Deltares

The Evolution of Storms on the Wadden Sea

Juliana Lopez de la Cruz
Annegien Tijssen
Joost Beckers

1200264-004

Title
The Evolution of Storms on the Wadden Sea

Project	Reference	Pages
1200264-004	1200264-004-HYE-0012	21

Keywords
Storm Evolution, Wadden Sea, Correlation Analysis and Multivariate Regression

Summary

This report presents a standard storm profile, giving a description of the typical temporal variations of wind speed and direction during the evolution of a storm. This model storm profile can be applied in a new setup of WAQUA and SWAN simulations for the Wadden Sea in WTI2011 (under development). In addition to this, a preliminary study by means of correlation analysis and multivariate regression is performed to obtain a first indication of the linear relations between the surge levels in the Wadden Sea and a number of meteorological variables, such as wind speed, wind direction, air pressure and temperature. This analysis is preliminary to the development of a new type of probabilistic model for storm surges in the Wadden Sea, which would resolve a number of limitations of the current Hydra-K model. The results of the correlation analysis show that high surges at Harlingen are mildly correlated with wind speed. Several other weather conditions (cloud cover, temperature, wind direction) show weak correlations, but the dependencies are not very pronounced and could be coincidental. Further studies to investigate the underlying relationship and dependencies between surge and weather conditions are necessary before the development of a new probabilistic model is started.

References
SBW-Belastingen 2009
1200264.004 Probabilistics

Versie	Datum	Auteur	Paraaf	Review	Paraaf	Goedkeuring	Paraaf
1.0	okt. 2009	Lopez/Beckers/Tijssen		G. Lipari			
2.0	Feb 2010	Beckers		G. Lipari		M. van Gent	

State
final

Contents

1 Introduction

1.1 Background

The SBW programme is aimed at improving the models and methods that are used for the five-yearly statutory assessment of the Dutch flood defences. Special attention is being paid to the Dutch Wadden Sea as many new methods have been developed for this area during the last few years. The main themes are wave growth and wave propagation on the Wadden Sea. For this purpose the wave propagation model SWAN has been improved and will be used for the Hydraulic Boundary Conditions (HBC) in 2011.

The SWAN model takes wind, water level and currents as input. In the HR2006 setup, the water level is assumed uniform over the Wadden Sea area and the currents are neglected. The results of many SWAN simulations have been stored in a database and used by the probabilistic model Hydra-K to obtain nearshore wave conditions from wind and water level. The correlation between wind and water level has been derived from observed storms and extrapolated to extreme events.

There are, however, indications that this Hydra-K version is not equipped to efficiently model the physical processes leading to high water levels in the Wadden Sea. Research of Alkyon (2009) for instance, showed that a schematized storm with a constant wind speed and direction (over time) could not reproduce the observed water levels. Also, Hydra-K does not model currents, while it is evident that currents have a great influence on wave characteristics of the Wadden Sea.

Taking into account the above, it is clear that there is a need for a new type of probabilistic model for the Wadden Sea. However, to develop such a model would take several years. For WTI2011 this will be too late, so it is therefore necessary to keep on using Hydra-K. To take some of the above mentioned physical aspects into consideration for HR2011, a standard storm profile can be used. In a new setup of SWAN calculations for Hydra-K, described in (Deltares, 2009b), the water level and current fields are calculated using the hydrodynamic model WAQUA, also taking wind as input. This model setup requires a typical storm profile (time evolution of wind speed and direction). Such a profile will be described in this report.

In order to find realistic wave conditions it is essential to use realistic wind fields. For the 2011 HBC, an average of several observed storm evolutions will be used that describe a gradual increase of the wind speed to a peak value and a subsequent gradual decrease. Meanwhile, the wind direction is changing and the spatial variation of the wind field is neglected. This analysis is described in Chapter 2. To further develop the method after 2011, more knowledge of the correlations between surge level and wind conditions is required. This is dealt with in Chapter 3. Conclusions and recommendations are given in Chapter 4.

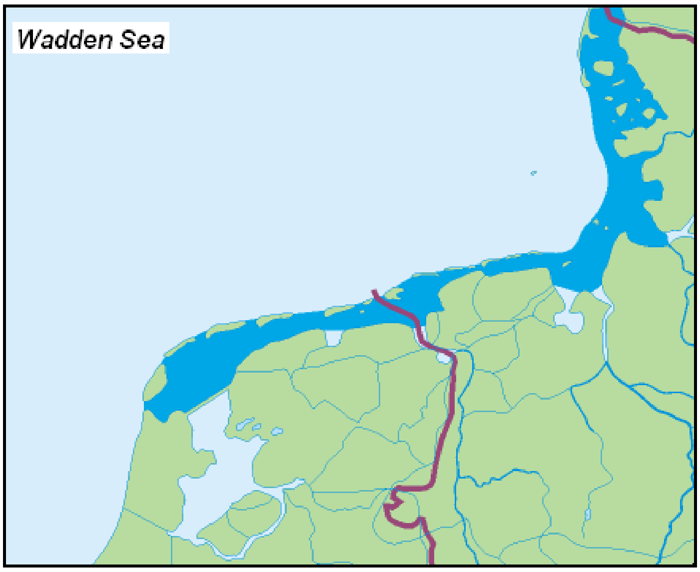

Figure 1.1 The Wadden Sea region.

1.2 Objectives

The objectives of this study are:

1. providing a standard model for the time dependency of wind speed and direction for usage in Hydra-K (WTI2011);

2. investigating the correlations between the surge levels in the Wadden Sea and a number of meteorological variables, such as wind speed, wind direction, air pressure and temperature;

3. making suggestions for further research and looking ahead to a possible new probabilistic approach of randomized storm evolution.

2 Standard storm profile for WTI2011

In this chapter a standard storm profile is discussed that can be used in a new setup of WAQUA and SWAN simulations for Hydra-K in WTI2011. This model setup requires a storm profile with temporal variations of wind speed and direction. Although Alkyon (2009) conjectured that spatial variation is also required to reproduce measured water levels, this will not be part of the storm profile for WTI2011. A spatial variation would complicate the connection to a wind speed exceedance probability. Moreover, it is not the purpose of the WAQUA and SWAN models to reproduce the water level statistics for Hydra-K. Rather, the SWAN model should produce realistic wave conditions.

Data were downloaded from the KNMI website: www.knmi.nl/samenw/hydra/index.html. For the analysis, the hourly means of the wind potential speed at De Kooy (Den Helder) from 1 January 1972 until 1 January 2005 have been used. In addition to speed, also the hourly wind directions at the measuring station De Kooy for the same period have been used. In the Hydra dataset, the wind directions are 10-minute averages, taken at the end of each hour.

2.1 Approach

An aggregate temporal evolution has been determined for both wind speed and wind direction. Both are described briefly in this section.

Storm selection

To determine the evolution, we first have to select a population of storms. Here, western (eastern) storms are defined as storms with a peak wind direction ranging from south to north by west (east). This selection is done by using the Peak-Over-Threshold method. For this method, a time window of one day is used to screen out storms with multiple peaks. A threshold value of 20 m/s leads to one storm a year on average. These storms, however, all have a western peak wind direction. In order to include eastern storms as well, a threshold value of 15 m/s has been used for the selection of eastern storms.

Temporal evolution of wind speed

A scaling method is used to derive the (schematic) wind speed evolution. For a more detailed description of this method please refer to (RIZA, 2004). The same method has also been used to determine the Hydra-VIJ storm duration (RIZA, 2006) and storm surge duration and storm duration at Hoek van Holland (Deltares, 2009a).

For all selected storms the wind speed histories are re-scaled, dividing the wind speed at each time instant by the peak wind speed of the storm. Next, for each storm the durations of the re-scaled wind speed exceeding levels 0.1, 0.2, 0.3, …, 1.0 are determined. These durations are averaged over the storms, yielding an average storm evolution. Finally, a symmetrical trapezium with a top duration of 1 hour is fitted to this average storm evolution. The trapezium is the desired (schematized) temporal evolution of the wind speed.

The above analysis is applied separately to groups of storms in several categories, according to the wind direction at the storm peak. The categories and the number of storms in each category are given in Table 2.1.

Table 2.1 Number of selected storms per peak wind direction.

Peak wind direction [degrees]	Threshold value [m/s]	Number of storms
180 – 360 of which:	20	43
180 – 270	20	33
270 – 360	20	10
0 – 180 of which:	15	19
0 – 90	15	11
90 – 180	15	8

Schematic temporal evolution of the wind direction

The temporal evolution of the wind direction has also been derived using the same storms as for wind speed. For every storm, the angle between the wind direction at a certain time and the wind direction at the peak is determined. Next, the time series of this angle for each category of storms is determined. The average is taken over all storms in the category and a Gaussian error function (erf) is fitted to the result. This fit is the desired temporal evolution of the wind direction.

The analysis is applied to groups of storms in several categories, according to the wind direction at the peak of the storm. The categories and the number of storms in each category are given in Table 2.2

Table 2.2 Number of selected storms per peak wind direction.

Peak wind direction [degrees]	Threshold value [m/s]	Number of storms
180 – 360 of which:	20	43
180 – 240	20	18
240 – 300	20	21
300 – 360	20	4
0 – 180 of which:	15	19
0 – 90	15	11
90 – 180	15	8

2.2 Results

The evolution of wind speed in time

Figure 2.1 shows that the wind-speed temporal evolution of each storm group does not differ too much. Therefore, an overall wind speed evolution is used irrespective of the wind direction at peak[1]. For application in Hydra-K, the hours around the peak of the storm are of main interest. A schematization of the storm evolution in time should therefore accurately describe this part. Figure 2.2 shows that a trapezium with a one-hour peak plateau and a 48-hour duration yields an adequate fit to the schematic wind speed history.

1. *Actually, the eastern storms typically last longer: of the order of 55 hours. However, since the uncertainty is large (there were only 19 selected eastern storms) and the eastern storms do not produce the highest loads in the Wadden Sea, no separate temporal evolution for eastern storms was defined.*

Recall that the average wind speed is computed by averaging the durations of exceedance. This means that the wind speed is averaged horizontally and not vertically, leading to horizontal error bars displayed in Figure 2.2. Also, a time window of plus or minus one day was used in the storm selection. Therefore, for relative wind speeds close to zero, the duration of exceedance is approximately 48 hours. The variations of these durations between storms are very small. This explains the small error bars in the lower region of Figure 2.2. More importantly, the fit reproduces well the storm evolution of the group near the peak, since it falls within the uncertainty embedded in the data.

The wind speeds are defined as hourly averages. Therefore, the duration of the peak value is set to 1 hour. The variations of these durations between storms are very small. This explains the small error bars in near the peak in Figure 2.2.

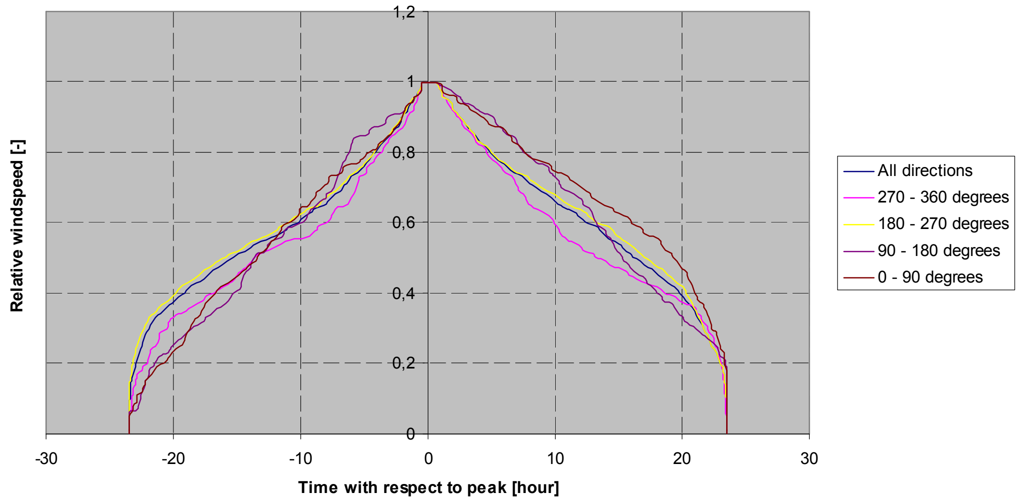

Figure 2.1 *Temporal evolution of the average wind speed for various peak wind directions*

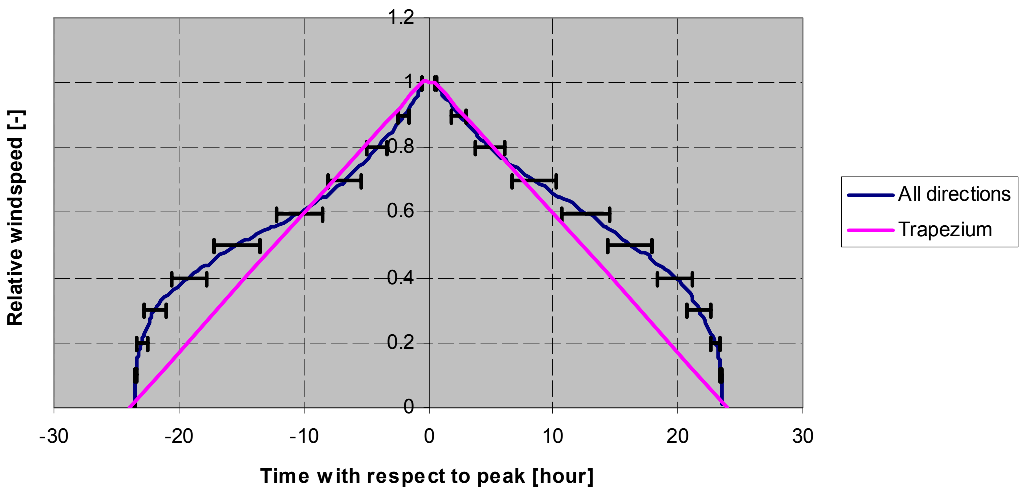

Figure 2.2 *Schematized temporal evolution of the average wind speed (all peak wind directions), with 95% confidence intervals.*

The evolution of the wind direction in time

Figure 2.3 shows that the temporal evolution of the wind turning indeed varies with the peak wind direction. Western storms have a clear pattern, a turning from southwest to northwest, whereas the temporal evolution of the wind direction for eastern storms is less clear. Figure 2.4 and Figure 2.5 show confidence limits for the mean turning of wind direction for east and west sectors. The confidence limits are based on the standard error of the mean ($\sigma/\sqrt{n}$, where σ is the estimated standard deviation of the population and n is the number of measurements).

The change in average wind direction for the western sectors is significant, whereas for the eastern storms the change in wind direction falls within the confidence limit. This is consistent with the general knowledge that western storms typically display winds turning overhead, caused by a low pressure system travelling west/southwest-wards over the North Sea. In contrast, winds from the eastern sector originate from a variety of meteorological situations, resulting in various patterns of changing wind direction. A standard temporal evolution of the wind direction has therefore been derived only for western storms (peak wind direction between 180 and 360 degrees). Eastern storms will not be considered in the remainder of this study.

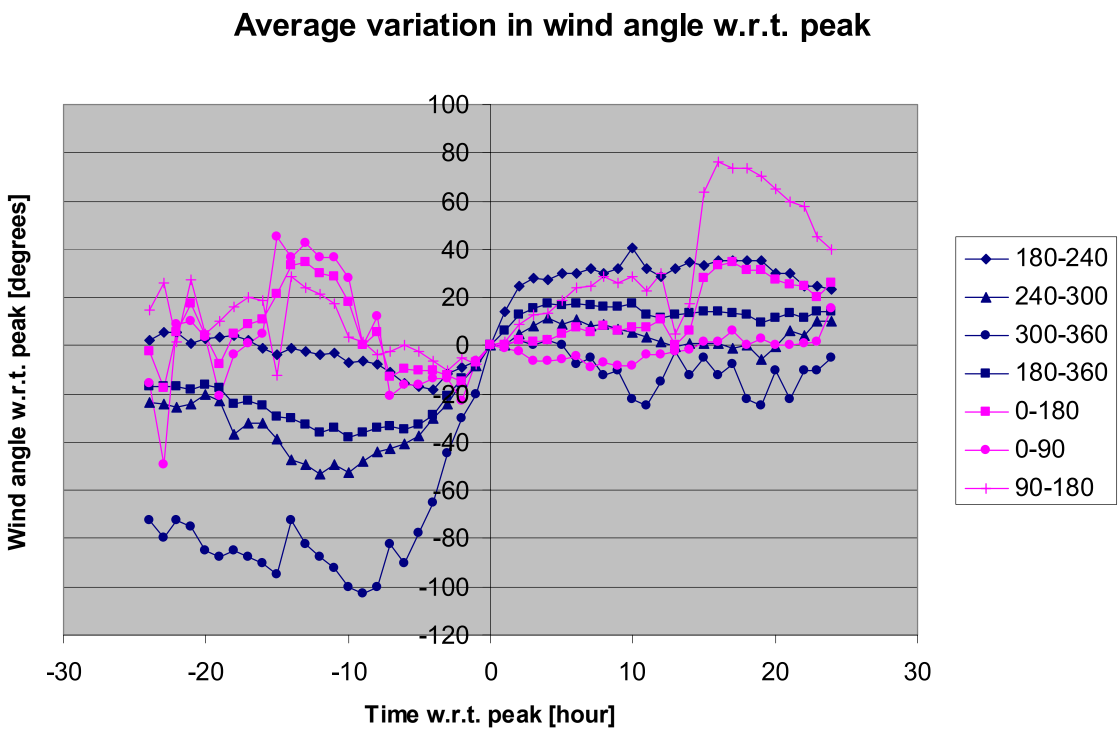

Figure 2.3 Average relative wind direction temporal evolution compared to peak wind directions for various peak wind directions

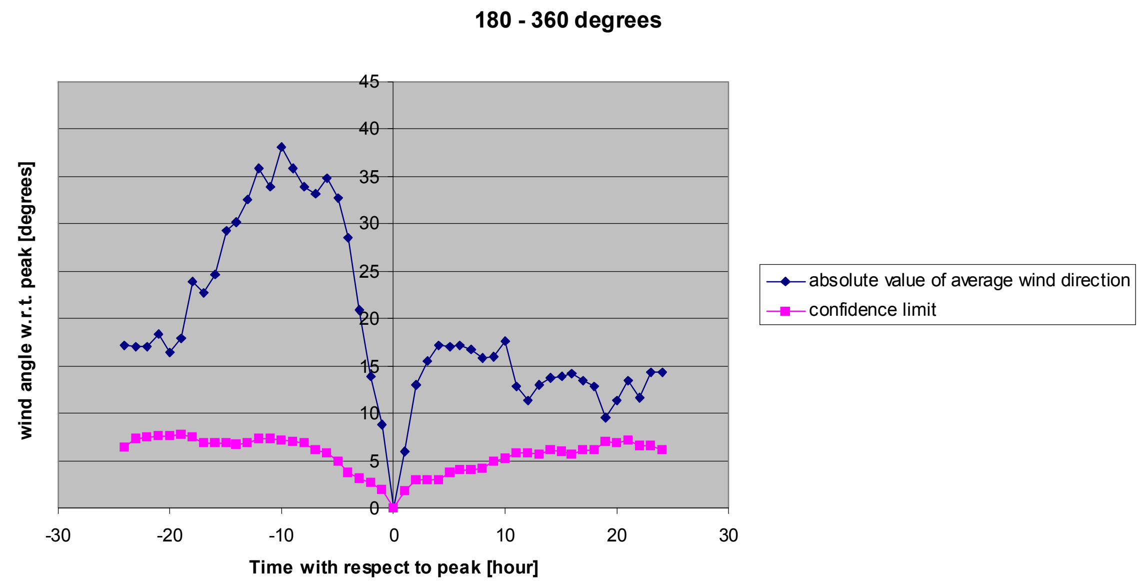

Figure 2.4 *Temporal evolution of the average wind direction with confidence limit for western storms (peak wind direction between 180 and 360 degrees)*

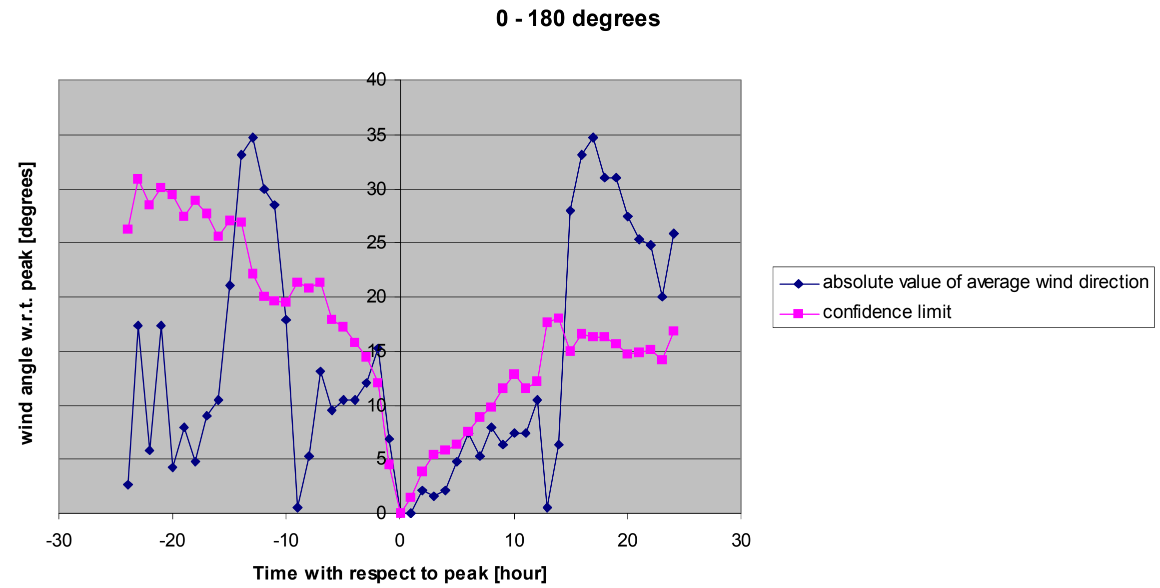

Figure 2.5 *Temporal evolution of the average wind direction with confidence limit for eastern storms (peak wind direction between 0 and 180 degrees)*

A Gaussian error function (erf) is fitted by hand to the average wind speed for the three sub-categories of western storms, resulting in the schematized temporal evolutions of the wind direction, shown in Figure 2.6 - Figure 2.9. These patterns can be used for the time-dependent wind forcing of the WAQUA simulations that produce water level and current fields for the SWAN simulations for Hydra-K.

We stress that this average pattern is rather crude and is only meant as a first improvement to the complete neglect of temporal variation of wind in the HR2006 setup. If the storm pattern is to be used to produce realistic water levels it should be determined to better accuracy. One improvement that should be considered is the possible dependency of the turning of wind direction on the wind sector and on the peak wind speed, or the severity of the storm. The fact that sub-categories 180-240, 240-300 and 300-360 show different patterns indicates that such dependencies indeed exist. More advanced modelling of such dependencies is recommended, but falls beyond the scope of this study.

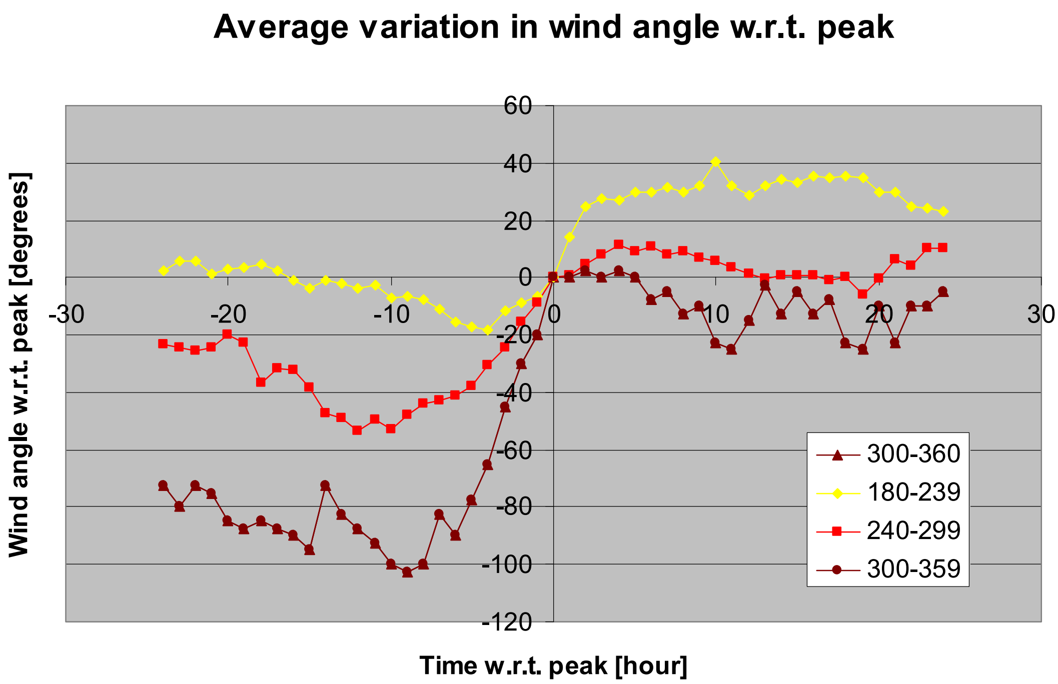

Figure 2.6 Schematized temporal evolution of the average wind direction for western storms

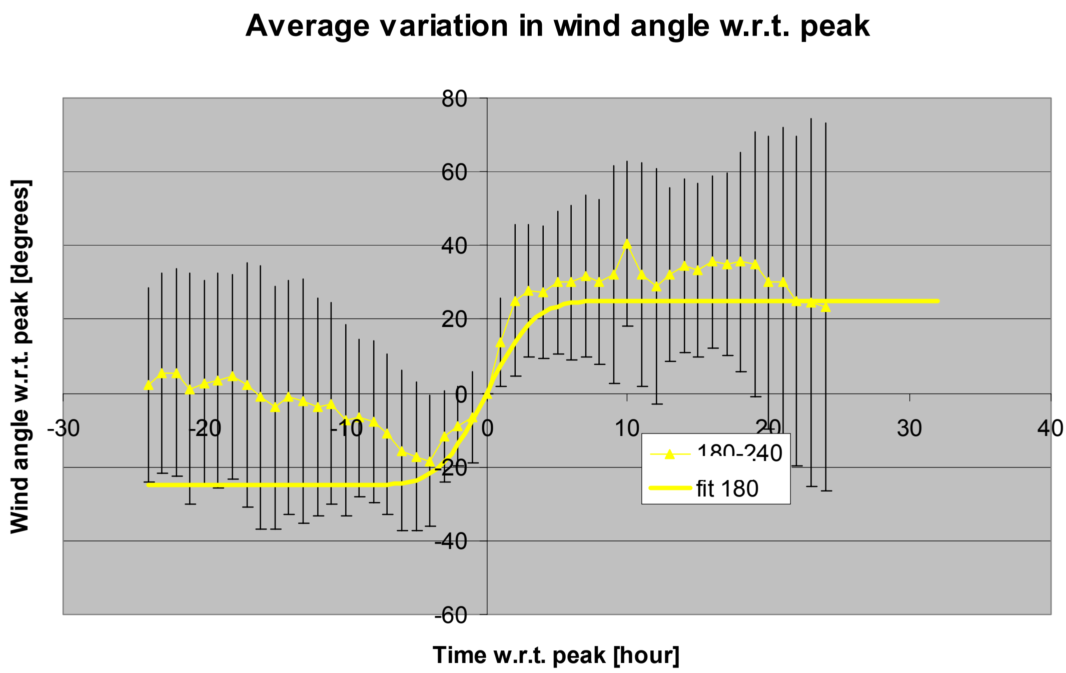

Figure 2.7 Schematized temporal evolution of the average wind direction with 95% uncertainty bands.

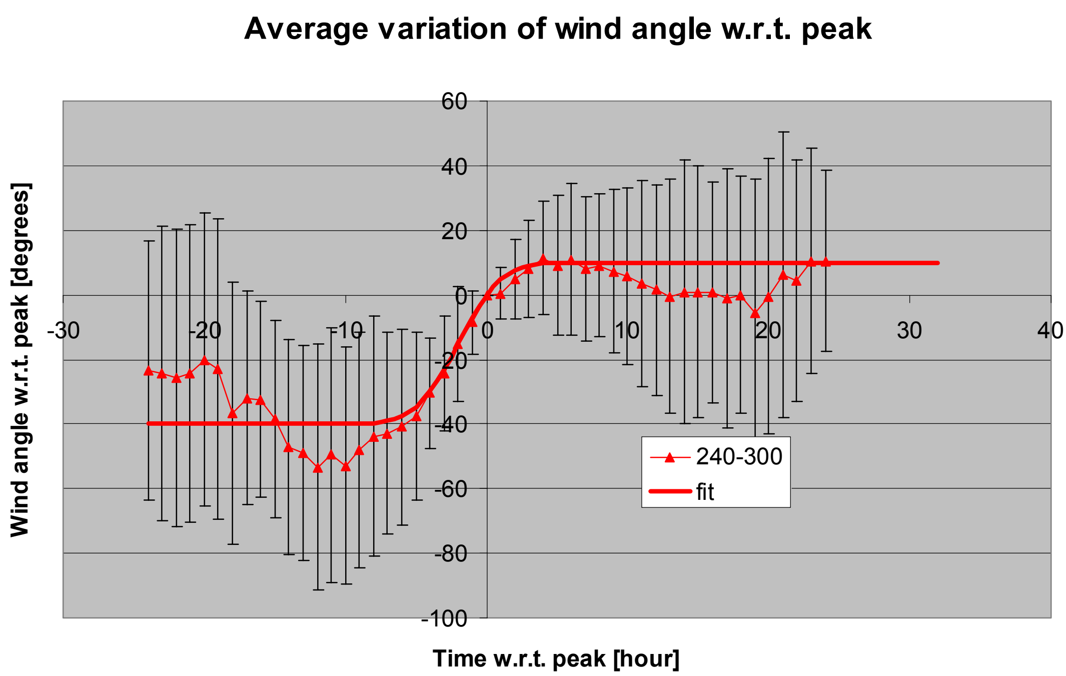

Figure 2.8 Schematized temporal evolution of the average wind direction with 95% uncertainty bands.

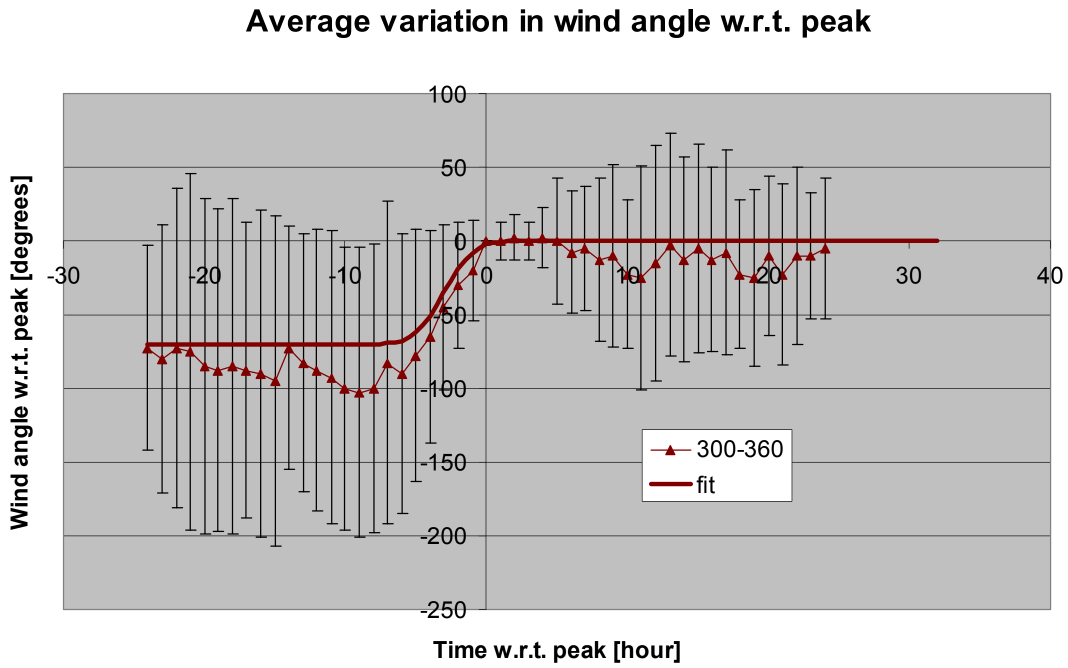

Figure 2.9 Schematized temporal evolution of the average wind direction with 95% uncertainty bands.

3 Multivariate regression and correlation analysis

In this Chapter, the correlation of meteorological variables or parameters with high surges in the Wadden Sea is investigated. Two statistical techniques are applied in order to identify the level of correlation between the extreme surge[2] values and the weather conditions that accompany them. This is a first step towards a new probabilistic model for extreme hydraulic loads (water level and waves) on coastal defences in the Wadden Sea. Section 3.1 presents the background theory on the statistical methods applied in this study. Section 3.2 provides a description of the data handling prior to the analysis and Section 3.3 presents the conclusion of the analysis and further recommendations for future research.

3.1 Statistical methodology

Several statistical methods are available to perform a pattern analysis of variables in a given case study. Generally, the analysis is simplified by investigating the correlation between pairs of variables, which is more easily performed as numerous methods are available and a least squares fit usually provides a good measure of their relationship. In cases that involve several variables, the analysis becomes more challenging as multidimensional fits require a different methodology.

In this report, the relationships among variables are studied by means of two statistical methods: c*orrelation analysis* and *multivariate regression.* These methods are widely used to study correlations between data series as they measure the linear relationships among the variables involved in the analysis.

Correlation Analysis

Correlation is a measure that indicates the strength and direction of the linear relationship between two variables. The correlation coefficient takes values within the interval [-1, 1]. Two independent variables have a correlation coefficient equal to zero[3]. The correlation coefficient of two variables is computed as:

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E((X-\mu_X)(Y-\mu_Y))}{\sigma_X \sigma_Y} \tag{3.1}$$

where E is the expected value operator, μ is the expected or mean value of the variable and σ is the standard deviation.

The interpretation of the correlation coefficient is somewhat arbitrary, as it requires consideration of the data quality and quantity and of the techniques used to gather it. However, an indication of the strength of the relation between the variables can be assessed following Table 3.1.

[2] *Surge is defined as water level minus astronomical tide.*

[3] *On the other hand, a correlation coefficient of zero does not mean that the variables are independent. Two variables can be strongly coupled in a non-linear way and have a correlation coefficient of zero.*

Table 3.1 Correlation coefficient interpretation

Correlation Strength	Negative	Positive
Weak	-0.3 to -0.1	0.1 to 0.3
Mild	-0.7 to -0.3	0.3 o 0.7
Strong	-1.0 to -0.7	0.7 to 1.0

Multivariate regression

Regression analysis is a statistical technique to model numerical data that consists of a dependent variable *Y* and one or more independent variables X_j. The dependent variable in the regression equation is modeled as a linear combination of independent variables times constants (beta weights) and an error term. The error term represents the part of the dependent variable that cannot be captured by the independent variables. The independent variables' beta weights, also called regression coefficients[4], are estimated to give the best fit of the data, which is usually evaluated applying the least squares method or a similar method. The multivariate regression is a useful tool to analyze the relative predictive importance of the independent variables by comparing the regression coefficients of each independent variable.

Some of the requirements of a multivariate regression are:

1) The sample must be representative of the population for the inference prediction;
2) The independent variables do not contain outliers;
3) The independent variables must be linearly independent (i.e. it must not be possible to express any of the variables as a linear combination of the others).

The first requirement is assumed for the time being. The second requirement is fulfilled by singling out and removing outliers. The third assumption is checked by computing the correlation coefficients among the independent variables. If a large correlation coefficient between two independent variables is observed, then *multicollinearity* may be assumed. Perfect multicollinearity occurs if the correlation coefficient between two independent variables is equal to 1 or -1. In practice, perfect multicollinearity in a data set is rarely encountered. However, in cases where the correlation coefficient between two variables is above 0.9, either of the pair should be suppressed from the analysis as it is already represented by the remaining variable.

The multiple regression equation deals with a set of variables. The unknown variables are the constants or regression coefficients denoted as β_j**.** The independent variables or measured variables are denoted as X_{ij}**.** The dependent variable is denoted as Y_i**.** Thus, the regression equation is a function of variables X_{ij} and regression coefficients β_j expressed in the form:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_p X_{ip} + e_i \qquad i = 1,...,n \tag{3.2}$$

where *e* stands for the error of the fit and *n* is the number of measurements. The multiple regression system has a unique solution if the number of measurements *n* is larger than the number of unknowns β_j. In the regression analysis, a solution is found for the unknown regression coefficients β_j that will minimize the distance between the measured and predicted values of the dependent variable *Y*. The regression coefficients β_j represent the amount that

4. *Not to be confused with the correlation coefficients mentioned in the former section on correlation analysis.*

the dependent variable Y changes when the corresponding independent variable changes by one unit, while the correlation coefficients measure the percent of variance of the dependent variable explained by the independent variables when all the independent variables are allowed varying.

The proportion of variability of the dependent variable Y that is explained by the regression model is computed as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}_i\right)^2} \tag{3.3}$$

where $\overline{Y}$ is the sample mean of Y and $\hat{Y}_i$ are the least square estimates of Y_i.

The variables in the data set have different magnitudes, dimensions and dispersions. In order to be able to compare the regression coefficients, each variable (weather conditions and surge) is standardized by subtracting its mean value and dividing by its standard deviation.

$$X_i^{'} = \frac{X_i - \overline{X}_i}{\sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(X_j - \overline{X}_j\right)^2}} \tag{3.4}$$

In this way, all variables are dimensionless and proportional in range and the regression coefficients can be compared. Since the surge is also standardized, the constant term (β_0 in equation 3.2) will become zero.

3.2 Results

This study aims at investigating the influence of wind speed, wind direction, temperature, air pressure, cloud cover and atmospheric humidity, in the surge levels for selected storms in the Wadden Sea. The analysis is performed applying the methods described in the previous Section. Correlation analysis and regression methods are based on the assumption that the data set is free of outliers. The input data for analysis are carefully selected and outliers are eliminated. In this study, outliers are defined as data points that deviate more than 3σ from other members of the sample.

Data conditioning

Surge levels from Harlingen and weather conditions from station De Kooy were used for the regression analysis. Since the objective is to make only a first selection of variables that are correlated with high surges no other stations in the Wadden Sea were considered. A pre-selection of storm periods between 1972 and 2002 was made, based on SVSD reports. 24 Storms were selected from the list of storms that caused the highest water levels at Harlingen over the past 39 years[5], listed in Table 3.2.

Table 3.2: Selected storms for the correlation and regression analysis at Harlingen.

Nr	Date	Max water level m+NAP	Max surge level m
1	3 Jan 1976	3.69	3.67
2	26-27 Feb 1990	3.66	3.19
3	1 Feb 1983	3.55	2.90
4	20 Jan 1976	3.53	2.76
5	9 Nov 2007	3.50	2.66
6	28 Jan 1994	3.44	2.77
7	18 Jan 2007	3.31	2.48
8	21 Feb 1993	3.31	2.89
9	12 Dec 1973	3.27	2.13
10	1 Nov 2006	3.26	2.24
11	18 Mar 2007	3.20	2.63
12	12 Jan 2007	3.09	2.20
13	3 Nov 1970	3.05	2.77
14	14 Nov 1977	3.04	2.89
15	30 Dec 1977	3.03	2.49
16	12 Dec 1990	3.00	2.35
17	24 Nov 1981	3.00	2.69
18	16 Nov 1973	3.00	2.40
19	20 Dec 1991	2.99	2.30
20	16 Dec 1982	2.97	2.43
21	25 Jan 1993	2.96	2.59
22	14 Feb 1989	2.96	2.72
23	13 Nov 1973	2.96	2.03
24	14 Mar 1994	2.95	1.87

[5] *According to the storm surge warning service (SVSD), see http://www.svsd.nl/*

The weather conditions were averaged over the five hours preceding the peak water level before usage in the statistical analyses. This period is expected to be mainly responsible for the high surge. As an alternative, the weather conditions at the peak of the surge and the average over 10 hours before the peak were investigated, which gave comparable results.

The selected significant weather variables for the analysis are shown in Table 3.3. These variables were selected from a set of 16 available time series of meteorological variables by expert judgement. The wind speed and the wind direction (average in time) are analyzed as separate variables. The air pressure is defined as the local atmospheric pressure reduced to the sea level.

Table 3.3 Selected weather variables for the surge analysis.

Weather variables	
S	Average wind speed [m/s]
D	Average wind direction [degrees]
T	Average temperature [in 0.1 degrees Celsius]
T_D	Average dew point temperature at 1.50 m [in 0.1 degrees Celsius]
P	Average air pressure reduced to mean sea level [in 0.1 hPa]
C	Average cloud cover [octants]
H	Average relative atmospheric humidity [percentage]

Correlation analysis

Figure 3.1 displays the scatter plots of the surge versus the wind speed and direction. One outlier, indicated by the arrows was removed (storm nr 17, 24 Nov 1981). The correlation coefficients for all combination of two variables were computed for the remaining data set of 23 storms.

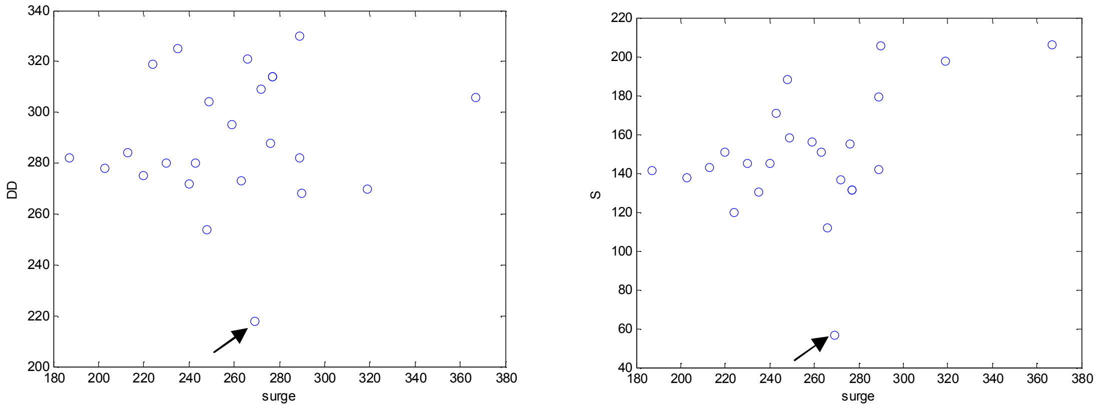

Figure 3.1: Scatter plot of the surge at Harlingen against the wind direction (left) and wind speed (right) at De Kooy.

The results of the correlation analysis of the data set are presented in Table 3.4. As mentioned, the correlation analysis measures only linear dependence among the variables. The correlation coefficients in the diagonal correspond to the self-correlation with the obvious value *one*. A negative correlation coefficient indicates that when either variable increases the other decreases.

The results in Table 3.4 illustrate that the surge is most strongly correlated with the wind speed, although the correlation coefficient is still 'mild' according to Table 3.1. Temperature, cloud cover and to a lesser extent wind direction show a weak correlation with the surge. Temperature has a negative correlation with the surge level. The air pressure and humidity have a correlation coefficient close to zero, which suggests that the surge does not have a linear relationship with the air pressure or humidity.

Table 3.4 Correlation analysis of the weather variables and surge for 24 storms in the Wadden Sea

	surge	**S**	**D**	**T**	**P**	**C**	**H**
surge	1,00	0,56	0,18	-0,29	-0,04	0,31	-0,11
S	0,56	1,00	-0,57	0,16	-0,65	0,19	0,27
D	0,18	-0,57	1,00	-0,52	0,68	0,11	-0,33
T	-0,29	0,16	-0,52	1,00	-0,43	-0,14	0,08
P	-0,04	-0,65	0,68	-0,43	1,00	-0,30	-0,36
C	0,31	0,19	0,11	-0,14	-0,30	1,00	0,40
H	-0,11	0,27	-0,33	0,08	-0,36	0,40	1,00

Interestingly, the correlation coefficients between the air pressure and the wind speed and wind direction are rather large (low-pressure systems are typically responsible for North Sea storms). Furthermore, although the latter are correlated with the surge, there seems to be no strong correlation between surge and air pressure. Indeed, the surge is more related to pressure *gradients,* through the wind speed, instead of the absolute value of the air pressure, through the adjustment of the free surface to the air pressure. However, such a relation cannot be untangled by the current correlation analysis.

The correlation coefficients between air pressure on the one hand and wind speed and wind direction on the other suggests an almost strong linear relationship between these variables, which can influence the regression coefficients of the multivariate regression by multicollinearity. However, it has been empirically established that only bivariate correlations above 0.90 might translate into multicollinearity in the data (Garson, 2009).

Multivariate linear regression

The multivariate regression is applied to the data set of 23 storms. The advantage of this method is that causality or predictability analyses can be performed by this technique (Hamilton, 2008). In the case at hand, a predictability analysis is carried out for high surge data observed at severe weather conditions where the individual contribution of each weather variable is assessed.

The results of the multivariate regression are presented in Table 3.5. Recall that the correlation coefficients can be compared because all variables were normalized by dividing by their mean values and standard deviations before the regression analysis was carried out.

The results in Table 3.5 show that the wind speed is the variable with the strongest influence on the surge. This finding is in agreement with the results from the correlation analysis. The wind direction and air pressure also have some influence on the surge. The temperature has the weakest influence in the surge of all variables.

Table 3.5: Multivariate regression coefficients and error.

Multivariate regression coefficients		
	Coefficients	***Std. error***
constant	0	0
S	1,030	0,170
D	0,426	0,212
T	-0,046	0.151
P	0,311	0,214
C	0,255	0,159
H	-0,236	0,144
Max error	0.90	

The max error in Table 3.4, measures the maximum distance between the standardize surge values and the values obtained by the regression model which is illustrated in Figure 3.2. After the data is standardized, a residual higher than *3.3* (Garson, 2009), indicates the presence of outliers or a poor fitted model to the variables of interest. In other words, a high error can illustrate that the assumption of a linear relationship between the variables is erroneous.

In order to identify the percentage of the surge variance that is explained by the regression model, the R^2 value is computed by means of Equation 3.3. The results show that 76% (R^2 = 0.765) of the variance is explained by the regression model which underlines a good performance of the selected model.

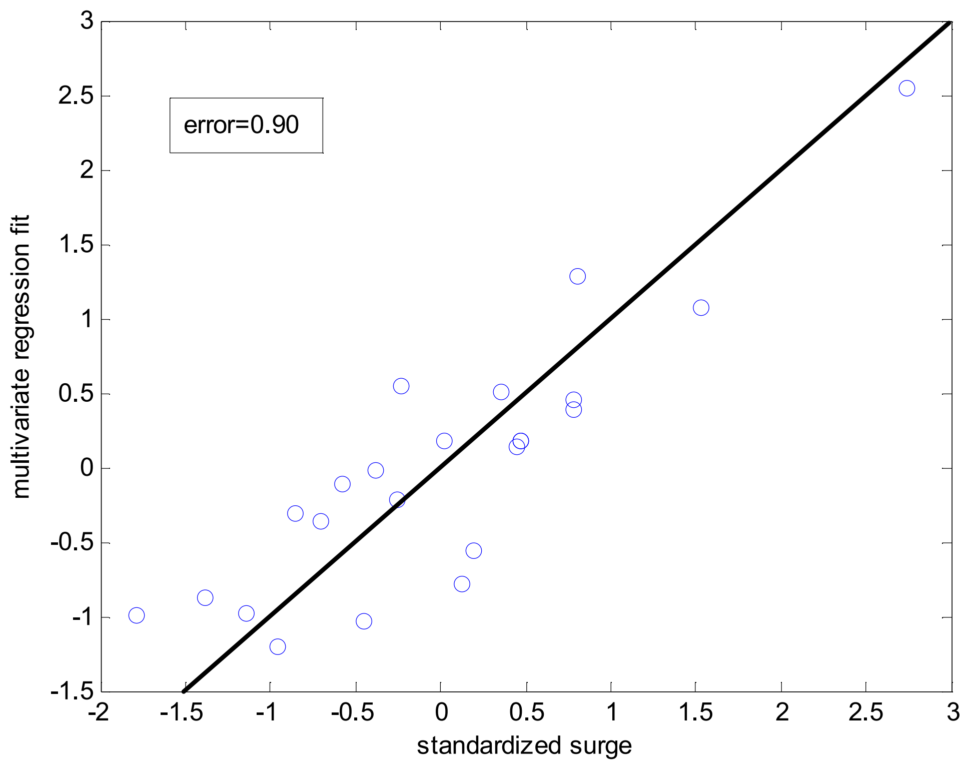

Figure 3.2: Regression fit of the standardized surge.

3.3 Conclusions of the correlation and regression analyses

The weather conditions and surge data of 23 storms were analyzed by means of two different statistical methods. Before the analysis, the data set was filtered and standardized in order to decrease the error induced by outliers and differences in data magnitude, dimension and dispersion.

The correlation and regression analysis indicate that there is a positive linear correlation between the wind speed and the surge. Other variables that show weak correlations with the surge are wind direction, cloud cover and temperature (negative correlation). The regression analysis suggests that air pressure is also an explanatory variable for the surge. However, this is not confirmed by the correlation analysis.

The results of the current regression analysis are only a preliminary step to a more advanced investigation of correlation analysis between meteorology and surge. In order to be useful, the analysis should be extended to take into account more knowledge of physical relationships between the variables, non-linear relationships and spatial correlations of the variables. Also, the dataset should be extended by including more storms (also before 1970). The variability range of the surge and explanatory variables should be increased in order to confirm or reject the dependencies. The variability in the current dataset is limited because only those that produced the highest water levels were selected.

4 General conclusion and recommendations

The standard storm profile described in Chapter 2 gives a rough estimate of the typical temporal variation of wind speed and direction during the evolution of a storm. This model storm profile will be used for WAQUA and SWAN simulations leading to WTI2011 for the Wadden Sea. We stress that this storm pattern is rather simplistic and is only meant as a first improvement to the complete neglect of temporal variation of wind in the HR2006 setup. More detailed descriptions of the storm pattern could include spatial variations and dependency of the temporal evolution on the type and severity of the storm. This is left for future research.

Chapter 3 revealed that the wind speed is most strongly linear correlated with high surges. Several other weather conditions (cloud cover, temperature and wind direction) have weak linear correlations with the surge. The dependencies and correlations between the meteorological variables and surge that were found are not very pronounced, even for processes whose general physics is well understood. This suggests that their occurrence in the Wadden Sea is further complicated by other non-meteorological, say geographical, factors.

The results of the current regression analysis are only a preliminary step to a more advanced investigation of correlation analysis between meteorology and surge. In order to be useful, the analysis should be extended to take into account more knowledge of physical relationships between the variables, non-linear relationships and spatial correlations of the variables. It is therefore recommended to perform further research in preparation for a new probabilistic model for Wadden Sea storm surges, which would resolve a number of limitations of the current Hydra-K model. This research should take into consideration the following issues:

- Knowledge of physical relationships between the meteorological variables, non-linear relationships and spatial correlations of the variables should be taken into account in the regression analysis. Two suggestions are:
  - Investigate correlations between the surge and combinations of meteorological variables (e.g. wave age) based on knowledge of physical processes.
  - Investigate the influence of storm tracks and the spatial structure of storms/depressions.

- The dataset should be extended by including more storms (also before 1970). Also, the variability range of the surge and explanatory variables should be increased in order to reveal dependencies. The variability in the current dataset is limited because only those that produced the highest water levels were selected.

- Additional information of the linear relationship among the weather conditions and the surge could be extracted by *Principal Component Analysis* (Smith, 2002). This technique reveals correlations between variables in terms of eigenvectors of the covariance matrix, also called principal components. A principal component that accounts for the largest variability of the surge describes the variables that are most correlated with this variation. The PCA technique does not assume linearity.

5 References

Alkyon (2009) *A viability study of a prototype windstorm for the Wadden Sea surges*, G. Lipari and G. van Vledder, Alkyon report A2239, May 2009

RIZA (2004) *Opschaling van afvoergolven en stormen*, V.A.W. Beijk en C.P.M. Geerse. RIZA-Werkdocument 2004.075x. RIZA Lelystad, April 2004

Deltares (2009a) *Storm surge duration and stom duration at Hoek van Holland*, Tijssen A. and F. Diermanse, Deltares report 1200264.001, November 2009

Deltares (2009b) *Probabilistic Model Wadden Sea.* J.V.L. Beckers, Deltares report 1200103.057, July 2009

Garson, G. (2009) *Multiple regression*. Statistical notes, North Carolina State University, http://faculty.chass.ncsu.edu/garson/PA765/regress.htm

RIZA (2006) *Hydraulische Randvoorwaarden 2006 Vecht- en IJsseldelta - Statistiek IJsselmeerpeil, afvoeren en stormverlopen voor Hydra-VIJ*. C.P.M. Geerse. RIZA-werkdocument 2006.036x. RIZA Lelystad, January 2006;

Daniel P. Haie B., Aubail X. (2009) *Operational forecasting of tropical cyclones storm surge at Meteo-France*. Marine Geodesy, 32, pp 233 – 242, April 2009

Hamilton, L. (2008) *Low-Tech causal modelling*, University of Hampshire, http://pubpages.unh.edu/~lch/causal2.pdf

Smith, L.I. (2002) *A tutorial on Principal Component Analysis*, http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf