomness contained in these methods makes the comparison among them difficult once it is almost not possible to reproduce the same reduced wave climate. Using the Maximum dissimilarity and Fixed bins algorithms as cluster initiation solves the issue of inconsistency and it improves the morphological evolution once it points the algorithm in the right direction. Cluster initiation with Fixed bins resulted in slightly better performance than MDA.
Using weighting functions such as H_{rms}^{p} and longshore sediment transport (S_y) could improve the performance of the methods Fixed bins and Crisp k-means. However it does not improve in the present study because when using H_{rms}^{p}, low wave cases were not selected and when using S_y the conversion of sediment transport centroids to wave height presented deficiencies. This was due to the absence of inverse function once the S_y used was obtained from the brute force simulation. Whereas for the Fuzzy k-means and K-harmonic means the weighting functions still do not overcome the dependency on the observation's frequency of occurrence.
Method 8: Maximum Dissimilarity has the disadvantage that only selects outliers observations resulting in a non-realistic morphological evolution. This method indeed selects high wave height conditions, however for its selection account for mild wave conditions the number of centroids (k) should increase considerably (*e.g.* higher than 32). And this is another disadvantage of the method because with the same number of centroids, other methods present better performance than method 8.
Method 9: Energy Flux is considered a good method because it has a pre-definition of bins that guarantees the spreading over direction and wave height. Also it aggregates the observation in equal energy flux bins which is an approximation of the sediment transport similar to the weighting function H_{rms}^{p}. This is a good physical background for the selection of wave cases, however its drawback is that it only accounts for one parameter (wave height) as proxy for sediment transport and allows negative and positive directions to belong in the same bin, causing averaging of opposite wave contributions. The application of this method is easy to be performed and it presented promising morphological results, even though there is still a need for improvement that can come through improving the definition of the directional bins.
Method 10: Sediment Transport Bins is considered the best method analyzed in the present study as it resulted in the best morphological results. As the Energy flux method, it has a pre-definition of bins and a good physical background for the selection of the wave cases, with 2 parameters (wave height and wave direction) being considered on the aggregation of the observations into bins. However, one disadvantage of this method is that its application is not so easily performed because it requires pre-knowledge of sediment transport rates. In the present study, longshore sediment transport were used to represent a predominantly cross-shore process such as bar morphodynamics. This is is explained by the fact that the cross-shore distribution of the longshore transport influences on bar growth or decay during offshore and onshore migration (Walstra et al. 2011). Moreover, the longshore sediment transport rates were obtained from the brute force simulation and this is very unlikely scenario. An alternative to the brute force simulation is to estimate the sediment transport rates from general sediment transport formulas such as CERC, Kamphuis, Bijker and others. Another alternative is to use an approximation for the sediment transport such as the wave power or energy flux that presented promising results in this study.
In contrast with the findings of the present study, Olij (2015) indicates Crisp k-means or K-harmonic means as good methods for morphological representation of Durban coast, South Africa. However, besides that the Durban coast is a rather energetic coast with significant wave height predominantly higher than 1 m, the mentioned study did not investigate the Energy Flux method nor the Sediment Transport Bins method. Benedet et al. (2016) investigated 3 methods, among others, analyzed in the present study: Fixed bins, Energy flux and CERC method which is analogous to the Sediment transport bins method of the present study, but with the sediment transport rates estimated by the CERC formula. The authors indicate the Energy Flux as the best method which partially corroborates with the findings of the present study once the Energy Flux method presented promising results in the Dutch coast. However, it is important to consider the discrepancies among both studies. Besides that the study was performed on the coast of Florida, U.S., Benedet et al. (2016) applied the CERC method without considering the shore-normal angle, therefore, positive and negative transport contributions could cancel themselves. Moreover, the author used alongshore sediment transport as evaluation measure for the input reduction methods while in the present study the focus is on morphology.

4.2. SEQUENCING

In the present study, random, Markov Chain (MC), Monte Carlo Markov Chain (MCMC) and Monte Carlo Markov Chain with repetition (MCMCR) sequencing methods were analyzed. The random, MCMC and MCMCR sequencing contain randomness and therefore were performed with 5 replicates in order to avoid a bias due sequence when analyzing and interpreting the results.

Overall, it was possible to note the influence that sequencing has on the morphological response of reduced wave climate simulations. The results presented considerable variation on the skill score according to the sequencing method as well as among the random replicates.

The random sequencing presents the best morphological response because the reduced time-series ordered randomly presents higher variability than when ordered by other methods. This occurs because the MC, MCMC and MCMCR sequencing make use of the Markov Chain transitions probabilities that are primarily associate with the statistical properties of the full dataset. This implies that the Markov Chain transitions have high probabilities of remaining in the same sea state and low probabilities of changing from calm to energetic wave conditions. Consequently, the reduced time-series ordered by the MC sequencing has the tendency to aggregate low wave height conditions at the beginning of the simulation and high wave height conditions at the end. This occurs because when sequencing the wave conditions, the MC method selects the wave case with highest Markov Chain transition probability conditional to the prior wave condition which means that calm wave conditions that have the highest probabilities are selected first. The MCMC and MCMCR methods attempt to overcome this effect with its randomness. The MCMC sequencing does generate reasonable variability in the reduced time-series while the MCMCR does not once it allows wave conditions to persist. Since the persistence of calm wave conditions has high probabilities, the tendency of the MCMCR sequencing is also to aggregate calm wave conditions at the beginning of the simulation and more energetic wave conditions at the end of the simulation. The MCMCR sequencing presented the worst morphological response.

Even though the MCMC sequencing presents reasonable variability on its reduced time-series, the random sequencing still has higher variability leading to a better morphological response once it resembles the natural variability of wave conditions that occurs in reality. Despite of its good performance, the random sequencing has the drawback that it is completely random and has no user-control which means that good sequences can be generated as well as bad sequences such as aggregation of energetic conditions. Besides that, it is important to note that the application of random sequencing has its limitations since the model is dependent upon the initial condition, *i.e.* initial profile. For instance, a winter profile evolves differently than a summer profile with the same sequence of wave conditions. This implies that the random sequencing for locations where the chronology of wave conditions is important has limitations on its application. However, for Noordwijk the bar dynamics is classified as inter-annual with a gradual offshore migration and not specifically linked to storm events (Walstra et al. 2013), so chronology is not so relevant and random sequences can be applied, aware of its limitations.

In the Durban coast, South Africa, the Markov Chain sequencing is recommended even though its sequence is similar to ordering the wave conditions from the smallest wave height to the largest wave height, namely Low to High sequencing (LtH) (Olij 2015). However, in the mentioned study the random sequencing was not investigated. Moreover, Walstra et al. (2013) found that randomly ordered synthetic time-series performs better than systematic sequencing of wave conditions such as ascending/descending wave height and wave angles towards positive/negative directions.

4.3. NUMBER OF CASES

Different combinations of directional and wave height bins were analyzed in the present study. The results demonstrated that increasing the number of directional bins improves the morphological performance of the reduce wave climate more than increasing the number of wave height bins. However, when decreasing the number of wave height bins to 2 ($k = 8$), it resulted in notably worse performance. Thus, 3 and 4 are acceptable and practical quantities of wave height bins.

The difference on the effect of the resolution of the wave height and directional bins is caused by the distinct effects that these parameters have on sediment transport. Although the relation between wave height and sediment transport is non-linear, the sediment transport is always proportional to the wave height in a power that can vary among 2-3. On the other hand, the wave direction influences the sediment transport in a sinusoidal manner in which a determined wave angle (*e.g.* 45° to the shore-normal in an idealized straight coastline) causes a maximum sediment transport and angles bellow and higher than the threshold generate lower sediment transport rates. Therefore, the resolution of direction bins must account for this behavior in

order to obtain a reasonable morphological representation.
The best morphological results were with $k = 32A$ which has 8 directional bins and 4 wave height bins. However, 32 wave conditions can be consider a high number of transitions ($NoT = 127$) and not feasible for application in more complex models. Moreover, $k = 10$ (2 directional bins) presented the worse results due to its lack resolution of directional bins.
For that reason, the amount of directional bins should be higher or equal to 4 and the choice upon the number of representative cases should balance the quality of morphology representation and the number of transitions that is associated with computational time.
Benedet et al. (2016) analyzed 5 input reduction methods with $k = 30$, $k = 20$, $k = 12$ and $k = 6$. With $k = 6$, all methods performed poorly whereas with higher number of representative wave conditions, better the performance of the methods. Furthermore, the authors indicate 12 as an optimal quantity to represent annual wave climate. Additionally, Olij (2015) concluded that doubling the representative wave conditions from 10 to 20 does mostly not lead to better performance, hence much more conditions are necessary to improve the input reduction performance.

4.4. WAVE CLIMATE DURATION

The wave climate duration turned out to be a very important aspect of reducing wave climate. It was observed that a less robust wave climate with 12 representative wave cases applied in a smaller time-scale such as 134 days presented a morphological evolution more similar to the full wave climate than a more robust wave climate with 32 wave conditions applied in a yearly time-scale (301 days). This points out that not only the robustness of the reduced wave climate is an important aspect but also the time-scale in which the reduced wave climate is applied. This is explained by the fact that morphology is not only dependent on the forcing (*e.g.* waves), but also on the duration of the forcing. The same wave condition applied in different time-scales will likely give rise to distinct morphology. Therefore, the choice upon the duration of the reduced wave climate should balance resemblance of the natural variability by applying the reduced wave climate in smaller time-scales and computational effort by reducing the number of transitions thus, applying the reduced wave climate in bigger time-scales.
Walstra et al. (2013) also point out the relevance of the duration of the forcing on reducing wave climate. The authors found reliable reduced wave climate predictions for Noordwijk given $T_{wc} < 301$ days. However, it is important to note that there is a lower limit of T_{wc} because the variability of the reduce wave climate should not be higher than the full wave climate. In other words, the wave conditions cannot last too long that would generate irreversible morphological changes but also not too short that morphology does not have enough time to adjust to the hydrodynamic conditions. The lower limit of T_{wc} can be estimated by the randomized time-series approach (Southgate 1995) where random time-series are generated by splitting the full time-series into a number of segments of constant lengths and randomly reordering these segments. In this way, all the conditions of the full wave climate are simulated and T_{wc} is comparable to the segment length. The lower limit of T_{wc} is based on the shortest segment length for which acceptable predictions are found and the lowest frequency of occurrence of the reduced wave climate. For Noordwijk, Walstra et al. (2013) found that the lower limit of T_{wc} should be about 10 days.

4.5. VALIDATION

The validation of the reduced wave climate was performed with the Sediment Transport Bins method, 5 random sequences and wave climate duration of 123 days in Anmok beach, South Korea. Despite small instabilities in few random sequences, the reduce wave climate reproduced overall the morphological evolution of the full wave climate. However, Anmok beach has considerable alongshore variation and only one profile does not well represent the whole area. The Dutch coast is relatively alongshore uniform and so for Noordwijk, the assumption that one profile is representative of the area of interest is reasonable. However, for other locations such as Anmok beach this might not be valid. Therefore, the one dimension assumption of the present study is a limitation of the methodology and a next step on the investigation of the input reduction implications on morphology should be analyzing two dimensional domains. Still, the findings of the present study are applicable as general guidelines for input reduction.

5

CONCLUSIONS

The main objective of the present study is to understand the influence of reduced wave climate on simulated morphological evolution. In this section, the conclusions obtained from this study are presented bellow in answers to the research questions.

1. Which input reduction method presents the best performance in terms of morphology?
   One of the findings of the present study is that having more control on the definition of the bins results in a better selection of representative wave conditions for morphology. Input reduction methods that are based on statistical properties of the dataset tend to define the centroids towards the most frequent observations which is a fair representation of the dataset, but not necessarily good for morphology. For instance, these methods do not select high wave height conditions that have an important role on morphological evolution. Thereby, the Sediment Transport Bins method is the best input reduction method once it has a pre-definition of bins and it uses a good physical background (equal sediment transport) when defining the bins. Yet, in order to apply this method pre-knowledge of sediment transport rates is required. This information can be obtained from brute force simulations as well as estimated by sediment transport formulas (*e.g.* CERC, Kamphius, Bijker, etc.) or by approximations (proxies) such as wave power or energy flux. One example of sediment transport proxy is the Energy Flux method. This method also comprises pre-definition of bins and good physical background when defining the bins (equal energy flux). Nonetheless, one drawback of the Energy flux method is that it only accounts for the influence of wave height on the definition of bins whereas the Sediment Transport Bins method accounts for both wave height and wave direction. Another disadvantage of the Energy Flux method is that it allows bins to cross the shore-normal angle resulting in averaging out of opposite wave contributions. Despite that, the Energy Flux method demonstrated consistent morphological evolution and can be seen as the second best input reduction method, after Sediment Transport Bins. The wave power as sediment transport proxy was not investigated in the present study, however it is expected to perform as good as the energy flux or even better once it accounts for wave height and wave direction.

2. What is the influence of sequencing of the representative wave conditions on the performance of the input reduction methods?
   Sequencing has a lot of influence on the performance of the input reduction methods. Some evidence of this are the considerable variations in skill scores and morphological evolution among the random replicates. Moreover, when applying the Monte Carlo Markov Chain with repetition sequencing (MCMCR), the skill scores were substantially lower and the morphological evolution represented poorly the full wave climate.
   Regarding sequencing, one of the findings of the present study is that the order of the representative wave conditions should resemble the natural variability of the full wave climate. Among the sequencing methods analyzed in the present study, random sequencing was the one that performed the best. This occurs because the other sequencing methods analyzed made use of statistical properties of the full wave dataset through Markov Chain transitions. This resulted in loss of variability once the highest probabilities are to remain in the same or similar states. The random sequencing is not influenced by the statistics of the full wave dataset and that is why it generates reduced wave time-series with more variability. Despite that, one drawback of random sequencing is that it still can aggregate energetic

wave conditions through its randomly selection. Therefore, random sequencing should be applied but with verification on the resultant reduced time-series.

3. What is the influence of the number of wave conditions on the performance of the input reduction methods?
   The number of wave conditions is associated with the robustness of the reduced wave climate. Although a wave climate with more classes is better represented, the number of bins has to increase considerably (*e.g.* from 12 to 32) in order to cause improvement on the morphological evolution of the reduced wave climate. Additionally, the increase of the directional bins resolution results in a better representation of the reduced wave climate for morphology than the wave height bins. This occurs due to the different relation that wave height and wave direction have with sediment transport. The wave height determines the sediment transport exponentially while the wave direction sinusoidally in which the maximum is smaller than 90° to the shore-normal.
   Nevertheless, since the number of representative wave conditions is directly associated with computational time (higher the amount of wave conditions, higher the computational time required), the choice upon the number of representative wave conditions must balance computation effort and quality of morphological prediction. Additionally, the performance of the number of representative wave conditions is influenced by the reduced wave climate duration. One of the findings of the present study is that a less robust reduced wave climate (*e.g* $k = 12$) applied in a smaller time-scale (*e.g* $T_{wc} = 134$ days) presents morphological evolution more representative of the full wave climate than a more robust wave climate (*e.g* $k = 32$) applied in longer time-scales (*e.g* $T_{wc} = 301$ days). Therefore, the time span in which the reduce wave climate is applied plays an important role on its performance. In general, the duration of the reduced wave climate should be long enough so morphology can adjust to the hydrodynamic conditions but not too long that could generate irreversible morphological changes.

4. How does the best input reduction method, sequencing and number of wave conditions performs in a validation case?
   The validation of the optimal input reduction set-up (Sediment Transport Bins with $k = 12$, $T_{wc} = 123$ days and random sequencing) was effectuated with a yearly wave dataset in Anmok, South Korea. Although with few small instabilities, the reduce wave climate reproduced overall the morphological evolution of the full wave climate. Therefore, the findings of the present study on the Dutch coast are considered validated for a different case. However, they should be interpreted rather as general input reduction guidelines given that the methodology of the present study has limitations such as one-dimensional domain.

6

RECOMMENDATIONS

In this chapter two types of recommendations are provided. First, suggestions on the implementation of the input reduction methodology analyzed in this study are given followed by considerations on how to improve the methodology applied in this study. Finally, Figure 6.1 presents a diagram with the guidelines for input reduction application.

6.1. APPLICATION OF INPUT REDUCTION METHODOLOGY

1. **Use Sediment Transport Bins method or Energy Flux method** - Grouping the wave observations in equal sediment transport bins is the best method of reducing wave climate for morphology. In case sediment transport rates are not available, estimating them with simple sediment transport formulas such as CERC or using proxies of sediment transport such as energy flux is an alternative way to best represent wave climate for morphology.

2. **Use random sequencing, but with verification** - Random sequencing is still the best way of imposing variability in the reduced wave climate, however aggregation of energetic wave conditions can occur and therefore verification of the randomly sequenced reduced wave time-series prior to implementation is recommended.

3. **Use at least 12 representative wave conditions and preferably duration of the reduced wave climate of approximately 100-150 days** - The morphological evolution from a reduced wave climate not only depends on the robustness of the wave climate but also on the duration of the forcing. A less robust reduced wave climate applied in a smaller time-scale is more appropriate for morphology rather than a more robust wave climate applied in a yearly time-scale.

4. **Be aware that input reduction always generate errors** - Input reduction can be interpreted as a schematization of past wave events to predict future scenarios which is a considerable conceptual assumption. Moreover, reducing the wave climate implies in such simplification of the forcing that is not possible to represent entirely the full wave climate morphological evolution. Therefore, when using reduced wave climate morphodynamic models as assessment tools it is important to be critical when interpreting the model outputs. Also, it is recommended to perform relative assessment, *i.e.* comparing scenarios, instead of using the outcomes of the model as absolute values.

6.2. IMPROVEMENT OF THE INPUT REDUCTION METHODOLOGY

1. **Implement the Energy Flux method with directional bins starting from the shore-normal angle** - One of the drawbacks of the Energy Flux method is that it allows directional bins to cross the shore-normal angle and this implies in aggregation of opposite wave contributions. Modifying the definition of the directional bins to not cross the shore-normal angle can improve the performance of the Energy Flux method.

2. **Investigate other proxies for sediment transport** - In the present study the Energy Flux method is pointed out as an alternative to the Sediment Transport Bins method because in fact the energy flux is

a proxy for sediment transport. However, the energy flux does not account for the influence of wave angle. Therefore, in order to improve the methodology of the present study it is recommended to investigate other proxies for sediment transport such as wave power or even rough estimates of sediment transport formulas such as CERC.

3. **Implement Markov Chain sequencing with conditional restrictions** - The Markov Chain sequencing orders the wave cases according to the highest Markov Chain transition probability given a prior sea state. This method tends to aggregate calm conditions at the beginning of the simulation and energetic conditions at the end of the simulation. Yet, it does make use of the statistical properties of the full wave dataset which is a more consistent way of sequencing the reduced wave climate rather than randomly. The Markov Chain sequencing could be improved by imposing conditional restrictions such as after a certain time interval of calm conditions, an energetic wave condition must be selected, still based on the Markov Chain transition probabilities.

4. **Use bins or clusters to assign centroids to observations when using Markov Chain sequencings methods** - In the present study all the Markov Chain sequencing methods initialized by assigning a centroid to every observation by the smallest distance to it. However, this allows to centroids being assigned to observations that are not from the class which the centroid represents. A more consistent way of initializing the Markov Chain sequencing methods is to assign to the observations of a bin or cluster the referent centroid.

5. **Furhter investigate morphological validation** - The validation of the findings on input reduction methodology at Noordwijk, Netherlands was performed in the present study in Anmok beach, South Korea by the comparison of the morphological results of a brute force model (Unibest-TC) and a reduced wave climate model. However few small instabilities were observed in the reduced wave climate results due to the coarse sediment applied in the model. Therefore, further investigations on theses instabilities as well as improvement of the brute force model calibration are recommended.

6. **Validate the proposed input reduction methodology with 2D or 3D domains and other models** - In this study, a brute force simulation with a simple 1D model is used as reference for bar morphology. In fact, since there is a brute force model available, the input reduction in not necessary. However, the application of the methodology here investigated would not be possible in rather complex models that would require unfeasible long run times. Therefore, a next step of the analysis of the input reduction framework would be to apply the findings of the present study in more complex cases such as 2D or 3D domains and with different models, for instance Delft3D.

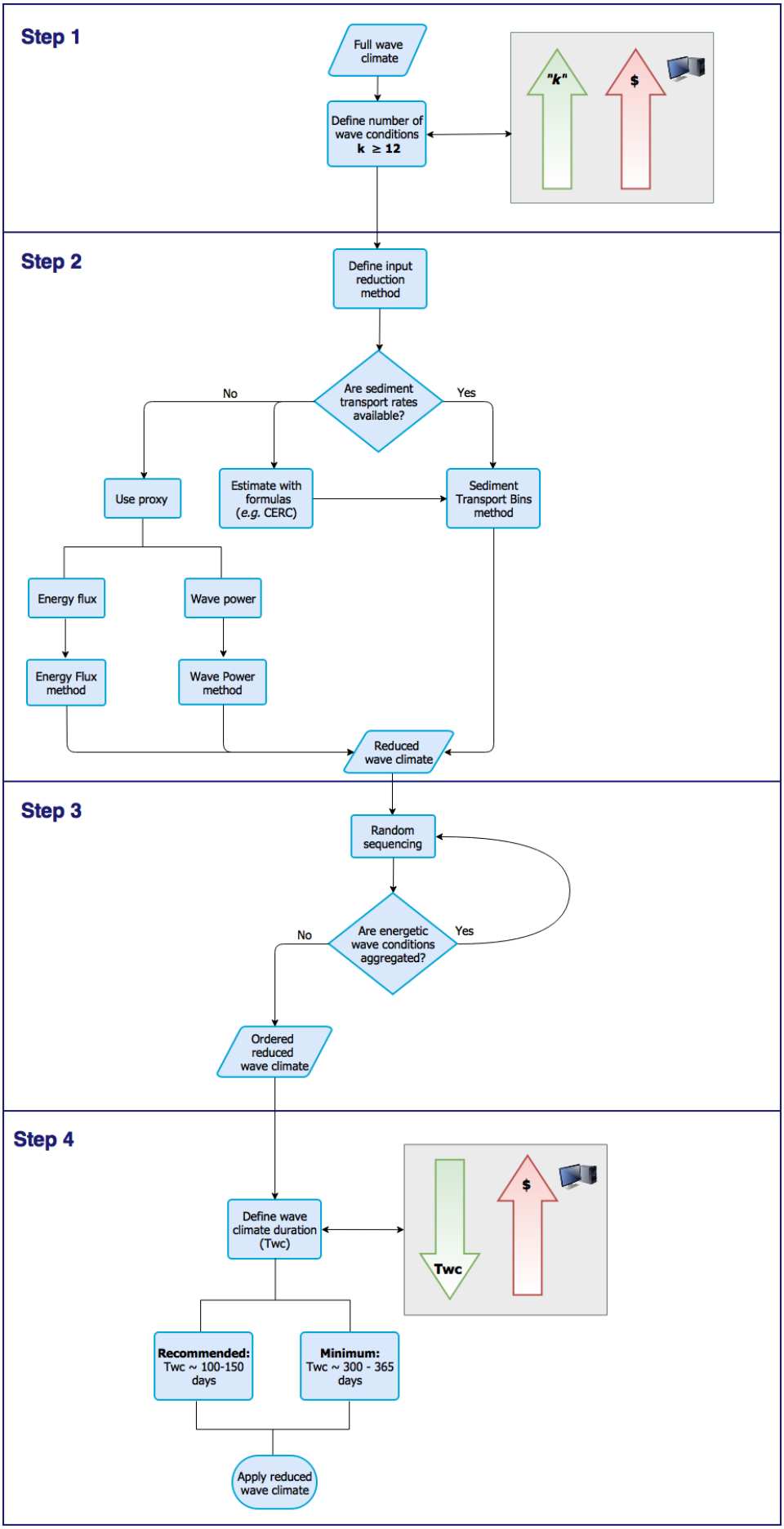


Figure 6.1: Diagram of input reduction guidelines

A

SENSITIVITY ANALYSIS

A sensitivity analysis of the parameters ϵ_{min} (minimal improvement of the centroids) and o (fuzzy parameter) of the Crisp k-means (Method 5), Fuzzy k-means (Method 6) and K-harmonic means (Method 7) methods was carried out in order to define their best set-up. The values of ϵ_{min} analyzed were $\epsilon_{min} = 10^{-5}$ and $\epsilon_{min} = 10^{-10}$. The analysis was carried out with all the input variables possible: H_{rms}, T_p, θ; $H_rms^{2.5}$, T_p, θ and S_y, T_p, θ and cluster initiation: Fixed bins once it allows the comparison between different set-ups and its computation is faster than MDA. Figures A.1 to A.18 present the selection of wave cases for the Crisp k-means, Fuzzy k-means and K-harmonic means with $\epsilon_{min} = 10^{-5}$ and $\epsilon_{min} = 10^{-10}$, respectively. Overall, decreasing the minimal improvement of the centroids (ϵ_{min}) did not produce relevant changes on the selection of the final centroids, thus the minimal improvement of the centroids adopted was $\epsilon_{min} = 10^{-5}$.

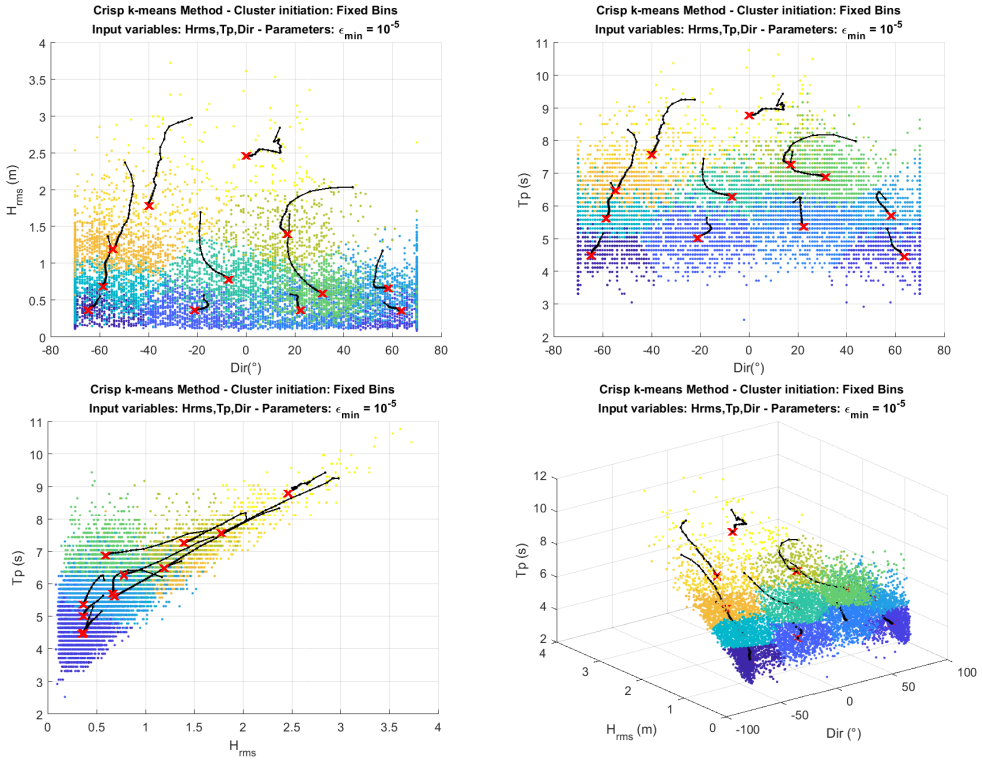


Figure A.1: Selection of method 5: Crisp k-means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

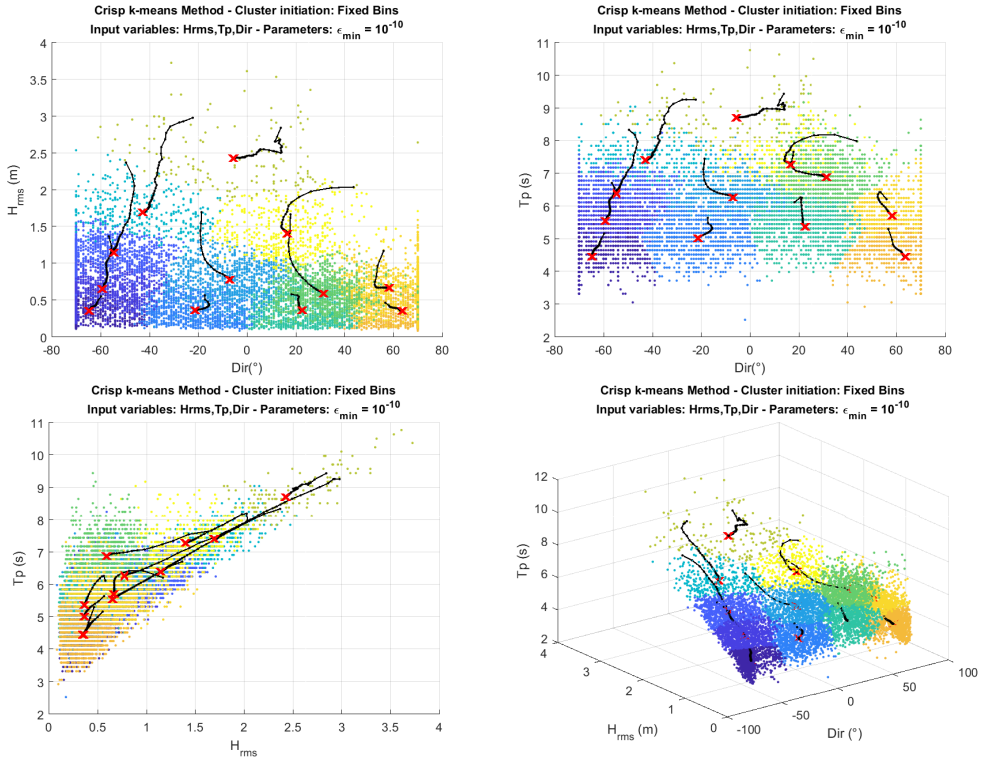


Figure A.2: Selection of method 5: Crisp k-means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

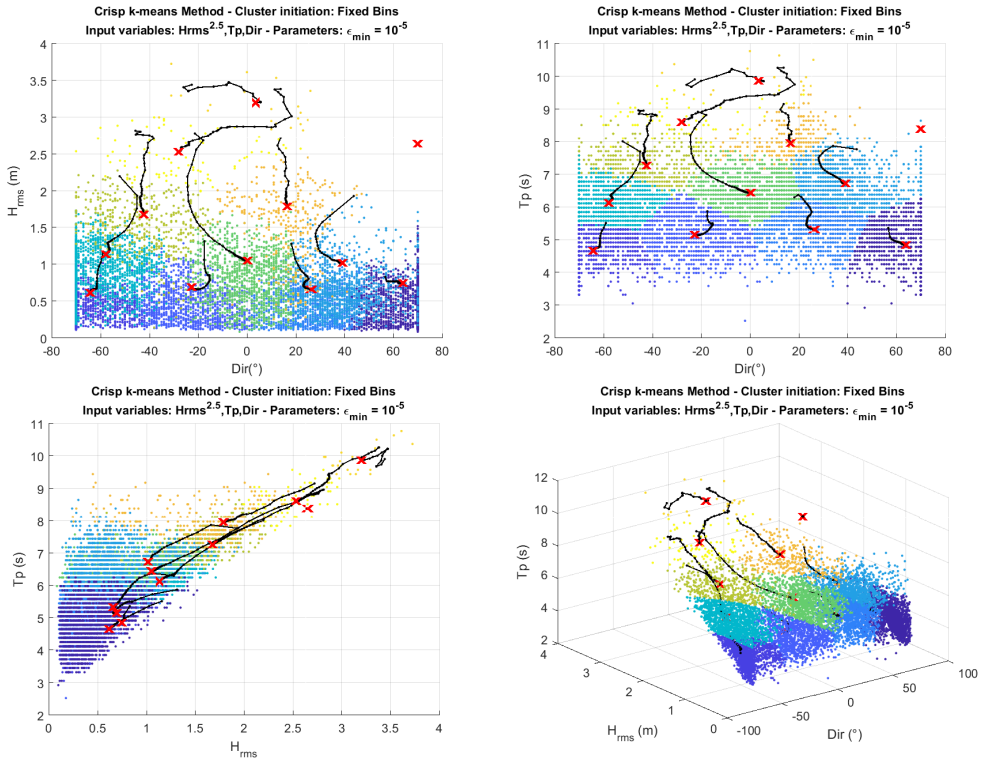


Figure A.3: Selection of method 5: Crisp k-means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

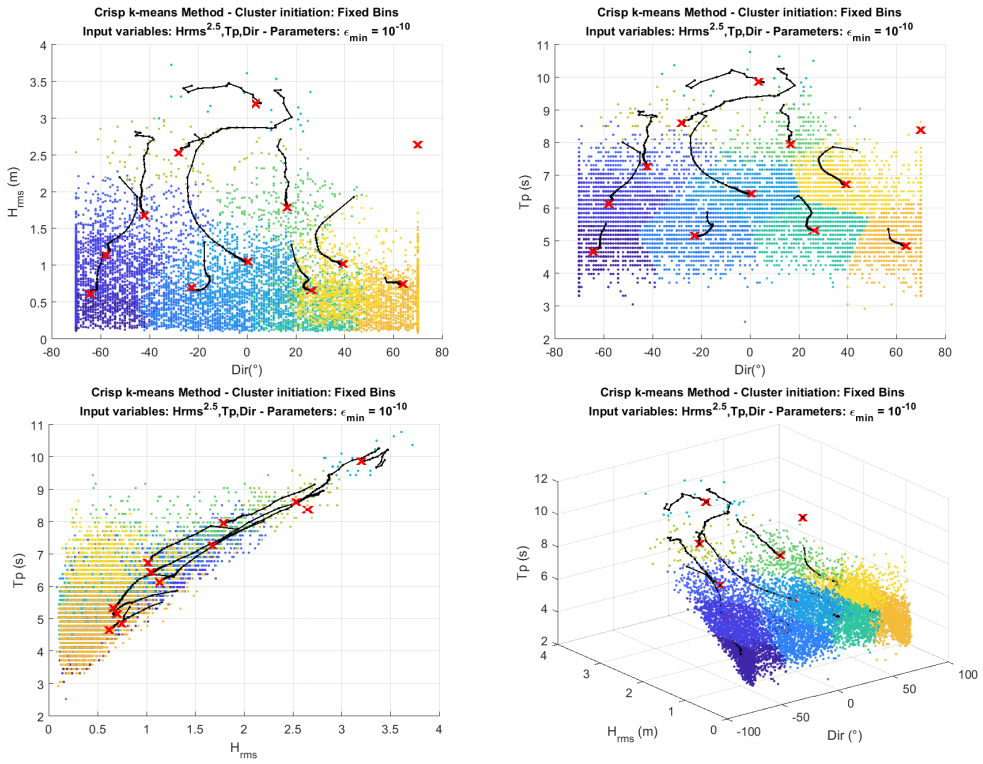


Figure A.4: Selection of method 5: Crisp k-means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

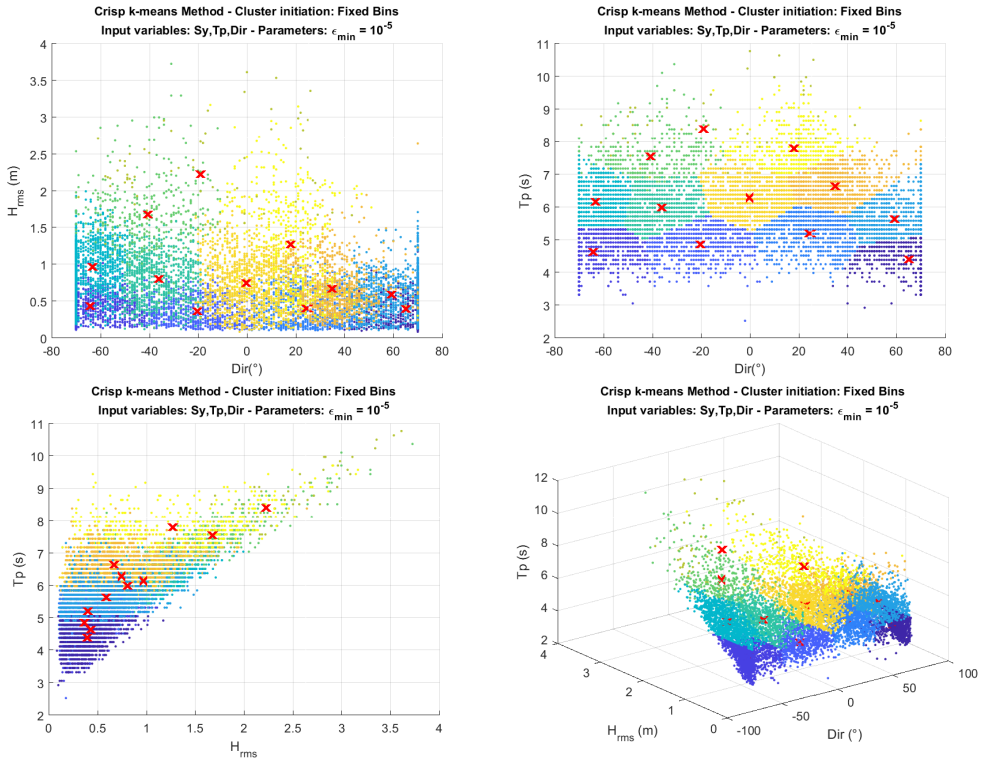


Figure A.5: Selection of method 5: Crisp k-means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

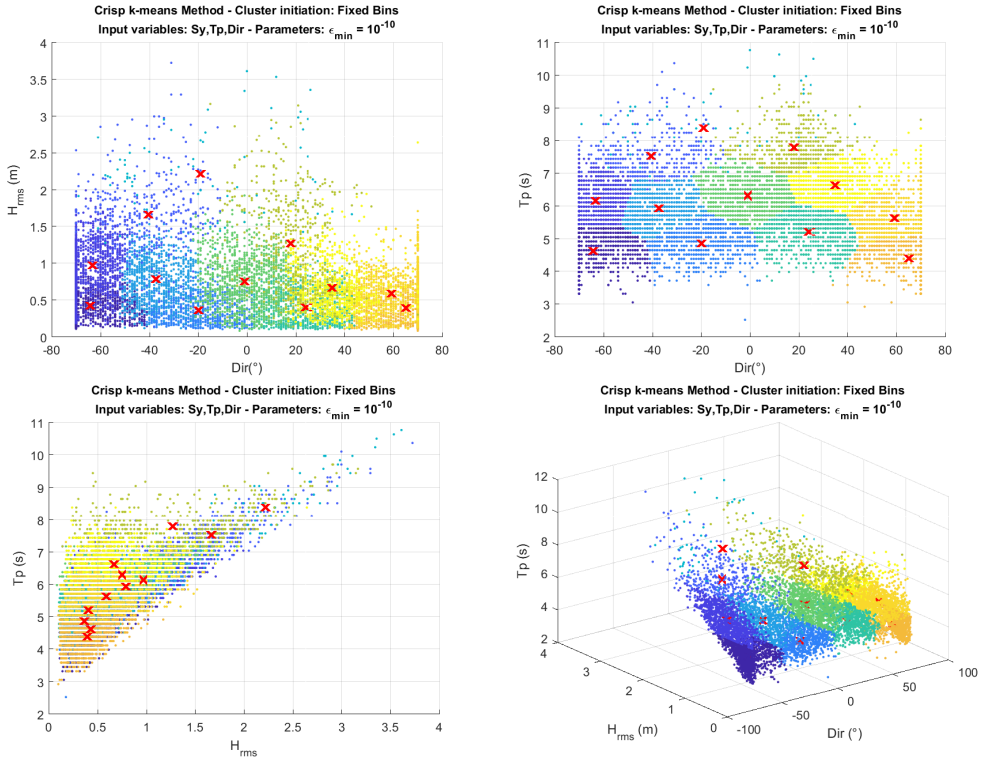


Figure A.6: Selection of method 5: Crisp k-means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

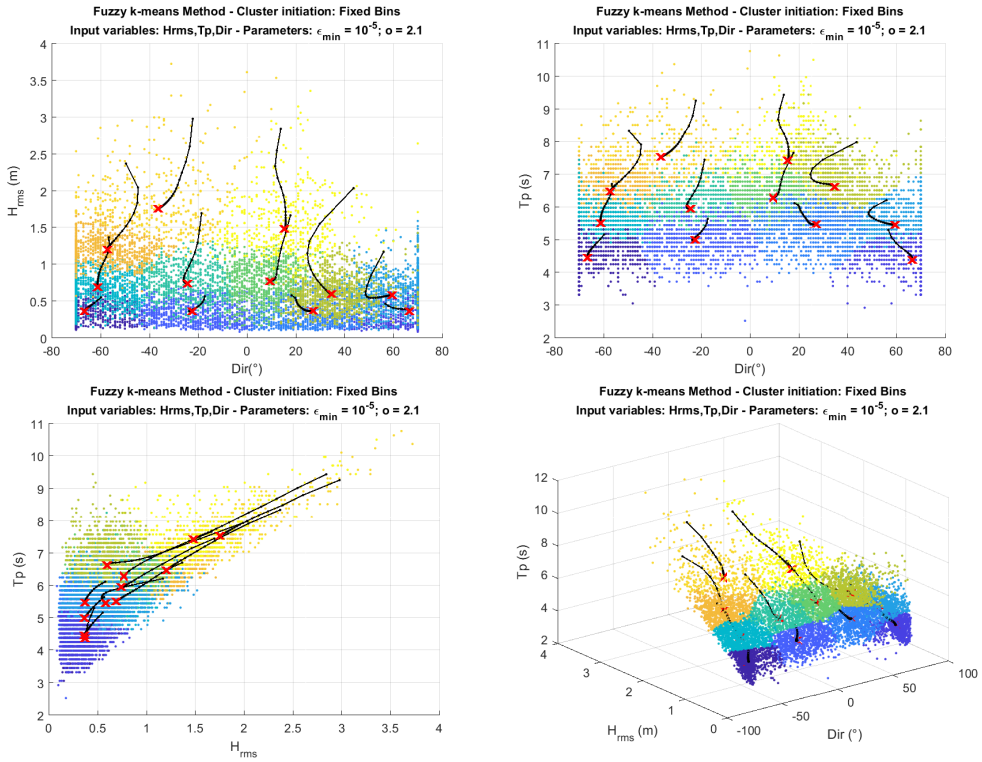


Figure A.7: Selection of method 6: Fuzzy k-means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

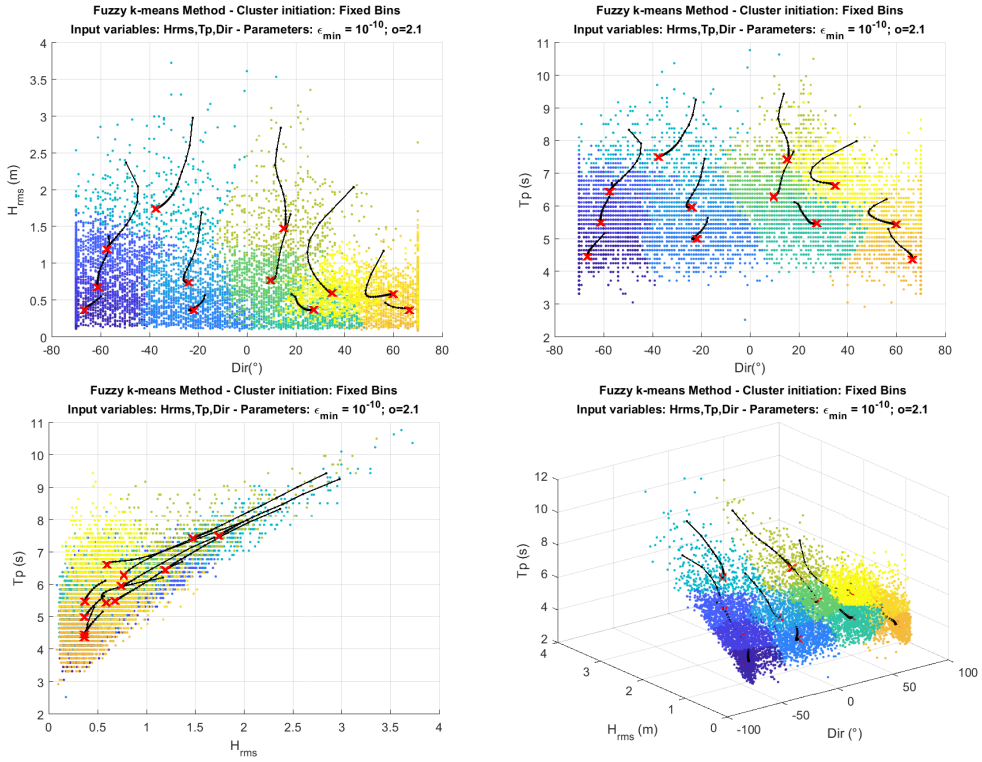


Figure A.8: Selection of method 6: Fuzzy k-means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

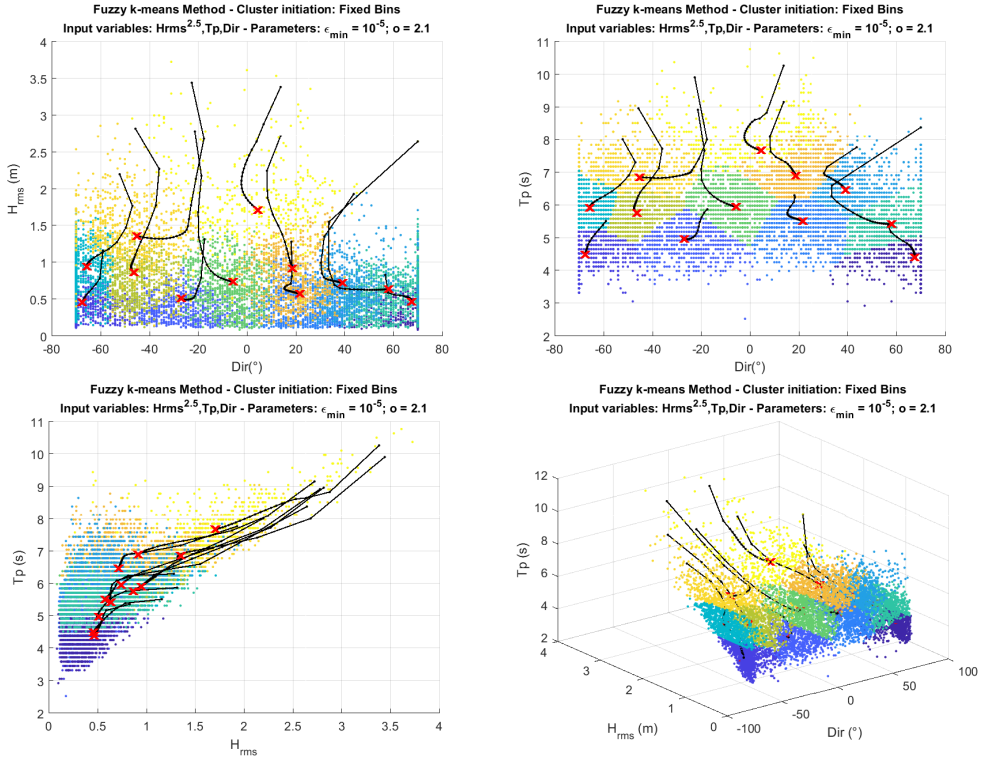


Figure A.9: Selection of method 6: Fuzzy k-means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

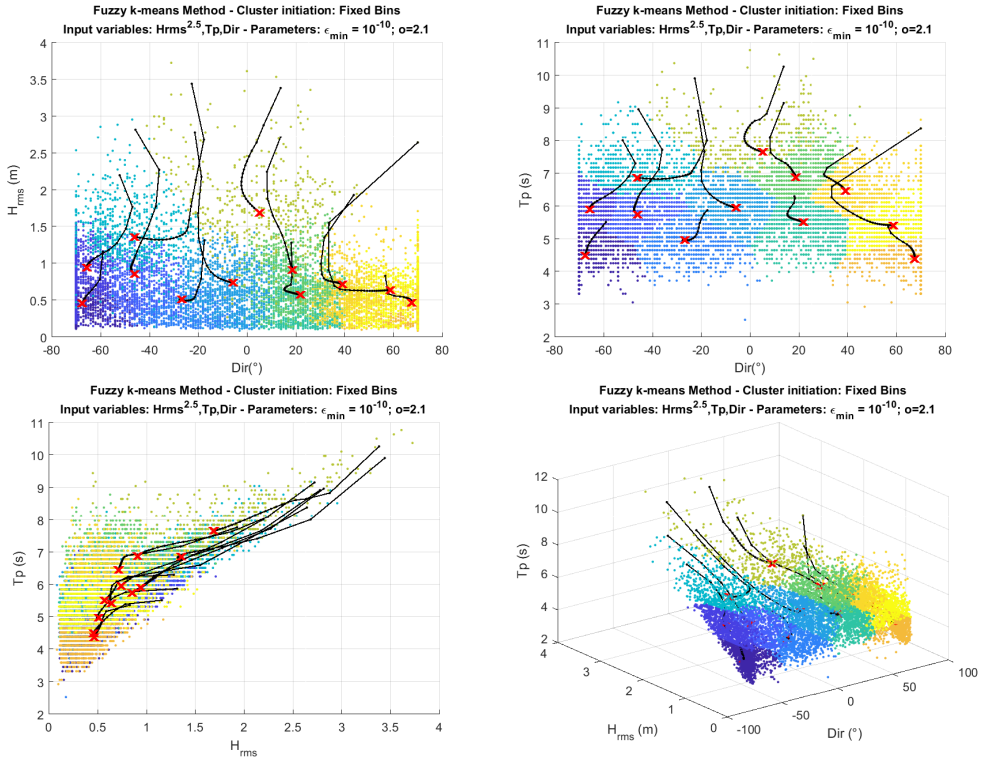


Figure A.10: Selection of method 6: Fuzzy k-means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

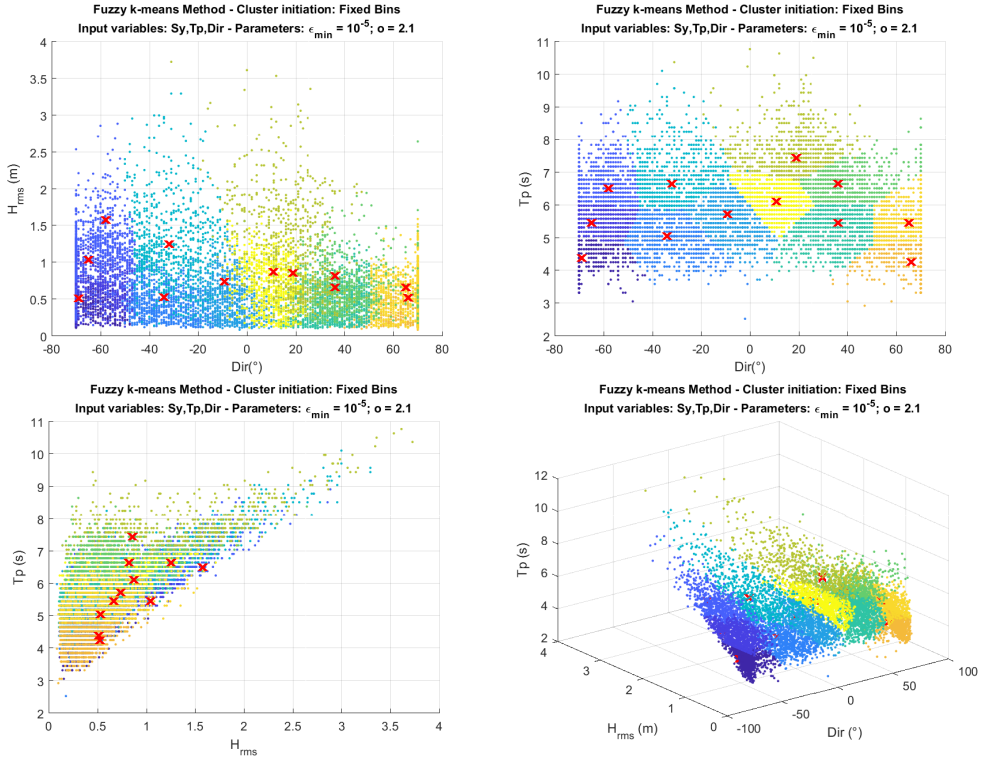


Figure A.11: Selection of method 6: Fuzzy k-means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

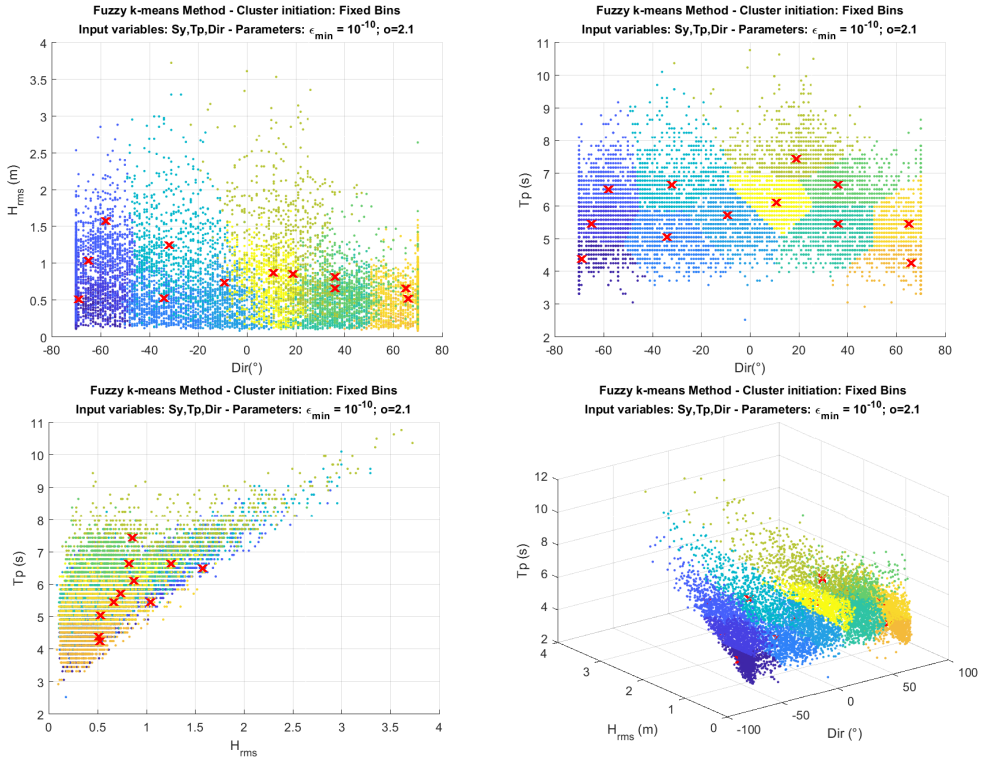


Figure A.12: Selection of method 6: Fuzzy k-means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

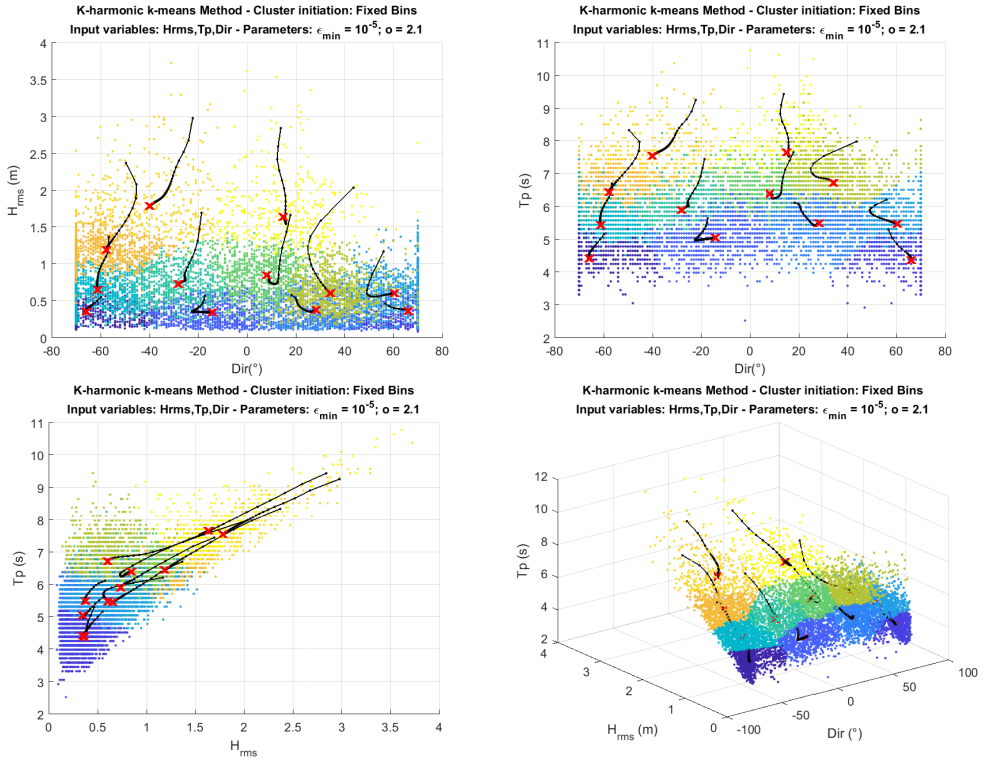


Figure A.13: Selection of method 7: K-harmonic means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

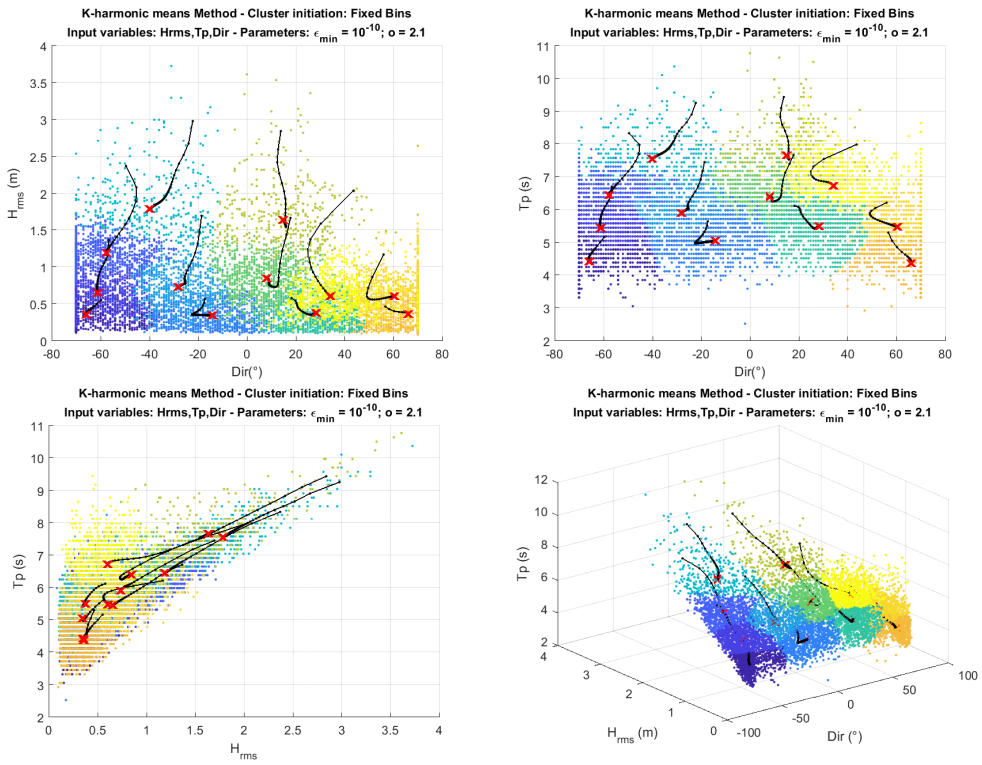


Figure A.14: Selection of method 7: K-harmonic means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

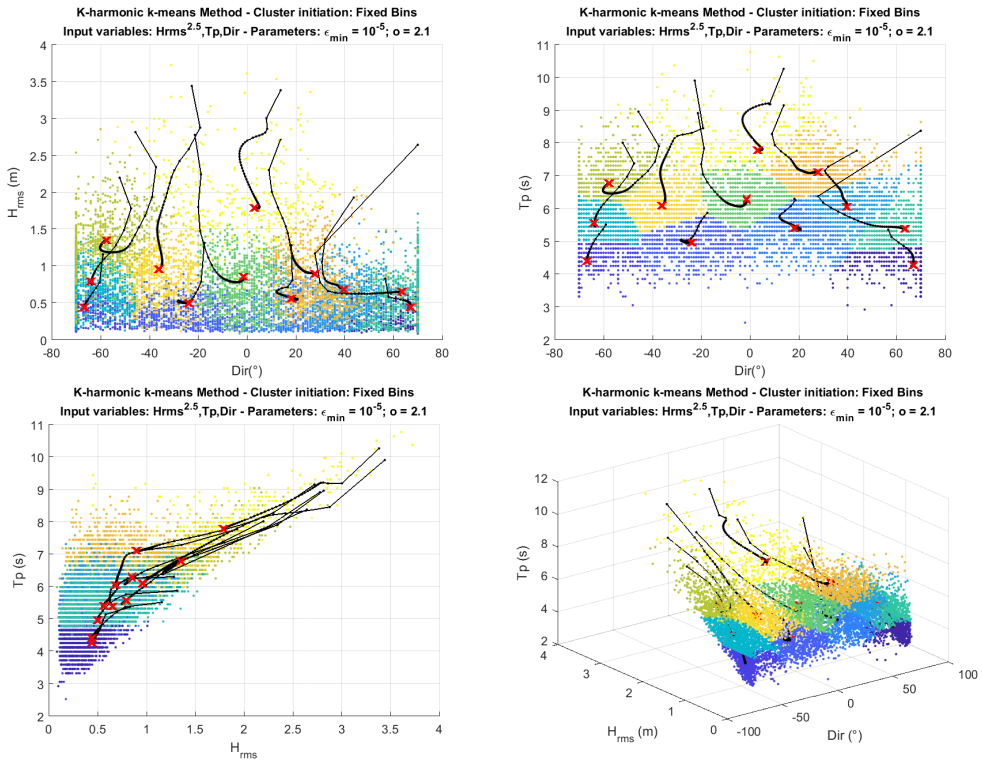


Figure A.15: Selection of method 7: K-harmonic means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

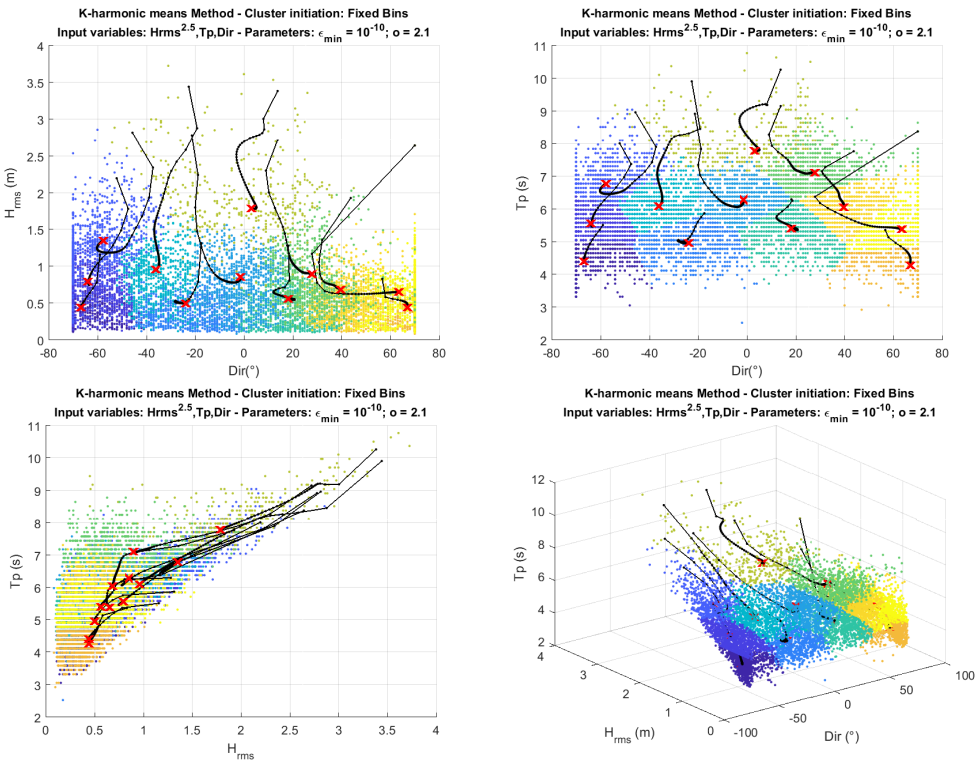


Figure A.16: Selection of method 7: K-harmonic means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

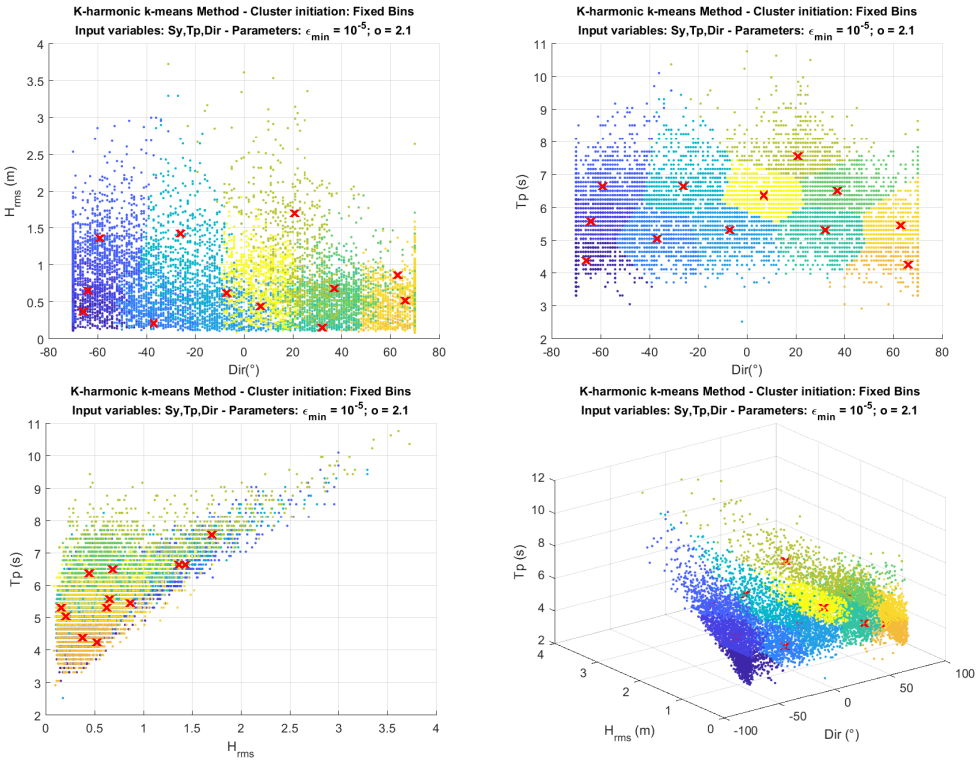


Figure A.17: Selection of method 7: K-harmonic means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

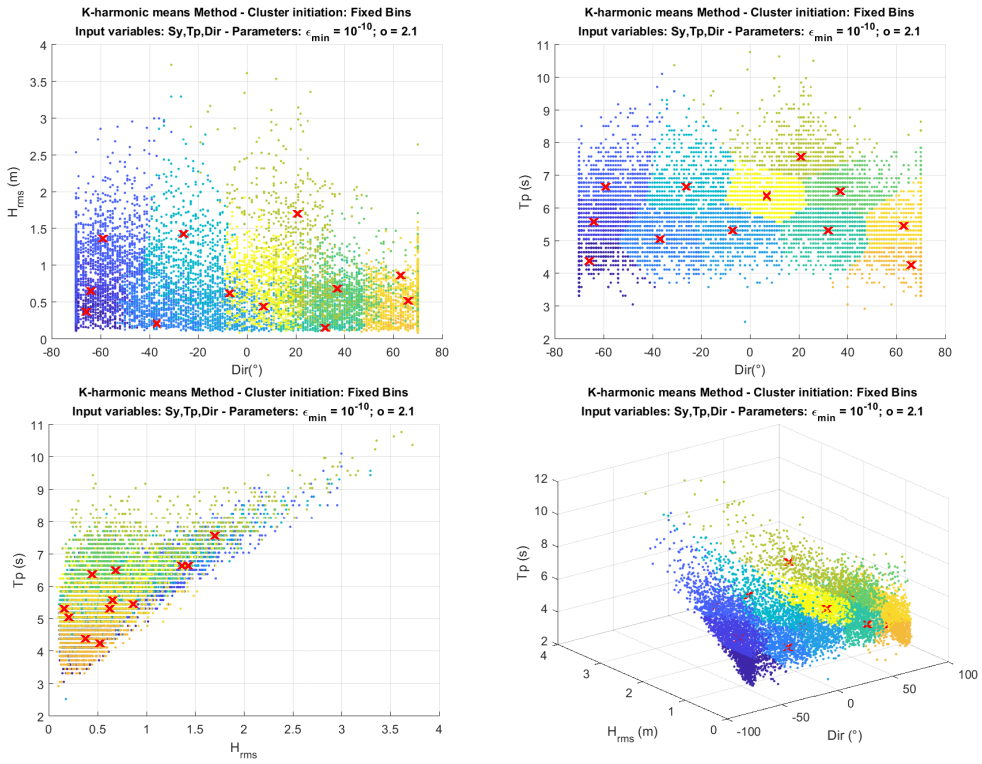


Figure A.18: Selection of method 7: K-harmonic means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

The fuzzy parameter o applies only to the Fuzzy k-means and the K-harmonic means methods. Moreover, the parameter influences in a different manner each method. For the Fuzzy k-means, $o \geq 2$ to perform as expected of membership function in which smaller distances have higher membership function. Figure A.19 shows the different behavior of the membership function according to different values of o.

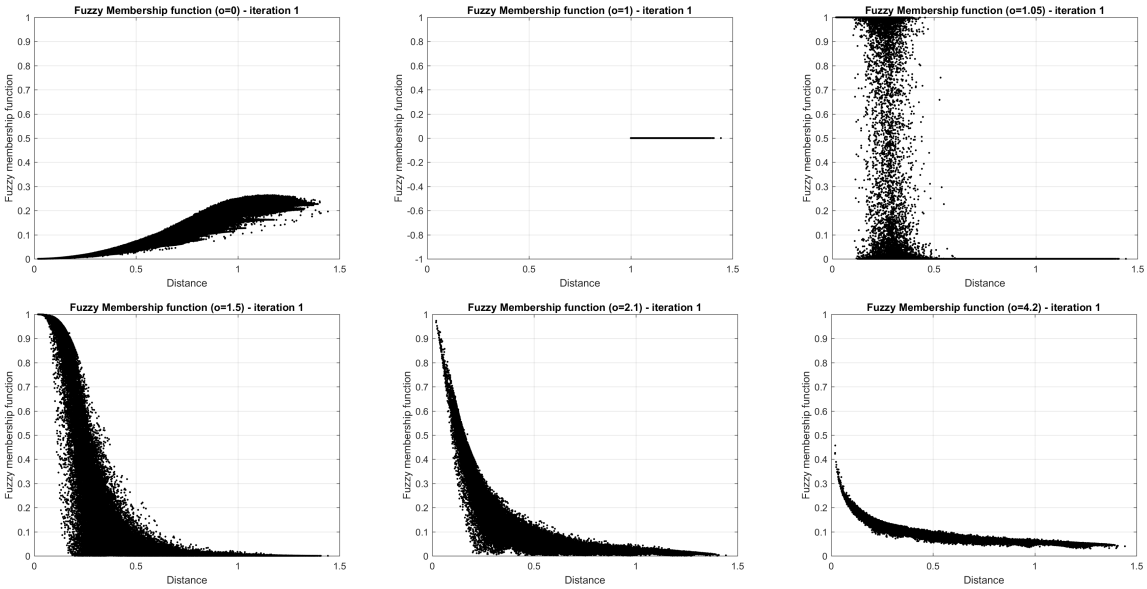


Figure A.19: Behavior of the fuzzy membership function of the Fuzzy K-means method according to different values of o. From left to right: $o = 0$, $o = 1$, $o = 1.05$, $o = 1.5$, $o = 2.1$ and $o = 4.2$.

The value $o = 2.1$ was adopted in Olij (2015) and it is used in the present study as a starting point for the

sensitivity analysis. Figures A.20 to A.28 present the selection of centroids of the Fuzzy k-means method with different input variables and cluster initiation: Fixed bins for the parameters $o = 1.5$, $o = 2.1$ and $o = 4.2$, respectively.

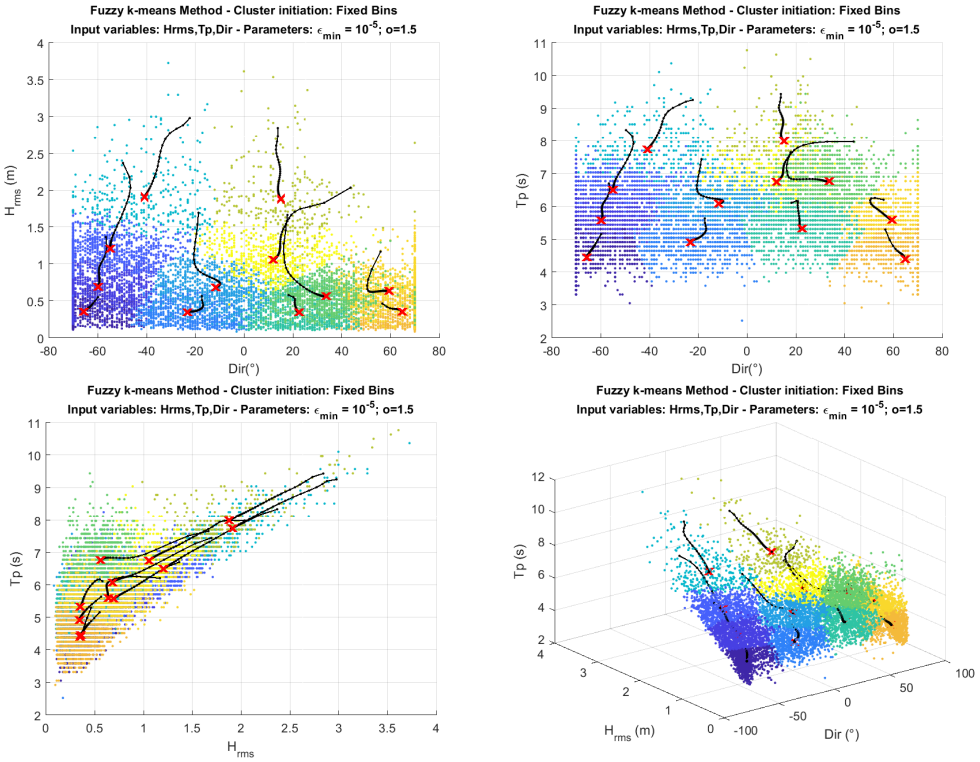


Figure A.20: Selection of method 6: Fuzzy k-means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

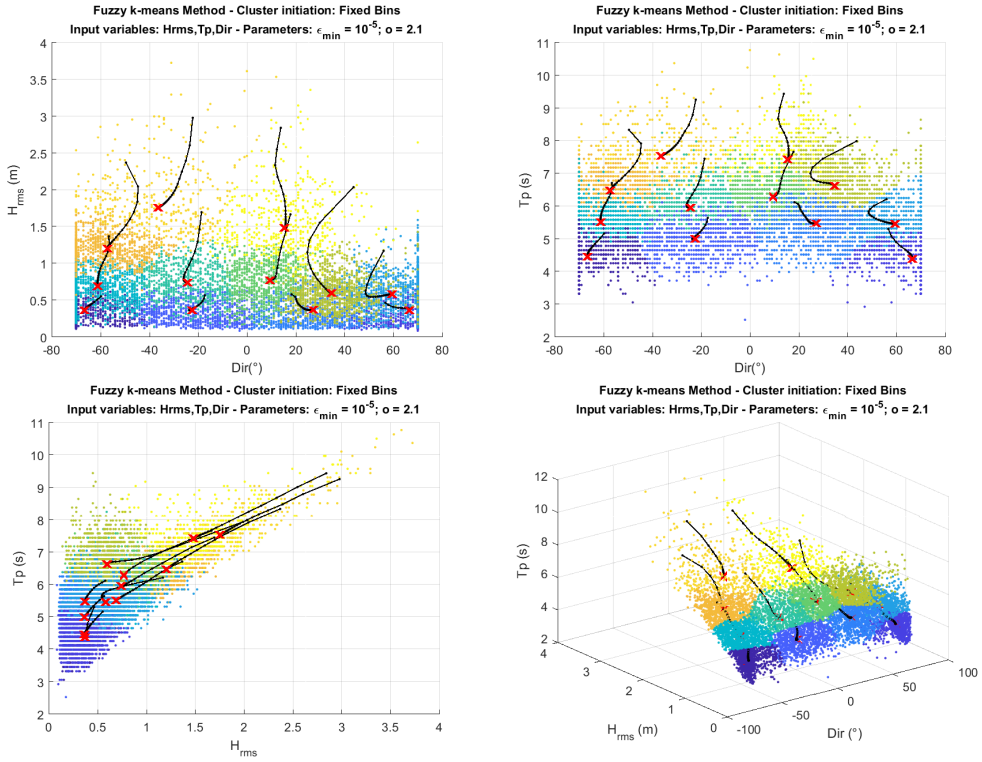


Figure A.21: Selection of method 6: Fuzzy k-means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

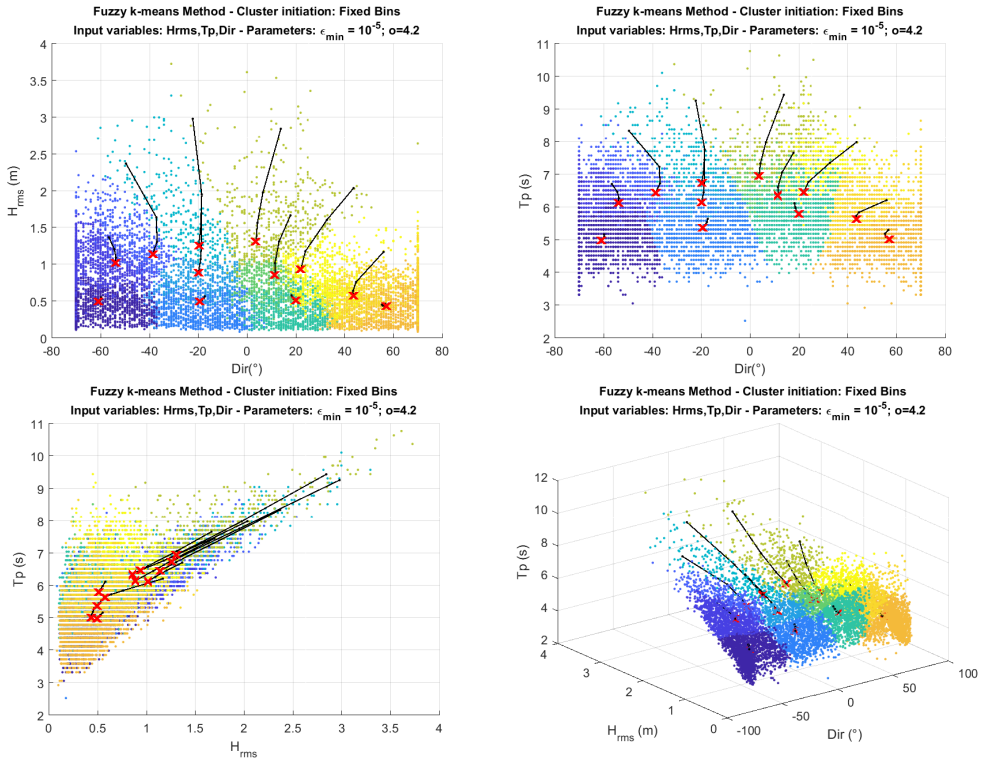


Figure A.22: Selection of method 6: Fuzzy k-means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

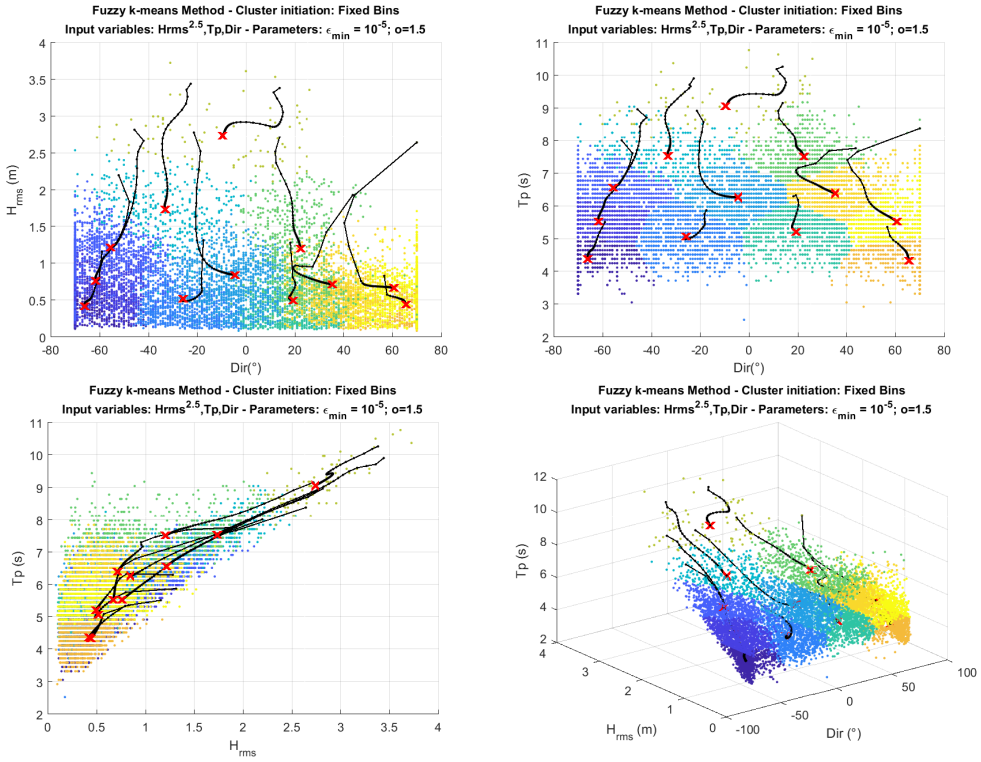


Figure A.23: Selection of method 6: Fuzzy k-means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

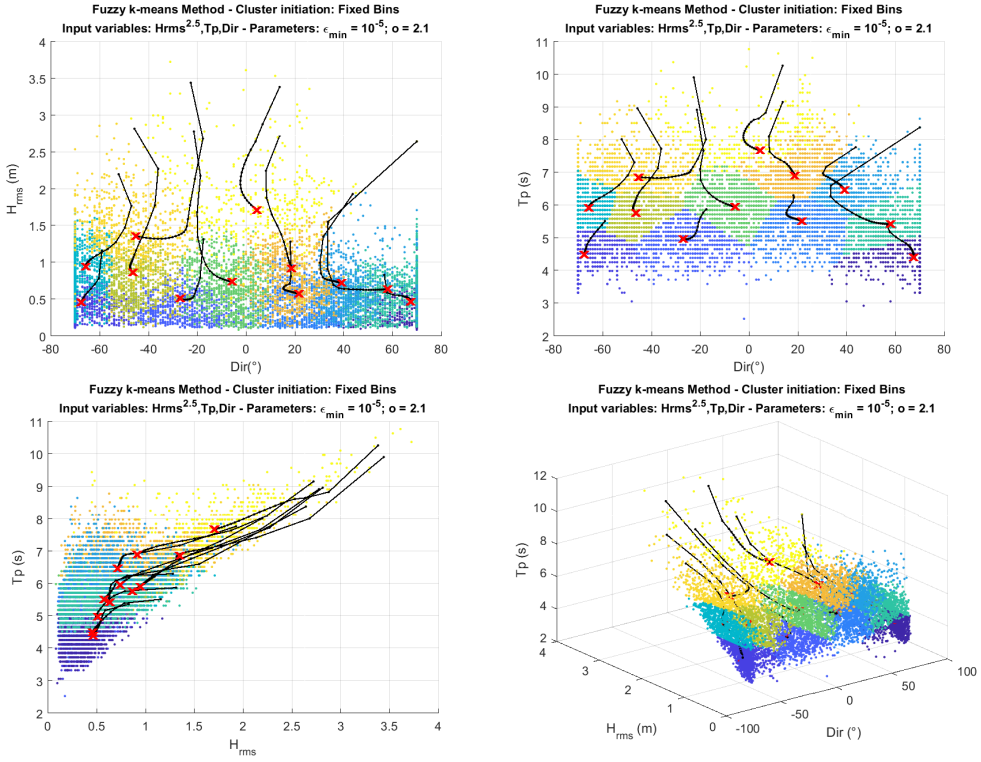


Figure A.24: Selection of method 6: Fuzzy k-means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

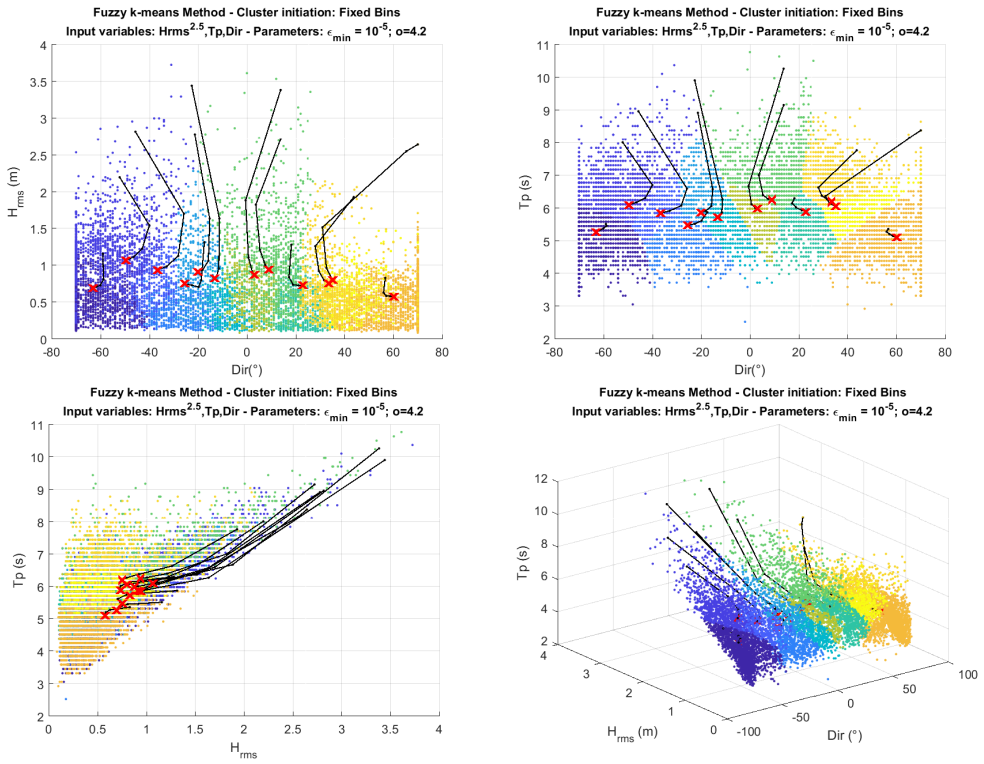


Figure A.25: Selection of method 6: Fuzzy k-means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

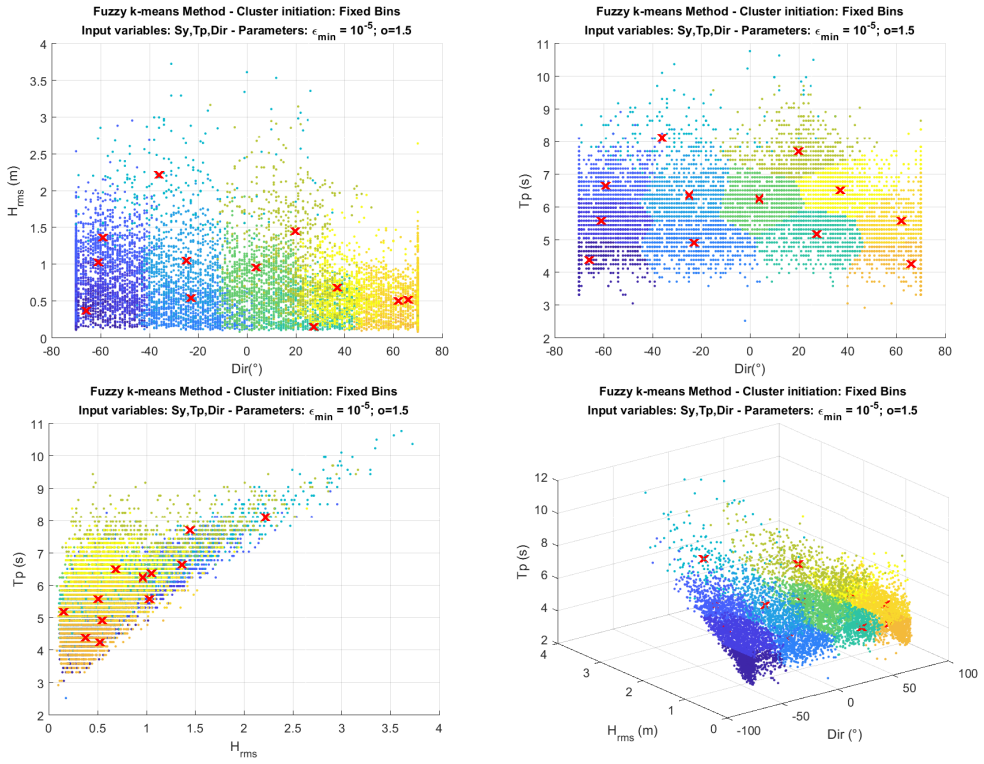


Figure A.26: Selection of method 6: Fuzzy k-means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

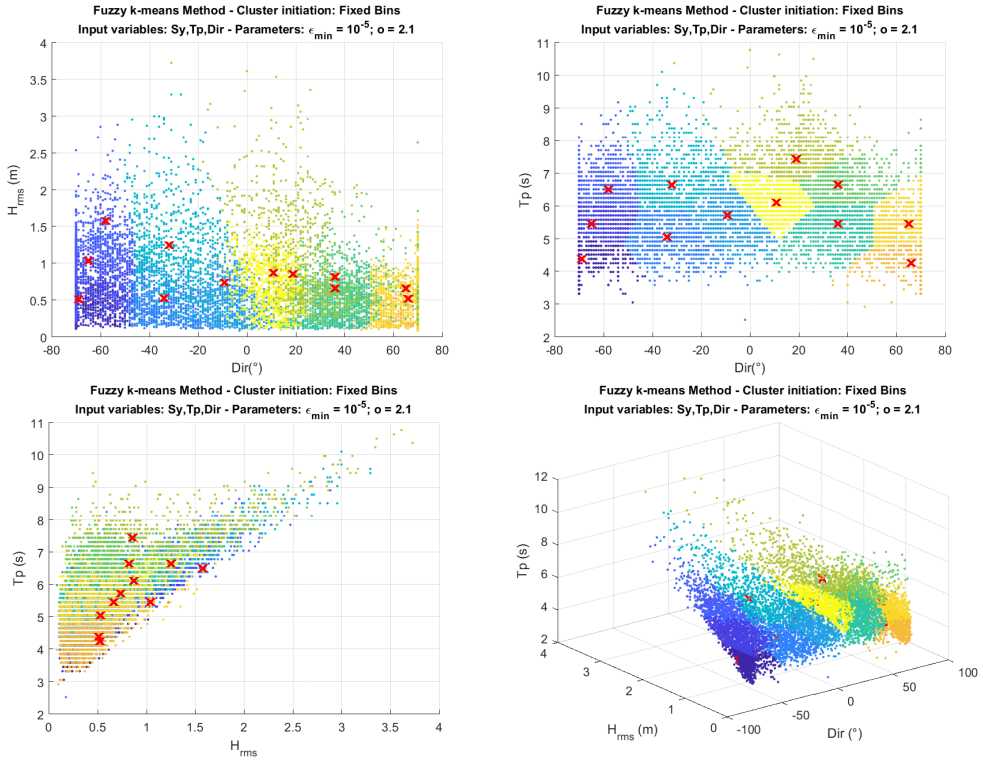


Figure A.27: Selection of method 6: Fuzzy k-means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

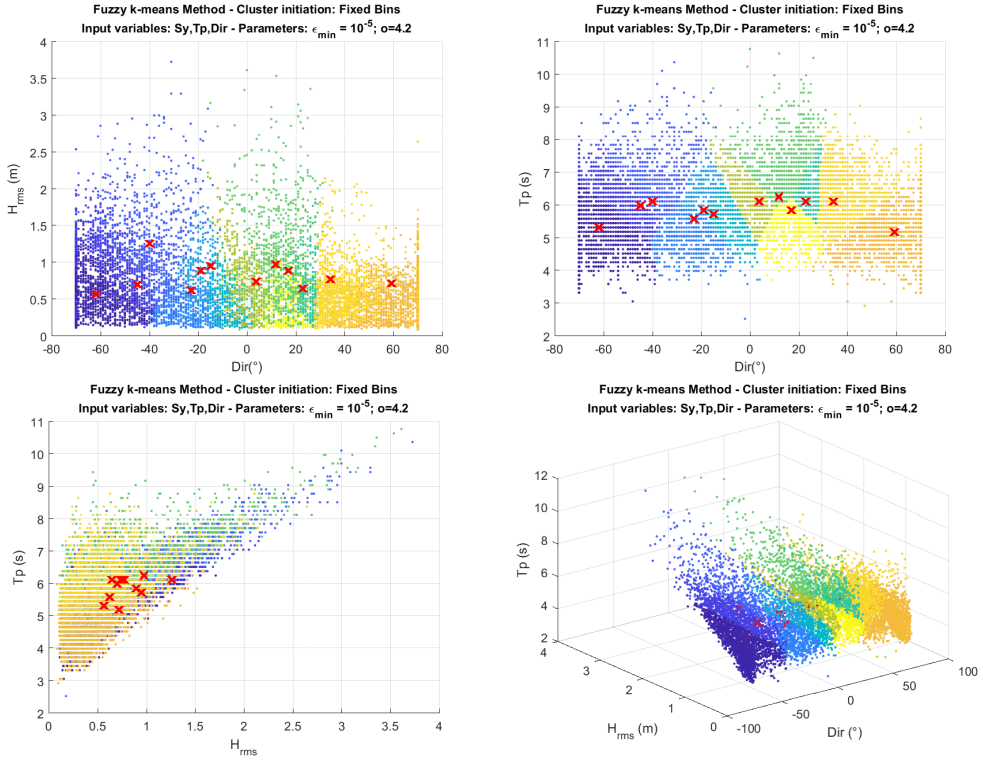


Figure A.28: Selection of method 6: Fuzzy k-means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

Based on the visualization of the selected centroids, the Fuzzy k-means methods with its different combinations of cluster initiation and input variables were simulated with $o = 1.5$ and $o = 2.1$. Table A.1 describes the simulated methods. The performance of the methods were evaluated by means of the cumulative skill

Table A.1: Simulated Methods 6 - Fuzzy k-means for sensitivity analysis of parameter o

Method	**Cluster initiation**	**Input variables**	**Parameter *o***
M6A1	K-means++	H_{rms}, T_p, θ	$o = 2.1$
M6B1	K-means++	$H_{rms}^{2.5}$, T_p, θ	$o = 2.1$
M6C1	K-means++	S_y, T_p, θ	$o = 2.1$
M6D1	MDA	H_{rms}, T_p, θ	$o = 2.1$
M6E1	MDA	$H_{rms}^{2.5}$, T_p, θ	$o = 2.1$
M6F1	MDA	S_y, T_p, θ	$o = 2.1$
M6G1	Fixed Bins	H_{rms}, T_p, θ	$o = 2.1$
M6H1	Fixed Bins	$H_{rms}^{2.5}$, T_p, θ	$o = 2.1$
M6I1	Fixed Bins	S_y, T_p, θ	$o = 2.1$
M6A2	K-means++	H_{rms}, T_p, θ	$o = 1.5$
M6B2	K-means++	$H_{rms}^{2.5}$, T_p, θ	$o = 1.5$
M6C2	K-means++	S_y, T_p, θ	$o = 1.5$
M6D2	MDA	H_{rms}, T_p, θ	$o = 1.5$
M6E2	MDA	$H_{rms}^{2.5}$, T_p, θ	$o = 1.5$
M6F2	MDA	S_y, T_p, θ	$o = 1.5$
M6G2	Fixed Bins	H_{rms}, T_p, θ	$o = 1.5$
M6H2	Fixed Bins	$H_{rms}^{2.5}$, T_p, θ	$o = 1.5$
M6I2	Fixed Bins	S_y, T_p, θ	$o = 1.5$

score R (see section 2.6). Figure A.29 and Figure A.30 present the skill scores of the different set ups of method 6 for each random sequence and the mean skill score of the 5 random sequences, respectively. Overall, the performance of method 6 presents higher skill scores when using $o = 1.5$. However, for some models, especially when using H_{rms}, T_p and θ as input variables, the set up of method 6 with $o = 2.1$ presented higher skill scores. This could be because changing the fuzzy parameters does not seems to influence much the selection of wave cases when compared to other input variables. Additionally, the sequencing can also have effects. The parameter $o = 1.5$ was selected for further analysis.

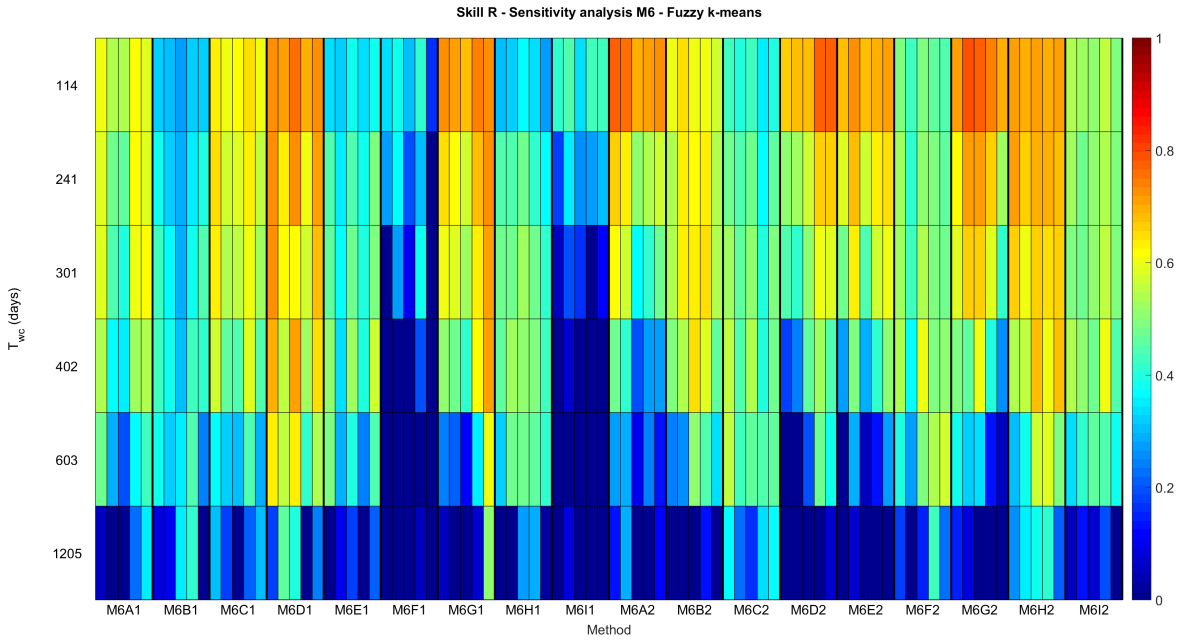


Figure A.29: Cumulative Skill Score of the 5 random sequences of the simulated set-ups of the Fuzzy k-means method for sensitivity analysis of parameter *o*.

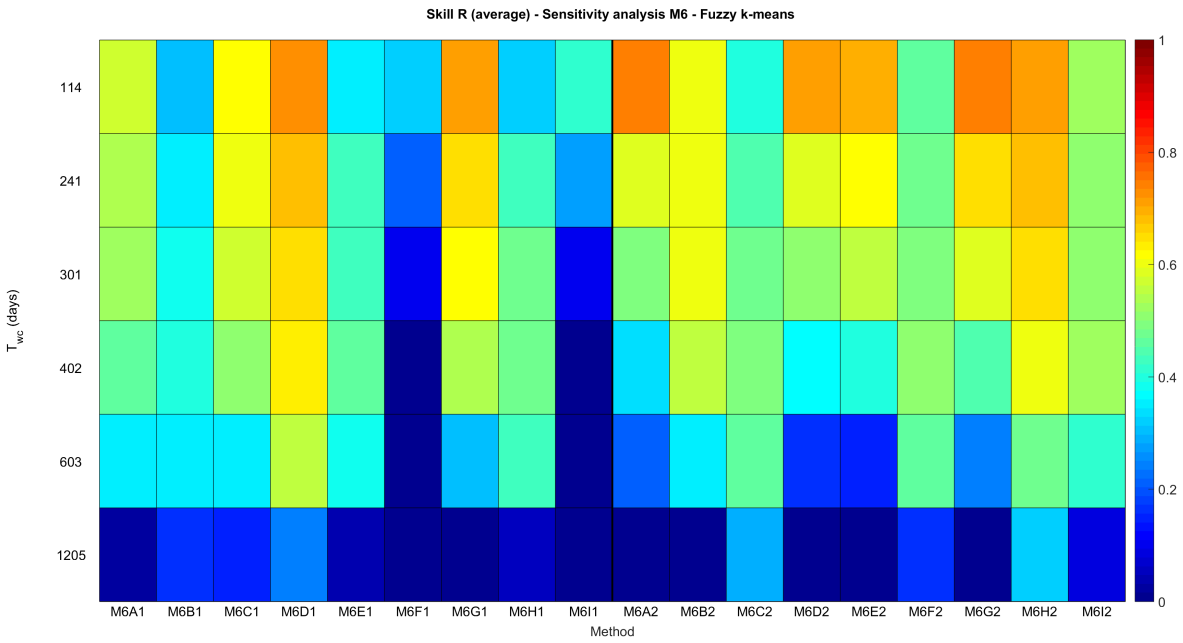


Figure A.30: Average cumulative Skill Score of the 5 random sequences of the simulated set-ups of the Fuzzy k-means method for sensitivity analysis of parameter *o*.

For the K-harmonic means method, the fuzzy parameter o influences on the dynamic weight, which is a balance between the membership function that gives higher weight to observations closer to the centroid and the boosting function that, gives higher weight to observations far from the centroids. According to Zhang (2000) in order to the dynamic weight effectively works $o \geq 2$, otherwise the membership function included in the dynamic weight overrules the boosting part. Furthermore, Zhang (2000) did not established an upper limit to the parameter o and states that higher the dimensionality, larger should be o. Figure A.31 demonstrates the different behavior of the dynamic weighting function for different values of o.

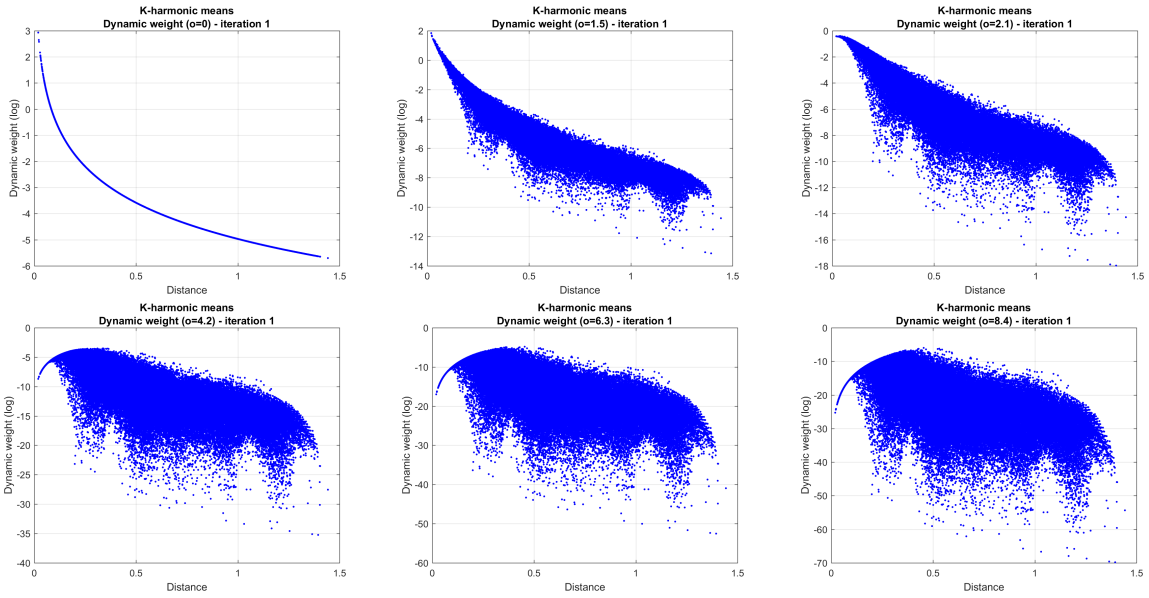


Figure A.31: Behavior of the fuzzy membership function of the Fuzzy K-means method according to different values of o. From left to right: $o = 0$, $o = 1.5$, $o = 2.1$, $o = 4.2$, $o = 6.3$ and $o = 8.4$.

The value $o = 2.1$ adopted in Olij (2015) was used again as a starting point for the sensitivity analysis. Figures A.32 to A.43 present the selection of centroids of the K-harmonic means method with different input variables and cluster initiation: Fixed bins for the parameters $o = 2.1$, $o = 4.2$, $o = 6.3$ and $o = 8.4$, respectively. It is possible to observe that when increasing the parameter o, higher wave cases are selected, however some of the spreading over direction is lost. In general, the method tens to behave similarly to the Maximum dissimilarity method when increasing o.

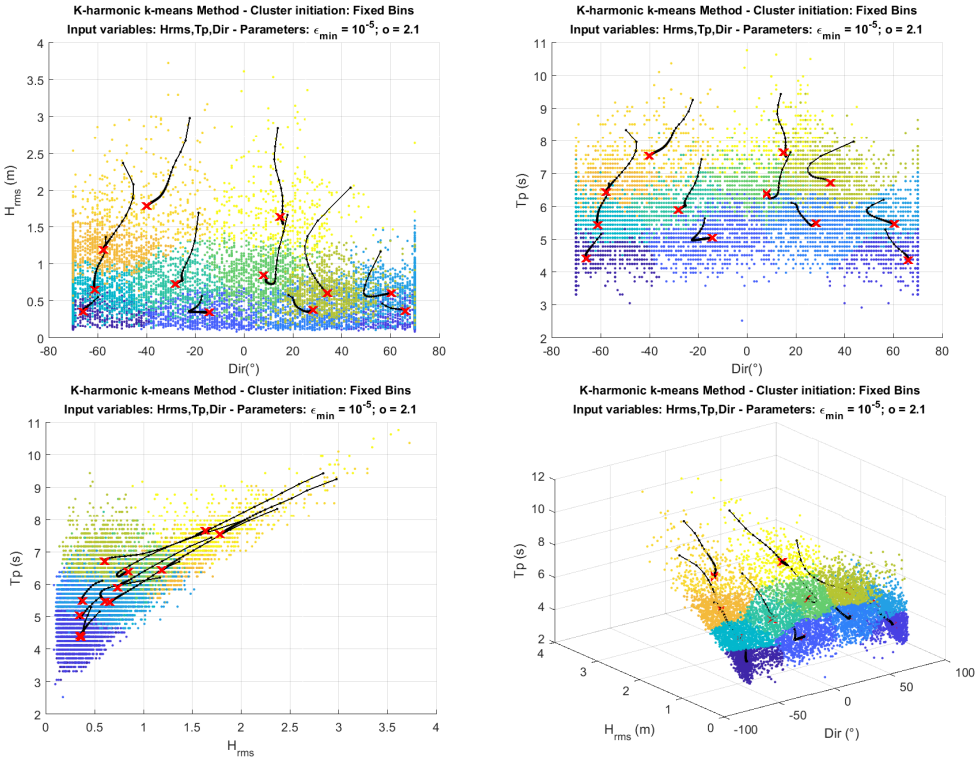


Figure A.32: Selection of method 7: K-harmonic means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

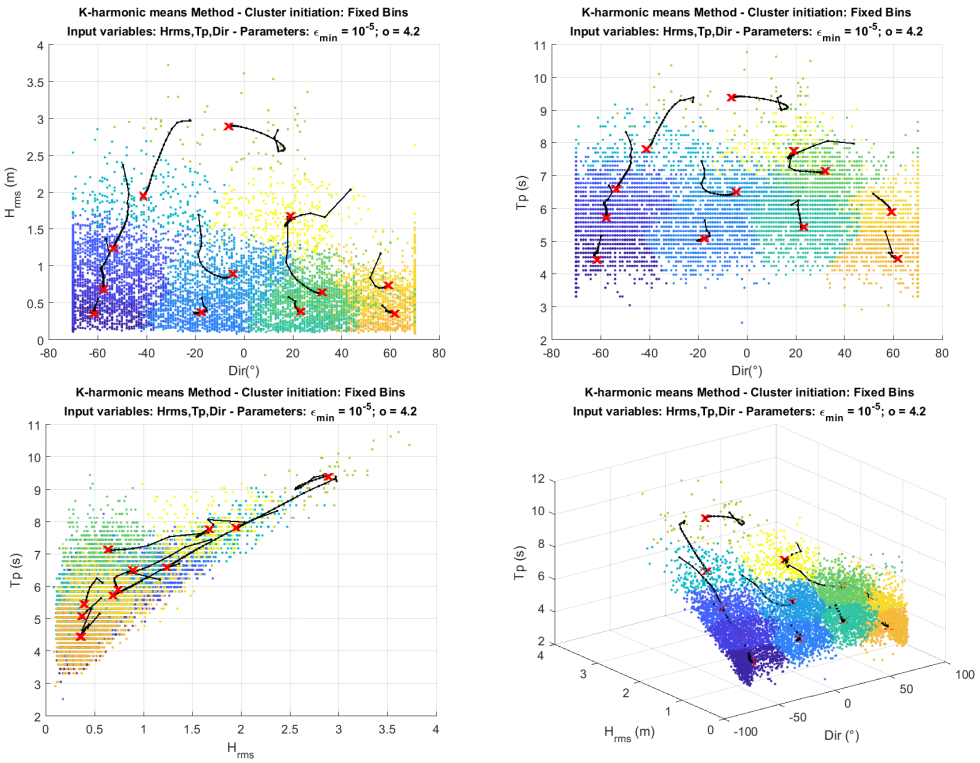


Figure A.33: Selection of method 7: K-harmonic means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

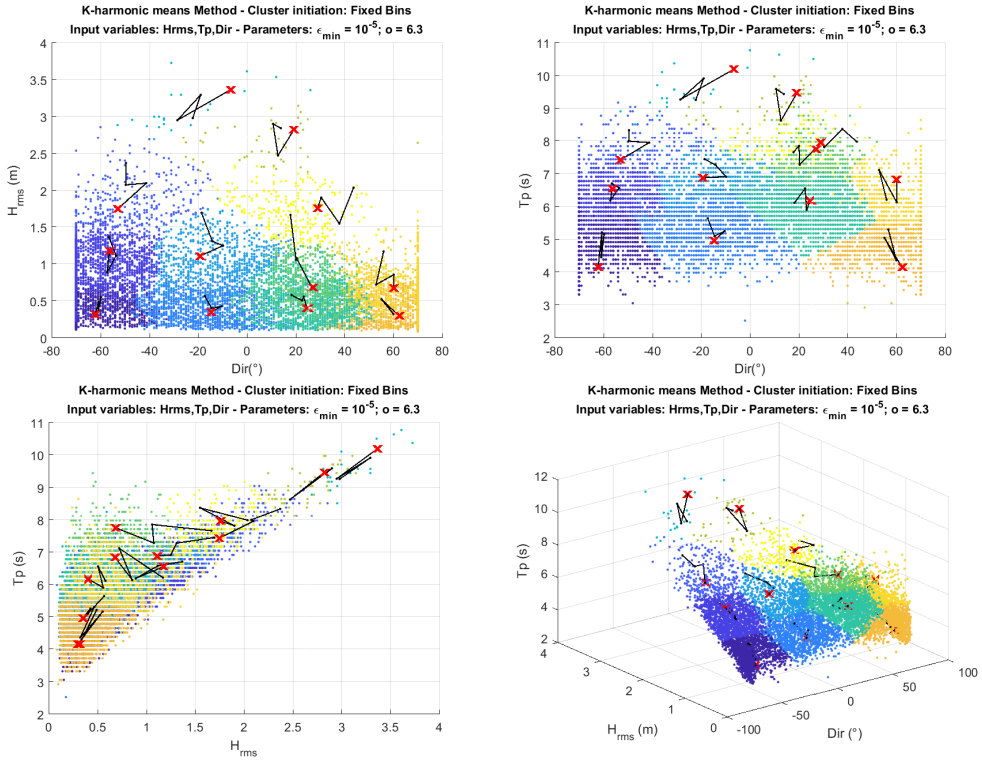


Figure A.34: Selection of method 7: K-harmonic means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

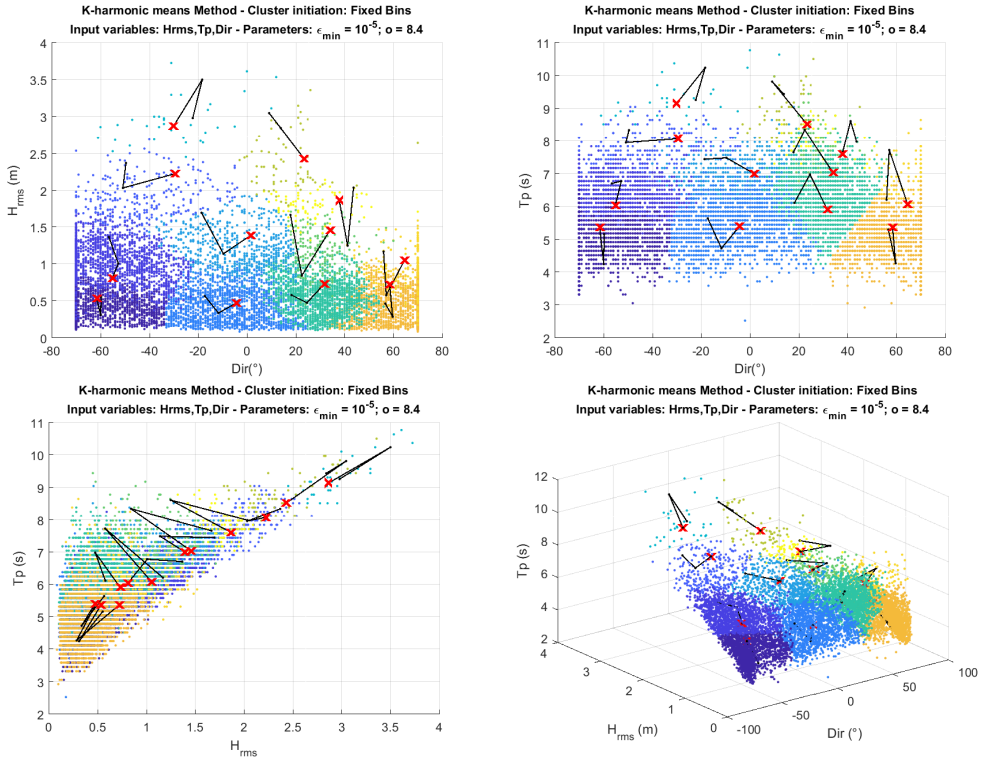


Figure A.35: Selection of method 7: K-harmonic means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

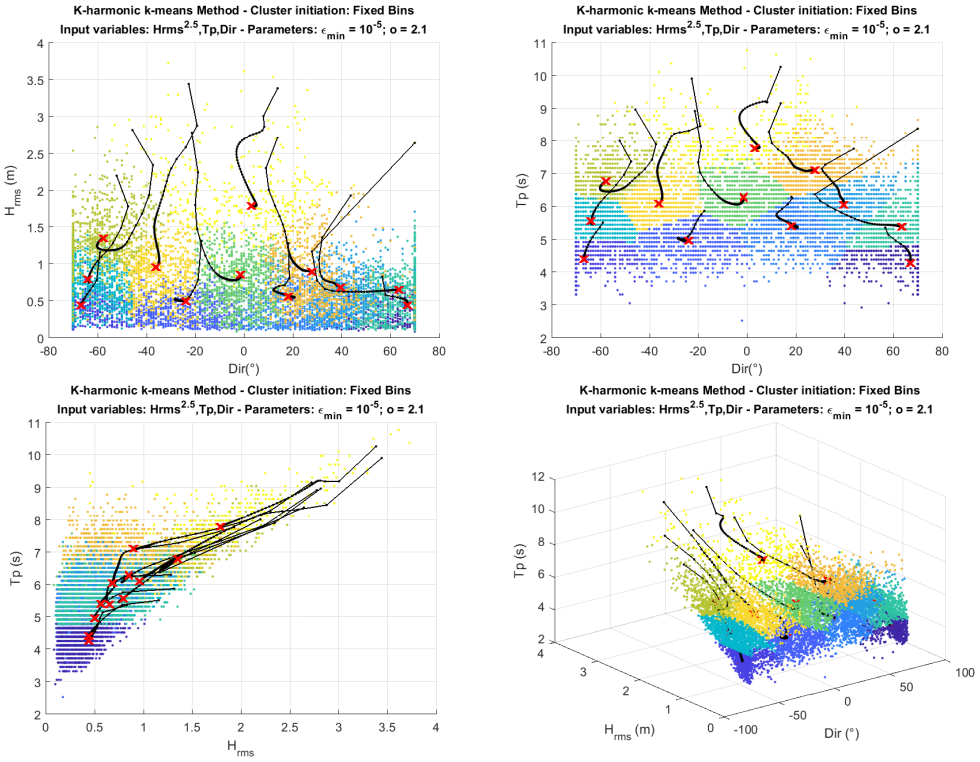


Figure A.36: Selection of method 7: K-harmonic means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

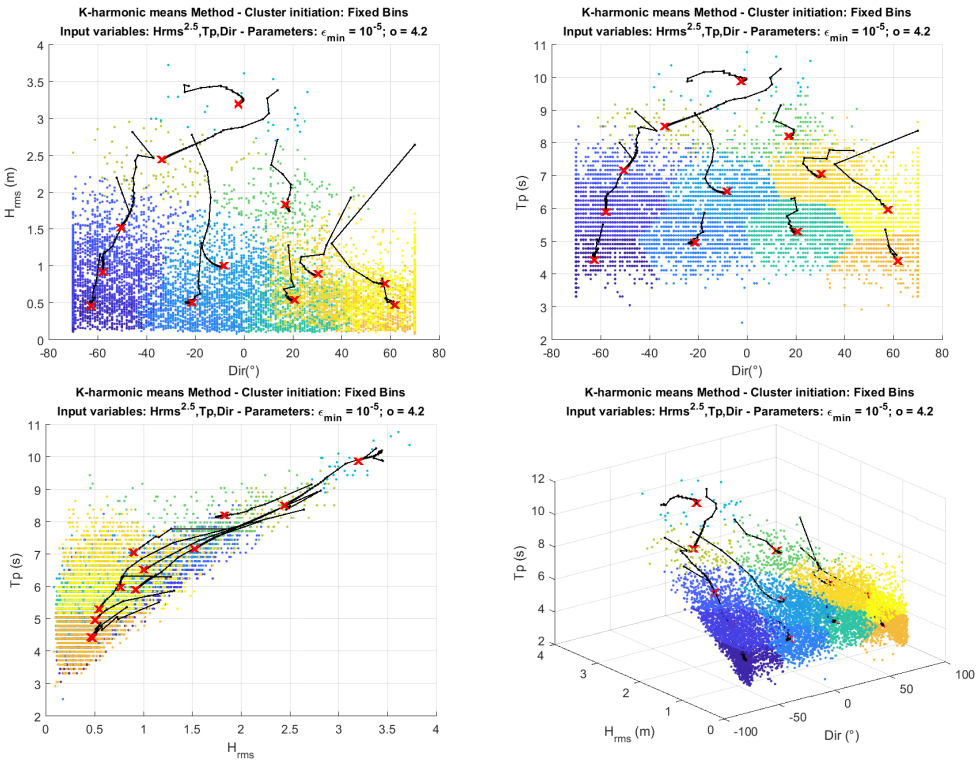


Figure A.37: Selection of method 7: K-harmonic means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

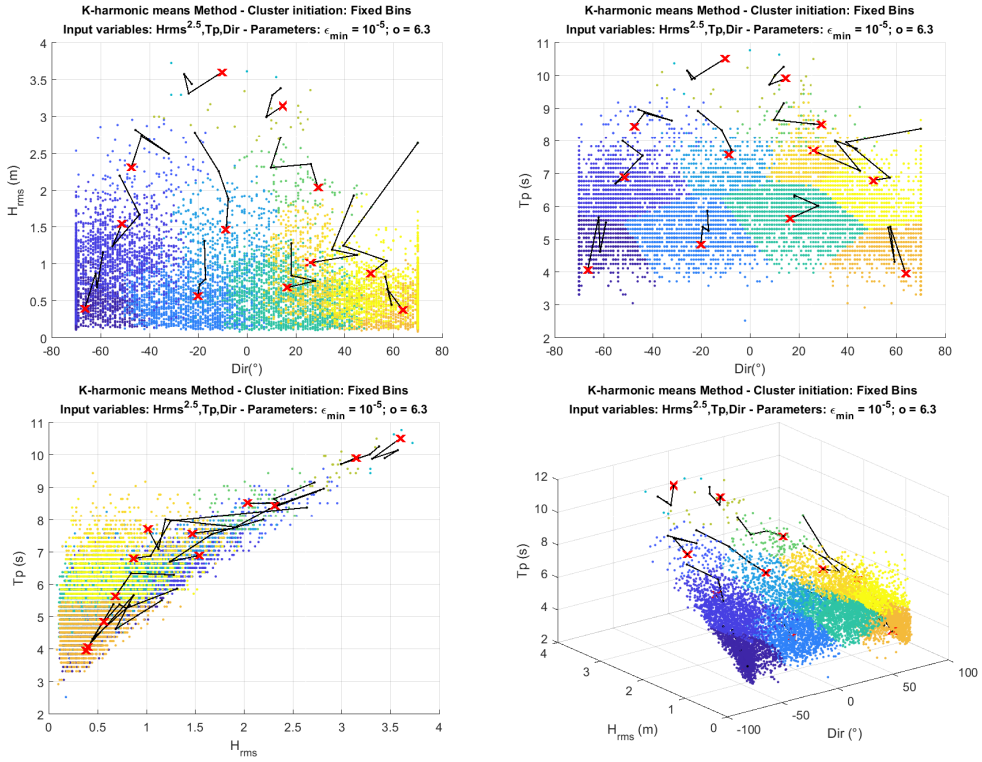


Figure A.38: Selection of method 7: K-harmonic means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

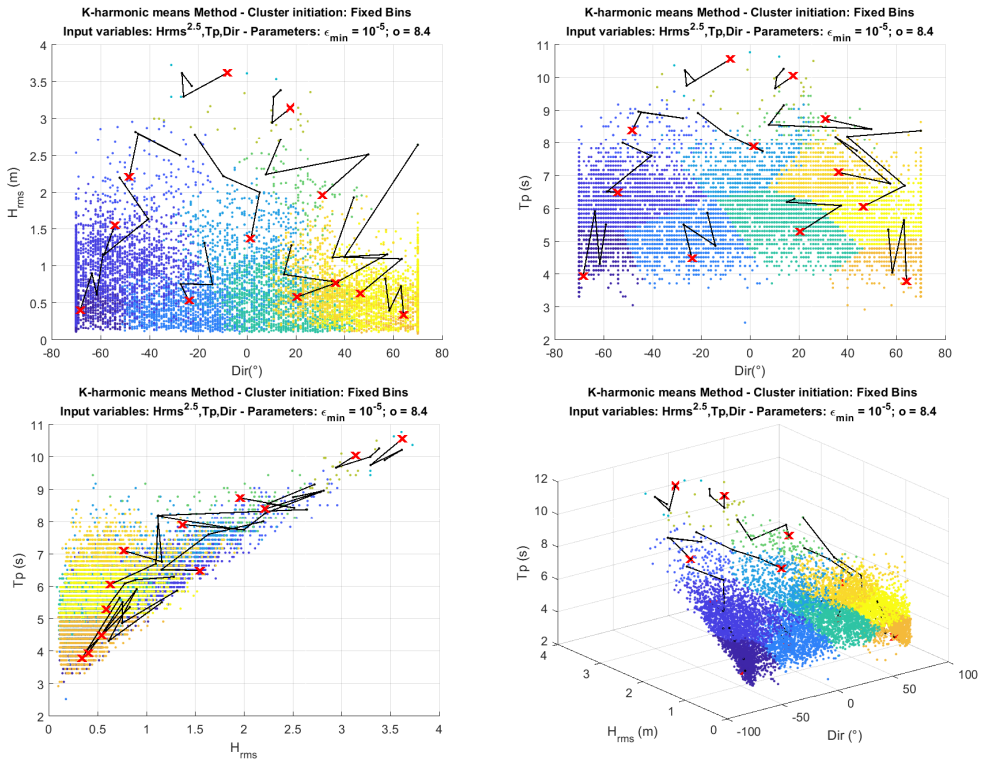


Figure A.39: Selection of method 7: K-harmonic means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters. The black lines represent the path followed by the centroids during the iterative process.

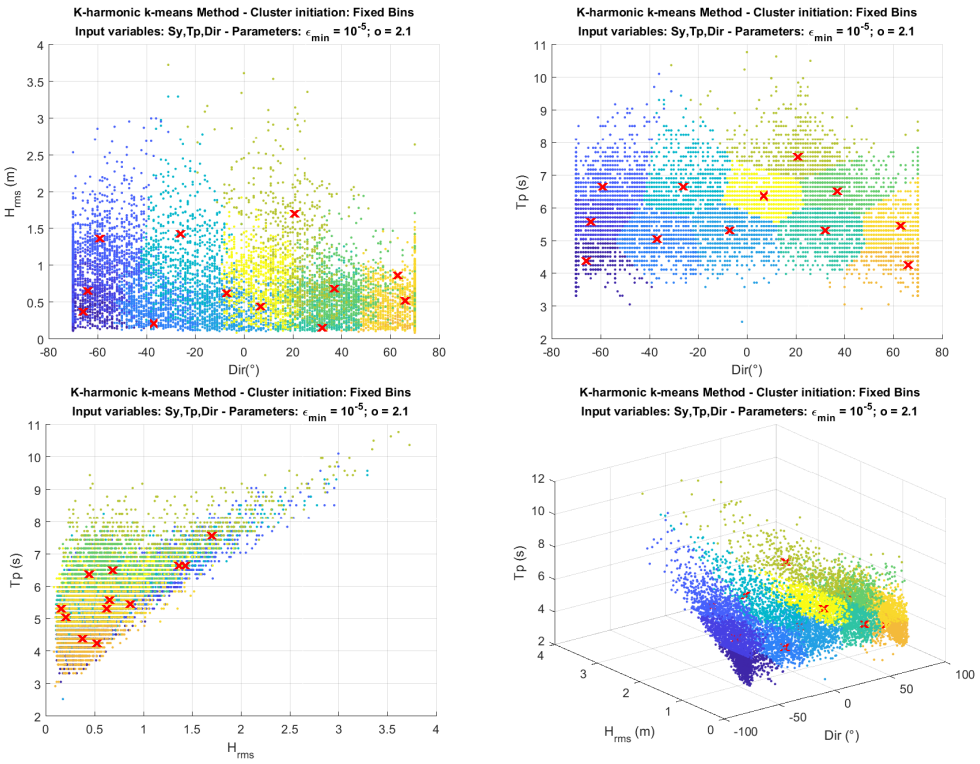


Figure A.40: Selection of method 7: K-harmonic means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

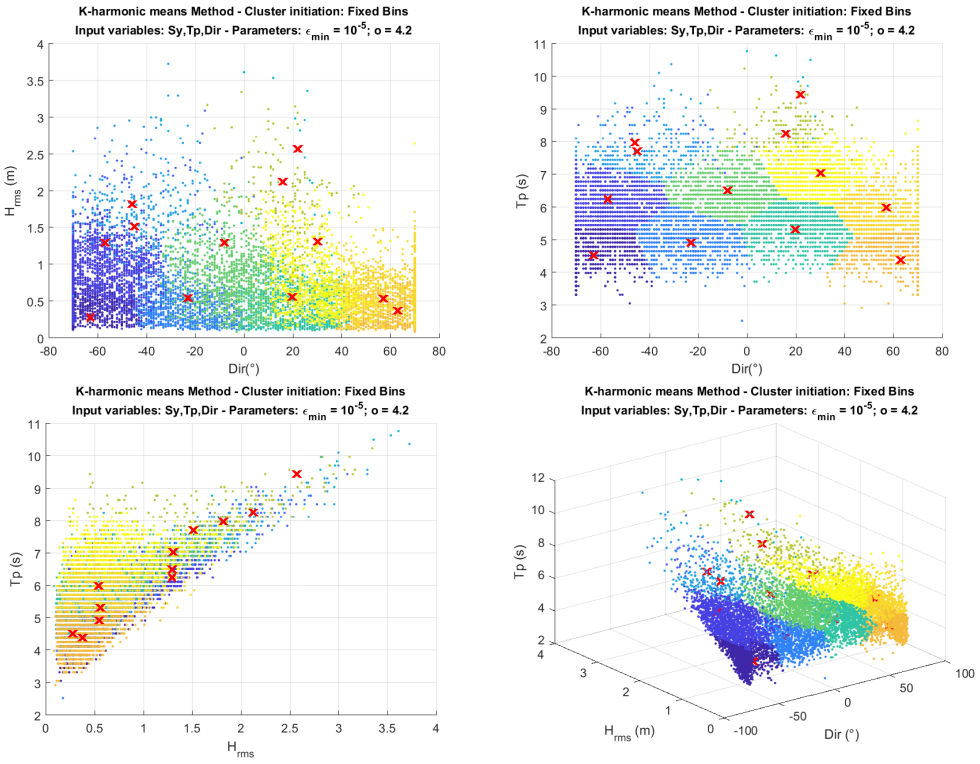


Figure A.41: Selection of method 7: K-harmonic means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

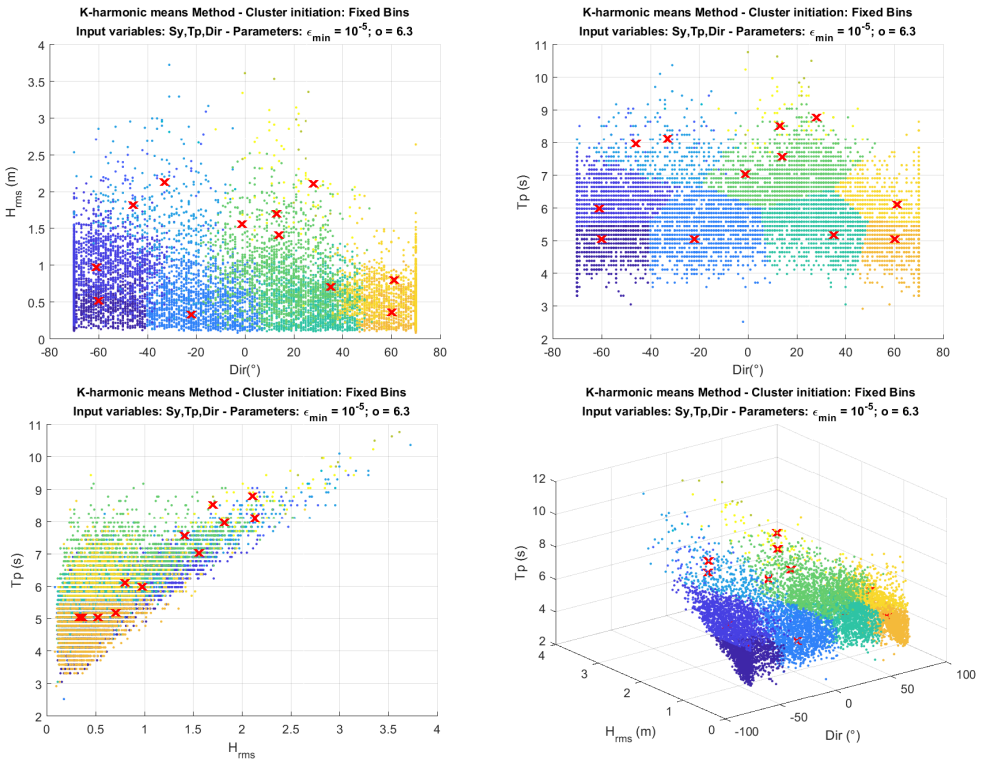


Figure A.42: Selection of method 7: K-harmonic means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

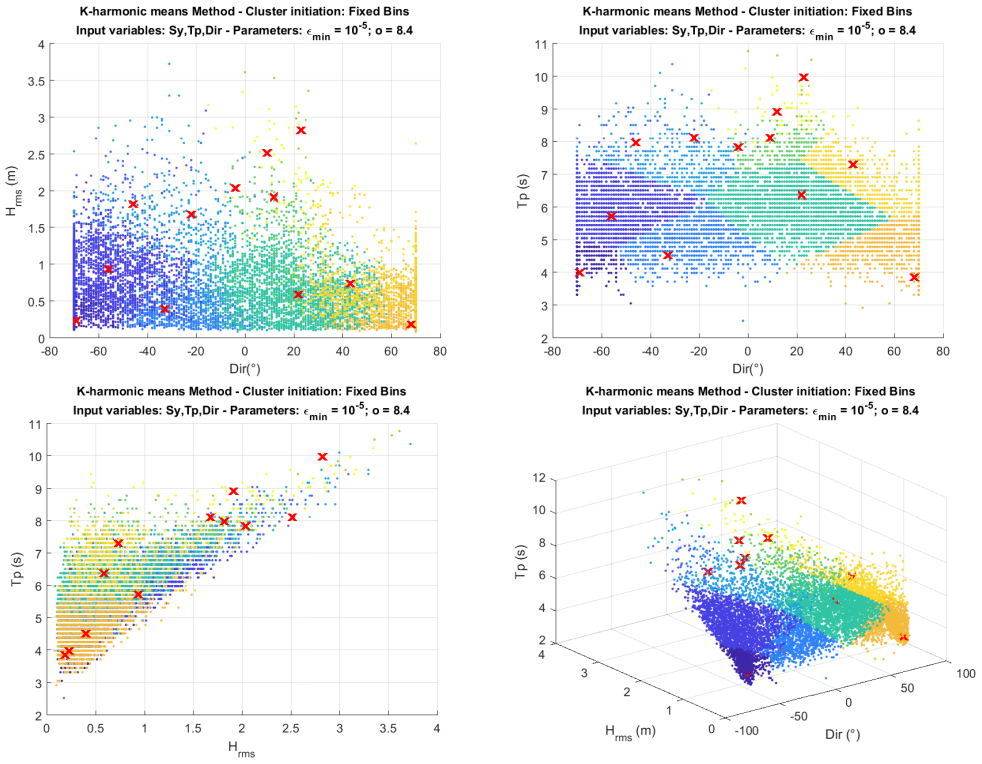


Figure A.43: Selection of method 7: K-harmonic means. The red crosses are the representative wave conditions and the small dots are the observations. The colors indicate the clusters.

Based on the visualization of the selected centroids, the K-harmonic means method with its different combinations of cluster initiation and input variables were simulated with $o = 2.1$, $o = 4.2$ and $o = 6.3$. Table A.2 describes the simulated methods. The performance of the methods were evaluated by means of the

Table A.2: Simulated Methods 7 - K-harmonic means for sensitivity analysis of parameter o

Method	**Cluster initiation**	**Input variables**	**Parameter *o***
M7A1	K-means++	H_{rms}, T_p, θ	$o = 2.1$
M7B1	K-means++	$H_{rms}^{2.5}, T_p, \theta$	$o = 2.1$
M7C1	K-means++	S_y, T_p, θ	$o = 2.1$
M7D1	MDA	H_{rms}, T_p, θ	$o = 2.1$
M7E1	MDA	$H_{rms}^{2.5}, T_p, \theta$	$o = 2.1$
M7F1	MDA	S_y, T_p, θ	$o = 2.1$
M7G1	Fixed Bins	H_{rms}, T_p, θ	$o = 2.1$
M7H1	Fixed Bins	$H_{rms}^{2.5}, T_p, \theta$	$o = 2.1$
M7I1	Fixed Bins	S_y, T_p, θ	$o = 2.1$
M7A2	K-means++	H_{rms}, T_p, θ	$o = 4.2$
M7B2	K-means++	$H_{rms}^{2.5}, T_p, \theta$	$o = 4.2$
M7C2	K-means++	S_y, T_p, θ	$o = 4.2$
M7D2	MDA	H_{rms}, T_p, θ	$o = 4.2$
M7E2	MDA	$H_{rms}^{2.5}, T_p, \theta$	$o = 4.2$
M7F2	MDA	S_y, T_p, θ	$o = 4.2$
M7G2	Fixed Bins	H_{rms}, T_p, θ	$o = 4.2$
M7H2	Fixed Bins	$H_{rms}^{2.5}, T_p, \theta$	$o = 4.2$
M7I2	Fixed Bins	S_y, T_p, θ	$o = 4.2$
M7A3	K-means++	H_{rms}, T_p, θ	$o = 6.3$
M7B3	K-means++	$H_{rms}^{2.5}, T_p, \theta$	$o = 6.3$
M7C3	K-means++	S_y, T_p, θ	$o = 6.3$
M7D3	MDA	H_{rms}, T_p, θ	$o = 6.3$
M7E3	MDA	$H_{rms}^{2.5}, T_p, \theta$	$o = 6.3$
M7F3	MDA	S_y, T_p, θ	$o = 6.3$
M7G3	Fixed Bins	H_{rms}, T_p, θ	$o = 6.3$
M7H3	Fixed Bins	$H_{rms}^{2.5}, T_p, \theta$	$o = 6.3$
M7I3	Fixed Bins	S_y, T_p, θ	$o = 6.3$

cumulative skill score R (See section 2.6). Figure A.44 and Figure A.45 present the skill scores of the different set ups of method 7 for each random sequence and the mean skill score of the 5 random sequences, respectively. Overall, method 7 with $o = 4.2$ presented slightly better performance than with $o = 2.1$ even though in some cases method 7 with $o = 2.1$ have higher scores due to the lost of spreading over direction caused by increasing o which explains why method 7 with $o = 6.3$ had the worst performance.

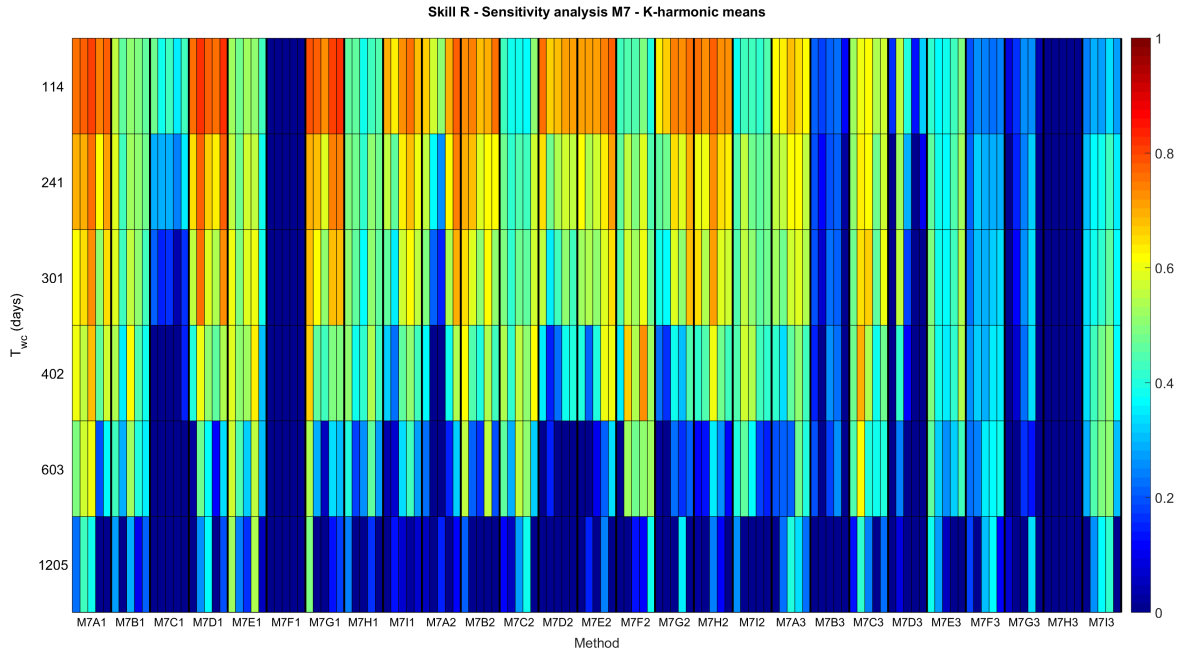


Figure A.44: Cumulative Skill Score of the 5 random sequences of the simulated set-ups of the Fuzzy k-means method for sensitivity analysis of parameter *o*.

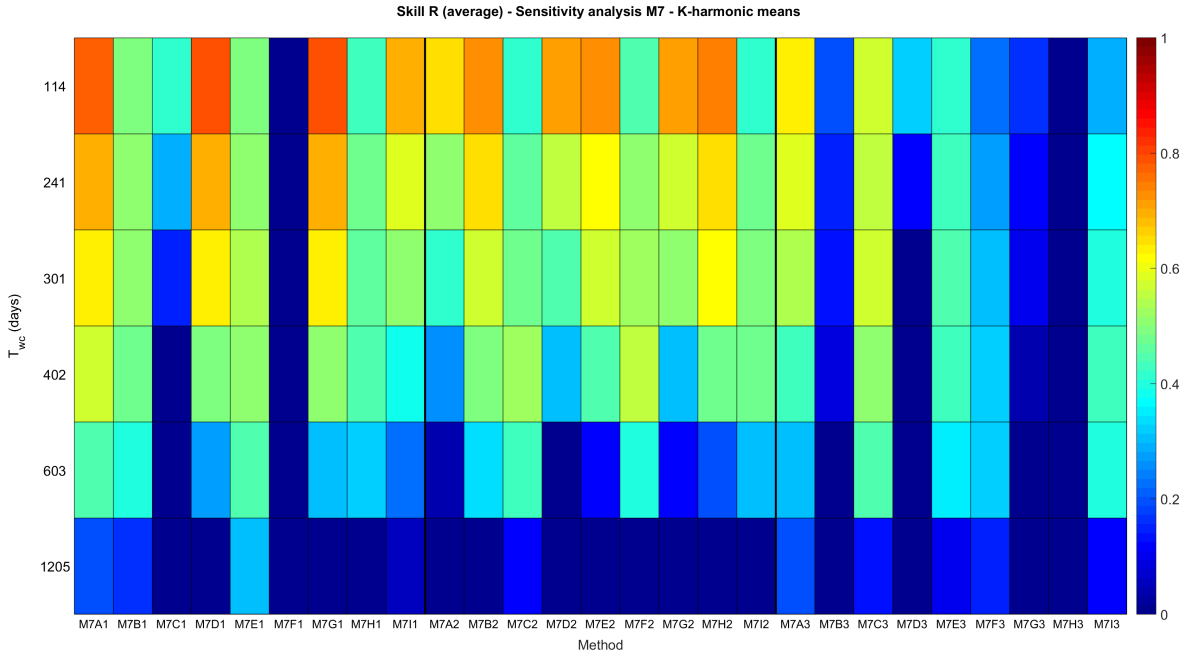


Figure A.45: Average cumulative Skill Score of the 5 random sequences of the simulated set-ups of the Fuzzy k-means method for sensitivity analysis of parameter *o*.

B

CALIBRATION OF UNIBEST-TC MODEL AT ANMOK BEACH, SOUTH KOREA

The calibration of the one year-long brute force model (06/16/2015-06/18/2016) of Anmok beach was performed by comparison of the profile at the end of the simulation (06/18/2016) with a measured profile of the survey campaign of 07/27/2016. The simulation period of the brute force model was selected according to wave and survey measurements availability. The time interval of approximately one month between the compared profiles is due to the lack of measured wave data. Moreover, it was assumed that none storm event neither significant morphological change occurred in this period.
The initial profile is based on survey measurements of 06/16/2015 and was interpolated onto a computational grid with resolution of 100 m offshore, gradually decreasing to 10 m from 20 m water depth towards the shore, including the active part of the profile. The model was forced with wave time-series (Hrms,Tp and θ) measured 850 m offshore (around 20 m water depth) with temporal resolution of 1 hour. The tidal forcing was not considered in the model because Anmok is characterized by micro-tidal regime, with tidal range of few decimeters (Swinkels et al. 2016). The D_{50} was taken as $400\mu m$ (medium sand) (Swinkels et al. 2016). The model contains some free parameters that were tuned in order to calibrate the simulated profile with measured profile. More information about the parameter and formulations of the Unibest-TC model can be found in Delft Hydraulics (2000). Table B.1 presents the main parameters set up for the calibration of Anmok's brute force Unibest-TC model.

Table B.1: Calibration parameters of Anmok's brute force model

Parameter	**Value**
Time-step (dt)	$dt = 0.04167$ days
Breaker-delay (λ)	$\lambda = 1$
Angle of repose 1 (ϕ_1)	$\phi_1 = 1.5$
Cross-shore location of ϕ_1 ($XF1$)	$XF1 = 400m$
Angle of repose 2 (ϕ_2)	$\phi_2 = 0.1$
Cross-shore location of ϕ_2 ($XF2$)	$XF2 = 150m$
Current-related roughness (k_c)	$k_c = 0.005593$
Wave-related roughness (k_w)	$k_w = 0.00045$

Figure B.1 presents the comparison of the simulated and measured profiles. The simulated and measured profiles do not match perfectly. Possible reasons for that are the time interval between measured and simulated profile and the alongshore variation observed in Anmok beach. Furthermore, it is recognized that the calibration of the Anmok's brute force needs some improvement. However, the profile obtained from

the brute force calibration was considered good enough for the purpose of validation of the input reduction framework.

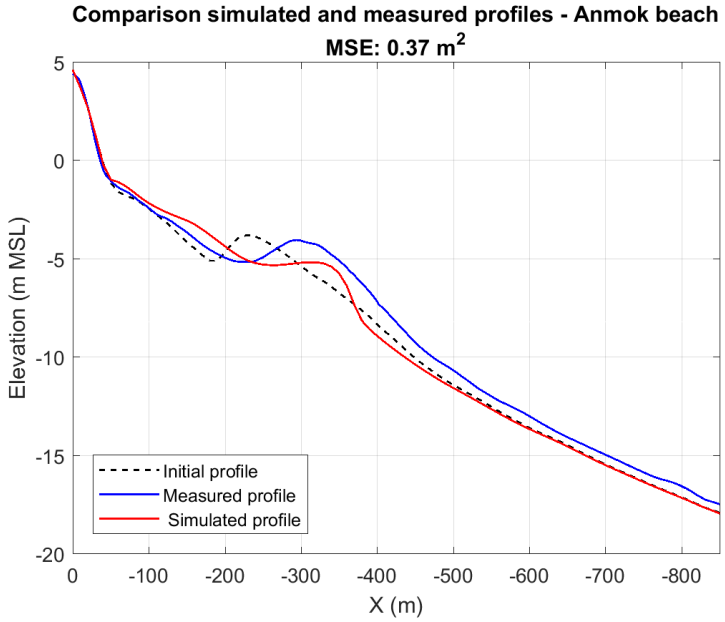


Figure B.1: Comparison of simulated and measured profiles at Anmok beach, South Korea. The black dashed like is the initial profile (measured), the blue line is the measured profile at 07/27/2016 and the red line is the simulated profile at the end of the simulation.

BIBLIOGRAPHY

Arthur, D. & Vassilvitskii, S. (2007), 'K-Means++: the Advantages of Careful Seeding', *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms* **8**, 1027–1025.

Benedet, L., Dobrochinski, J. P. F., Walstra, D. J. R., Klein, A. H. F. & Ranasinghe, R. (2016), 'A morphological modeling study to compare different methods of wave climate schematization and evaluate strategies to reduce erosion losses from a beach nourishment project', *Coastal Engineering* **112**, 69–86.
URL: *http://dx.doi.org/10.1016/j.coastaleng.2016.02.005*

Bosboom, J., Reniers, A. J. H. M. & Luijendijk, A. P. (2014), 'On the perception of morphodynamic model skill', *Coastal Engineering* **94**, 112–125.
URL: *http://dx.doi.org/10.1016/j.coastaleng.2014.08.008*

Bosboom, J. & Stive, M. J. (2011), *Coastal Dynamics I - Lecture Notes CT4305*, VSSD.

Brown, J. M. & Davies, A. G. (2009), 'Methods for medium-term prediction of the net sediment transport by waves and currents in complex coastal regions', *Continental Shelf Research* **29**(11-12), 1502–1514.

Chang, C., Lai, J. & Jeng, M. (2011), 'A Fuzzy K-means Clustering Algorithm Using Cluster Center Displacement.', *Journal of Information Science and Engineering* **1009**, 995–1009.
URL: *http://www.iis.sinica.edu.tw/page/jise/2011/201105_12.pdf*

Delft Hydraulics (2000), Unibest-TC 2.0x - Overview of Model Formulations, Technical Report January.

Dobrochinski, J. P. H. (2009), Redução E Esquematização De Clima De Ondas Para Modelagem Morfológica, Technical report.

Grunnet, N. M. & Ruessink, B. G. (2005), 'Morphodynamic response of nearshore bars to a shoreface nourishment', *Coastal Engineering* **52**(2), 119–137.

Guillen, J., Stive, M. & Capobianco, M. (1999), 'Shoreline evolution of the Holland coast on a decadal scale', *Earth Surface Processes and Landforms* **24**, 517—-536.

Kennard, R. W. & Stone, L. A. (1969), 'Computer aided design of experiments', *Technometrics* **11**(1), 137–148.
URL: *http://www.tandfonline.com/doi/abs/10.1080/00401706.1969.10490666*

Ojeda, E., Ruessink, B. G. & Guillen, J. (2008), 'Morphodynamic response of a two-barred beach to a shoreface nourishment', *Coastal Engineering* **55**(12), 1185–1196.
URL: *http://dx.doi.org/10.1016/j.coastaleng.2008.05.006*

Olij, D. J. C. (2015), 'Wave climate reduction for medium term process based morphodynamic simulations with application to the Durban Coast', p. 131.

Ruessink, B. G., Kuriyama, Y., Reniers, A. J. H. M., Roelvink, J. A. & Walstra, D. J. R. (2007), 'Modeling cross-shore sandbar behavior on the timescale of weeks', *Journal of Geophysical Research: Earth Surface* **112**(3), 1–15.

Ruessink, B. G., Pape, L. & Turner, I. L. (2009), 'Daily to interannual cross-shore sandbar migration: Observations from a multiple sandbar system', *Continental Shelf Research* **29**(14), 1663–1677.

Southgate, H. N. (1995), 'The effects of wave chronology on medium and long term coastal morphology', *Coastal Engineering* **26**(3-4), 251–270.
URL: *http://linkinghub.elsevier.com/retrieve/pii/0378383995000283*

Sutherland, J., Peet, A. H. & Soulsby, R. L. (2004), 'Evaluating the performance of morphological models', *Coastal Engineering* **51**(8-9), 917–939.

Swinkels, C., Friedman, J., de Boer, W., Mccall, R. & Huisman, B. (2016), CoMIDAS - East Coast Case - Numerical model study of Anmok Beach, Technical report, Deltares, Delft.

van Arkel, M. (2016), 'Towards an efficient sensitivity analysis of wave forcing in coastal erosion studies', (May), 100.
URL: *http://repository.tudelft.nl/islandora/object/uuid%3A3bfe1ad0-426d-4671-810f-2110724c3385*

van Duin, M. J. P., Wiersma, N. R., Walstra, D. J. R., van Rijn, L. C. & Stive, M. J. F. (2004), 'Nourishing the shoreface: Observations and hindcasting of the Egmond case, The Netherlands', *Coastal Engineering* **51**(8-9), 813–837.

Walstra, D. J. R., Hoekstra, R., Tonnon, P. K. & Ruessink, B. G. (2013), 'Input reduction for long-term morphodynamic simulations in wave-dominated coastal settings', *Coastal Engineering* **77**, 57–70.
URL: *http://dx.doi.org/10.1016/j.coastaleng.2013.02.001*

Walstra, D. J. R., Reniers, A. J. H. M., Ranasinghe, R., Roelvink, J. A. & Ruessink, B. G. (2011), 'On bar growth and decay during interannual net offshore migration', *Coastal Engineering* **60**(1), 190–200.
URL: *http://dx.doi.org/10.1016/j.coastaleng.2011.10.002*

Wijnberg, K. M. & Terwindt, J. H. J. (1995), 'Extracting decadal morphological behaviour from high-resolution, long-term bathymetric surveys along the Holland coast using eigenfunction analysis', *Marine Geology* **126**(1-4), 301–330.

Willett, P. (1996), 'Molecular diversity techniques for chemical databases'.
URL: *http://www.informationr.net/ir/2-3/paper19.html*

Zhang, B. (2000), 'Generalized K -Harmonic Means – Boosting in Unsupervised Learning 1'.

Zhang, B., Hsu, M. & Dayal, U. (1999), 'Clustering Algorithm K -Harmonic Means -A Data Clustering Algorithm', *Hewlett-Packard Labs Technical Report HPL-1999-124* .