Development of an adjoint free 4Dvar and application to storm surge models

Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Xiangyang Zheng

Kiel, 2016

Erster Gutachter:	Prof. Dr. Roberto Mayerle
Zweite Gutachter:	Prof. Dr. Thomas Slawig
Tag der mündlichen Prüfung:	05.12.2016
Zum Druck genehmigt:	07.12.2016

gez. Prof. Dr. Natascha Oppelt, Dekanin

Declaration

I declare that

- apart from the supervisor's guidance the content and design of the thesis is all my own work;
- the thesis has not been submitted either partially or wholly as part of a doctoral degree to any other examining body and it has not been published or submitted for publication;
- the thesis has been prepared subject to the Rules of Good Scientific Practice of the German Research Foundation.

Signed:

_____

Date:

_____

Acknowledgements

First of all, I would like to express my sincere gratitude to my supervisor Prof. Dr. Roberto Mayerle for his great support and guidance during my PhD study. He has provided me opportunities in conducting researches in the project "Decision support system for the sustainable coastal development of the Shandong Peninsula, China" and in the fields of coastal ocean modeling and data assimilation. He gives me strong motivation and critical feedbacks in writing this thesis.

I would like to thank my Chinese supervisor Prof. Shi Ping for his support for my study in Germany. After I finished my master study in Yantai, he encouraged me to continue PhD study in Kiel University, which is the most important decision in my life. I appreciate his help over these years.

My special gratitude goes to Dr. Karl Heinz Runte. He helps me a lot not only in the project, PhD study but also in my life in Germany. He is the first to show me around Kiel and we often have interesting talks about the world.

Thank Dr. Jose Manuel Fernandez Jaramillo for his guidance in learning data assimilation and writing this thesis. He is knowledgeable and warmhearted. The discussions with him are very interesting and helpful.

I would like to thank three colleagues in the Chinese project, Dr. Gerd Bruss, Guilherme Dalledonne and Natacha Fery. We had good collaboration in the project about Shandong Peninsula. The measurement campaigns we performed together for the project is an unforgeable experience.

I appreciate the help and supports from Dr. Xing Qianguo, Dr. Liu Xin, Dr. Tang Cheng and other friends from Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences.

Thank all colleagues Kadir Orhan, Qingyang Song, Dr. Peter Weppen, Marcos Carvajalino, Katarina Niederndorfer and Dr. Simon Adriaan van der Wulp in CORELAB for their kindly help during these years.

Finally, I must thank my family, my parents, my wife Ding Yana and my two-year-old daughter Zheng Nuoxuan. It is their expectation and love that support me to complete the PhD study.

Abstract

Data assimilation is a methodology which optimizes the extraction of reliable information from observations and combine it with, or assimilate it in, numerical models. The developments of observation techniques, numerical models and computation abilities have made data assimilation an important and attractive field within the communities of oceanography and meteorology.

In this thesis a data assimilation scheme based on the adjoint free Four-Dimensional Variational(4DVar) method was proposed. This method works in a similar way to the incremental 4Dvar but avoids adjoint equations, which are difficult to develop and maintain for large scale models. The tangent linear equations are estimated explicitly by running a set of models. Theoretically, this method is equal to the adjoint method in terms of effectiveness and efficiency for low dimensional model parameters. For high dimensional parameters, the dimension needs to be reduced to avoid the under-determined problem, and principal components analysis (PCA) are used for dimension reduction. A series of twin experiments indicate that the proposed adjoint free 4Dvar is effective for parameter estimate. It is able to recover the contaminated low dimension model parameters to their true values.

The proposed method is then applied to an existing operational coastal forecasting system for the German North Sea. Since storm surges are major threats to the coast of the German Bight, the study is focused on the improvement of storm surge simulation. The data assimilation scheme is applied to a severe storm which occurred in the North Sea on December 5, 2013. Sensitivity tests show that the uncertainties of wind drag coefficient C_d has significant effect on the accuracy of storm surge model. This is probably due to the missing wind-wave coupling process in the storm surge model. Waves during a strong storm also enhance the sea bed shear stress, which would influence the storm surge simulation. C_d and bed drag coefficient are adjusted by the proposed data assimilation, and model skills are improved significantly. Compared to the storm surge model with function of C_d proposed by Smith, the root mean square deviations between model and observations decrease by 60% - 90%. The updated C_d is compared with the C_d calculated from a wave

model, showing that C_d after data assimilation are more reasonable than before data assimilation.

Zusammenfassung

Datenassimilation ist eine Methode, welche die Extraktion verlässlicher Informationen aus Naturmessungen optimiert und diese mit numerischen Modellen kombiniert oder in letztere assimiliert. Die Entwicklung von Messmethoden, numerischen Modellen und computergesteuerter Datenverarbeitung, hat Datenassimilation innerhalb der ozeanographischen und meteorologischen Disziplinen zu einem attraktiven und wichtigen Forschungsgebiet gemacht. In dieser Arbeit wird ein Datenassimilationsverfahren vorgeschlagen, welches auf der adjungierten freien Methode der 4-dimensionalen variationalen Datenassimilation (4Dvar) basiert. Sie funktioniert ähnlich wie die schrittweise 4Dvar, vermeidet jedoch adjungierte Gleichungen, die für grossskalige Modelle schwer zu entwickeln und zu führen sind. Die linearen Tangentialgleichungen werden über Modellläufen explizit ermittelt. Im Bezug auf die Effektivität und Effizienz, ist die Methode für niedrigdimensionale Modellparameter theoretisch gleich der adjungierten Methode. Für hochdimensionale Parameter muss die Dimension reduziert werden, um unterbestimmte Gleichungssysteme zu vermeiden. Hierfür wird die Hauptkomponentenanalyse (Prinicipal Component Analysis, PCA) verwendet. "Twin" Experimente zeigen erfolgreiche Parameterabschätzungen durch das vorgeschlagene adjungierte freie 4Dvar-Verfahren. Es kann fehlerbehaftete niedrigdimensionale Modellparameter an die wahren Werte annähern. Die vorgeschlagene Methode wird in einem bestehenden operativen Vorhersagesystem für die Deutsche Nordseeküste eingesetzt. Da starke Sturmfluten die größte Bedrohung für den Küstenschutz in der Deutschen Bucht darstellen, konzentriert sich die vorliegende Studie auf die Verbesserung von Sturmflutsimulationen. Das Datenassimilationsschema wird auf den starken Orkan in der Nordsee vom 5. Dezember 2013 angewendet. Sensitivitätsstudien zeigen, dass Unsicherheiten beim Windschubkoeffizienten (C_d) einen signifikanten Effekt auf die Genauigkeit des Sturmflutmodells haben. Dies ist höchstwahrscheinlich auf der fehlenden Wind-Wellen Kopplung im Sturmflutmodell zurückzuführen. Zusätzlich verstärken Wellen in einem starken Sturm die Schubspannung am Boden, was einen Effekt auf das Sturmflutmodell hat. C_d und Sohlschubkoeffizient werden mit dem vorgeschlagenen Datenassimi-

lationsschema angepasst, wodurch die Modellleistung signifikant verbessert wird. Im Vergleich zu einem Sturmflutmodell mit einer Funktion zur Berechnung von C_d nach Smith, kann die mittlere quadratische Abweichung der Modellergebnisse von den Messwerten um 60% bis 90% reduziert werden. Der aktualisierte C_d wird außerdem mit dem im Wellenmodell berechneten, C_d verglichen, und es zeigt sich, dass ersterer nach der Datenassimilation bessere Ergebnisse erzielt als davor.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Coasts are the transition areas between land and seas. About half of the world's population live within 200 kilometers of coastlines in 2008 (Dahl and Støttrup, 2012). Coastal areas have advantages over inland areas in terms of transportation, industry developments, tourism and fish farming. However, large amount of population and economic activities impact the coastal environments significantly, making coastal areas increasingly vulnerable to natural disasters, such as tsunamis, storm surge and coastal erosions.

In order to make a better use of ocean resources and to protect the ocean environment, it is necessary to investigate the ocean. Studying movement and physical attribute of ocean water is often the first step. This includes water level, current velocity, temperature of sea water and salinity. In the last decades, the measurement techniques are improved considerably. Tidal gauge networks along the coasts measure water level change; buoys on the sea surface can obtain the surface current velocity, sea water temperature, salinity and wave information; coast based high frequency radars and X-band radars measure surface current velocity and waves in the range from tens to hundreds of kilometers; satellite can measure ocean in a larger areas. These instruments provide better understanding of ocean. The developments of telecommunication technique facilitate data transfer. Therefore, real-time measurements can be available even if the instruments are deployed in a very remote area.

Numerical ocean modeling is another tool to study ocean. Ocean modeling is a relatively new method proposed only after the fluid mechanics was well developed and the modern computers were powerful enough. The great progress of ocean observations and modeling in the recent decades facilitates the set-up of operational

ocean forecasting systems, which is comparable to a weather center in the sense that it provides ocean state forecast over several days in the future.

One of the most concerned problems is the forecast accuracy, which mainly relies on how well the model can represent the real processes and the quality of model input data. Observations are regarded to be more reliable but usually very sparse and can not represent the complete ocean state both in time and space. Therefore, only models are used to obtain the general ocean state and observations are often used for the model validation by comparing the model results and observations at some locations. Apart from models and observations, data assimilation is another crucial component of an ocean forecast system. Data assimilation has been used in meteorological modeling since 1950's, aiming to improve the weather forecast by adjusting the initial conditions since weather evolution is mainly an initial condition problem. There are similarities between ocean models and meteorological models. Both of them are based on geophysical fluid equations, the model state dimension is usually very large and both of them require large amount of computation resource. Therefore, ocean forecasters borrow many concepts from their meteorological counterparts. Data assimilation methods used in meteorological models, such as optimal interpolation (OI), three-dimensional variational (3Dvar) and four-dimensional variational (4Dvar) are more and more used in the ocean model study. On the other hand, ocean models have special properties due to the unique properties and processes of the ocean. In addition, the coverage of ocean measurements is much smaller than the weather observations, making the ocean data assimilation less effective than that in weather forecast.

Data assimilation improves model skill by combining models and observations in geoscience. There are various forms of data assimilation, and least square method is very commonly used among them. As early as in 1801, the German mathematician Johann Carl Friedrich Gauss used least square method to successfully predict the future path of Ceres (a dwarf planet which orbits between Mars and Jupiter) by fitting the observations to an orbit equation of Ceres (Lewis et al., 2006). However, the implementation of data assimilation to meteorological models was started only after the computer was invented in 1940's. The complexity of meteorological or ocean models makes the implementation very difficult, especially for the method of 4Dvar. The difficulty is based on the fact that 4Dvar uses the models as the constraint and thus requires adjoint equations. The objective of this thesis is to develop a practical method to implement 4DVar to a coastal ocean model.

1.2 Aims of the study

In order to decrease the implementation difficulty of data assimilation, the first aim of this thesis is to propose a new scheme which follows the principle of 4Dvar but in an easier way than adjoint method. Another aim of this thesis is to implement the proposed scheme into a storm surge model, which is a part of an existing coastal forecasting system for the German Bight developed by Fernández Jaramillo (2014) in the Research and Technology Centre Westcoast, Kiel University (FTZ).

The German Bight is the southeast part of the North Sea (Figure 1.1), and adjoins the Netherlands, Germany and Denmark. Storm surge is an abnormal change of sea water level caused by strong wind and low air pressure. The German Bight is particularly vulnerable to storm due to the coastal geometry. On January 31 1953, a combination of a high spring tide and a severe European windstorm over the North Sea caused a strong storm surge; it led to extensive flood in the Netherlands, Belgium, England and Scotland, and 2533 deaths in the surrounding countries of the North Sea (Wikipedia, 2016). On February of 1962, a severe storm swept the German Bight, causing severe damage along the German North Sea coast and costing more than 300 lives in Hamburg (von Storch et al., 2008). In the early December 2013, the winter storm Xaver attacked the North Sea area, resulting in severe flood in England, France, the Netherlands, Germany and Denmark. (See Figure 1.2 for the impacts of these storm surge events)

The great losses brought by storm surge indicate the necessity of storm surge forecasting system. Actually, in the past 20 years, the death toll directly caused by storm surges has decreased a lot compared to previous storm surge events all over the world. One reason is the strengthening of the coastal dyke systems. The second reason is the great improvements of weather forecast models together with the storm surge early warning systems; guided by the early warning, people can be evacuated to a safer place in advance (Webster, 2008).

With respect to the operational system of FTZ, the model Delft3D is used for the numerical simulations of water level. Observations from tidal gauges are used for the model validation and data assimilation. The operational system integrates several data assimilation methods, including linear regression, artificial neural networks and fuzzy logic. Using these methods, observations of water level are assimilated to improve the open boundary conditions and thus improve the nowcasting and forecasting of water level.

Open boundary is the artificial interface between the open sea and the domain of a coastal ocean model. Open boundary conditions are significant for water level simulation in the modeled area. The tide generating forces inside the modeled area

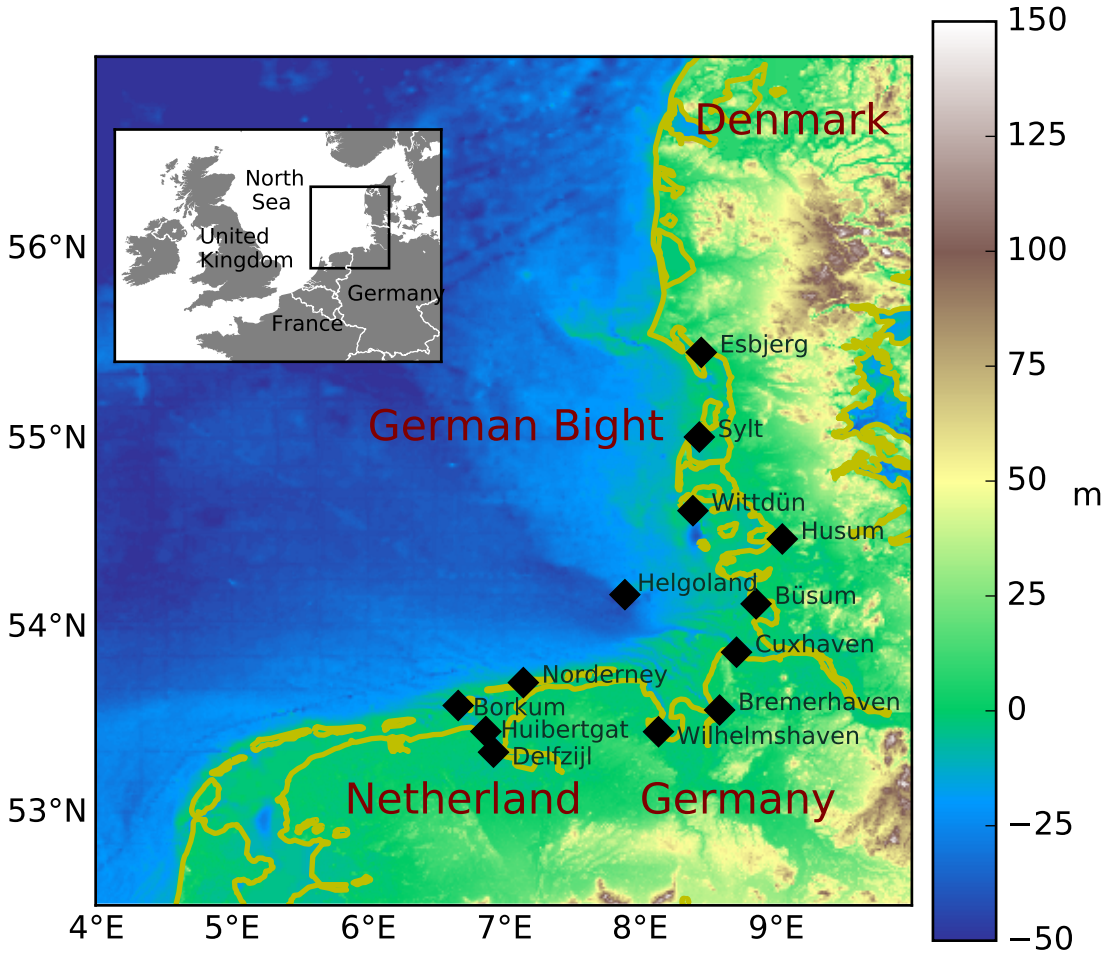

Figure 1.1: German Bight

are negligibly small and most tidal energy comes from the deep ocean through open boundary. Wind stress imposed over the sea surface is another driving force in a coastal model. Under normal weather conditions, water level induced by wind stress is insignificant compared with tide in most coastal areas. But the effects of wind may be higher than the tide during strong storms. Up to the present day, the mechanism of air-sea energy and momentum transfer is still not well understood (Letchford and Zachry, 2009), which may result in significant uncertainties of storm surge models. In this thesis, the new data assimilation method based on 4Dvar is used to adjust wind stress of the storm surge model. The method proves to be effective to improve storm surge simulations and relatively easy to be implemented.

Figure 1.2: A: A breach at Erith, England after the 1953 flood. B: One street in Hamburg after the 1962 storm surge. C: The flooded Hamburger Fishmarket in December 2013.

1.3 What is data assimilation?

Broadly speaking, data assimilation means bringing data from different sources together and reaching a better estimate. One example is using two thermometers to measure temperature in a room. Suppose the temperature measured by them is T_1 and T_2 respectively. The two values can be assimilated to obtain a new value of temperature T by simply averaging T_1 and T_2:

$$T = \frac{T_1 + T_1}{2} \tag{1.1}$$

In the meteorological or oceanographic communities, data assimilation often means assimilating observations to numerical models and improves the model skills. Complex model dynamics and high dimension add enormous difficulties to data assimilation. Robinson and Lermusiaux (2000) defined the data assimilation as "a combination of observational data with the underlying dynamical principles governing the system under observation". They emphasized that data assimilation can make possible efficient, accurate and realistic estimations which might not otherwise be feasible. This means that the ordinary method may not work on the large scale meteorological or ocean models. Novel methods are necessary to implement

the data assimilation. They also summarized the purposes of data assimilation for ocean models as listed below:

- the control of errors for state estimates;
- the estimation of parameters;
- the elucidation of real ocean dynamical processes;
- the design of experimental networks;
- and ocean monitoring and prediction.

The control of errors for state is also known as state estimate. It is usually used in weather forecast to improve the initial conditions. Therefore, state estimate is a special case of parameter estimate. In this thesis, we only discuss parameter estimate using data assimilation.

1.4 4Dvar and its implementation problem

4D variational method is the latest proposed among the common data assimilation. "4D" indicates the method uses observations over both space (3D) and time (1D). "variational" indicates it is based on the variational calculus, which is a mathematical analysis that deals with maximizing or minimizing functionals. The advantage of 4Dvar over sequential data assimilation methods is that it considers model dynamics. Therefore, the adjusted model parameters are dynamically more consistent than sequential methods, such as optimal interpolation and 3Dvar. The second advantage is that 4Dvar uses the adjoint equations and thus efficient high dimension data assimilation becomes possible.

However, it is difficult to develop and maintain the adjoint equations for meteorological or ocean models. One approach of reducing this difficulty is using automatic differentiation tools to generate the adjoint equations given the source codes of numerical models. Models are usually very complex and have been used for many years. It is also a big job to tidy the code before giving to the automatic differentiation tools. Therefore, in practice, most adjoint models are still developed by hand, rather than using automatic tools (Lawless, 2013).

1.5 Methodology

In this thesis, a new data assimilation method is developed. This method is a variant of 4Dvar. It also assimilates observations over a given period. Unlike

traditional 4Dvar, the new method uses a model ensemble to represent the tangent linear equations explicitly and thus avoids the adjoint equations. Therefore, the implementation difficulty of 4Dvar is decreased.

The method is then used to improve the storm surge model skills by adjusting critical model parameters. Before applying the data assimilation method to the storm surge model, several sensitivity tests are carried out to investigate the effect of model parameters on model skills. After that, data assimilation are used to adjust those model parameters which bring most uncertainties in the storm surge simulation. The ability of the method in improving accuracy is tested using data assimilation for recent severe storm surges. In this study, the storm Xaver in early December 2013 was used.

1.6 Outline

The outline of this thesis is:

Chapter 2 will review the previous studies on data assimilation and storm surge modeling.

Chapter 3 will introduce of the model Delft3D and show the preliminary results of the storm surge model for the German Bight.

Chapter 4 will propose the data assimilation method of adjoint free 4Dvar.

Chapter 5 will validate the proposed adjoint free 4Dvar using a series of twin experiments.

Chapter 6 will investigate the effects of some model parameters on the storm surge simulations.

Chapter 7 will apply the proposed adjoint free 4Dvar to the storm surge model.

Chapter 8 will conclude this thesis and give recommendations for the future study.

Chapter 2

Literature review

2.1 The history of data assimilation

2.1.1 Data assimilation in early stage (Daley, 1993)

Data assimilation dates back to the year 1854. A severe snow storm resulted in large losses in Europe. The French meteorologist LeVerrier collected all the observations around Europe before and after this snow storm and made a series of synoptic charts, showing the weather state all over Europe. People realized that these synoptic charts were helpful for forecasting weather. In the beginning of the 20th century, the invention of telegraph made the real-time synoptic charts possible.

Synoptic charts were constructed by meteorologists manually based on their judgments and experiences. This process now is called subjective analysis. Subjective analysis reached to its summit after the second world war. The synoptic charts do not fit to the concept of modern data assimilation, but it is still a form of data assimilation because it combines the observed data to the people's experiences and can improve weather forecast.

Bjerknes thought the weather forecast can be resolved as an initial condition problem and Richardson put this idea into practice in 1922. However, Richardson's attempt failed due to the bad initialization in the way of subjective analysis. In 1950, the second attempt was carried out in Princeton Institute of Advanced Study. The results were much better than the previous experiment. But the prediction took much longer time than the simulation time because the subjective analysis was very time-consuming.

2.1.2 Objective analysis

To speed up the weather forecast process, objective analysis was proposed.

The first objective analysis was performed by Panofsky (1949). He used a polynomial expansion to fit all the observations in an area. The coefficients of the polynomial expansion were solved by least square method and the weight is related to the accuracy of the observations. The number of the coefficients can control the smoothness of the analysis. Cressman (1959) also used the polynomial expansion to fit the observations, but he introduced the localization and physical balance; he also suggested that the results from the previous forecast can be combined with the observations to improve the analysis. This was the first time the prior estimate was proposed.

Berthorsson and Doos put forward a new scheme, successive correction, in 1955 (Daley, 1993). They used observations to correct a background field which may be from previous forecast or climatological data. The background field was subtracted from the observations and the result is called observation increment. Then the background value on each grid point is added to observation increment multiplied by a coefficient. The coefficient is inversely proportional to the distance between grid point and observation points. They also used automatic data quality control for the first time.

Optimal interpolation (OI) is a more advanced objective analysis (Gandin and Hardin, 1965). It is similar to the successive correction in terms of the background field and observation increment. The difference lies in how to get the coefficient of the observation increments. The coefficient of OI is not simply related to the distance, but represents a linear best unbiased estimate given the error of background field and observations. OI was very common in 1970's and 1980's in the weather centers all over the world for operational forecasting.

2.1.3 Model dynamics related data assimilation

The objective analysis combines the background field and observations, improving the initial condition. It is not difficult to implement for operational forecasting. But one problem is that data assimilation at one fixed time may impact the dynamic balance of different variables, because they do not consider the dynamic balancing or just use over-simplified relations. To solve this problem, the model itself must be used in the data assimilation to keep the dynamic balance. There are two directions to integrate the model in the data assimilation.

One is Kalman filter, which was developed by Kalman (1960). For linear models and linear observation operators, Kalman filter gives the best estimate in the sense of unbiasedness and least variance, provided the background initial condition and its covariance. Meteorological models are nonlinear in most cases. Therefore extended

Kalman filter was developed for nonlinear problems. But the high dimension of meteorological model prevents the application of Kalman filter into operation. Evensen (1994) proposed the ensemble Kalman filter(EnKF) as a solution. The probability features of model results can be represented by ensemble and evolve with the model dynamics. The typical number of ensemble ranges from 50 to 100. The ensemble members can run independently and in parallel. This made the EnKF relatively efficient and easy to implement. From that time, a number of ensemble based Kalman filter were proposed (Anderson, 2001; Bishop et al., 2001; Nerger et al., 2012).

The other widely used dynamics related data assimilation is four dimension variational (4Dvar) method. It is used to estimate the poorly known model parameters and initial conditions. For a given model, the model results can be viewed as the function of model parameters. The discrepancies between model results and the observations are denoted as a scalar J. The problem becomes the minimization of J with the model as a constraint. The gradient of J with respect to model parameters is required to find the optimal value (Figure 2.1). However, if the model is complex and has a high dimension, it is not an easy task to minimize J. Sasaki (1958) proposed a variational analysis method with a model as the constraint in late 1950's, but at that time it was impossible to implement this method to a high dimension model. Then he proposed the adjoint method to compute the gradient of a cost function (a scalar measuring the discrepancy of model results and observations), which became the standard method to compute the gradient in 4Dvar (Sasaki, 1970). In 1994, the introduction of the incremental 4Dvar method made the application of 4Dvar for large models in practice (Courtier et al. (1994)). With the progress of computation ability, many regional weather service centers began to use 4Dvar for their operational forecasting systems (Rabier et al. (2000), Clayton et al. (2013)).

2.2 The previous researches of ensemble methods and 4Dvar for meteorological and ocean models

Since Evensen (1994) proposed the EnKF, many researchers began to improve Evensen's method or develop a variant of it. Houtekamer and Mitchell (1998) mentioned that in order to prevent filter divergence, perturbations should be added to the observations; they also found 100 ensemble members are enough to accurately represent the local anisotropic, baroclinic correlation structures. Burgers et al. (1998) further explained theoretically that perturbations must be added to

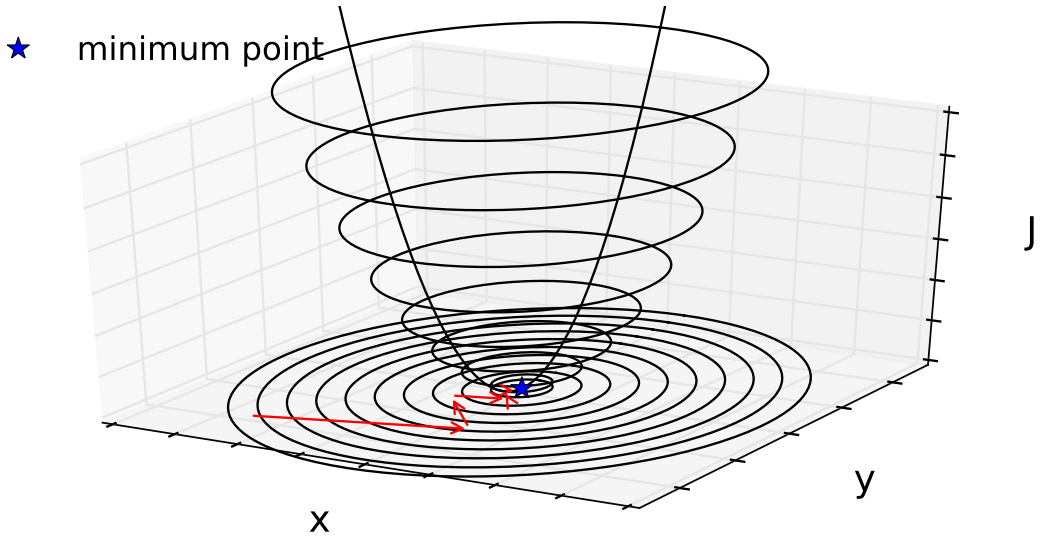

Figure 2.1: Illustration of using gradient method to find the minimum point. Red arrows indicate the inverse gradient directions

the observations in the Evensen's EnKF method, otherwise the EnKF would not be consistent with the standard Kalman filter and result in filter divergence. However, the perturbations of observations may be bound to sampling errors. To avoid such errors, another ensemble method called ensemble square root filter (EnSRF) was developed. There are different filters in the family of EnSRF. Anderson (2001) devised the ensemble adjusted Kalman filter (EAKF). Bishop et al. (2001) developed the ensemble transform Kalman filter (ETKF). Nerger et al. (2012) unified the common EnSRF filters and pointed out that they are very similar in terms of precision and computation time.

The ensemble methods mentioned above assimilate model results and observations at a certain time other than over a certain period. When the observations are fairly frequent, the model must pause and restart frequently. This not only takes a long time for operational application but also may destroy the smoothness of model results. An alternative ensemble based method is 4DEnKF, or asynchronous EnkF, proposed or implemented by Hunt et al. (2004), Fertig et al. (2007), Sakov et al. (2010). The time when the data assimilation is being performed is different from the time at which observations are available. Therefore, the observations within a period can be assimilated simultaneously. They demonstrated that 4DEnKF is identical with 4Dvar in principle but avoids the linear tangent model and adjoint model. The way to implement 4DEnKF is almost the same as standard EnKF, making it a potential alternative to 4Dvar.

Since the proposal of adjoint model by Sasaki (1970), the study on 4Dvar is

much more than Kalman filter in the meteorological models. Many weather service centers are using 4Dvar to get a better estimate of initial condition for forecasting as mentioned above. 4Dvar has also been used in ocean model for parameter estimate. Ngodock and Carrier (2014a) developed an assimilation system with 4Dvar and the Navy Coastal Ocean Model; this system can improve the initial conditions and other external model forces; they have validated the system by assimilating the satellite measured sea surface temperature and sea surface height(Ngodock and Carrier (2014b)). Janeković et al. (2013) applied 4Dvar to a coastal model for the island of Oahu, Hawaii; he examined the influences of initial conditions, meteorological forces and forces from open boundaries and pointed out that coastal studies may not be initial value problems, rather they are forced problems that require a knowledge of the large-scale energy propagated into the region. Zhang (Zhang and Lu (2008a), Zhang and Lu (2008b), Zhang and Wang (2014)) published a series of study about applying adjoint method to estimate the open boundary conditions, bottom friction for coastal ocean models.

2.3 Previous study of storm surge model

2.3.1 History

The efforts of forecasting storm surge date back to Conner et al. (1957). They used an empirical method for the water level peak forecasting in tropical storms. They proposed a relation between the surge h and the air pressure p_0 at the center of the storm, $h = B(p_n - p_0)^b$. p_n is normally equal to 1005; B and b can be solved with least square method given data of p_0 and h. 30 pieces of data from the stations on the coast of Gulf of Mexico were used and the resulting equation is $h = 0.867(1005 - p_0)^{0.618}$. Then the relation was applied on stations on the east coast of the US. It fit reasonably well for the storms which move on land definitely. This method is simple as only one parameter is used for forecasting but it can account for half of the variability of the storm surge. The authors further mentioned that other parameters were also important but they cannot be used in this method.

Christiansen and Siefert (1978) used an empirical method to do the surge prediction in the German Bight. The method is based on the correlation of water level between the data of wind speed and wind direction at two stations. Although the physical mechanisms was not clear, the method was effective as the authors mentioned that the accuracy was 30 minutes in time and 20cm in height.

Jelesnianski (1966) is one of the earliest authors who used numerical methods to simulate the storm surge. He mentioned that numerical models are preferable

than the empirical methods because it considered more factors. He used a linearized model without bottom stress to study the influences of model parameters on storm surge. His model was useful for computing short-duration transient coastal surges. He found the coastal surge is not sensitive to any bottom friction law. He also found a correlation between the center air pressure and peak surge similar to what Conner et al. (1957) proposed.

Heaps (1969) made a storm surge simulation in the North Sea surrounding the British islands with a 2D linearized model. The wind and air pressure data were from hourly weather charts. The total water level was assumed to be the summation of tide, wind surge and barometric surge. Only the wind surge was computed by the model.

In the late 1960's, high-speed computers allowed the development of numerical approximations to the governing equations using structured computational grids first in two spatial dimensions (vertically integrated) and subsequently in three dimensions. More realistic forces can be added to storm models. Researchers started to investigate the mechanism of storm surge in more detail.

Prandle (1975) had a grid network extending through the Strait of Dover into the English Channel. These models had finer mesh than the shelf model of Heaps (1969). They included non-linear terms with tide as well as surge prescribed along the open boundaries. Prandle specified tidal elevations along the northern and southern boundaries in terms of seven important harmonic constituents including the major semi diurnal components M2 and S2. The storm surge of January 31 1953 was reproduced satisfactorily by Prandle's model.

The increasingly powerful and affordable computation and storage resources made the three dimensional (3D) storm surge simulations much easier. Although water level is a two dimensional (2D) variable, 3D models can produce more detailed current velocity. In 2D models, the bottom shear stress is computed with depth averaged current velocity whereas in 3D models the current velocity on the bottom layer is used. Obviously, 3D models represent more realistic mechanisms. Cooper and Pearce (1982) applied a 3D model to hindcast two tropical storms. They pointed that the bottom shear stress is much larger than the surface shear stress during the storms. They proposed that waves are the dominant reason for the large bottom shear stress displayed in the data. Zheng et al. (2013) applied the Finite Volume Coastal Ocean Model (FVCOM) to simulate the Hurricane Ike in the US. They compared the results of 2D model and 3D model. The surge height in the 3D model is higher than that in 2D model due to the larger current velocity in the computation of bottom stress in the 2D model. 2D model and 3D model have similar

results after calibrations. They thought 2D model is a more straight-forward and efficient approach to storm surge simulation.

Most storm surge modelers only used a 2 dimensional shallow water model, and they can obtained fairly good results. But the contributions of waves to the storm surge is also important. Bruss and Mayerle (2009) showed a comparison of storm surge hindcast with and without the effect of waves. They found simulation with waves can be improved significantly. The effects of wave on storm surge mainly consist of radiant stress, wind drag coefficient and bottom roughness. Longuet-Higgins and Stewart (1964) discussed how the radiation stress is generated. Mastenbroek et al. (1993) applied the radiation stress in his wave-surge coupled model of the North Sea. They found the radiation stress can increase the storm surge about 5% in one storm but can be negligible for another two storm surge events. Mastenbroek et al. (1993)'s work also showed that the wind drag coefficient can be improved by the wave effect compared with the model using wind drag coefficient from (Smith and Banke, 1975). Thus the surge simulation can be improved. They found the improvement was substantial during a fast-moving storm. In recent year, there are many more sophisticated wave-surge coupled model. Sheng et al. (2010) used a storm surge modeling system, CH3D-SSMS, which includes coupled coastal and basin-scale storm surge and wave models; they successfully simulated measured winds, waves, storm surge, currents, and inundation during the hurricane Isabel. They revealed noticeable effects of waves on storm surge but found a contradicting effect to Mastenbroek et al. (1993); they pointed that the radiation stress is more important than the wave induced bottom stress. In shallow water, the orbital velocity induced by wave has a significant influence on the bottom, therefore, the variation of bottom shear stress during storm surge cannot be neglected. Wu et al. (1994) showed the bottom shear stress was increased when the wave effect was added to the storm surge model.

2.3.2 The interactions of tide and surge

One of the interests in the storm surge studies is the tide-surge interactions. A tendency that surge peaks in the Thames estuary occur most frequently on the rising tide has been recognized for a long time (Doodson and Dines, 1929). Rossiter (1961) assumed idealized surges with diurnal periodicity and showed how a negative surge would retard tidal propagation whereas a positive surge would advance high water. Prandle and Wolf (1978) examined tide gauge data from nine ports along the UK's east coast over the period 1969–1973. They confirmed the tendency for surge peaks to occur most often on the rising tide, and used numerical models to draw

a conclusion that this pattern arises irrespective of the phase relationship between tide and surge in the northern North Sea. The models made it possible to separate the contribution to interaction from shallow water and bottom friction. Horsburgh and Wilson (2007) analyzed water level data from five tidal gauges. They found the surge peaks occurred 3 to 5 hours before the nearest high water level. They also revealed that the surge peak will avoid high water for any finite tidal phase shift and increasing tidal range can make the surge peak farther from the high water. Numerical experiments indicate that nonlinear bottom friction (described by the quadratic formula) is a major factor to predict these oscillations while the nonlinear advective terms and the shallow water effect have little contribution (Zhang et al., 2010). All the conclusions of the tide-surge interactions studies indicate that tide is necessary for a reliable operational forecasting of storm surge.

2.3.3 The wind field model

Wind data is one of the most important factors for a good storm surge forecasting. Storm surge models work on a large domain, and it is impossible to measure wind in such a big domain. Therefore, modeled wind is always used. In the early days, idealized wind fields were used in the hindcast studies. In the storm surge simulation of Heaps (1969), the wind data with two-hour interval were generated by geostrophic winds from weather charts (weather charts showing air pressure and weather fronts) and then adjusted to the surface winds. Flather (1984) used a similar method to derive the wind field in the storm on January 31 1953. Since the late 1970s', the numerical weather models can provide meteorological data for storm surge model. Flather and Proctor (1983) used the wind and air pressure data from the Meteorological Office's fine mesh 10-level weather prediction model. Beardsley et al. (2013) used the surface forcing fields which were computed on a 9 km $\times$9 km grid using the National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) Weather Research and Forecast (WRF) mesoscale model and the Medium Range Forecast planetary boundary layer scheme. The modeled wind data have been working well for most extra tropical storm surge models. However, in some areas where weather models were not available, the weather charts are still very good data sources (Johnson and Kowalik, 1986). But the modeled wind fields in a tropical storm are still not good enough in an operational application. Most storm surge models for hurricanes used the parametric wind fields.

The tropical cyclone numerical modeling is very complex. Reconstructing a tropical cyclone normally requires four parameters: (1) central pressure p_c (2) radius

to maximum wind R. (3) speed and direction of movement of the cyclone eye V_{FM} and θ_{FM} (4) the wind speed U_{10} 10m above the sea level. Many studies (Mulligan et al., 2008) used the method proposed by Holland (1980) to model the wind and air pressure field.

2.3.4 Wind drag coefficient

Air transfers momentum to ocean by wind shear stress. It is the force on the sea surface per unit area. Wind shear stress (τ_s) is calculated by the quadratic drag law:

$$\tau_s = C_d \rho_a |U_{10}| U_{10} \tag{2.1}$$

where C_d is the wind drag coefficient and ρ_a is the air density. Almost all the storm surge models use a linear function of C_d with respect to the wind speed U_{10},

$$C_d = 10^{-3}(a + bU_{10}) \tag{2.2}$$

where a and b are two unknown parameters. They are determined by measurements of wind shear stress. The valid range of equation (2.2) in most cases is under intermediate wind speed, about from 8 to 22 m/s. Wu (1982) proposed that the function can apply the wind ranging between 2 m/s and 50 m/s. Smith (1988) argued that the linear function can not be used to high wind speed.

2.3.5 Data assimilation on the storm surge models

Yu and O'Brien (1991) applied the 4Dvar to a modified Ekman model. The control variables were wind drag coefficient and eddy viscosity. By assimilating observations, they obtained the optimal estimates of model field. They found that the small interval of data can improve the results.

Heemink et al. (1995) pointed out that the method of optimal interpolation cannot be used for a coastal shallow water flow model because it can result in numerical instabilities and contaminate the model results. The reason is that the error covariance needed by optimal interpolation is inadequate for a dynamically consistent correction. They proposed to use Kalman filter on a linearized shallow water model. They added a term to denote the errors from the meteorological data for the evolutions of the model result covariance. Using the extended Kalman filter, the time varying Kalman gain is obtained. They found the Kalman gains are almost invariant. Therefore, he did not need to use a full Kalman filter, but directly use the Kalman gain to update model state when there are observations.

The data assimilation brought improved forecast of the surge, but the improvement vanished very fast. The reason may be that only improving the initial conditions is not enough.

The method of Verlaan and Heemink (1997) is also based on the Kalman filter. The algorithm used a reduced rank approximation of the error covariance matrix. This can decrease the burden of integrating the covariance. Their experiments show that 10 modes are enough for an effective error reduction. The authors claimed that they will apply the method to a non-linear storm surge model. Canizares et al. (1998) used a similar method to assimilate the storm surge in the North Sea in February 1993.

Canizares et al. (2001) implemented the ensemble Kalman filter to the model MIKE21 in the Danish waters. They added errors in the meteorological forcing terms. They mentioned the error covariance matrix becomes nearly invariant after about 2 days of simulation. Due to the assumption of the constant accuracy of the observations, the Kalman gains become invariant as well. Therefore, once the quasi constant Kalman gain is calculated, it is not necessary to propagate the error covariance matrix, which is the most time-consuming in Kalman filter. It is similar to the method of Heemink et al. (1995). Evensen (2009) called this method the ensemble optimal interpolation (EnOI).

Lionello et al. (2006) applied adjoint method to a storm surge model in the northern Adriatic Sea. The control variable is initial condition and the analysis length is 3 days. Their results show that the data assimilation is able to decrease model errors (due to both inaccurate meteorological forcing and model shortcomings) and effectively improves the reliability of the storm surge forecast.

Butler et al. (2012) made a series of twin experiments with the data assimilation of singular evolutive interpolated Kalman (SEIK) filter. Twin experiments were performed on two hurricanes Ike and Katrina in the Gulf of Mexico respectively. Data assimilation can improve the water level forecasts up to 48 hours before landfall of hurricane.

Chapter 3

Model

3.1 Introduction of Delft3D-FLOW

The ocean model used for storm surge simulation is Delft3D. This model suite includes flow, wave, sediment, particle tracking and morphodynamic modules. These modules can run independently or be coupled with other modules. In this thesis, only the flow module is used for the storm surge simulation in the German Bight. Delft3D-FLOW is a multi-dimensional (2D or 3D) hydrodynamic (and transport of sea water temperature and salinity) simulation program which calculates non-steady flow and transport phenomena that result from tidal and meteorological forcing on a rectangular or a curvilinear grid. In 3D simulations, the vertical grid is defined following the sigma (σ) coordinate. For more detailed features, readers are referred to the Delft3D-FLOW manual (Hydraulics, 2007). Delft3D-FLOW has been implemented in a number of areas ranging from continental shelf areas to river mouth areas and lakes (Elias et al., 2001; Mayerle and Zielke, 2005; Jiao, 2014). It proves to be able to simulate hydrodynamic feature well. Delft3D has a complete set of modules for pre- and post-processing, such as mesh generator RGFGRID, bathymetry module QUICKIN, nesting tool NESTHD and MATLAB based post-processing tool QUICKPLOT. These tools decrease the modeling difficulties and make Delft3D very suitable for an operational coastal forecasting system.

3.1.1 Primitive equations for storm surge models

Like many other coastal models, Delft3D-FLOW solves the large scale Navier Stokes equations for an incompressible fluid. Hydrostatic pressure, Boussinesq and shallow water assumptions are used to simplify the equations. Delft3D-FLOW has several turbulence schemes and can deal with the sub-grid turbulence process.

Delft3D solves the hydrodynamic equations with finite difference method. The Navier Stokes equations are discretized on an orthogonal curvilinear grid horizontally (in ξ and η directions). For 3D simulation, Delft3D provides both σ coordinate and z coordinate system in the vertical direction.

The depth-averaged continuity equation is given by:

$$\frac{\partial \zeta}{\partial t} + \frac{1}{\sqrt{G_{\xi\xi}}G_{\eta\eta}}\frac{\partial[(d+\zeta)U\sqrt{G_{\eta\eta}}]}{\partial \xi} + \frac{1}{\sqrt{G_{\xi\xi}}G_{\eta\eta}}\frac{\partial[(d+\zeta)V\sqrt{G_{\xi\xi}}]}{\partial \eta} = Q \tag{3.1}$$

The 2D momentum equations in ξ and η directions are given by:

$$\begin{aligned}\frac{\partial u}{\partial t} + \frac{u}{\sqrt{G_{\xi\xi}}}\frac{\partial u}{\partial \xi} + \frac{v}{\sqrt{G_{\eta\eta}}}\frac{\partial u}{\partial \eta} + \frac{w}{d+\zeta}\frac{\partial u}{\partial \sigma} - \frac{v^2}{\sqrt{G_{\xi\xi}}\sqrt{G_{\eta\eta}}}\frac{\partial \sqrt{G_{\eta\eta}}}{\partial \xi} + \\ \frac{uv}{\sqrt{G_{\xi\xi}}\sqrt{G_{\eta\eta}}}\frac{\partial \sqrt{G_{\xi\xi}}}{\partial \eta} - fv = -\frac{1}{\rho\sqrt{G_{\xi\xi}}}P_\xi + F_\xi + \frac{1}{(d+\zeta)^2}\frac{\partial}{\partial \sigma}(\nu_v\frac{\partial u}{\partial \sigma}) + M_\xi\end{aligned} \tag{3.2}$$

and

$$\begin{aligned}\frac{\partial v}{\partial t} + \frac{u}{\sqrt{G_{\xi\xi}}}\frac{\partial v}{\partial \xi} + \frac{v}{\sqrt{G_{\eta\eta}}}\frac{\partial v}{\partial \eta} + \frac{w}{d+\zeta}\frac{\partial v}{\partial \sigma} - \frac{uv}{\sqrt{G_{\xi\xi}}\sqrt{G_{\eta\eta}}}\frac{\partial \sqrt{G_{\eta\eta}}}{\partial \xi} + \\ \frac{u^2}{\sqrt{G_{\xi\xi}}\sqrt{G_{\eta\eta}}}\frac{\partial \sqrt{G_{\xi\xi}}}{\partial \eta} - fu = -\frac{1}{\rho\sqrt{G_{\eta\eta}}}P_\eta + F_\eta + \frac{1}{(d+\zeta)^2}\frac{\partial}{\partial \sigma}(\nu_v\frac{\partial v}{\partial \sigma}) + M_\eta\end{aligned} \tag{3.3}$$

where

- ζ: water level above some horizontal plane of reference
- d: depth below some horizontal plane of reference
- $G_{\xi\xi}$, $G_{\eta\eta}$: coefficients used to transform curvilinear to rectangular coordinates
- U, V: depth-averaged velocity in ξ direction and η direction
- Q: global source or sink per unit area
- u, v, w: flow velocity in the ξ, η and σ direction
- f: Coriolis parameter
- P_ξ, P_η: gradient hydrostatic pressure in ξ and η directions
- F_ξ, F_η: turbulent momentum flux in ξ and η directions

3.1.2 Sea surface boundary conditions

On the sea surface, wind is an important driving force for ocean hydrodynamics, and wind transfers momentum from air to ocean by shear stress on the sea surface, which provide the sea surface boundary conditions for the momentum equations:

$$\left.\frac{\nu_v}{H}\frac{\partial u}{\partial \sigma}\right|_{\sigma=0} = \frac{1}{\rho_0}|\vec{\tau}_s|\cos(\theta) \tag{3.4}$$

$$\left.\frac{\nu_v}{H}\frac{\partial v}{\partial \sigma}\right|_{\sigma=0} = \frac{1}{\rho_0}|\vec{\tau}_s|\sin(\theta) \tag{3.5}$$

where ν_v is the vertical eddy viscosity, H is water depth and θ is the angle between the wind stress vector and the local direction of the grid-line. The magnitude of wind shear-stress τ_s is determined by the following widely used quadratic expression:

$$|\vec{\tau}_s| = \rho_a C_d U_{10}^2 \tag{3.6}$$

Where:

ρ_a**:** the density of air

C_d**:** the wind drag coefficient

U_{10}**:** the wind speed 10 meter above the free surface (time and space dependent)

U_{10} can be obtained by numerical meteorological models or parametric wind models (Holland, 1980). Wind drag coefficient C_d is often derived by field measurements of wind stress. In the past decades, a number of investigations have been carried out on this subject (Smith and Banke, 1975; Smith, 1980; Large and Pond, 1981). One consensus is that the wind drag coefficient over sea surface is not a constant number, but varies with respect to sea state, wind speed and the atmosphere stratification. Under weak wind conditions, sea surface is smooth and C_d can be taken as a constant. When the wind speed is larger than about 8 m/s, wave height increases and the sea surface becomes rougher; wind over this rougher sea surface transfers more momentum to the ocean, therefore, C_d increases with the increase of wind speed. When wind speed is extremely strong, wave height stops increasing, and wave break and whitecapping will result in energy dissipation with increase of wind speed. Experiments with a bubble annulus indicate that bubble layers in salt water impede transfer of momentum from the wind (M. D. Powell, 2003) if the wind speed is more than 33 m/s. C_d will level off or even decrease.

A number of researches have been done to quantify the relation between C_d and wind speed. The ways to obtain the relation include field measurements, laboratory

experiments and remote sensing. Smith and Banke (1975) used eddy correlation method and the wind measurements on Sable Island to investigate the wind drag coefficient; he found the drag coefficient increases with wind speed at least up to 21 m/s; he obtained a linear relation between wind drag coefficient and wind speed $C_d = 10^{-3}(0.63 + 0.066U_{10})$, $3 \leq U_{10} \leq 21$ m/s; he also mentioned that this relation was valid in the shallow water around Sable Island; in the deep water the drag coefficient may be increased by 10%. Smith (1980) did another investigation about the wind drag coefficient on a platform off the coastline of Holland; he obtained the relation $C_d = 10^{-3}(0.61+0.063U_{10})$ in the range of from 6 to 22 m/s . Large and Pond (1981) applied the Reynolds flux and dissipation methods on wind data measured from a stable tower located in a deep water area; they obtained $C_d = 10^{-3}(0.49+0.065U_{10})$ with U_{10} between 11 and 25 m/s; C_d is a constant 1.2×10^{-3} if the wind speed is lower than 11 m/s. But during a storm when the wind speed is usually larger than 25 m/s, there is no field measurements of wind shear stress and wind speed due to the difficulty of installing the in situ instruments. M. D. Powell (2003) used GPS sonde to measure the wind profiles; these wind profiles were measured in the hurricane eye walls in the Atlantic, Eastern and Central Pacific basins; his results are similar to Large and Pond (1981)'s results or extrapolation of Large and Pond (1981)'s formula when the wind speed is below 35 m/s; however, C_d is much smaller than the extrapolation of Large and Pond (1981)'s formula. Powell (2008) used more than 2400 profiles to study the wind drag coefficient when the wind speed is below 70 m/s. He obtained C_d under different wind speed both in shallow water and deep water, shown in table 3.1.

Table 3.1: Wind drag coefficient obtained by Powell (2008)

shallow water($\leq$20m)		deep water([20m,160m])	
U_{10}	$C_d \times 10^3$	U_{10}	$C_d \times 10^3$
20.3	1.19	19.9	1.01
26.6	1.85	26.8	1.72
33	2.05	33.5	1.95
39.2	3.22	40.6	2.27

There is a wide variation of the break points in equation 3.14 according to different authors. Some of them are shown in table 3.2. In this table a and b are the intercept and slope in the expression $10^3 C_d = a + b|U_{10}|$ respectively.

Table 3.2: Wind drag coefficient formula from different authors

Author	a	b (s/m)
Sheppard (1958)	0.800	0.114
Deacon (1962)	1.000	0.070
Miller (1964)	0.750	0.067
Zubkovskii and Kravchenko (1967)	0.720	0.120
Brocks and Krugermeyer (1970)	1.180	0.016
Sheppard et al. (1972)	0.360	0.100
Wieringa (1974)	0.860	0.058
Kondo (1975)	1.200	0.025
Garratt (1977)	0.750	0.067
Smith (1980)	0.610	0.063
Wu (1969)	0.800	0.065
Large and Pond (1981)	0.490	0.065
Donelan (1982)	0.960	0.041
Geernaert et al. (1987)	0.578	0.085
Yelland and Taylor (1996)	0.600	0.070

Bruss and Mayerle (2009) made a scatter diagram between wind speed U_{10} and wind drag coefficient C_d (Figure 3.1). They set up a flow-wave coupled model for the storm surge simulations in the Baltic Sea. The model was driven by meteorological data from the Ensemble Prediction System (EPS) of the European Centre for Medium-Range Weather Forecasts (ECMWF). The storm surge model can output wind shear stress spatially and temporally, which is then used to compute wind drag coefficient by means of equation (3.6). In Figure 3.1, the sample density is shown in percentage of the total number in logarithmic gray scale. It can be found that C_d indeed increases with the increase of U_{10} in the wind speed range $10 \sim 33$ m/s. However, the relation between them is quite scattered. The linear function of C_d with respect to U_{10} can not represent their relation well when the wind speed ranges between 10 and 33 m/s. Bruss and Mayerle (2009) explained that the scatter of the C_d values is related to the spatial and temporal variation of the relevant factors such as fetch length, wind duration and wave age.

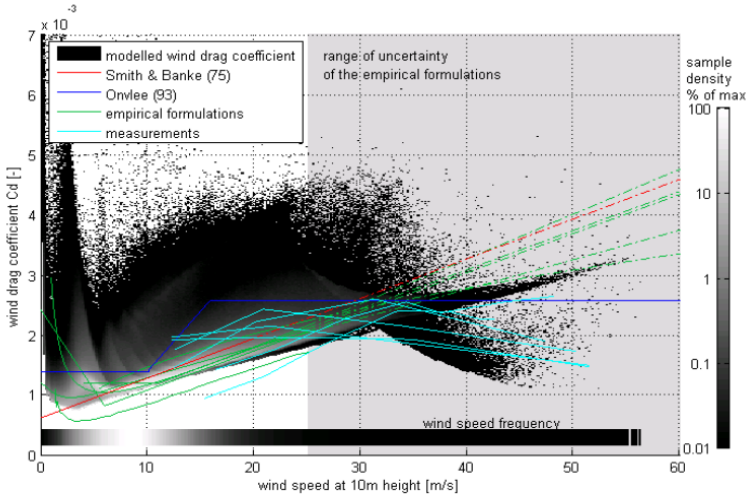

Figure 3.1: The scatter of U_{10} and C_d(Bruss and Mayerle, 2009)

In order to have a better understanding of the relation between C_d and its influencing factors, a more in-depth review of the relation between C_d and wind speed is presented below.

The wind profile is approximately a logarithm curve with respect to height in the atmospheric boundary layer and it can be expressed as,

$$U(z) = \frac{u_*}{\kappa}\ln(\frac{z}{z_0}) \tag{3.7}$$

where $U(z)$ is the wind speed at the height of z; z_0 is the aerodynamic surface roughness; $\kappa = 0.4$ is the von Karman coefficient; u_* is wind friction velocity, which is defined by:

$$u_* = \sqrt{\frac{\tau_s}{\rho_a}} \tag{3.8}$$

Combining equation (3.6) and equation (3.8), the relation between u_* and C_d can be obtained:

$$u_*^2 = C_d U_{10}^2 \tag{3.9}$$

Combining equation (3.7) and (3.9) yields the relation between z_0 and C_d,

$$z_0 = z_{10}\exp(-\kappa/C_d^{1/2}) \tag{3.10}$$

where $z_{10} = 10$m. This means C_d is determined by z_0. Specifying z_0 is equal to specifying C_d. The value of z_0 at a given location on land is considered to be fixed. But on the sea surface z_0 varies with sea surface state. Charnock (1955) proposed a non-dimensional relation between z_0 and u_*, i.e. $gz_0/u_*^2 = \alpha$, where g is the

gravitational acceleration and α is the Charnock coefficient. Charnock (1955) took it as a constant. Combining all the relations above, the relation between C_d and U_{10} are obtained as below,

$$\alpha^{1/2}U_{10}/(gz_{10})^{1/2} = C_d^{-1/2}\exp(-\frac{\kappa}{2}C_d^{-1/2}) \tag{3.11}$$

When C_d is in the range of $(1.0 - 4.0) \times 10^{-3}$, the equation (3.11) is almost a linear function expressed in equation (3.12)

$$C_d = (a + bU_{10}) \tag{3.12}$$

where $b = 0.475\alpha^{1/2}$. There are various values of α; $\alpha = 0.012$(Charnock, 1955); $\alpha = 0.035$(Kitaigorodskii and Volkov, 1965); $\alpha = 0.0144$(Garratt, 1977). Stewart (1974) proposed that α is not a constant but dependent on the wave age C_p/U_λ, where C_p is the peak phase speed from wave spectrum, and U_λ is the wind speed at the height of λ. In practice, U_{10} is often used instead of u_*. Using the wind stress and wave data at a platform in the North Sea during the Humidity Exchange over the Sea (HEXOS) program Smith et al. (1992) found that α is inversely proportional to the wave age.

Donelan et al. (1993) combined the HEXOS and his own data and obtained a formula,

$$z_0/\sigma = 6.7 \times 10^{-4}(U_{10}/C_p)^{2.6} \tag{3.13}$$

where σ is the locally wind-generated wave height. A simplified form when the wind and wave are towards the same direction is $z_0 = 3.7 \times 10^{-5}(U^2/g)(U/C_p)^{0.9}$. This formula indicates that z_0 is dependent on the wind speed, wave height and wave age. Sea surface with young waves has larger C_d than that with mature waves. However, this formula cannot apply to the sea with swell.

It is well known that C_d is not only a function of wind speed, but also of the wave state and other factors. Therefore, C_d is changing spatially and temporally in a very complex way. However, in practice, most ocean models still use wind speed dependent wind drag coefficient, which is the case in the model Delft3D. Delft3D applies a piecewise function to represent variants of C_d with respect to wind speed

U_{10} as follows (also see Figure 3.2),

$$C_d(U_{10}) = \begin{cases} C_d^A & U_{10} \leqslant U_{10}^A \\ C_d^A + (C_d^B + C_d^A)\dfrac{U_{10} - U_{10}^A}{U_{10}^B - U_{10}^A} & U_{10}^A \leqslant U_{10} \leqslant U_{10}^B \\ C_d^B + (C_d^C + C_d^B)\dfrac{U_{10} - U_{10}^B}{U_{10}^C - U_{10}^B} & U_{10}^B \leqslant U_{10} \leqslant U_{10}^C \\ C_d^C & U_{10}^C \leqslant U_{10} \end{cases} \tag{3.14}$$

Where:

C_d^i : are the user-defined wind drag coefficients at respectively the wind speed $U_{10}^i (i = A, B, C)$.

U_{10}^i : are user-defined wind speeds$(i = A, B, C)$.

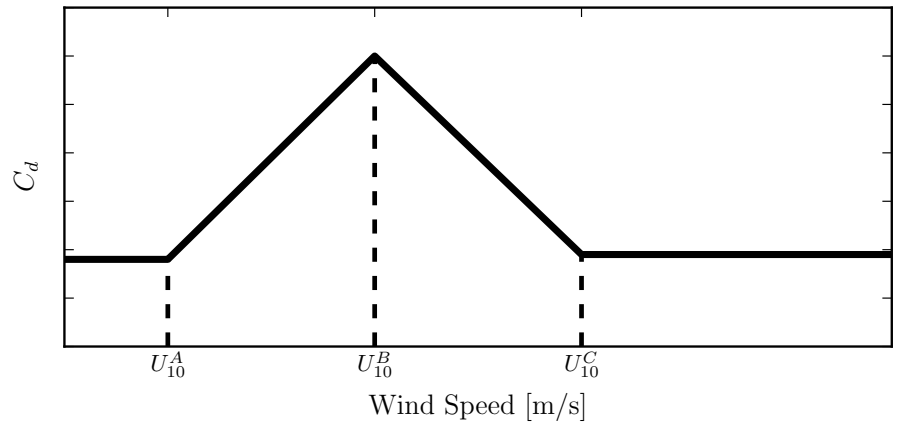

Figure 3.2: The piecewise function of wind drag coefficient to wind speed

3.1.3 Sea bed boundary condition

At the seabed, the boundary conditions for the momentum equations are:

$$\frac{\nu_v}{H}\frac{\partial u}{\partial \sigma}\bigg|_{\sigma=-1} = \frac{1}{\rho_0}\tau_{b\xi} \tag{3.15}$$

$$\frac{\nu_v}{H}\frac{\partial v}{\partial \sigma}\bigg|_{\sigma=-1} = \frac{1}{\rho_0}\tau_{b\eta} \tag{3.16}$$

where $\tau_{b\xi}$ and $\tau_{b\eta}$ are the components of the bed stress in ξ and η direction, respectively. The bed stress may be the combined effect of flow and waves. In this section only flow effect is introduced. For 2D depth-averaged flow the shear-stress

at the bed induced by a turbulent flow is assumed to be given by a quadratic friction law:

$$\vec{\tau}_b = \frac{\rho_0 g \vec{U}|\vec{U}|}{C_{2D}^2} \tag{3.17}$$

where $|\vec{U}|$ is the magnitude of the depth-averaged horizontal velocity. The 2D-Chezy coefficient C_{2D} can be determined according to one of the following two formulations:

- Chezy formulaton:

$$C_{2D} = \text{Chezy coefficient}[\text{m}^{1/2}/\text{s}] \tag{3.18}$$

- Manning's formulation:

$$C_{2D} = \frac{\sqrt[6]{H}}{n} \tag{3.19}$$

where:

H is the total water depth[m].

n is the Manning coefficient [$\text{m}^{-1/3}\text{s}$]

Manning's coefficient n is related to the alluvial bed properties. A summary of Manning coefficient is shown in table 3.3.

Table 3.3: Manning coefficient n ($\text{m}^{-1/3}\text{s}$)

Channel Material	n	Channel Material	n
Neat cement, smooth metal	0.010	Natural channels in good condition	0.025
Rubble masonry	0.017	Natural channels with stones and weeds	0.035
Smooth earth	0.018	Very poor natural channels	0.060

3.2 Domain and data of the storm surge model

3.2.1 Model domain

The study area is the German Bight, which is in the southeast of the North Sea. The German Bight model (GBM) covers the whole German Bight for the high resolution storm surge simulation. In order to have a better water level on the open boundary of the GBM exerted by the storm, the GBM is nested to a larger model, which covers the most Northwest European continental shelf (CSM) (Verboom et al.,

1992). The coverage of the two model domains are illustrated in Figure 3.3. More information about the grids of the CSM and GBM are in table 3.4. As discussed above, open boundary, wind and bottom roughness are the three most important parameters for storm surge simulations. The more accurately they are added to the model, the better results the model are expected to reach.

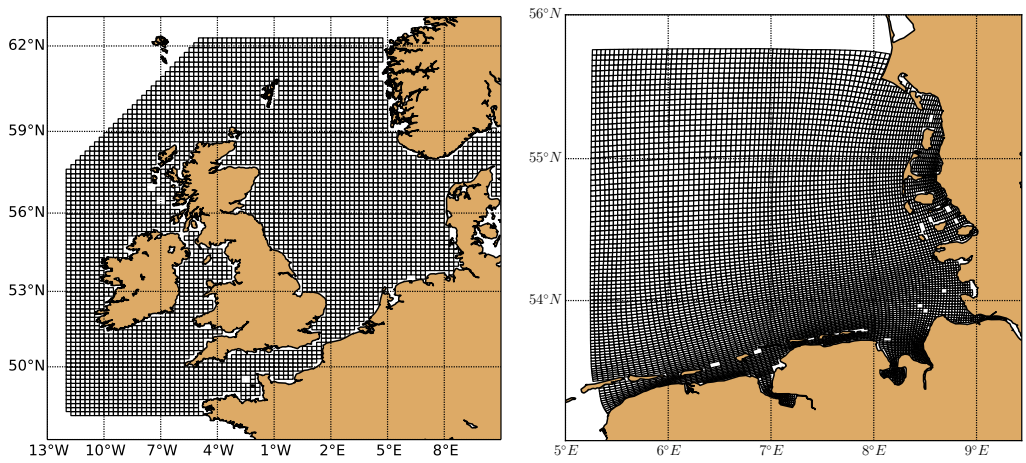

Figure 3.3: Model grids. left: the CSM grid; right: the GBM grid

Table 3.4: Model settings for the storm surge model

Parameter	CSM	GBM
Number of grid point ξ-direction	201	227
Number of grid point η-direction	173	266
Spatial resolution	ca. 9km	$0.5 \sim 1.7$ km

3.2.2 TPXO tidal harmonic constants

The open boundary of the CSM are forced by water level, which are represented by the astronomical tide without consideration of wind and density effects. This is because 1. it is almost impossible to obtain sufficient data along the open boundary of a big model; 2. in most areas, tide accounts for the largest fraction of water level change. Therefore this is a reasonable solution to the data insufficiency on the open boundary. 14 tidal harmonic constituents (M2, S2, N2, K2, K1, O1, Q1, P1, MF, MM, M4, MS4, MN4) are interpolated to the open boundary of the CSM. These harmonic constants are from a global model of ocean tides, which best-fits, in a least-squares sense, the Laplace Tidal Equations and along track averaged data

from TOPEX/Poseidon and Jason (on TOPEX/POSEIDON tracks since 2002). The spatial resolution in Northwest Europe continental shelf is 1/30°. The methods used to compute the tide model are described in detail by Egbert and Erofeeva (2002). The open boundary conditions of the GBM are provided by the simulated water level of the CSM . Therefore, it contains the effects of both tide and wind.

3.2.3 Wind data

In this model study the wind and air pressure data are from a meteorological model Consortium for Small Scale Modeling (COSMO-EU) operated by the German Weather Service Center (DWD) (COSMO, 2007). The spatial resolution is 0.0625° and the time interval is one hour. Besides this hourly data (DWD1), another two meteorological data with time interval of three hours are available and they will be used in section 6.4.1; one is also from DWD (DWD3) and the third is from European Centre for Medium-Range Weather Forecasts (ECMWF). The three sets of data are assessed on the basis of measurements at several stations along the German North Sea coast. Figure 3.4 shows comparisons of modeled and observed air pressure and wind speeds at the station Sylt (see Figure 1.1) from the 1st to the 10th of December 2013. For DWD1, the root mean square deviation of wind speed is about 1.62 m/s and the correlation coefficient is 0.90. The comparisons indicate that the modeled data can capture the real wind field quite well thus being able to provide adequate meteorological forcing to the storm surge model.

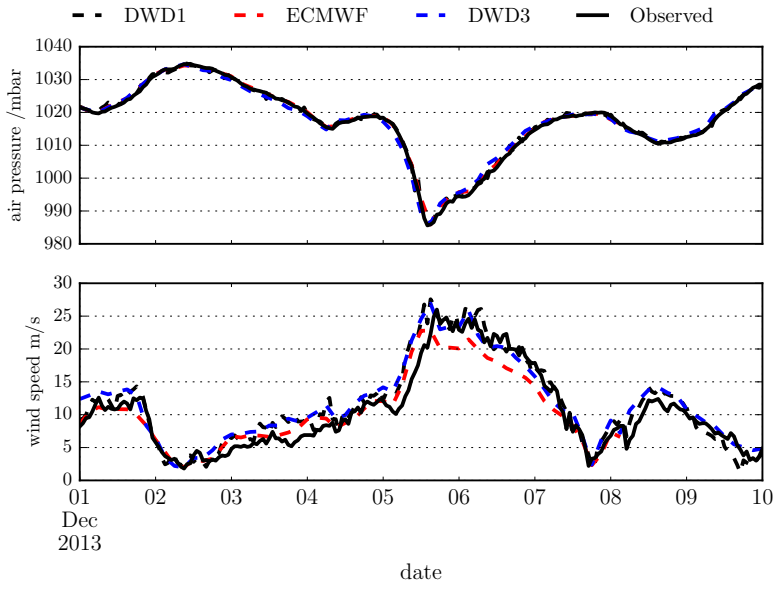

Figure 3.4: Comparison of DWD modeled data and observed data at the station Sylt (see Figure 1.1). The solid line is the observed data and the dashed lines are the modeled data

3.2.4 Bathymetry

The bathymetry data of the North Sea is taken from the model provided by Verboom et al. (1992). The bathymetry near the German coast has been updated by a series of measurements by the Bundesamt für Seeschifffahrt und Hydrographie (BSH), Wasser-und-Schiff fahrtsverwaltung des Bundes (WSV), Kuratorium für Forschung im Küsteningenieurwesen (KFKI) and Amt für ländliche Räume (ALR) Husum. The bathymetrical information covers the German coast for the period 1969-2009. (Fernández Jaramillo, 2014)

3.2.5 Tidal gauges measured data

Water level data measured by tidal gauges along the coast of the German Bight are used in this thesis for the model calibration and data assimilation (Figure 1.1). Quality control is carried out before using the data:

- Remove the invalid data or find the gap of missing data.
- If the gap is less than one hour, linear interpolation was used to fill the gap.
- If the gap is larger than one hour, this section of data will not be used.

3.3 Calibration of bottom roughness coefficient for tide simulation

Tide is the main component of water level variation in most coastal waters and also plays an important role in the storm surge due to the tide-surge interaction. Therefore, tide simulation should be validated before the deployment of the storm surge model. The tide model is forced by 14 tide harmonics on the open boundary of the CSM . Both Chezy and Manning coefficient are used for the calculation of bottom shear stress. The values of Chezy coefficient are 65 and 85 $m^{1/2}/s$, which represented the most rough and most smooth sea bed used by previous researchers (Horsburgh and De Vries, 2011). The Manning coefficients are 0.02 (smooth) and 0.03 $m^{-1/3}s$ (rough), which are also the lower limit and upper limit used in the previous storm surge models. The simulated water level at the locations of the selected tidal gauges are stored every 15 minutes from December 1, 2011 to January 1, 2012. Then harmonic analysis are performed on these modeled and observed water level.

The comparisons between them are shown in table 3.5 and 3.6 at Helgoland and Buesum respectively. More comparisons are shown in Appendix A. Only three dominating tide harmonics are shown in these tables. One obvious conclusion is that the amplitudes of harmonics are decreased when the bottom is rougher due to the more energy diffusion by the bottom friction. The phase is also well related to the bottom roughness; the phase lag is smaller in the case of smoother bottom; but the tide phase is less sensitive to the bottom roughness than amplitude.

Table 3.5: Comparison of harmonic constants at Helgoland

	M2		S2		O1	
	amp(m)	pha(°)	amp(m)	pha(°)	amp(m)	pha(°)
tidal_gauge	1.11	310	0.33	36	0.07	231
chezy_65 [a]	1.03	304	0.27	23	0.04	241
chezy_69	1.14	302	0.31	20	0.04	239
chezy_85	1.46	296	0.41	14	0.04	233
manning2 [b]	1.53	293	0.43	10	0.04	227
manning25	1.16	300	0.32	18	0.04	236
manning3	0.93	305	0.25	23	0.04	241

Table 3.6: Comparison of harmonic constants at Buesum

	M2		S2		O1	
	amp(m)	pha(°)	amp(m)	pha(°)	amp(m)	pha(°)
tidal_gauge	1.57	331	0.49	61	0.08	245
chezy_65	1.40	344	0.37	71	0.05	258
chezy_69	1.55	341	0.42	67	0.05	256
chezy_85	2.00	332	0.56	58	0.05	250
manning2	1.96	335	0.55	61	0.05	250
manning25	1.44	346	0.39	73	0.05	259
manning3	1.13	354	0.29	82	0.05	265

The sensitivity tests provide some clues to calibrate the model parameters. The harmonic M2 is the largest at all the seven stations. In models with Chezy coefficient, the amplitudes of M2 from measured data are between the amplitudes calculated from the two models Chezy_65 and Chezy_85. Therefore one value of

[a] Model with Chezy coefficient of 65 $m^{1/2}/s$

[b] Model with Manning coefficient of 0.02 $m^{-1/3}s$

Chezy coefficient between 65 and 85 $m^{1/2}/s$ can be used for a new simulation and is expected to improve model results. It is assumed that the amplitude of M2 is linearly related with the Chezy coefficient (in both ξ and η direction and both for the CSM and GBM) :

$$\mathrm{Amp_{M2}} = eC_{2D} + f \tag{3.20}$$

Where e and f are unknown constants, which are determined by the Chezy coefficient and the corresponding amplitude of M2 at each station. Then a new Chezy coefficient is obtained by

$$C_{2D} = \frac{\mathrm{Amp_{M2}^{tidalgauge}} - f}{e} \tag{3.21}$$

The new Chezy coefficient C_{2D} can be obtained from each station and the average of Chezy coefficients is 69.88 $m^{1/2}/s$. A similar method is used to calibrate the Manning coefficient n and a calibrated $n = 0.02525$ $m^{-1/3}s$ is obtained.

The harmonic constants from the model with Chezy coefficient 69.88 $m^{1/2}/s$ are also shown in tables from 3.5 to A.5. The amplitudes of M2 at the seven stations are much closer to the observations than before calibrating. However, the phase of M2 and other harmonics are not improved and even worse. But It can be expected the improvement of the water level simulation, because M2 has the largest energy in the German Bight. The improvements are illustrated in table 3.7, in which the RMSD of the models with calibrated bottom roughness are reduced compared with original model at most stations. The mean RMSD shows the model with Chezy coefficient 69.88 $m^{1/2}/s$ is the best among all the sensitivity tests. This value will be used to simulate the storm surge and as the first guess in the data assimilation.

Table 3.7: Root mean square deviation of model with different bottom roughness

	chezy_65	chezy_69	chezy_85	manning2	manning25	manning3
Helgoland	0.152	0.161	0.374	0.459	0.194	0.196
Buesum	0.346	0.267	0.368	0.374	0.374	0.581
Husum	0.263	0.208	0.440	0.414	0.352	0.546
Wittduen	0.142	0.122	0.358	0.378	0.136	0.239
Esbjerg	0.252	0.268	0.369	0.383	0.260	0.242
Havneby	0.218	0.241	0.400	0.399	0.234	0.232
Bremerhaven	0.330	0.238	0.453	0.373	0.362	0.616
mean_RMSE	0.243	0.215	0.395	0.397	0.273	0.379

3.4 Preliminary simulation of the storm surge event in early December 2013

A storm surge event in early December 2013 is modeled as a study case. The storm named Xaver is regarded as the most serious storm in the past 60 years in the North Sea. The resulted surge was comparable to the storm surge in January 1953 in the North Sea. The United Kingdom, Germany, the Netherlands and Scandinavia Peninsula were most heavily affected by the storm. The significant wave height was reported to be six meter high in the German coast, which was the second highest on record since 1825.

The low pressure storm system started on December 4 to the south of Greenland. As it moved to the east, its intensity was increasing. On December 5 it passed the north of Scotland. From December 5 to December 6, Xaver cut across southern parts of Norway and Sweden, reaching its peak intensity over the Baltic Sea with a minimum pressure of 960 mb. The peak gusts are up to 44 m/s on December 5th/6th, which reaches the hurricane force. Figure 3.5 and Figure 3.6 show the distribution of air pressure and wind vector at 2013-12-05 08:00:00 UTC. The time series of air pressure and wind speed at station Sylt are illustrated in Figure 3.4.

The storm resulted in very strong surge on the coast of German Bight. In Helgoland, the maximum surge over the mean high water (MHW) was 2.3m; in Emden, the value was 3.52m; Bremerhaven, 3.16m; Cuxhaven, 3.11m; Buesum, 2.97m; Husum, 3.28m. According to the storm surge classification in the German Bight [1], the storm surge caused by Xaver is a severe storm surge (schwere Sturmflut) or very severe storm at some stations (sehr schwere Sturmflut).

[1] see http://www.bsh.de/de/Meeresdaten/Vorhersagen/Sturmfluten/Berichte/index.jsp

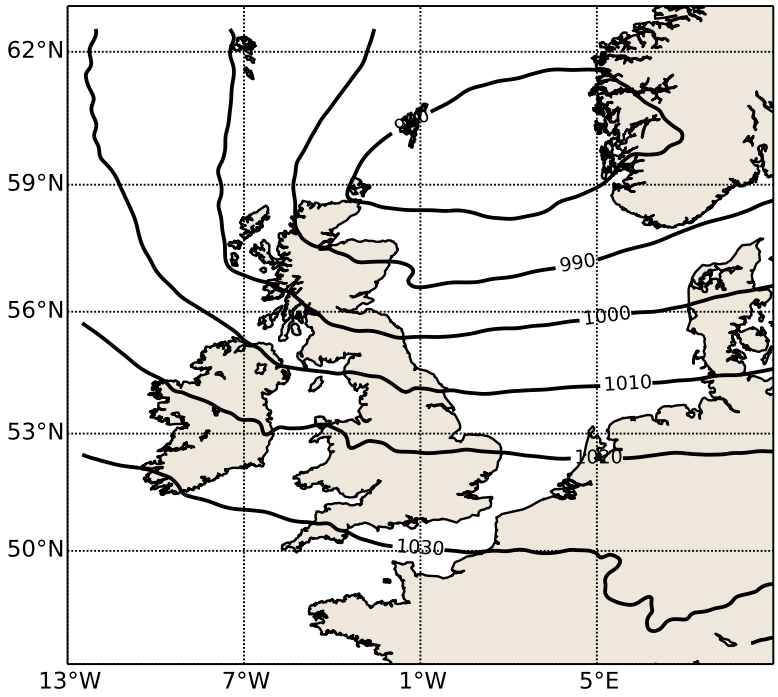

Figure 3.5: The distribution of air pressure at 2013-12-05 08:00:00 UTC

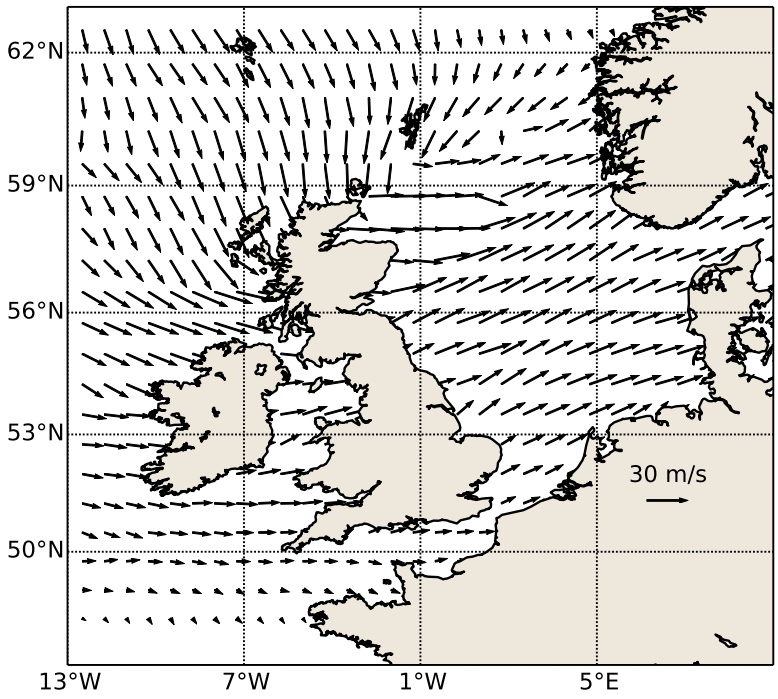

Figure 3.6: The wind field at 2013-12-05 08:00:00 UTC

The storm surge simulation starts at 00:00:00UTC on Dec.1 2013 and ends at

00:00:00UTC on Dec.10 2013. The bottom roughness is the Chezy coefficient 69.88 $m^{1/2}/s$ calibrated in section 3.3. The model is forced by astronomical tides on the open boundary, wind stress and air pressure over the sea surface. The wind stress is calculated by the formula proposed by Smith (1980) because the formula was derived from measurements in the German Bight. Four tidal gauges are chosen for detailed comparisons between the modeled and observed water level. They are Helgoland, Busuem, Husum and Cuxhaven (see Figure 1.1).

Table 3.8 shows that the RMSD of water level between model and tidal gauges in different stages of the storm surge. From 2013-12-03 00:00 UTC to 2013-12-05 12:00 UTC is the period before the storm; from 2013-12-05 12:00 UTC to 2013-12-08 12:00 UTC is the period in which there was high surge; from 2013-12-08 12:00 UTC to 2013-12-10 00:00 UTC is the period after storm. The RMSD in the period of storm is more than twice larger than that in normal weather conditions. This can be demonstrated more clearly in Figure 3.7. Generally speaking , the increased RMSDs during the storm period are due to the obviously increased peak water level, decreased low water level and phase lag. There are five tide cycles in the whole storm period(the time between two peaks is a tide period). The common features of the water level comparisons in the four stations are:

- The observed water level is higher than the modeled water level in the first two tide cycles. And there are obvious phase lag difference. The simulated water level peak is earlier than the observed.

- The model agrees much better to the observations between the third peak and third trough.

- The observed troughs in the last two cycles are lower than the simulated troughs. The simulated phase is more close to the observations than that in the first period.

Table 3.8: Root mean square deviation of water level (m) between model and tidal gauges in different stage of storm surge

	before storm	storm period	after storm
Helgoland	0.162	0.335	0.214
Buesum	0.285	0.497	0.258
Husum	0.198	0.491	0.208
Cuxhaven	0.143	0.487	0.215

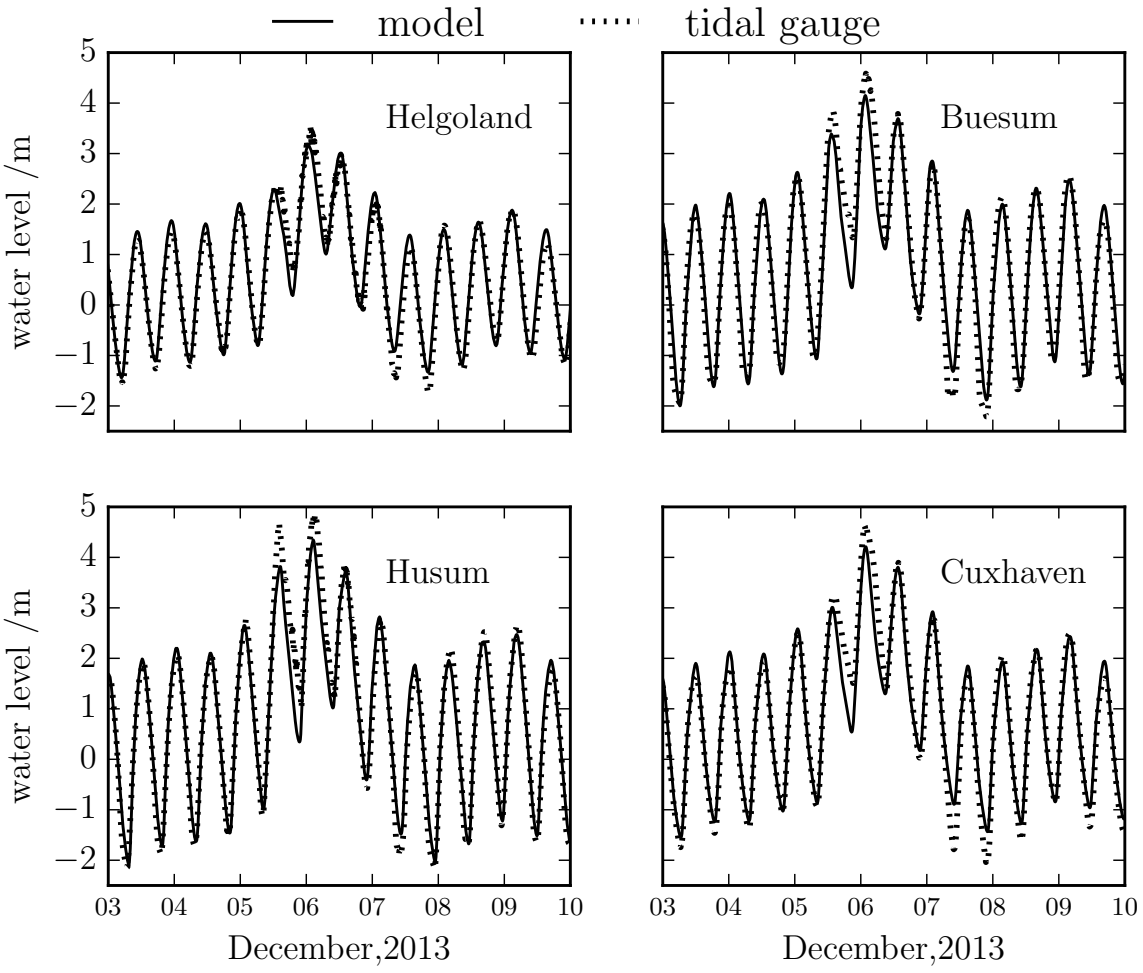

Figure 3.7: Comparison of modeled and observed water level in the German Bight

In order to know the reason why the RMSDs increase significantly in the period of storm, the surge, which is obtained by removing the tide from the total water level is investigated below. Then the observed surge and modeled surge are compared at the four tidal gauges. The comparisons of surge are shown in Figure 3.8. It can be seen that the modeled and observed surge have good correlation at all the four stations, especially around the largest two surge peaks. But the modeled surge peak is much smaller than the observed, which is the reason why the phase of the total water level occurs earlier than the observed value in Figure 3.7. One possible reason for that why the model underestimates the peak is that the sea water did not get enough energy from the wind stress. Recall that wind stress is calculated by wind speed and wind drag coefficient. The underestimate of wind stress is due to inaccuracy of both wind speed data and wind drag coefficient. Figure 3.4 indicates that the wind speed data has a high accuracy at the station Sylt. But the wind data are discretized in space and time, and there must be missing information due

to the insufficient resolution. The wind drag coefficient also has uncertainties, which are from measurement of wind speed and wind stress, representative errors and the valid range of the formula. The linear formula is derived from intermediate wind speed less than 21 m/s. While during the storm Xaver, the wind speed in German Bight can be up to 44 m/s. These uncertainties may lead to significant errors during strong storm period. A series of sensitivity tests will be carried out to investigate the error sources of storm surge simulations. Once the important error sources have been detected, data assimilation can be used to improve the model skill. Before the sensitivity tests, the adjoint free 4Dvar will be proposed in Chapter 4.

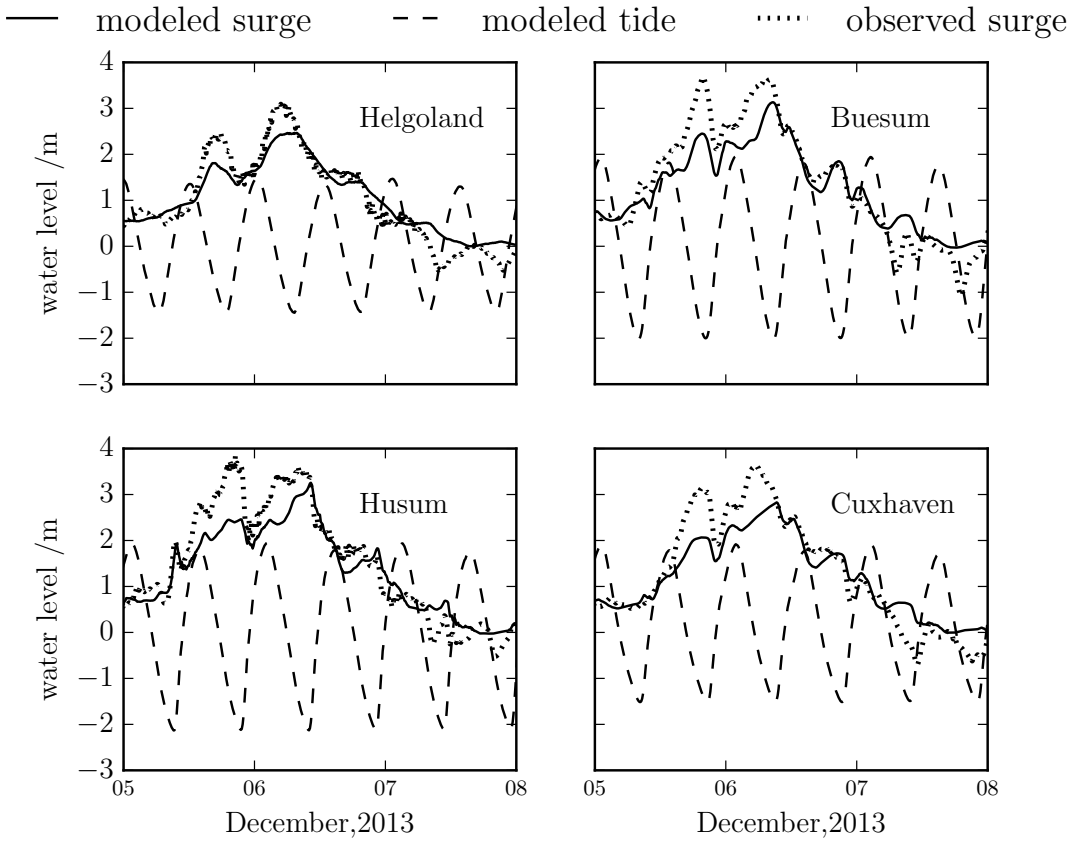

Figure 3.8: Comparison of modeled and observed surge

Chapter 4

Adjoint free 4Dvar

As mentioned in the Chapter 1, 4Dvar is currently the most effective data assimilation method. It has been implemented by ECWMF, UK met office and many other weather centers to improve the weather forecast. One advantage of 4Dvar is that it can keep the data assimilation results consistent with the model dynamics. This chapter will begin with a brief introduction of the variational calculus for the general minimization problems, which is the foundation of the method 4Dvar. Then the cost function of 4Dvar and the adjoint method for the minimization of the cost function are introduced. After that, a new method based on the incremental 4Dvar and ensemble method will be proposed. The proposed new method has comparable performance to adjoint method for low dimension parameters. The advantage is that it does not need adjoint equations and therefore can reduce the difficulty of implementation. The way to deal with data assimilation of high dimension parameters is also discussed in this chapter.

4.1 Variational calculus

Variational calculus is a mathematical method to find a function which can optimize its functional. Functional is the integration of one or several functions. A basic functional, in which the high order of the derivative is one, is denoted as $J[f(x)] = \int_{x_1}^{x_2} F(x, f(x), f'(x))\,\mathrm{d}x$, which contains the independent variable x, the function $f(x)$ and the first order derivative of $f'(x)$. The basic functional will be used to show how to get the optimal function $f(x)$. It should be noted that to distinguish the functional and function, [ ] is used instead of ( ).

The variational problem is to find an optimal $f(x)$ which makes the functional J minimum or maximum. Suppose $f(x)$ has fixed boundary conditions, that is, $f(x_1) = y_1, f(x_2) = y_2$.

δJ is the variation of J, which is derived by the variation of all functions in the functional.

$$\delta J = J[f(x) + \delta f(x)] - J[f(x)] = \int_{x_1}^{x_2} (F_f \delta f + F_{f'} \delta f') \, \mathrm{d}x \tag{4.1}$$

where F_f and $F_{f'}$ are the derivative of F relative to f and f'. In order to replace $\delta f'$ with δf, integrate by parts on term $F_{f'}\delta f'$ in equation (4.1) and obtain:

$$\delta J = \int_{x_1}^{x_2} (F_f \delta f - \frac{\partial F_{f'}}{\partial x} \delta f) \, \mathrm{d}x + (F_{f'} \delta f)|_{x_1}^{x_2} \tag{4.2}$$

$f(x)$ has fixed boundary conditions, thus $\delta f(x_1)$ and $\delta f(x_2)$ is zero. The second term on the right side of equation (4.2) can be eliminated.

$$\delta J = J[f(x) + \delta f(x)] - J[f(x)] = \int_{x_1}^{x_2} (F_f - \frac{\partial F_{f'}}{\partial x}) \delta f \, \mathrm{d}x \tag{4.3}$$

When $f(x)$ is optimized, δJ is equal to zero, that is, the right side in equation (4.3) is equal to zero. δf is arbitrary, so $F_f - \frac{\partial F_{f'}}{\partial x}$ must be equal to zero.

$$\begin{cases} F_f - \frac{\partial F_{f'}}{\partial x} = 0 \\ f(x1) = y_1 \\ f(x2) = y_2 \end{cases} \tag{4.4}$$

Now the problem of minimizing the functional J is converted to solving a partial differential equation as shown in equation (4.4). More extensive descriptions of variational calculus can be found in Troutman (2012).

4.2 Cost function of 4Dvar

Numerical models are common tools to study oceans. Models are derived by discretizing the primitive partial differential equations on space, time or frequency. Despite of the different degree of complexity of ocean models, all of them can be denoted by a general form as follows:

$$\begin{cases} x_{i+1} = \mathcal{M}_i(x_i, p) \\ x_0 = x_{0b} \\ p = p_b \end{cases} \tag{4.5}$$

where $x_i \in R^n$ is the n-dimensional model state vector at time t_i. i is the time index ranging from 0 to N_T, where N_T is the number of time step. $p \in R^k$ is the model parameter vector, consisting of initial conditions, boundary conditions and all other model parameters. $\mathcal{M}_i$ is the model forward operator, converting model state at t_i to model state at t_{i+1}. If the model and all the input parameters have no error, equation (4.5) is a well posed problem. Given these parameters (x_{0b} and p_b, the subscript b means background value or first guess), a unique model result can be obtained. This is the way that models are used in most situations.

Models are approximations of the real world, therefore they always have errors. The errors are from imperfect model dynamics, insufficient model resolution, parametrization schemes or inaccurate model parameters. Therefore, in order to improve model skills, observations are needed to calibrate and validate models. On the other hand, an automatic improvement of the model skills can be made using variational method. This is the so-called data assimilation in the meteorological and oceanographic communities. By adding the observations into the model (4.5), a new model system is obtained:

$$\begin{cases} x_{i+1} = \mathcal{M}_i(x_i, p) \\ x_0 = x_{0b} \\ p = p_b \\ y_i = Hx_i \end{cases} \tag{4.6}$$

where $y_i \in R^m$ is the observations vector at time t_i. In most cases, m is much smaller than n due to the difficulty of obtaining observations, which is often the case in the ocean study; $H \in R^{m \times n}$ is the observation operator, which converts the model state space to observations space. y_i can be direct observations of model state, such as water level from tidal gauges for an ocean model, or indirect observations, such as radar reflectivity for a meteorological model. Addition of observations makes the system to be an over-determined system which has no solution. As mentioned above, there are uncertainties in the model, model parameters and observations. Therefore the over-determined system can be converted to an under-determined system by adding error terms to equation (4.6):

$$\begin{cases} x_{i+1} = \mathcal{M}_i(x_i, p) + \epsilon_{x_i} \\ x_0 = x_{0b} + \epsilon_{x_0} \\ p = p_b + \epsilon_p \\ y_i = Hx_i + \epsilon_y \end{cases} \tag{4.7}$$

where $\epsilon_{x_i}, \epsilon_p, \epsilon_{x_0}$ and ϵ_y are errors of model, parameter p, initial condition and observations respectively. The poorly known parameters x_i, p and x_0 can be adjusted to decrease the summation of the square of errors in equation (4.7). A cost function (functional) J is introduced to represent the summation of errors square:

$$\begin{aligned} J[x_0, x_i, p] =& \frac{1}{2}\sum_{i=1}^{N_T}(x_{i+1} - \mathcal{M}_i(x_i, p))^T W_i(x_{i+1} - \mathcal{M}_i(x_i, p)) + \\ & \frac{1}{2}(x_0 - x_{0b})^T W_0(x_0 - x_{0b}) + \frac{1}{2}(p - p_b)^T W_p(p - p_b) + \\ & \frac{1}{2}\sum_{i=1}^{N_T}(y_i - Hx_i)^T Q(y_i - Hx_i) \end{aligned} \tag{4.8}$$

where W_i, W_0, W_p and Q are the weight matrix for the terms of model, initial condition, model parameters and observations. They represent the accuracy of the model, model parameters and observations. Terms in equation (4.8) with larger weight have higher accuracy. The poorly known x_0, x_i and p which are to be adjusted are called control variables. By minimizing J, the optimal control variables can be solved.

In more situations, the model in equation (4.5) is assumed to be perfect. That is, the error term ϵ_{x_i} in equation (4.7) is 0. This is the so-called strong constraint 4Dvar. The cost function of strong constraint 4Dvar is simplified from equation (4.8):

$$\begin{aligned} J[x_i, p] =& \frac{1}{2}(x_0 - x_{0b})^T W_0(x_0 - x_{0b}) + \frac{1}{2}(p - p_b)^T W_p(p - p_b) \\ & + \frac{1}{2}\sum_{i=1}^{N_T}(y_i - Hx_i)^T Q(y_i - Hx_i) \end{aligned} \tag{4.9}$$

The adjoint free 4Dvar method is proposed in the framework of strong constraint 4Dvar.

4.3 Gradient method to minimize the cost function J

The cost function J in equation (4.9) is quadratic if the model operator and observation operator are linear, or it is quasi quadratic if the model operator and observation operator are weak non-linear. This can be satisfied for most ocean models. The minimum of J is located where the gradient of J with respect to the control variables is zero. To find this point, a direction can be chosen, along which the minimum point can be found most efficiently. There are mature optimization algorithms

to find the global minimum. Among them, Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm is an effective and efficient method when the dimension of control variables is very large (Nocedal and Wright, 2006). The crucial step is to find the gradient of cost function with respect to the control variables. Adjoint method is the common method to get the gradient for meteorological and ocean models.

4.4 Adjoint method

For convenience, further simplification is made on the equation (4.9). It is assumed that observations are much more accurate than first guess of parameters and all observations have the same accuracy. As a result, only the observation term of equation (4.9) is kept, and the simplified cost function is:

$$J[p, x_i] = \frac{1}{2}\sum_{i=1}^{N_T}(H_i x_i - y_i)^T(H_i x_i - y_i) \tag{4.10}$$

x_i between $i = 1$ and N_T are not independent but subject to the model in equation (4.5). Hence, 4Dvar is a constrained optimization problem. Compared with 3Dvar, 4Dvar uses observations over a certain period; and in addition, model dynamics is also considered in the process of data assimilation. Therefore, results of 4Dvar are more consistent with model dynamics than 3Dvar or optimal interpolation. The constrained optimization problem can be converted to an unconstrained optimization problem by the Lagrangian multiplier method as follows:

$$\begin{aligned} L[p, \lambda_i, x_i] =& \frac{1}{2}\sum_{i=1}^{N_T}(H_i x_i - y_i)^T(H_i x_i - y_i) + \\ & \sum_{i=1}^{N_T}\lambda_{i-1}^T(x_i - \mathcal{M}_{i-1}(x_{i-1}, p)) \end{aligned} \tag{4.11}$$

where $\lambda_i \in R^n$ is the Lagrangian multiplier at time t_i. The variation of equation (4.11) is

$$\begin{aligned} \delta L =& \sum_{i=1}^{N_T}(H_i x_i - y_i)^T H_i \delta x_i + \\ & \sum_{i=1}^{N_T}(x_i - \mathcal{M}_i(x_{i-1}, p))^T \delta\lambda_{i-1} + \\ & \sum_{i=1}^{N_T}(\lambda_{i-1}^T \delta x_i - \lambda_{i-1}^T \frac{\partial \mathcal{M}_{i-1}}{\partial x_{i-1}}\delta x_{i-1} - \lambda_{i-1}^T \frac{\partial \mathcal{M}_{i-1}}{\partial p}\delta p) \end{aligned} \tag{4.12}$$

In order to get the gradient of L with respect to p, the derivative of L with respect to x_i is set to 0, giving:

$$\lambda_{i-1} - (\frac{\partial \mathcal{M}_i}{\partial x_i})^T \lambda_i = -H_i^T (H_i x_i - y_i) \tag{4.13}$$

Equation (4.13) is the so-called adjoint model. λ_i is the adjoint variable. Given the first guess of model parameter p, the ocean model (equation 4.5) integrates forward (from t_{i-1} to t_i) leading to model results. λ_{N_T} is then set to zero. Equation (4.13) is integrated backward (from t_i to t_{i-1}) forced by $H_i^T(H_i x_i - y_i)$, which are the deviations between observations and model results at t_i. All the λ_i can be solved. After that λ_i are substituted into equation below:

$$\frac{\partial L}{\partial p} = \sum (-\frac{\partial \mathcal{M}_i}{\partial p})^T \lambda_i \tag{4.14}$$

Note that the gradient $\nabla_p^J = \frac{\partial L}{\partial p}$. Once ∇_p^J is calculated, gradient based optimization methods such as the BFGS quasi-Newton minimization algorithm can be used to obtain the optimal p which minimizes equation (4.10) . This is usually an iterative process and it stops once the termination condition is satisfied.

4.5 Incremental 4Dvar

4Dvar methods are extremely time-consuming and expensive computationally in that it needs to run both the primitive model and adjoint equation for several times and adjoint equation is normally 1.5 times more time-consuming than the primitive model. 4Dvar was not implemented for operational weather forecast until the incremental 4Dvar was proposed. Incremental 4Dvar uses the increment of p, i.e. Δp, as the control variable instead of p. Assume that the optimal p is close to the first guess of $p = p_b + \Delta p$. x_i^b is the model result at time t_i given $p = p_b$ and Δx_i is the increment of x_i due to increment of p. Δx_{i+1} is represented by Δp and Δx_i:

$$\Delta x_{i+1} = \mathcal{M}_i(x_i^b + \Delta x_i, p_b + \Delta p) - \mathcal{M}_i(x_i^b, p_b) \tag{4.15}$$

Apply Taylor's formula to the first term of the right hand side of equation (4.15) and only keep the terms of first order:

$$\mathcal{M}_i(x_i^b + \Delta x_i, p_b + \Delta p) = \mathcal{M}_i(x_i^b, p_b) + \frac{\partial \mathcal{M}_i}{\partial x_i}\Delta x_i + \frac{\partial \mathcal{M}_i}{\partial p}\Delta p \tag{4.16}$$

Substituting equation (4.16) into equation (4.15), the tangent linear equation is obtained as below:

$$\Delta x_{i+1} = \frac{\partial \mathcal{M}_i}{\partial x_i}\Delta x_i + \frac{\partial \mathcal{M}_i}{\partial p}\Delta p \tag{4.17}$$

where $\frac{\partial \mathcal{M}_i}{\partial x_i}$ and $\frac{\partial \mathcal{M}_i}{\partial p}$ are linear operators. A new cost function with respect to Δx_i can be obtained by substituting $x_i = x_i^b + \Delta x_i$ into equation (4.11) as follows:

$$J_1[\Delta p, \Delta x_i] = \frac{1}{2}\sum_{i=1}^{N_T}(H_i\Delta x_i - d_i)^T(H_i\Delta x_i - d_i) \tag{4.18}$$

$d_i = y_i - H_i x_i^b$ is the discrepancy between observations and model results when $p = p_b$. Equation (4.18) is minimized subject to the equation (4.17), which is a linear model. Therefore, equation (4.18) is quadratic and has a unique global minimum at Δp^a. The process that solves Δp^a represents an inner loop. p is updated in an outer loop by $p_1 = p_b + \Delta p^a$ and then a new inner loop starts with the new first guess p_1. The two loops proceed until Δp^a is small enough.

The linear tangent equation (4.17) can be a simplified model or solved on a coarser model grid, which reduces the computation burden and enables to apply 4Dvar to meteorological or ocean models. Operational weather forecast centers such as ECMWF (Rabier et al., 2000), UK Met Office (Lorenc et al., 2000) and the ocean model ROMS (Moore et al., 2011) have developed their data assimilation system using incremental 4Dvar.

4.6 Ensemble Kalman filter

Kalman filter is a well known sequential data assimilation method which takes model dynamics into account. This section shows how Kalman filter is implemented to meteorological or ocean models, providing a hint for developing the adjoint free 4Dvar.

For a linear model:

$$x_{i+1} = \phi x_i + w_i \tag{4.19}$$

where x is the model state vector with dimension n ($x_i \in R^n$) and ϕ is the linear model operator which integrates model state at time t_i to model state at time t_{i+1}, Kalman filter gives the best least unbiased estimate (BLUE) of the model state. w_i is model error at time t_i. When the model is deterministic, w_i is 0, that is, given first guess of x_0, model state x_i at any time can be obtained. However, x_0 is not exactly known and is actually a random vector. It is assumed that x_0 is in normal distribution. The model operator ϕ is not perfect, therefore, model errors always

exist in the course of model integrations. In Kalman filter, it is assumed that w_i has the following features,

$$< w_i >= 0, < w_i w_i >= Q_i, < w_i w_{i+1} >= 0$$

where $<>$ is the mean sign of a random vector.

Assume at each time step t_i, observation y_i of x_i is available,

$$y_i = H_i x_i + v_i \tag{4.20}$$

where y_i is an m dimensional random vector. $H_i \in R^{m \times n}$ is the observation operator from model space to observation space. v_i is the observation error. v_i is assumed to have features as follows,

$$< v_i >= 0, < v_i v_i >= R_i, < v_i v_{i+1} >= 0 < w_i v_i >= 0$$

Kalman filter works starting from specifying a normal distribution of initial condition x_0, $\mathcal{N}(x_{0b}, P_0)$, i.e., the mean of x_0 is x_{0b} and the covariance is P_0. After that, Kalman filter integrates the distribution of x_0 at each time step (forecast step). When there is observation available, Kalman filter obtains a new model state vector (analysis step or data assimilation step) which is a linear combination of model state and observations at that time with least variance summation. The two steps are shown in equation (4.21) and (4.22).

Forecast step

$$\begin{cases} x_{i+1}^f = \phi x_i^a \\ P_{i+1}^f = \phi P_i^a \phi^T + Q_i \end{cases} \tag{4.21}$$

Analysis step

$$\begin{cases} x_i^a = x_i^f + K_i(y_i - H_i x_i^f) \\ P_i^a = (I - K_i H_i) P_i^f \\ K_i = P_i^f H_i^T (H_i P_i^f H_i + R_i)^{-1} \end{cases} \tag{4.22}$$

The superscript f means forecast, a means analysis and T means transposing in equation (4.21) and (4.22). $K_i \in R^{n \times m}$ is the Kalman gain; I is an identity matrix.

In practice, the model state dimension is often too large to integrate forward P_i in the forecast step, making it difficult to apply Kalman filter.

Evensen (1994) gave a new method to make the difficulty less challenging. He combined a Monte Carlo method with Kalman filter and proposed the ensemble Kalman filter (EnKF). Instead of using ϕ or linearized ϕ to advance the covariance

P_i directly, EnKF propagates P_i using the spread of a group of model results (ensemble) which are obtained by integrating a group of models (ensemble) with different initial conditions. Below only two important steps are listed to illustrate how EnKF advances model state error mean and covariance. For more details of EnKF, readers are referred to Evensen (2009).

Creation of ensemble of initial condition

N ensemble members of initial condition are generated from the distribution $\mathcal{N}(x_{0b}, P_0)$. Denote x_0^j as initial condition of the the jth model.

Ensemble forecast step

N models are run with initial conditions ranging from x_0^1 to x_0^N. A matrix $X_i \in R^{n\times N}$ is constructed, whose column vectors are the model state of all N models. The ensemble mean matrix of X_i is defined as $\bar{X}_i = X_i\bar{1}_M$, where $\bar{1}_M \in R^{N\times N}$ and all elements are $1/N$. Model state perturbation matrix X_i' is defined as $X_i' = X_i - \bar{X}_i$. The mean and covariance of model state at equation (4.21) is represented approximately as:

$$\begin{cases} x_i^f \approx \bar{X}_i[:,1] \\ P_i^f \approx \frac{X_i' X_i'^T}{N-1} \end{cases} \tag{4.23}$$

where $\bar{X}_i[:,1]$ is the first column of the matrix $\bar{X}_i$.

4.7 Adjoint free (AF) 4Dvar

Adjoint methods are very powerful tools for parameter estimate. The gradients of cost functions with respect to the model parameters can be calculated by integrating both forward and adjoint models regardless of the dimension of the model parameters. As the coding job of adjoint models is almost the same as the forward models, the development and maintenance of adjoint models for complex ocean models is very demanding. Some novel methods are needed to make implementation of 4Dvar easier. As shown in Section 4.6,ensemble method makes it possible to implement Kalman filter to meteorological and ocean models. Inspired by the EnKF, a method based on the incremental 4Dvar but without the adjoint equations is proposed.

In equation (4.17), Δx_{i+1} is calculated sequentially in a numerical model given Δx_i and Δp. Δx_{i+1} can also be represented by an implicit function with respect to

Δp, as follows:

$$\Delta x_i = M_i(\Delta p) \tag{4.24}$$

Substituting equation (4.24) into equation (4.18), a cost function with respect to Δp is obtained,

$$J_2[\Delta p] = \frac{1}{2}\sum_{i=1}^{N_T}(H_i M_i(\Delta p) - d_i)^T(H_i M_i(\Delta p) - d_i) \tag{4.25}$$

To calculate the gradient $\nabla_{\Delta p}^{J_2}$, the essential step is to represent M_i explicitly. M_i represents the linear tangent model, which converts perturbations of p to the perturbations of model results at time t_i. Therefore, M_i is a matrix and the parentheses in equation (4.25) can be neglected. The idea is to add perturbations to the control variables just like adding perturbations to the initial condition in EnKF. Then the ensemble models are run and the perturbations of model results are obtained.

Assume that the first guess of p is p_b and the corresponding model results are x_b. Then N perturbations of p, stored in a matrix $P \in R^{k\times N}$ are generated. Adding p_b to each column in P results in N models with different parameters p. Running these models leads to an ensemble of model results. The ensemble is denoted as $E_i \in R^{n\times N}$ at time t_i. Recall that n is the dimension of the model state. Subtracting x_i^b from each row of E_i, an ensemble perturbation of model state denoted as $X_i \in R^{n\times N}$ is obtained. Both X_i and P are small perturbations, therefore,

$$X_i \approx M_i P \tag{4.26}$$

Equation (4.26) is a problem of multivariate linear regression. Columns of P represent the predictor variable, and columns of X_i represent the response variable. The number of sample is N. The regression coefficient M_i is the estimate of tangent linear model. Once M_i is estimated explicitly, the cost function (4.25) can be represented explicitly. It is an exact quadratic equation. Therefore the unique optimal Δp can be solved by BFGS. This is the inner loop like in the incremental 4Dvar. The outer loop is the same as in the incremental 4Dvar.

When the dimension of parameters k is a small number compared with the number of sample N, the ordinary linear regression can be used to obtain the regression coefficient M_i. However, the problem "Curse of Dimensionality" arises when k is larger than the number of samples or there is strong correlations in the predictors. In this case, M_i can not be determined. High dimension has been a problem for data assimilation in the meteorological and ocean modeling communities. Special methods have been applied for the practical implementation of Kalman filter or

4Dvar on large dimensional models. As mentioned in Section 4.5, in the incremental 4Dvar, a low resolution linear tangent model and its adjoint model are used to get the parameter increment and then the increments are interpolated back to the original model grid; The same strategy of reduced resolution was used by Hoteit et al. (2005) on an extended Kalman filter; Cohn and Todling (1995) assumed that most changes in the error covariance can be accounted for by a small collection of singular modes and only these modes are evolving in the model; Evensen (1994) used Monte Carlo method to generate a number of perturbations fitting to the error covariance and get the analysis of model state on the subspace spanned by these perturbations. The assumption in the above-mentioned studies is based on the fact that the errors are correlated and the errors in the full dimension can be represented by a much smaller number of components. Hence the dimension of parameters can be reduced without losing much information. Dimension reduction will also be used to relax the problem of "Curse of Dimensionality" in equation (4.26). Principal components analysis (PCA) is a common method for dimension reduction. In the following paragraphs, PCA will be introduced briefly and then how to apply PCA to adjoint free 4Dvar is explained in detail. For more introduction of PCA, please refer to Vidal et al. (2005) or other textbooks on multivariate linear regression.

PCA is a method to find a set of uncorrelated principal components in a multivariate random variable. In the matrix P in equation (4.26), each row vector is a set of perturbations of one component; each column represents a realization of a random vector. The first principal components w_1 is a unit vector and defined in the way that along this principal components, the variance of p is the largest.

$$w_1 = \text{argmax}((P^T w)^T P^T w) \tag{4.27}$$

where w is an arbitrary k dimensional unit vector. Therefore ,its norm is equal to 1.

$$\|w\| = w^T w = 1 \tag{4.28}$$

The solving of w_1 is a constrained optimization problem subject to equation (4.28). Using Lagrangian method, it can be solved that w_1 is an eigenvector of the matrix PP^T and the corresponding eigenvalue is the largest. The other principal components can be obtained in a similar method and they are the other eigenvectors of the matrix PP^T. According to the definition in equation (4.27), most variances are on the first principal components. They represent the true correlation between different dimensions. The variances in the last principal components are often very small and most of the time they are the noises when sampling. The removal of the

last principal components not only reduces dimension but also makes the system more stable. Once all the principal components are obtained, the original dataset P can be projected to the first r (r is smaller than k) principal components. By this means, the dimension of P is reduced.

In practice, due to the high dimension of parameters, singular value decomposition (SVD) is often used to find the principal components. Apply SVD on P (equation 4.26), obtaining:

$$P = USV^T \tag{4.29}$$

where $U \in R^{k\times k}$. Columns of U are the eigenvectors of matrix PP^T. $S \in R^{k\times N}$. The diagonal elements of S are the singular values of P. The first r columns of U are the directions along which the variances are largest. r can be determined by a ratio:

$$\mathrm{s}_{\mathrm{ratio}} = \frac{\sum_{i=1}^{r} s_i}{\sum_{i=1}^{N} s_i} \tag{4.30}$$

where s_i is the singular values sorted by the descending order. $\mathrm{s}_{\mathrm{ratio}}$ normally should be larger than 90%. Columns of U are also sorted by the the descending order of s_i. Only the first r eigenvectors in U are used. These vectors form a matrix $U_r \in R^{k\times r}$. A new parameter matrix $P_r \in R^{r\times N}$ with reduced dimension ($r \ll k$) can be obtained by:

$$P_r = U_r^T P \tag{4.31}$$

Now P_r is used instead of P. Substituting $P = U_r P_r$ into equation (4.26), obtaining:

$$X_i \approx M_i^r P_r \tag{4.32}$$

where $M_i^r = M_i U_r$. The dimension of P_r is smaller than the number of sample N. Therefore, ordinary linear regression can be used to solve M_i^r.

The cost function (4.25) is adjusted using a new control variable Δp_r, i.e., the increment of parameters with reduced dimension:

$$J_3[\Delta p_r] = \frac{1}{2}\sum_{i=1}^{N_T}(H_i M_i^r \Delta p_r - d_i)^T (H_i M_i^r \Delta p^r - d_i) \tag{4.33}$$

The optimum Δp_r, Δp_r^a can be solved by optimizing cost function (4.33). Then the optimum Δp is obtained by:

$$\Delta p^a = U_r \Delta p_r^a \tag{4.34}$$

The model state perturbation matrix X_i is the model response to the perturbations of parameters. Thus the proposed adjoint free (AF) 4Dvar is consistent with

model dynamic similar to 4Dvar and Kalman filter. The proposed method is comparable to EnKF in terms of the easy implementation. The model ensemble run in parallel without data exchange. Therefore, the whole process of data assimilation can speed up by simply using more cores. There are also differences between AF 4Dvar and EnKF, which are summarized as follows:

- One obvious difference is that EnKF assimilates observations at a time point while AF 4Dvar over a period. The observation frequency in coastal waters is often high. It is better to assimilate these observations over a period synchronously. Otherwise, models is interrupted for data assimilation and then restart frequently. Too frequent data assimilation may destroy smoothness of model results. In this sense, AF 4Dvar is preferable for coastal model data assimilation.

- EnKF is used to adjust model state to provide better initial conditions for the following simulation cycle. AF 4Dvar is not only able to adjust initial condition but also any model parameter. Coastal phenomena like tide and storm surge are boundary condition problems and effect of initial condition uncertainty is often neglected in practice. Therefore, AF 4Dvar and 4Dvar are more suitable for coastal models.

- Another difference is the way how perturbations are generated. Perturbations for EnKF represent how the error distribution of initial condition evolves with model dynamics. They must be generated randomly and follow a predefined Gaussian distribution. The number of perturbations for EnKF must be high enough for a proper representation of the error covariance. While for AF 4Dvar, perturbations of p are used to express explicitly the numerical tangent linear equations (Equation 4.17) in a matrix. Perturbations can be any relatively small values compared with the first guess. However, if perturbations are too small, the round-off errors become significant in X_i and would affect the accuracy of the estimation of M_i in Equation (4.26). Therefore, it is necessary to calibrate the perturbations before implementing the data assimilation.

The steps of implementing the proposed AF 4Dvar are listed below:

1. Prepare all the input files and reasonable model parameters for the model which is named model_0 here.

2. Select the model parameters to be considered as the control variables p_b in the data assimilation.

3. Run model_0.

4. Generate N sets of perturbations of control variables and add the perturbations to the p_b in the model_0. An ensemble with N sets of models is formed in this step.

5. Run the N models in the ensemble.

6. Estimate the tangent linear equation by perturbations of control variables and model results and get the cost function with respect to the increment of control variables.

7. Minimize the cost function and get the optimal increments for the first guess of control variables.

8. Update the control variables by adding the optimal increments to the first guess of control variables.

9. Run model_0 with updated parameters to obtain new model results.

10. Check whether the termination condition is satisfied. If true, exit from the outer loop; otherwise repeat the sequence from step 4.

The value $\text{abs}(J_{i+1} - J_i)/J_{i+1}$ is used as the termination condition. If $\text{abs}(J_{i+1} - J_i)/J_{i+1} < 0.01$ (i is the index of the outer loop), the outer loop will exit.

The programming language Python is used to implement the proposed adjoint free 4Dvar. Python is a free, multi-platform (Windows, Linux, Mac), object orientated and general purpose script language. Python is able to call Fortran and C codes in an easy way. And in addition, Python has a number of open source numerical and scientific packages. These features make Python a good tool to manage data assimilation.

One Python script and related functions control the whole data assimilation process indicated in the steps above, including generating model parameter perturbations, executing the model Delft3D, reading model output files and observation files, minimizing cost function and update model parameters. One Perl script is used to run model ensemble (in step 5) with a number of cores, which decreases the total data assimilation time significantly. Python packages “numpy”, “pandas” and “scipy” are heavily used for matrix operations, time series data process and cost function minimization (the function scipy.optimize.fmin_bfgs()).

4.8 Summary

In this chapter, the data assimilation method 4Dvar which uses adjoint equations has been reviewed. It can be seen that adjoint equations for an ocean model is difficult to develop and implement. One practical adjoint free 4Dvar is proposed based on the incremental 4Dvar. This method represents the tangent linear equations explicitly to avoid the adjoint equations using ensemble method. For model parameters of high dimension, PCA is used to reduce the dimension and overcome the under-determined problem in the calculation of the explicit tangent linear equation. In the next chapter, several numerical experiments will be performed to validate the adjoint free 4Dvar.

Chapter 5

Twin experiments

In the ocean model community, observations sufficiency is always a limitation due to the difficulty and high cost of installing instruments. Observation System Simulation Experiments (OSSE) is a possible solution to ease this problem. OSSEs have been conducted by various scientists for different purposes (Atlas, 1997; Masutani et al., 2010). The most common motivation for conducting OSSEs is to estimate the cost effectiveness of new observations. In addition, a data assimilation system can be tested by OSSEs because of the existence of a known "truth" in the context of OSSEs. The OSSE technique uses a model-generated proxy for the real sea state, commonly called the Nature Run. This generation is performed using a realistic ocean model in a free-running mode without data assimilation. Simulated observations are generated by virtual instruments from the Nature Run and used as real observations in data assimilation experiments. The generation of these simulated observations includes the addition of realistic errors. In these idealized experiments, the Nature Run provides the "truth" of the ocean, which is never obtained from the real observations. In OSSEs, all the data assimilation results can be validated against this "truth". The aim of this study is to assess whether the data assimilation method proposed in Chapter 4 has the ability to produce a better water level simulation using the OSSE technique. If the model and its configurations are identical for the Nature Run and the data assimilation system, the system is referred to as an identical twin OSSE (twin experiment).

A typical twin experiment for validation of data assimilation consists of:

The nature run is a reference model without data assimilation whose results are taken to be the truth.

Pseudo-observations are extracted from the results of the nature run plus the assumed errors.

The control run is a model without data assimilation with different model settings than the nature run.

Data assimilation combines the model and observations based on the estimate theory or optimization control.

Data assimilation assessment criteria The performance of data assimilation can be assessed by the decrease of cost function, the improvement of model skill at the stations which are not used for the data assimilation and whether the parameters are adjusted by the data assimilation in a reasonable way.

5.1 Twin experiment settings

A data assimilation method based on 4Dvar have been developed in Chapter 4. This method does not need the adjoint equations. Therefore, it is easier to implement this method than the adjoint method. In this chapter this method will be validated . In addition, it will be investigated how feasible this method can be implemented for operational use of ocean models. Twin experiments will be performed to investigate these problems. The Continental Shelf model (CSM) and German Bight model (GBM) described in Chapter 3 are used for the nature run and control runs in the twin experiments.

5.1.1 Nature run

The model settings of the nature run is the same as the preliminary storm surge model. For detailed description of this model, see the Section 3.4.

5.1.2 Pseudo-observations

Pseudo-observations are defined as the model results of nature run. Several stations along the coast of German Bight are employed in twin experiments. Some of them are used for data assimilation and some of them are used for validation of data assimilation. It should be noted that in this chapter, the observations are directly extracted from nature run without adding errors, because the purpose of the twin experiments is to see how well data assimilation can pull the control runs to the nature run. “Perfect” observations facilitates the assessments of data assimilation.

5.1.3 Scenarios of control runs

A number of scenarios of control runs are designed in the twin experiments, which are listed in Table 5.1. These scenarios are classified into three basic scenarios. The basic scenarios are denoted by the integer part of the scenario number in Table 5.1.

1. In the first basic scenario, the error sources of models are exactly known; the control variables (model parameters to be adjusted) of data assimilation contain all the imperfect model parameters[1]. The first basic scenario tests the effectiveness of the proposed adjoint free 4Dvar method. Only after making sure this method is correct can it be implemented into practice.

2. The second basic scenario represents more realistic situations. Only parts of the imperfect model parameters are in the control variables. The second basic scenario helps understanding problems in a realistic use of data assimilation.

3. The third basic scenario is designed to investigate the effect of imperfect model parameters which are not the control variables on the data assimilation.

Each basic scenario contains some number of subordinate scenarios, which are indicated by the decimal part of the scenario number (see the first column in Table 5.1). These subordinate scenarios vary due to the different combinations of control runs, observation types, data assimilation time windows, observation stations or control variables. Control runs are defined by changing the model parameters in the nature run. These model parameters include Chezy coefficient in the direction ξ and η of the model grid, a and b in the linear function of wind drag coefficient formula (Table 3.2). In Table 5.1, they are denoted by C_ξ, C_η, wind_a and wind_b. In most scenarios, observations of water level are used for data assimilation except in scenario 1.4 and 2.5. The time window of data assimilation is between 00:00 and 23:00 on December 6 2013 in most scenarios, except that in scenario 1.3 and 2.4 the time window is between 00:00 and 12:00.

5.2 Discussion

5.2.1 Scenario 1

In scenario 1.1, the differences between control run and nature run are Chezy coefficients in the ξ and η directions of the model curvilinear grid. Therefore, the

[1]Imperfect model parameters are those model parameters whose values in the control run are different from their counterparts in the nature run

Table 5.1: All scenarios of the twin experiments

scenario number	**control run**	**obs type**	**time window**	**obs stations**	**control variables**
1.1	GBM_C_ξ=50 GBM_C_η=60	water level	00:00 to 23:00	Helgoland Buesum	GBM_C_ξ and GBM_C_η
1.2	GBM_C_ξ=50 GBM_C_η=60	water level	00:00 to 23:00	Helgoland	GBM_C_ξ and GBM_C_η
1.3	GBM_C_ξ=50 GBM_C_η=60	water level	00:00 to 12:00	Helgoland Buesum	GBM_C_ξ and GBM_C_η
1.4	GBM_C_ξ=50 GBM_C_η=60	current_ξ	00:00 to 23:00	Helgoland Buesum	GBM_C_ξ and GBM_C_η
1.5	GBM_C_ξ=20 GBM_C_η=20	water level	00:00 to 23:00	Helgoland Buesum	GBM_C_ξ and GBM_C_η
1.6	CSM_Cu=50 GBM_C_ξ=60	water level	00:00 to 23:00	Helgoland Buesum	CSM_Cu and GBM_C_ξ
1.7	GBM_C_ξ=50 GBM_C_η=60 wind_a=0.8 wind_b=0.08	water level	00:00 to 23:00	Helgoland Buesum	CSM_Cu, GBM_C_ξ, wind_a and wind_b
2.1	GBM_C_ξ=50 GBM_C_η=60 wind_a=0.8 wind_b=0.08	water level	00:00 to 23:00	Helgoland Buesum	GBM_C_ξ and GBM_C_η
2.2	GBM_C_ξ=50 GBM_C_η=60 wind_a=0.8 wind_b=0.08	water level	00:00 to 23:00	Buesum	GBM_C_ξ and GBM_C_η
2.3	GBM_C_ξ=50 GBM_C_η=60 wind_a=0.8 wind_b=0.08	water level	00:00 to 23:00	Helgoland	GBM_C_ξ and GBM_C_η
2.4	GBM_C_ξ=50 GBM_C_η=60 wind_a=0.8 wind_b=0.08	water level	00:00 to 12:00	Helgoland Buesum	GBM_C_ξ and GBM_C_η
2.5	GBM_C_ξ=50 GBM_C_η=60 wind_a=0.8 wind_b=0.08	current_ξ	00:00 to 23:00	Helgoland Buesum	GBM_C_ξ and GBM_C_η
2.6	GBM_C_ξ=50 GBM_C_η=60 wind_a=0.8 wind_b=0.08	water level	00:00 to 23:00	Delfzijl	GBM_C_ξ and GBM_C_η
2.7	GBM_C_ξ=50 GBM_C_η=60 wind_a=0.8 wind_b=0.08	water level	00:00 to 23:00	Delfzijl Buesum	GBM_C_ξ and GBM_C_η
2.8	GBM_C_ξ=50 GBM_C_η=60 wind_a=0.8 wind_b=0.08	water level	00:00 to 23:00	Helgoland Buesum	wind_a wind_b
3.1	GBM_C_ξ=67 GBM_C_η=67 wind_a=0.8 wind_b=0.08	water level	00:00 to 23:00	Helgoland Buesum	wind_a wind_b

control run will generate model results different from the nature run. As the differences between control run and nature run are exactly known, data assimilation

and pseudo-observations can be used to adjust Chezy coefficients in the control run to decrease the difference between them. The data assimilation time window in scenario 1.1 is the period between 2013-12-06 00:00 and 2013-12-06 23:00 UTC. The observations are the water level data extracted at the station Buesum and Helgoland. Figure 5.1 shows the variations of cost function and control variables in scenario 1.1. The initial value of the cost function is 21.46. After one iteration in the outer loop, the value has a sharp decrease to 0.35. In the third iteration, the value is almost zero, meaning that the model results over the time window at the station Buesum and Helgoland have adjusted to the same results of the nature run. Figure 5.1 also shows the variations of Chezy coefficients. They change from (50.0, 60.0) to (69.88, 69.88) in the third iteration, which is exactly the same as the Chezy coefficients in the nature run. Scenario 1.1 indicates that the proposed adjoint free 4Dvar is correct, it is implemented properly and it is effective for parameter estimation.

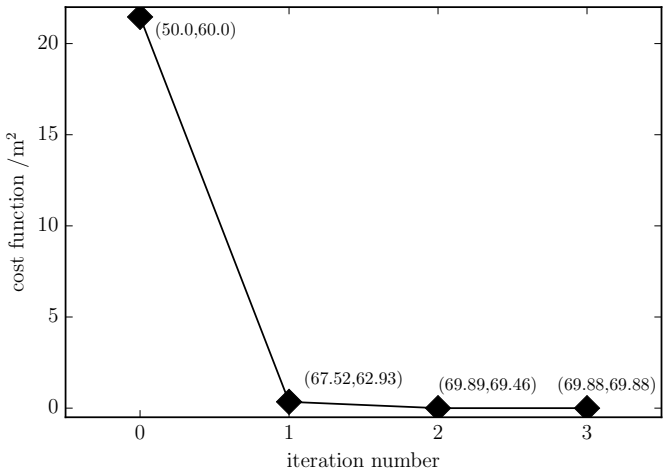

Figure 5.1: Cost function in the experiment scenario 1.1. The value pairs are the Chezy coefficients [$m^{1/2}/s$] in the ξ and η direction of model grid

Another three experiments, which are variants of scenario 1.1, are performed to test the influences of observation features on the data assimilation. In scenario 1.2 observations only at the station Helgoland are used; in scenario 1.3 the time window is 12 hours; in scenario 1.4 the current velocity in the ξ direction is used instead of water level as observations. Data assimilation in these three experiment scenarios has similar performance to the scenario 1.1. The values of the cost function decrease to almost 0 at the third iteration. The control variables are adjusted to almost the same as in the nature run no matter what type of observations, the time window or the number of stations are used. Scenario 1.1 and 1.4 indicate that different types of observations are possible to be assimilated to improve model parameters. This

is important for a coastal operational system, as coastal operational systems often have various types of observations.

Scenario 1.5 is the same as scenario 1.1 but with different first guess of Chezy coefficient in the control run. In other words, the initial values of the control variables are different in scenarios 1.1 and 1.5. Chezy coefficients in scenario 1.1 are closer to the true values than that in scenario 1.5. Figure 5.2 shows the comparison of scenario 1.1 and scenario 1.5. Both of them can converge to the nature run. Cost function in scenario 1.1 converges at the third iteration but more slowly in scenario 1.5. This implies that in order to converge faster, the first guess of control variables should be given as accurately as possible.

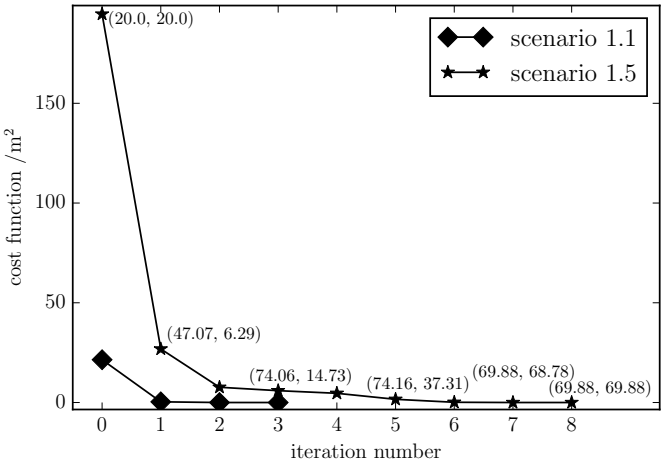

Figure 5.2: Comparison of cost function in the experiment scenario 1.1 and 1.5. The value pairs are the Chezy coefficients [$m^{1/2}/s$] in the ξ and η direction of model grid

As introduced in Chapter 3, the GBM is nested into the CSM. Therefore, it is helpful if the parameters of the CSM can be adjusted using the observations and the model results of the GBM. Scenario 1.6 is designed to test whether the proposed data assimilation method can be applied in a nesting model system. The Chezy coefficients in the ξ direction of both the CSM and GBM are changed to 50.0 and 60.0 $m^{1/2}/s$ respectively. The other model parameters are kept the same as in the nature run. The way to apply the data assimilation to a nesting model is exactly the same as that in a single model system without extra efforts. These two parameters in the control run converged to 69.88 $m^{1/2}/s$ in the fourth iteration of the outer loop.

Scenario 1.7 is designed to test the effectiveness of the data assimilation on the control variables with different units (Table 5.2). Both Chezy coefficient and wind drag coefficient (a and b) are changed based on the nature run. The observation setting is the same as in scenario 1.1. The value of the cost function gets to 2.0×10^{-6}

from 32.65 in the fourth iteration, indicating that the model parameters in the control run have been adjusted to their counterparts in the nature run.

Table 5.2: The units of the model parameters used in scenario 1.7

	CSM_Cu	GBM_C_ξ	wind_a	wind_b
unit	$m^{1/2}/s$	$m^{1/2}/s$	1	$(m/s)^{-1}$

5.2.2 Scenario 2

Experiments in basic scenario 1 show that the data assimilation works well for parameter estimate in an ocean model system if the error sources are exactly identified and the control variables contain all the inaccurate model parameters. However, model errors can be from the model parameters, model input data, insufficient resolutions and imperfect primitive equations. It is impossible to put all the error sources in the control variables of data assimilation. Therefore, the experiments in scenario 2 are designed to test the performance of data assimilation in realistic situations.

The control run in scenario 2.1 is the same as in scenario 1.7 except that the control variables only contain the Chezy coefficient. Figure 5.3 shows the values of cost function in scenario 2.1. The cost function converges to the value 4.76 after the second step. The Chezy coefficient pair converges near the value of (74.24, 58.23), which are not equal to the Chezy coefficient (69.88, 69.88) in the nature run. The decrease of the cost function indicates that model results of control run at Buesum and Helgoland over the time window are closer to the nature run after data assimilation, as shown in Figure 5.4. But outside the time window, the updated water level in the control run is improved as well. Table 5.3 shows the RMSD of model and observations before and after data assimilation at four stations. Station Husum and Cuxhaven are not used in data assimilation. There are also improvements after data assimilation both within and outside time window, but relative decrease of RMSD within the time window are larger than outside time window. This is because data assimilation tries to decrease the differences of model results and observations only within the time window. Although wind drag coefficient in control run 2.1 has errors compared with the nature run, it is kept unadjusted. Therefore, Chezy coefficient in control run 2.1 cannot be adjusted back to the true values in the nature run. It is expected that the optimal control variables are dependent on the observation types, time window and number of observations. Some following experiments are carried

out to test the influence of observations on data assimilation.

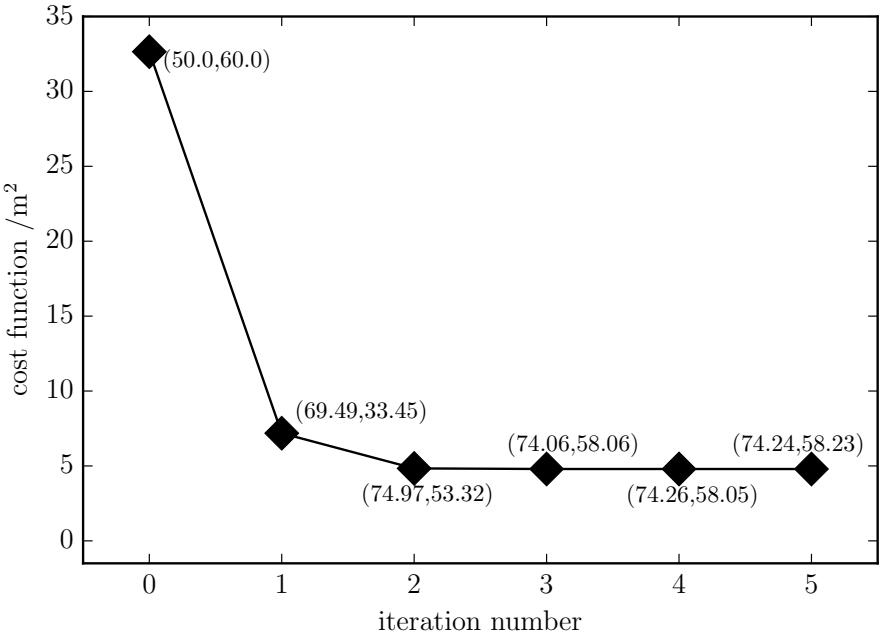

Figure 5.3: Cost function in the experiment scenario 2.1. The value pairs are the Chezy coefficients [m$^{1/2}$/s] in the ξ and η direction of curvilinear grid

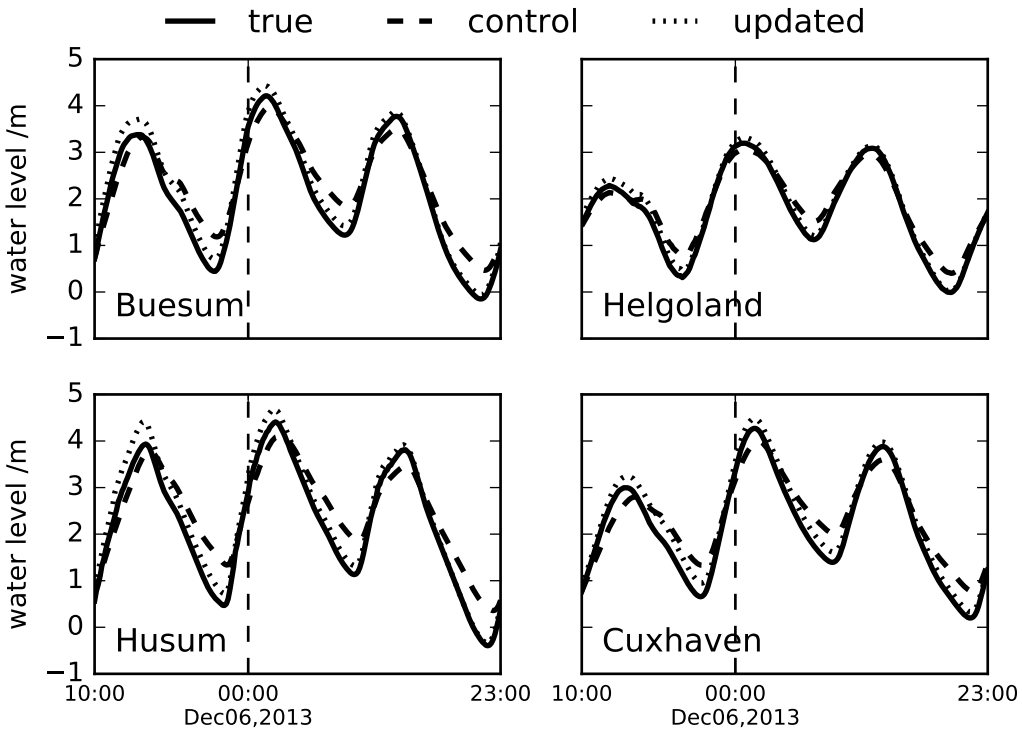

Figure 5.4: Time series of water level in nature run, control run in scenario 2.1

Table 5.3: Root mean square deviation(m) of model and observations before and after data assimilation in scenario 2.1

	out of time window			within time window		
	before DA	after DA	relative decrease	before DA	after DA	relative decrease
Buesum	0.442	0.329	0.255	0.421	0.168	0.600
Helgoland	0.254	0.168	0.339	0.239	0.078	0.673
Cuxhaven	0.417	0.275	0.340	0.400	0.184	0.539
Husum	0.529	0.351	0.337	0.496	0.175	0.647

In scenario 2.2, the water level data at station Buesum is used for data assimilation and in scenario 2.3 the station Helgoland is used. Scenario 2.4 has a shorter time window than that in scenario 2.1. In scenario 2.5, current velocity data is used instead of water level. Table 5.4 shows the results of these scenarios. The Chezy coefficient pair in scenario 2.3 is updated to (71.10, -16.87). However, negative Chezy coefficient is not possible physically. Therefore, in scenario 2.3 data assimilation fails to adjust Chezy coefficient and the model results will not be shown in the following analysis. Table 5.4 shows the updated Chezy coefficient and the RMSD after data assimilation in four scenarios. Note that RMSDs before data assimilation are in Table 5.3. The adjusted Chezy coefficient in the four scenarios does not converge to the same value, but the RMSDs of water level after data assimilation are decreased in a very similar way. In scenario 2.5, in which the current velocity is used for data assimilation, the RMSDs are larger than at the other three stations. But the differences are still small. This indicates that observations of current velocity from ADCP or high frequency radar can also be used for data assimilation purposes. Figure 5.5 shows the time series of water level after data assimilation at the four stations. They are very close, indicating that the updated Chezy coefficients have similar effect on the modeled water level at the four stations even though their values are very different.

Table 5.4: Root mean square deviation(m) of model and observations over the time window before and after data assimilation in scenario 2.1, 2.2, 2.4 and 2.5

	updated Chezy coefficient	Buesum	Helgoland	Cuxhaven	Husum
scenario 2.1	(74.25,58.18)	0.168	0.078	0.184	0.175
scenario 2.2	(73.72,62.52)	0.168	0.080	0.184	0.176
scenario 2.4	(72.49,61.69)	0.169	0.082	0.186	0.175
scenario 2.5	(72.82,75.53)	0.171	0.087	0.191	0.188

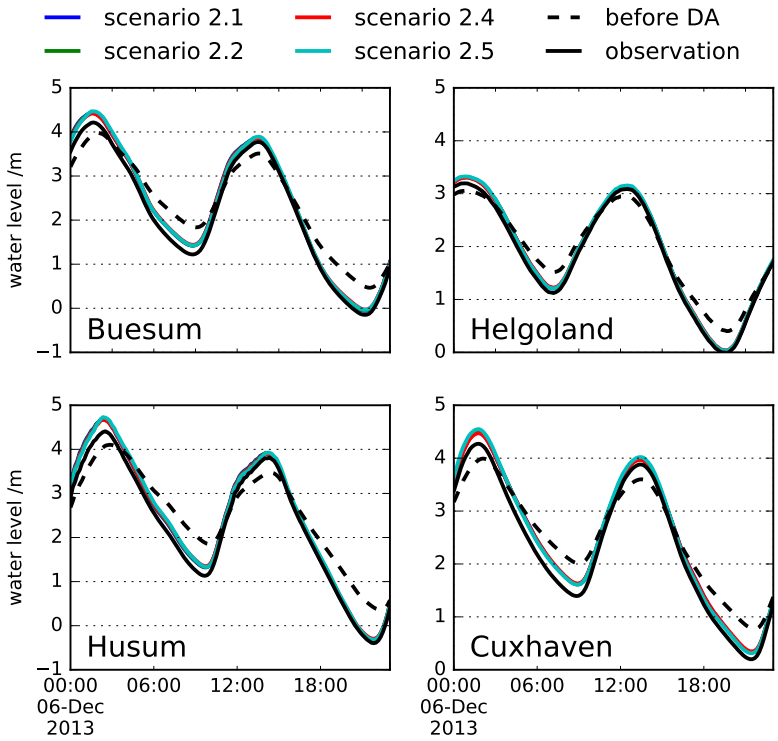

Figure 5.5: Comparison of water level in scenario 2.1, 2.2, 2.4, 2.5

In the method of optimal interpolation or Kalman filter, one widely used assumption for the construction of the covariance matrix is the distance-dependent spatial correlation. In these methods, data assimilation at one station can only influence the adjacent area. If one given point is too far away from the assimilation station, the effect of data assimilation is negligible. Scenario 2.2 and scenario 2.6 are compared in order to investigate the spatial effect of 4Dvar on the water level correction. In scenario 2.2, water level data at Buesum is used while in scenario 2.6 the water level data at Delfzijl is used. The converged Chezy coefficients in scenario 2.2 are (73.71, 62.50) and (72.30, 73.09) in scenario 2.6. The difference between the Chezy coefficient in the η direction is 10.6, which in most cases is big enough to result in obvious variations of model results. Table 5.5 shows that data assimilation in both scenarios is effective to decrease the errors. There are obvious improvements at the station Borkum, Huibertgat and Wilhelmshaven, which are either closer to the station Delfzijl or on the same side of the German Bight. However, at List_sylt, Toenning and Hoernum, which are distant from Delfzijl, there are also evident improvements. Therefore, the 4Dvar data assimilation improves the model results not fully related to the distance but based on the model dynamics.

In the above-mentioned experiments in scenario 2, the control runs are obtained

Table 5.5: Root mean square deviation of water level(m) from model and observations over the time window before and after data assimilation in scenario 2.2, 2.6 and 2.7

	before DA	scenario 2.2	scenario 2.6	scenario 2.7
Hoernum	0.240	0.076	0.060	0.063
List_sylt	0.207	0.099	0.086	0.087
Toenning	0.359	0.177	0.158	0.166
Huibertgat	0.128	0.060	0.051	0.051
Borkum	0.192	0.099	0.084	0.084
Wilhelmshaven	0.424	0.197	0.150	0.148

by contaminating the Chezy coefficients and wind drag coefficient in the nature run. Then only Chezy coefficients are adjusted in the data assimilation. Scenario 2.8 is made to test how effective to reduce the RMSD by adjusting the wind drag coefficient C_d instead of Chezy coefficient. The data assimilation setting is the same as in scenario 2.1. The value of cost function in scenario 2.8 converged to 20.18, much larger than 4.80, which is the converged value in scenario 2.1. The comparisons of RMSD of water level are shown in Table 5.6. There are also improvements in scenario 2.8 over the time window, but the RMSDs at all the stations are larger than that in scenario 2.1. This indicates that adjusting wind drag coefficient is less effective than adjusting Chezy coefficient. a and b in scenario 2.8 converged approximately to the value of 1.837 and -0.0037 respectively, showing that C_d decreases with the increase of wind speed. But the updated b contradicts the widely accepted relation between wind drag coefficient and wind speed that is described in Chapter 3.

This implies a potential risk when using 4Dvar for parameter estimate. The model parameters may be adjusted to values which are out of the valid range and as a consequence the model may be crashed, which arises in scenario 2.3. In order to avoid such risk, control variables should contain as many model parameters as possible. However, in most cases, control variables only contain small number of parameters which are considered to be most uncertain and sensitive to model results. In weather forecasting, the initial conditions are adjusted to improve the model results in a time window and therefore the forecasting in a short future period after the time window is expected to be improved.

5.2.3 Scenario 3

A new basic scenario, scenario 3 is designed to discuss the effect of parameter (not to be adjusted by data assimilation) accuracy on the data assimilation of control

Table 5.6: Root mean square deviation of water level(m) from model and observations over the time window before and after data assimilation in scenario 2.1 and 2.8

	before DA	scenario 2.1	scenario 2.8
Buesum	0.422	0.168	0.334
Cuxhaven	0.400	0.184	0.306
Helgoland	0.239	0.078	0.183
Husum	0.496	0.175	0.409

variables. Only a and b for wind drag coefficient formula will be adjusted by data assimilation. Scenario 3.1 is obtained by contaminating the Chezy coefficient and wind drag coefficient of the nature run. The Chezy coefficient is given to (67.0, 67.0), which are more accurate than those in scenario 2. a and b are the same as in scenario 2.8, as well as the data assimilation setting.

To investigate the relative importance of inaccurate Chezy coefficient, a and b, three models are designed as listed in Table 5.7. They are named as Model A, Model B and Model C. Model errors (discrepancies between results of control run and nature run) caused by inaccuracy of Chezy coefficient are obtained by subtracting results of nature run from results of Model A or Model B; Model errors caused by error of wind drag coefficient are obtained by subtracting results of nature run from results of Model C. In scenario 2.8, the model errors are approximately equal to the summation of model error of the Model A and C; In scenario 3.1, the model errors are approximately equal to the summation of model error of the Model B and C. Figure 5.6 shows these model errors at station Buesum and Helgoland. It can be found that at both stations, the model error caused by Chezy coefficient (50, 60) is much larger than the other two. The model error caused by a and b is larger than Chezy coefficient (67, 67). Therefore, in the control run of scenario 2.8, Chezy coefficient is the main error source. While in the control run of scenario 3.1, a and b bring more errors than the inaccurate Chezy coefficient.

Table 5.7: Chezy coefficient and wind drag coefficient in model A, B and C

	Chezy coefficient	a	b
nature run	(69.88,69.88)	0.61	0.063
Model A	(50,60)	0.61	0.063
Model B	(67,67)	0.61	0.063
Model C	(69.88,69.88)	0.8	0.08

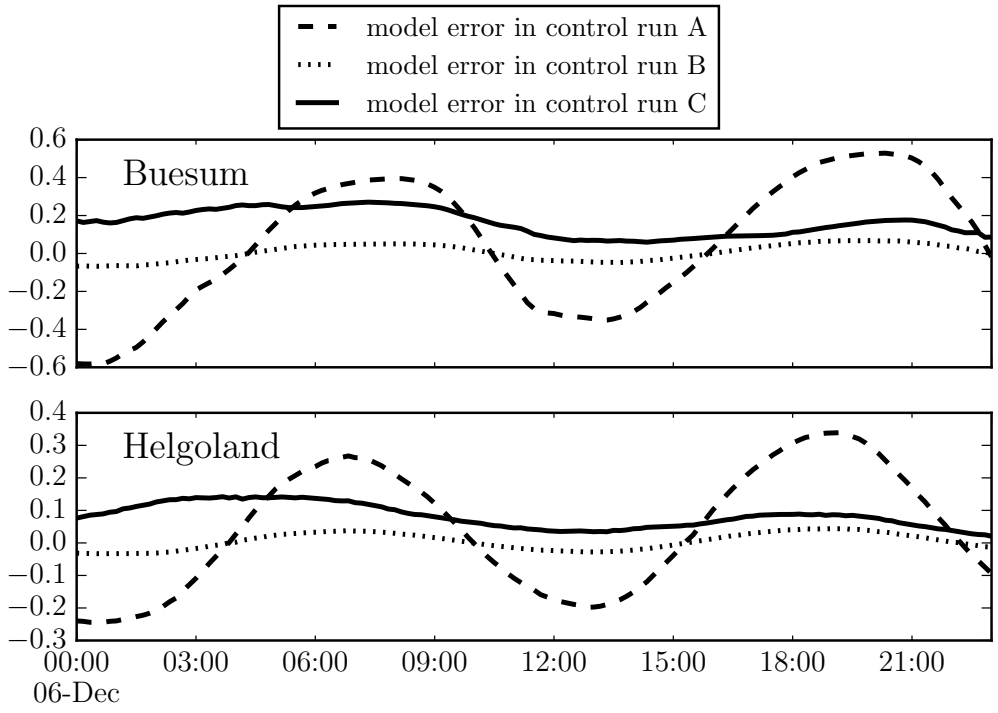

Figure 5.6: Model errors (m) in the model A, B and C at Buesum and Helgoland

In both scenario 2.8 and 3.1, the first guess of a is 0.8 and the first guess of b is 0.08. After data assimilation, the updated a and b are different considerably in scenario 2.8 and scenario 3.1 (Figure 5.7). As mentioned above, in scenario 2.8 b became negative, which is not physically meaningful. In scenario 3.1, the updated a and b are closer to their true values than scenario 2.8. This is because in scenario 3.1, Chezy coefficients is much more accurate compared with in scenario 2.8 and most model errors result from inaccuracy of wind drag coefficient. In 4Dvar, the discrepancies between model results and observations are the sources of the model parameters correction. If one parameter which brings large errors to the model results is not adjusted by the data assimilation, the corrections of the control variables will be too much and then some parameters may become meaningless. Another consequence is that the model results cannot be improved effectively because the model parameters with large uncertainty have not been reasonably adjusted.

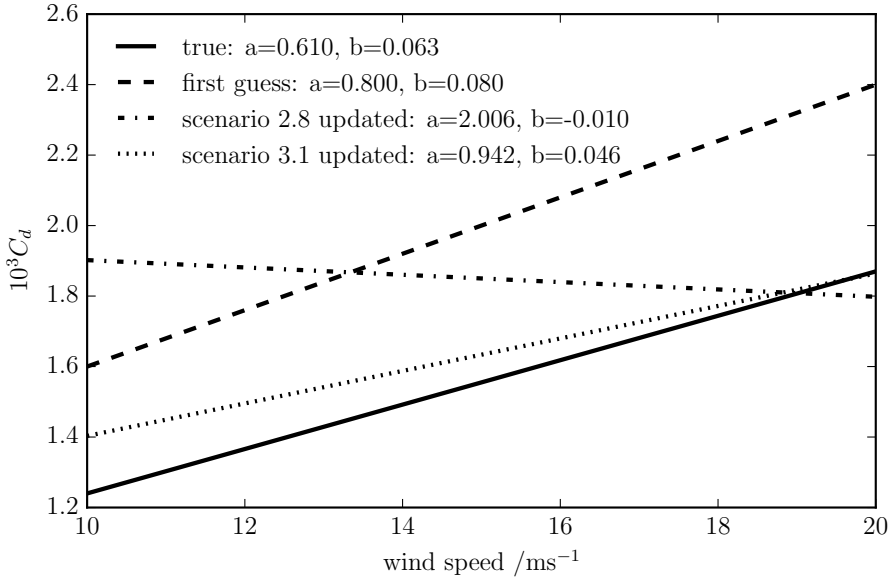

Figure 5.7: The linear function of wind drag coefficient with respect to wind speed in nature run, and different scenarios

5.3 Conclusions of the twin experiments

The twin experiments in this chapter show the correctness and effectiveness of the proposed adjoint free 4Dvar for ocean models parameter estimation. If the control variables contain all the inaccurate model parameters and the observations are error-free, 4Dvar is able to pull all the parameters to their true values and the model becomes error-free. This is not the case in reality, but 4Dvar is still able to effectively reduce the RMSD between model results and observations. Twin experiments also indicate that model parameters with higher uncertainty should be adjusted by data assimilation with higher priority, otherwise the model cannot be improved effectively. The performance of 4Dvar in an operational system could also be influenced by the observation number, quality and type. Therefore, it is necessary to investigate the influencing factors of model accuracy. This is the topic of Chapter 6.

Chapter 6

Sensitivity tests of the storm surge model

In Chapter 3, the model Delft3D is introduced, which is used to simulate the storm surge in this thesis. The preliminary model results of a storm surge event occurring in early December 2013 are also presented, showing good correlation with observations but underestimate the peak surge.

The natural processes of storm surge can be represented in mathematical equations. Those equations are either simple or sophisticated, but they can never be equal to the natural processes. Numerical models are obtained by discretizing the mathematical equations on time, space or spectrum. As a consequence, small scale processes may not be resolved by models. Therefore, there are always discrepancies between model results and the true values. Model skill can be increased by adopting more processes, using more sophisticated primitive equations or higher model resolution.

Ocean models have a large number of parameters. Some parameters are better known, such as gravitational acceleration and sea water density. Some parameters are empirical and less well known like parameters relevant to turbulence. These less well known parameters will largely affect model skills.

Once the model grid and parameters are fixed after a calibration process, model results are uniquely determined by the boundary conditions, including initial conditions, open boundary conditions and sea surface and bottom conditions. The quality of these conditions will affect the model accuracy. They are specified with observed data or empirical values with different degree of errors. One example is the initial conditions for weather forecast. Initial conditions have significant effect on the forecast. Therefore, meteorologists try to improve the initial conditions to improve the skill of weather models.

The data assimilation described in Chapter 4 falls into the category of strong constraint 4Dvar, which only deals with the errors of model parameters and boundary conditions. In Chapter 5, it has been shown that model results can be improved by adjusting the model parameters using 4Dvar. Although any model parameter can be adjusted by data assimilation, numerical experiments indicate that those parameters with high inaccuracy and large effects on the model results should be adjusted in order to avoid the excessive adjustment and to decrease the model errors more effectively. Therefore, it is necessary to examine the accuracy of the model parameters and perform sensitivity tests before implementing the data assimilation. In the following, a series of sensitivity tests will be shown with respect to open boundary conditions, initial conditions, wind shear stress, sea bed conditions.

The general steps of a sensitivity test for a given model parameter are:

1. Investigate the accuracy of the model parameter. Accuracy indicates the discrepancy between the first guess of the parameter and its true value. As it is impossible to know the true values, observations are used instead of true values. In the situation when several first guesses of the model parameter from different sources are available, these first guesses can form an ensemble. It is assumed that these first guesses are unbiased estimate of the true values. Therefore, the standard deviation (STD) of these first guesses can represent the accuracy. If the standard deviation is small, these first guesses have a high accuracy.

2. Generate a group of perturbations according to the accuracy, add them to the first guess of model parameters and form an ensemble of parameter, or directly use the ensemble of first guesses of model parameters from different sources.

3. Run the model ensemble.

4. Analyze the sensitivity of model results to the model parameters. The sensitivity to a certain model parameter is measured by the standard deviation(STD) of model results obtained in step 3. STD is calculated by the method below:

$$\mathrm{STD} = \sqrt{\frac{1}{N * N_T - 1} \sum_{i=1}^{N} \sum_{j=1}^{N_T} (\eta_{ij} - \bar{\eta})^2}$$

$$\bar{\eta} = \frac{1}{N * N_T} \sum_{i=1}^{N} \sum_{j=1}^{N_T} \eta_{ij}$$

where η denote the simulated water level over a certain period, N is the number

of ensemble members and N_T is the number of time steps over the period. In the sensitivity tests below, the period is from 2013-12-05 12:00 UTC to 2013-12-06 12:00 UTC and the time step is 10 minutes, in which the surge peak occurred in the German Bight.

6.1 The role of the larger model for the storm surge simulation

As mentioned previously in section 3.2.2, the CSM provides water level on the open boundary of the GBM. The water level on the open boundary of the GBM includes the effect of both tide and wind-induced water level variations. The domain of the GBM is much smaller than the CSM. It is expected that the wind effect along the open boundary of the GBM should not be neglected in the storm period. One test is designed to examine the wind effect on the open boundary of the GBM. This test consists of two models based on the storm surge model in Chapter 3:

Model I: the CSM with wind stress

Model II: the CSM without wind stress

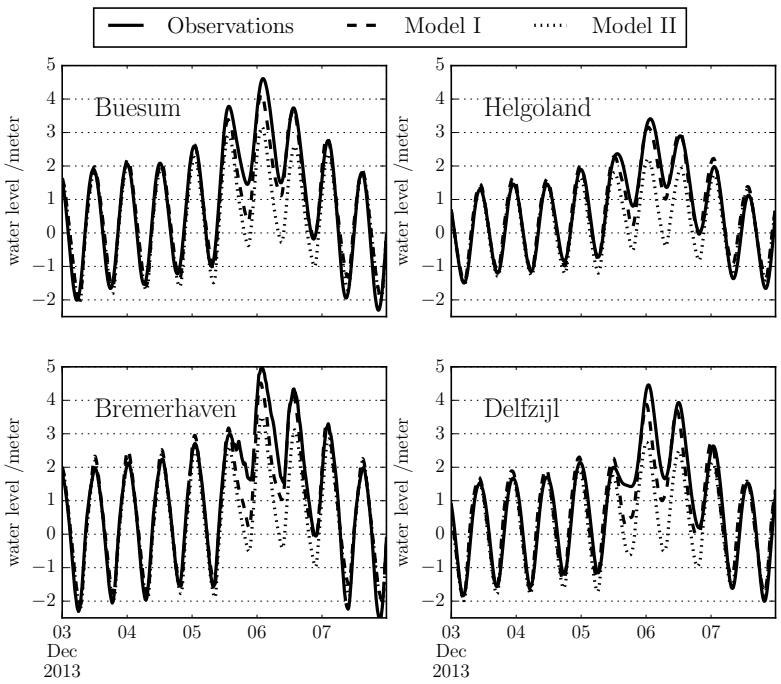

Figure 6.1: Time series of water level with and without wind effect on the open boundary of the GBM. Solid line is the observed water level; dashed line is from model I; dotted line is from model II

The time series of water level at four stations from model I, model II and the observations are shown is Figure 6.1. In normal weather condition, the simulated water level in model I in model II are similar. The RMSD between model I, model II at the four stations within December 3 and December 4 is 0.168m but increases to 0.861m in the storm period. Therefore, a model covering a sufficiently large area should be preferred for a more accurate storm surge modeling. Otherwise, the surge is strongly underestimated.

6.2 The interactions between tide and surge

One approach for predicting water level in the storm is running a storm surge model without tide. Then the total water level is obtained by adding the modeled surge to the tide predicted by the harmonics. Tide is the dominant hydrodynamic process for most coastal waters, and it is expected to have some influences on the storm surge simulations. To investigate the role of tide on storm surge simulation, three models are run and the model settings are listed in Table 6.1.

Table 6.1: Models for the test of tide-surge interaction

	drving forces
Model_Total	wind and tide
Model_Tide	tide only
Model_Wind	wind only

From the three models listed in table 6.1, two different surges are obtained:

- **Surge1** is obtained by subtracting the modeled water level of Model_Tide from modeled water level of Model_Total.
- **Surge2** is the model results of Model_Wind.

The time series of **Surge1** and **Surge2** at the station Buesum are shown in Figure 6.2 and **Surge1** includes the effect of tide-surge interaction. **Surge1** and **Surge2** have similar responses to the storm. Both of them have two water level peaks between 2013-12-05 12:00 UTC and 2013-12-06 12:00 UTC and they have similar peak values. However, they also have evident differences. The peak value of **Surge1** occurs about 150 minutes later than **Surge2** in the storm period; the peak value in **Surge1** is 14.7cm lower than that in **Surge2**. This difference is

caused by the bottom shear stress in the model; bottom shear stress is proportional to the square of depth-averaged current velocity in a 2D model. Due to the tide open boundary conditions in the Model_Total, the magnitude of current velocity is much larger than that in the Model_Wind (Figure 6.3). Therefore, the bottom shear stress in Model_Total is much larger than that in Model_Wind. The energy dissipation on the sea bed is larger in the former model and this results in the smaller peak value and phase lag. In the period of 2013-12-06 12:00 UTC and 2013-12-07 12:00UTC there are three smaller peaks in the **Surge1**; while water level of **Surge2** decreases smoothly. This may be explained by the more variable water level in the Model_Total. Pugh (1987) put forward a simplified formula to estimate the water level slope, $\frac{\partial \eta}{\partial x} = CW^2/H$. C is a constant including the effect of gravity and density; W is the constant wind speed blowing on the sea surface; H is the total depth. This formula indicates that the smaller the water depth is, the higher will be the surge. For a real simulation with spatially varying wind, the water level change is much more complex. But the water depth change is the most probable reason for the small peak in the **Surge1** as the peaks always happened at the time of low water level.

To sum up, the tide is a necessary driving force in the storm surge model due to the influences of tide on the surge simulation shown above.

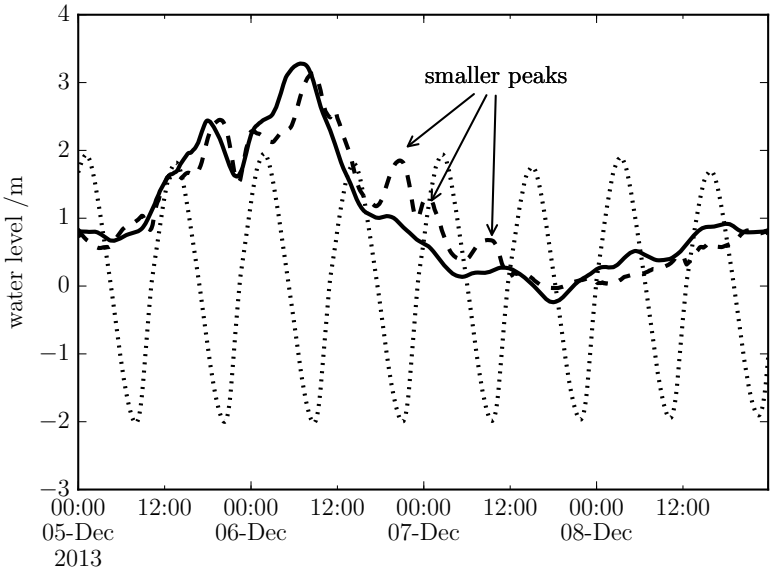

Figure 6.2: The surge simulation with and without tide effect at the station Buesum. Dashed line is the **Surge1**; solid line is the **Surge2**; dotted line is the tide.

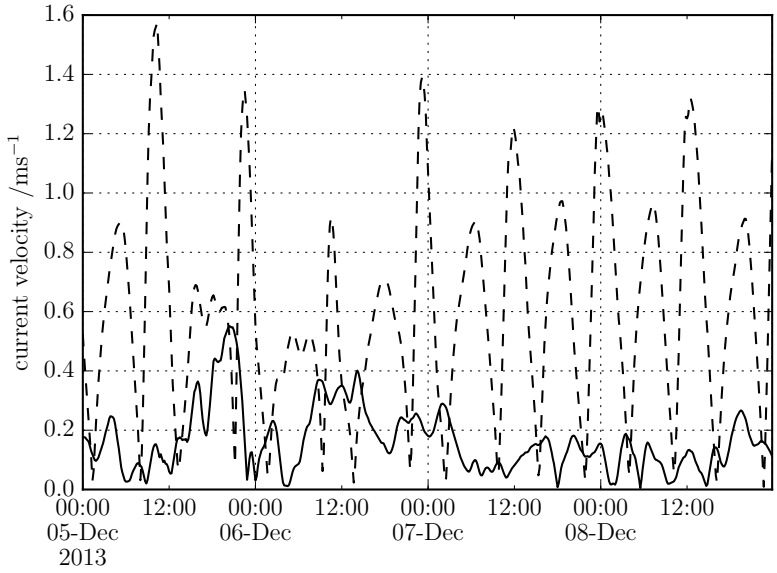

Figure 6.3: The magnitude of current velocity at the station Buesum. Dashed line is the **Surge1**; solid line is the Surge2; dotted line is the tide.

6.3 Open sea boundary conditions of the CSM

As described in Chapter 3, open sea boundary conditions of the CSM are specified with tidal harmonic constituents of some dominate harmonics. Here an ensemble of harmonic constituents extracted from four different tide models will be added to the CSM to investigate the accuracy of harmonic constituents and the effect on water level simulations. The four databases are TPXO7.2, OUS12, DTU10 and HAM12. The tidal harmonic constituents are calculated by means of numerical global tide models and data assimilation of water level data measured by satellite altimeters. See Stammer et al. (2014) for a detailed descriptions of these global tide models.

Figure 6.4 shows the comparison of amplitude and phase of four main harmonics from the four tide models. Among the four harmonics, M2 is the dominant harmonic in terms of amplitude and S2 is the second largest. Amplitude and phase of M2 and S2 from the four tide models have very good agreements; Amplitude and phase of K1 and O1, which are about one order smaller than S2 and M2, also have similar values from the four databases. These comparisons indicate that the modeled harmonic constituents along the open boundary of the CSM have a high accuracy.

Four models are run with different open sea boundary conditions extracted from databases of TPXO7.2, OUS12, DTU10 and HAM12. Figure 6.5 shows the simulated water level of these models at Helgoland and Buesum. The differences between the four models are very small, even hardly distinguishable in Figure 6.5. The STDs

at all the stations are around 0.03m, which is a very small value compared with the errors of the storm surge in Chapter 3. Therefore, it can be concluded that the simulated water level is not sensitive to the uncertainties of tidal harmonics imposed at the open boundaries of the CSM .

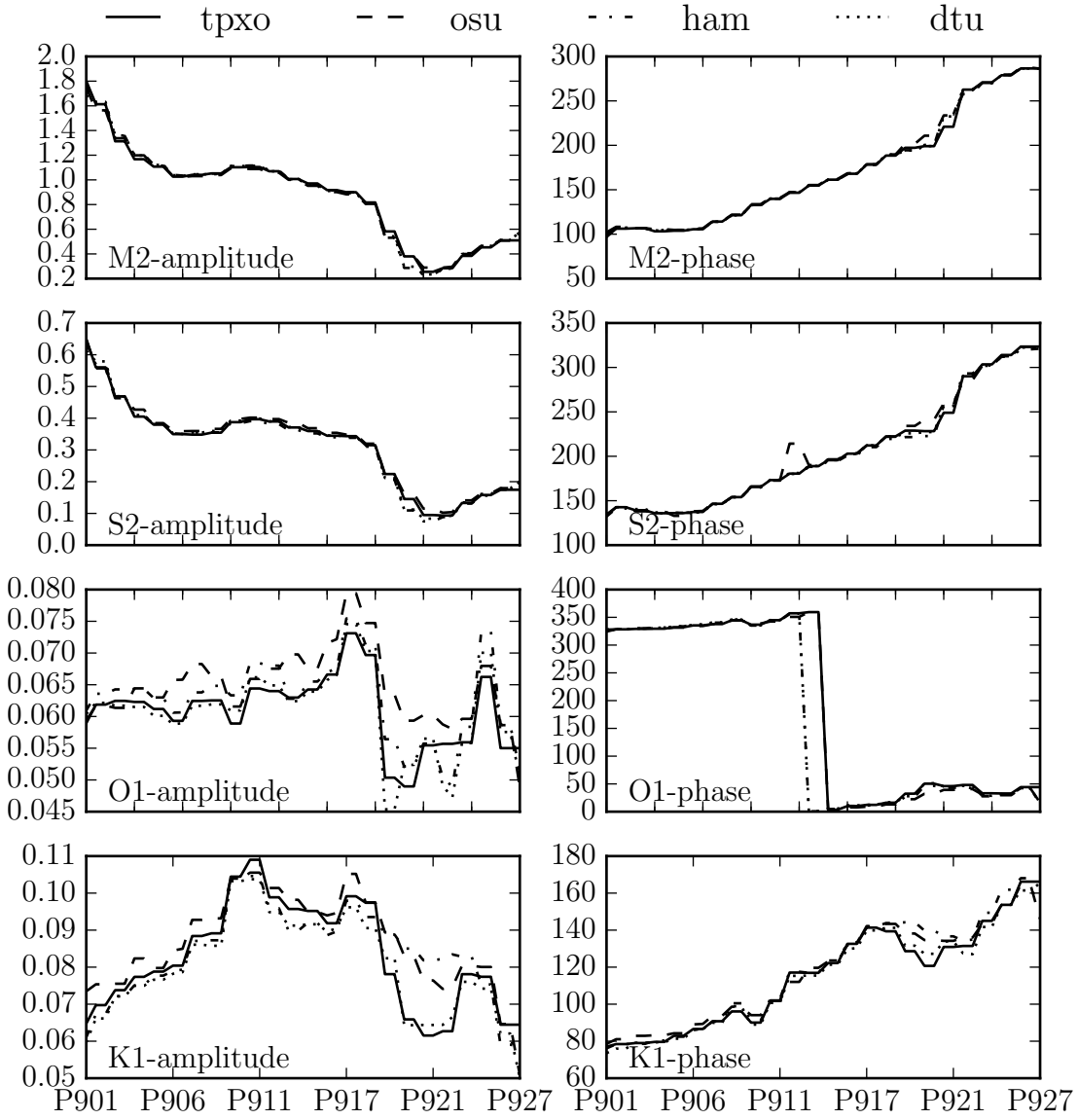

Figure 6.4: Amplitude (m) and phase (degree) of M2, S2, O1 and K1 from four databases along the open boundary from location P901 to P927

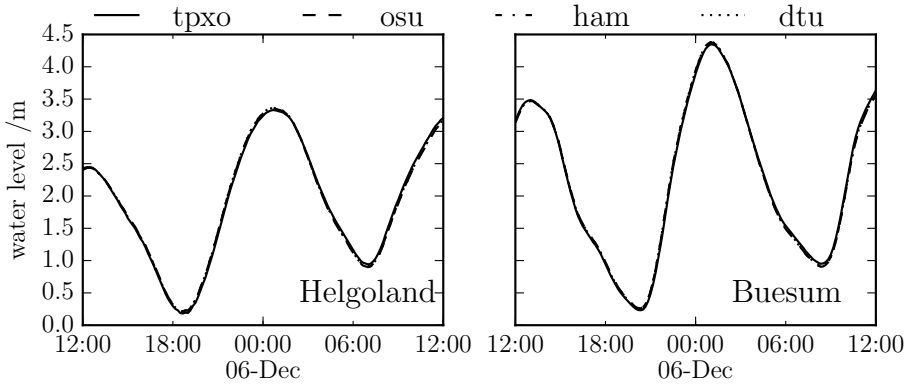

Figure 6.5: Comparison of water level (m) from four models at Helgoland and Buesum

6.4 Wind shear stress

Wind stress over sea surface is the main driving force of the surge during a strong storm (Horsburgh and De Vries, 2011). It is calculated given air density, wind drag coefficient and wind speed. Abdalla and Cavaleri (2002) studied the influences of air density on the wave height and found that the variations of wave height resulting from air density variability are usually smaller than 1 cm. Therefore, air density is expected to have only minor effect on storm surge simulation. Wind speed data is from the meteorological model COSMO-EU with spatial resolution of 0.0625° and time interval of 1 hour. Although the modeled wind data agree well with observed wind speed (Figure 3.4), they still contain errors and cannot resolve completely all the wind features. Wind drag coefficient is derived from wind measurements and contain large uncertainties. In this section, the sensitivities of storm surge simulation to the modeled wind data and wind drag coefficient will be investigated separately.

6.4.1 Wind speed

In this sensitivity test, wind speed data is used from three meteorological models. Besides the hourly wind data obtained from DWD (named DWD1 below) described in Chapter 3, another two coarser data are available from DWD global model (named DWD3 below) and the European Centre for Medium-Range Weather Forecasts (named ECMWF3 below) respectively. Time interval of DWD3 and ECMWF3 is three hours. Figure 3.4 shows the their comparisons in the period of the storm in early December 2013. In order to know the effect of wind stress on the water level simulations in a normal weather condition, the model is also run in mid-July,

2011. As shown in Figure 6.6, the wind speed at Sylt in mid-July ranges between 7 to 15 m/s, which is an intermediate value in the year 2011; while the wave height at Sylt is among the smallest in the year 2011. Therefore, the mid July 2011 is a representative period of the normal weather conditions.

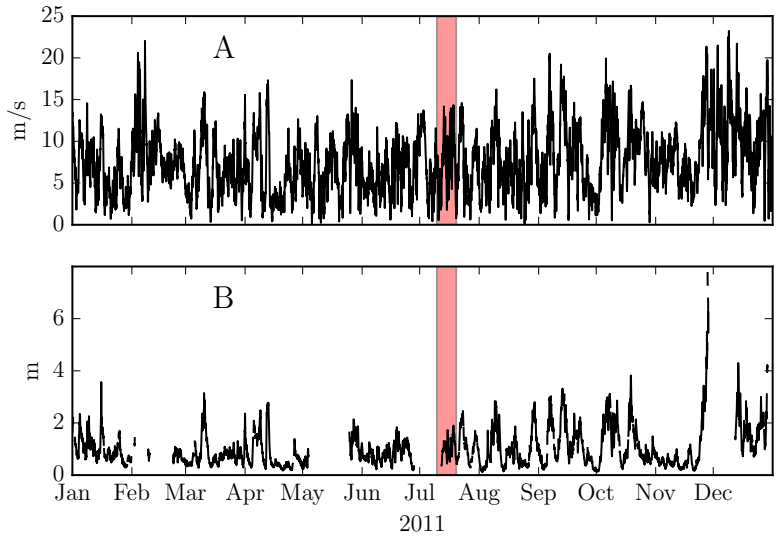

Figure 6.6: A: wind speed at Sylt (DWD1) in 2011; B: wave height at Sylt in 2011. The red patched zone indicates the period of mid July 2011.

In the early December 2013, all the air pressure data from meteorological models have very good agreements with the observed air pressure. Air pressure affects water level in a straightforward way, i.e., the inverse barometric effect; Higher air pressure results in low sea water level. Therefore it is expected that air pressure contribute very small errors to the total errors of storm surge simulations. The modeled wind is fairly consistent with the observations, but not as good as the air pressure data, especially the ECMWF modeled wind. The peak wind speed has an underestimate of about 5 m/s compared with observed wind speed at the station Sylt. The two modeled wind data by DWD is better in terms of peak wind speed, but they are 2-3 hours earlier than the observed. This comparison is only at one point, but it is estimated that the modeled meteorological data also have good accuracy over the storm surge model domains.

The wind speed in mid July 2011 at Sylt (Figure 6.7) was much smaller than that in the storm of December 2013. However, the correlations of wind speed between modeled and observed wind are smaller than that in early December 2013. One possible reason is that the local processes in a weather system are more significant under normal weather condition. As a result, the interpolated wind speed data

at Sylt from the three meteorological models with different spatial resolution have smaller correlation than that in the storm period.

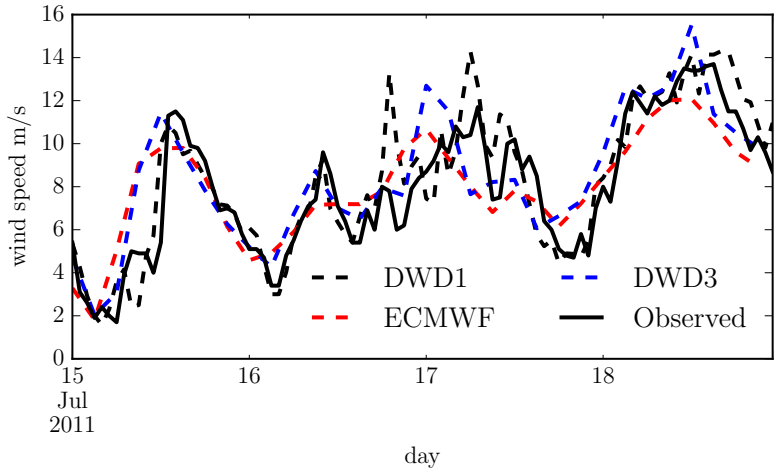

Figure 6.7: Wind speed at Sylt in the mid July 2011 - normal weather conditions

Six models are run with meteorological data from DWD1, DWD3 and ECMWF during the two periods respectively. The results are shown in Figure 6.8. The left two figures in Figure 6.8 show comparisons in the period of mid July 2011, over which there was no storm in the North Sea. It can be seen that the modeled water level show very small discrepancies both at Buesum and Helgoland. The STD between the three models at Buesum and Helgoland are 0.021m and 0.015m respectively. The RMSDs (between models and observations) in this period at Buesum are 0.405m, 0.413m and 0.423m for the models with DWD1, DWD3 and ECMWF3 data respectively; for Helgoland, the RMSDs are 0.137m, 0.139m and 0.147m. The above-mentioned STD and RMSDs indicate modeled water level are not sensitive to the wind data under normal weather conditions. DWD1 wind data is of better quality than the other two, but the improvement of the model results is very slight.

The RMSDs in early December 2013 at Buesum are 0.609m, 0.889m and 0.964m for the models with DWD1, DWD3 and ECMWF; and the RMSDs at Helgoland are 0.319m, 0.569m and 0.593m. Model wind DWD1 wind is obviously the best among the three wind data. Although the DWD3 wind data agrees well with the observations at Sylt as shown in Figure 3.4, more detailed wind information is lost than DWD1 data which has better resolution. The averaged STD is 0.175m and 0.145m at Buesum and Helgoland over the period between 2013-12-05 12:00 and 2013-12-06 12:00, which are much larger than the STD in the mid July 2011. This indicates that the quality of meteorological data is important for storm surge simulation. If it is possible, data of higher resolution should be used. However,

there are still large deviations between model results with DWD1 and observations. In a strong storm, hourly wind data may be still not enough to represent some important processes that change rapidly and thus result in errors of storm surge simulation. The errors are also caused by inaccuracy of wind drag coefficient, which is investigated in the next section.

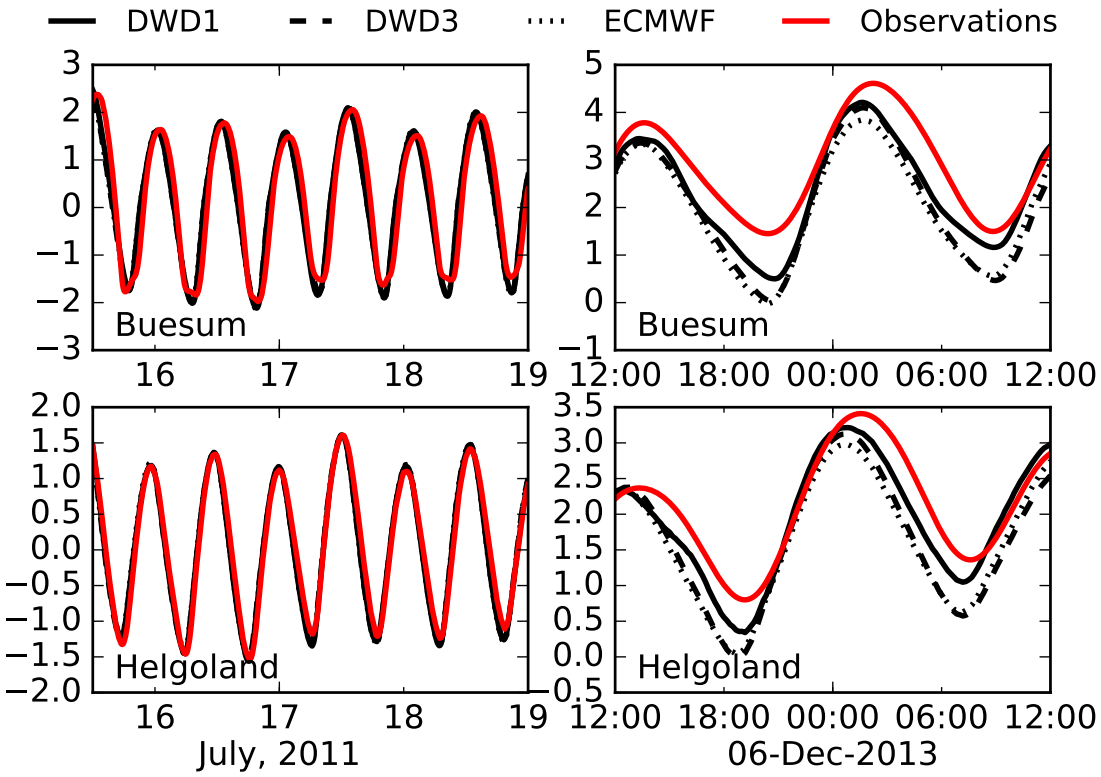

Figure 6.8: Comparison of water level(m) from four models at Helgoland and Buesum in mid July 2011 (left) and early December 2013 (right)

6.4.2 Wind drag coefficient

Wind drag coefficient C_d also affects the wind stress over the sea surface. As described in Chapter 3, in the model Delft3D the wind drag coefficient is a piecewise function of wind speed. The function is uniform over the model domain, but C_d varies spatially with wind speed. There have been a number of functions of C_d with respect to wind speed defined by different authors. These functions are derived based on the measurements of wind stress and wind speed under different sea conditions, e.g. wave state and bathymetry. The measurement errors could also produce uncertainties of the functions. These linear functions are shown in Table 3.2, in which the mean values of a and b are 0.787 and 0.069 and their standard deviations

are 0.238 and 0.029. Similar to the sensitivity test of water level to wind field data, the storm surge models are run over the period of mid July 2011 and early December 2013 respectively. In these models three functions of C_d proposed by Smith (1980) (Smith2), Sheppard et al. (1972) (Sheppard1), Donelan (1982). A constant C_d 0.0025 independent of wind speed is also used, which was often applied in storm surge models (Horsburgh and De Vries, 2011).

Figure 6.9 shows the water level from models with different wind drag coefficient and observations. Table 6.2 shows the STD and RMSD in this sensitivity test. In mid July 2011, the water level is not sensitive to the wind drag coefficient C_d both at Buesum and Helgoland. The STDs at the two stations are 0.036m and 0.027m respectively, which are similar to the STD in the sensitivity tests of wind data in section 6.4.1. In the early December 2013, water level becomes much more sensitive to C_d. STDs are as high as 0.626m and 0.502m at Buesum and Helgoland. Figure 6.9 shows that model with Sheppard has the largest contribution to STD, indicating Sheppard formula of C_d is less reasonable than other three. The model with constant C_d 0.0025 is the best in terms of RMSD at Buesum and Helgoland in the early December 2013. But in the mid July 2011, The constant C_d 0.0025 is a little worse than the other three. Although a constant C_d is not correct in physical sense, it is more practical in both normal and stormy weather. This implies that C_d can be adjusted to improve the storm surge simulation even if it is no longer valid. Smith formula is the second best after the constant 0.0025. The possible reason is that Smith formula was obtained from the measured wind data at a platform off the coastline of Holland, which is close to the German Bight. That may indicate that functions of C_d relative to wind speed is location dependent. The functions in Table 3.2 are obtained at different locations and as a result they have large variability.

Table 6.2: The STD (m) and RMSD (m) from models with different function of C_d at Buesum and Helgoland

	Buesum		Helgoland	
	Mid July 2011	Early December 2013	Mid July 2011	Early December 2013
STD	0.036	0.626	0.027	0.502
Smith_RMSD	0.405	0.609	0.137	0.319
Sheppard_RMSD	0.413	0.767	0.146	0.797
Donelan_RMSD	0.405	0.739	0.137	0.415
constant_RMSD	0.419	0.377	0.154	0.283

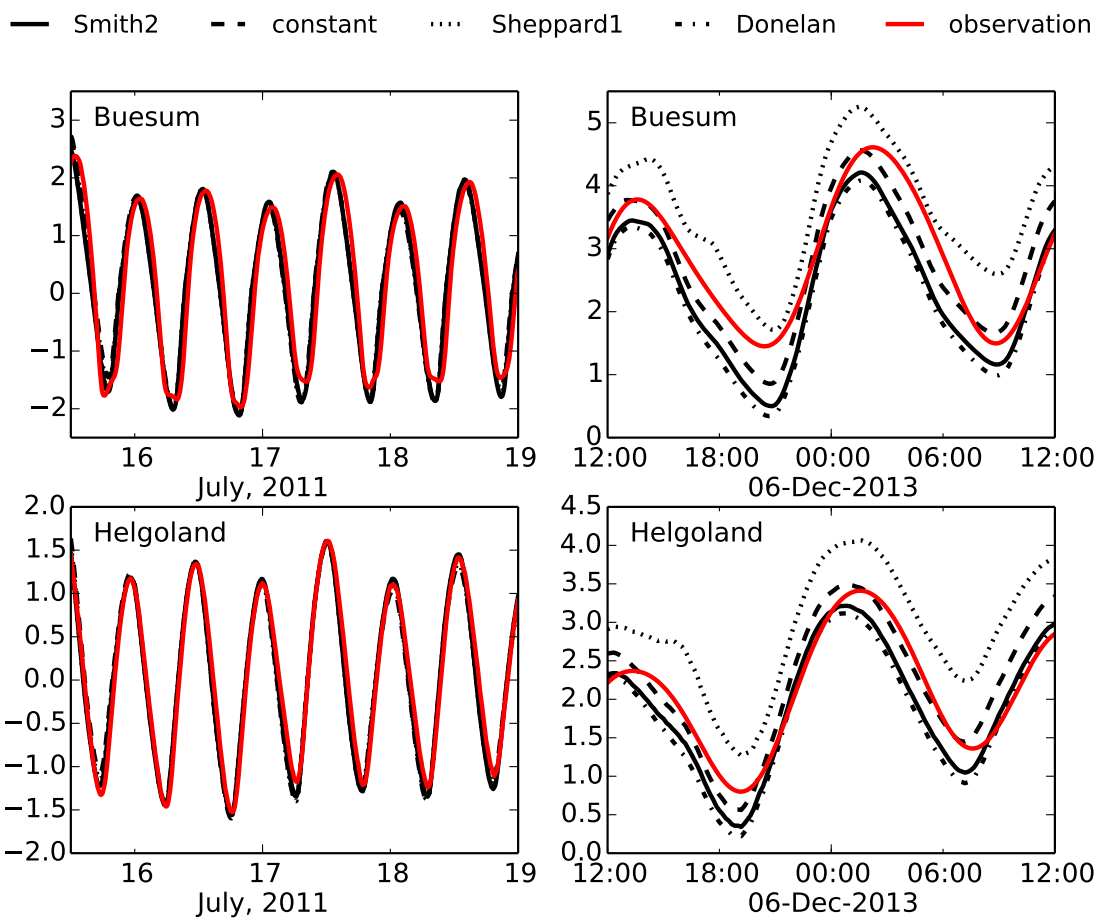

Figure 6.9: Comparison of water level(m) from four models at Helgoland and Buesum

The distributions of C_d derived from Smith formula and Charnock number (Equation (3.12)) at 2011-07-17 00:00 (Figure 6.10) and 2013-12-06 00:00 are shown in Figure 6.11. Charnock number is calculated by an air-wave coupled model by ECMWF, therefore C_d derived from Charnock number is regarded to be more reasonable. Smith formula underestimates C_d in both periods. In mid July 2011, both C_ds are far smaller than the constant value of 0.0025. However, due to the relatively weak wind, modeled water level is not sensitive to C_d. In early December, the Charnock calculated C_d over the German Bight is around 0.0025. Yet Smith formula calculated C_d is around 0.002. Hence, model with constant C_d 0.0025 can provide better results.

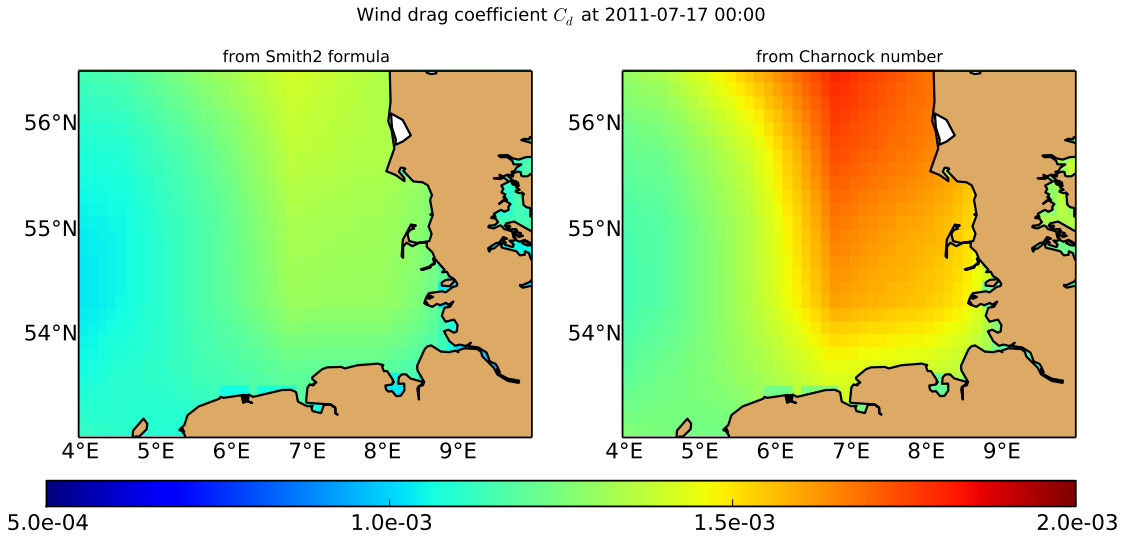

Figure 6.10: Distribution of C_d over German Bight at 2011-07-17 00:00

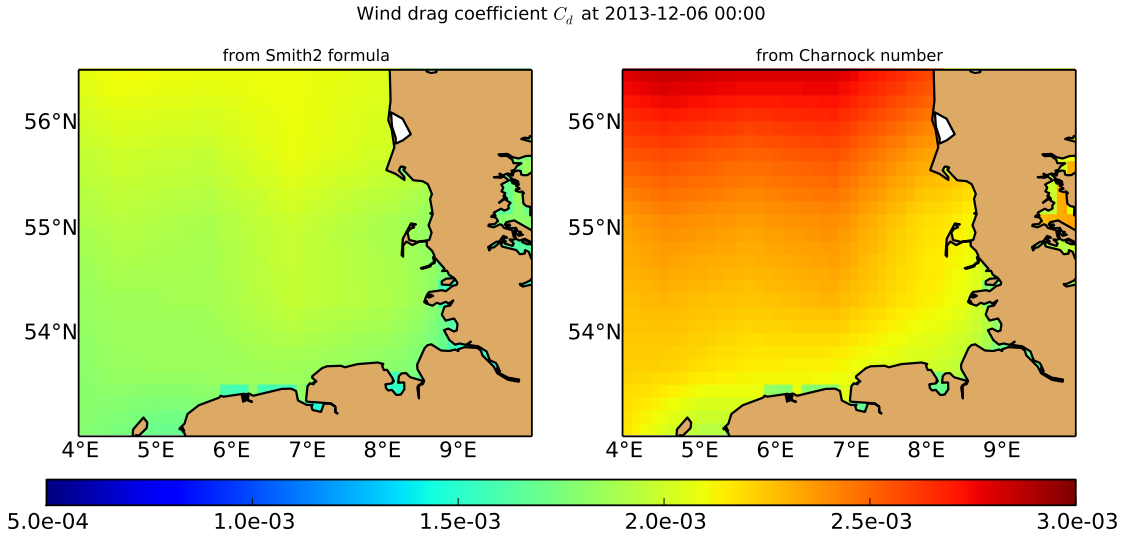

Figure 6.11: Distribution of C_d over German Bight at 2013-12-06 00:00

6.4.3 The energy of surge and total water level

It has been found that under normal weather conditions simulated water level is not sensitive to both wind speed and wind drag coefficient. While in a stormy weather condition, wind stress becomes important for water level. The surge simulation is sensitive to the existing proposed linear function of wind drag coefficient with respect to wind speed. The reason why in period of storm, the errors of wind stress estimate have a considerable effect on storm surge simulations is explained in

the viewpoint of energy.

The potential energy of a water column per unit horizontal area E in a given period between t_1 and t_2 is defined by

$$E = \frac{1}{2}\int_{t_1}^{t_2} \rho_w g \zeta^2 \mathrm{dt}. \tag{6.1}$$

Here $\rho_w = 1025\mathrm{kgm}^{-3}$ is the sea water density, g is gravitational acceleration and ζ is the water level relative to the mean water level. Total energy E_t is calculated with the total water level including the effect of wind and tide, while the surge energy E_s is the calculated only with the surge. A percentage ratio $\alpha = 100\frac{E_s}{E_t}$ is used to indicate the relative importance of wind effect in total water level.

Figure 6.12 shows the spatial distribution of percentage ratio α over the period between 2011-07-15 and 2011-07-19(a) as well as the period between 2013-12-05 12:00 and 2013-12-06 12:00(b). There is very clear difference between the two periods. Under normal weather conditions (Figure 6.12a), the percentage of energy from wind is lower than 10% over most areas in German Bight; there is an amphidromic point of semi-diurnal tides near the northwest corner of the GBM domain, therefore the percentage ratio is large in that area. But in the stormy weather, the percentage ratio becomes much larger; in the deeper water, wind accounts for more than 90% of energy input. Even in the German coastal areas, where the tide range is larger than in the deeper areas, the wind can still account for more than 70% of energy input.

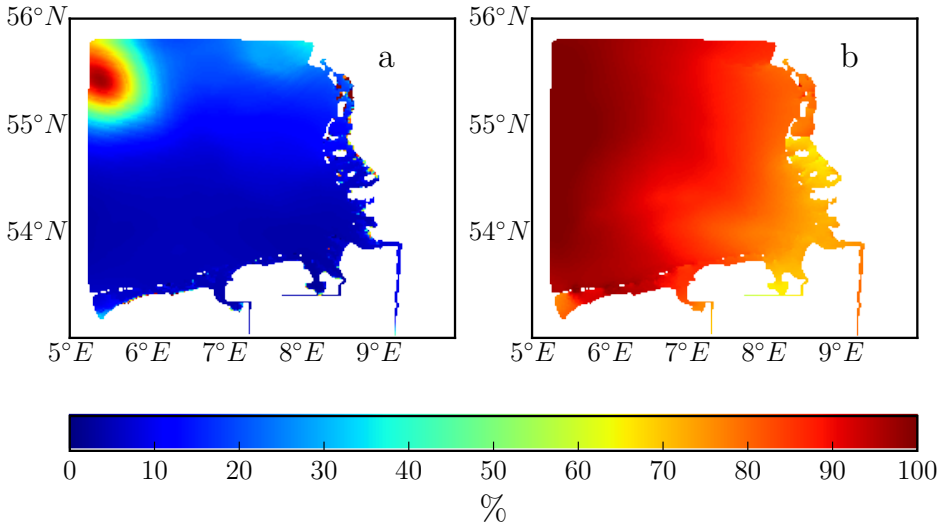

Figure 6.12: The percentage ratio of energy of surge out of total water level in mid-July 2011 (a) and early-December 2013 (b)

6.5 Bottom roughness

In Chapter 3 the tide simulation has been calibrated by adjusting Chezy coefficient or Manning coefficient. One proper Chezy coefficient of 69.88 $m^{1/2}/s$ for both the CSM and the GBM was obtained. Apart from bed form, waves can also affect the bottom shear stress. Especially in shallow water, wave induced orbital velocity can touch the bottom and increase the rate of vertical transfer of horizontal momentum. This is equivalent to an increase of the bottom drag coefficient. Due to the lack of bed form information, bottom drag coefficients become tuning parameters in the range of 2×10^{-3} and 3×10^{-3}. In the following, results of models with different combination of bottom roughness coefficient will be presented. Wind drag coefficient C_d is calculated with Smith formula. The settings of bottom drag coefficients are given in Table 6.3.

Table 6.3: The settings for the sensitivity tests of Chezy coefficient [$m^{1/2}/s$]

test	the CSM	the GBM
1	69.88	69.88
2	69.88	80.0
3	69.88	60.0
4	60.0	69.88

Figure 6.13(a) shows the modeled water level at Buesum in test 1, 2 and 3, in which the Chezy coefficient of the GBM varies around 69.88 $m^{1/2}/s$. The peak water level in test 2, which has the smoothest bottom, is 0.445m larger than that in test 3, indicating the water level in storm surge is sensitive to the bottom shear stress. Figure 6.13(b) demonstrates that the varying Chezy coefficient in the the CSM can also lead to the obvious change of water level. The difference of peak value in test 1 and test 4 is 0.23m, which is not a negligible value.

6.6 Summary

In this chapter, the effects of several processes and model parameters on the storm surge simulation have been tested. It is found that tide is an important factor for surge simulation in shallow waters and should be added in the model. The tidal harmonics from global tide models are imposed on the open boundary

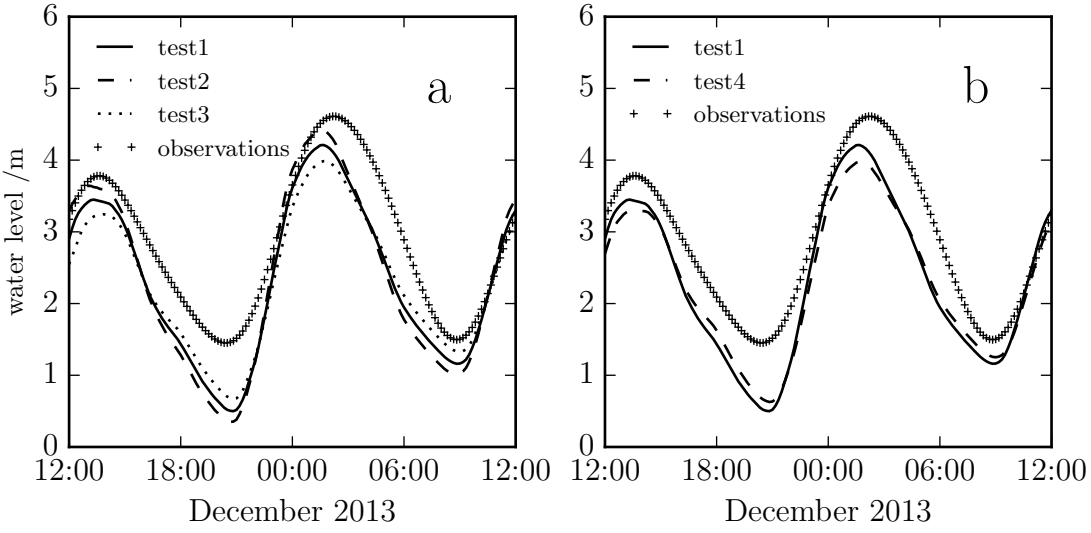

Figure 6.13: Water level(m) at Buesum in the sensitivity tests

of the CSM; tidal harmonics from different sources show insignificant discrepancy, indicating the good accuracy of tidal harmonics along the open boundary of the CSM. As shown in section 6.4.3, wind stress accounts for most energy in the total potential energy of sea water in a strong storm while in normal weather conditions wind stress only accounts for small fraction of total energy. Therefore, the calibration of storm surge model should be focus on the wind stress. In the model Delft3D, wind stress is calculated by providing wind field data and a linear function of C_d with respect to wind speed. Sensitivity tests show that both wind speed and wind drag coefficient C_d could lead to uncertainties of storm surge simulation. Wind field data is obtained from German's national weather service center and agrees well with observed wind while the variation of C_d over sea surface with wind speed is still poorly understood. Hence, it is assumed that C_d account for most of error in the calculation of wind stress. The uniform Chezy coefficient is also not well known because the characteristic of the sea bed is not well investigated; the sensitivity tests show that Chezy coefficient affect the storm surge simulation, especially the peak water level. In the next chapter, data assimilation is used to adjust the wind drag coefficient C_d and Chezy coefficient to improve the storm surge simulation.

Chapter 7

Application of data assimilation for the storm surge model

In Chapter 6, it has been shown that in a 2D storm surge model, water level is sensitive to the wind drag coefficient and the bottom shear stress. Both parameters are not well understood and are exposed to a high degree of uncertainty. Therefore, these parameters are suitable to be the control variables of data assimilation. In the framework of 4Dvar, any model parameter can be a control variable and adjusted to improve model results, but in practice, only those "Suitable parameters" are adjusted.

In this chapter, the data assimilation method proposed in Chapter 4 will be applied to the storm surge model described in Chapter 3. Before the implementation of data assimilation, a wave model is introduced. The wave model results will be used for the calculation of C_d, which will be compared with the C_d value adjusted by the data assimilation. Several tests are performed to study the data assimilation settings, such as model parameter accuracy, control variable and time window. With proper data assimilation settings, the model can be improved significantly. Apart from the storm surge event in the early December 2013, models for another two storm surge events are also effectively improved, showing the applicability of the proposed data assimilation method. Finally, a storm surge nowcasting and forecasting system will be demonstrated.

7.1 Estimate of wind drag coefficient from a wave model

The wind drag coefficient and bottom drag coefficient will be the control variables in the data assimilation. In order to know the validity of these parameters, other information is necessary to be provided to compare the parameters adjusted by data assimilation. It has been mentioned in Chapter 3 that the wind drag coefficient is highly related to the wave state, including wave age, wave height and wave direction. The storm surge model in Chapter 3 is based on Delft3D FLOW without coupling a wave model. In this section, a wave model for the German Bight and the method to estimate wind drag coefficient and bottom shear stress based on wave and wind information will be described.

Many researchers have studied the quantitative relation between wind shear stress and wave state over the sea surface. In Chapter 6, Charnock number α is used to calculate the wind drag coefficient C_d. Charnock number is dependent on the weather and wave conditions and thus the function $C_d = 10^{-3}(a + bU_{10})$ is variable in space and time. It is widely accepted that Charnock number is not constant but well related with wave age(Stewart, 1974). Wave age can be denoted as C_p/U_λ, where C_p is the wave speed at the spectral peak and U_λ is the wind speed at the height of λ m. Donelan (1982) concluded that the roughness z_0 is proportional to the root mean square (RMS) wave height σ. From the data measured at a platform in the North Sea in the program Humidity Exchange Over the Sea (HEXOS) , Smith et al. (1992) found the Charnock number α is inversely proportional to the wave age. Donelan et al. (1993) proposed a regression function between roughness length, significant wave height and wave age as below:

$$z_0/\sigma = 6.7 \times 10^{-4}(U_{10}/C_p)^{2.6} \tag{7.1}$$

Equation (7.1) is based on the Donelan's own data and data from the program HEXOS. z_0 can be calculated given wave information and wind speed. Wave information is simulated from the Delft3D's WAVE module. Wind speed data is also from DWD hourly data introduced in Chapter 3. Once z_0 is calculated, equation (3.10) is used to calculate wind drag coefficient C_d.

The wave model domain as well as model grid and bathymetry are the same as the storm surge model . The nesting sequence consists of the continental shelf model (CSM) and the German Bight model (GBM). The CSM provides open boundary conditions to the GBM. The wave model is validated by comparing the significant

wave height at Sylt (see Figure 1.1) as shown in Figure 7.1. The modeled significant wave height fits well with the observations outside the storm period (The storm happened between 2013-12-05 12:00UTC and 2013-12-06 12:00UTC). The observations within storm period became very noisy, but the model results are still close to the observations, indicating the wave model is able to provide reasonable wave information which is used for the following calculation of C_d.

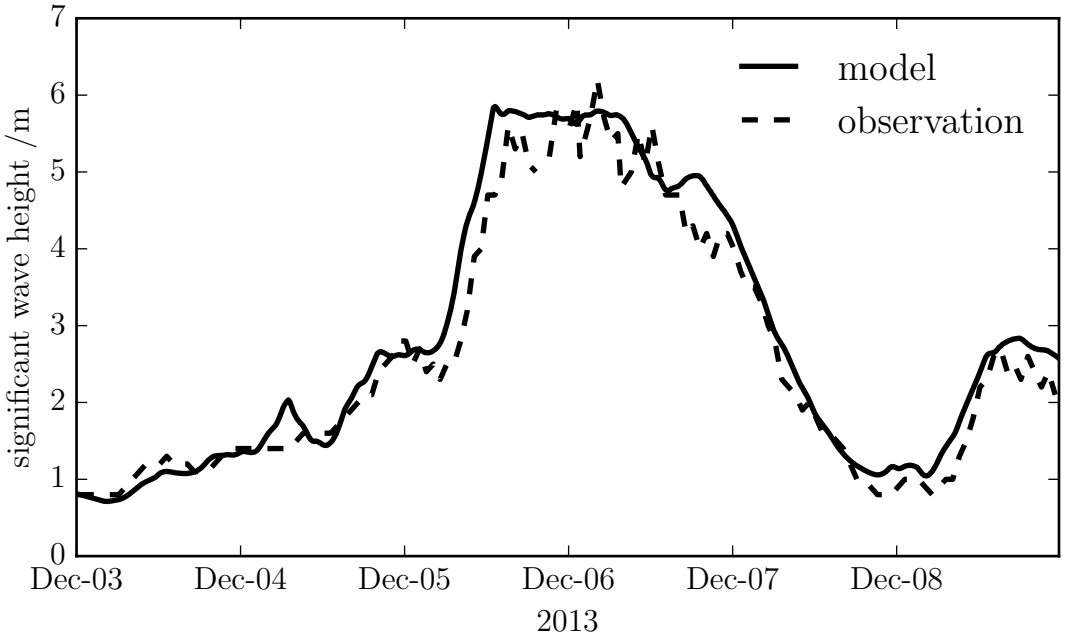

Figure 7.1: Comparison of significant wave height between model and observation at Sylt

According to linear wave equation, C_p in equation (7.1) can be calculated according to the formula:

$$C_p = \sqrt{\frac{gC_pT}{2\pi}}\tanh(\frac{2\pi h}{C_pT}) \tag{7.2}$$

where T is the peak wave period and h is water depth, which is obtained from the wave model. C_p is calculated by solving equation (7.2). Figure 7.2 shows the wave speed with respect to wave period at different water depth. It is seen that the wave speed increases almost linearly with the increase of wave period and then converges to $\sqrt{gh}$ regardless of the wave period, where $\sqrt{gh}$ is wave speed in shallow water.

The wave model can provide the significant wave height $H_{1/3}$. Holthuijsen (2007) proposed a relation between RMS wave height and significant wave height:

$$\sigma = \frac{1}{2}\sqrt{2}H_{1/3} \tag{7.3}$$

Equation 7.3 is applicable if the waves are not too steep and not in a very shallow water.

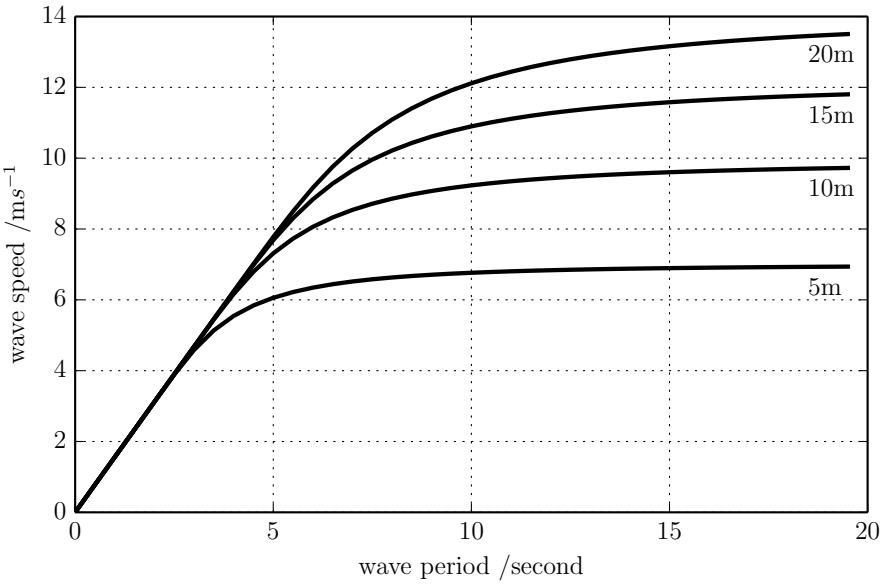

Figure 7.2: Wave speed with respect to wave period at different water depth

7.2 Effects of wave on bottom shear stress

As introduced in Chapter 3, in the Flow module of Delft3D, the bottom shear stress is calculated given water density, bottom drag coefficient and depth-averaged current velocity. Bottom drag coefficient is dependent on the bed sediment type. However, due to the lack of bed sediment information and considerable uncertainties in the measurements of bed sediment, it is very common to specify a spatially uniform bottom drag coefficient. In the module Flow, bottom shear stress is related to current velocity. But in reality, waves also contribute to the bottom shear stress, especially in the shallow water. wave-induced orbital velocity increases the turbulence close to the bottom, which is equivalent to an enhancement of bottom shear stress.

Many authors have worked on the effect of waves on the bottom shear stress in the past decades. Here, an expression by Wu et al. (2010) is used to estimate the total bottom shear stress τ.

$$\tau = \lambda_w \tau_c \tag{7.4}$$

where $\lambda_w \geq 1$ is the shear stress enhancement factor and τ_c is the current related

bottom shear stress. λ_w is given by:

$$\lambda_w = \frac{\sqrt{U^2 + c_w u_w^2}}{U} \tag{7.5}$$

where U is the current velocity; c_w is an empirical coefficient; u_w is the wave orbital velocity amplitude on the bottom layer. Here the value of c_w is equal to 0.65, which is calibrated by the Coastal Model System (CMS), a flow-wave-sediment coupled coastal model. u_w is a direct output from the wave model described in Section 7.1. Therefore, the enhancement factor λ_w can be calculated to investigate the range of bottom shear stress and drag coefficient.

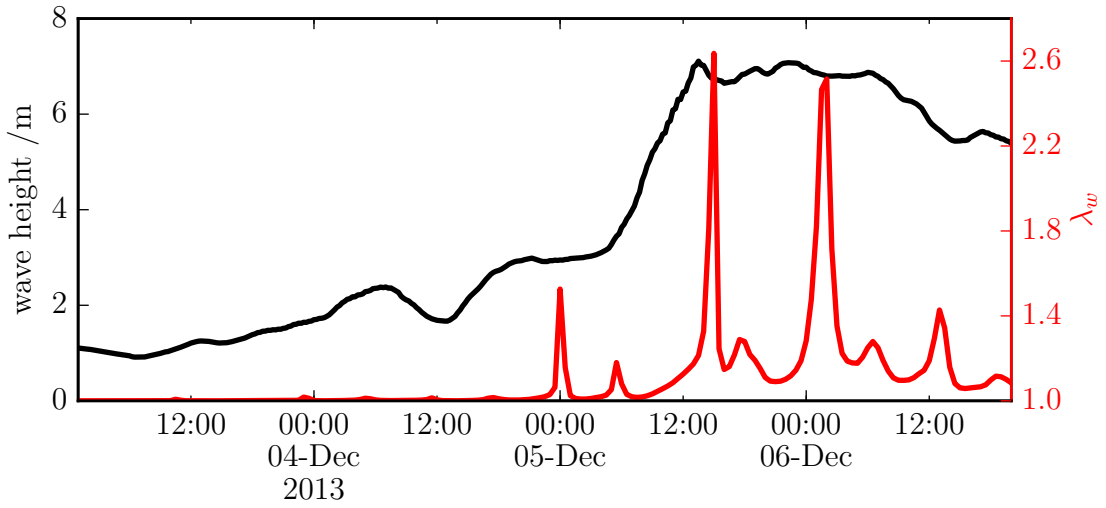

Figure 7.3: Significant wave height (black) and enhancement factor λ_w (red) at Helgoland

Figure 7.3 shows the enhancement factor λ_w at Helgoland in early December, 2013. It can be seen that if the significant wave height is less than 2.5m, λ_w is smaller than 1.02; λ_w increases up to 2.65 in the period of storm when the significant wave height is about 7m. The current velocity mainly varies with tide, which can be represented by an ellipse, therefore, λ_w ranges between 1.15 to 2.65 when the significant wave height is more than 6m. Bottom drag coefficient can be decomposed into bed-sediment-related drag coefficient and wave-induced drag coefficient. It is a reasonable assumption that the bed-sediment-related drag coefficient is constant in a normal sea condition. But there may be a distinct change of wave-induced drag coefficient if the wave height becomes larger as shown in Figure 7.3.

In the storm surge model of this study, bottom drag coefficient is represented by Chezy coefficient (Equation (3.18)). When the wave effect is taken into account,

the range of Chezy coefficient variation is estimated as follows:

$$\tau = \frac{\rho_0 g \vec{U}|\vec{U}|}{C_b^2} = \lambda_w \tau_c = \lambda_w \frac{\rho_0 g \vec{U}|\vec{U}|}{C_s^2} \tag{7.6}$$

where C_b is the Chezy coefficient corresponding to the total bottom drag coefficient and C_s is the Chezy coefficient corresponding to the bed-sediment-related drag coefficient. Their relation can be expressed:

$$C_b = \frac{C_s}{\sqrt{\lambda_w}} \tag{7.7}$$

The Chezy coefficient C_s is set to 69.88 $m^{1/2}/s$ from a preliminary calibration of tide simulation in Section 3.3. This value will be viewed as the bed-sediment-related Chezy coefficient. The range of the total Chezy coefficient approximately varies between 43 and 65 given the range of λ_w in the storm surge in early December of 2013 at Helgoland.

7.3 Data assimilation of the model under normal weather conditions (Mid July 2011)

It is seen that under normal weather conditions, the wind drag coefficient plays very small roles in the water level simulation. Similarly, the wave-induced bottom drag coefficient is very small under normal weather condition. Therefore the wave-induced drag coefficient can be negligible. In this section, data assimilation is used to adjust the Chezy coefficient in the GBM. The first guess value of Chezy coefficient is 69.88 $m^{1/2}/s$, which is calibrated in section 3.3. The model simulation time is mid July 2011. In this period the wave height is among the smallest (see Figure 7.4) in the year 2011. The adjusted Chezy coefficient can be regarded as the optimal value for the bed-sediment-related Chezy coefficient.

Several data assimilation tests are carried out with different time windows (see first column of Table 7.1). Water level data at the tidal gauges of Helgoland and Cuxhaven are used for data assimilation. The cost function values before and after data assimilation are shown in the Figure 7.5. The percentages above the bars denote the relative decrease of cost function. The updated Chezy coefficient of the GBM are in Table 7.1. The optimal Chezy coefficients in different time windows are also different, ranging between 70.81 and 76.28. The averaged RMSDs at stations Buesum, Cuxhaven, Helgoland, Mittelgrund and Husum over time window in each test before and after data assimilation are also shown in Table 7.1. Generally speak-

ing, data assimilation improves model skill in all the tests in terms of RMSD. But the improvements are not impressive, especially in test 1, 5 and 6. It is well known that Chezy coefficient affects the tidal range and the RMSD due to inaccurate Chezy coefficient normally gets the peak when the water level is at high water or low water. RMSDs in a short period may not reflect the general errors resulting from Chezy coefficient. In test 1-6, the updated Chezy coefficients are different, but they tend to be closer to their first guess value 69.88 $m^{1/2}/s$. Recall that the value 69.88 $m^{1/2}/s$ is calibrated by comparison of tidal harmonic constituents, which are obtained by three-month data. It implies that data assimilation on Chezy coefficient with time windows of several days is not an effective way to improve model. Therefore, the value 69.88 $m^{1/2}/s$ is still be used as the bed-sediment-related Chezy coefficient.

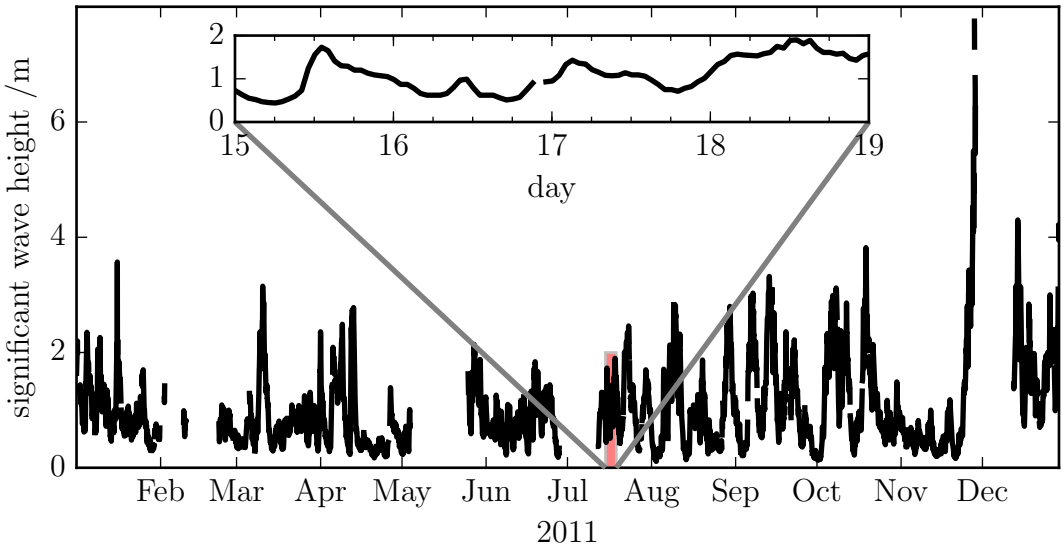

Figure 7.4: Significant wave height at Sylt in 2011

Table 7.1: Data assimilation results for the simulation in Mid July 2011

	time window (mmddHH)	updated Chezy coefficient	RMSD (m) before DA	RMSD (m) after DA
test 1	071512∼071515	70.81	0.111	0.115
test 2	071512∼071518	75.50	0.481	0.392
test 3	071512∼071600	74.85	0.380	0.323
test 4	071512∼071612	76.28	0.414	0.351
test 5	071512∼071712	73.23	0.392	0.362
test 6	071512∼071812	71.68	0.370	0.355

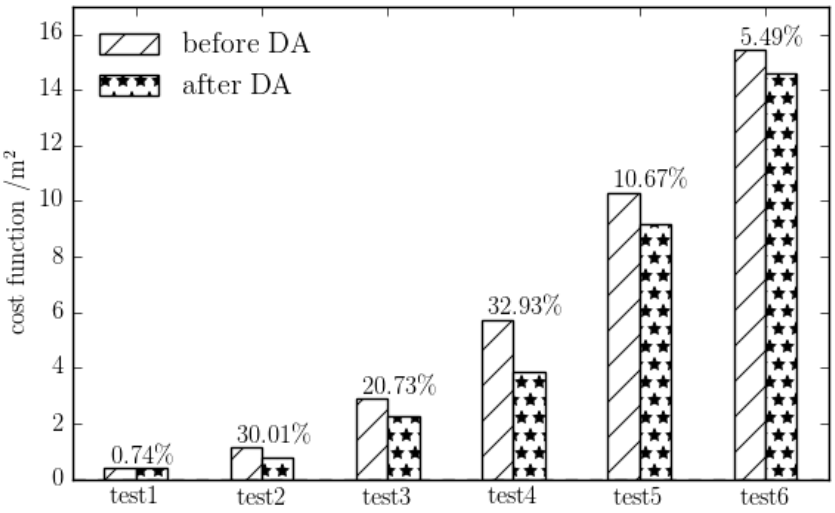

Figure 7.5: The cost function values before and after data assimilation

7.4 Data assimilation of storm surge model in early December 2013

As described in Section 3.4, the storm surge in early December 2013 over the North Sea is the strongest storm in this area since 1953. Severe surge happened from 2013-12-05 12:00UTC to 2013-12-06 12:00UTC. Several tests are performed before implementing the data assimilation into the storm surge model. These tests will investigate the effect of parameter accuracy, control variables, number of observations and time window length on data assimilation.

7.4.1 Effect of the parameter accuracy on data assimilation

In equation (4.9), W_p and Q represent the weight of the terms corresponding to model parameters and observations in the cost function. Those parameters with less uncertainty have higher weight. The observations used for the data assimilation are measured by tidal gauges, the accuracy of which is usually very high. According to Woodworth and Smith (2003), the accuracy of a new radar tidal gauge and a conventional bubbler pressure gauge are comparable and in the order of 0.01m. The water level data used in this paper are resampled every ten minutes. Resampling could decrease the accuracy of the data. Therefore, it is assumed that all the observations of water level have an accuracy of 0.02m.

It is more difficult to specify the accuracy of a, b and Chezy coefficient. a and b are derived from measurements of wind stress and wind speed. Their accuracy

depends not only on the measurement errors but also on the understanding of the air-sea interaction. As mentioned in Section 3.1.2, it is difficult to measure wind shear stress in a storm. Therefore, large errors are expected when applying a and b to calculate wind drag coefficient. The poor understanding of air-sea interaction also brings more uncertainties. Chezy coefficient is given a uniform value over the model domain according to the previous studies, making it difficult to know its accuracy. A sensitivity test is performed to investigate the effect of the accuracy of a, b and Chezy coefficient on the data assimilation skill. Tests with different accuracy are listed in Table 7.2. a, b and Chezy coefficient are adjusted only for the GBM. The accuracy for each parameter in test1 has the same order as the first guess; from test2 to test4, the accuracy is increasing by ten times than the previous. In these four tests, the observations of water level at Buesum and Helgoland are used for data assimilation, Husum and Cuxhaven for validation.

Table 7.2: Parameter accuracy in the test 1-4

	a	b	Chezy coefficient
test1	0.6	0.06	70
test2	0.6×10^{-1}	0.06×10^{-1}	70×10^{-1}
test3	0.6×10^{-2}	0.06×10^{-2}	70×10^{-2}
test4	0.6×10^{-3}	0.06×10^{-3}	70×10^{-3}

Table 7.3 shows the RMSDs of water level in the four tests listed in Table 7.2. The column "ref" contains the RMSDs before data assimilation. In test1, test2 and test3, data assimilation decreases the RMSDs significantly. While in test4, the RMSDs after data assimilation just decrease very slightly due to the high accuracy of a, b and Chezy coefficient. In test2, the accuracy is ten times smaller than that in test1, but the RMSD is almost the same after data assimilation; this means that accuracy of observations is higher than parameters in test2. In test3, the higher accuracy of parameter begins to play a role, making the RMSD obviously larger than that in test1 and test2. According to the discussion in Chapter 6, the accuracy of the three parameters are not high and should be lower than 10% of their first guess. Therefore, for the data assimilation of storm surge model, the terms of first guess in the cost function 4.9 have small weight.

Table 7.3 also shows a, b and Chezy coefficient adjusted by data assimilation. The adjusted a and b seem quite different in the four tests. The calculated C_d from $10^3 C_d = a + b|U_{10}|$ are used to show the effect of parameter accuracy on data assimilation in Figure 7.6. The grey patch in Figure 7.6 indicates the wind speed range in the storm. Within the grey patch, C_d in test1, test2 and test3 are closer

Table 7.3: RMSD of water level (m) and parameter values in the four tests listed in Table 7.2

	ref	test1	test2	test3	test4
Buesum	0.554	0.204	0.207	0.226	0.515
Helgoland	0.283	0.142	0.140	0.141	0.263
Husum	0.589	0.196	0.196	0.237	0.546
Cuxhaven	0.526	0.186	0.190	0.229	0.492
a	0.61	1.50	0.95	0.74	0.62
b	0.063	0.071	0.095	0.094	0.064
Chezy coefficient	69.88	62.46	62.81	63.05	69.01

to each other than the reference C_d, while C_d in test4 is almost the same as the reference C_d. This implies that data assimilation adjusts a and b regardless of the physical significance in a large range of wind speed, but focusing on adjusting C_d in a narrow range of wind speed in the time window. Chezy coefficient is adjusted in a similar way to a and b. In test 1, 2 and 3, the adjusted values of Chezy coefficient are around 63.00.

In summary, the accuracy of a, b and Chezy coefficient can be specified to 1/10 of the first guess of each parameter. Such roughly estimated accuracy not only fits to the estimate of uncertainty but also makes sure that the model results are closer to observations after data assimilation, because observations are regarded to be highly accurate.

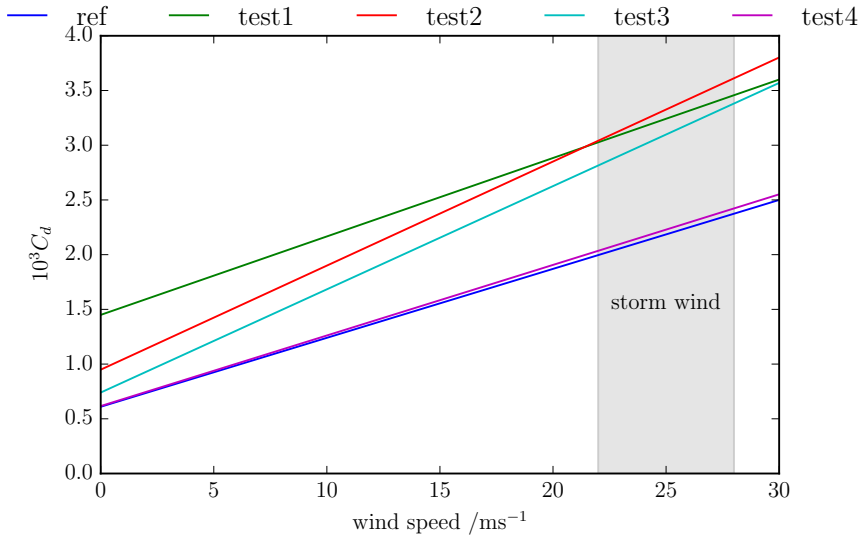

Figure 7.6: Linear function of C_d with respect to wind speed before and after data assimilation

7.4.2 Effect of the control variables choice on data assimilation

In order to test the effect of choosing control variables on the data assimilation, six tests are designed as shown in Table 7.4. The data assimilation settings in these tests are same in the six tests except the control variables. The observed water level used in these tests are from Buesum and Helgoland; the accuracy of water level is a constant value 0.02m. Observations at Cuxhaven and Husum are used for validation. The time window is the period between 2013-12-05 12:00UTC and 2013-12-06 12:00UTC. The accuracy of a and b in the wind drag coefficient formula is 0.6 and 0.06; the accuracy of Chezy coefficient is 7.0.

Table 7.4: Data assimilation settings for different control variables

	control variables
test1	GBM bottom
test2	GBM wind drag
test3	CSM wind drag
test4	GBM bottom, GBM wind drag
test5	CSM bottom, CSM wind drag
test6	CSM bottom, CSM wind drag, GBM bottom, GBM wind drag

The values of the cost functions in the six tests are illustrated in Figure 7.7. In all the six tests, the values of cost function are decreased after data assimilation. In test 1, in which only the Chezy coefficient of the GBM is the control variable, the cost function decreases from 72.53 to 68.12. In other tests, the decrease of the cost function is much larger, indicating that wind drag coefficient C_d has more uncertainties and it is more effective to adjust it. In test 2, the cost function converges to 15.36. In test 3 it converges to 41.86 and in test 5 the converged value is 32.67, indicating only adjusting parameters of the CSM is not a good choice. In test 4, both Chezy coefficient and C_d of the GBM are adjusted; the cost function levels off on the value of 13.46. In test 6, Chezy coefficient and C_d of both the CSM and GBM are adjusted and the cost function value converge to 12.91. It is concluded from these tests that adjusting more model parameters can better decrease the model errors but also requires more computation resources. From the viewpoint of effectiveness and efficiency, test 4 is the best among all the control variables choices listed in Table 7.4.

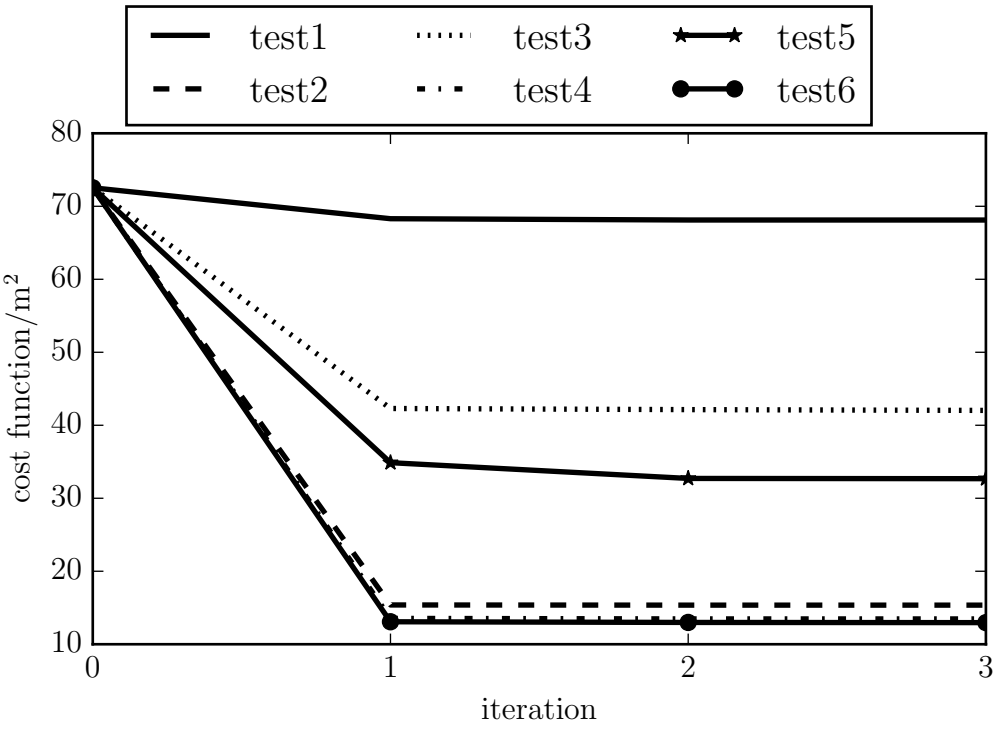

Figure 7.7: Cost function in the tests listed in Table 7.4

Figure 7.8 shows the time series of water level in the time window. "ref" denotes the modeled water level before data assimilation. It is found that at all the four stations the modeled peak water level are improved significantly if C_d of the GBM is adjusted (in test 2, 4 and 6). The improvements can be seen both at data assimilation stations and validation stations. Quantitative assessments of the improvements by the data assimilation are displayed in Table 7.5. In test 1, where only the Chezy coefficient of the GBM is adjusted, the improvements of water level at Buesum and Helgoland after data assimilation are very slight; At Husum and Cuxhaven, the RMSDs even become larger than before data assimilation. There are evident improvements in the tests 2, 4 and 6. The relative improvement at all the three stations is more than 50%. In test 4, RMSDs at all stations are smaller than that in test 2, indicating the inclusion of Chezy coefficient in the control variables are helpful to reduce the model errors. However, test 3 does not perform better than test 2 even though three parameters of the CSM are in the control variables. Performance of test 4 is comparable to test 6 in terms of RMSDs, but the number of control variables is much less than in test 6. Therefore, among the six tests described in Table 7.4, test 4 is the most practical scheme.

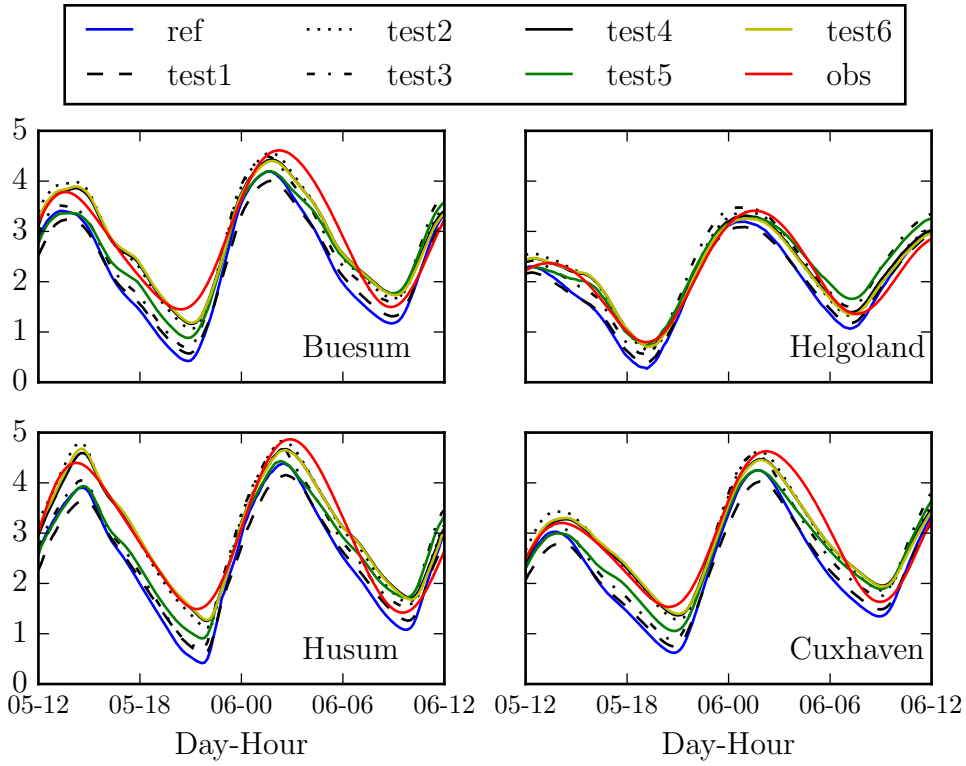

Figure 7.8: Time series of water level before (blue) and after (black, green and yellow) data assimilation and the observations (red) in the time window from 2013-12-05 12:00UTC to 2013-12-06 12:00UTC at four tidal gauges

Table 7.5: RMSDs of modeled water level (m) before and after data assimilation with different control variables

	ref	test1	test2	test3	test4	test5	test6
Buesum	0.554	0.552	0.230	0.389	0.208	0.368	0.213
Helgoland	0.283	0.267	0.155	0.235	0.139	0.184	0.129
Husum	0.589	0.591	0.192	0.441	0.191	0.439	0.186
Cuxhaven	0.526	0.542	0.203	0.357	0.189	0.349	0.197

In the twin experiments in Chapter 5, one conclusion is that if the control variables account for most of model errors, the adjusted control variables can be close to the true value and hence fall into their reasonable range. Otherwise, there is a risk that the adjusted model parameters are out of the valid range even if the model results at some stations can be improved. Table 7.6 shows the adjusted model parameters after data assimilation in the six tests. Although the adjusted Chezy coefficients vary from test to test, all of them are smaller than the first guess 69.88 $m^{1/2}/s$, which is derived from a model in normal weather condition. It is mentioned

in Section 7.2 that strong waves increase the bottom shear stress and that is equivalence of decreasing Chezy coefficient. Thus, the Chezy coefficients are reasonably adjusted in test 1, 4 and 6. As mentioned previously, wind drag coefficient C_d is the function of a and b. It can be found from Table 7.6 that the adjusted a and b for the GBM are very close to each other, which can be confirmed in Figure 7.9 (right). However, the adjusted a and b for the CSM present more discrepancies; in test 3 and 5, C_d is evidently larger than that in Smith formula ($C_d = 10^{-3}(0.61 + 0.063U_{10})$) if the wind is stronger than 15m/s; while in test 6, the function of C_d is very close to Smith formula.

Table 7.6: Model parameters before and after data assimilation in the tests for control variables choice

	before DA	test1	test2	test3	test4	test5	test6
GBM bottom	69.88	61.47	*	*	63.85	*	65.85
GBM wind a	0.610	*	0.628	*	0.706	*	0.772
GBM wind b	0.063	*	0.112	*	0.108	*	0.104
CSM bottom	69.88	*	*	*	*	57.52	67.12
CSM wind a	0.61	*	*	0.284	*	0.515	0.654
CSM wind b	0.063	*	*	0.101	*	0.097	0.056

Wind drag coefficient C_d is also calculated from Charnock number and the Donelan method described in Section 7.1. They will be compared with the DA-adjusted (adjusted by data assimilation) C_d. Figure 7.10 shows the spatial distribution of C_d at 2013-12-06 00:00 computed from different methods in the domain of the GBM. Cooler color denotes a lower value of C_d. Generally speaking, C_d obtained from Smith formula has the smallest value at 2013-12-06 00:00 over the German Bight. C_d obtained from Charnock number is larger, and from Donelan's method the largest; especially in the coastal areas, C_d from Donelan's method can be as high as 0.004. The differences result from whether and how the wave effect is considered for the C_d calculation. Smith formula is fixed regardless of wave state. It is a special case when the Charnock number is equal to 0.119. However, as Charnock number from ECMWF demonstrated, Charnock number is not constant but varying spatially and temporally. Therefore, C_d calculated from variable Charnock number should be more reasonable. Donelan's method takes wave age into account, making it the most reasonable for the C_d estimation among the three method. It is highly probable that the true value of C_d is closer to the C_d obtained by Donelan's method than the other two methods. The DA-adjusted C_d is shown in Figure 7.10 (lower right); it is closer to C_d computed with the Donelan's method at 2013-12-06 00:00.

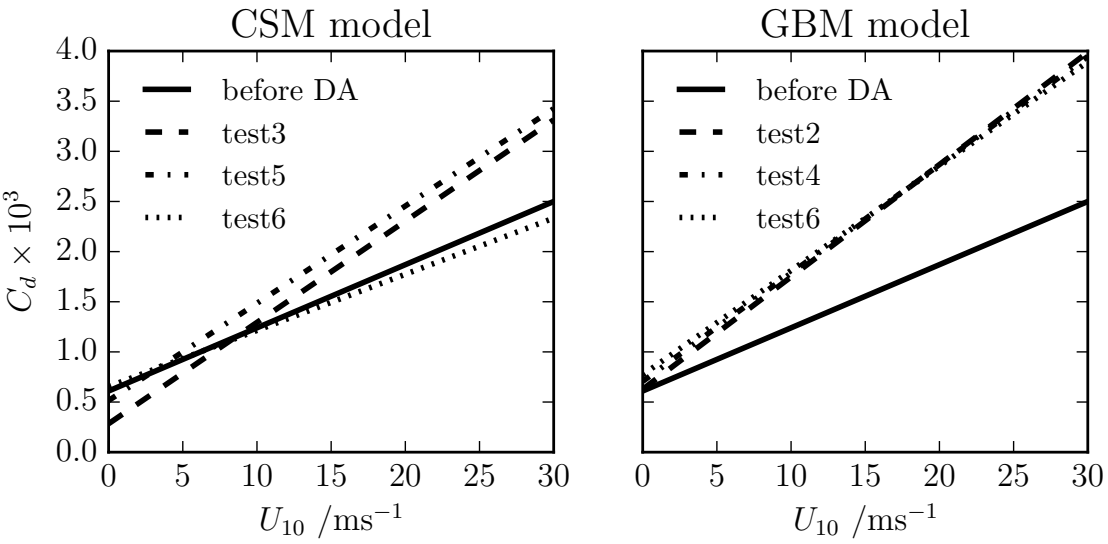

Figure 7.9: Function of C_d in the CSM and GBM

Comparisons of time series of C_d at some locations are also carried out. The locations are labeled as A, B, C, D, E and F shown in Figure 7.10 (upper left), roughly covering the GBM domain from offshore to the near-coast areas. The comparisons are illustrated in Figure 7.11. At all the six locations, data assimilation can adjust C_d towards the values computed by Donelan method. At location A and C, the DA-adjusted C_d are almost the same as the Donelan-method-computed C_d. At locations B, D and F, DA-adjusted C_d are still much smaller than the Donelan-method-computed C_d. There are also obvious differences between DA-adjusted C_d and Donelan-method-computed C_d. In shallow water, the wave height tends to be higher and the wave speed tends to be smaller. Both of them can result in larger roughness length z_0 and thus enhance the wind drag coefficient. In Delft3D, the function of C_d is uniform over the model domain. Therefore, C_d computed by the Delft3D (linear function with respect to wind speed) tends to be more uniform than Donelan-method-computed C_d. This is why there are larger differences in the near-shore and off-shore areas.

As discussed in the twin experiments in Chapter 5, control variables in 4Dvar will be adjusted to be close to their true values if the control variables are the main error sources of model. Here since the observed water level data has high accuracy and wind drag coefficient has considerable uncertainties in a storm, the DA-adjusted wind drag coefficient is directed to be closer to the true value, as confirmed by the comparison between DA-adjusted C_d and Donelan-method-computed C_d.

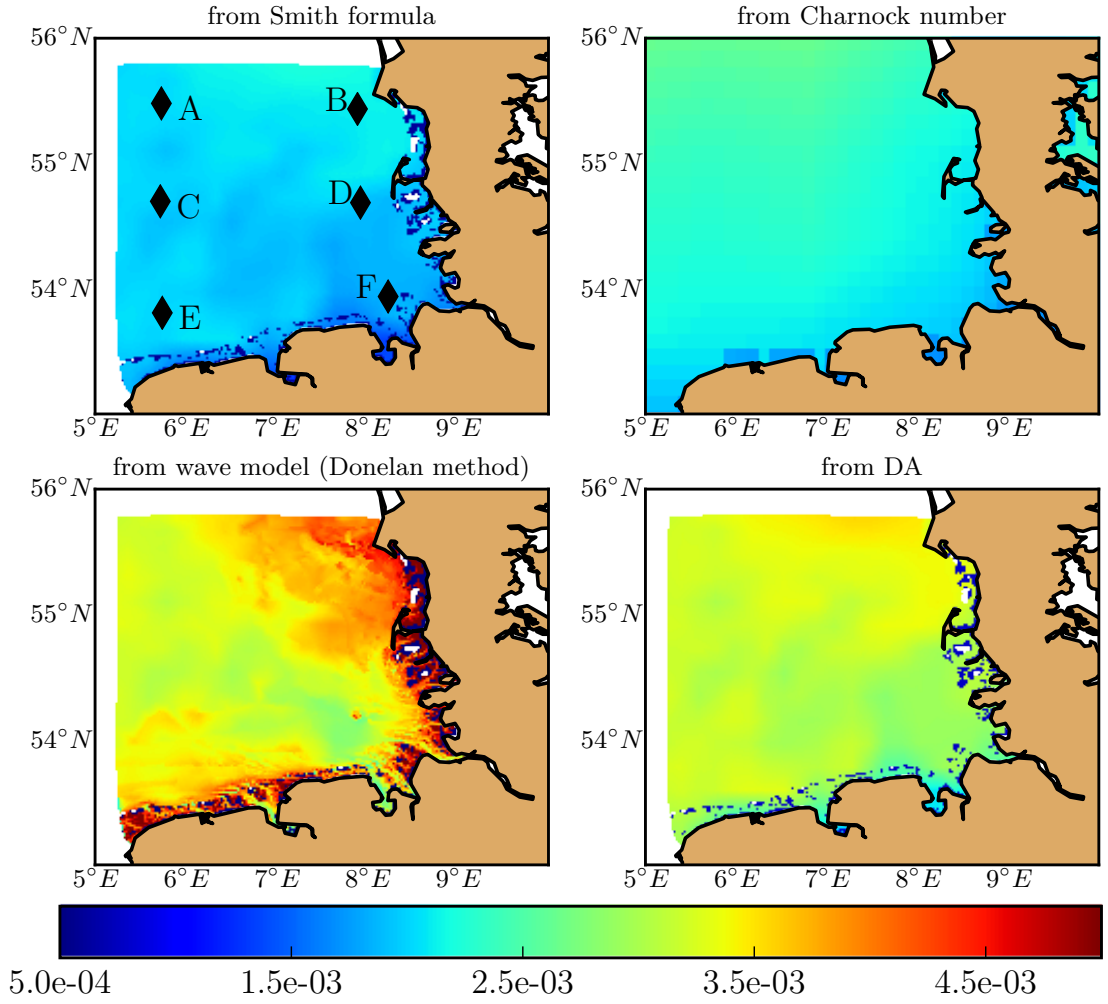

Figure 7.10: Distribution of C_d at 2013-12-06 00:00 in German Bight

The distribution of C_d in the CSM domain at 2013-12-06 00:00 is shown in Figure 7.12. At this moment, Smith formula also underestimate C_d compared with Charnock number method and Donelan's method. DA-adjusted C_d in test 3 and 5 are larger than that obtained by Smith formula. However, DA-adjusted C_d of the CSM in test 6, in which C_d of both the CSM and the GBM are in the control variables, are almost the same as the Smith formula. In other words, if C_d of the CSM and C_d of the GBM are control variables at the same time, the increment of C_d of the CSM is much smaller than C_d of the GBM. Therefore, it is more practical to only adjust C_d of the GBM. A new test (test7) is added to find whether a can be get rid of from control variables list. The adjusted b and Chezy coefficient are 0.109 and 63.09 respectively, almost the same as the adjusted b and Chezy coefficient in test4. Therefore, in the following tests, the control variables are b and Chezy coefficient of the GBM.

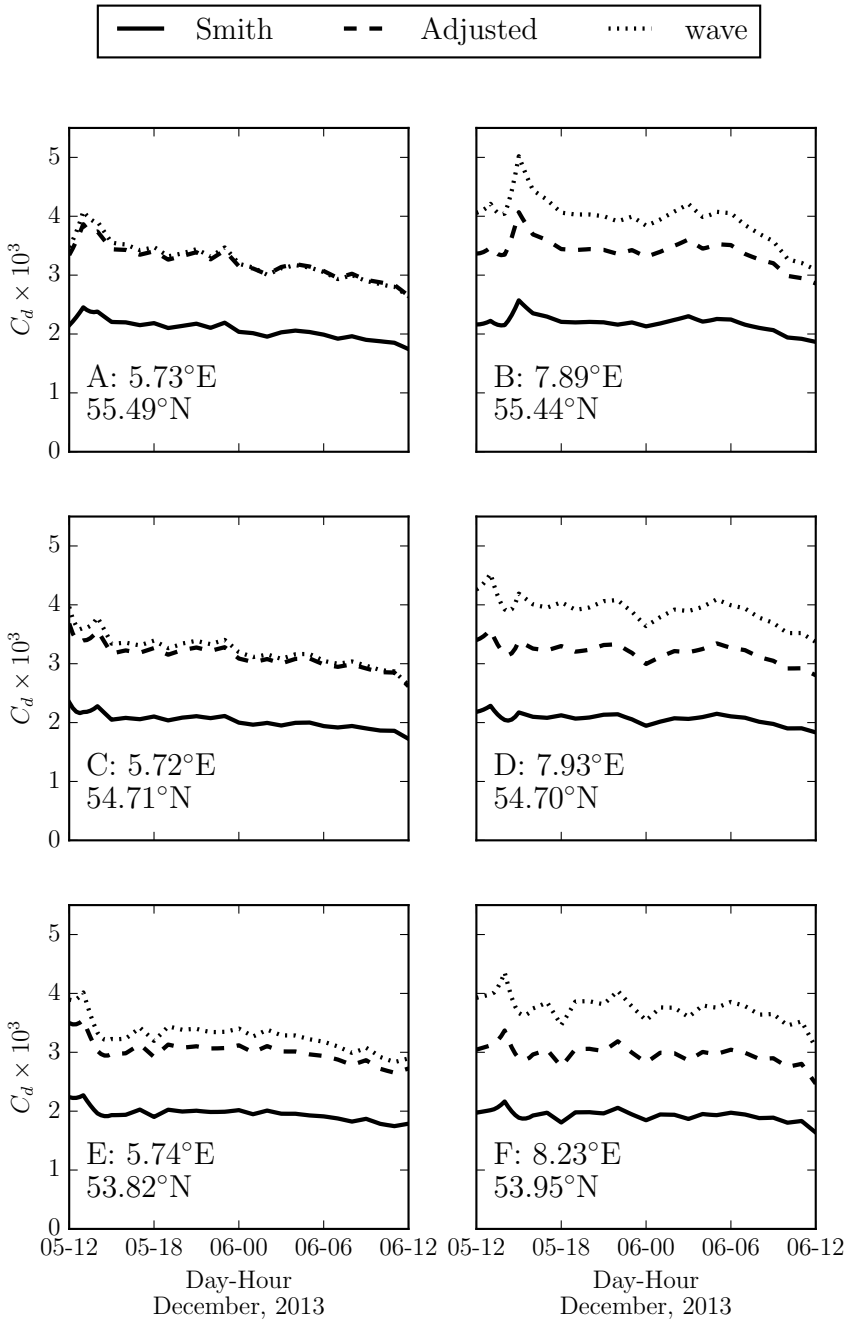

Figure 7.11: Comparison of C_d of the GBM over the time window obtained from different methods

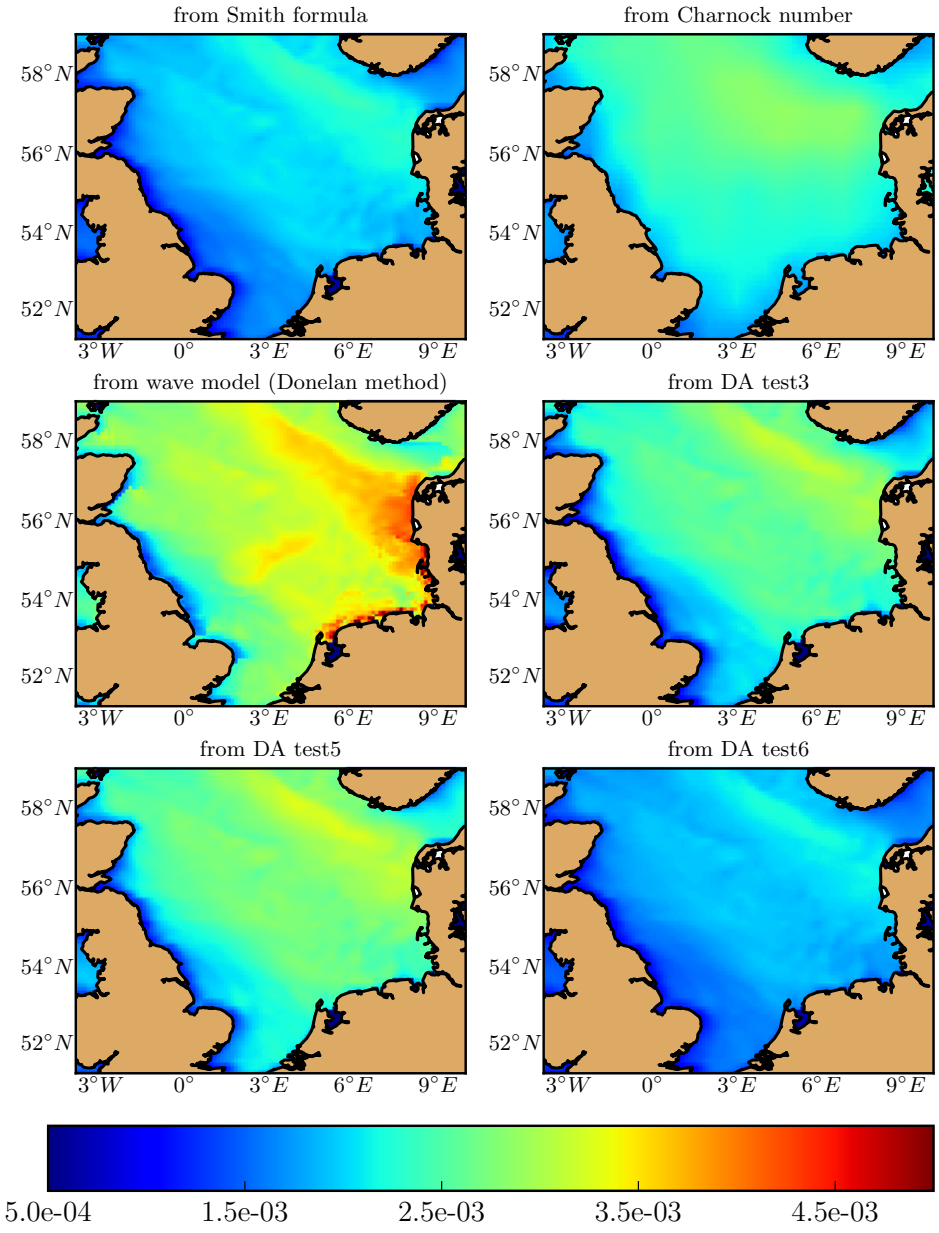

Figure 7.12: Distribution of C_d at 2013-12-06 00:00 in the domain of CSM

7.4.3 Effect of the number of observations

It is not easy to measure the state of ocean, therefore, the number of observations is a concern in ocean study. It is always attractive to use small number of observations to obtain comparably good parameter estimate. In this section some single-station tests for data assimilation of C_d and Chezy coefficient of the GBM will be shown. The settings of model and data assimilation are the same as in the test 4 in Section 7.4.2 except that the observation stations are replaced by one sin-

gle station. Six stations on the coastline of the German Bight are used in six tests respectively (test 2-7). See Table 7.7 for more details.

Table 7.7: Settings for the single-station tests

	station
test 1 (ref)	Helgoland, Buesum
test 2	Buesum
test 3	Helgoland
test 4	Cuxhaven
test 5	Husum
test 6	Wilhelmshaven
test 7	Delfzijl

Figure 7.13 shows how the cost function decreases in the seven tests. All cost functions show sharp decreases. Figure 7.13 also shows the relative decrease in percentage, quantifying the effectiveness of the data assimilation. The initial values of cost function in the 7 tests are different, indicating the RMSDs of model results are not distributed uniformly. RMSDs at Helgoland is obviously smaller than those at the other five stations. The relative improvement at Helgoland are also smaller. Relative improvements signify roles that C_d and Chezy coefficient play in the model at a certain station. Low relative improvements mean that other parameters or factors play relatively larger roles for the accuracy of model, such as poorly understood model processes or model grid resolution. These factors can not be fixed by adjusting C_d and Chezy coefficient using 4DVar.

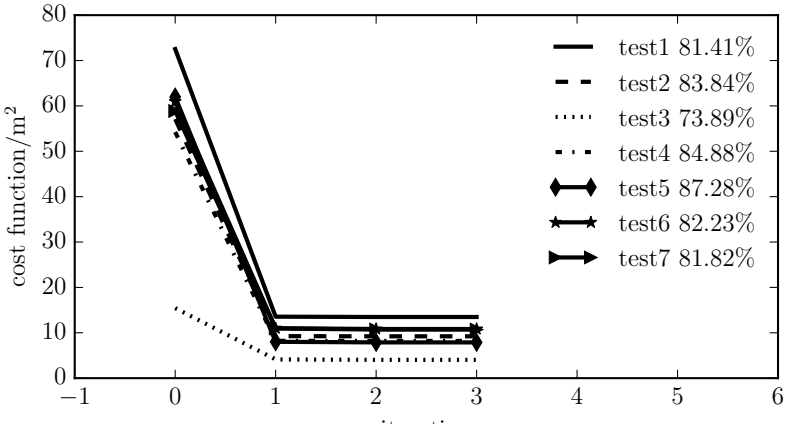

Figure 7.13: Variations of cost functions in the data assimilation tests listed in Table 7.7

Table 7.8 lists the adjusted b and Chezy coefficient in the 7 tests. It can be found that the adjusted a and b are similar in the first six tests, in which the stations used

are on the east coast of the German Bight. In test 7, the station Delfzijl is used, the adjusted b is 0.143, much larger than that in other tests. The adjusted Chezy coefficient has a similar feature, that is, the adjusted Chezy coefficient is the smallest among all the tests. One possible reason is that station Delfzijl is located in the south coast of the German Bight. As mentioned above, both wind drag coefficient and Chezy coefficient are not uniform spatially in reality. These two parameters at Delfzijl are much different from other stations. In the model Delft3D, the function of C_d and Chezy coefficient are uniform, therefore one value cannot be the optimal for all the stations. Such limitation can be relaxed by using the spatially varying Chezy coefficient. But insufficient information of bed sediment over model domains makes it unpractical for storm surge models.

Table 7.8: Adjusted b and Chezy coefficient of the GBM in the single-station tests

	wind b	Chezy coefficient
test 1	0.111	63.88
test 2	0.113	63.88
test 3	0.105	63.50
test 4	0.113	65.73
test 5	0.111	66.80
test 6	0.119	60.94
test 7	0.143	59.16

Figure 7.14 shows the RMSD of water level before and after data assimilation in the tests shown in Table 7.7. It can be found that at stations Buesum, Helgoland, Husum, Cuxhaven and Wilhelmshaven, the reductions of RMSD have similar pattern. Any station among the five used for data assimilation (the first six tests) has similar effect on the RMSD reduction at other four stations. This also implies that water level in the east coast of German Bight can be improved effectively by using just one station in this area. However, there is less improvement at Delfzijl in the first six tests. In test 7, things are opposite. There is more improvement at Delfzijl but less improvement at Buesum, Helgoland, Husum, Cuxhaven and Wilhelmshaven.

Table 7.8 and Figure 7.14 point to a conclusion that the linear relation between C_d and wind speed is location-dependent due to the spatially varying wave state distribution. The stations used for data assimilation should be located in the study areas of interest. For the coastal areas of the German Bight, any station or stations in the first six tests is suitable for the data assimilation and will not result in much difference.

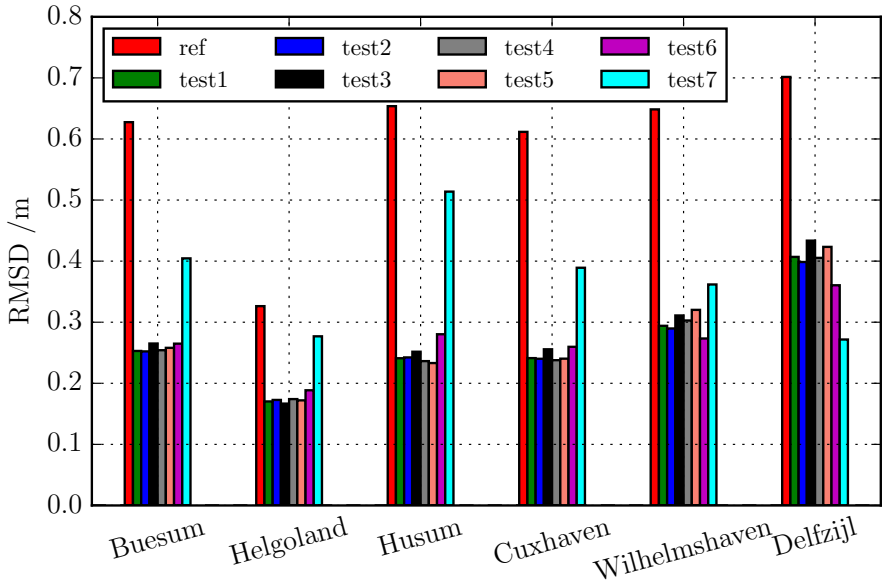

Figure 7.14: RMSE of water level before and after data assimilation in different tests. "ref" is the RMSD before data assimilation

7.4.4 Effect of the time window

The effects of control variables choice and number of observation stations have been shown in the sections above. In these tests, the time window is the period between 2013-12-05 12:00 and 2013-12-06 12:00, meaning that the adjusted control variables are optimal over 24 hours in terms of the minimum of cost function. Both C_d and Chezy coefficient are dependent of wave state. In a storm, the wave state may experience fast change. Therefore, time window of 24 hours may be too long to capture the fast changing features of C_d and Chezy coefficient, which have important effect on the water level simulation. As a result, the improvement may be not on the whole period, as shown in Figure 7.8. Several data assimilation tests with different time window periods are shown in this section. The control variables include b and Chezy coefficient of the GBM. The stations used for data assimilation are Buesum and Helgoland. The settings are indicated in Table 7.9. In these tests the study period is also between 2013-12-05 12:00 and 2013-12-06 12:00.

Figures from 7.15 to 7.17 show the RMSD of modeled water level after data assimilation with different time windows.

It can be found from Figure 7.15 that data assimilation with time window of both 24 hours and 12 hours is effective to decrease the RMSD at all the six stations. In the period from Dec05 12:00 to Dec06 00:00, the RMSDs at Buesum, Helgoland and Wilhelmshaven are slightly smaller in test 2 and and at the other three stations the

Table 7.9: Setting of tests with different time window for data assimilation

	time window(hour)
test 1	24
test 2	12
test 3	6
test 4	3

RMSDs in test 2 are slightly larger. In the period from Dec06 00:00 to Dec06 12:00, the RMSDs at all the stations in test 2 are slightly smaller. Generally speaking, data assimilation with time window of 12 hours can improve the accuracy of model but the improvement is insignificant.

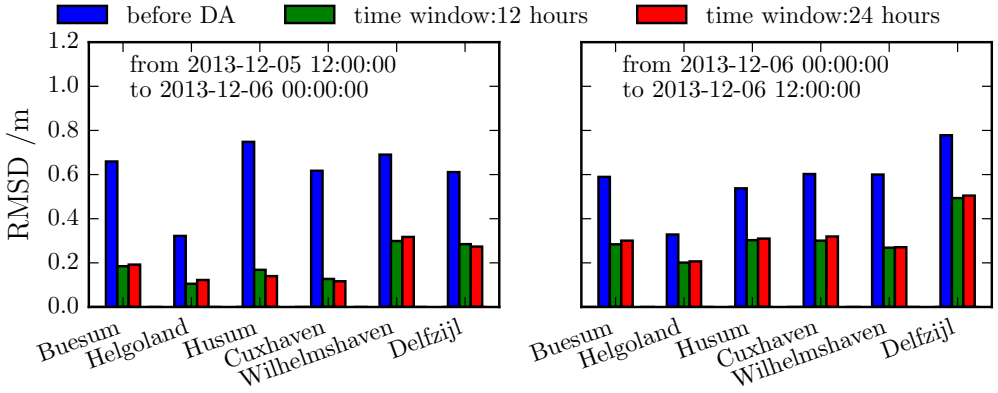

Figure 7.15: RMSD of water level before and after data assimilation in test 1 and 2. The RMSD is calculated over 12 hours

RMSD over six hours from test 1, 2, and 3 are shown in Figure 7.16. In the first two periods, data assimilation in test 3 (with time window of six hours) has similar performance to test 1 and 2. In the period from Dec06 00:00 to Dec06 06:00, in which the surge peak happened, RMSDs at all the six stations in test 3 are much smaller than that in test 1 and 2. In the period from Dec06 06:00 to Dec06 12:00, the RMSDs before data assimilation are much smaller than that in the first three periods; as a result, the improvements due to data assimilation are also smaller than that in the first three period. Nevertheless, RMSDs in test 3 are smaller than test 1 and 2 at the stations Buesum, Helgoland, Cuxhaven and Husum.

RMSD over three hours from test 1, 2, 3 and 4 are shown in Figure 7.17. In most 3-hour periods and at most stations, data assimilation can reduce the RMSD very effectively. But in the period from Dec05 21:00 to Dec06 00:00 at Helgoland, RMSDs are increased after data assimilation with any time window. Figure 7.8

shows that modeled water level at Helgoland are not sensitive to model parameters and RMSDs of all models are relatively low. Worse results after data assimilation also happen in the period Dec06 09:00 to Dec06 12:00, in which data assimilations with time window of 12 hours and 24 hours make the RMSDs larger at all stations but Wilhelmshaven; however, data assimilations with time window of 3 hours and 6 hours can still improve the model at station Buesum, Helgoland, Husum and Cuxhaven. The major advantage of data assimilation of 3-hour time window is in the period from Dec06 03:00 to Dec06 06:00, in which the surge gets to peak in the German Bight.

In summary, shorter time window for storm surge data assimilation is preferable to longer time window, especially for the period when the surge peak occurs. This is because wind drag coefficient C_d changes rapidly at this moment and shorter time windows can capture fast-changing features more accurately.

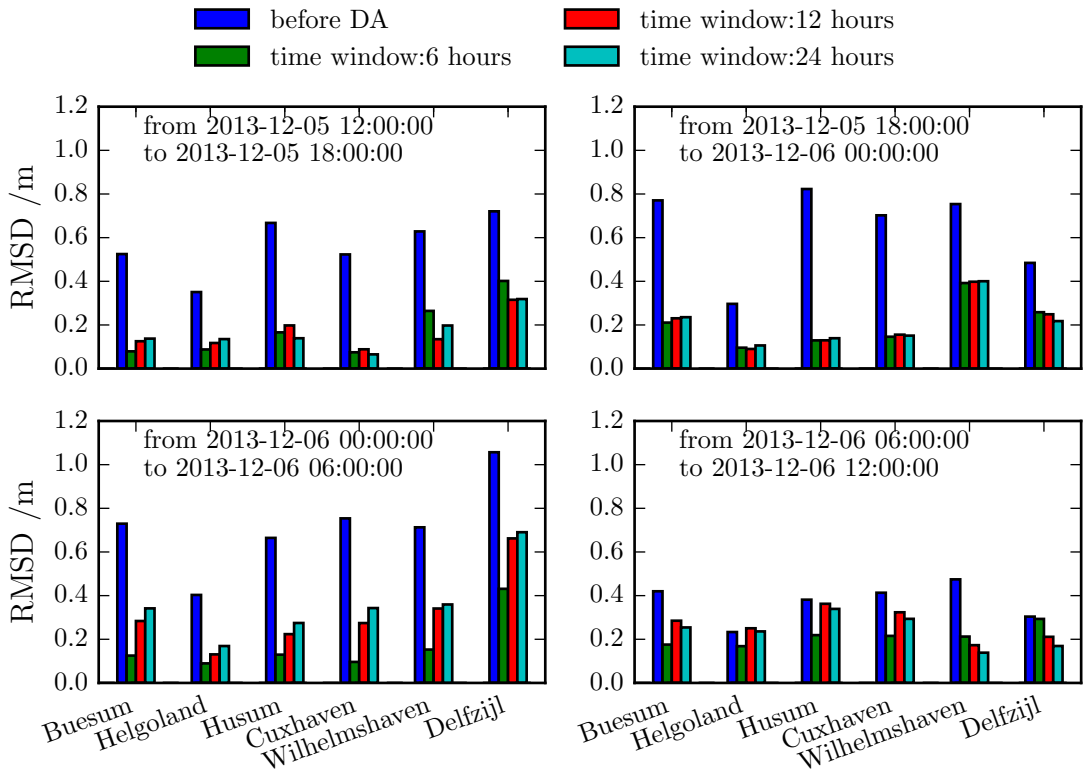

Figure 7.16: RMSD of water level before and after data assimilation in test 1, 2 and 3. The RMSD is calculated over 6 hours

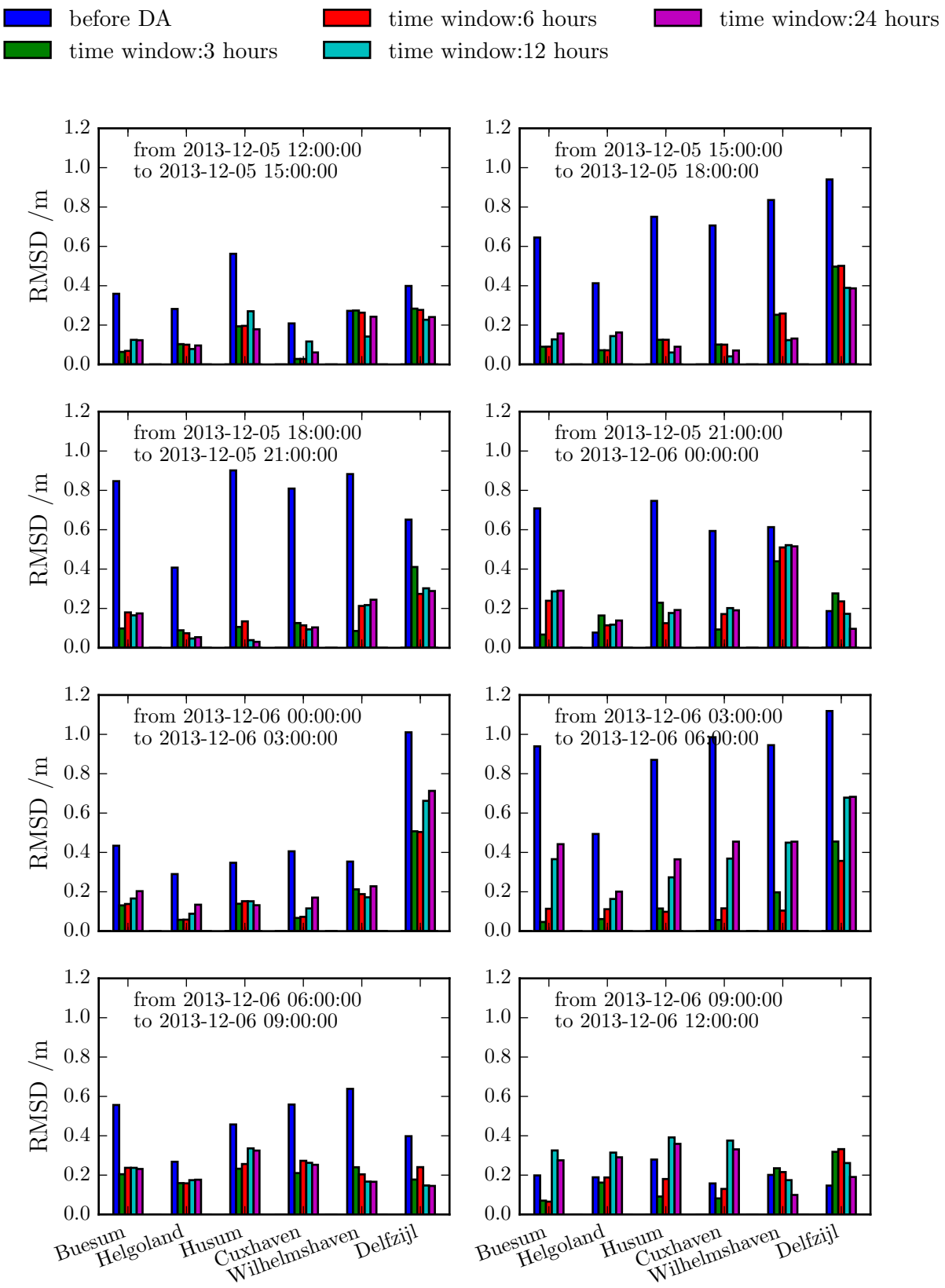

Figure 7.17: RMSD of water level before and after data assimilation in test 1, 2, 3 and 4. The RMSD is calculated over 3 hours (to be continued)

7.5 Data assimilation on the other two storm surge events

The aforementioned tests show that the proposed data assimilation method can improve the model skills significantly by adjusting the wind drag coefficient and Chezy coefficient of the GBM. In this section data assimilation will be applied to another two storm surge events in recent years. These events of storm surge are described briefly in Table 7.10. According to the definition of BSH, both the two storm surge events classified as "severe storm surge". The control variables are b for the function of wind drag coefficient and Chezy coefficient of the GBM. The water level data from the tidal gauge at Buesum are used for data assimilation. It can be expected that all the stations in the east coast of German Bight can be improved effectively. The time windows of data assimilation cover the periods of peak water level in each event of storm surge.

Table 7.10: Two events of storm surge in the German Bight

	occurring time	Peak surge at Buesum (m)
event 1	October 28 2013, 02:00UTC	2.32
event 2	January 11 2015, 02:30UTC	2.39

The results of data assimilation for the two storm surge events are shown in Figure 7.18 and 7.19. Figure 7.18 shows the time series of modeled water level before and after data assimilation and their comparison with observed water level at Buesum and Husum in storm event 1 with the time window between 2013-10-28 12:00UTC and 2013-10-28 20:00UTC. Data assimilation improves the model skill considerably at both the two stations. The RMSD of water level can decreased by 68.8% at Husum even if only observations at Buesum are used for data assimilation (see Table 7.11). Data assimilation improves the storm surge simulation more significantly for the second storm surge event (time window between 2015-01-10 11:00UTC and 2015-01-10 23:00UTC), as the relative decrease of RMSD is up to 91.8% at Cuxhaven.

Table 7.11: RMSD of water level before and after data assimilation in the storm in October 28 2013

station	RMSD (before DA)	RMSD (after DA)	relative decrease of RMSD
Buesum	0.668	0.138	79.3%
Husum	0.747	0.233	68.8%

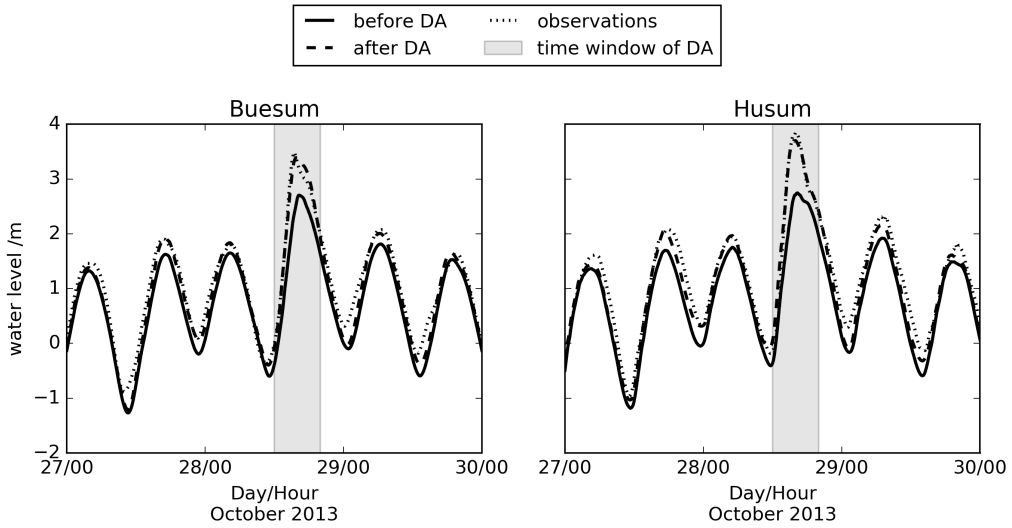

Figure 7.18: Time series of water level before and after data assimilation in the storm in October 28 2013

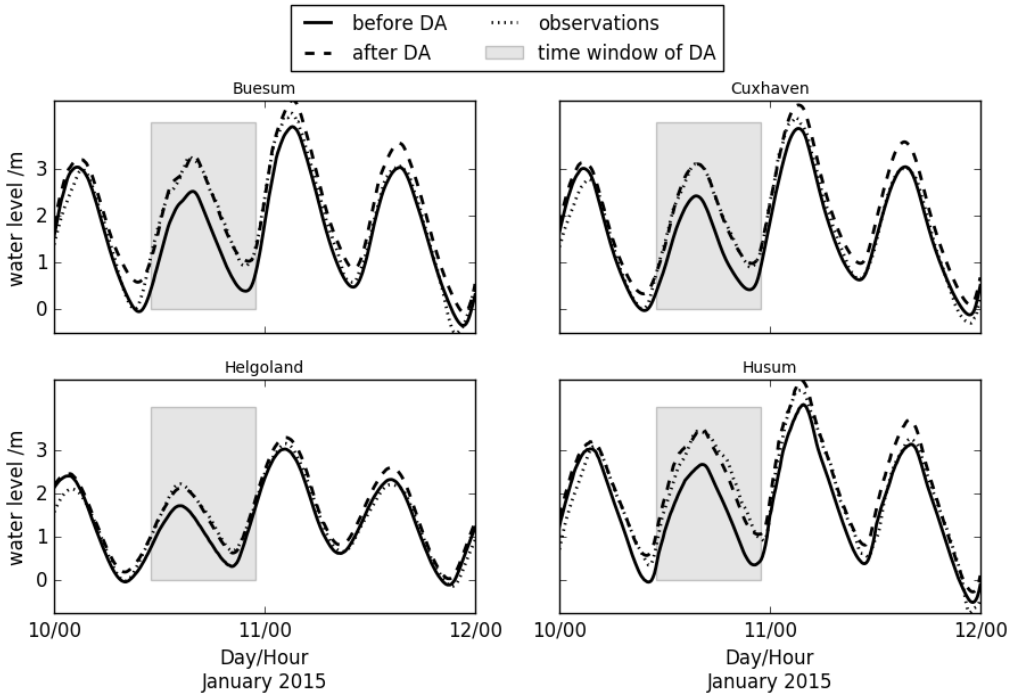

Figure 7.19: Time series of water level before and after data assimilation in the storm in January 10 2015

Donelan's method is used again to assess b adjusted by data assimilation. Before calculating C_d using Donelan's method, the wave state in these two storm events are simulated. Figure 7.20 shows the comparison of wind drag coefficient derived from different methods at location C and D (see Figure 7.10). For both storm events,

Table 7.12: RMSD of water level before and after data assimilation in the storm in January 11 2015

station	RMSD (before DA)	RMSD (after DA)	relative decrease of RMSD
Buesum	0.706	0.078	88.9%
Cuxhaven	0.633	0.052	91.8%
Helgoland	0.410	0.057	86.0%
Husum	0.832	0.223	73.2%

data assimilation can increase the Smith formula derived C_d and pull it to the wave model derived C_d, which is regarded to be a more reasonable value.

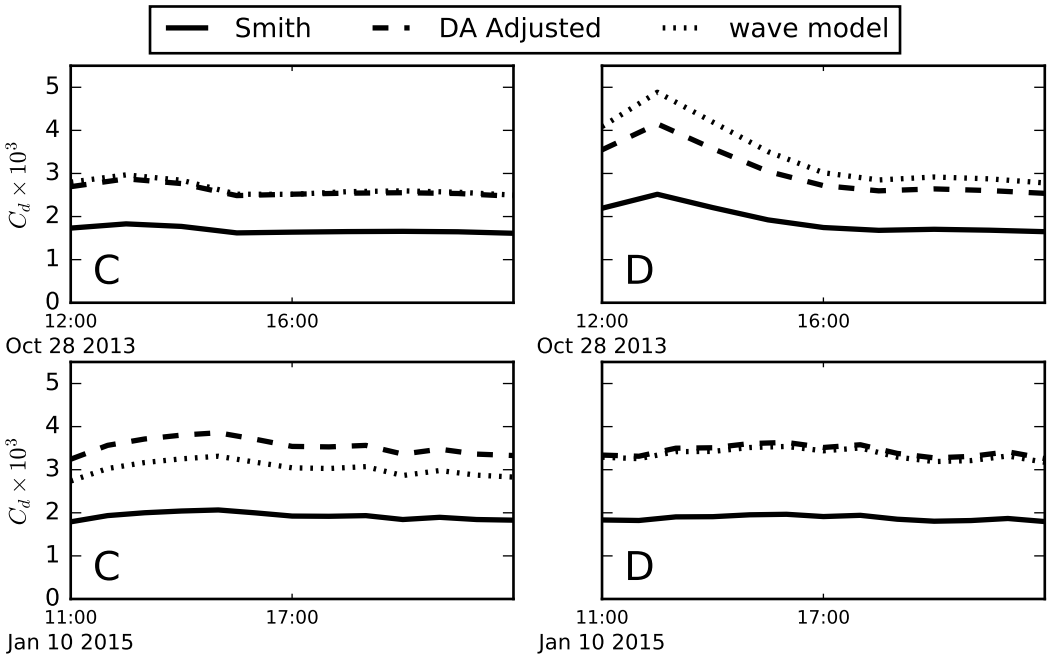

Figure 7.20: Time series of C_d in the two storm surge events

After three storm surge events are investigated by means of data assimilation, comes a question that whether an optimal b and Chezy coefficient for the GBM can be determined. Table 7.13 lists the values of b and Chezy coefficient before and after data assimilation. The values of b in all the three storm events are increased to values larger than 0.1 from the first guess of 0.063 after data assimilation. The optimal b for the storm_201501 differs more significantly. It seems that there is no optimal b for all the storm surge events. One possible reason is that the storm intensity and features vary in the three events. Figure 7.21 shows the wind roses in the three storms. All the three storms swept the German Bight from west to east. The storm_201312 has the largest intensity; the most frequent magnitude is in the

range of 23-25m/s; the largest wind speed can be up to 29m/s. The storm_201310 is of the smallest intensity; most frequent wind magnitude is 17-19m/s. The main wind direction of these storms are also different. The simple uniform linear function for C_d can not well describe the actually spatial-varying and complex process of air-sea momentum transfer. There is also no unique optimal Chezy coefficient for all the storm surge events. This makes sense because bottom shear stress would be enhanced by wave. That is the reason why in storm_201312 the optimal Chezy coefficient is smaller than in the other two.

Table 7.13: The values of b and Chezy coefficient before and after data assimilation for the three storm events

	before DA		**after DA**	
	b	Chezy coefficient	b	Chezy coefficient
storm_201310	0.063	69.88	0.117	73.88
storm_201312	0.063	69.88	0.108	63.06
storm_201501	0.063	69.88	0.141	73.66

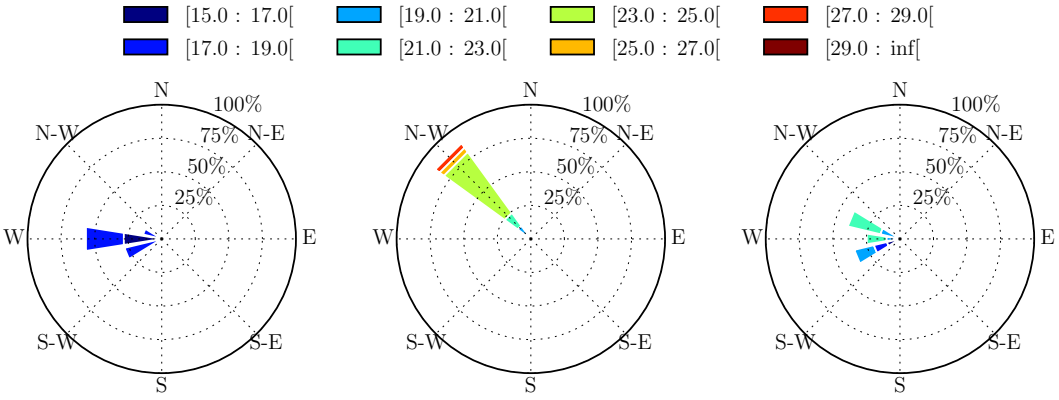

Figure 7.21: Wind rose of the three storm events. The wind data is spatially averaged over the domain of the GBM. Left: storm_201310; middle: storm_201312; right: storm_201501

7.6 Application of data assimilation into the operational system for storm surge in the German Bight

With the development of computer ability and increasing number of ocean data, operational ocean forecasting systems (operational system for simplicity in the fol-

lowing) have been setup in many countries. Operational system can provide reliable and punctual forecast of ocean state. A typical operational system consists of:

1. Well developed observation networks to measure the ocean state continually.
2. Computer systems supporting the numerical ocean models, data storing, data analysis, data assimilation and visualization.
3. Data communication facilities for fast transfer of observations to the control center.

Among these components, data assimilation is the essential part for reliable estimate and forecast of ocean state. In the above-mentioned sections, it has been shown that the data assimilation method developed in Chapter 4 can improve the storm surge model skill effectively. Sensitivity tests have been performed to investigate the effects of control variables, number of tidal gauges and time window length on the data assimilation.

Fernández Jaramillo (2014) set up an operational system for the coastal area of the German Bight. He used the technique of linear regression, artificial neural networks and fuzzy logic to assimilate the tidal gauges measured water level to improve the water level imposed on the open boundary of models. The data assimilation can reduce the RMSDs in the model of the Dithmarschen Bight by 33%. The neural network were also used for the water level forecast without consuming too much computation time. Besides the new data assimilation method, he developed a web interface for the data visualization. Forecast of water level and wave parameters can be obtained from this website.

Since the adjoint free 4Dvar developed in Chapter 4 can improve the storm surge simulation effectively, it is added to the operational system of Fernández Jaramillo (2014) as a supplement for storm surge simulation. In normal weather, the operational system works without using 4Dvar. When strong storm comes, 4Dvar will be activated to adjust the wind drag coefficient and Chezy coefficient, which contain more uncertainties in strong storms. The strong storms means the peak wind speed is more than 20 m/s, which can be predicted by numerical weather models several days before the storm. Using the predicted wind and air pressure field, a preliminary simulation of surge (without tide) can be simulated. By examining the time series of simulated surge at a location, a period named DA period can be defined in which the surge is larger than a certain value(for example, 1.5m above the mean water level). The wind drag coefficient and Chezy coefficient will be adjusted in the DA period. DA period is split into shorter periods. These shorter periods are time window for data assimilation. When there are new observations in one time window,

4Dvar will start to adjust wind drag coefficient and Chezy coefficient using model results and observations. The adjusted model parameters will be used as a first guess for the next time window and also used for the forecast in the rest time of the DA period. This process is illustrated in Figure 7.22, in which the black solid double arrow indicates the first nowcasting and the black dashed double arrow indicates the first forecasting. Figure 7.7 shows that the cost function almost levels off after the first iteration of the outer loop. Therefore, To speed up the data assimilation, the number of outer loop will be only one in the operational system.

The storm event in the early December 2013 is used here as an example to demonstrate how well the operational system can perform. From data assimilation test of the three storm surge events in Section 7.5, it has been found that $b = 0.122$ ($a = 0.61$) for the wind drag coefficient can improve the storm surge model than using the Smith formula. Data assimilation experiments with time window longer than 6 hour showed that Chezy coefficient ranging 66 to 73 $m^{1/2}/s$ is proper for storm surge simulation. The value 69.88 $m^{1/2}/s$ for the Chezy coefficient will still be used. The DA period is between Dec.05 12:00 and Dec.06 12:00. The time window is 30 minutes because it is found in Section 7.4.4 that using shorter time window can generally improve model skill. Observations at Buesum are used for data assimilation.

In order to justify the performance of the operational system, the model with the Smith formula and the updated $b = 0.122$ will be run. The data assimilation skill is measured by means of RMSD over time windows and forecast skill over short periods after time windows. These short periods are 30 minutes, one hour, 1.5 hours and two hours long after data assimilation respectively. The bars in Figure 7.23 show the RMSDs at Buesum, Helgoland, Husum and Cuxhaven. Among the four stations, Buesum is used for data assimilation in the operational system.

The RMSD at Buesum over the time window is 0.025m, which is much smaller than models with Smith formula and $b = 0.122$. The mean RMSD over the following 30-minute period is 0.07m, still much smaller than other two models. The mean RMSD over the next 30-minute period increases to 0.14m, about half of the RMSD of the model with $b = 0.122$. The mean RMSD becomes comparable to the RMSD of the model with $b = 0.122$ in the period two hours after the time window of data assimilation, about half of the RMSD of the model with the Smith formula.

The RMSD at other three stations over the time window are larger than that at Buesum, but still lower than the model with $b = 0.122$. The RMSDs also increase with the periods away from the time window. After two hours, the mean RMSDs become similar to the model with $b = 0.122$.

Figure 7.23 indicates that nowcasting with smaller time window can reduce model errors and increase the accuracy of forecast. However, the improvements of forecast can last about two hours then the model errors become similar to the model with $b = 0.122$. This implies the wind drag coefficient and Chezy coefficient change with wave state very fast. Short time window is needed for data assimilation.

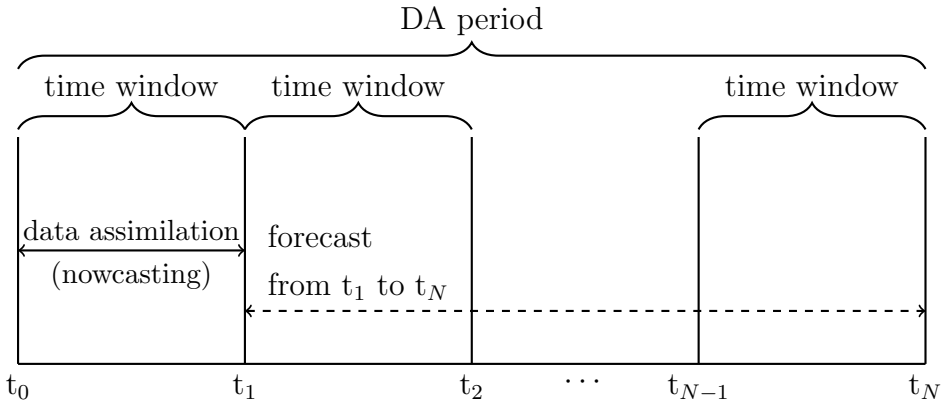

Figure 7.22: The work flow of the operational system for storm surge. The black solid double arrow denotes the time window for data assimilation between t_0 and t_1 (nowcasting); the black dashed double arrow denotes the forecast between t_1 and t_N after the nowcasting between t_0 and t_1. The following nowcasting is perform between t_1 and t_2 and forecasting between t_2 and t_N. The cycle of nowcasting and forecasting continues until t_N

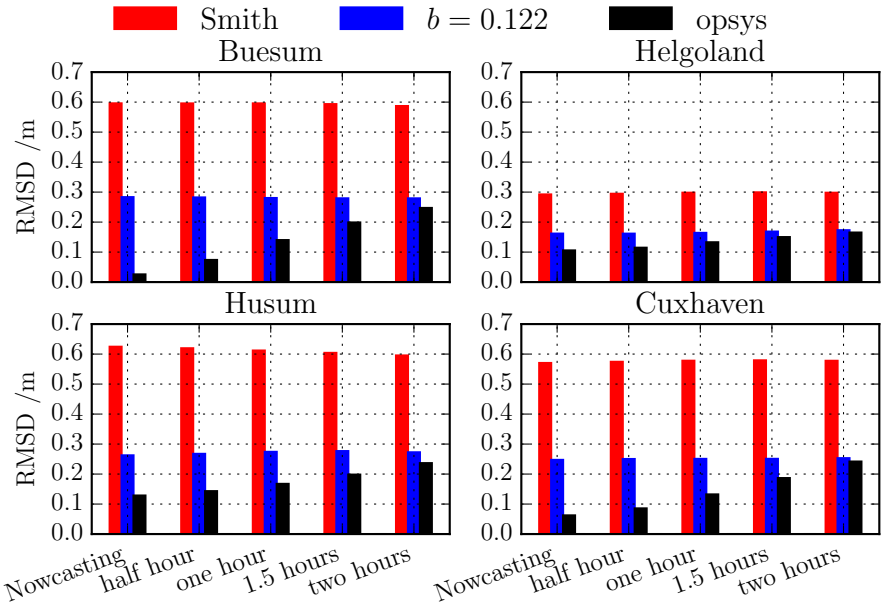

Figure 7.23: Nowcasting and forecasting skills using different function of C_d with respect to wind speed

Chapter 8

Conclusions and future work

In the recent years, more and more operational coastal ocean forecasting systems are developed. They provide vital information for ocean resource exploitations, human lives and environment protections. A typical coastal forecasting system usually consists of a numerical model and an observing system. Many coastal forecasting systems also include a data assimilation module, which helps improving forecasting skills. In this thesis, a new data assimilation method is proposed using the principles of incremental 4Dvar and inspired by the ensemble method. It has a comparable ability to the adjoint method for low dimension parameters estimation but is easier to implement. The data assimilation is then used to improve a storm surge model.

8.1 Conclusions

Uncertainties of storm surge simulation

Using Smith's formula ($C_d = 10^{-3}(0.61 + 0.063U_{10})$) for wind drag coefficient C_d, Delft3D underestimates surge compared with observations during a strong storm. Sensitivity tests show that C_d accounts for the largest uncertainties of storm surge simulation. The uncertainty of C_d mainly comes from the poor understanding of the air-sea momentum transfer and over-simplified representation of C_d. The difficulty in measuring wind shear stress normally for high wind speed above sea surface also results in uncertainty of C_d. For such problem, data assimilation is used to decrease the uncertainty and improve storm surge models.

Adjoint free 4Dvar

An adjoint free 4DVar is proposed using the principle of incremental 4Dvar. By representing the tangent linear equations using ensemble of model parameters and ensemble of model results, the constrained optimization of cost function (Equation 4.18) in 4Dvar becomes an unconstrained problem and the adjoint equations are avoided.

The number of ensemble members should be larger than the number of control variables. Otherwise, the estimation of tangent linear equation becomes an under-determined problem, which does not have an unique solution. The problem arises for high dimension model parameters because it is impossible to run a high number of models. In this case, principal components analysis (PCA) is used to reduce the parameter dimension and figure out the under-determined problem.

The proposed adjoint free 4Dvar is validated by a series of numerical twin experiments. It effectively improves model skills by adjusting model parameters which have large uncertainties. It also works well on a nesting model system. Twin experiments indicate that model parameters with low accuracy should be adjusted, otherwise, the adjusted parameters become far from reasonable.

Data assimilation for storm surge simulations

Adjusting C_d and Chezy coefficient of the GBM is the best scheme in the perspective of both effectiveness and efficiency. Due to fast-changing sea state during a strong storm, C_d also changes rapidly over space and time. Shorter time window resolves better temporal change of C_d and improves storm surge model skill, especially at the wind peak time.

The data assimilation is performed on three storm events, being able to improve storm surge simulations effectively. The relative improvement ranges from 60% to 90% compared to the model with first guess of C_d and Chezy coefficient. Spatial-temporal varying C_d is calculated by considering sea state, which is simulated by a well-calibrated wave model. Data assimilation directs the C_d calculated using Smith's formula to the value of wave-related C_d, indicating C_d is suitable control variable for storm surge simulation.

From data assimilation on the three storm events, a new linear function between C_d and U_{10} is derived.

$$C_d = 10^{-3}(0.61 + 0.122U_{10}) \tag{8.1}$$

This formula only works for the German Bight model for storm surge simulation. The function between C_d and U_{10} is location dependent. Therefore, for other study

areas, data assimilation should be performed on those specific areas.

Data assimilation for nowcasting and forecasting

The proposed data assimilation method is integrated experimentally to the operational system developed by FTZ. By nowcasting of observations, the storm surge model skill is better than just using equation (8.1). The forecasting skill is also improved within two hours after nowcasting.

Guidelines of implementing data assimilation

This thesis presents a set of procedures to improve storm surge models using adjoint free 4Dvar. The procedures are applicable to improving other models which have different mechanism than storm surge. Guidelines of implementing data assimilation are generalized from this study as follow:

1. Prepare observations according to the phenomenon of interest.

2. Models should be fully understood, carefully constructed and well calibrated for most parameters. Resolutions should be high enough to capture the variations in time, space or spectrum.

3. Find out parameters which have large uncertainties and affect model skills significantly.

4. Adjust those model parameters found in step 3 using data assimilation.

5. Investigate the validity of the adjusted parameters.

8.2 Future work

The presented researches focus on the proposal of adjoint free 4Dvar and its application to a storm surge model. Future works on data assimilation and coastal models should be continued and emphasized as follows:

- For high dimensional parameters with spatial correlation with each element, the error covariance must be provided for dimension reduction and correction spreading. One of the future study will be the construction of a proper error covariance.

- The development of storm surge model in the German Bight should be continued and improved. As the computation ability is increasing, spatial resolutions of the model grid can be increased to better capture fast changing surges. More updated bathymetry data will be available every year to increase the model accuracy, especially in the shallow waters.

- As indicated in the previous chapters, wind drag coefficient C_d is the most important parameter for storm surge simulation. However, the model Delft3D cannot calculate C_d using wave state. The Flow-wave coupled model does not have such process either. This is a model limitation. It needs to be fixed for better model accuracy. Therefore, in the future, based on the better understanding of the wind-wave process, the flow-wave coupled module will be improved. The wind-wave process will be added to that module, making it able to improve the wind drag coefficient in the model by itself. The further plan is to develop an air-sea coupled model for storm surge simulations. Air-sea coupled models can not only provide higher resolution wind field data, but also improve the wind drag coefficient.

Appendix A

Comparison of tidal harmonic constants

Table A.1: Comparison of harmonic constants at Husum

	M2		S2		O1	
	amp(m)	pha(°)	amp(m)	pha(°)	amp(m)	pha(°)
tidal_gauge	1.59	356	0.49	91	0.08	260
chezy_65	1.43	1	0.37	89	0.05	265
chezy_69	1.59	358	0.42	87	0.05	263
chezy_85	2.04	351	0.57	80	0.05	264
manning2	1.94	357	0.54	86	0.05	265
manning25	1.44	6	0.38	94	0.05	269
manning3	1.13	13	0.28	101	0.05	273

Table A.2: Comparison of harmonic constants at Wittduen

	M2		S2		O1	
	amp(m)	pha(°)	amp(m)	pha(°)	amp(m)	pha(°)
tidal_gauge	1.13	342	0.33	69	0.07	253
chezy_65	1.06	343	0.26	66	0.05	259
chezy_69	1.18	341	0.29	64	0.05	258
chezy_85	1.54	335	0.41	57	0.04	252
manning2	1.53	333	0.41	53	0.04	251
manning25	1.13	339	0.29	60	0.04	258
manning3	0.89	344	0.22	66	0.04	261

Table A.3: Comparison of harmonic constants at Esbjerg

	M2		S2		O1	
	amp(m)	pha(°)	amp(m)	pha(°)	amp(m)	pha(°)
tidal_gauge	0.77	32	0.19	122	0.07	269
chezy_65	0.68	10	0.16	88	0.04	264
chezy_69	0.75	9	0.18	86	0.04	262
chezy_85	0.94	5	0.23	83	0.04	258
manning2	0.95	4	0.24	81	0.04	256
manning25	0.73	9	0.18	87	0.04	263
manning3	0.59	14	0.14	93	0.04	267

Table A.4: Comparison of harmonic constants at Havneby

	M2		S2		O1	
	amp(m)	pha(°)	amp(m)	pha(°)	amp(m)	pha(°)
tidal_gauge	0.84	24	0.23	115	0.07	265
chezy_65	0.77	8	0.18	90	0.05	260
chezy_69	0.85	6	0.21	87	0.05	259
chezy_85	1.10	1	0.28	81	0.04	255
manning2	1.08	360	0.27	80	0.04	253
manning25	0.80	6	0.20	87	0.04	260
manning3	0.64	12	0.15	95	0.04	263

Table A.5: Comparison of harmonic constants at Bremerhaven

	M2		S2		O1	
	amp(m)	pha(°)	amp(m)	pha(°)	amp(m)	pha(°)
tidal_gauge	1.73	343	0.52	77	0.08	250
chezy_65	1.54	352	0.40	82	0.05	263
chezy_69	1.71	349	0.45	78	0.05	261
chezy_85	2.23	338	0.61	66	0.05	254
manning2	2.12	342	0.58	70	0.04	257
manning25	1.55	355	0.41	84	0.05	266
manning3	1.20	3	0.31	93	0.05	272

Bibliography

Abdalla, S. and Cavaleri, L. (2002). Effect of wind variability and variable air density on wave modeling. *Journal of Geophysical Research: Oceans (1978–2012)*, 107(C7):17–1.

Anderson, J. L. (2001). An ensemble adjustment kalman filter for data assimilation. *Monthly weather review*, 129(12):2884–2903.

Atlas, R. (1997). Atmospheric observations and experiments to assess their usefulness. *Journal of the Meteorologzcal Society of Japan*, 75(1B):111–130.

Beardsley, R. C., Chen, C., and Xu, Q. (2013). Coastal flooding in scituate (ma): A fvcom study of the 27 december 2010 nor'easter. *Journal of Geophysical Research: Oceans*, 118(11):6030–6045.

Bishop, C. H., Etherton, B. J., and Majumdar, S. J. (2001). Adaptive sampling with the ensemble transform kalman filter. part i: Theoretical aspects. *Monthly weather review*, 129(3):420–436.

Brocks, K. and Krugermeyer, L. (1970). The hydrodynamic roughness of the sea surface. *Ber. Inst. Radiometeorol. and Marit. Meteorol*, (14):55.

Bruss, G. and Mayerle, R. (2009). Investigations on the influence of the wind drag coefficient in storm surge models. In *Proceedings of the 3rd International Conference in Ocean Engineering*, pages 325–332.

Burgers, G., Jan van Leeuwen, P., and Evensen, G. (1998). Analysis scheme in the ensemble kalman filter. *Mon. Wea. Rev.*, 126(6):1719–1724.

Butler, T., Altaf, M., Dawson, C., Hoteit, I., Luo, X., and Mayo, T. (2012). Data assimilation within the advanced circulation (adcirc) modeling framework for hurricane storm surge forecasting. *Monthly Weather Review*, 140(7):2215–2231.

Canizares, R., Heemink, A., and Vested, H. (1998). Application of advanced data assimilation methods for the initialisation of storm surge models. *Journal of Hydraulic Research*, 36(4):655–674.

Canizares, R., Madsen, H., Jensen, H., and Vested, H. (2001). Developments in operational shelf sea modelling in danish waters. *Estuarine, Coastal and Shelf Science*, 53(4):595–605.

Charnock, H. (1955). Wind stress on a water surface. *Quarterly Journal of the Royal Meteorological Society*, 81(350):639–640.

Christiansen, H. and Siefert, W. (1978). Storm surge prediction by combined wind and tide data. *Coastal Engineering Proceedings*, 1(16).

Clayton, A., Lorenc, A. C., and Barker, D. M. (2013). Operational implementation of a hybrid ensemble/4d-var global data assimilation system at the met office. *Quarterly Journal of the Royal Meteorological Society*, 139(675):1445–1461.

Cohn, S. E. and Todling, R. (1995). Approximate kalman filters for unstable dynamics. In *Second International Symposium on the Assimilation of Observations in Meteorology and Oceanography*, pages 241–246. Citeseer.

Conner, W., Kraft, R., and Harris, D. L. (1957). Empirical methods for forecasting the maximum storm tide due to hurricanes and other tropical storms. *Monthly Weather Review*, 85(4):113–116.

Cooper, C. and Pearce, B. (1982). Numerical simulations of hurricane-generated currents. *Journal of Physical Oceanography*, 12(10):1071–1091.

COSMO (2007). Operations at dwd. Technical report, cosmo-eu.

Courtier, P., Thépaut, J.-N., and Hollingsworth, A. (1994). A strategy for operational implementation of 4d-var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120(519):1367–1387.

Cressman, G. P. (1959). An operational objective analysis system. *Monthly Weather Review*, 87(10):367–374.

Dahl, E. and Støttrup, J. (2012). *Global challenges in integrated coastal zone management*. John Wiley & Sons.

Daley, R. (1993). *Atmospheric data analysis*. Number 2. Cambridge university press.

Deacon, E. (1962). Small-scale interaction in the sea. *Interchange of properties between sea and air*.

Donelan, M. (1982). The dependence of the aerodynamic drag coefficient on wave parameters.

Donelan, M. A., Dobson, F. W., Smith, S. D., and Anderson, R. J. (1993). On the dependence of sea surface roughness on wave development. *Journal of Physical Oceanography*, 23(9):2143–2149.

Doodson, A. T. and Dines, J. S. (1929). *Report on Thames floods*. HM Stationery Office.

Egbert, G. D. and Erofeeva, S. Y. (2002). Efficient inverse modeling of barotropic ocean tides. *Journal of Atmospheric and Oceanic Technology*, 19(2):183–204.

Elias, E., Walstra, D., Roelvink, J., Stive, M., and Klein, M. (2001). Hydrodynamic validation of delft3d with field measurements at egmond. In *COASTAL ENGINEERING CONFERENCE*, volume 3, pages 2714–2727. ASCE AMERICAN SOCIETY OF CIVIL ENGINEERS.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162.

Evensen, G. (2009). *Data assimilation: the ensemble Kalman filter*. Springer Science & Business Media.

Fernández Jaramillo, J. M. (2014). *Development of an Operational System for a Coastal Area on the German North Sea using Artificial Intelligence*. PhD thesis, Kiel, Christian-Albrechts-Universität, Diss., 2014.

Fertig, E. J., Harlim, J., and Hunt, B. R. (2007). A comparative study of 4d-var and a 4d ensemble kalman filter: Perfect model simulations with lorenz-96. *Tellus A*, 59(1):96–100.

Flather, R. and Proctor, R. (1983). Prediction of north sea storm surges using numerical models: Recent developments in uk. In *North Sea Dynamics*, pages 299–317. Springer.

Flather, R. A. (1984). A numerical model investigation of the storm surge of 31 january and 1 february 1953 in the north sea. *Quarterly Journal of the Royal Meteorological Society*, 110(465):591–612.

Gandin, L. S. and Hardin, R. (1965). *Objective analysis of meteorological fields*, volume 242. Israel program for scientific translations Jerusalem.

Garratt, J. (1977). Review of drag coefficients over oceans and continents. *Monthly weather review*, 105(7):915–929.

Geernaert, G., Larsen, S., and Hansen, F. (1987). Measurements of the wind stress, heat flux, and turbulence intensity during storm conditions over the north sea. *Journal of Geophysical Research: Oceans*, 92(C12):13127–13139.

Heaps, N. S. (1969). A two-dimensional numerical sea model. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 265(1160):93–137.

Heemink, A., Bolding, K., and Verlaan, M. (1995). Storm surge forecasting using kalman filtering.

Holland, G. J. (1980). An analytic model of the wind and pressure profiles in hurricanes. *Monthly weather review*, 108(8):1212–1218.

Holthuijsen, L. H. (2007). *Waves in oceanic and coastal waters*. Cambridge University Press.

Horsburgh, K. and De Vries, H. (2011). Guide to storm surge forecasting.

Horsburgh, K. and Wilson, C. (2007). Tide-surge interaction and its role in the distribution of surge residuals in the north sea. *Journal of Geophysical Research: Oceans (1978–2012)*, 112(C8).

Hoteit, I., Korres, G., and Triantafyllou, G. (2005). Comparison of extended and ensemble based kalman filters with low and high resolution primitive equation ocean models. *Nonlinear Processes in Geophysics*, 12(5):755–765.

Houtekamer, P. L. and Mitchell, H. L. (1998). Data assimilation using an ensemble kalman filter technique. *Mon. Wea. Rev.*, 126(3):796–811.

Hunt, B., Kalnay, E., Kostelich, E., Ott, E., Patil, D., Sauer, T., Szunyogh, I., Yorke, J., and Zimin, A. (2004). Four-dimensional ensemble kalman filtering. *Tellus A*, 56(4):273–277.

Hydraulics, D. (2007). User manual delft3d-flow: Wl. *Delft Hydraulics*.

Janeković, I., Powell, B., Matthews, D., McManus, M., and Sevadjian, J. (2013). 4d-var data assimilation in a nested, coastal ocean model: A hawaiian case study. *Journal of Geophysical Research: Oceans*, 118(10):5022–5035.

Jelesnianski, C. P. (1966). Numerical computations of storm surges without bottom stress. *Monthly Weather Review*, 94(6):379–394.

Jiao, J. (2014). *Morphodynamics of Ameland Inlet: Medium-term Delft3D Modelling*. PhD thesis, TU Delft, Delft University of Technology.

Johnson, W. R. and Kowalik, Z. (1986). Modeling of storm surges in the bering sea and norton sound. *Journal of Geophysical Research: Oceans (1978–2012)*, 91(C4):5119–5128.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45.

Kitaigorodskii, S. and Volkov, Y. A. (1965). On the roughness parameter of the sea surface and the calculation of momentum flux in the near-water layer of the atmosphere. *Izv. Atmos. Oceanic Phys*, 1:973–988.

Kondo, J. (1975). Air-sea bulk transfer coefficients in diabatic conditions. *Boundary-Layer Meteorology*, 9(1):91–112.

Large, W. and Pond, S. (1981). Open ocean momentum flux measurements in moderate to strong winds. *Journal of physical oceanography*, 11(3):324–336.

Lawless, A. S. (2013). Variational data assimilation for very large environmental problems. *Large Scale Inverse Problems: Computational Methods and Applications in the Earth Sciences*, 13:55.

Letchford, C. and Zachry, B. (2009). On wind, waves, and surface drag. In *5th European and African conference on wind engineering, Florence, Italy*.

Lewis, J. M., Lakshmivarahan, S., and Dhall, S. (2006). *Dynamic data assimilation: a least squares approach*, volume 13. Cambridge University Press.

Lionello, P., Sanna, A., Elvini, E., and Mufato, R. (2006). A data assimilation procedure for operational prediction of storm surge in the northern adriatic sea. *Continental shelf research*, 26(4):539–553.

Longuet-Higgins, M. S. and Stewart, R. (1964). Radiation stresses in water waves; a physical discussion, with applications. In *Deep Sea Research and Oceanographic Abstracts*, volume 11, pages 529–562. Elsevier.

Lorenc, A., Ballard, S., Bell, R., Ingleby, N., Andrews, P., Barker, D., Bray, J., Clayton, A., Dalby, T., Li, D., et al. (2000). The met. office global three-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 126(570):2991–3012.

M. D. Powell, P. J. V. . T. A. R. (2003). Reduced drag coefficient for high wind speeds in tropical cyclones. *Reduced drag coefficient for high wind speeds in tropical cyclones*, 422:279–283.

Mastenbroek, C., Burgers, G., and Janssen, P. (1993). The dynamical coupling of a wave model and a storm surge model through the atmospheric boundary layer. *Journal of Physical Oceanography*, 23(8):1856–1866.

Masutani, M., Woollen, J. S., Lord, S. J., Emmitt, G. D., Kleespies, T. J., Wood, S. A., Greco, S., Sun, H., Terry, J., Kapoor, V., et al. (2010). Observing system simulation experiments at the national centers for environmental prediction. *Journal of Geophysical Research: Atmospheres*, 115(D7).

Mayerle, R. and Zielke, W. (2005). Promorph–predictions of medium-scale morphodyanimcs: Project overview and executive summary. *Die Küste*, 69:1–24.

Miller, B. I. (1964). A study of the filling of hurricane donna (1960) over land. *Monthly Weather Review*, 92(9):389–406.

Moore, A. M., Arango, H. G., Broquet, G., Powell, B. S., Weaver, A. T., and Zavala-Garay, J. (2011). The regional ocean modeling system (roms) 4-dimensional variational data assimilation systems: Part i–system overview and formulation. *Progress in Oceanography*, 91(1):34–49.

Mulligan, R. P., Hay, A. E., and Bowen, A. J. (2008). Wave-driven circulation in a coastal bay during the landfall of a hurricane. *Journal of Geophysical Research: Oceans (1978–2012)*, 113(C5).

Nerger, L., Janjic, T., Schröter, J., and Hiller, W. (2012). A unification of ensemble square root kalman filters. *Monthly Weather Review*, 140(7):2335–2345.

Ngodock, H. and Carrier, M. (2014a). A 4dvar system for the navy coastal ocean model. part i: System description and assimilation of synthetic observations in monterey bay*. *Monthly Weather Review*, 142(6):2085–2107.

Ngodock, H. and Carrier, M. (2014b). A 4dvar system for the navy coastal ocean model. part ii: Strong and weak constraint assimilation experiments with real observations in monterey bay*. *Monthly Weather Review*, 142(6):2108–2117.

Nocedal, J. and Wright, S. (2006). *Numerical optimization*. Springer Science & Business Media.

Panofsky, R. (1949). Objective weather-map analysis. *Journal of Meteorology*, 6(6):386–392.

Powell, M. (2008). High wind drag coefficient and sea surface roughness in shallow water. *NOAA/AOML Hurricane Research Division Final Rep. to the Joint Hurricane Testbed*.

Prandle, D. (1975). Storm surges in the southern north sea and river thames. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 344, pages 509–539. The Royal Society.

Prandle, D. and Wolf, J. (1978). The interaction of surge and tide in the north sea and river thames. *Geophysical Journal International*, 55(1):203–216.

Pugh, D. (1987). Tides, surges and mean sea-level: a handbook for engineers and scientists, 472 pp.

Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F., and Simmons, A. (2000). The ecmwf operational implementation of four-dimensional variational assimilation. i: Experimental results with simplified physics. *Quarterly Journal of the Royal Meteorological Society*, 126(564):1143–1170.

Robinson, A. R. and Lermusiaux, P. F. (2000). Overview of data assimilation. *Harvard reports in physical/interdisciplinary ocean science*, 62:1–13.

Rossiter, J. R. (1961). Interaction between tide and surge in the thames. *Geophysical Journal International*, 6(1):29–53.

Sakov, P., Evensen, G., and Bertino, L. (2010). Asynchronous data assimilation with the enkf. *Tellus A*, 62(1):24–29.

Sasaki, Y. (1970). Some basic formalisms in numerical variational analysis. *Monthly Weather Review*, 98(12):875–883.

Sasaki, Y. K. (1958). *An objective analysis based on the variational method*.

Sheng, Y. P., Alymov, V., and Paramygin, V. A. (2010). Simulation of storm surge, wave, currents, and inundation in the outer banks and chesapeake bay during hurricane isabel in 2003: The importance of waves. *Journal of Geophysical Research: Oceans (1978–2012)*, 115(C4).

Sheppard, P. (1958). Transfer across the earth's surface and through the air above. *Quarterly Journal of the Royal Meteorological Society*, 84(361):205–224.

Sheppard, P., Tribble, D., and Garratt, J. (1972). Studies of turbulence in the surface layer over water (lough neagh). part i. instrumentation, programme, profiles. *Quarterly Journal of the Royal Meteorological Society*, 98(417):627–641.

Smith, S. D. (1980). Wind stress and heat flux over the ocean in gale force winds. *Journal of Physical Oceanography*, 10(5):709–726.

Smith, S. D. (1988). Coefficients for sea surface wind stress, heat flux, and wind profiles as a function of wind speed and temperature. *Journal of Geophysical Research: Oceans (1978–2012)*, 93.

Smith, S. D., Anderson, R. J., Oost, W. A., Kraan, C., Maat, N., De Cosmo, J., Katsaros, K. B., Davidson, K. L., Bumke, K., Hasse, L., et al. (1992). Sea surface wind stress and drag coefficients: the hexos results. *Boundary-Layer Meteorology*, 60(1-2):109–142.

Smith, S. D. and Banke, E. G. (1975). Variation of the sea surface drag coefficient with wind speed. *Quarterly Journal of the Royal Meteorological Society*, 101(429):665–673.

Stammer, D., Ray, R., Andersen, O. B., Arbic, B., Bosch, W., Carrère, L., Cheng, Y., Chinn, D., Dushaw, B., Egbert, G., et al. (2014). Accuracy assessment of global barotropic ocean tide models. *Reviews of Geophysics*, 52(3):243–282.

Stewart, R. (1974). The air-sea momentum exchange. *Boundary-Layer Meteorology*, 6(1-2):151–167.

Troutman, J. L. (2012). *Variational calculus and optimal control: optimization with elementary convexity*. Springer Science & Business Media.

Verboom, G., de Ronde, J., and van Dijk, R. (1992). A fine grid tidal flow and storm surge model of the north sea. *Continental Shelf Research*, 12(2–3):213 – 233.

Verlaan, M. and Heemink, A. (1997). Tidal flow forecasting using reduced rank square root filters. *Stochastic hydrology and Hydraulics*, 11(5):349–368.

Vidal, R., Ma, Y., and Sastry, S. (2005). Generalized principal component analysis (gpca). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(12):1945–1959.

von Storch, H., Gönnert, G., and Meine, M. (2008). Storm surge–an option for hamburg, germany, to mitigate expected future aggravation of risk. *Environmental Science & Policy*, 11(8):735–742.

Webster, P. J. (2008). Myanmar's deadly daffodil. *Nature Geoscience*, 1(8):488–490.

Wieringa, J. (1974). Comparison of three methods for determining strong wind stress over lake flevo. *Boundary-Layer Meteorology*, 7(1):3–19.

Wikipedia (2016). North sea flood of 1953 — wikipedia, the free encyclopedia. [Online; accessed 2-September-2016].

Woodworth, P. L. and Smith, D. E. (2003). A one year comparison of radar and bubbler tide gauges at liverpool. *The International hydrographic review*, 4(3).

Wu, J. (1969). Froude number scaling of wind-stress coefficients. *Journal of the Atmospheric Sciences*, 26(3):408–413.

Wu, J. (1982). Wind-stress coefficients over sea surface from breeze to hurricane. *Journal of Geophysical Research: Oceans (1978–2012)*, 87(C12):9704–9706.

Wu, W., Sanchez, A., and Zhang, M. (2010). An implicit 2-d depth-averaged finite-volume model of flow and sediment transport in coastal waters. Technical report, DTIC Document.

Wu, X., Flather, R. A., and Wolf, J. (1994). A third generation wave model of european continental shelf seas with depth and current refraction due to tides and surges and its validation using geosat and buoy measurements.

Yelland, M. and Taylor, P. K. (1996). Wind stress measurements from the open ocean. *Journal of Physical Oceanography*, 26(4):541–558.

Yu, L. and O'Brien, J. J. (1991). Variational estimation of the wind stress drag coefficient and the oceanic eddy viscosity profile. *Journal of physical oceanography*, 21(5):709–719.

Zhang, J. and Lu, X. (2008a). On the simulation of m2 tide in the bohai, yellow, and east china seas with topex/poseidon altimeter data. *Terrestrial, Atmospheric & Oceanic Sciences*, 19.

Zhang, J. and Lu, X. (2008b). Parameter estimation for a three-dimensional numerical barotropic tidal model with adjoint method. *International journal for numerical methods in fluids*, 57(1):47–92.

Zhang, J. and Wang, Y. P. (2014). A method for inversion of periodic open boundary conditions in two-dimensional tidal models. *Computer Methods in Applied Mechanics and Engineering*, 275:20–38.

Zhang, W.-Z., Shi, F., Hong, H.-S., Shang, S.-P., and Kirby, J. T. (2010). Tide-surge interaction intensified by the taiwan strait. *Journal of Geophysical Research: Oceans (1978–2012)*, 115(C6).

Zheng, L., Weisberg, R. H., Huang, Y., Luettich, R. A., Westerink, J. J., Kerr, P. C., Donahue, A. S., Crane, G., and Akli, L. (2013). Implications from the comparisons between two-and three-dimensional model simulations of the hurricane ike storm surge. *Journal of Geophysical Research: Oceans*, 118(7):3350–3369.

Zubkovskii, S. and Kravchenko, T. (1967). Direct measurements of some turbulence in the near-water layer. *Izv. Atmos. Oceanic Phys*, 3:127–135.