

## UNIVERSITY OF TWENTE.

**Faculty of Civil Engineering** 

Department of Water Engineering and Management

# Extrapolation of morphodynamics for the estimation of future seabed changes in the Europlatform area

Theodora Vinni M.Sc. Thesis Report October 2019

Graduation committee: Prof. dr. S.J.M.H. Hulscher (University of Twente) Dr. ir. P.C. Roos (University of Twente) Dr. ir. G.H.P. Campmans (University of Twente)



#### SUMMARY

**Background** Sand waves typically occur in sandy shallow seas such as the North Sea and they are induced by tidal currents. Sand waves are dynamically active features with migration rates up to a few meters per year posing a serious threat to the safety of submarine structures. Thus, a detailed study of the sand wave characteristics (shape and migration rates) could be used in the strategy formulation process of offshore wind farms.

**Objectives** This study aims to calculate the main characteristics (height, length, asymmetry, migration and growth rate) of a sand wave field located in the vicinity of the Europlatform (North Sea). Furthermore, it aims to identify correlation patterns between the migration rate and the sand wave shape. The measurements of 13 surveys (from 2006 to 2018) carried out at the study area have been used in the bathymetric analysis.

The second objective of the study is to create a statistical predictive model that can be used to estimate future seabed levels. The predictive model has been developed, using the findings of the bathymetric analysis.

**Methods** We use the Fourier analysis to filter bedforms, with shorter and longer wavelengths, from the sand wave signal. From the filtered seabed profiles, we determine the positions of crests and troughs and we calculate the sand wave characteristics using the method suggested by Knaapen (2005).

The least square method has been used to define the best fitted regression model. The selected regression model has been extrapolated and we use the prediction bounds to define future minimum and maximum vertical positions of sand wave crests and troughs.

**Results** From the bathymetric analysis, we found that the characteristics of the sand wave field show little, if any, variation over the study period (13 years). Furthermore, the predominant direction of the migration of the sand wave field is North East, which coincides with the steeper slope. The field comprises individual sand waves with average lengths and heights ranging from 117 to 347m and from 1.9 to 8.4m respectively. The mean (per individual sand wave) migration rates vary between 0.3 and 1.86m/yr to the North East. Lastly, it has been observed that a few sand waves migrate in the opposite direction of the steeper slopes.

A statistically weak negative correlation has been revealed between migration rate and sand wave height and length. This study found no correlation between migration rate and sand wave asymmetry. Based on the predictive model and the suggested uncertainty bounds, the future seabed level changes have been estimated. Specifically, we found that the maximum potential rise and lowering of the seabed level in 2030 is 0.9m and -0.4m, respectively. The predicted seabed rising and lowering have been estimated with respect to the most recent bathymetrical survey (2018).

**Conclusion** The proposed model could be used to improve the decision-making process by predicting minimum and maximum seabed level changes. However, the prediction model is not valid outside of the study area.

#### CONTENTS

1	. INT	ROD	UCTION	1
	1.1.	PRO	OBLEM DEFINITION	1
	1.2.	RES	SEARCH APPROACH	2
	1.2.	.1.	RESEARCH OBJECTIVES AND RESEARCH QUESTIONS	2
	1.2.	.2.	METHODOLOGY	2
	1.3.	RE/	ADING GUIDE	2
2	. STI	JDY	AREA	3
	2.1.	LOC	CATION OF THE STUDY AREA	3
	2.2.	EN\	/IRONMENTAL PARAMETERS OF THE STUDY AREA	4
	2.2.	.1.	SUMMARY OF FINDINGS	6
3	. ANA	ALYS	SIS OF BATHYMETRIC DATA	7
	3.1.	INT	RODUCTION	7
	3.2.	EST	TIMATION OF SAND WAVE CHARACTERISTICS	8
	3.2.	.1.	DATA PROCESSING	8
	3.2.	.2.	DEFINITION OF THE SAND WAVE CHARACTERISTICS	9
	3.2.	.3.	SUMMARY OF FINDINGS	11
	3.3.	EST	TIMATION OF THE CORRELATION BETWEEN THE MIGRATION RATION	TE AND THE
	SAND	WAV	E SHAPE	17
	3.3.	.1.	SUMMARY OF FINDINGS	18
	3.4.	COI	NCLUSIONS	20
4	. FU1	TURE		22
	4.1.	MAG	GNITUDE OF UNCERTAINTY BANDS	22
	4.1.	.1.	SOURCES OF UNCERTAINTIES	22
	4.1.	.2.	UNCERTAINTY BANDS	23
	4.1.	.3.	SUMMARY OF FINDINGS	23
	4.2.	PRE	EDICTIVE MODEL	24
	4.2.	.1.	SELECTION OF THE BEST REGRESSION MODEL	24
	4.2	2.1.1	. ESTIMATION OF THE ADJUSTED COEFFICIENT OF	MULTIPLE
	DE	ETER	RMINATION	24
	4.2	2.1.2	. SUMMARY OF FINDINGS	25
	4.2.	.2.	EXTRAPOLATION OF THE BEST REGRESSION MODEL	26

4.2.2.1. ESTIMATION OF PREDICTION INTERVALS	26
4.2.2.2. SUMMARY OF FINDINGS	27
4.3. ESTIMATION OF FUTURE SEABED LEVEL RISING AND LOWERING	31
4.3.1. SUMMARY OF THE FINDINGS	31
4.4. CONCLUSIONS	32
5. DISCUSSION	33
6. CONCLUSION	34
REFERENCES	
APPENDIX A: BATHYMETRIC DATA ANALYSIS	
APPENDIX B: REGRESSION ANALYSIS	42

#### NOMENCLATURE

#### Notations

Symbol	Parameter/variable	Unit
А	Sand wave asymmetry	[-]
b	Regression coefficients	[-]
С	Sand wave migration rate	[m/yr]
D <sub>50</sub>	Mean particle size	[µm]
G	Sand wave growth rate	[m/yr]
Н	Sand wave height	[m]
Hd	Water depth	[m]
Hs	Significant wave height	[m]
L	Sand wave length	[m]
PI	Prediction Intervals	[m]
R	Correlation coefficient	[-]
R <sub>cr</sub>	Critical correlation coefficient	[-]
Rc versus H	Correlation coefficient between migration rate and sand wave height	[-]
Rc versus L	Correlation coefficient between migration rate and sand wave length	[-]
R <sub>c versus A</sub>	Correlation coefficient between migration rate and sand wave asymmetry	[-]
$R^2_{adj}$	Adjusted coefficient of multiple determination	[-]
Т	Wave period	[s]
$t_{1-\alpha,n-2}$	Critical value of the Student t-distribution for the n-2 degrees of freedom and significance level $\alpha$	[-]
U <sub>M2</sub>	Tidal velocity amplitude	[m/s]
V	Wind velocity	[m/s]
Х	Independent variable in regression analysis	[yr]
Ŷ	New independent variable in regression analysis	[yr]
X	Mean of the independent variable in regression analysis	[yr]
Y	Observed dependent variable in regression analysis	[m]
Ŷ	Predicted dependent variable in regression analysis	[m]
Ŧ	Mean of the dependent variable in regression analysis	[m]
θ <sub>wave</sub>	Wave direction	[°]
$\theta_{wind}$	Wind direction	[°]
σ	Standard deviation	[-]

#### Abbreviations

Cov[b]	Covariance of the regression coefficients
DFT	Discrete Fourier Transform
ED50	European Datum 1950
EPL	Europlatform
HSBL	Highest Seabed Level
LAT	Lowest Astronomical Tide
LSBL	Lowest Seabed Level
MSE	Mean Square of Error
SSR	Sum of squares of the regression
SST	Total sum of squares
UTM	Universal Transverse Mercator

#### 1. INTRODUCTION

In sandy coastal areas, various bedforms such as ripples, sand waves and sand banks can be found. Sand waves have shown to be the biggest threat to human interventions mainly due to their dimensions and migration activity (Morelissen et al., 2003; Drago et al., 2014).

Sand waves typically do not appear alone but in patches forming sand wave fields. Sand wave fields occur in water depths of approximately 10 to 50m and are induced by tidally driven currents (McCave, 1971; Knaapen et al., 2001; Németh et al., 2002). They are relatively big sea bed forms with wavelength and height dimensions, ranging from 100m up to 1km and from 1 to 10m, respectively. They occur as rhythmic patterns and they have migration rates up to several meters per year. They usually migrate with their crests perpendicular to the principal axis of the tidal ellipse of the dominant tidal constituent (Terwindt, 1971).

#### 1.1. PROBLEM DEFINITION

Sand waves pose a significant threat to economic activities that take place in offshore regions. Buried offshore pipelines and cables could be laid into sand wave fields. Due to the migration of the sand waves, pipelines or cables, that were positioned on a sand wave crest during the construction phase, could find itself in a trough later on. This process may create free spanning pipelines, see Figure 1-1. The change in sand wave asymmetry can also cause free spans (Nemeth et al., 2003), see Figure 1-1. There is a danger of damage to the exposed free span pipeline or cable.



Figure 1-1: Illustration of free span due to the migration and alteration in the asymmetry of sand waves (Nemeth et al., 2003) Pipeline is denoted by the dotted line and the sand wave is represented by the solid line.

Furthermore, offshore wind farms may be located in active seabed areas with significant sand wave migration. Therefore, the wind turbine foundations may be subjected to seabed lowering and rising. The consequences of a lowering seabed can be lowering of the pile fixation level for piled foundations (Deltares, 2017).

The study area is a sand wave field in the North Sea (more details about study area are given in Chapter 2) that future offshore structures could be built on.

Due to the strong growth in offshore wind energy, more and more locations of the Dutch continental shelf could be used for wind farms (RVO, 2015). Therefore, it is necessary to understand the morphodynamic behaviour of the area to ensure integrity of a potential structure during its design life.

#### 1.2. RESEARCH APPROACH

#### 1.2.1. RESEARCH OBJECTIVES AND RESEARCH QUESTIONS

In this study, we aim at a better understanding of the morphodynamics in the investigated area and creating a statistical model, that can predict seabed changes. Integrity management strategies and policies can be developed based on this model. The following research questions will be answered to reach the general objective:

RQ1: What is the magnitude of the main morphodynamic characteristics (sand wave height, sand wave length, asymmetry, migration rate of crests and growth rate) of the study area?

RQ2: Is there any correlation between migration rate and sand wave shape (sand wave height and length)?

RQ3: What is the most adequate model that can make accurate predictions of the vertical positions of sand wave crests and troughs?

RQ4: What is the magnitude of the maximum seabed rise and lowering of the study area over a period of 12 years?

#### 1.2.2. METHODOLOGY

In order to achieve the research objective and to answer the research questions, two methods have been used: literature review and data analysis.

Literature review:

- The purpose of the literature review is to understand the hydrodynamic and morphodynamic processes governing the formation of sand waves.
- Environmental parameters have been determined from the literature study.

Statistical analysis:

- A detailed data analysis of bathymetry is carried out and the investigated sand wave characteristics are determined.
- A regression analysis is carried out based on the finding of the morphodynamic analysis. The regression procedure has been employed to estimate future changes in the seabed levels.

#### 1.3. READING GUIDE

This report is organized as follows. In Chapter 2, the location of the study area is identified in regional scale and a brief description of the environmental conditions is presented. In chapter 3, the morphodynamic analysis is carried out and the main sand wave characteristics are presented. In chapter4, various uncertainties are addressed and vertical uncertainty bounds are proposed. Additionally, it is presented the mothodology, followed to generate a predictive model that can estimate future vertical position of sand wave crests and troughs. Finally, future seabed levels are estimated. Chapter 5 contains the discussion of uncertainties and assumptions made during this research. Finally, chapter 6 presents the conclusions by summarizing the answers to the research questions.

#### 2. STUDY AREA

This chapter presents the location and the magnitude of the key environmental parameters of the study area. The key environmental parameters are: tidal velocity amplitude ( $U_{M2}$ ), grain sizes ( $D_{50}$ ), water depths ( $H_d$ ), wind velocity (V), wind direction ( $\theta_{wind}$ ), significant wave height ( $H_s$ ), wave period (T) and wave direction ( $\theta_{wave}$ ).

#### 2.1. LOCATION OF THE STUDY AREA



Figure 2-1: The location of the study area and the Europlatform (EPL) in the North Sea (top right). The map shows the mean water depth of the Netherlands Continental Shelf relative to lowest astronomical tide (LAT) (Damen et al., 2018). A plan view of the sand wave field and the location of the investigated transect for the 2006 survey (top left). The cross section of the transect for the 2006 survey (bottom). The asterisk and the triangle indicate the start and end points of the transect respectively.

Data availability is important to the robustness of the seabed predictions. Thus, as location we have selected a sand wave field in the North Sea (ranging from 5757 to 5760km northing and from 500 to 502km easting in UTM ED50), where annual measurements of bathymetry are available, see Figure 2-1. Additionally, the investigated transect is orientated almost perpendicular to the sand wave crests. The location of the transect and the bottom profile along the cross-section for 2006 survey is given in Figure 2-1.

The selected area is situated in the vicinity of the Europlatform (EPL). The EPL is an offshore platform (5761km northing and 519km easting in UTM ED50) that mainly serves as beacon for ships, see Figure 2-1. It is also equipped with instruments that monitor the wind and wave conditions.

#### 2.2. ENVIRONMENTAL PARAMETERS OF THE STUDY AREA

#### Literature study

The tidal velocity and direction are key parameters of the study since they determine the transport and deposition of sediment. Tidal currents in the study area are mainly semi-diurnal with M2 as the dominant constituent. The M2 tidal velocity amplitude is around 0.7m/s according to the study of Damen et al. (2018).

Finally, the sand waves occur at locations with sandy grain sizes (Hulscher & Van den Brink, 2001) and the sand wave characteristics depend on grain size. The mean particle size at the study area is around 393µm (Damen et al., 2018).

#### Data analysis

The wind and wave parameters are defined based on the measurements collected at the ELP. Additionally, the mean water depth has been defined by the datasets provided by the Royal Dutch Navy. The Royal Dutch Navy conducted annual surveys covering the period from 2006 to 2018. More details regarding the surveys are given in section 3.1.

The box plots are used to define the mean of the datasets of Hs, T, V and Hd. The Figure 2-2 presents the various box plots and the defined mean values of the parameters. The mean values for the  $H_s$ , T, V and  $H_d$  are 1.2m, 4.4s, 6.2m/s and 33.5m, respectively.



Figure 2-2: Box plots of significant wave height (top left), wave period (top right), wind velocity (bottom left) and water depth (bottom right). The boxes show the variability of the data sets. The positions of the mean and median values are indicated with green diamonds and red horizontal lines respectively. The data points that exceed a distance of 1.5 times the height of the box from either end of the box are called extreme values and are shown with red crosses.

Finally, the prevailing wave and wind directions have been defined based on the frequencies of occurrence by direction. The rose graph displays the frequency of winds and waves coming from particular directions, see Figure 2-3. From the rose plot, it can be concluded that the dominant direction for both wind and wave is the South West (SW).



Figure 2-3: Frequency of occurrence of wave direction (red) and wind direction (blue) for the Europlatform area.

#### 2.2.1. SUMMARY OF FINDINGS

All the environmental parameters are summarized in Table 2-1.

Literature study				Data a	nalysis		
U <sub>M2</sub> in [m/s]	D₅₀ in [µm]	H₄in [m]	H₅ in [m]	T in [s]	θ <sub>wave</sub> in [-]	V in [m/s]	θ <sub>wind</sub> in [-]
0.7	393	33.5	1.1	4.3	SW	5.8	SW

Table 2-1: Summary of all environmental parameters

#### 3. ANALYSIS OF BATHYMETRIC DATA

This chapter describes the data analysis of the bathymetry. The current chapter consists of three parts: introduction, methods and results. The second part presents the methods used to process the data and how the dimension parameters of the sand wave characteristics, sand wave height (H), sand wavelength (L), sand wave asymmetry (A), migration (c) and growth rates (G), are defined. Lastly, the third part presents the outcomes obtained from the analysis.

#### 3.1. INTRODUCTION





Bathymetry is the study of the submarine topography (seabed). Nowadays, the seabed topography is mapped using acoustic sounding systems, see Figure 3-1. Acoustic sounding systems are usually mounted underneath a ship and they transmit and receive a fan of acoustic pulses (beams). The time it takes each beam to return to the sounders is accurately measured and these measurements are translated, using the speed of sound in water, to derive the bathymetric data (Simons & Snellen, 2014). Echo sounding is usually performed by single-beam and multibeam systems.

The datasets used here are from the Rijkswaterstaat. They have been also used for the definition of the median water depth in section 2.2.

#### 3.2. ESTIMATION OF SAND WAVE CHARACTERISTICS

#### 3.2.1. DATA PROCESSING

The signal, used to generate the seafloor, can be represented as the superposition of various amplitudes, wave numbers and phases different. The data processing is performed in order to investigate these signal parameters. In this study, the Discrete Fourier Transform (DFT) is used to separating the original signal of compound bed forms into signal of individual bed form types. The plot in Figure 3-2 reveals three clear groups of wavelengths that each represents a different bed form type, sand bank, sand waves and megaripples. The same method has been used in various studies (see Knaapen 2005; Van Dijk et al. 2008; Van Santen et al. 2011).

The next step in the analysis is to determine the low and high cutoff wavelengths. After trying various possible wavelengths, the 100m and 700m have been selected as the most adequate low and high cutoff wavelengths, respectively, see Figure 3-2. They have been selected because they minimize the number of the identified crests and trough points, but the filtered seabed profile still approaches the sand wave morphology well (Van Dijk et al. 2008).



Figure 3-2: Amplitude versus wave length of data set of survey 2006. The vertical dashed lines indicate the low and high cutoff wavelengths, used to filter bedforms with shorter and longer wavelengths, from the sand wave signal.

Finally, the Low and High pass Fourier Filters (MATLAB algorithms) are used to filter all the seabed topographies from 2006 to 2018. An example (2006 survey) of the resulting profiles is provided in Figure A-1 (Appendix A).

#### 3.2.2. DEFINITION OF THE SAND WAVE CHARACTERISTICS

The resulting filtered seabed are used to define the positions of crests and troughs. The positions of crest and trough points in all surveys (from 2006 up to 2018) have been determined by manual inspection of the filtered data. The selected positions for the year 2006 are presented in Figure 3-3. The sand waves are numbered from South West to North East.

It should be noted that filtering gives poor results on the edges and on crests of the sand wave, see Figure A-1 (Appendix A). Therefore, the vertical local maxima are retrieved from the raw, unfiltered data (Knaapen, 2005). By following this method, an underestimation of the sand wave heights due to the filtering procedure can be avoided. Finally, after defining the position of crests and troughs, the magnitude of the main sand wave characteristics, sand wave height (H), sand wavelength (L), migration rate of crests (c) and growth rate (G) can be estimated.



Figure 3-3: The positions of the sand wave crests and troughs for the survey 2006. The crests and troughs are indicated with red upward and blue downward pointing triangles respectively. The sand waves are numbered from South West to North East and the numbers assigned to each sand wave are included in the figure.

Additional graphs can be found in the Appendix A. Specifically:

- the seabed development between 2006 and 2018 is given in Figure A-2
- and the positions of sand wave crests and troughs throughout the year are presented in Figure A-3

#### Sand wave shape

Several definitions have been found in the literature regarding the sand wave dimension parameters. In this study, the shape characteristics of the sand waves have been determined similar to the method of Knaapen (2005), see Figure 3-4.



Figure 3-4: Definition of morphologic parameters (sand wave wavelength, height and asymmetry) (Knaapen, 2005).

The *length*, *L*, of a sand wave is calculated as:

$$L = |x_2 - x_1|$$

Where:

 $x_1$  refers to the horizontal coordinate of the trough which is located North East to the crest  $x_2$  refers to the horizontal coordinate of the trough which is located South West to the crest

Furthermore, the *height, H,* is estimated as follows:

$$H = z_{crest} - \frac{z_1 * L_2 + z_2 * L_1}{L}$$

Where:

 $\boldsymbol{z}_{crest}$  refers to the vertical coordinate of the crest

 $\mathbf{z}_1$  refers to the vertical coordinate of the trough which is located North East to the crest

 $\boldsymbol{z}_2$  refers to the vertical coordinate of the trough which is located South West to the crest

L1 is the horizontal distance between the crest and the trough which is located North East to the crest

L<sub>2</sub> is the horizontal distance between the crest and the trough which is located South West to the crest

Finally, the asymmetry, A, is defined as:

$$A = \frac{L_2 - L_1}{L}$$

Where:

 $L_1$  is the horizontal distance between the crest and the trough which is located North East to the crest  $L_2$  is the horizontal distance between the crest and the trough which is located South West to the crest L is the total length of the sand wave

The asymmetry has values between -1 and 1. Sand waves are symmetrical when A is zero. The positive sign indicates that the steeper side of the sand wave is toward the North East direction, see Figure 3-5.

#### Migration and growth rate

The *migration* describes the motion of the sand waves horizontally. The migration rate of the crests, c, is estimated by the differences in horizontal position of the same sand wave crest divided by the time difference between two successive surveys, see Figure 3-5. In this study, it is considered that the surveys have been conducted the same day of each year thus the time interval is 365 days. It should be stressed that the migration rates have been estimated using backwards in time method. For instance, the resulted migration rates have been obtained by subtracting the required measurements of the current year from the ones of the previous year. The crest of sand wave is moving in the North East direction when the migration is positive and in the opposite direction (South West) when the migration is negative, see Figure 3-5.



Figure 3-5: Method to define migration

Furthermore, *growth* rate, G, is defined as the difference of sand wave height divided by the time difference between two successive surveys. Like the migration rate, the backwards in time method has been used. The growth rate is positive when the sand wave height increases and negative when the height decreases.

#### 3.2.3. SUMMARY OF FINDINGS

The results are presented visually using box plots. Box plots have been chosen due to their simplicity and the fact that no assumptions or knowledge of the underlying statistical distribution of the data are needed. Particularly, the width of the box gives an indication of the variability of a data set. In addition, the position of the box in its whiskers (lines extending vertically from the box) and the position of the line across the box (the median) show if a data set is normally distributed or skewed. For instance, when the median is closer to the bottom of the box and the whisker is shorter on the lower end of the box, then the distribution is positively skewed (skewed right). Finally, extreme values are defined as data points that are located 1.5 times the width of the box below the lower end of the box and 1.5 times the width of the box.

The box plots can be classified into two groups. The first group includes the sand wave characteristics averaged per year over the sand wave field. Particularly, the ranges of sand wave length, height and asymmetry are depicted in Figure 3-6 and the ranges of the migration and growth rates are presented in Figure 3-7. The second group comprises the characteristics averaged per sand wave over the whole period. Specifically, the Figure 3-8 shows the box plots of sand wave length, height and asymmetry and the Figure 3-9 illustrates the migration and growth rates.



Figure 3-6: The top, middle and bottom plots display the annual ranges of sand wave length, height and asymmetry of the sand field, respectively. The mean and median values are indicated with green diamonds and red horizontal lines respectively.



Figure 3-7: The top and bottom plots show the annual ranges of migration and growth rate of the sand field, respectively. The mean and median values are indicated with green diamonds and red horizontal lines, respectively. In 2008, 2011, 2015, 2017 and 2018, the migration rates (top plot) appear to have no variability (boxes of zero width) since the vast majority of the data points is zero.



Figure 3-8: The top, middle and bottom plots display the ranges of sand wave length, height and asymmetry per individual sand wave, respectively, over the years 2006-2018. The mean and median values are indicated with green diamonds and red horizontal lines respectively. The sand waves are numbered from South West to North East.



Figure 3-9: The top and bottom plots give the rages of migration and growth rates per individual sand wave, respectively, over the years 2006-2018. The mean and median values are indicated with green diamonds and red horizontal lines, respectively. The sand waves are numbered from South West to North East. The migration rates (top plot) of the 1<sup>st</sup>, 3<sup>rd</sup>, 4<sup>th</sup>, 7<sup>th</sup>, 11<sup>th</sup> and 15<sup>th</sup> sand waves show no variability (boxes of zero width) since the vast majority of the data points is zero.

Below is a summary of the most important findings from the first group of box plots (see Figure 3-6 and Figure 3-7):

- During the study period, the yearly mean values of the sand wave length and height range from 188.4m to 190.6m and from 4.4m to 4.7m respectively.
- Most of the time, the sand wave field tends to migrate North East (positive sign) with migration rates ranging from 0.2 to 2.2m/yr. However, in the years 2007 and 2011, the sand wave field migrates South West with migration rates 0.44m/yr and 0.22m/yr respectively.
- The sand wave field is always asymmetrical oriented to the North East, meaning that the steeper slopes are facing North East (positive sign).
- The transect covers 17 sand waves. Neither generation nor extinction of sand waves have been observed. Therefore, the total number of sand waves remains 17 during the investigated period (from 2006 to 2018).

The key findings of the second group of box plots (see Figure 3-8 and Figure 3-9) are given below:

- The 3<sup>rd</sup> sand wave is the biggest with length and height equal to 381.6m and 7.3m respectively.
   On the other hand, the 14<sup>th</sup> sand wave is the smallest with length and height equal to 129m and 1.6m respectively.
- The biggest sand wave migrates with the slowest rate (0.31m/yr), while the smallest sand wave migrates with the fastest rate (1.86m/yr).
- All sand waves within the sand wave field migrate towards North East (positive sign).
- The majority of the sand waves presents positive asymmetry, northeast facing steep slope. Only 6 out of 17 sand waves show negative asymmetry, southwest facing steep slope.
- 8 out of 17 sand waves tend to grow. On the contrary, 7 out of 17 sand waves tend to decay and one sand wave remains stable.

In the Appendix A, it can be found summary tables with the average values of sand wave characteristics with a measure of uncertainty, the standard deviation. Specifically:

- the sand wave characteristics averaged per year over the sand wave field are summarized in Table A-1.
- the sand wave characteristics averaged per individual sand wave over the whole period are listed in Table A-2.

#### 3.3. ESTIMATION OF THE CORRELATION BETWEEN THE MIGRATION RATE AND THE SANDWAVE SHAPE

Herein, the correlation between the migration rate and the shape of the sand waves (sand wave height, length and asymmetry) is investigated. The datasets, which are used in the analysis, consist of the mean values per individual sand wave, see Table A-2 in Appendix A. Two method are used to analysis the strength of the relationships:

- The scatter plots have been created to visualise the relationship between the variables.
- The correlation coefficient (R) has been estimated to measure the strength and direction of the relationship between the variables. The formula suggested by Pearson (1896) has been used. The formula can be written:

$$R = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

Where

- x<sub>i</sub> and y<sub>i</sub> are the investigated variables
- $\bar{x}$  and  $\bar{y}$  are the mean values of the investigated variables

Next the significance of the correlation coefficient should be assessed. The estimated value of R is compared with the critical value  $R_{cr}$  given in Table 3-1. If the absolute value of the estimated R is greater than the critical value, then the statistical relationship between the variables is considered significant. In this study, the two-tailed test is used to determine significance at the 95% level ( $\alpha$ =0.05). The degree of freedom (df) is defined by subtracting 2 from the sample size (n). The critical value for df 15 (17-2), highlighted in yellow in Table 3-1, is 0.48.

Table 3-1: Table of Critical Values for Pearson's R. Only the positive critical values are tabulated in the table.

		Level of Sig	mificance of a T	Two-Tailed or M	Nondirectional T	`est
df	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.02$	$\alpha = 0.01$	$\alpha = 0.001$
1	0.9511	0.9877	0.9969	0.9995	0.9999	0.9999
2	0.8000	0.9000	0.9500	0.9800	0.9900	0.9990
3	0.6870	0.8054	0.8783	0.9343	0.9587	0.9911
4	0.6084	0.7293	0.8114	0.8822	0.9172	0.9741
5	0.5509	0.6694	0.7545	0.8329	0.8745	0.9509
6	0.5067	0.6215	0.7067	0.7887	0.8343	0.9249
7	0.4716	0.5822	0.6664	0.7498	0.7977	0.8983
8	0.4428	0.5494	0.6319	0.7155	0.7646	0.8721
9	0.4187	0.5214	0.6021	0.6851	0.7348	0.8470
10	0.3981	0.4973	0.5760	0.6581	0.7079	0.8233
11	0.3802	0.4762	0.5529	0.6339	0.6835	0.8010
12	0.3646	0.4575	0.5324	0.6120	0.6614	0.7800
13	0.3507	0.4409	0.5140	0.5923	0.6411	0.7604
14	0.3383	0.4259	0.4973	0.5742	0.6226	0.7419
15	0.3271	0.4124	0.4821	0.5577	0.6055	0.7247
16	0.3170	0.4000	0.4683	0.5425	0.5897	0.7084
17	0.3077	0.3887	0.4555	0.5285	0.5751	0.6932
18	0.2992	0.3783	0.4438	0.5155	0.5614	0.6788
19	0.2914	0.3687	0.4329	0.5034	0.5487	0.6652
20	0.2841	0.3598	0.4227	0.4921	0.5368	0.6524
21	0.2774	0.3515	0.4132	0.4815	0.5256	0.6402
22	0.2711	0.3438	0.4044	0.4716	0.5151	0.6287
23	0.2653	0.3365	0.3961	0.4622	0.5052	0.6178

#### 3.3.1. SUMMARY OF FINDINGS

The scatterplots and the trendlines have been created and displayed in Figure 3-10 (c versus H), Figure 3-11 (c versus L) and Figure 3-12 (c versus A).



Figure 3-10: Scatter plot of migration rate versus sand wave height. Trend is indicated by red line. The data consists of the mean values per individual sand wave averaged over the period 2006-2018.



Figure 3-11: Scatter plot of migration rate versus sand wave length. Trend is indicated by red line. The data consists of the mean values per individual sand wave averaged over the period 2006-2018.



Figure 3-12: Scatter plot of migration rate versus sand wave asymmetry. Trend is indicated by red line. The data consists of the mean values per individual sand wave averaged over the period 2006-2018.

The estimated correlation coefficient values (R) are summarized in the Table 3-2.

Table 3-2: Overview of the estimated correlation coefficients					
D	р	D			

R <sub>c versus H</sub>	R <sub>c versus L</sub>	R <sub>c versus A</sub>
-0.14	-0.21	-0.001

The main findings associated with the correlation between the migration rate and the shape parameters of the sand waves can be summarized as follows:

- A weak negative correlation between the migration rate and the sand wave height and length has been specified, see Figure 3-10, Figure 3-11.
- The correlation coefficient between the migration rate and sand wave length has the greatest absolute value among all, see Table 3-2.
- The critical value R<sub>cr</sub> has been determined equal to **0.48** (see Table 3-1), meaning that:
  - $R_{cr} > |R_{c \text{ versus } H}|$
  - $\quad R_{cr} > |R_{c \ versus \ L}|$
  - $\quad R_{cr} > |R_{c \, versus \, A}|$

Thus, the results revealed that the investigated correlations are statistically non-significant.

#### 3.4. CONCLUSIONS

This section draws conclusions based on the findings of the bathymetric analysis. The main conclusions are the following:

- The sand waves field retains more or less their shape during the investigated period.
- The outcomes of this study (sand wave length, L, sand wave height, H, and migration, c,) are compared with two previous studies (Van Dijk et al., 2011 Menninga 2012), carried out approximately at the same location, see Table 3-3. Additionally, all studies used the Fourier analysis to select the position of crests and troughs after truncating the signal of mega ripples and large scale bedforms.

Table 3-3: Comparison of the findings of current study with two other relevant studies. The table includes the characteristics of individual sand waves. Columns contain minimum, average and maximum values of the sand wave length, L, sand wave height, H, and the migration rate, c. The finding of the current study are in bold.

		L in [m]			H in [m	]		c in [m/	yr]
Studies	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
Van Dijk et al. (2011)	127	215	415	1.3	3.3	5.7	-1.6	0.6	4
Menninga (2012)	125	270	480	1.3	4.1	7.6	-1	0.4	1.4
This study	117	190	347	1.9	4.6	8.4	-0.4	0.7	1.5

From the Table 3-3, it can be concluded that the mean sand wave height obtained by this study is 1.3m greater than the outcomes of Van Dijk at et. (2011). A possible explanation can be that the current study estimated the sand wave height from the raw data (local vertical maxima), while Van Dijk at et. (2011) used the filtered seabed profiles to estimate the height. However, the mean sand wave height found in the study of Menninga (2012) is in close agreement with the results of the current study.

Furthermore, our average sand wavelength is smaller than the average lengths found in both studies. This difference may be explained by the fact that the investigated sites of Menninga (2012) and Van Dijk at et. (2011) are located at some distance to our site.

In terms of migration rate, the average migration rate calculated by Menninga (2012) and Van Dijk et al. (2011) agrees very well with the findings of this study.

• The sand wave field and the individual sand waves migrate usally in the direction of the steeper slopes. This is in agreement with the studies of Knaapen (2005) and Ernstsen et al. (2005) that found similar relation between the migration rate and steeper slope.

 Herein, the bathymetric analysis revealed that the lowest migration rate corresponds to the biggest sand wave and the greatest migration rate corresponds to the smallest sand wave. This is in agreement with the study of Németh et al. (2003). However, a weak negative correlation (not statistically significant) has been indentifed between the migration rate and sand wave length and height of the individual sand waves.

Additionally this study found no correlation between the migration rate and the sand wave asymmetry. Contrary to the study of Knaapen (2005) that showed strong positive correlation



Figure 3-13: The study of Knaapen (2005) found clear relation between the migration rate and the asymmetry of the sand waves

between the asymmetry and migration rate, see Figure 3-13. This difference could be related to the fact that Knaapen (2005) examined various sand wave fields situated at different location, assuming that the sand waves in the sand wave fields migrate as a group. While this study analysed individual sand waves within the same sand wave field (one location).

#### 4. FUTURE BATHYMETRIC CHANGES

This chapter provides a detailed description of the methodology for calculating future bathymetries and the corresponding seabed changes. Particularly, the first section addresses the various sources of uncertainty and suggests a vertical uncertainty band. The second and third sections present the methodology used to build the predictive model and estimate the maximum rising and lowering of the sand wave field respectively. The final part of this chapter interprets the findings of the analysis.

#### 4.1. MAGNITUDE OF UNCERTAINTY BANDS

Various uncertainties need to be taken into consideration in the assessment of the future bed level changes and the corresponding bathymetries. Knowledge of the source and magnitude of uncertainties is important for a meaningful interpretation of the outcomes and the usefulness of results in decision making processes.

#### 4.1.1. SOURCES OF UNCERTAINTIES

Three main sources of uncertainty can be distinguished: uncertainty due to data collection, uncertainty in the preprocessing of data and uncertainty due to the applied analysis methods (Deltares, 2015).

Echo-sounding measurements, which is used to generate the submarine topography, usually involve uncertainty. For instance, the intersection between the acoustic beam and the seafloor is called footprint, and the size of the footprint area significantly affects the uncertainty of the measurements (Monahan & Wells, 2002). As the size of the footprint area increases, the accuracy of the collected data decreases.

Furthermore, errors could be incurred as a result of the processing of the row echo sounding signals. Typical examples of such pre-processing are corrections for the movement of the vessel and for the tidal signal during the survey (Deltares, 2015).

The last source of uncertainty is associated with the choice of the analysis methods. For instance, the exact timing of the surveys is unknown and it has been assumed that they conducted on the same day of each year. This leads to an uncertainty of the estimates of the migration velocities. Furthermore, the surveys were performed once per year and therefore short-term changes of the seabed are averaged out. Temporal variation in the migration rate and the shape of sand waves may occur due to storm events (van Dijk & Kleinhans, 2005; Sterlini et al., 2012). The current analysis applies wide prediction bands that are sufficiently accounted for the short-term changes. Thus, this type of uncertainty can be ignored in the estimate of future seabed levels. Lastly, megaripples are often superimposed on the flanks of sand waves. Megaripples migrate faster than sand waves with average migration rate in order of 100m/yr (Morelissen et al., 2003; Knaapen, 2005). Their maximum length and height are approximately 30m and 1m respectively (Morelissen et al., 2003; Knaapen, 2005). Because of their fast movement, the larger megaripples may affect the foundation design or the burial depth of cables (Larsen et al., 2016).

#### 4.1.2. UNCERTAINTY BANDS

A band associated with the addressed sources of uncertainty should be applied in the analysis. The total vertical uncertainty band consists of the following contributions:

- Survey inaccuracies
- Migration of megaripples

The uncertainty related to the measurement and processing methods is considered equal to 0.18m (Knaapen, 2005; Deltares, 2015). The quantified uncertainty band is applied both upward and downward. The uncertainty bands associated with the migration of megaripples are 0.2m upwards and 0.15m downwards (Deltares, 2015).

#### 4.1.3. SUMMARY OF FINDINGS

The suggested vertical uncertainty bands are given in the Table 4-1.

Table 4-1: Summary of the uncertainty bands

	Uncertainty band upward in [m]	Uncertainty band downward in [m]
Survey inaccuracies	0.18	-0.18
Migration of megaripples	0.2	-0.15
Total	0.38	-0.33

#### 4.2. PREDICTIVE MODEL

In this study, the regression analysis is used to build an accurate predictive model. The input data to the predictive model consists of the yearly average values of the vertical positions of sand wave crests and troughs that has been specified in Chapter 3. The input data are given in Appendix B. Particularly, the datasets of vertical positions of sand wave crests and troughs are summarized in Table B-1 and Table B-2, respectively. For the purposes of the current analysis, the selected regression model needs to be compatible with a 12-year time-horizon (short term).

#### 4.2.1. SELECTION OF THE BEST REGRESSION MODEL

The first step in the analysis is to define the basic form of the regression model. The adjusted coefficient of multiple determination  $(R_{adj}^2)$  is used to specify the most adequate regression function.

4.2.1.1. ESTIMATION OF THE ADJUSTED COEFFICIENT OF MULTIPLE DETERMINATION Initially, the regression analysis is performed by incorporating three types of regression functions, the *first order polynomial, exponential* and *power* functions. The method of least squares has been used to estimate the optimal coefficients of each regression function. The basic idea behind the method is to find a function that minimizes the sum of the squares of the error. Error is the vertical difference between each observation and its true function (Kutner et al., 2005). More details about the method are given in Appendix B (section *Least Square Method*).

Next, the adjusted coefficient of multiple determination  $(R_{adj}^2)$  is estimated to define the best fitting function. Specifically, the  $R_{adj}^2$  measures how successful the variability of data can be explained by the fitted function (Kutner et al., 2005). Variation measures how far a set of random numbers are spread out from their average value (Kutner et al., 2005).

The  $R_{adj}^2$  is estimated as follows:

$$R_{adj}^2 = 1 - \frac{SSR}{SST}$$

Where:

• The SSR is the sum of squares of the regression and it is computed as:

$$SSR = \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2$$

In which:

Y<sub>i</sub> is the observed dependent value

- $\widehat{\boldsymbol{Y}}_i$  is the dependent value predicted by the regression function
- And SST is the total sum of squares and it estimated as:

$$SST = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

In which:

 $\overline{Y}$  is the mean of the observed dependent values

 $R_{adj}^2$  values vary between -1 and 1. Values closer to 1 indicate a good fit. While negative values indicate that the chosen regression line fits worse than a horizontal line.

#### 4.2.1.2. SUMMARY OF FINDINGS

The calculated values of  $R_{adj}^2$  of the possible regression functions are presented in Table 4-2 and Table 4-3. In particular, the Table 4-2 and Table 4-3 summarize the outcomes of the analysis of the position of sand wave crests against time and the analysis of the position of sand wave troughs against time respectively.

Table 4-2: Overview of the estimated  $R_{adj}^2$  of the potential regression functions between position of sand wave crests and time

Regression functions	R <sub>adj</sub> in [-]
$\hat{Y} = 0.015 \cdot X - 26.52$	0.242
$\hat{Y} = 0.00022 \cdot e^{0.005X}$	0.241
$\hat{Y} = 0.000001 \cdot X^{9.6}$	0.242

The predicted dependent variable is denoted by  $\hat{Y}$  and the independent variable is denoted by X

Table 4-3: Overview of the estimated  $R_{adj}^2$  of the potential regression functions between position of sand wave troughs and time

Regression functions	R <sub>adj</sub> in [-]
$\hat{Y} = -0.002 \cdot X + 1.8$	0.091
$\hat{Y} = -0.178 \cdot e^{0.0011X}$	0.089
$\hat{Y} = 0.0000007 \cdot X^{2.18}$	0.091

The predicted dependent variable is denoted by  $\hat{Y}$  and the independent variable is denoted by X

The results reveal that:

- The  $R_{adj}^2$  is almost the same for all the investigated regression functions between position of sand wave crests and time. The  $R_{adj}^2$  is equal to 0.24 which means that the functions can explain about 24% of the variation of the data.
- Similarly, the same  $R_{adj}^2$  values have been computed for all the investigated regression functions between position of sand wave troughs and time. In this case the estimated  $R_{adj}^2$  is lower and equal to 0.09 (<0.24).
- The very low values of  $R^2_{adj}$  indicate that the regression functions do not fit the data very well.

#### 4.2.2. EXTRAPOLATION OF THE BEST REGRESSION MODEL

The linear regression model has been selected to quantify the relationship between independent and dependent variables. The general linear function is given by:

$$\hat{Y} = b_0 + b_1 X$$

Where:

- Ŷ denotes the predicted dependent variable,
- X denotes the independent variable
- b<sub>0</sub> and b<sub>1</sub> are the regression coefficients as derived previously in section 4.2.1.2

The best-fitting linear model for the measurements is extrapolated into the future. The basic assumption of the extrapolation method is that the patterns of the past observations continue into the future (Kutner et al., 2005).

#### 4.2.2.1. ESTIMATION OF PREDICTION INTERVALS

The basic idea of prediction intervals is to estimate a range in the distribution of the predicted dependent variable  $(\hat{Y})$ , in which a new data point (predicted) will fall. It is assumed that  $\hat{Y}$  follows a Student t distribution. The prediction intervals must account for (Kutner et al., 2005):

- Variation in the value of the mean of the distribution of  $\hat{Y}$ .
- Variation within the distribution of  $\widehat{Y}$ .

In this study, the prediction intervals have been estimated based on the formulas suggested in Kutner et al. (2005) as follows:

$$PI = \hat{Y} \pm t_{1-a,n-2} \sqrt{MSE + \hat{X}^T \hat{X} Cov[b]}$$

Where:

- t<sub>1-a,n-2</sub> is the critical value for the Student t-distribution for the n-2 degrees of freedom and significance level α. Where n is the population size. The critical t values can be found in the Table B-3 (Appendix B). In this study, the confidence level set to 95% (α=0.05). The 95% confidence level means that future values fall into the prediction envelope 95% of the time.
- MSE is the mean square of error that can be estimated as follows:

$$MSE = \frac{\sum_{i}^{n} (Y_{i} - \widehat{Y}_{i})^{2}}{n - 2}$$

•  $\widehat{X}$  and  $\widehat{X}^{T}$  are the vectors of new (future) independent values, defined as:

$$\widehat{\mathbf{X}} = \begin{bmatrix} \mathbf{1} \\ \widehat{\mathbf{X}}_1 \\ \vdots \\ \vdots \\ \widehat{\mathbf{X}}_n \end{bmatrix} \text{ and } \widehat{\mathbf{X}}^T = \begin{bmatrix} \mathbf{1} & \widehat{\mathbf{X}}_1 & \widehat{\mathbf{X}}_2 & \cdots & \widehat{\mathbf{X}}_n \end{bmatrix}$$

T symbol denotes the transpose of the matrix (exchange the rows and columns)

• Cov[b] is the covariance of the b parameters that can be calculated as:

$$Cov[b] = \frac{-MSE\,\bar{X}^2}{\sum_{1}^{n}(X_i - \bar{X})^2}$$

In which:

 $\overline{\mathbf{X}}$  is the mean of the independent values

More information about the formulas, used in this section, can be found in Appendix B (section *Estimation of prediction intervals*).

4.2.2.2. SUMMARY OF FINDINGS

The best fitted regression models have been defined and presented as:

- equations of the regression functions are presented in Table 4-4
- and the regression curves are illustrated in Figure 4-1.

Table 4-4: Summary of the most suitable regression functions.

Datasets	Linear regression model
Averaged vertical position of sand wave crests versus time	$\hat{Y} = 0.015X - 26.52$
Averaged vertical position of sand wave troughs versus time	$\hat{Y} = -0.002X + 1.8$



Figure 4-1 The graphs illustrate the scatterplots and the best-fit regression functions. The top and bottom graphs show the average vertical position of sand wave crests versus time and the average vertical position of sand wave troughs versus time respectively.

Lastly, the extrapolated regression models and the estimated prediction bounds for vertical position of sand waves crests versus time and position of sand wave troughs versus time are depicted in Figure 4-2 and Figure 4-3, respectively. The predicted average minimum and maximum vertical positions of crests and troughs for the year 2030 are summarized in Table 4-5.

Table 4-5: Overview of the minimum and maximum	n predicted values in 2030
--	----------------------------

Average vertic sand wave o	al position of crests in [m]	Average vertic wave tro	al position of sand oughs in [m]
Max	Min	Max	Min
3.65	3.08	-1.54	-1.68



Figure 4-2: Extrapolation of the regression equation and prediction bands of the model (average vertical position of sand wave crests against time). The red circles highlight the minimum and maximum predicted values in 2030.



Figure 4-3: Extrapolation of the regression equation and prediction bands of the model (average vertical position of sand wave troughs against time). The red circles highlight the minimum and maximum predicted values in 2030.

#### 4.3. ESTIMATION OF FUTURE SEABED LEVEL RISING AND LOWERING

In the current section, the future bathymetries and the corresponding seabed changes are estimated over the considered 12 years period. The analysis is based on the predicted values of the mean vertical position of sand wave crests and troughs and the uncertainty bounds. The datasets included in the analysis are:

- The maximum predicted vertical position of sand wave crests and the minimum predicted vertical position of the sand wave trough, see Table 4-5
- The suggested vertical uncertainty bands see Table 4-1.

#### Maximum rise of the seabed for time span 12 year

The maximum seabed level (HSBL) in the year 2030 is equal to the maximum predicted vertical position of sand wave crests plus the upward uncertainty band. Thus, the HSBL is calculated as follows:

• 3.65 + 0.38 = 4m

By calculating the difference between the HSBL and the most recent bathymetry (2018), the maximum rise of the seabed can be determined, as follows:

• 4 - 3.15 = 0.85m

#### Maximum lowering of the seabed for time span 12 year

Similarly, the lowest seabed level (LSBL) can be estimated by adding the minimum predicted vertical position of sand wave troughs in the year 2030 and the downwards uncertainty band. Therefore, the LSBL is estimates as follows:

• -1.68 - 0.33 = - 2m

The maximum seabed lowering can be calculated by subtracting the LSBL from the 2018 bathymetry, as follows:

• -2 - (-1.6) = 0.4

#### 4.3.1. SUMMARY OF THE FINDINGS

The predicted maximum rise and lowering of the seabed with respect to the bathymetry of 2018 are presented in Table 4-6.

Table 4-6: Predictions of seabed level changes with respect to the bathymetry of 2018

Maximum seabed level rising in [m]	Maximum seabed level lowering in [m]
0.9	-0.4

#### 4.4. CONCLUSIONS

On the basis of the findings, it can be observed that:

- Out of the three types of regression models, the linear regression model is selected because it is simple and readily extendible over time (Harrell & Frank, 2015).
- The magnitudes of the R<sup>2</sup><sub>adj</sub> values indicate that the best-fitting linear functions are not very good predictors of future vertical position of sand wave crests and troughs. Thus, the prediction bounds based on the linear regression model has been used to define maximum and minimum future seabed levels.

#### 5. DISCUSSION

A change in the shape of the sand wave can cause a threat to the safety of offshore structures. This may happen, for example, when the asymmetry of the sand wave changes (Nemeth et al., 2003). This study assumes that all the sand waves included in the investigated sand wave field move vertically as a group. Therefore, the asymmetry variations of the individual sand waves have not been taken into account.

In this study, future seabed levels have been calculated based on the prediction bands. Prediction bands have been selected since the linear regression function presents significant deviation. However, this approach could be considered quite conservative, leading to overestimating of potential seabed levels (Harrell & Frank, 2015). In addition, model conservatism has been increased by selecting significant level 95% and adding uncertainty bounds. Consequently, conservatism results in costly designs, leading to investment hesitation on behalf of the stakeholders.

The proposed prediction model has been created without taking into account environmental parameters such as sediment characteristics, tidal velocities and wave characteristics. However, the effect of the various environmental parameters is already included in the shape characteristics and migration activities (Knaapen, 2005).

This study assumed that no significant change in the yearly average horizontal position of the sand wave crests will occur over a period of 12 years (prediction time horizon), and consequently the specific variable (horizontal position of sand wave crests) has not been included in the prediction model. This assumption is based on the fact that the annual mean migration rates of the sand wave field are low, ranging from -0.4 to 2.2m/yr (see Table A-1). Therefore, the proposed modelling approaches cannot be used for the estimation of potential seabed levels outside of the study area.

Herein, the low and high cutoff wavelengths have been selected equal to 100m and 700m, respectively. The chosen cutoff wavelengths eliminate the crests and trough points and the resulted seabed profile approaches adequately the sand wave morphology (Van Dijk et al. 2008). The aforementioned wavelengths have been selected after testing various wavelengths (sensitivity tests). From the sensitivity tests, it can be pointed out that the sand wave length and height were affected slightly from changing the cutoff wavelengths. However, the migration rate was more sensitive to the wavelengths. Specifically, we found that declining the low cutoff wavelength the migration rate increases. Lastly, all the tested cutoff wavelength resulted in the underestimation of the vertical position of the crests.

#### 6. CONCLUSION

In this chapter, the main findings with regard to the research questions are summarised and general conclusions are described.

### RQ1: What is the magnitude of the main morphodynamic characteristics (sand wave height, sand wave length, asymmetry, migration rate of crests and growth rate) of the study area?

It has been analysed the bathymetric surveys from 2006 to 2018 at the Europlatform area (North Sea) to study the key morphodynamic characteristics. The results suggest that the sand wave field retains more or less its shape during the investigated period, with total average sand wave length and height equal to 190m and 4.6m respectively. The field migrates with total average rate 0.7m/yr to the North East, coinciding the steeper slopes.

The individual sand waves present heights ranging between 1.9 and 8.4m and wavelengths varying between 117 and 347m. Most sand waves are asymmetrical with values varying from -0.4 to 0.8. Furthermore, all the sand waves migrate with rates varying between 0.3 and 1.9m/yr in directions to the North East. Most of the time, the steeper slopes face the North East direction.

## RQ2: Is there any correlation between migration rate and sand wave shape (sand wave height and length)?

The correlation between the migration rate and the sand wave shape of the individual sand waves has been investigated. The data used in the analysis consist of the mean values of individual sand waves, averaged over the whole period. The results of the data analysis reveal a weak negative correlation between the migration rate and the sand wave height and length. Lastly, it has been found that the sand wave asymmetry does not correlate with the migration rate.

## RQ3: What is the most adequate model that can make accurate predictions of the vertical positions of sand wave crests and troughs?

The linear model has been selected mainly because it is simply and readily extendible. However, the selected model cannot adequately fit the data and consequently the extrapolation of the model could give poor predictions. Therefore, the prediction bounds of the extrapolated linear regression model are used to define the maximum and minimum vertical positions of sand wave crest and troughs.

## RQ4: What is the magnitude of the maximum seabed rise and lowering of the study area over a period of 12 years?

The future bathymetries and the corresponding seabed level rising and lowering are estimated based on the outcomes of the prediction model and the proposal uncertainty bounds. The maximum rise and lowering of the seabed in 2030, with respect to the 2018 bathymetry, have been estimated equal to 0.9m and 0.4m, respectively.

Overall, the suggested model could provide practical support for the design of offshore structures at the study area. The magnitude of the proposed highest and lowest seabed levels could be used to determine the optimal burial depth of pipelines/cables and the extend and depth of scour protection around the monopile foundations. In fact:

- Due to the sand wave migration, the initial burial depth (the vertical distance from the seabed to the top of the pipe/cable) should be increased to guarantee the safety of the submarine pipes/cables (Deltares, 2015). The definition of the extra burial depth can be based on the outcomes of the prediction model.
- The role of the scour protection is to ensure a constant fixation depth of the monopile. The scour protection should be able to cope with the predicted seabed level lowering. Specifically, the edges of the scour protection should be sufficiently flexible (consist of rock materials able to roll down easily) to stabilise the slope and prevent further lowering (Deltares, 2017). Furthermore, if the extent of the scour protection is sufficiently large, the amount of soil remaining around the foundation can ensure the integrity of the scour protection (Deltares, 2017). The outcomes of the prediction model can help the decision-makers to decide the diameter of the rock material and the optimal length of the scour protection.

However, as already mentioned in the discussion section the model is not valid outside of the study area.

#### REFERENCES

- Damen, J. M., Van Dijk, T. A. G. P., & Hulscher, S. J. M. H. (2018). Spatially varying environmental properties controlling observed sand wave morphology. *Journal of Geophysical Research: Earth Surface*, *123*.
- Deltares. (2015). Morphodynamics of Borssele Wind Farm Zone.
- Deltares. (2017). Scour and scour mitigation for Hollandse Kust (zuid).
- Drago, M., Mattioli, M., Bruschi, R., & Vitali, L. (2014). Insights on the design of free-spanning pipelines. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 373.
- Ernstsen, V. B., Noormets, R., Winter, C., Hebbeln, D., Bartholomä, A., Flemming, B. W., & Bartholdy, J. (2005). Development of subaqueous barchanoid-shaped dunes due to lateral grain size variability in a tidal inlet channel of the Danish Wadden Sea. *Journal of Geophysical Research: Earth Surface, 110.*
- Harrell, J., & Frank, E. (2015). General aspects of fitting regression models. In Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Springer.
- Hulscher, S. J. M. H., & Van den Brink, G. M. (2001). Comparison between predicted and observed sand waves and sand banks in the north sea. *Journal of Geophysical Research: Oceans*.
- Knaapen, M. A. F. (2005). Sandwave migration predictor based on shape information. *Journal of Geophysical Research: Earth Surface*, *110*.
- Knaapen, M. A. F., Hulscher, S. J. M. H., Vriend, H. J., & Stolk, A. (2001). A new type of sea bed waves. *Geophysical Research Letters*, 28.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models* (Vol. 5). McGraw-Hill Irwin Boston.
- Larsen, S. M., Roulund, A., Brooks, A. J., & Chaffey, A. (2016). Sandwaves and megaripples at Race Bank (UK) offshore windfarm. In Scour and Erosion: Proceedings of the 8th International Conference on Scour and Erosion (Oxford, UK, 12-15 September 2016).
- McCave, I. N. (1971). Sand waves in the North Sea off the coast of Holland. Marine Geology, 10.
- Monahan, D., & Wells, D. (2002). The Use of the International Hydrographic Organisation's' Standards for Hydrographic Surveys' As a Measure of Depth Accuracy in Continental Shelf Determinations. *The International Hydrographic Review*, *3*(1).
- Morelissen, R., Hulscher, S. J. M. H., Knaapen, M. A. F., Németh, A. A., & Bijker, R. (2003). Mathematical modelling of sand wave migration and the interaction with pipelines. *Coastal Engineering*.
- Németh, A. A., Hulscher, S. J. M. H., & de Vriend, H. J. (2002). Modelling sand wave migration in shallow shelf seas. *Continental Shelf Research*, 22.
- Nemeth, A., Hulscher, S. J. M. H., & de Vriend, H. J. (2003). Offshore sand wave dynamics, engineering problems and future solutions. *Pipeline and Gas Journal, 230*.
- Pearson, K. (1896). Mathematical Contributions to the Theory of Evolution. III. Regression. *Heredity and Panmixia. Philosophical.*

- RVO (2015). Offshore wind energy in the Netherlands: The roadmap from 1,000 to 4,500 MW offshore wind capacity, *Netherlands*
- Simons D. G., & Snellen M. (2014). Lecture notes: Seafloor mapping. Delft University of Technology.
- Sterlini, F., van Dijk, T. A. G. P., IJzer, S., & Hulscher, S. J. M. H. (2012). Seasonal changing sand waves and the effect of surface waves. In *the 5th International Short Conference on Applied Coastal Research, SCACR 2011, Aachen, Germany.*
- Terwindt, J. H. . J. (1971). Sand waves in the southern bight of the North Sea. Marine Geology, 10.
- van Dijk, T. A. G. P., & Kleinhans, M. G. (2005). Processes controlling the dynamics of compound sand waves in the North Sea, Netherlands. *Journal of Geophysical Research: Earth Surface*, *110*.
- Van Dijk, T. A. G. P., Lindenbergh, R. C., & Egberts, P. J. P. (2008). Separating bathymetric data representing multiscale rhythmic bed forms: A geostatistical and spectral method compared. *Journal of Geophysical Research: Earth Surface*, *113*.
- Van Dijk, T. A. G. P., Van der Tak, C., De Boer, W.P., Kleuskens, M.H.P., Doornenbal, P. J. & Noorlandt,
   R. P (2011). The scientific validation of the hydrographic survey policy of the Netherlands
   Hydrographic Office, Royal Netherlands Navy
- Van Santen, R. B., De Swart, H. E., & van Dijk, T. A. G. P. (2011). Sensitivity of tidal sand wavelength to environmental parameters: A combined data analysis and modelling approach. *Continental Shelf Research*, 31.

#### APPENDIX A: BATHYMETRIC DATA ANALYSIS

The defined cutoff wavelengths have been converted to frequencies and the low and high Fourier pass have been applied in the raw bathymetric data. An example of the filtering process is given in Figure A-1, which presents the results of the method of the survey 2006.







Figure A-2: The seabed development between 2006 and 2018. The filtered seafloor profiles of the surveys of 2006 and 2018 are depicted with solid and dashed lines respectively. The blue upward-pointing triangles indicate the crests and the red solid circles indicate the troughs. The sand waves are numbered from South West to North East and the numbers assigned to each sand wave are included in the figure.



Figure A-3: Position of sand wave crests (yellow) and troughs (blue) over the years.

Mean values of the investigated parameters with their standard deviations,  $\sigma$ , are summarised in Table A-1 (yearly average values) and in Table A-2 (average values per individual sand waves).

	Н	$\sigma_{H}$	L	$\sigma_L$	Α	$\sigma_{As}$	С	σ	Gr	$\sigma_{Gr}$
Year	[m]	[m]	[m]	[m]	[-]	[-]	[m/yr]	[m/yr]	[m/yr]	[m/yr]
2006	4.5	1.3	189.1	52.0	0.13	0.3	_	1	_	_
2007	4.5	1.3	189.5	51.6	0.12	0.3	-0.44	3.0	0.02	0.1
2008	4.7	1.3	190.2	51.2	0.10	0.3	1.31	2.5	0.15	0.2
2009	4.5	1.2	188.4	51.0	0.11	0.3	1.53	1.8	-0.17	0.2
2010	4.5	1.2	189.1	53.4	0.11	0.3	1.53	2.7	-0.07	0.1
2011	4.5	1.2	189.7	56.3	0.15	0.4	-0.22	1.2	0.04	0.3
2012	4.4	1.2	189.5	57.2	0.16	0.4	0.65	3.1	-0.09	0.2
2013	4.5	1.3	189.5	56.2	0.18	0.3	0.00	5.1	0.11	0.3
2014	4.7	1.3	190.6	58.1	0.19	0.3	2.18	4.2	0.24	0.1
2015	4.7	1.3	189.5	57.0	0.19	0.3	0.65	1.5	-0.07	0.1
2016	4.7	1.3	189.1	56.2	0.20	0.3	0.65	2.3	0.04	0.1
2017	4.7	1.3	189.1	54.1	0.19	0.3	0.22	2.6	0.01	0.1
2018	4.7	1.1	189.1	53.3	0.18	0.3	0.44	2.0	-0.06	0.1

Table A-1: Annual mean values, averaged over the field

Table A-2: Mean values of individual sand waves, averaged over the period

Sand	н	σн	L	$\sigma_{L}$	Α	σ <sub>As</sub>	С	σ	Gr	$\sigma_{Gr}$
wave number	[m]	[m]	[m]	[m]	[-]	[-]	[m/yr]	[m/yr]	[m/yr]	[m/yr]
1	5.4	0.2	237.8	9.1	0.25	0.05	1.24	2.5	-0.01	0.27
2	4.2	0.4	165.6	6.8	0.44	0.08	0.31	2.8	0.06	0.46
3	8.4	0.3	346.9	6.2	-0.20	0.03	0.31	1.5	0.04	0.49
4	4.6	0.3	211.8	10.1	-0.07	0.06	0.93	4.0	-0.03	0.33
5	4.7	0.3	180.1	7	0.44	0.06	1.55	4.1	-0.02	0.31
6	2.7	0.2	146.5	7.9	-0.18	0.09	0.31	2.8	-0.02	0.21
7	5.9	0.4	225.0	10.8	0.44	0.03	0.31	2.3	0.12	0.28
8	4.3	0.2	179.9	6.2	0.42	0.06	0.31	3.1	0.03	0.32
9	5.2	0.2	154.2	10	0.38	0.06	0.93	1.3	-0.02	0.32
10	4.0	0.2	140.7	3.7	-0.40	0.11	0.93	2.1	-0.04	0.40
11	5.5	0.3	233.2	8.7	-0.12	0.05	0.00	3.2	-0.02	0.40
12	3.9	0.2	150.2	10.4	0.24	0.07	0.31	3.7	0.00	0.31
13	3.4	0.3	126.2	4.8	0.12	0.09	1.24	4.6	0.07	0.40
14	1.8	0.1	117.3	7.9	0.33	0.12	1.86	1.9	0.02	0.12
15	6.4	0.3	202.1	10.4	0.16	0.07	0.62	1.1	-0.01	0.37
16	4.3	0.4	261.8	13.5	0.80	0.12	0.00	3.4	0.02	0.53
17	3.2	0.2	140.5	9.2	-0.42	0.08	0.93	1.8	0.01	0.29

#### APPENDIX B: REGRESSION ANALYSIS

#### Data used in the regression analysis

A regression analysis is performed between the vertical position of sand wave crests/troughs (dependent variables) and time (independent variable). The vertical position of sand wave crests and troughs have been estimated by the bathymetric analysis and their yearly mean values are presented in the Table B-1 and Table B-2, respectively.

Years	Mean vertical position of sand wave crests
	[m]
2006	3.00
2007	3.05
2009	3.06
2008	3.24
2010	3.00
2011	3.05
2012	2.96
2013	3.01
2014	3.23
2015	3.18
2016	3.22
2017	3.23
2018	3.15

Table B-1: Yearly average values of the vertical positions of sand waves crests

Table B-2: Yearly average values of vertical positions of sand waves troughs

	Mean vertical position of					
Years	sand wave troughs					
	[m]					
2006	-1.61					
2007	-1.58					
2009	-1.58					
2008	-1.55					
2010	-1.55					
2011	-1.53					
2012	-1.58					
2013	-1.60					
2014	-1.61					
2015	-1.58					
2016	-1.59					
2017	-1.59					
2018	-1.60					

#### **Least Squares Method**

The data consists of n paired observations of the predictor variable  $X_i$  and the predicted variable  $Y_i$ . The fitted regression equation yields:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Consider now writing the equation for each measurement (n):

$$Y_1 = \beta_0 + \beta_1 X_1 + \varepsilon_1$$
$$Y_2 = \beta_0 + \beta_1 X_2 + \varepsilon_2$$
$$\vdots \vdots \vdots$$
$$Y_n = \beta_0 + \beta_1 X_n + \varepsilon_n$$

The above equations can be grouped into the following matrices:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ \vdots \\ \beta_0 + \beta_1 X_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

or

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

In matrix terms:

$$Y = X\beta + \varepsilon$$

Where:

- Y is the predicted vector
- X is the predictor matrix
- β is the vector of parameters
- $\boldsymbol{\epsilon}$  is the error vector or deviation from the linear regression function

The sum of square error (SSE) is defined as follows:

$$SSE = \sum_{i=1}^{n} \varepsilon_i^2$$

Or in vector format:

$$SSE = \varepsilon \varepsilon^{T}$$
$$= (Y - X\beta)^{T} (Y - X\beta)$$
$$= (Y^{T} - \beta^{T} X^{T})(Y - X\beta)$$

$$= (Y^{T}Y - Y^{T}X\beta - \beta^{T}X^{T}Y + \beta^{T}X^{T}X\beta)$$
$$= (Y^{T}Y - 2\beta^{T}X^{T}Y + \beta^{T}X^{T}X\beta)$$

where the T symbol denotes the transpose of the matrix (exchange the rows and columns). Notice that  $\mathbf{Y}^{T}\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}^{T}\mathbf{X}^{T}\mathbf{Y}$  (1×1 matrix)

#### Estimation of the regression coefficients b

The optimum regression coefficient vector **b** can be calculated by setting the derivative of SSE with respect to  $\varepsilon$  equal to zero.

$$\frac{dSSE}{d\varepsilon} = \left(\frac{d}{d\varepsilon}(\mathbf{Y}^T\mathbf{Y}) - 2\frac{d}{d\varepsilon}(\mathbf{b}^T\mathbf{X}^T\mathbf{Y}) + \frac{d}{d\varepsilon}(\mathbf{b}^T\mathbf{X}^T\mathbf{X}\mathbf{b})\right)$$
$$= (0 - 2\mathbf{X}^T\mathbf{Y} + 2\mathbf{X}^T\mathbf{X}\mathbf{b})$$
$$= (\mathbf{X}^T\mathbf{X}\mathbf{b} - \mathbf{X}^T\mathbf{Y}) = 0$$

The final form:

$$X^T X b = X^T Y$$

The solution of the above equation gives the least squares regression coefficients:

$$\boldsymbol{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

X<sup>T</sup>X can be expressed as:

$$\boldsymbol{X}^{T}\boldsymbol{X} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{1} & X_{2} & \dots & X_{n} \end{bmatrix} \begin{bmatrix} 1 & X_{1} \\ 1 & X_{2} \\ \vdots & \vdots \\ 1 & X_{n} \end{bmatrix} = \begin{bmatrix} n & \sum_{i}^{n} X_{i} \\ \sum_{i}^{n} X_{i} & \sum_{i}^{n} X_{i}^{2} \end{bmatrix}$$

**X<sup>T</sup>Y** is defined as:

$$\mathbf{X}^{\mathsf{T}}\mathbf{Y} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum_{i}^n Y_i \\ \sum_{i}^n X_i Y_i \end{bmatrix}$$

And  $\mathbf{X}^{T}\mathbf{X}\mathbf{b} = \mathbf{X}^{T}\mathbf{Y}$  can be rewritten as:

$$\boldsymbol{X}^{T}\boldsymbol{X}\boldsymbol{b} = \boldsymbol{X}^{T}\boldsymbol{Y} = > \begin{bmatrix} n & \sum_{i=1}^{n} X_{i} \\ \sum_{i=1}^{n} X_{i} & \sum_{i=1}^{n} X_{i}^{2} \end{bmatrix} \begin{bmatrix} b_{0} \\ b_{1} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} Y_{i} \\ \sum_{i=1}^{n} X_{i} Y_{i} \end{bmatrix}$$

Therefore, the  $b_0$  and  $b_1$  coefficient are:

$$b_1 = \sum_{1}^{n} (X_i - \bar{X}) (Y_i - \bar{Y}) / \sum_{1}^{n} (X_i - \bar{X})^2$$

And

$$b_0 = \bar{Y} - b_1 \bar{X}$$

Where  $\overline{X}$  and  $\overline{Y}$  are the mean values of the predictor and predicted vectors, respectively. The scalar form yields:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$
 and  $\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$ 

#### The variance-covariance matrix of b

The assumptions for the simple linear regression model are:

Assumption 1: fixed regressors.

All elements of the  $n \times 2$  matrix **X** are considered fixed in repeated samples (non-stochastic).

$$E[X] = XX^T$$

Assumption 2: random disturbances,  $\varepsilon_i$ , with zero mean.

The n×1 vector  $\boldsymbol{\varepsilon}$  consists of random disturbances (errors) with zero mean. Symbolically, we have:

$$\frac{1}{n}\sum_{1}^{i}\varepsilon_{i}=0 \quad \text{or} \quad E[\varepsilon_{i}]=0$$

Assumption 3: homoskedasticity or equal variance of  $\varepsilon_i$ .

The diagonal elements of the variance-covariance matrix of the disturbances  $E[\varepsilon \varepsilon^T]$  are equal to  $\sigma^2$ . They are called variance and they can be written:

$$E[\varepsilon\varepsilon^T] = E[\varepsilon_i^2] = \sigma^2$$

Assumption 4: no correlation between the  $\varepsilon_i$ 

The off-diagonal elements of the variance-covariance matrix of the disturbances  $E[\varepsilon \varepsilon^T]$  are called covariance and they are all equal to zero. Symbolically covariance can be written:

$$\mathbf{E}[\varepsilon\varepsilon^T] = \mathbf{E}[\varepsilon_i\varepsilon_j] = 0$$

Assumption 5: normality

- The disturbances  $\varepsilon_i$  have a normal distributed. •
- The predicted variable  $Y_i$  and the disturbances,  $\varepsilon_i$  are jointly normally distributed. The joint normality implies that the assumption 2, 3 and 4 are valid for Y<sub>i</sub>.

Taking into account the assumption, the variance-covariance matrix of **b** can be computed as

.

$$b = (X^T X)^{-1} X^T Y \implies$$

$$E[bb^T] = (X^T X)^{-1} X^T E[YY^T] X (X^T X)^{-1}$$

$$= (X^T X)^{-1} X^T (X\beta + \varepsilon)$$

$$= (X^T X)^{-1} X^T \sigma^2 X (X^T X)^{-1}$$

$$= \sigma^2 (X^T X)^{-1}$$

Or

$$E[\boldsymbol{b}\boldsymbol{b}^{T}] = \begin{bmatrix} \frac{\sigma^{2}}{n} + \frac{\sigma^{2}\bar{X}^{2}}{\sum_{1}^{n}(X_{i} - \bar{X})^{2}} & \frac{-\sigma^{2}\bar{X}^{2}}{\sum_{1}^{n}(X_{i} - \bar{X})^{2}} \\ \frac{-\sigma^{2}\bar{X}^{2}}{\sum_{1}^{n}(X_{i} - \bar{X})^{2}} & \frac{\sigma^{2}}{\sum_{1}^{n}(X_{i} - \bar{X})^{2}} \end{bmatrix}$$

#### Fitted vector

The vector of fitted (predicted) values is:

 $\widehat{Y} = Xb$ 

#### **Residual vector**

The vector of residuals, e, is:

$$e = Y - \hat{Y}$$

An unbiased estimator of the error variance is the mean square of error (MSE):

$$MSE = \frac{\sum_{i}^{n} \boldsymbol{e}_{i}^{2}}{n-2} = \frac{\sum_{i}^{n} (Y_{i} - \widehat{Y}_{i})^{2}}{n-2}$$

For first degree linear regression model the  $\sigma^2 = MSE$ . Thus, the covariance matrix of the coefficients becomes:

$$E[\boldsymbol{b}\boldsymbol{b}^{T}] = \begin{bmatrix} \frac{MSE^{2}}{n} + \frac{MSE^{2}\bar{X}^{2}}{\sum_{1}^{n}(X_{i} - \bar{X})^{2}} & \frac{-MSE^{2}\bar{X}^{2}}{\sum_{1}^{n}(X_{i} - \bar{X})^{2}} \\ \frac{-MSE^{2}\bar{X}^{2}}{\sum_{1}^{n}(X_{i} - \bar{X})^{2}} & \frac{MSE^{2}}{\sum_{1}^{n}(X_{i} - \bar{X})^{2}} \end{bmatrix}$$

#### Estimation of prediction intervals

The prediction intervals can be estimated as follows:

$$PI = \hat{Y} \pm t_{1-a,n-2} \sqrt{MSE + \hat{X}^T \hat{X} Cov[b]}$$

Where:

- $t_{1-a,n-2}$  is the critical value for the Student t-distribution for the n-2 degrees of freedom and significance level  $\alpha$ . Where n is the population size. The critical t values can be found in the Table B-3 (Appendix B). In this study, the confidence level set to 95% ( $\alpha$ =0.05). The 95% confidence level means that future values fall into the prediction envelope 95% of the time.
- MSE is the mean square of error that can be estimated as follows:

$$MSE = \frac{\sum_{i}^{n} (Y_{i} - \widehat{Y}_{i})^{2}}{n - 2} \widehat{X}$$

•  $\widehat{X}$  and  $\widehat{X}^{T}$  are the vectors of new independent values, defined as:

$$\widehat{\mathbf{X}} = \begin{bmatrix} 1\\ \widehat{X}_1\\ \widehat{\mathbf{X}}_2\\ \vdots\\ \widehat{\mathbf{X}}_n \end{bmatrix} \text{ and } \widehat{X}^T = \begin{bmatrix} 1 & \widehat{X}_1 & \widehat{\mathbf{X}}_2 & \cdots & \widehat{\mathbf{X}}_n \end{bmatrix}$$

• Cov[b] is the covariance of the b parameters that can be calculated as:

$$Cov[b] = \frac{-MSE\,\bar{X}^2}{\sum_{1}^{n}(X_i - \bar{X})^2}$$

Table B-3: Critical t distribution table. The corresponding t critical value of both models has been highlighted in yellow.

	Level of Significance of a Two-Tailed or Nondirectional Test								
	$\alpha = 0.20$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.02$	$\alpha = 0.01$	$\alpha = 0.001$			
df	$1-\alpha=0.80$	$1-\alpha=0.90$	$1 - \alpha = 0.95$	$1-\alpha=0.98$	$1 - \alpha = 0.99$	$1 - \alpha = 0.999$			
1	3.078	6.314	12.706	31.821	63.656	636.578			
2	1.886	2.920	4.303	6.965	9.925	31.600			
3	1.638	2.353	3.182	4.541	5.841	12.924			
4	1.533	2.132	2.776	3.747	4.604	8.610			
5	1.476	2.015	2.571	3.365	4.032	6.869			
6	1.440	1.943	2.447	3.143	3.707	5.959			
7	1.415	1.895	2.365	2.998	3.499	5.408			
8	1.397	1.860	2.306	2.896	3.355	5.041			
9	1.383	1.833	2.262	2.821	3.250	4.781			
10	1.372	1.812	2.228	2.764	3.169	4.587			
11	1.363	1.796	2.201	2.718	3.106	4.437			
12	1.356	1.782	2.179	2.681	3.055	4.318			
13	1.350	1.771	2.160	2.650	3.012	4.221			
14	1.345	1.761	2.145	2.624	2.977	4.140			
15	1.341	1.753	2.131	2.602	2.947	4.073			
16	1.337	1.746	2.120	2.583	2.921	4.015			
17	1.333	1.740	2.110	2.567	2.898	3.965			
18	1.330	1.734	2.101	2.552	2.878	3.922			
19	1.328	1.729	2.093	2.539	2.861	3.883			
20	1.325	1.725	2.086	2.528	2.845	3.850			
21	1.323	1.721	2.080	2.518	2.831	3.819			