Development of an Operational System for a Coastal Area on the German North Sea using Artificial Intelligence

Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Christian-Albrechts-Universität zu Kiel

vorgelegt von

José Manuel Fernández Jaramillo

Kiel, July 2014

Referent:	Prof. Dr. Roberto Mayerle
Koreferent:	Prof. Dr. Athanasios Vafeidis
Tag der mündlichen Prüfung:	14.11.2014
Zum Druck genehmigt:	14.11.2014
Der Dekan:	Prof. Dr. Wolfgang J. Duschl

Erklärung

Hiermit erkläre ich, dass die Abhandlung, abgesehen von der Beratung durch meine akademischen Lehrer, nach Inhalt und Form meine eigene Arbeit ist. Diese Arbeit hat an keiner anderen Stelle im Rahmen eines Prüfungsverfahrens vorgelegen. Außerdem erkläre ich, dass diese mein erster Promotionsversuch ist.

Kiel, den 01.07.2014. José Manuel Fernández Jaramillo

To all my ancestors, from my parents to the first eukaryote that populate the oceans. Without their effort in surviving I would not have been able to write this dissertation. And to my wife with whom I am going to populate the planet.

Acknowledgements

I want to express my gratitude to Prof. Dr. Roberto Mayerle for the opportunity to carry out my research in the University of Kiel and his support along this time. To the German state of Schleswig-Holstein that provided the financial support for the development of the present study.

Also I want to thanks the people that helped me with the development of this work, Dr. Talal Etri that have expended countless hours discussing about model development, Dr. Carlos Escobar for his help at the beginning of this work, as well as Dr. Rangaswarni Narayanan, Dr. Sven Jacobsen and Dr. Peter Weppen.

To my colleagues in the institute, specially to my room mates, Katherina and Simon, and also Kahle and many others for their help.

During this journey I counted with the support of many friends, Talal, Marco, Greys, Nestor, Henner, Leslie, Oscar, Joaquim, Julia, Aidan. Thanks for all the time that we have spent together.

My wife Tina, for all her support along this time.

And last but not the least, the people that provide the grounds that allow me to reach this stage. My family, specially my grandparents Germán and Guillermo that supported my ideas and experiments during my childhood, impulsing my curiosity for learning new things every day. My grandmothers Blanca and Adelaida for all the cares and love that have had with me. My parents, they have been always there for me. My uncles and aunts for their patience. Also my former colleagues in

the Computer Center in the National University of Colombia, and my professors Dr. Mauricio Toro and Dr. Carlos Mejía, I have learned a lot from all of them. GE for everything.

Abstract

This thesis presents a methodology of developing a quasi real time flow model in a coastal area. The accuracy of predicted water levels by such a model depends on the ability to assimilate field measurements in the model results. This model improves the water level conditions along the boundary of the flow model by the implementation of a set of neural network short term forecast. The availability of a coastal model that can provide short term predictions is important to manage extreme conditions and reduce the impacts of storms in the region. It is also indispensable to provide early results to ensure the availability of the information with enough time to make decisions based on it.

The model requires integration of several components including flow and wave models, data acquisition, data assimilation, post-processing and data visualization. A description of the components required for such system is included. For the numerical flow model the set-up, calibration and validation of the model in the area of the Dithmarschen Bight in the German North Sea is presented. In the case of data assimilation, changes in the boundary conditions are imposed to improve the model results. For such changes, the information from real time data is employed. Linear regression, neural networks and fuzzy logic are used to improve the estimation of boundary conditions.

The visualization of the information can be done through a web interface. Time series of water levels, wave parameters are available for selected locations, as well as an interactive map with most important model parameters including water levels, significant wave height, flow direction and wind information.

The imposition of water levels along the boundaries of a coastal model by the use of neural networks to assimilate measurements, permits the enhancement of the accuracy of the model results. In the case of the Dithmarschen Bight region the

root mean square error (RMSE) of the water levels can be reduced by 33%. The forecast of water level with neural network does not consume too much computational time; this helps in the reduction of the required time to obtain outputs. In addition, the proposed algorithm to split the calculation of the nesting sequence of hydrodynamic models is able to reduce the computational time to half of that required if normal calculation is done.

Zusammenfassung

Diese Arbeit präsentiert eine Methode für die Entwickung eines Echtzeit-Modells in einer Küstenregion. Die Präzision der vorhergesagten Wasserstände mittels eines solchen Modells hängt von seiner Eignung ab, Feldmessungen in den Modellergebnissen zu assimilieren. Dieses Modell verbessert die Wasserstände an der offenen Grenze des Strömungsmodells, indem es eine Kurzzeit-Vorhersage mit neuralen Netzwerken ausführt. Die Verfügbarkeit eines Küstenmodells, das Kurzzeit-Vorhersagen bereitstellt, ist wichtig, um auf extreme Bedingungen zu reagiren und um die Auswirkungen von Stürmen in der Region zu reduzieren. Außerdem ist es unumgänglich, frühe Ergebnisse bereitzustellen, um sicherzustellen, dass die Informationen zeitnah verfügbar sind und die darauf basierenden Entscheidungen getroffen werden können.

Das Modell erfordert die Integration von mehreren Komponenten, darunter Strömungs- und Wellenmodelle, Datenbeschaffung, -verarbeitung und -visualisierung. Eine Beschreibung der Komponenten, die für solch ein System notwendig sind, ist mit einbezogen. Für das numerische Strömungsmodell der Bucht von Dithmarschen in der Deutschen Nordsee werden die Konfiguration, Kalibrierung und Validierung aufgezeigt. Bei der Datenaufnahme werden die Randbedingungen verändert, um die Modellergebnisse zu verbessern. Für solche Veränderungen werden die Informationen von Echtzeit-Daten verwendet. Lineare Regression, neurale Netze und Fuzzy-Logik werden benutzt, um die Anpassung des Randbedingungen zu verbessern.

Die Visualisierung der Informationen kann mithilfe eine Web Interface erfolgen. Zeitreihen von Wasserständen und Wellenparametern, sind für ausgewählte Orte verfügbar, ebenso wie eine interaktive Karte mit den wichtigsten Modellparametern wie Wasserstände, Signifikante Wellenhöhen, Strömungsrichtungen und Windinformationen.

Die Vorgabe von Wassserständen an den offenen Grenzen eines Küstenmodells durch die Anwendung von neuralen Netzwerken zur Assimilation von Messungen, erlaubt die Steigerung der Genauigkeit der Modellergebnisse. Im Fall des Gebietes der Dithmarscherbucht kann der RMAE „root mean square error" der Wasserstände um 33 Prozent reduziert werden. Die Vorhersage von Wasserständen mit neuralen Netzwerken benötigt nicht viel Rechenzeit; und ermöglich so eine schnelle Ergebnisausgabe.

Der vorgeschlagene Algorithms zur Aufteilung der Berechnungen der Nesting-Sequenz hydrodynamischer Modelle, ermöglicht eine Reduktion der Rechenzeit auf die Häfte der von „normalen" Berechnungen.

Contents

1. Introduction **1**

1.1. Description of the thesis 1

1.2. Objectives 4

1.3. Study area 5

1.4. Methodology 6

1.5. Thesis outlines 9

2. Hydrodynamic and wave models **11**

2.1. Mathematical and physical formulations for flow and waves 11

2.1.1. Navier-Stokes Equations 12

2.1.2. Driving Forces 13

2.1.3. Water level 13

2.1.3.1. Astronomical Constituents 14

2.1.4. Flow and Wave model solvers 16

2.1.4.1. Flow model 16

2.1.4.2. Wave model 18

2.2. Numerical aspects related to the design of the system 20

2.2.1. Spatial Approximation 20

2.2.1.1. Grid resolution 21

2.2.1.2. Nesting 22

2.2.1.3. Domain Decomposition 23

2.2.2. Speeding up the numerical model computation 24

2.2.2.1. Parallelization of the equations 24

2.2.2.2. Adaptive methods 25

3. Operational models and data assimilation 27

3.1. Operational Model 28

3.1.1. Collection of sample data 30

3.2. Data Assimilation 32

3.2.1. Modification of the model formulation 35

3.2.2. Statistical Methods 37

3.2.3. Ensemble methods 38

3.2.4. Optimization of the boundary condition 39

3.2.4.1. Amplification factor at the Open Sea Boundaries . 39

3.2.4.2. Use of a nesting sequence for the Gulf of Lion in the Mediterranean Sea 41

3.2.5. Shooting Method 41

3.3. Time series analysis 44

3.3.1. Autoregressive model 45

3.3.2. Dynamical systems and Iterative maps 46

3.3.2.1. Embedded dimension 48

3.3.2.2. Cross-correlation and its relation with embedded dimension 49

3.4. Artificial Neural Network (ANN) 50

3.4.1. Previous uses of ANN in hydrodynamics 51

3.4.2. Architecture of a feed-forward network 51

3.4.3. Training 54

3.4.3.1. Error backpropagation by the gradient descent method 55

3.4.3.2. Resilient Back-Propagation RPROP 57

3.4.3.3. Simulated annealing 57

3.4.4. Hidden neurons 58

3.5. Fuzzy Logic 59

3.6. Measuring the quality of the fit 61

4. Study area and available data sources 63

4.1. Study area 64

4.2. Available field measurements 68

4.2.1. Bathymetry 68

4.2.2. Water level 69

4.2.3. Wind information 69

4.2.4. Storms 75

4.2.5. Waves 76

4.2.6. Currents 77

4.2.7. Flows from rivers 78

4.3. Summary 79

5. Hydrodynamic Model for the Dithmarschen Bight 81

5.1. General description 82

5.1.1. Nesting sequence for pre-operational model 85

5.2. Sensitivity analysis 87

5.2.1. Warming period 89

5.2.2. Updating the bathymetric information 91

5.2.2.1. Sensitivity of the time step for wave modelling . . . 97

5.3. Calibration 98

5.3.1. Wind drag coefficient with low wind speed 98

5.3.2. Roughness 100

5.3.3. Combination of upgraded model parameters and Wave activation 101

5.3.4. Improve the water level boundary conditions with astronomical constituents from TOPEX 102

5.3.5. Wind drag coefficient at high wind speed 104

5.4. Validation 106

5.5. Summary 107

6. Data assimilation in the Dithmarschen Bight model 111

6.1. Assimilation strategies used in the present work 112

6.2. Translation of information to the boundary 113

6.2.1. Spatial correlation maps 116

6.2.1.1. Correlation maps for water level model outputs . . 117

6.2.1.2. Correlation maps for current model outputs 120

6.2.1.3. Correlation maps for significant waves height model outputs 120

6.2.2. Treatment of corrections along the boundary 122

6.2.3. Dynamic analysis of water level series 124

6.2.4. Linear translation of boundary conditions 126

6.2.5. Non linear approach 128

6.2.5.1. Required parameters for building a neural network. 130

6.2.5.2. Establishing the period length for the observation samples 132

6.2.5.3. Treatment of datasets for an artificial neural network 133

6.2.5.4. Reduction of the sample pool 134

6.2.5.5. Activation function 138

6.2.5.6. Correction size for simulating annealing 138

6.3. Building of local models for water level 139

6.3.1. Water level reconstruction Rochelsteert from Büsum measurements 140

6.3.2. Construction of boundary conditions from translation of an alternative station 141

6.3.2.1. South East boundary condition from Brunsbüttel . 142

6.3.2.2. South East boundary condition from Büsum 142

6.3.2.3. South West boundary condition from Büsum . . . 146

6.3.2.4. South West boundary condition from Cuxhaven . . 147

6.3.2.5. North West BC from Büsum 150

6.3.3. Forecast at the open sea boundary 152

6.3.3.1. Water level forecast at Rochelsteert using the nesting sequence with Delft3D model 153

6.3.3.2. ANN for forecasting water levels at Rochelsteert . . 153

6.3.3.3. ANN combined with fuzzy logic for forecasting water levels at Rochelsteert 157

6.4. Ensemble of forecast at Rochelsteert 159

6.5. Sensitivity to the inputs of an ANN 161

6.6. Corrected model results after assimilation 165

6.7. Summary 168

7. Hydrodynamical Operational Model for the Dithmarschen Bight 173

7.1. Some considerations about model engine and the operational model design 174

7.1.1. Domain decomposition vs. nesting in the operational model 174

7.1.2. Parallelization of the nesting sequence 175

7.1.3. Adaptive time step 177

7.1.4. Selection of number of cores for SWAN simulations. 179

7.2. Execution of the operational model 180

7.3. Dissemination of the information 182

7.4. Summary 185

8. Conclusions and recommendations for future work 189

8.1. Conclusions 190

8.2. Future work. 195

List of Figures

1.1. Wadden Sea and North of Germany 3

1.2. Study area 7

2.1. Nesting vs. Domain Decomposition data flow 22

3.1. General components of an Operational Model 30

3.2. Integration Domain for a PDE, Ω denotes the region where the system of equations are solved, Γ_b is the boundary condition. 33

3.3. Data ensemble of several model execution 38

3.4. Solution of a BVP by the use of Finite Differences and the equivalent IVP by the shooting method 44

3.5. Stability cases (a. and c. are adapted from Strogatz [1994]) 47

3.6. Poincaré-Bendixon 48

3.7. General Neuronal Network 52

3.8. Single processing unit or neuron 54

3.9. Scatter plot of a non linear set. 60

3.10. Water level error measurement. 61

4.1. Bathymetry of the study area with the location of measurements . . 65

4.2. Bathymetry information (Year of availability) for the German area (top) and a closeup to the Dithmarschen Bight (bottom) 70

4.3. Wind Roses for available stations data from 2008 to 2011. Bruns-büttel, Cuxhaven, Rochelsteert 73

4.4. DWD and Wind measurements at Büsum 74

4.5. Wind correction 74

4.6. Current measured by Electromagnetic Current Meter 78

4.7. Mean Elbe flow 79

5.1. Grids used in the nesting process. from Etri [2007] 82

5.2. Nowcasting system description pre-operational version 85

5.3. Wind conditions for selected periods. Left the calm period (from 01.04.2008 to 01.05.2008). Right the strong period (from 10.02.2008 to 10.03.2008) 88

5.4. Warming period for CSM model 90

5.5. Difference in bathymetry: positive erosion, negative deposition . . . 92

5.6. Effects of changing the bathymetry available on 1996 to an updated with information until 2009 in the shape of the flow tide, also compared with the astronomical tide obtained from a harmonic analysis of measurements at Büsum 93

5.7. Effect of the time step for wave model, in the left the difference in the water level between using 60 and 120minutes and 60 and 30 minutes. In the right a zoom showing the significant wave height obtained by using 30,60 and 120minutes 96

5.8. Model results versus Measurements during the validation period in July 2008 108

6.1. Maximum correlation coefficients and time delay in minutes for water level at Büsum 118

6.2. Maximum correlation coefficients and time delay in minutes for water level at Büsum, Cuxhaven, Rochelsteert and Brunsbüttel 119

6.3. Maximum correlation coefficients and time delay in minutes for velocity at Büsum, Cuxhaven, Rochelsteert and Brunsbüttel 121

6.4. Maximum correlation coefficients and time delay in minutes for Hs at Büsum, Cuxhaven, Rochelsteert and Brunsbüttel 123

6.5. Bathymetry profiles at boundary and their location 125

6.6. Space phase for two tidal gauge stations, left few phase points allow to see a more clear orbit, in the right side more phase points permits to observe how the orbit moves along a virtual axis -around 45° in the plot- 127

6.7. Water level correction at Brunsbüttel, Büsum and Cuxhaven 129

6.8. Autocorrelation Water Level at Büsum 132

6.9. Sampling strategy for an ANN 133

6.10. Example of the values contained in a single input set for a neural network 134

6.11. Filtered samples using different values of $cos\,(\alpha)$ 137

6.12. Neuronal network with 30 inputs and 12 hidden neurons W values . 141

6.13. South East BC from Büsum (ordered by descending quality of the outputs). 145

6.14. South West BC from Büsum. In the time series ▮ represents the neural network output, ▮ the measurements and ▮ the error calculated as the difference between the measurement and the network output 148

6.15. South West BC from Cuxhaven model 1 without dry state 151

6.16. North West BC from Büsum 152

6.17. Comparison of water levels obtained from different model formulations with the measurements at Rochelsteert in the North-West boundary of the DBM model 156

6.18. Ensemble of: a) four ANN models for hindcast and b) forecast 12 hours 160

6.19. Sensitivity analysis for hindcasting 164

6.20. Water levels at Büsum from original and corrected boundaries, compared to the measurements 167

6.21. Water level correction 169

6.22. Improved model results versus Measurements during the validation period in July 2008, see Figure 5.8 to compare results 170

7.1. Parallelization of a nesting execution 176

7.2. Evaluation of computational time for a parallelized nesting sequence 177

7.3. Automatic time step adjustment 179

7.4. Computational time of parallel realizations of SWAN relative to the serial version (left side) and the percentage of relative speed up of the computational time when a processor is added (right side) . . . 180

7.5. Nowcasting system general description 181

7.6. Sequence of model running showing the original forecast (NF), the hindcast (H) and the improved forecast (IF). 182

7.7. Main window of the screen 185

7.8. Details of the available options in the left menu 186

7.9. Zoom to the North Sea, showing water level map 186

7.10. Water level time series at Büsum 187

7.11. Wind maps in combination with water level time series at Büsum . 187

7.12. Wind fields at the North sea 187

7.13. Camera attached to the system (Operated by X-H2O) 188

List of Tables

4.1. Location of available Tide Gauges 71

4.2. Douglas Sea Scale, classification of the state of the sea 77

5.1. General Parameters used in each model 86

5.2. Error magnitude and time required 91

5.3. Comparison of average error in ebb and flow tide (cm) for different bathymetries comparing model results without imposing wind and astronomical tides 94

5.4. Comparison of average error in water level (cm) at different locations 94

5.5. Water level averaged errors and RMSE for the four stations considered in the evaluation of DBM model with low wind speed 96

5.6. Averaged errors and RMSE of water levels for the four stations considered in the evaluation of DBM model. Wind drag coefficient for low wind speed 99

5.7. Averaged errors and RMSE for the four stations considered in the evaluation of DBM model. Roughness coefficient for low wind speed 100

5.8. Averaged errors and RMSE for the four stations considered in the evaluation of DBM model. Effect of activate the wave coupling for low wind speed and using the bathymetry of 2009 102

5.9. Averaged errors and RMSE for the four stations considered in the evaluation of DBM model 103

5.10. Averaged errors and RMSE for the four stations considered in the evaluation of DBM model changing the wind drag coefficient with high wind speed 104

5.11. Averaged errors and RMSE for the four stations considered in the evaluation of DBM model with high wind speed 105

5.12. Averaged errors and RMSE for the four stations considered in the evaluation of DBM model 106

5.13. Error in DBM with nesting and corrections for July 2008 107

6.1. Reduction of sample pool for ANN training 136

6.2. Regressive Model for evaluation of South East boundary condition from Büsum 143

6.3. Errors for neural network evaluation of South East BC from Büsum values 144

6.4. Autoregressive Model for evaluation of South West BC from Büsum 146

6.5. Errors for neural network evaluation of South West BC from Büsum values 147

6.6. Errors for neural network evaluation of South West BC from Büsum values excluding samples where cell is dry 149

6.7. Autoregressive Model for evaluation of South West BC from Cuxhaven149

6.8. Errors for neural network evaluation of South West BC from Cuxhaven values . . . 150

6.9. Errors for neural network evaluation of South West BC from Cuxhaven values excluding samples where cell is dry . . . 150

6.10. Errors for neural network evaluation of North West BC from Büsum values . . . 151

6.11. Errors for neural network of North West BC forecast (0h forecast corresponds to a hindcast) . . . 154

6.12. Errors for neural network evaluation of North West BC forecast from Büsum values . . . 158

6.13. Standard deviation (cm) for the sensibility analysis of the hindcast of Rochelsteert water levels by the use of Büsum water levels and wind information . . . 163

6.14. Standard deviation (cm) for the sensitivity analysis of the 24 hours forecast of Rochelsteert water levels by the use of Büsum and Cuxhaven levels and wind information . . . 165

6.15. Averaged errors and RMSE for the four stations considered in the evaluation of DBM model in cm . . . 166

6.16. Error in DBM with nesting and corrections from assimilation for July 2008 . . . 167

Chapter 1

Introduction

1.1. Description of the thesis

For centuries, people have been living in coastal areas around the world. Nowadays, about 75% of the megacities with a population of more than ten million and approximately 45% of the total world population live near the coast [Gelpke and Visbeck, 2010]. As a result, the demand on coastal resources causing pollution and hazards to properties in the coastal line and problems with navigation, have been increasing. The coast of the State of Schleswig-Holstein in northern Germany, the area of this investigation, has been subjected to major pressures in the last century. For example, the severe storm of 1962 has destroyed several dikes and flooded about 12,500 ha near Hamburg, killing some 350 people [Stadelmann, 2008]. As a consequence of climate change and sea level rises it is expected that the frequency of such events would increase [Hofstede,1997].

The accident of the Schiff MS Pallas transporting about 142 tons of diesel and 580 tones of heavy fuel oil on the coast of Schleswig-Holstein at the end of the year 1998 [Maier, 1998] is another example of coastal hazard. Sixty tons of oil were released into the Wadden Sea [Liu, 2000]. Although remedial measures could be taken, the consequences of such accident for the sensitive environments could have been serious and lasting [Fleet, 2000]. This coastal area is also the access to the Harbour of Hamburg and the Kiel Canal connecting the North and Baltic Seas (See Figure 1.1). There is intense traffic of ships. Furthermore the area is part of the National Park of the Wadden Sea which is one of the main sensitive environments along the German North Sea [Kohlus and Agatha, 1998]. A better understanding of the flow and transport processes, and the development of early warning systems for extreme events [Krzhizhanovskaya et al., 2011] is therefore vital for reducing the risk of hazards and thus to save lives and properties [Tingsanchali, 2012]. This PhD research deals with the development of an operational model for the prediction of water levels, currents and waves along the Dithmarschen Bight on the German North Sea coast (see Figure 1.1). Emphasis is given to the data assimilation techniques for improving the accuracy of coastal models for simulation of flow.

The hydrodynamics in coastal areas are driven mainly by the combined effects of tides, winds and waves. The understanding of these processes started long time ago with the observation of the conditions at sea. In the early stages the prediction of storms relied mainly on wind observations. Due to the poor mathematical application such descriptive models used were not useful [Delmar-Morgan, 1959]. This was followed by slightly more precise observation based on comparable measurements of wind speeds and water levels. With the intensification of data collection, empirical models relating water levels to wind speeds were built leading to better predictions. Physical models also helped to improve the understanding of the physical processes. But, they are usually restricted to smaller areas and the simulation of wind driven conditions is extremely difficult. Only with the development of digital computers, numerical models started to play a decisive role in the

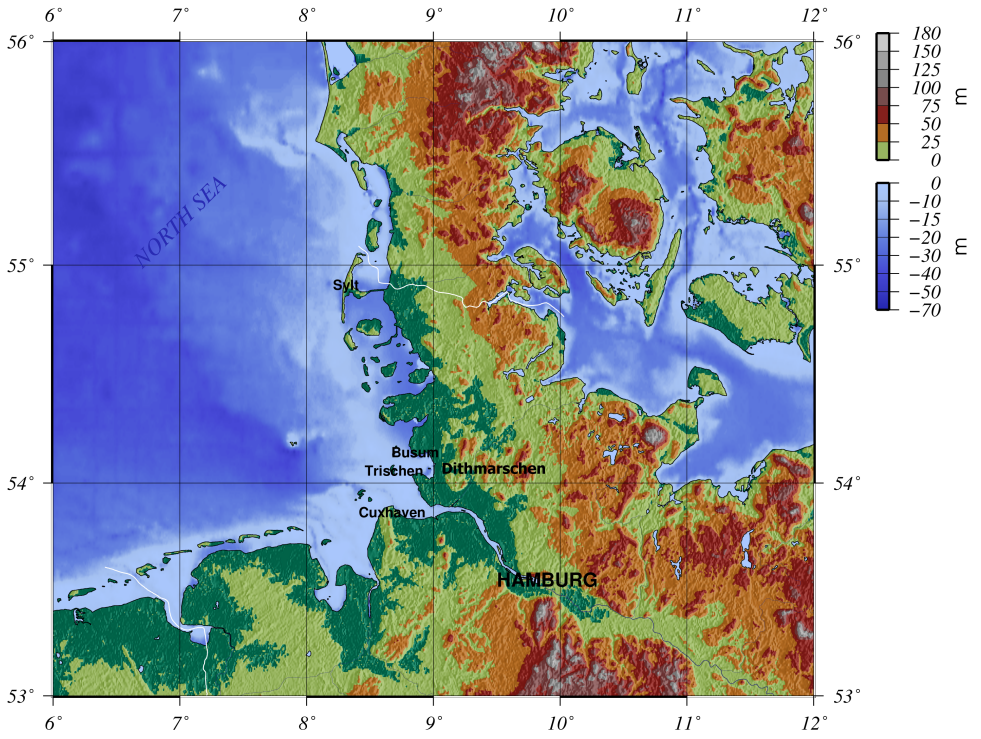


Figure 1.1: Wadden Sea and North of Germany

prediction of water levels, currents and waves. Flow and wave models are becoming important tools. Computational capabilities have been growing exponentially in recent times, enabling the handling of more complex problems in shorter time. During the last decades, data acquisition has also evolved from analogue plots to real time digital information. The improvement of the digital communications has allowed accessing data from remote areas in real time from operational stations located in different places simultaneously. As a result, new applications involving data monitoring, and more accurate nowcasting and forecasting based on data assimilation techniques are being developed.

In this study, emphasis is given to data assimilation techniques particularly for coastal area models. An existing large-scale model sequence covering the northwest European Continental Shelf, the German Bight and the Dithmarschen Bight is used. Wind forcing stem from atmospheric models operated by the German Meteorological Service. The flow and wave models are updated with recent mea-

surements and the use of improved flow and wave solvers. An operational model implementing a new data assimilation technique is developed. Several operational tidal gauges along the coast were used to improve the conditions specified along the open sea boundaries of the Dithmarschen Bight Model. The resulting operational model implements filters to reduce the presence of errors and outliers. It also includes data ensemble to reduce the deviations in the improved forecast algorithm. The data assimilation technique lead to improved hindcasts and forecasts of water levels on the Dithmarschen Bight. The system can be embedded in an early warning system for supporting authorities in the decision making process to prevent disasters.

1.2. Objectives

The overall goal of the thesis is the development and testing of a data assimilation technique particularly suited for setting-up operational models of coastal areas. The methodology proposed here should be applicable to coastal areas worldwide. As the accuracy of coastal area models depends very much on the proper prescription of the values specified along the open boundaries, the technique proposed here consists of adjusting these values on the basis of measured values within the model domain. This is done with regards to real time measurements available at several locations within the model domain. In this study the data assimilation technique proposed is applied to the Dithmarschen model that belongs to the nesting sequence covering the entire North Sea. Several operational tidal gauges along the coast are used to improve the water levels specified along the open sea boundaries of the Dithmarschen Bight Model.

To achieve the target of the thesis the following studies are carried out:

1. Update of the existing set of models for the study area. The direct coupling between the flow and wave models is included to take into the account the flow-wave interaction. More recent bathymetrical surveys and higher spatial and temporal resolution from the atmospheric models, are included. Calibration and validation of the large-scale model nesting sequence using water level measurements at several locations are carried out.

2. Development of a data assimilation technique that is able to deal with nonlinear processes. A combination of Artificial Neural Networks (ANN) and fuzzy logic is used to handle the nonlinearities particularly during drying and flooding. The ensemble of several ANN enabled the open sea boundary conditions to be improved leading to better predictions.

3. Establishment of functions relating to the information within the domain and at the open sea boundaries of the coastal model.

4. Design and set-up the operational model for the study area by integrating the numerical models, data observation, and data assimilation strategies.

5. Assessment of the accuracy of the system for hindcasting, nowcasting and forecasting.

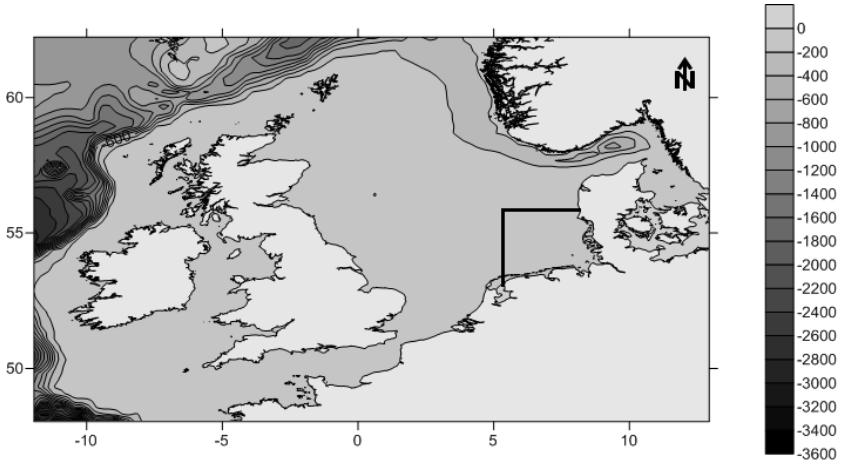
1.3. Study area

The study area is the Dithmarschen Bight located in the south-eastern part of the German Bight as is shown in Figure 1.2. The area was selected due to its relevance to coastal protection, availability of operational tidal gauges and existence of applicable process based models. The area of the Dithmarschen Bight stretches from the mouth of the Elbe estuary in the South to the Eiderstedt peninsula on the North and forms the southern compartment of the Wadden Sea. The domain is

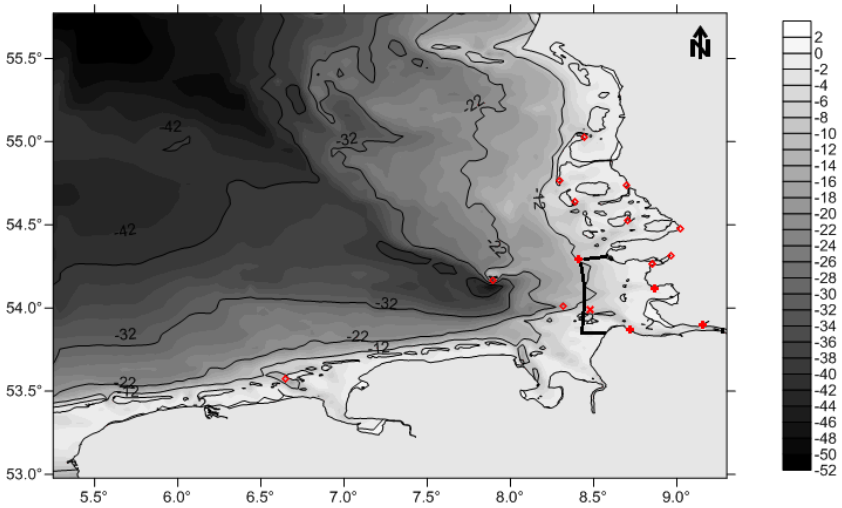
bordered by the coastline in the east and the 8° 25' meridian in the West which coincides approximately with the 13m isobath of the open North Sea. It borders the States of Schleswig-Holstein and of Lower Saxony at the South-East corner of the German Bight (see Figure 1.1). About half of the area consists of tidal flats which fall dry during the ebb phase. The median grain size varies from 90 to 110μm in the tidal channels [Vela-Diez, 2001] to 110 to 250μm in the tidal flats Reimers [2003] Figge [1981]. The interchange of water between the open sea and the near coastal areas takes place through several tidal channels. The flow in the area is driven mainly by the combined effects of tides, waves and winds. The fresh water discharges of the Elbe River, ranging from about 400 to 1200m^3/s also affects the flow particularly in the vicinity of the mouth. The mean tidal range in the area varies between 3.1m and 3.4m. The currents in the tidal channels can reach up to 2.8m/s at spring tides [Toro et al., 2005]. The winds are predominantly from the West reaching speeds of up to about 15m/s. Due to strong wind conditions, storm surges can increase the water level significantly. During the storm in 1967 the water level exceeded 5m above the normal null [EAK,2002]. Wave heights of up to 4m were observed in the outer region of the investigation area. Under normal conditions, however, the influence of waves on the flow is moderate on the tidal flats and negligible in the tidal channels. Wave breaking generally occurs along the edge of the tidal flats [Mayerle et al., 2005]. The analysis of information from different buoys shows a high dissipation of the energy in the transition zone between the open sea and the tidal flats [Toro et al., 2005].

1.4. Methodology

In this study a data assimilation technique particularly suited for coastal area models is developed and applied. As coastal area models cover only a part of the adjacent sea under investigation, open sea boundaries for the enclosed sea are in-



a) North Sea indicating the region of the German Bight



b) German Bight, indicating the Dithmarschen Bight area

Figure 1.2: Study area

troduced to limit the size of the modelled domain and hence reduce computational effort. Open sea boundaries are virtual "water-water" boundaries representing the influence of the sea area surrounding the modeled domain. The predictive capability of process-based models depends very much on the accurate description of the environmental forcing along the open sea boundaries. The type of open sea environmental forcing applied in practice often depends on the available data. Measured values or the results of simulations performed with larger-scale models covering the adjacent sea area are usually used for this purpose [Mayerle et al, 2005]. In this study a data assimilation technique which consists of adjusting the values imposed at the open sea boundaries using water levels measurements within the model domain was developed. The strategy adopted comprises of several steps. Initially a set of existing models for the study area are improved to increase its quality based in newer bathymetry, and more accurate wind field forecasts. With these models an analysis of the spatial behavior of tides, water levels and waves for the area in question is carried out on the basis of correlation maps. This helps in the identification of regions along the model open boundaries which are better represented by given observation cells within the domain thus facilitating the zoning of representative areas. On the basis of this information, functions relating grid points along the open sea boundary and within the domain are obtained to enable the nowcasting. Finally functions are built to reconstruct the missing information (hindcast) and to deliver forecasts. In this study the functions are created using neural networks and fuzzy logic taking into account the model water level and wind direction during the learning process of the network (calibration of the equations) and with those functions then the measurements are used for correcting the boundary conditions of the model.

1.5. Thesis outlines

The thesis consists of eight chapters. In Chapter 2 some concepts of fluid models and numerical approaches to calculate the flow fields are provided, this will help in the discussion of the design of the system in Chapters 5 and 7. In Chapter 3 a review of the data assimilation techniques usually adopted in conjunction with the process based models for simulation of flow and wave models is summarized. The required concepts and tools including an overview of the components of operational models are provided. In Chapter 4 details of the study area with and inventory of the available data used in the present study is presented. A review of the previous investigations carried out in the coastal area is also provided. Chapter 5 contains a description of the original settings of the nesting sequence covering the north-west European Continental Shelf, the German Bight and the Dithmarschen Bight. The approaches adopted to improve to the existing model settings and details of the coupling of the flow and wave models are described. Results of the calibration and validation of the models for simulation of flow and waves using field measurements is also summarized. Chapter 6 introduces the data assimilation technique employed in this study for nowcastings and forecastings. Functions relating measured water levels at different locations with the conditions imposed along the open sea boundaries of the models are derived. To handle nonlinearities, the effectiveness of artificial neural networks (ANN) is investigated. In Chapter 7 the set-up of the operational model for the study area is presented. The transfer of data between the model solver and the data assimilation module is described. Conclusions and recommendations for further investigations with emphasis to data assimilation techniques for hydrodynamic models are summarized in Chapter 8.

Chapter 2

Hydrodynamic and wave models

This Chapter is provided a brief description of the physical processes that are important in the hydrodynamics of coastal areas. This part also illustrates the capabilities and limitations of a process based model. The understanding of the physical processes is also a requirement in the selection of the parameters passed to an artificial neural network in order to obtain a reasonable result. The presence of **linear** and **nonlinear** relations among hydrodynamic variables has an important role in the selection of different approaches to obtain the flow conditions.

2.1. Mathematical and physical formulations for flow and waves

In this section the main terms involved in the hydrodynamics are described. Furthermore, a description of the capabilities of the solvers for the flow and wave

models are given in order to illustrate how linear and nonlinear terms deal with the water level estimation.

2.1.1. Navier-Stokes Equations

The Navier-Stokes equations are the general formulation of the conservation of momentum applied to parcel of fluid. This can be done over a control volume or an infinitesimal fluid element, leading to integral or differential representation of the equations. (Batchelor [1967] and Anderson [1995])

The Navier-Stokes equations have nonlinear terms due to the convective acceleration. Finding a general solution for the Navier Stokes equations is not possible [Constantin, 2000]. Except for simple flows, analytical solution is not possible. For complex cases the use of numerical solution is required. However, even with numerical solutions some simplifications are made. One example is the linearization of the nonlinear equations, which permits using idealized numerical techniques in the solver.

The roughness of the sea floor introduces friction at the bottom and is responsible for the bed shear stress. Roughness depends not only on grain size at the sea bed, but also on bed forms as dunes, mega-ripples and mini-ripples [Rijn, 1993]. The shear stress terms for the friction are responsible for nonlinear speed gradients in the boundary layer where velocity gradients are present. At higher depth, the Euler equations are valid [Novak, 2010] and for constant density and low speeds the current displays more linear behavior. This distinction explains why nonlinear approaches are required to create the correction equations for the data assimilation. Also depending on the value of the friction mainly in shallow conditions, the

occurrence of the ebb tide can be brought forward or backward some minutes.

2.1.2. Driving Forces

There are three main agents that can trigger the motion of sea water: gravity, air pressure and wind. Gravity is one of the most important, because it is the source of potential energy due to gradients in the level in the water surface level. It also generates tidal oscillation in the water level due to the interaction of gravitational forces between the Earth, the Moon and the Sun. Density currents are related as well with the gravity as this force makes denser fluid sink under lower density fluid. If density is considered as a constant then its effect is negligible in the solution of the flow. On the other hand the wind, which is generated by a gradient in the air's pressure field [Rampanelli et al., 2004], also provides energy to a fluid through shear stress at the surface. Depending on duration, pattern, direction and length of action over the surface fetch, several effects are be induced in the form of waves, storm surge and seiche among others [Kamphuis, 2000]. If a sufficiently large area is exposed to differential air pressure it is transferred to the water level as a barometric surge [Kamphuis, 2000]. The change in the water level due to the surges is proportional to the square of the wind speed and to the gradient of pressure over the water surface.

2.1.3. Water level

The water level in ocean and coastal areas is characterized by the oscillations associated with a variety of acting factors. Each one of those elements imposes an associated frequency. The most important physical processes are gravitational

forces, wind stress and air pressure. Gravitational forces depend on the relative position of the Earth, the Moon and the Sun, they create an oscillatory movement that is acting permanently over the water body. Air pressure and wind fields impose forces that are acting only during short periods, but that can modify the water level enough to create disasters at the shoreline.

2.1.3.1. Astronomical Constituents

Shureman [1958] developed an elegant approach to the tide prediction. By the combination of the potential provided by the gravitational forces of the Moon and the Sun over each particle on the earth he was able to construct a linear equation that describes the most important elements of water level variation in the long term. His methodology assumes a number (up to 140) of artificial satellites orbiting the Earth in a circular orbit around the earth. Each of those bodies has its unique mass and orbital period. The linear combination of those individual forces is able to reproduce the effect of elliptical orbits, Earth's precession or the declination of the Moon and the Sun among other astronomical movements.

$$h\left(t_j\right) = \sum_{i=0}^{l} a_i b_{ij} cos\left(\omega_i t_j + \left(V_0 + u\right)_{ij} - \kappa_i\right) \tag{2.1}$$

The terms represented in this equation are: l: number of frequencies used in the analysis, $h\left(t_j\right)$: water level at time t_j, a_i: empirical amplitude, b_{ij}: theoretical amplitude associated to the frequency ω_i, ω_i: angular frequency in degrees per hour $(^\circ/h)$ corresponding to the astronomical constituent i, $\left(V_0 + u\right)_{ij}$: It is the argument's phase of the astronomical constituent i and κ_i: phase lag or epoch of the constituent.

The angular frequency of the astronomical component is given in degrees per hour, this has normally a physical meaning as for example "S1" with a value of $15°/hour$ or equivalent to one pass of the sun over the observer's meridian once per day. The term b_{ij} is the theoretical amplitude associated to the astronomical component. It takes into account mass factors, relative distance, latitude, obliquity and amplitude constant. The port age considers the time shift of the tidal bulge from the pass of the astronomical constituent in front of the observer's meridian. This shift is created by the orographic shapes that alter the flat bottom assumption. The argument's phase of the astronomical constituent represents the reference location where the constituent is located at a given time. This is a continuous function of the time, but for simplicity in the original theory this is taken as a constant for a time frame. In Shureman [1958], tables to help with the estimation of these parameters are provided. Some computational tools like Delft3D or Xtide Flater [1996] follow this strategy. Gnotide [Cardona and Fernandez Jaramillo, 2001] instead calculate the argument's phase for every time step involved in the tidal analysis or the tidal prediction. It is required to take into account this limitation in the design of the operational model. If the reference time is kept constant, after two years the astronomical boundary conditions do not produce correct results.

With linear approach astronomical harmonics are good when nonlinear effects are neglected. The linear assumptions are valid where the water is deep enough to avoid frictional effects and the wind forces are limited. The astronomical constituents can explain most of the water variations and are good in evaluating water level and, the occurrence time of ebb and flow tide in such cases. As an example, in Buenaventura in the Pacific Ocean's coast of Colombia where continental slope is pronounced, harmonics theory represents 98% of the level fluctuation [Cardona and Fernandez Jaramillo, 2001]. As this method neglects the meteorological effects, there is a reduction in the ability to make good forecasts when strong winds prevail.

2.1.4. Flow and Wave model solvers

The model solver is the component of the system that is responsible to calculate water level, currents and waves for a simulation period. In this system two different models are coupled, so that they interchange information during the computation. Delft3D for flow and SWAN for waves are implemented. Those models have been used and tested in several projects along the North Sea coast in Germany, showing good results.

2.1.4.1. Flow model

Delft3D is the flow model solver used in the present work. It has been developed by Delft Hydraulics and the documentation for it is available in [WL , 2003a]. In this section a brief view of the flow processes that are considered in the solution is given. The hydrodynamical module "'Delft3D-FLOW"' simulates two dimensional (2D depth-integrated equations) as well as three-dimensional (3D) unsteady flow resulting from tidal and meteorological forcing. The development of this tool has been made for cases where horizontal processes are dominating in comparison with the vertical ones. In the modeled area, vertical density is considered homogeneous. Also the shallow water and the Boussinesq (Eddy viscosity) assumptions are applied. Delft3D includes a robust set of mathematical formulations that are used in the operational model, some which are taken from the user guide [WL , 2003a] are listed next:

1. Free surface gradients (barotropic effects).
2. The effect of the Earth's rotation (Coriolis force).

3. Tidal forcing at the open boundaries.
4. Space and time varying wind shear-stress at the water surface.
5. Space varying shear-stress at the bottom.
6. Space and time varying atmospheric pressure on the water surface.
7. Time varying sources and sinks (e.g. river discharges).
8. Drying and flooding of tidal flats.
9. Tide generating forces.
10. Vertical exchange of momentum due to internal waves.
11. Wave induced stresses (radiation stress) and mass fluxes.
12. Wind driven flows including cyclonic / hurricane / typhoon winds.

Density is determined by sediment in suspension, temperature and salinity concentration (affected by temperature and fresh water input from rivers or rainfall). Delft3D also includes some other features related to density driven currents that affect the model result more or less depending on the spatial scale. Those features are:

1. Water with variable density (equation of state).
2. Horizontal density gradients in the pressure (baroclinic effects).
3. Transport of salt, heat and other conservative constituents.
4. Heat exchange through the free surface.
5. Evaporation and precipitation.

2.1.4.2. Wave model

The Delft3D-wave model is based on SWAN available under GPL license. The basic technical information is described in Chapter 7 of "Delft3D-wave user's guide" [WL , 2003b] and in more detail in "SWAN scientific and technical documentation"[Del, 2009].

The computational grid of SWAN uses four dimensions, x and y in spatial space and, σ and θ in the spectral and directional space respectively to describe the action balance equation. In SWAN, the waves are described with a two-dimensional wave action density spectrum $N(\sigma, \theta)$ [WL , 2003b], which is linear. The use of this linear relation even in nonlinear conditions presented in areas like surf zone is justified by the ability to predict the spectral distribution of the second order moment of the waves [WL , 2003b]. The limitation of this assumption is that the statistical description of the wave has limited accuracy. The action density spectrum $M(\sigma, \theta)$ is preferred to the energy density spectrum $E(\sigma, \theta)$ because $N(\sigma, \theta)$ (calculated as $N(\sigma, \theta) = E(\sigma, \theta)/\sigma$) is conserved in the presence of currents. The evolution of the wave spectrum is described by the spectral action balance equation(Hasselman et al,1973):

$$\frac{\partial}{\partial t}N + \frac{\partial}{\partial x}c_x N + \frac{\partial}{\partial y}c_y N + \frac{\partial}{\partial \sigma}c_\sigma N + \frac{\partial}{\partial \theta}c_\theta x N = \frac{S}{\sigma} \tag{2.2}$$

The terms on the left-hand side represent the change of action density. The first one is the local rate of change in time, $\frac{\partial}{\partial x}c_x N$ and $\frac{\partial}{\partial y}c_y N$ represent propagation of action in geographical space (c_x and c_y are propagation velocities), $\frac{\partial}{\partial \sigma}c_\sigma N$ represents the shifting of the relative frequency due to variations in depths and currents (c_σ propagation velocity in σ-space) and $\frac{\partial}{\partial \theta}c_\theta x N$ represents the depth-induced and

current-induced refraction (c_θ propagation velocity in space). The expressions for these propagation speeds are taken from the linear wave theory [WL , 2003b] (refers to Whitham [1974], Mei [1989] and Dingemans [1997]). $S(\sigma, \theta)$ corresponds to the source term that includes generation, dissipation and nonlinear wave-wave interactions.

The source term S in the action balance equation is expressed in terms of energy density and represents the effect of several physical phenomenon. As described in the user guide of SWAN, the energy transfer from wind to the waves is carried by a resonant [Phillips, 1957] and a feed-back [Miles, 1957] mechanisms. It gives a combination of a linear and exponential growth given by:

$$S_{in}(\sigma, \theta) = A + BE(\sigma, \theta) \tag{2.3}$$

where A and B depend on wave frequency and direction and wind speed and direction.

Dissipation of the energy is given by the combination of three processes: white-capping, bottom friction and depth-induced breaking. In the formulation of such processes non linearities are also present [Komen et al., 1996].

2.2. Numerical aspects related to the design of the system

As was pointed out before in section 2.1.1, an analytical solution for the Navier-Stokes equations is not possible except in very simple cases. This creates a scenario for the use of the numerical solution. Numerical methods are based on finding a discrete solution for complex problems, usually by the use of numerical formulations such as finite differences or finite elements methods. The discretization can be spatial and/or temporal. For temporal discretization the "time step" Δt is the main parameter. For the spatial discretization, a grid, where the calculations for the model values are made, should be defined. In the simplest case a rectangular grid (2D problems) or a cuboid (3D problems) where the spatial increase in each dimension (Δx,Δy,Δz) is provided.

The spatial grid is one of the most important decisions that should be taken into account, because this controls the resolution of the simulation and the time step in order to guarantee the stability. This election has consequences in the global performance of the final model in terms of quality and speed of finding a solution.

2.2.1. Spatial Approximation

For the development of the model, the spatial domain can be defined in different ways in order to get a representation of the reality. Aspects like the grid resolution, the use of a single grid, a nesting sequence of domains, domain decomposition should be defined when a model is built for a specific area. The selected grid representation for the evaluation of the model for the domain, has an impact in

the quality and also in the performance of the model calculation. Also this must be considered in the design of the automation of the models. Next, some aspects that can affect the final set up of the system are presented.

2.2.1.1. Grid resolution

The grid defines the desired spatial resolution and imposes restrictions to the coarser time resolution that can be used. The Courant number relates the maximum time step with the inverse of the spatial resolution [Courant et al., 1928]. Then, when the cell side is halved, the courant number forces to double the required time steps as the minimal time step is also halved. By reducing the spatial resolution the number of points to be solved is also increased by a power of 2. Taking both together it can be seen that halving dx, the number of equations to be solved -and the total computational time- is increased in an order of 2^3.

A fine grid is required when small areas are to be represented for the processes that is in study or when large gradients in the variables are expected. Some examples that require fine grids are: narrow channels, rough bathymetry, vertical temperature distribution, density currents in estuaries and vertical distribution of velocities when $3D$ equations are used. For uniform fields of flow a coarser grid can be used. When there are areas where both uniform and high gradients are expected in the same domain, the use of a uniform discretization is not recommended because of the high simulation time required. Instead, selecting a mechanism that allows the coexistence of coarse and fine grid resolution is preferred. There are several strategies to do this. One is applying a non uniform grid, another approach is to split the domain in two or more sub-domains and then use nesting or domain decomposition to connect them, or a combination of both. The two concepts are elucidated in more detail in the following paragraphs.

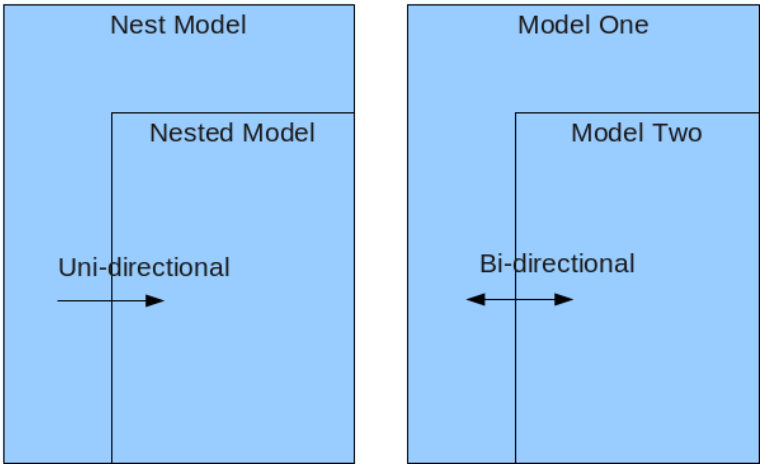


Figure 2.1: Nesting vs. Domain Decomposition data flow

2.2.1.2. Nesting

The nesting approach makes use of a nest model -generally with a coarse grid- and from its results, the boundary conditions for one or more embedded models (nested models) -generally with a finer grid than the nest- are imposed. The nesting follows a sequence doing the calculation in series. It is, that the nest model is evaluated first and then with the outputs of it, it is possible to evaluate the finest one. Also, the information is going from the coarse to the fine grid but not in the opposite way (See Figure 2.1). These characteristics of nesting have the advantage that the coarse domain is independent of the finer domain. When several scenarios in the nested model are tested but without modifications in the nest, only the nested ones need to be evaluated again. The design of the system takes benefit of this characteristic from nesting. The time step in the coarse grid can also be kept bigger, leading to a faster calculation (Delft3D allows to interpolate the time series at boundary condition). On the other hand, if only one calculation is planned, nesting cannot benefit from modern machines where multi-core is almost present in all new computers. From the modeling part, the lack of communication of results from the nested model to the nest, limits its use when bi-directional feedback between the domains is important. This lack of interchange, also forces to create domains covering a bigger area with a fine grid

than what is possible with domain decomposition. This approach is used in the models included in the present work.

2.2.1.3. Domain Decomposition

In the domain decomposition approach, two (or more) independent domains are evaluated at the same time. Those models should either overlap or have a common boundary. One domain may have a coarser grid than the other, but this is not necessary. In general, the overlapping nodes do not need to be located at the same positions and an interpolation method is used for providing the input value to each domain from the results of the other. However in Delft3D shared grid nodes are required along each boundary. This procedure is done in parallel: the numerical model calculates each domain for one time step and waits for the other domain to interchange the results and then continues with the next time step. As a result it takes benefit from multi-core processors, but to get the maximum performance both models should take approximately the same time to finish one iteration. Otherwise the total simulation time is going to be determined by the slowest model.

If several scenarios are required in only one of the two domains, both must be evaluated again because of bi-directional communication. Although a time space interpolation can be included in the computational tool, Delft3D forces to employ the same time step in all the sub-domains. Thus the smallest cell in any sub-domain controls the global time step. In comparison with nesting, domain decomposition permits to develop a smaller sub-domain around the region of interest. This is why as communication is carried in both directions (see Fig. 2.1), the effect of boundary conditions in theory is minimal.

2.2.2. Speeding up the numerical model computation

Model calculation can be accelerated by the use of algorithms that split the work among several processors or by taking the coarsest possible grid both in time and space. Two strategies that can be used by models to work faster are the parallelization of the equation and adaptive time step. The difference between these two strategies with nesting and domain decomposition is that they are transparent to the modeler. He only needs to define a single grid and the optimization is done internally by the numerical model solver.

2.2.2.1. Parallelization of the equations

Multi-core computers are widely available nowadays. Taking advantage of this computational power allows to narrow the required time to evaluate a model. Instead of splitting the domain it is possible to use algorithms that solve the matrixes delivering the task automatically to several processors. This is the strategy available in SWAN [Del, 2009]. The advantage is that the modeler can take care of setting one single domain and the code takes the responsibility to speed up the results. Depending on the available systems there are algorithms that use distributed memory (a cluster of several machines connected by a network) or shared memory (one single computer with multiple cores). Parallel Virtual Machine (PVM) Dongarra et al. [1991], Poplawski et al. [1989], Message Passing Interface (MPI) MPI Forum [1993] or Open Multi Processing (OMP) OpenMP Architecture Review Board [2011] are some examples of parallelization libraries available for both paradigms.

2.2.2.2. Adaptive methods

Adaptive methods are a modification of the standard methods which, after estimating the evolution of the function to be evaluated, are able to change the spatial and/or temporal resolution. They reduce the number of calculations where smooth solutions are obtained but increase the resolution when high gradients appear in the outputs. An example is the adaptive integration that is well illustrated in Press et al. [2007]. This approach is not available in Delft3D; As is proposed later in section 7.1.3, it is possible to make the operational model run a period with different time steps in order to reduce steep gradients and avoid instabilities in the solution.

Chapter 3

Operational models and data assimilation

This chapter deals with operational models and data assimilation techniques applied to the prediction of water levels in coastal areas. The chapter has been structured into two parts. First, the main components of operational models are introduced. A description of the most relevant numerical aspects and spatial approaches is provided. In this way, the possible improvements for enhancing the quality of short-term predictions were identified. In the second part **data assimilation** techniques applicable in conjunction with field measurements for improving the accuracy of model predictions are presented. The assimilation of data at the boundaries of the domain is emphasized in section 6.1. Attention is also given to the solution of partial differential equations, showing the links between the initial value problems and the boundary value problems by one example with the help of the shooting method. This is convenient in understanding how temporal and spatial problems can be interchanged. A description of **artificial intelligence algorithms** for dealing with nonlinear problems is also introduced and approaches

for assessing the quality of model results is provided.

3.1. Operational Model

In the course of a day a numerical model can be executed automatically several times. If the model delivers short-term forecasts in quasi real time, it is usually called **operational**. Such models have several components. The main components or modules are Data Acquisition, Data Storage, Data Delivery, Model Engine (or solver) and Data Assimilation (see Fig. 3.1). The components can be executed independently or integrated leading to a much more powerful tool for the assessment of the real world problems.

The **Data Acquisition** module is responsible for the management of the network of sensors relevant to the process that is being monitored. The transmission of data from remote locations to the data center is also controlled by this component (see sec. 3.1.1). **Data Storage** is the module that collects and stores the information available from the model, measurements and other sources. **Data Delivery** relates to the way the information is disseminated and made available to the users or external products (see section 7.3).

The **Model Engine or Model Solver** is the actual numerical model solver of the set of equations relevant to the problem in question. In the context of operational models, the application of a numerical approach to a specific scenario is called pre-operational model. Finally, **Data Assimilation** is the technique integrating measurements and model outputs with the aim of providing enhanced predictions (see sec. 3.2).

Nowadays, operational models for weather forecast are very common. COSMO-EU (Consortium for Small Scale Modeling)[COSMO, 2007] is a typical example for an operational weather forecast model. It is carried out by the German Weather Service (DWD - Deutscher Wetterdienst). It consists of 40 vertical layers, and it provides forecasts of up to 78 hours with a horizontal resolution of 7km. The model covers the whole continent of Europe, within the ranges of latitude $30°N$ to $70°N$ and longitude $10°W$ to $42°E$. The integration of field observations within the operational model is done every three hours. The method of forcing the open model boundaries with information from a larger scale adjacent model is applied every hour.

Operational models for simulation of flow and waves are still quite scarce. Allard et al. [2007] summarizes the experiences in the development of a real-time near shore model for the prediction of tides, waves and currents. The model covers the Portuguese coastal waters and was developed in the framework of the project Maritime Rapid Environmental Assessment 04 (MREA04). Ko et al. [2008] compared water levels, temperatures and salinities using the same data collected during the MREA04 campaign. Both systems are not actual operational models as they were not developed to work automatically and continuously.

Beside the models, a system to optimize the operation of vessels in the San Francisco Bay in order to avoid incidents was developed [Cheng and Smith, 1998]. It is based on the Physical Oceanographic Real-Time System (PORTS) developed by the National Oceanographic and Atmospheric Administration of the United States (NOAA). The system integrates a flow model to improve the understanding of the complex currents in the area. The information is available online at `http://sfports.wr.usgs.gov/sfports.html`.

The Federal Maritime and Hydrographic Agency in Germany (BSH which stands

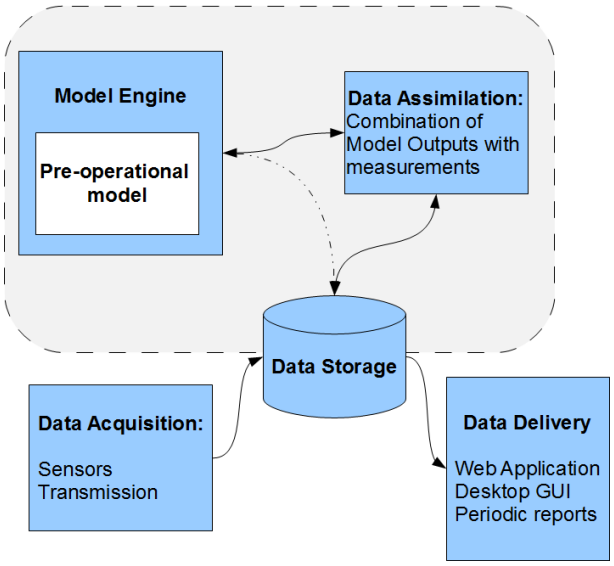


Figure 3.1: General components of an Operational Model

for its German name: "Bundesamt für Seeschifffahrt und Hydrographie") runs an operational model covering the North and the Baltic Seas. It predicts water levels, currents and waves, and takes into account density gradients and, the formation and melting of ice. A Model Output Statistic (MOS) is applied for the assimilation of measurements. The functionally of MOS is presented in section 3.2.2. Model results are available at `http://www.bsh.de/en/Marine_data/Forecasts/Prediction_models/index.jsp`.

3.1.1. Collection of sample data

To obtain real time field data, a network of sensors must be deployed. To achieve this, in this study, devices for measurement of water levels, current velocities, wind speed and direction were employed. Details of the devices are given in Chapter 4. Information about the protection of the sensors against weathering, and the transfer of data to a data center are summarized hereafter.

Data acquisition is the process of collecting information from analogue sources and converting the data to a digital form that can be manipulated for transmission and stored by computers [Taylor, 1997]. Although the main details of data acquisition are specific to electronic engineering, some general aspects should be taken into consideration in the planning and deployment of measuring devices. A classification of devices with the corresponding levels of protection against intrusion from solid objects and liquid ingress can be found in the weathering protection IP Code (ingress protection rating) defined by the international standard IEC 60529. Biological activity is of major concern for the protection and care of devices exposed to coastal environments. Proper measures to minimize the deterioration of the devices and the supporting structures so that maintenance is reduced thus extending their life, should be adopted.

After the sensors have been located in the field collect the data, the information has to be transferred to a central repository for further use. Basically there are three main alternatives to transmit information remotely. For sensors located at less than 10km from a cellphone antenna, a GSM or a GPRS modem can be employed. This method requires least effort for deployment and reasonable maintenance costs as data plans with competitive prices are becoming more common. Connecting to the telephone network the information is accessible from any place in the world via INTERNET. In the reception site, another GSM or GPRS modem needs to be installed to receive the data. Alternatively this step can be done with specialized software provided by the manufacturer.

For devices located farther away, in areas in which telephone network is not available, radio systems can be used. The distance that can be reached by radio systems depends on the power of the transmitter, antenna gains and physical properties of the frequency selected. With ground wave propagation over the surface of earth or water, microwaves (around 1GHz) travel only in a line-of-sight. Thus they are

able to travel long distances providing that there are no obstacles. On the other hand, High Frequency (HF, in the range $3 - 30MHz$) is reflected in the ionosphere being able to travel longer distances [Relay, 2010]. Attention should be payed to the national regulations for the use of radio spectra. The International Telecommunication Union (ITU) is the institution that coordinates the policies of frequency used around the world. As an example Datawell offers a wave-rider system that can reach distances up to $50km$ from the receiver location [Dat, 2010].

For devices located at remote locations, satellite communication is required. The advantage here is that it needs less installation efforts than a radio transmission but transmission is very expensive. This technology has been used mainly for remote wave riders, met-ocean buoys or tracking drifters to follow current paths. If meteorological information needs to be sent to vessels, Automatic Identification Systems (AIS) can be used [Lessing et al., 2006].

3.2. Data Assimilation

The data assimilation is the process of incorporating measured data in numerical models in order to improve model predictions. Before introducing the current techniques a brief notation about partial differential equations will be given. A partial differential equation for a state variable $u(x_1, \ldots, x_n)$ has the form:

$$F\left(x_1, \ldots, x_n, u, \frac{\partial u}{\partial x_1}, \ldots, \frac{\partial u}{\partial x_n}, \frac{\partial^2 u}{\partial x_1 \partial x_1}, \ldots, \frac{\partial^2 u}{\partial x_1 \partial x_n}\right) = 0 \tag{3.1}$$

where F is a function of u (also called state variable, see section 3.3.2 for more

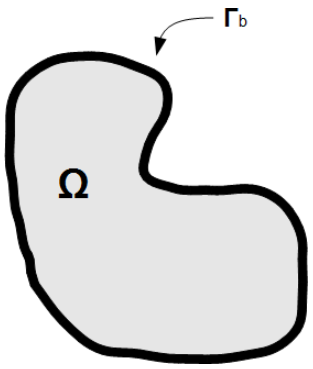


Figure 3.2: Integration Domain for a PDE, Ω denotes the region where the system of equations are solved, Γ_b is the boundary condition.

details), its derivatives and the independent variables x_i. The region of integration (Ω) in the real space of order n ($\Re^n$, where n is the number of state variables) is surrounde by a boundary Γ_b (see Fig. 3.2). When more than one state variable is involved, it can be defined:

$\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$ with $\mathbf{x} \in \Omega$ is the vector of independent terms.
$\mathbf{u} = [u_1, u_2, \ldots, u_l]^T$ the state vector.
$\mathbf{F} = [F_1, F_2, \ldots, F_l]^T$ is the vector with the partial difference operator.

Then it is possible to define the system of partial differential equations as:

$$\mathbf{F}(\mathbf{u}, \mathbf{x}) = 0 \tag{3.2}$$

The final goal is to find an expression for the state vector in terms of the independent variables.

$$\mathbf{u} = \mathbf{G}(\mathbf{x}) \tag{3.3}$$

The system of differential equations that relates the variables within the model domain needs to be known in order to fully define a Partial Differencial Equation (PDE). The same applies to the initial conditions, i.e. the values of all the variables at the starting time and to the boundary conditions which are the values specified at the open boundaries of the domain. When only the values at the starting time are known, the solution is given by the calculated evolution of the variables in time. This is a typical example of an Initial Value Problem (IVP). On the other hand, when only fixed values at the boundaries of the domain are known, and the solution does not change in time (a so-called steady state problem). Then a spatial solution is obtained, constituting a Boundary Value Problem (BVP). Initial-Boundary Value Problems (IBVP) or mixed problems appear when both temporal and spatial variation of the variables is included. Initial values are used in combination with values at the boundaries provided as a function of time [Gockenbach, 2002]. IVP are also described in section 3.3.2, in conjunction with the discussions about the changes of transient systems. The field of dynamic analysis provides the basis for the assimilation strategy proposed in this study.

For a mixed problem of the form $\mathbf{F}(\mathbf{u}, \mathbf{x}) = 0$, there is a function ($\mathbf{F}$) that rules the problem, some initial (defined at Ω at $t = 0$) and boundary conditions (defined at Γ_b as a function of the time). A final solution is obtained from the integration of the system of equations ($\mathbf{G}$ or its numerical equivalence). If any of those elements is altered, a change in the final results from the model is expected. Depending on the proposed model and availability of data, several strategies can be used. The most common ones can be grouped into the following five categories, according to where the changes in the solution are made. A combination of some of the strategies is also possible.

1. Modification of the formulation of the model equations to be solved. The observation is included as part of the independent function. Variational calculus approach and least square methods are typical examples.

2. Modification of boundary conditions by feeding it with measurements.

3. Post processing of model outputs. This strategy includes statistical analysis.

4. Ensemble of several scenarios using alternative models, different parametrization of the same model or different initial conditions.

5. Definition of proper initial conditions by interpolation or through a warming up period by using other assimilation technique.

The first two strategies are considered deterministic as they provide an explicit function to be used by the solver. Strategies 3 and 4 are statistical. The last one depends on the interpolation technique adopted to obtain the initial conditions. Some types of interpolations such as inverse of the distance are deterministic, while Kriging is a statistical interpolation. In the following paragraph a brief description of some of the methods available is provided.

3.2.1. Modification of the model formulation

Sasaki [1955] introduced variational methods in data assimilation. In this case a cost function proportional to the square of the errors between the values analyzed (estimation including the observation) as well as background (previous estimation) and observations is minimized. A Newtonian relaxation term, which accounts for the differences between calculated and observed values, is added to the right hand side of the system of equations. This method has been adopted in the fifth-

generation mesoscale model (MM5 NWP model) for weather forecast Hahmann et al. [2010].

A general expression for the optimal least-squares estimation is given by the interpolation equation shown below [Park and Xu, 2009]:

$$X^a = X_b + K\left(y - H\left[X_b\right]\right) \tag{3.4}$$

Here K is a linear operator called weight matrix of the analysis and is represented by

$$K = BH^T\left(HBH^T + R\right)^{-1} \tag{3.5}$$

In these expressions X^a is the analysis model state, H is an observation operator, B is the covariance matrix of the background errors $X_b - X$, X is the time model state, X_b is the background model state and R is the covariance matrix of observation errors.

One advantage of this strategy is that it uses the physical formulation of the problem being solved. Besides it allows the assimilation of the information from locations within the domain during the simulations. This is particularly relevant for larger domains as it gives higher flexibility regarding the selection of measurement locations. Such strategy is particularly suited in conjunction with atmospheric models in which boundary conditions are imposed from the monitoring locations.

3.2.2. Statistical Methods

In this method, model outputs are evaluated against measurements for several scenarios and statistical relations are obtained. The resulting relations measure the performance of the model. Sets of equations are derived from the statistical relations between model results and measurements. The results are used to improve the predictions of the operational model.

[Müller-Navarra and Knüpffer, 2010] developed a model output statistics (MOS) in conjunction with an operational model covering the German North Sea coast. The MOS is used to correct systematic errors of the outputs of the hydrodynamic model in the forecast mode [Müller-Navarra and Bork, 2010]. It is used primarily to reduce the variance of the resulting forecasts at measurement locations. The equations are solved for hourly outputs and predictions are done up to $T + 33$ hours, T being the present time. Altogether seven groups of predictors are applied according to the conditions in question. It includes predictors for the latest surge observation at one station; Direct Model Output (DMO) surge forecast; wind surge from a Global Forecast system, one measurement location and the $Pers2D_korr$. The latter is a parameter that keeps the last known initialization error of the 2D model. A classification strategy, based on the tidal phase of the surge (high or low water level) is also applied. As a result of the application of this method the variance was reduced up to 28%. The use of Pers2D correction reduces the variance by 45% and 15% respectively at $T + 3$ and $T + 33$.

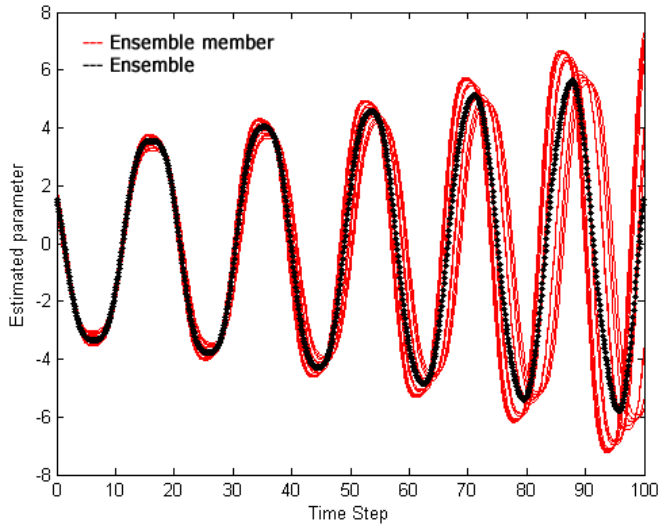


Figure 3.3: Data ensemble of several model execution

3.2.3. Ensemble methods

In this method, several scenarios are executed, with small differences in the initial conditions and in the driving forces or by using different types of models. Every result is called a member of the ensemble. The prediction of the desired parameter is a statistical combination of the outputs of the members. In Figure 3.3 a synthetic example of a sinusoidal model with intentional random error growing is shown. The dark line is the average of all the members. For this combination of outputs the concept of the central limit theorem [Montgomery and Runger, 2010] was adopted. The Ensemble Kalman Filtering (EnKF) is a typical example. Basically it is a sample based statistical-dynamical filter developed by Evensen [1994] and Houtekamer et al. [1998]. In theory, for obtaining optimal solution it requires an infinite number of forecast scenarios. As this is not feasible, the underestimation in the variance should be accounted for. In section 6.4 a simplified ensemble method using the variance of individual models will be proposed to improve the predictions of the coastal model developed in the framework of this study.

3.2.4. Optimization of the boundary condition

Whenever reliable information at open sea boundaries of numerical models is available, it is recommended to use it for forcing purposes. However there are several restrictions that should be taken into account to obtain good quality data. One is the absence of measurements covering the whole simulation period for all the representative locations or segments along the open sea boundaries. This motivates the development of alternative strategies. In this study, a strategy for the estimation of the values of a variable at a specific location was proposed. For this purpose local functions were developed to reconstruct missing information in the time series and to extrapolate single measurements to larger domains. The strategy can also be extended to enable forecasts of a variable for specified periods. The use of a nesting sequence can give a reasonable estimation of the change of the variable along the open sea boundaries. However, due to the lower resolution of the coarser models, the accuracy is likely to be reduced. Therefore a combination of the local model and nesting sequence was proposed. In the next two subsections, examples of previous studies dealing with this subject are presented.

3.2.4.1. Amplification factor at the Open Sea Boundaries

Echeverri [2004] proposed a strategy to improve the accuracy of a nesting sequence of flow models covering the German North Sea. Values at the open sea boundary of the coastal model were linked with measurements at gauge stations located within the model domain. The discrepancies between measured and modeled tidal amplitudes and time lags of the extreme water levels were estimated. To improve the accuracy of the coastal model, the amplitude of the values at the open sea boundaries of the coast model were adjusted with an amplification factor for the

water level and a time shift for correcting the times lags:

$$t_{adjusted} = t_{nesting} - \overline{\Delta t} \tag{3.6}$$

$$H_{adjusted} = \frac{1}{\overline{\Delta A}} \cdot H_{nesting} - \overline{\Delta H} \tag{3.7}$$

where $t_{adjusted}$ is the adjusted peak occurrence time, $t_{nesting}$ is the peak occurrence time in the nested boundary condition, $\overline{\Delta t}$ is the average of the differences in peak occurrence time $\Delta t = t_{nesting} - t_{measurement}$, $H_{adjusted}$ is the adjusted water level, $\overline{\Delta A}$ is the average amplitude ratio $\overline{\Delta A} = \frac{A_{nesting}}{A_{measurement}}$, $H_{nesting}$ is the water level from nested water levels in the boundary condition and $\overline{\Delta H}$ is the average peak water level difference $\overline{\Delta H} = H_{nesting} - H_{measurement}$.

This strategy improves the representation of the extreme water level values and reduces the time shifts during high and low tides. However, there are some limitations associated with it. One is that as a result of the application of a constant amplification factor, the modulation does not respond to the changes in amplitude during a given period. Besides, it keeps the shape of the water level variation, making it impossible to improve the results between tidal extremes. In addition to that, due to the time differences between the simulated water levels at the imposed open sea boundaries, the time frame applied in the two models of the nesting sequence is different. This can be resolved by manually extending the simulation period so that it covers the required time shift. However it increases the complexity in the implementation of an automatic modeling system.

3.2.4.2. Use of a nesting sequence for the Gulf of Lion in the Mediterranean Sea

Vandenbulcke et al. [2006] applied a nesting sequence of flow models to identify the best place to implement the data assimilation technique. A Reduced-Rank Square Root (RRSQRT) assimilation scheme described by Verlaan and Heemink [1997] was adopted. This technique is based on the assumption of a quasi-linear process and Gaussian error. According to Auclair et al. [2003] it does not reflect a real situation but it is acceptable for short time forecasts only. In general, the filter shows good results, but in places with high variability, the filter is unable to capture the variation properly. Such problems were encountered for example in the plume of the Rhône River and at some locations along the coast. To improve the results, a scale of correction was implemented. Corrections were carried out where the filter is effective and decreased to zero, when discrepancies are high. Vandenbulcke calculated the Buoyancy frequency N^2 (also known as Brunt-Väisälä) [Turner, 1973], which allows for the determination of the stability of the stratified layers at each grid point. It was found out that it is more effective to implement the assimilation technique in the detailed model grid instead of making corrections in the global model.

3.2.5. Shooting Method

The shooting method is a numerical algorithm that allows solving Boundary Value Problems (BVP) as a system of Initial Value Problems (IVP). It is presented here to illustrate the possibility of interchanging the problems, and the error distribution depending on the problem in question. The method helps also to understand the approach proposed in this study in the set-up of local functions to improve

the values at the open sea boundaries of a near coastal area model. Keller [1968] Presents a formal description of the method. Marzulli and Gheri [1989] provide an estimation of the error variation in the application of a shooting method for solving in a linear boundary problem. Ghelardoni et al. [1995], on the other hand, estimate the errors when applying the shooting method to an initial value problem (IVP) converted to a boundary value problem (BVP) to take advantage of the delimited error in BVP. An example of this method will be presented hereafter to illustrate how the solutions of an IVP and a BVP are related. Later in section 6.2, a strategy based on the assumption that the solutions from the flow model can be interpreted as several boundary value problems is adopted to interrelate values in different parts of the domain.

An example of the interchangeability of BVP into IVP is shown here. The pattern in the propagation of the errors is also illustrated. The differential equation that expresses the vertical position of an object moving in a uniform gravitational field as a function of time is given by:

$$\frac{d^2y}{dt^2} = -g \tag{3.8}$$

where y is the vertical position, t is the time and g is the gravity constant.

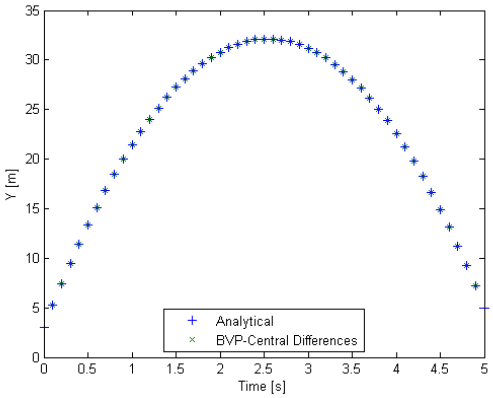
If only the initial position of the object is known and the aim is to reach a desired target at a given time, Eqn. 3.8 can be solved as a BVP subjected to the conditions $y(t = 0) = y_0$ and $y(t = t_n) = y_n$. The BVP can be converted into an IVP and solved with an appropriate numerical method. Defining the velocity as $\frac{dy}{dt} = v$, then the gravitational acceleration is given by $\frac{d^2y}{dt^2} = \frac{dv}{dt} = -g$ and the original set of equations can be rewritten as follows:

$$\begin{aligned} \frac{dv}{dt} &= -g \\ \frac{dy}{dt} &= v \end{aligned} \tag{3.9}$$

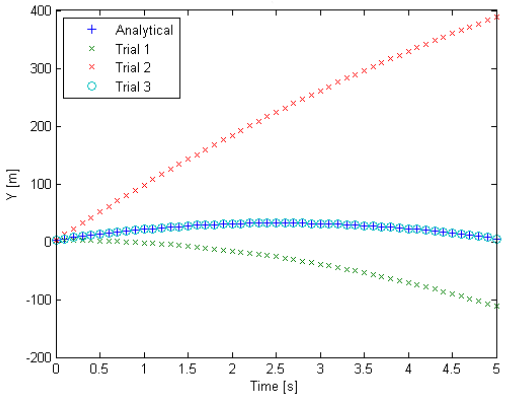
$$\begin{aligned} y(t=0) &= 3 \\ v(t=0) &= v_0 \end{aligned} \tag{3.10}$$

where v_0 is the initial velocity, g is the gravity constant, v is the velocity, y is the vertical position and t is the time.

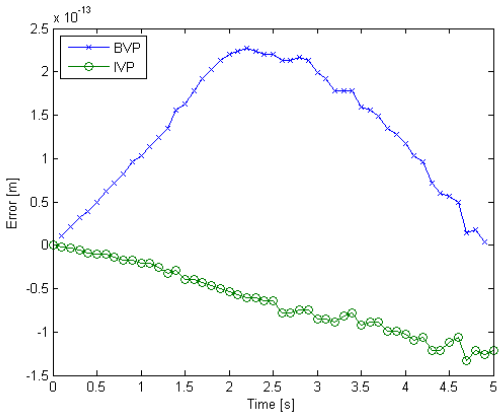
In the first step v_0 is assumed as an arbitrary but constant initial value. This system can be solved with a simple method such as an Euler or using a more sophisticated scheme such as a 4th order Runge-Kutta (e.g. Press et al. [2007]). When t_n is reached, the value of y_n is obtained and compared with the target value at the right hand side boundary. Then the system is solved for the second time using another arbitrary value for v_0. If the equations are linear, as in the example at hand, an appropriate value for v_0, fulfilling the original conditions, can be easily obtained by linear interpolation. In case of nonlinear problems, some extra guesses for v_0 to feed a nonlinear interpolation method or an iterative procedure can be applied until the desired value is reached. Figure 3.4.a shows the numerical solution obtained from the BVP formulation. In Figure 3.4.b the intermediate results and the final solution obtained from the application of the IVP are presented. The differences between the analytical and numerical solutions are shown in Fig. 3.4c. The error variation resulting from the application of the different schemes is clearly seen. For the IVP the error is zero in one extreme, as is the given initial condition, and it grows until it reaches a maximum value in the opposite extreme. When solving the equation as BVP, the solution is fixed at both extremes, and the error at the imposed boundaries is zero, reaching its maximum value in an intermediate location. To be noticed that if the applied numerical flow model is reasonably accurate, forcing the boundary with an appropriate value will



a. Finite Differences BVP



b. Shooting method IVP



c. Error comparison using both methods

Figure 3.4: Solution of a BVP by the use of Finite Differences and the equivalent IVP by the shooting method

lead to a solution that shall adequately describe the whole domain.

3.3. Time series analysis

In this section some basic concepts of time series analysis relevant to the present study are introduced. Focus is given to the way the time series can be treated as dynamical systems by removing the time dimension. This enables the development of local functions that use artificial intelligence tools such as neural networks.

Wei [1990] defined time series as an ordered sequence of observations. Usually the order is determined by the time component, but other dimensions can also be used. Several techniques such as Auto Regressive (AR) models, Moving Average (MA) or a combination of both into an ARMA process are widely used. For the study of water levels in the present work, one of the proposed strategies is based on AR models. In this case the description of dynamical analysis will help in the evaluation of time series from a point of view where the temporal variable is eliminated.

3.3.1. Autoregressive model

Autoregressive models (AR) are linear methods for predicting the output of systems based on previous observations of the same variable [Wei, 1990]. An autoregressive model of order p -$AR(p)$- is defined as follows:

$$Y_t = c + \sum_{i=1}^{p} \varphi_i Y_{t-i} + \epsilon_t \tag{3.11}$$

where Y_t is the value of Y at time t, c is a constant, φ_i are the parameters, ϵ_t is the white noise and p is the number of parameters used in the model.

Later in section 6.2.4. a modification of the AR model is presented to be used as the basis for the development of improved local functions for the prediction of water levels.

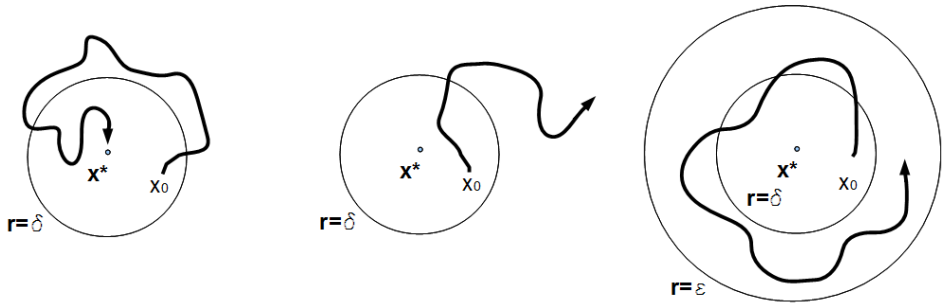
3.3.2. Dynamical systems and Iterative maps

The study of dynamical systems is a discipline dedicated to the investigation of the changes of systems and their variation in time. Meiss [2007] defines dynamical systems as a rule for time evolution on a state space. Unlike time series analysis, this methodology focuses on a spatial evolution of the system. It is applied to figure out whether an equilibrium is reached or whether cycles are being repeated or display complex movements. The study of dynamical systems is derived from the analysis of a set of first-order differential equations of the form:

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}) \tag{3.12}$$

$\tilde{\mathbf{x}}$ is a vector in the n-dimensional real space ($\tilde{\mathbf{x}} \in \Re^n$) whose elements are all the dependent variables of the system. The variables included in $\tilde{\mathbf{x}}$ are called state variables, and each one provides one degree of freedom to the system. Although the system is time dependent, most of the tools provided by dynamical analysis methods are related to the use of spatial evolution. Hence, the temporal variable is removed from the relations and used as the parameter in a set of parametric equations or iterative maps.

For a given discrete time step, it is possible to describe the evolution of the time series by a function f, also called mapping function, at a time $t = m + 1$ which depends on the value at time $t = m$ with $m \in \mathbb{Z}$. This representation is called recurrence relation or iterative map as it enables the time evolution of x to be obtained by repetitive applications of the mapping function using the relation $x_{m+1} = f(x_m)$. $\mathbf{x}$ can be a scalar if n is equal to one, or a vector $\in \Re^n$. Findlay [1911] defined the phase space as the representation of all possible states of a



a. Attracting fixed point b. Repeller fixed point c. Liapunov stability

Figure 3.5: Stability cases (a. and c. are adapted from Strogatz [1994])

system. Every degree of freedom should be represented as an axis. Every single point represented by the vector $\mathbf{x}$ is called phase point.

Attention is payed to two special cases, i.e. closed orbits and fixed points. In the first case, a phase point can orbit around a region in the phase space and it returns to the original point at some time in the future. The solution is periodic and has the form $x(t+T) = x(t)$. The second important case is when $x_{m+1} = x_m$. The recurrence relation where x_m has no change to lead to the solution becomes fixed to that point. That is the reason for calling this value of x a fixed point and it can be denoted by X^*. Then, for a fixed point there is no more evolution of the system. However for trajectories starting in its vicinity, the state points may converge to X^*, or diverge from it. In the first case it is called an attractor and it is considered a stable solution (see Fig. 3.5.a). In the second case it is named a repeller (see Fig. 3.5.b). X^* is a Liapunov stable (see Fig. 3.5.c) if there is a region around it, where all the trajectories starting within that region, are kept orbiting around in a distance r, with $0 < \delta < r < \epsilon$. This forces the trajectories to be closed for all time.

The theorem of Poincaré-Bendixson implies that when a trajectory is confined to a closed region with no fixed points within the region, at the limit when the time goes to infinite it is going to evolve to a closed and periodic orbit (see Fig. 3.6).

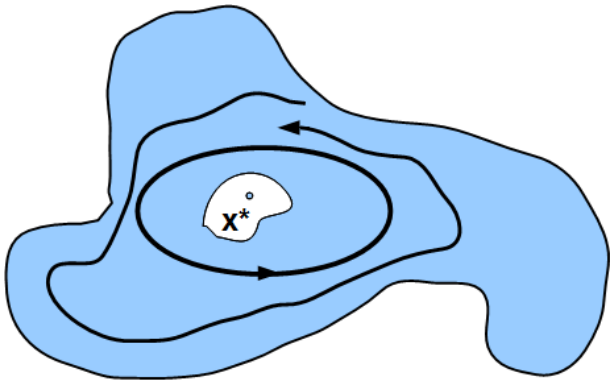


Figure 3.6: Poincaré-Bendixon

For more details the reader is referred to Grant [1999].

The theorem is not valid for space dimensions higher than 2, as the trajectories can describe infinite paths without reaching a fixed point or a closed orbit. "Trajectories can be attracted to complex geometry called a strange attractor" [Strogatz, 1994]. Strange attractors are highly sensitive to the initial conditions, making the prediction unfeasible in the long term. For a point P contained in a phase space the coordinates are given by:

$$\vec{x}(p) = [x_1(p), x_2(p), ..., x_n(p)]$$

where $\vec{x}$ is the state vector, $x_i(p) \forall i = 1, \ldots, n$ are the state variables -they are the components of the state vector- and p is the point in the phase space, given by position vector in the phase space.

3.3.2.1. Embedded dimension

Time-Delay Embedding theorem:

> "Consider a dynamical system with a δ-dimensional space and an evolving solution $g(t)$. Let x_t be some observation $x(g(t))$ and τ_i be specified time lags. The lag vector can be defined as $x(t) \equiv x_t, x_{t-\tau_1}, x_{t-\tau_2}, x_{t-\tau_3}, \ldots, x_{t-\tau_{\delta-1}},$. In general, the space of vectors $x(t)$ generated by the dynamics contains all of the information of the space of solution vectors $g(t)$. The mapping between them is smooth and invertible. This property is referred to as an embedding. Thus, the study of the time series $x(t)$ is also the study of the solutions of the underlying dynamical system $g(t)$ via a particular coordinate system given by the observable x." [Babovic et al., 2005]

The fact that the solution of a time series is interchangeable with the solution of the dynamical system is crucial for the use of artificial neural networks (ANN) in the solution of time series. The inputs of ANN are considered as not necessarily time dependent state variables.

3.3.2.2. Cross-correlation and its relation with embedded dimension

The cross-correlation measures the similarity of two series as a function of a time lag (τ). For two stochastic processes represented by X_t and Y_t the normalized cross-correlation function is given by [Wei, 1990]:

$$\rho_{xy}(\tau) = \frac{E\left[(X_t - \mu_x)(Y_{t+\tau} - \mu_y)\right]}{\sigma_x \sigma_y} \tag{3.13}$$

where σ_x and σ_y are the standard deviations of X_t and Y_t.

Autocorrelation is the particular case of a cross-correlation in which the same serie is introduced twice ($X_t = Y_t$ in the eq. 3.13). As shown by Estévez Báez et al. [2008], the shape of the autocorrelogram plot gives hints about the type of process that controls the time series. In theory, for pure random processes, autocorrelation after lag 1 shall be zero, meaning that there is no relation at all between the value and its predecessors. For periodic processes and with hidden periodicities, it gives a periodic autocorrelogram without damping. When the series stem from a moving average, the value of the coefficient decays to zero, with short oscillations of low amplitude and values close to zero until the end of the autocorrelogram. An autoregressive process has a periodic oscillation, which reduces amplitude without reaching zero. The autocorrelation can be used as an estimator of the dimension δ of the dynamical system when time series are considered.

3.4. Artificial Neural Network (ANN)

Artificial neural networks are artificial intelligence algorithms inspired in the nervous system of animals. In such systems the number of neurons is interconnected in a network which shares information among their elements. Some neurons can act as sensors, others belong to the central system of processing the information and the remaining reveal the response of the network to the outer world. The information entering a neuron can activate new connections and the response of the whole system to some inputs can determine specific actions.

3.4.1. Previous uses of ANN in hydrodynamics

Tirozzi et al. [2006] carried out studies on the reconstruction of hourly variations of waves and water levels. Wenzel and Schröter [2010] used ANNs to connect the coastal sea levels with the regional and global mean values and also to reconstruct time series to estimate the sea level rise based on stations located at the coastlines around the world. Bazartseren et al. [2003] carried out short-term forecasts of about 10 hours in rivers based on ANNs. An ANN was also integrated within a decision support system for the management of the navigation of the Magdalena River in Colombia Nelson Fernández and Altamiranda [2010]. Herman et al. [2007] used Principal Component Analysis (PCA) to accelerate the simulation of wave scenarios in a morphodynamic model. The PCA was used in combination with wind and wave model results to train an ANN. Babovic et al. [2005] developed an error correction for a predictive ocean wave model. To predict the nonlinear errors from a wave model a process based model was combined with the chaos theory. Genetic algorithms were applied to find the optimal parameters of the local model.

3.4.2. Architecture of a feed-forward network

There are at least three layers of neurons with different properties in a feed-forward network (see Figure 3.7). In the input layer, each node represents one element of the input vector $x \in \Re^n$ that can be represented in its original form or scaled to a specific range (some authors take $x \in [0, 1]$). In the present work water levels are the main inputs. However in some cases, wind components are also used.

The output layer is the last level available and it represents the outputs or the results from the network. A single or several simultaneous outputs can be ob-

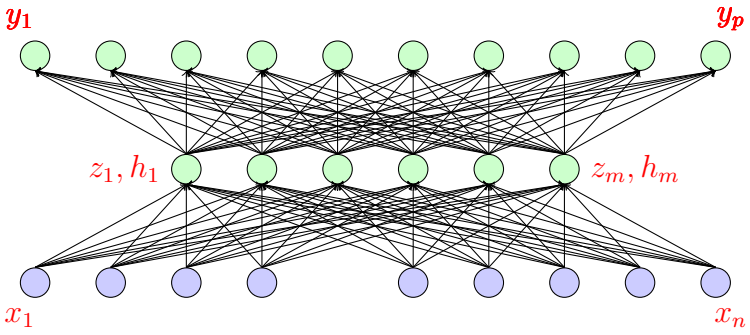


Figure 3.7: General Neuronal Network

tained. In this study attention is given to one single output that corresponds to the estimation of the water level. The output neurons collect the nonlinear terms from lower levels and through linear combination (see eq. 3.14) a response to the stimulus introduced in the input layer is obtained. Any other layers, other than input and output, are hidden layers. The neurons contained in such layers can be considered as the brain. Those elements interact with the inputs. A linear combination of the inputs (eq. 3.17) is transformed by a nonlinear activation function (eq. 3.16) to create outputs (eq. 3.15), giving nonlinear capabilities to the neural network. A detailed representation of a single hidden neuron is shown in Figure 3.8.

The role of activation functions is to combine the linear inputs z_k with the trigger value t_k in order to establish active synapses under some conditions or deactivate them. To activate or deactivate synapses the sigmoid function is applied. This function allows to intermediate nonlinear values. Other activation functions commonly used in ANNs are the Heaviside step function (unit step function). It provides only binary values, and a linear activation function that permits a soft transition from zero to one. The latter alternatives are not considered in the present work, as nonlinear support is not provided. The output of the neuron is a value in the range between 0 and 1.

The neuronal network is defined by:

$$\bar{y_p} = \sum_{k=1}^{m} V_k h_k \tag{3.14}$$

$$h_k = \sigma(z_k - t_k) = \sigma(\theta) \tag{3.15}$$

$$\sigma(\theta) = \frac{1}{1 + e^{-\alpha\theta}} \tag{3.16}$$

$$z_k = \sum_{j=1}^{n} w_{k,j} x_j \tag{3.17}$$

where $\bar{y_p}$ is the estimation of the output value p, V_k is the weight coefficient for the hidden layer behind the output layer, h_k is the output from neuron k, z_k is the linear combination of the nodes of the previous layer, t_k is the trigger value for the activation function, θ is the argument of the sigmoid function -it is given by the combination of the z_k and t_k-, α is the parameter that changes the transition width of the sigmoid function, z_k is the linear combination of the outputs from the previous layer, $w_{k,j}$ is the weight coefficient that connects neuron k in layer i with neuron j in layer $i-1$, x_j is the input element, σ is the sigmoid function -this is the activation function-, p is the number of output elements, m is the number of hidden neurons and n is the number of input elements.

In these equations, V_k, t_k and $w_{k,j}$ are the parameters to be optimized. The input data is $x_j \in \Re^n$. The output data is y_p. α is a constant to be selected beforehand. It controls the sensitivity of the activation function. Values between 0 and 1 are

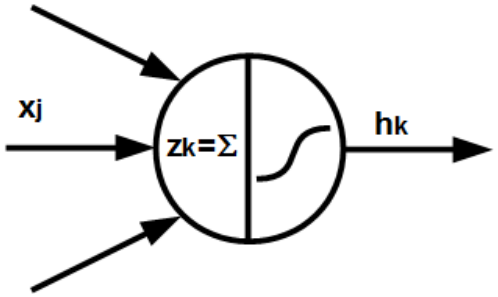


Figure 3.8: Single processing unit or neuron

assigned respectively for disabled and maximum weighted neurons. Once the σ function is totally activated ($\sigma = 1$) or deactivated ($\sigma = 0$), small changes in the argument θ does not produce any effect in the calculation of neurons in a higher layer.

3.4.3. Training

The training of an ANN is the process of finding the parameters of the network on the basis of sample data sets. Nonlinear optimization techniques enable the minimization of the error term of the output function with respect to measurements. To achieve this, a sample subset is used to fit the network until a desired tolerance is reached. Then the network is validated with the remaining samples. If the errors are of the same order of magnitude, the solution is considered to be sufficiently accurate. Due to the number of parameters and the nonlinearity, there is no guarantee that the global minimum is found, limiting the solutions to local minima.

3.4.3.1. Error backpropagation by the gradient descent method

The gradient descent method is a deterministic procedure that takes the error between the network estimation and the measurements into account. A correction of the function parameters proportional to the partial derivative is calculated. A detailed mathematical formulation can be found in Hecht-Nielsen [1989]. Here only the main steps are presented.

The procedure starts assuming that the error of the estimation for each set of inputs p is defined as follows:

$$e_p = \frac{1}{2}(\bar{y_p} - y_p)^2 \tag{3.18}$$

where e_p is the estimation error for input p, $\bar{y_p}$ is the estimation for the output p and y_p is the observation p.

The correction terms are calculated according to:

$$V_k\,(t+1) - V_k\,(t) = \Delta V_k = -\beta \frac{\partial e}{\partial V_k} \tag{3.19}$$

where $V_k\,(t+1)$ is the parameter V_k at iteration $t+1$, $V_k\,(t)$ is the parameter V_k at iteration t and β is the advance speed modulator.

Rearranging the terms in the equation, the new value for the parameter is given

by:

$$V_k\left(t+1\right) = V_k\left(t\right) - \beta\frac{\partial e}{\partial V_k} \tag{3.20}$$

The constant β enables the modulation of the time efficiency of the method. A small value of β modifies the parameters slowly. This value must be carefully selected. Attention should be payed to high gradients in the error term as it could lead to instabilities, or to random jumps in the solution space without convergence.

After evaluating the error for each parameter in the neural network, the final set of equations can be written as:

$$\frac{\partial e}{\partial V_{\underline{k}}} = \left(\bar{y_p} - y_p\right) h_{\underline{k}} \tag{3.21}$$

$$\frac{\partial e}{\partial t_{\underline{k}}} = -\left(\bar{y_p} - y_p\right) \alpha e^{-\alpha\theta}\sigma^2\left(\theta_{\underline{k}}\right) \tag{3.22}$$

$$\frac{\partial e}{\partial w_{\underline{k},\underline{j}}} = -\frac{\partial e}{\partial t_{\underline{k}}} x_{\underline{j}} \tag{3.23}$$

Finally the new parameter set is tested. If the error is lower than in the previous iteration, the set is updated and the model advances to the next iteration. Otherwise, the optimization process stops, as it assumes to have reached an optimal solution. One of the drawbacks of this method is that it cannot ensure that the

result is only a local minimum. This is bound to happen when a large number of parameters are involved. Moreover, if the value of β is too high, random solutions are obtained.

3.4.3.2. Resilient Back-Propagation RPROP

It is a modification of the gradient descent method that intends to avoid the problems with descent gradients. Here the magnitude of the correction step is not defined by the magnitude of the derivative of the error [Riedmiller, 1994]. Instead it is defined by the change in sign of the error term. If in two successive iterations, the sign remains the same, it is assumed that the direction of the changes and the correction term is increased to speed up the convergence of the solution. In case of a sign inversion, the correction step is decreased to slow down the modifications.

3.4.3.3. Simulated annealing

Simulated annealing is a derivative free optimization (DFO) [Nocedal and Wright, 2006]. It is used when the derivative is difficult to be obtained or unreliable in the presence of noise. In general DFO employs sampled function values to estimate the new iteration. Some DFOs use a linearization of the function to be optimized in order to search for the next value. In the case of simulated annealing, it is a non-deterministic approach.

Every parameter to be updated is selected by a random function and the new parameter value is given by $w_{k,j}^{i+1} = w_{k,j}^{i} + a \cdot h\left(i\right)$ with $a \in [-1, 1]$ being a random number, $h\left(i\right)$ is a cooling function that gives smaller corrections as long as the

time -number of iterations- advance.

3.4.4. Hidden neurons

If too many neurons are included in the ANN and the number of available samples for the training is limited, memorization problems might occur. As a result the ANN loses its ability to respond to new scenarios. Besides, outputs based entirely on what can be remembered from the inputs used during the training are delivered.

In order to avoid memorization and to keep the capacity of generalization of the network, the number of neurons in the hidden layer must be limited. The Baum-Haussler rule see Baum and Haussler [1989] enables the estimation of the hidden neurons as follows:

$$N_{hidden} \leq \frac{N_{train} E_{tolerance}}{N_{pts} + N_{output}} \tag{3.24}$$

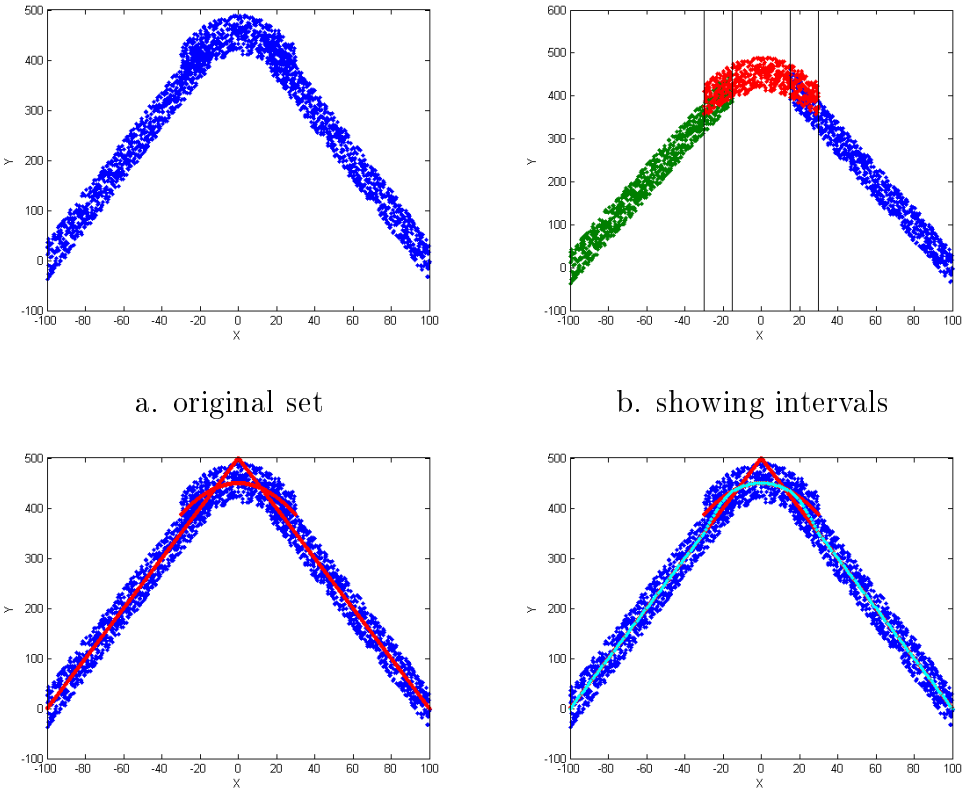
Where N_{hidden} is the number of hidden neurons, N_{train} the number of training examples, $E_{tolerance}$ the error tolerance, N_{pts} the number of data points per training example (input nodes) and N_{output} the number of output neurons.

3.5. Fuzzy Logic

If the samples present different patterns for different regions in the domain, they can be adjusted to different functions by splitting the samples into several sub-domains. In standard logic, limits of intervals are fixed by a single value. However in many practical cases those limits are well defined. The idea behind fuzzy logic is that there are some areas where the behavior of a given parameter could be easily defined whereas in other parts this is not possible. Figure 3.9 shows an example of a linear relation with positive gradient for negative and positive x values. However in the vicinity of $x = 0$, corresponding to the intersection of the two lines, there is not a clear angle but a sort of parabolic variation. Therefore the problem is to define the limits of a segment function representing the global shape. In Figure 3.9.b the three original data sets can be seen. The first group covers values with $x \in [-100, -20]$, the second for $x \in [-30, 30]$ and the last for $x \in [20, 100]$. When standard logic is used, there is a risk in obtaining a non-continuous function (see Figure 3.9.c).

$$y(x) = \begin{cases} 500 + 5x & if\ x \in [-100, -30] \\ (1 - w1)(500 + 5x) + w1(450 - 0.07x^2) & if\ x \in [-30, -15] \\ 450 - 0.07x^2 & if\ x \in [-15, 15] \\ w2(500 - 5x) + (1 - w2)(450 - 0.07x^2) & if\ x \in [15, 30] \\ 500 + 5x & if\ x \in [30, 100] \end{cases} \tag{3.25}$$

with: $w1 = \frac{x-(-30)}{-15-(-30)}$ and $w2 = \frac{x-(15)}{30-(15)}$



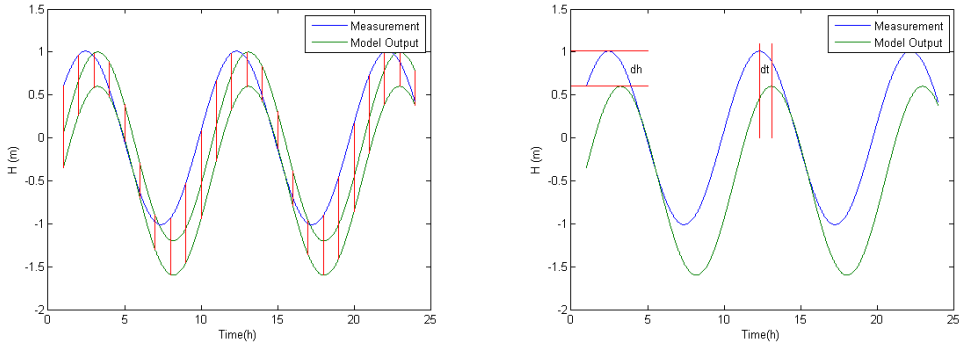
a. original set

b. showing intervals

c. with regression functions

d. with fuzzy function

Figure 3.9: Scatter plot of a non linear set.



a. Considering as a time series b. Using only extreme values

Figure 3.10: Water level error measurement.

3.6. Measuring the quality of the fit

To assess the quality of models in predicting water levels, it is possible to analyze model results and measurements as simple time series and evaluate them with standard statistical tools. For this purpose linear regression, root-mean-square deviation (RMSE), correlation coefficients or standard deviation can be used. Besides, if only the extreme values are of interest, the high and low water level magnitudes and times of occurrence can be considered to verify how close the series are related (see Figure 3.10). However, one of the limitations of evaluating only extreme values is that inter-tidal discrepancies between the model results and measurements are omitted. These differences could affect the patterns of flow velocity with respect to the real conditions.

Chapter 4

Study area and available data sources

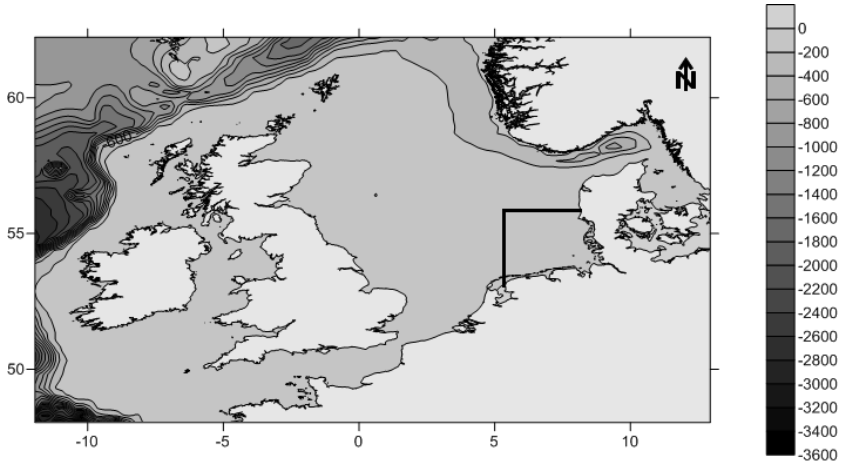
The theoretical concepts have been presented in previous pages; in this chapter the Dithmarschen Bight, which is the area where the present study is carried out, is in the focus of interest. Therefore, physical aspects as well as the available meteorological data for this region are described. This information is important for the development of hydrodynamical numerical models and for the data assimilation that are discussed in Chapters 5 and 6. The morphology and meteorological conditions are also of interest. So, an inventory of measured data to be integrated in the automatization of the model is presented. This method is followed by a description of the wind characteristics in Brunsbüttel, Cuxhaven and Rochelsteert (see Figure 4.3). In addition to that, a comparison of wind measurements at Büsum with wind model values at the same location is highlighted to emphasize the quality of the wind forecast provided by the German Weather Service (DWD).

4.1. Study area

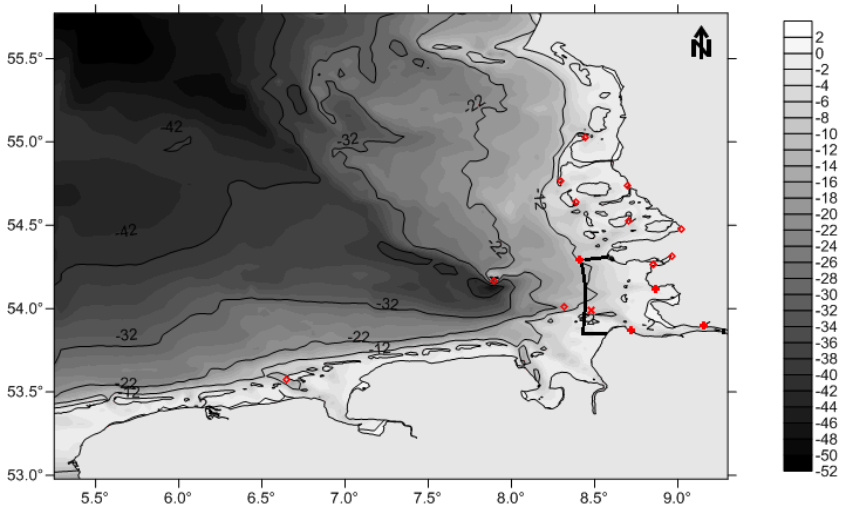
The Dithmarschen Bight (see Figure 4.1) is located in the Wadden Sea. The Wadden Sea belongs to the North Sea and covers the coast from Ho Bay in Denmark to the island of Texel in the Netherlands. In Germany, the Wadden Sea includes the states of Schleswig-Holstein, Hamburg and Lower Saxony. This area is important for the economy of Germany as it is the entrance to the port of Hamburg and the artificial channel that connects the North and the Baltic Sea ("Nord-Ostsee-Kanal" for its name in German). The Dithmarschen Bight area consists of tidal mudflats, channels, sandbanks, dunes, islands and salt-marshes [Clorious, 2008].

The mudflats are wide areas of the Wadden Sea which are constantly flooded during high tide and dried in ebb tide. This extremely variable environment is exposed to wind in low tide, and to waves and water currents in high tide. There are a number of microorganisms that take nutrients from the water and the soil, and consume the pollutants that exist in the area. In this process they become an important source of food for other organisms settled in this area [of Lower Saxony, 2008]. Bigger organisms like marine mussels, bivalves (Macoma Balthica), snails, shrimps, worms (Nerelis Diversicolor, Arenicola Marina) and several marine mammals also live in this environment [Reise, 1981]. Due to the special conditions of the ecosystem, the UNESCO has included some parts of the Wadden Sea in the World Heritage List since June 2009 [UNESCOPRESS, 2009].

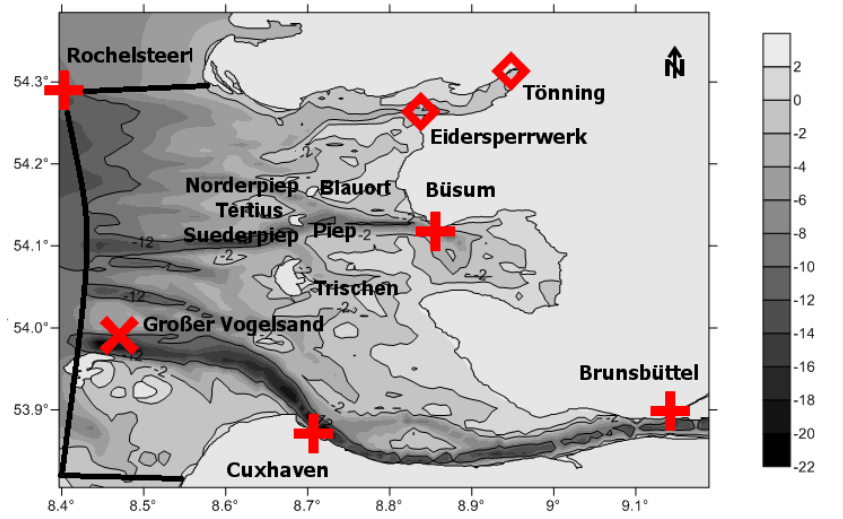
The strength of mud beds is determined by physical, chemical and biological composition [Verreet et al., 1986]. As in the Wadden Sea there are a variety of hydrodynamic conditions; the distribution of sediments is a mixture of mud and sand in diverse proportions depending on the location of the tidal flats, the tidal channels, estuaries or deeper areas. Houwing [1999] carried out studies on critical erosion



a) Continental shelf of the North Sea



b) German Bight and the Wadden Sea



c) Dithmarschen Bight and the available stations in this area

Figure 4.1: Bathymetry of the study area with the location of measurements

of cohesive sediments on the intertidal mudflats in the Dutch Wadden Sea and found that the ratio of sand to mud had a mean value of $80 \pm 10\%$, but in the pioneer zone, in the vicinity of the salt marsh the amount of sand is reduced to 60%. The sand fraction is well sorted with fine sand ($d_{10} = 65\mu m$, $d_{50} = 80\mu m$ and $d_{90} = 110\mu m$). Wohlenberg [1937] studied the sediments at Königshaven at the island of Sylt. In this location it was found a coarse grain with a median of $456\mu m$ and a sorting coefficient as a ratio of the first to the third quartiles by weight Q_1/Q_3 of 1.5 and a variable amount of detritus.

In the mudflats, there are three layers of sediments that can be easily differentiated by the color. At the surface there is a brown layer with a thickness between $0.2cm$ and $2cm$. This color is due to the iron hydroxides in the presence of oxygen (oxic condition). Then there is a black layer with a thickness of $10cm$ to $15cm$, and at the bottom a gray sediment. The colors black and gray are the result of the presence of iron sulfides and iron bisulfides in an anaerobic environment [Reise, 1981].

Salt marshes are sedimentary deposits that serve as a transition between the sea bed and terrestrial habitats [Beeftink, 1977]. The salt marshes are important for the coastal defense, absorbing wave energy [Niemeyer, 1977]. In a study on the effect of the sea level rise on the salt marshes in the Netherlands Wadden Sea carried out by Dijkema et al. [1990] it was found that the pioneer zone for the foreland salt marshes (created by the people) have an accretional deficit. On the other hand, salt marshes have a positive accretionary balance due to the protection of sediments provided by the vegetation.

The Dithmarschen Bight morphology is dominated by tidal flats with sand islands namely Blauort, Tertius and Trischen (see Figure 4.1.c). Fifty percent of the area becomes dry during the low tide, and is submerged again in the high tide

[Mayerle and Palacio, 2002]. Besides, there is a migration of islands. Trischen for example moves some $29cm$ per year to the East [Wieland, 1972]. The three main tidal channels Piep, Nordpiep and Süderpiep dominate the central area of the bay. These channels have maximum depths of $23m$, $14m$ and $17m$ and their widths are in a range from a few meters to $4km$ [Pramono, 2004].

Dittmer [1938] obtained a description of the vertical structure of the sea bed in the Dithmarschen Bight. By using core analysis he found a layer consisting of Pleistocene sands at a depth of 25 to $30m$. This layer is covered by cohesive silty clay deposits with a thickness of $10m$. On top there is a layer of sand with a thickness of $20m$ in the tidal flats and nonexistent in some parts of the tidal channels. Figge [1981] provided a horizontal distribution map of sediment distribution for the German Bight but with a coarse resolution that does not provide information about the spatial variation inside the Dithmarschen Bight. His work reveals that 50% of the sand has a diameter in the range between 63 and $250\mu m$. The sand fraction can represent from 20 to 90% of the total sediment.

Reimers [2003] measured sediment sizes with a median size in the range of 150 to $230\mu m$ on the region located at South West of Tertiussand, and between 110 to $150\mu m$ on the West of Blauortsand and North of Bielshoevensand. And for the East of Blauortsand and the tidal flats, values of d_{50} between 90 and $110\mu m$. Vela-Diez [2001] concentrate its efforts in the tidal channels of Norderpiep, Süderpiep and Piep. In these channels, sediments consists of medium to fine sand and also consolidated mud. The d_{50} in the tidal channels are in the range between 80 to $230\mu m$.

4.2. Available field measurements

In this section, the measurements integrated in the operational model are described, as well as the sources of bathymetrical information and the covered areas. This bathymetry is used to update the model. In situ measurements of currents, water level, waves and sediment collected during the PROMORPH project and used for the calibration of the original model are described by Palacio et al. [2005].

4.2.1. Bathymetry

The bathymetry of the North Sea is taken from the original model provided by Verboom et al. [1992]. In addition measurements with a dense sampling are provided by the Bundesamt für Seeschifffahrt und Hydrographie (BSH), Wasser- und Schifffahrtsverwaltung des Bundes (WSV), Kuratorium für Forschung im Küsteningenieurwesen (KFKI) and Amt für ländliche Räume (ALR) Husum. This bathymetrical information covers the German coast for the period 1969-2009. The depth measurements from the different sources are combined to create the bathymetry in the domain. To use the latest available source of data, an algorithm to optimize the selection of the newest sample in the vicinity of a cell grid has been developed. For every cell in the area, a search of samples from the newest available dataset is carried out. A variable radio around the cell -depending on the cell size- is used to avoid that samples affect the estimation of the bathymetry in grids located far away. Once the bathymetry data is interpolated in a cell, its value is fixed for the rest of the procedure. The next dataset is then loaded and a new search is done for the remaining empty cells. When all the samples are exhausted, the original bathymetry from 1968 is used to fill the rest of the domain in the German area (see Figure 4.2 to see the most update information available). An inspection is

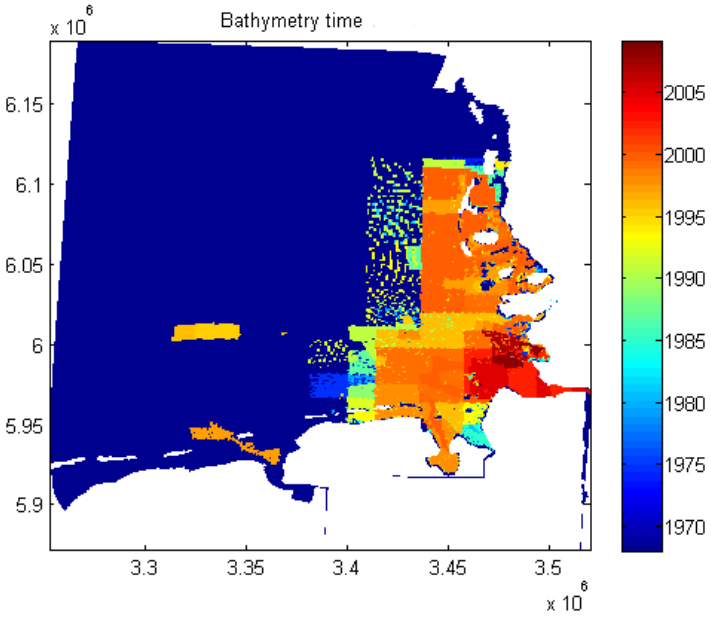
done to remove artifacts and improve some areas to guarantee the continuity of the channels, mainly in the Dithmarschen Bight.

4.2.2. Water level

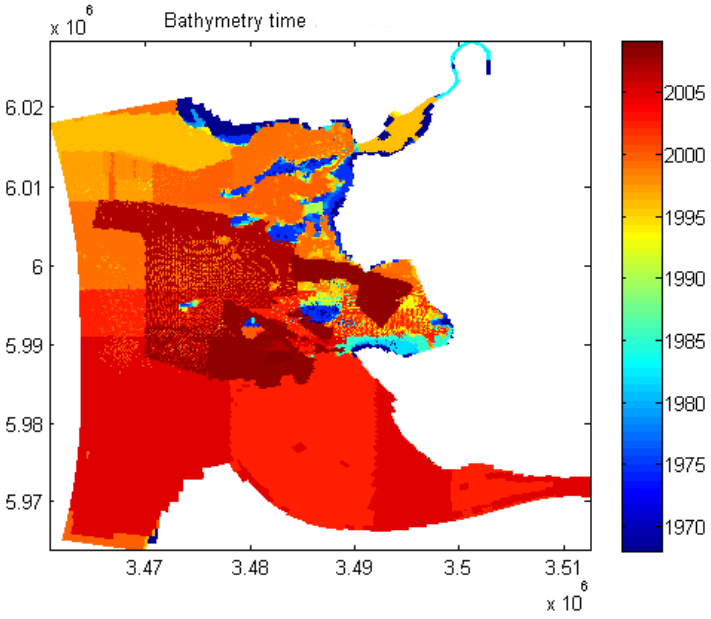
On-line water level stations in the area are provided by the radio transmission of the 'Wasser- und Schifffahrtsverwaltung des Bundes' (WSV). Their locations are listed in Table 4.1 and shown in Figure 4.1. Additionally, the FTZ has installed -in the frame of the OMS project (Fernandez Jaramillo and Mayerle [2009])- its own device at the location Rochelsteert. This station is listed in the same table. Out of these stations the most relevant for this work are those located in the vicinity of the Meldorf Bight. Measurements carried out at Rochelsteert, Cuxhaven, Brunsbüttel, Büsum and Großer Vogelsand are used to improve the models.

4.2.3. Wind information

Wind measurements in real time are available at the FTZ institute in Büsum. This station is located 15m above the ground level close to the coastal line between the FTZ building and the protecting dike [Vanselow, 2008]. In this station, sensors connected to a mechanical anemometer manufactured by Theodor Friedrichs are installed with an operational range of wind speed (catalog number 4035.0000) from 0 to 70m/s and for direction (catalog number 4123.0000) in the range 0-360° [Friedrichs and und Systeme GmbH, 2004]. As the wind speed used in the Delft3D model is defined at 10m above the free surface [WL , 2003a], measurements should be modified to obtain equivalent values to that reference level. This step is made by using the Air-Sea version 2.0 developed by Beardlsley and Pawlowicz



a) German Bight domain



b) Dithmarschen Bight domain

Figure 4.2: Bathymetry information (Year of availability) for the German area (top) and a closeup to the Dithmarschen Bight (bottom)

Table 4.1: Location of available Tide Gauges

Location	Latitude	Longitude
Bake_Z	54°00'50.00"N	8°18'59.90"E
Büsum	54°07'15.30"N	8°51'35.09"E
Borkum	53°35'17.00"N	6°40'11.00"E
Brunsbüttel	53°53'70.00"N	9°08'40.00"E
Cuxhaven	53°52'30.00"N	8°42'40.48"E
Dagebüll	54°43'49.99"N	8°41'27.70"E
Eidersperrwerk_Außenpegel	54°15'53.47"N	8°50'28.44"E
Eidersperrwerk_Binnenpegel	54°15'53.47"N	8°50'28.44"E
Glückstadt	53°47'10.00"N	9°24'40.00"E
Großer Vogelsand	53°59'40.00"N	8°28'30.20"E
Helgoland	54°10'26.00"N	7°53'37.00"E
Hörnum	54°45'30.47"N	8°17'45.48"E
Husum	54°28'32.47"N	9°00'23.34"E
List_Sylt	55°01'00.03"N	8°26'23.44"E
Pellworm	54°31'16.11"N	8°41'42.39"E
Rochelsteert (pile) FTZ	54°17'30.00"N	8°24'30.00"E
Tönning	54°18'51.33"N	8°57'04.01"E
Wittdün	54°37'55.90"N	8°23'08.90"E

[1999] based on Large et al. [1995]. For the usual ranges of wind speed observed in this area the correction is shown in Figure 4.5. Other stations provided by BSH are: Bake_a, Brunsbüttel, Cuxhaven, Eidersperrwerk Außenpegel, Hörnum, Rochelsteert, Borkum and Wittdün with the same locations listed for the tidal gauges in the Table 4.1.

To describe the wind characteristics, the wind roses with information from 2008 to 2011 of Brunsbüttel, Cuxhaven and Rochelsteert are presented in Figure 4.3. At Brunsbüttel the wind is from a wide range of directions from the North-West to the South. The maximum speed for wind from the west is 12.4m/s, while from the south it does not exceed 7.5m/s. Wind speed coming from North and East is negligible. At Cuxhaven the wind is also coming from North-West to South, but the speed can reach up to 14.3m/s. The maximum speed for winds is coming from North-West and from South-Southwest. For Rochelsteert and Eidersperrwerk, the dominant winds are covering from West to North East and the speeds are distributed more uniformly. As can be observed, the stronger wind is blowing from the sea side where friction is lower than on the land. Wind coming to the coast is also a more critical condition in case of storm surges.

Spatial wind is provided by the COSMO-EU model from the German weather service as it has been described at section 3.1. The received dataset is provided at intervals of 12 hours. The first record is created by analysis of measurements, while the next 48 hours is a forecast. A temperature forecast for a longer period of 72 hours is also available. The covered area is in the square limited by the coordinates 12.5°W, 13.5°E, 47.5°N, and 62.5°N. Comparisons of model wind with measurements at the available stations have been made. They show a good agreement for 48 hours forecast as can be seen in Figure 4.4

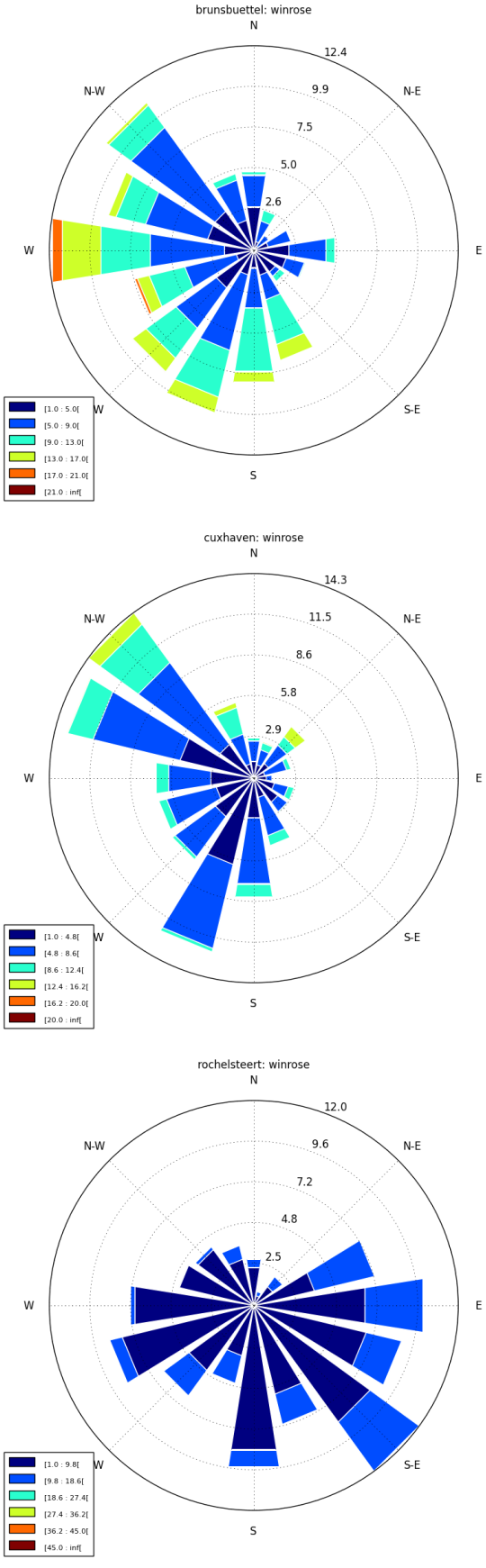
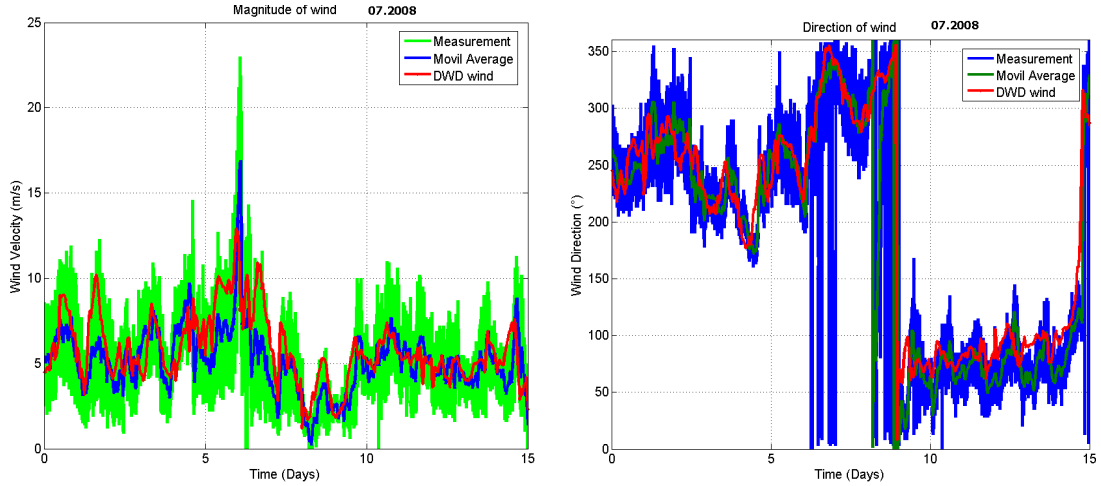


Figure 4.3: Wind Roses for available stations data from 2008 to 2011.

Brunsbüttel, Cuxhaven, Rochelsteert



a) Velocity b) Direction

Figure 4.4: DWD and Wind measurements at Büsum

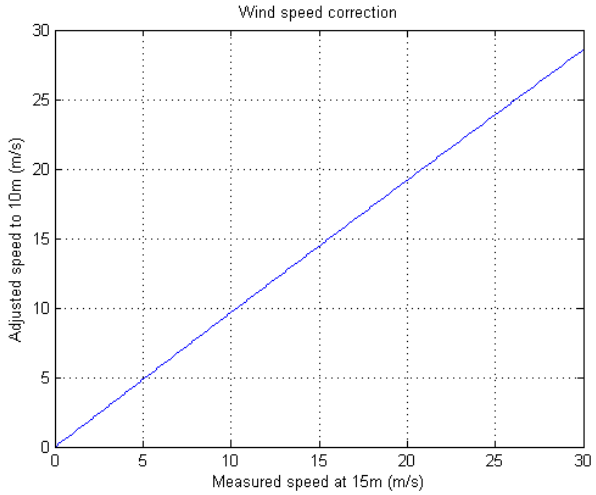


Figure 4.5: Wind correction

4.2.4. Storms

During winter, strong wind conditions exist in the North Sea. This phenomena makes the region vulnerable to the appearence of storms. Flooded land in the coastal region has been occurring for a long time. Stadelmann [2008] has made an inventory of some of the storms that occurred in the last part of the 20th century. Since 2009 BSH has a report of storms that is updated after every important event Müller-Navarra et al. [2013]. Next is a list of storms provided by both sources:

- 16/17 February 1962: 12500ha of urban areas in Hamburg were flooded and 350 people died. In addition two dikes in Amrum, one in the Dockkoog in front of Husum and one in Eiderstedt have failed. The water reaches values of $NN+4.94m$ or $MThw+3.43m$ at Büsum, $NN+5.21m$ or $MThw+3.73m$ at Husum.
- 23 February 1967: Relative small heights were observed because it occurred during the ebb tide.
- 13,16,19 November and 6,14,17 December 1973: A series of storms have been occurred.
- 3 January 1976.
- 21 January 1976: The water level has remained high during the ebb tide.
- 24 November 1981.
- 26-28 February 1990.
- 23 January 1993, 28 January 1994, 10 January 1995: For these three years, water levels of $NN + 4.31m$ $MThw + 2.71m$, $NN + 4.56m$ $MThw + 2.96m$, $NN + 4.27m$ $MThw + 2.67m$ were reached.

- 18 January 2007: The storm Kyrill with winds in the range of 100k/h to 198km/h had produced a relative low inundation in the city of Hamburg compared with the magnitude of the storm [Lange, 2012].
- 4 October 2009: During the storm Sören, water levels of $NN + 4.11m$ were reached in Hamburg , Bremen $NN + 4.29m$, Cuxhaven $NN + 3.44m$ and Büsum $NN + 3.39m$.
- 4 and 5 February 2011: Hamburg NN+4.26m, Bremen NN+4.13m, Cuxhaven NN+3.21m, Büsum NN+3.49m
- 5 June 2012: The Hurricane Andrea drove the water leves to reach NN+4.33m in Hamburg , NN+4.20m in Bremen , NN+3.22m in Cuxhaven and 3.16m Büsum
- 31 January 2013: The storm Lennart. Hamburg NN+4.77, Bremen NN+4.35m, Cuxhaven NN+3.53, Brunsbüttel NN+3.71m and Büsum NN+3.78m.
- 28 October 2013: Storm Christian Hamburg NN+3.85m, Bremen NN+3.74m Cuxhaven 3.06m, Brunsbüttel NN+3.35m and Büsum NN+3.47m.

4.2.5. Waves

As mentioned in the Introduction, wave heights of up to $4m$ have been observed in the outer region of the investigation area [Mayerle and Zielke, 2005]. But because of the wave breaking along the edge of the tidal flats, the influence of the waves in the tidal channels is negligible [Toro et al., 2005]. To obtain wave information close to the boundary of the model, at the pier located at Rochelsteert, water levels are measured by an acoustic transducer with a sampling rate of 200 milliseconds. From those measurements a frequency analysis is done to obtain the significant wave height (Hs) and wave period (tp). To have an idea of the

Table 4.2: Douglas Sea Scale, classification of the state of the sea

Degree	Height(m)	greater than	Class percentage	Description
9	14.00+	100.00	0.00	Phenomenal
8	9.00 - 14.00	100.00	0.00	Very High
7	6.00 - 9.00	100.00	0.00	High
6	4.00 - 6.00	100.00	0.01	Very Rough
5	2.50 - 4.00	99.99	1.39	Rough
4	1.25 - 2.50	98.60	15.60	Moderate
3	0.50 - 1.25	83.00	39,90	Slight
2	0.10 - 0.50	43.10	37.40	Smooth
1	0 - 0.10	5.70	3.60	Calm (Rippled)
0	no wave	2.10	2.10	Calm (Glassy)

prevalent conditions in the area, the wave measurements can be classified according to the Douglas Sea Scale (proposed by H.P. Douglas on 1920 and adopted by the International Meteorological Conference in Copenhagen in 1929 [Office, 2010]). Table 4.2 presents the percentage of occurrence of the significant wave height in the each category. The dominant conditions are between smooth and moderate that represent 92.9% of the observations. Calm conditions are present only 5.7% of the time, while rough and very rough occurs only 1.40% of the time.

4.2.6. Currents

During the PROMORPH project currents reaching up to $2.8m/s$ have been measured in the tidal channels of the Dithmarschen Bight [Toro et al., 2005]. With the aim to establish a continuous program of measurement of currents, an electromagnetic current meter has been deployed in the Rochelsteert pier. The maximum

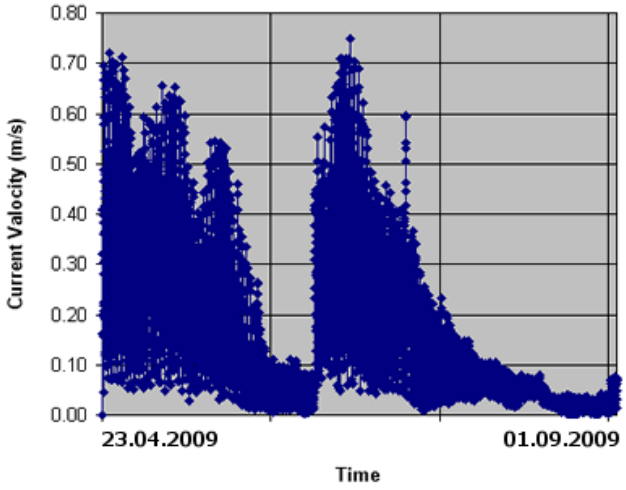


Figure 4.6: Current measured by Electromagnetic Current Meter

current measured on this location was $0.75 m/s$. As this device should be submerged all the time to perform the measurements, it is susceptible to be colonized by oceanic organisms like mollusks that reduce its measurement ability. When this is the case, a permanent maintenance is required. In the installed pile, the colonization starts to affect the measurements as soon as two or three weeks after the maintenance. This can be observed in Figure 4.6 obtained from measurements at Rochelsteert Pier. As the device is fixed, divers are required to do a routine maintenance.

4.2.7. Flows from rivers

Flows from rivers Ems, Weser, Eider and Lauders located along the coast of the North Sea at present time are assumed as permanent, however, for the Elbe, the daily flow is available for Neu-Darchau since 1874 [Geesthacht, 2008]. In the forecasting of flow, it is assumed that the flow for the next day is equal to the day before where information is already available. In case that data is not available online for a period longer than two days, the average flow for that specific day

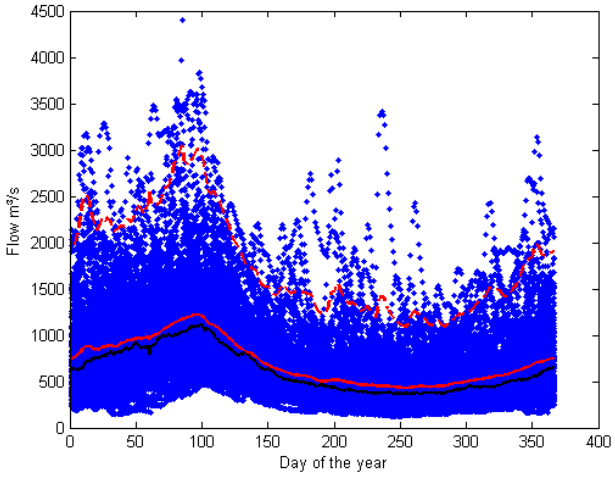


Figure 4.7: Mean Elbe flow

of the year is assumed. To calculate the daily mean value, information since 1874 has been analyzed. This mean value has a variation in a range from 400 to $1200m^3/s$ approximately, but extreme values are in the range of 130 to $3800m^3/s$ (see Figure 4.7). The standard deviation for the flow is in the range 450-650 m^3/s from January to April and 200-320m^3/s from June to October. This variability is due ice melting at the end of winter. This increases the surface runoff but with different magnitude and occurrence according to the variability of the winter conditions. In the summer a lower flow is present.

4.3. Summary

The Dithmarschen Bight is subject to a tidal cycle that creates a dry condition in the tidal flats during the ebb tide. In this condition the effect of bottom friction is high. The strong winds in the North Sea are able to create rough waves over 1.39% of the time. Historical information in the last century, shows the vulnerability of this region to large flows. Because of this, a proper knowledge of the sea state and short time prediction is a valuable tool to prevent damages.

Chapter 5

Hydrodynamic Model for the Dithmarschen Bight

This chapter presents a set of three hydrodynamic models (covering the continental shelf of the North Sea, the German bight and the Dithmarschen Bight) that interacting through a nesting sequence, captures the hydrodynamics in the study area. These models have been set in previous investigations by Mayerle and Zielke [2005]; here updates are made to incorporate more up to date information and a new wind source. At first, a brief description of the previous studies in the development of the models is given. Then calibration required for the addition of coupling of the wave and flow models follows. The updates of this model include a new bathymetry with the information until 2009 and a higher resolution -both spatial and temporal- wind field information. Improvements of the astronomical terms used as the boundary conditions in the continental shelf model that covers the North Sea (see Figure 5.1) are made, and some adjustment of the roughness and the choice of the wind drag coefficient are carried out. Finally, with the validation, the quality of the new settings of the model is discussed. In the context

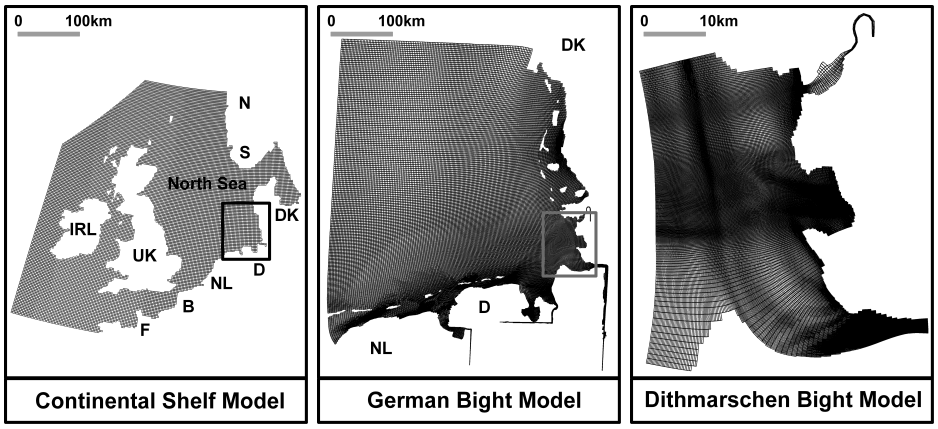


Figure 5.1: Grids used in the nesting process. from Etri [2007]

of an operational model, to avoid confusion, the hydrodynamic model is referred as a pre-operational model.

5.1. General description

Delft3D has been used to create a nesting sequence of three domains for the hydrodynamic model in order to provide proper boundary conditions for the area of interest. The models are defines as the Continental Shelf Model (CSM), the German Bight Model (GBM) and the Dithmarschen Bight Model (DBM)(see Figure 5.1).

CSM had been developed by Verboom et al. [1992] to obtain boundary conditions in deep water at the North sea. They implemented six harmonics along the boundary (M2, S2, N2, K2, O1, K1). The calibration period covers the JONSDAP-1976 measuring campaign with an extensive number of measurements available. The authors were uncertain about the bathymetry due to the lack of detailed information. For the calibration process they tuned four characteristics (local topography, bathymetry, boundary condition and shear stress). The changes in the parameters

have been made in the next sequence:

1. assume a constant roughness and make only small changes in the bathymetry to have a better local representation of the tide,
2. bathymetry and topography were kept constant while tuning the roughness, and
3. boundary conditions are adjusted within the limits of accuracy.

As it is difficult to make measurements in deep waters with the constant change of sediments present on the bed, detailed measurement of roughness is not feasible for the big areas that must be covered. Therefore, some simplifications are required to cover the modeled area. Verboom et al. [1992] used criteria based only on water depth. The roughness has been assumed as a function of the depth. The model is built in geographic coordinates, to cover a vast area.

The original GBM and DBM models were developed in 1997 for the Niedersächsisches Landesamt für Ökologie, Forschungsstelle Küste, Norderney and Amt für Land- und Wasserwirtschaft Heide by Delft Hydraulics Hartsuiker [1997]. In this report a complete description of the set-up, calibration and verification of the tidal model is presented. This was carried out with tidal information obtained during the period from 1976 to 1986. The conclusions from that study can be summarized as: the developed GBM shows a good behavior with its ability to predict the vertical tide without meteorological influence. The DBM also represented the vertical tide, but at lower water depths some differences between the measured and modeled water levels were observed. These differences are due to the schematization of the relatively narrow channels. The model also shows a similar behavior in the magnitude and direction of measured velocities. But the absolute value

of the velocity has some differences. Normally, the model is over-estimates the measurements of water levels taken at the internal channels.

In the project Promorph 2005 Mayerle and Zielke [2005] and Toro et al. [2005], decoupled flow and wave models were applied. More astronomical constituents (M2, S2, N2, K2, O1, K1, Q1, P1, NU2, L2) had been added at the boundary of the Continental Shelf Model. Meteorological information was included by the use of results from PRISMA, which is an interpolation model developed by the Max Planck Institute of Meteorology in Hamburg [Luthardt, 1987]. It provides synoptic wind fields from measurements along the coastal line and some offshore stations. The data is available in a regular mesh of 42km with a resolution of 3 hours. As this is very coarse when compared with the size of the DBM model, in the PROMORPH settings the DBM uses uniform wind. Escobar Sierra [2007] developed an approach by using model results to estimate a pattern of sediment and bed forms distribution in order to obtain an improved roughness that includes local effects.

All the three models are horizontal $2D$ models with velocity vertically averaged based on the shallow water assumptions. Rahbani [2011] has studied the same area to evaluate the use of a 3D model -with 10 vertical layers-, and found out that for the water levels the improvements are not considerable, but with an increase of the computational cost.

With the availability of new versions of Delft3D it is possible to calculate **coupled flow and wave models**. To include this feature in the system, a new process of calibration and validation for the new conditions is carried out. The wind forecast is obtained from a different source, and it is not available for the same periods used in previous studies with this model. It is required to do a comparison of the modification of the model based on new periods where the new forecast wind is

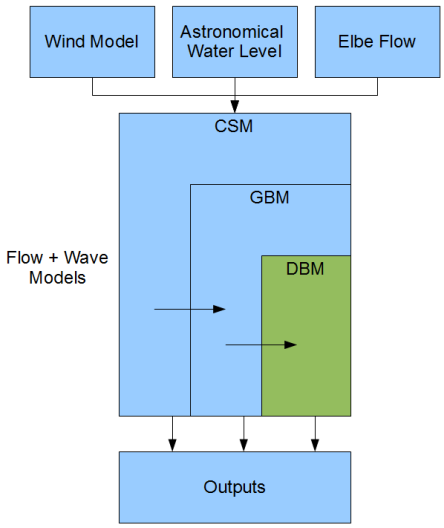


Figure 5.2: Nowcasting system description pre-operational version

available with the measured data.

5.1.1. Nesting sequence for pre-operational model

The three models shown in Figure 5.1 are executed by a nesting sequence (see Figure 5.2). First, the CSM is calculated only for the flow model. It employs astronomical water levels at the open sea boundary, the discharges of the incoming flow from several rivers along the coast line is also an imposed boundary condition and the wind field from DWD (see section 3.1) as a driving force. Once CSM computes its outputs in terms of water levels, they are passed to the next model in the process (GBM). The GBM is driven by the wind and it couples flow and waves. The river flow is also imposed. Finally, from GBM results, water levels are enforced at North, South and East boundaries. Wave values are imposed at North, West and South, while currents are employed at the West. As DWD has a finer grid than PRISMA, in the new version, a spatial wind distribution is included for the DBM. This new version of the model also couples flow and waves. All the outputs from each model are available independently of the others if required.

The Table 5.1 presents the basic setting at the starting point of the work after PROMORPH. Grids has been maintained unmodified.

Table 5.1: General Parameters used in each model

Parameter	CSM	GBM	DBM
Coordinate system	spherical	cartesian	cartesian
Grid points m-direction	201	227	268
Grid points n-direction	173	266	227
Grid	csm.grd	gbm2.grd	dbm2.grd
Bathymetry	csm-new.dep	gbm2.dep	dbm2-96.dep
Wave grid		gbm2c.grd	dbmc.grd
Wave dep		gbm2c-96.dep	dbmc-96.dep
Wave Grid points m-direction		112	134
Wave grid points n-direction		132	113
Gravity	9.813	9.813	9.813
Water density	1023	1024	1024
Air density	1.25	1	1
Spatial distribution of wind field	Variable	Variable	Uniform
wind drag coef			
First breakpoint coeff	0.00063	0.00063	0.00063
First breakpoint wind speed	0	0	0
Second breakpoint coeff	0.00723	0.00723	0.00723

Second breakpoint wind speed	100	100	100
Roughness file	c96.rgh	gbm2.man	dbm2.rgh
W roughness type	Chezy	Manning	Manning
Wave			
Gravity		9.81	9.81
Water density		1025	1025
Generation mode		3-rd gen.	3-rd gen.
Depth induced breaking alpha		1	1
Depth induced breaking gamma		0.73	0.73
Nonlinear triad interactions alpha		0.1	0.1
Nonlinear triad interactions beta		2.2	2.2
Bottom friction type		Jonswap	Jonswap
Bottom friction coeff		0.067	0.067

5.2. Sensitivity analysis

Depending on the wind condition, which is one of the elements that has more impact on the results of the water level calculation, three scenarios have been selected to carry out the sensitivity analysis and calibration of the model. One of the selected periods corresponds to low wind velocities (most of the records were lower than $5m/s$). This low wind speed period is used to create two scenarios. In

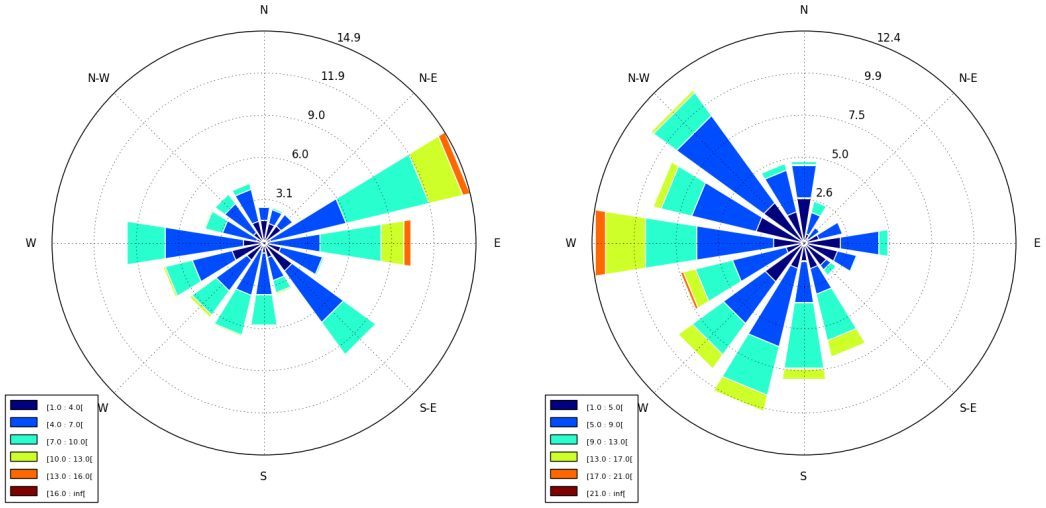


Figure 5.3: Wind conditions for selected periods. Left the calm period (from 01.04.2008 to 01.05.2008). Right the strong period (from 10.02.2008 to 10.03.2008)

one of them the model is run turning off the wind effect and is used to evaluate the importance of the astronomical tides. In a second scenario, the model is calculated taking into account the wind in the simulations. A third scenario is formed with higher wind velocities dominated by values between $10m/s$ and $20m/s$ and the model is calculated with the wind activated. See Figure 5.3.

During the calm period from 01.04.2008 to 01.05.2008 most of the wind records were lower than $5m/s$ blowing in all directions, but the strongest wind was from the coast line to the open sea (East-North-East wind), limiting its effects over the water conditions (fetch limited). In the strong wind period from 10.02.2008 to 10.03.2008, the wind was from in a wide range of directions from North-North-East to South-East, where the effects over water levels and currents were higher.

5.2.1. Warming period

The warming period is the time elapsed from the starting of the simulation, until the effects of the selected initial conditions are negligible due the importance of the boundary conditions of the model. To establish the warming period for the model sequence, three initial conditions for uniform water level (defined as Z_0) have been used to determine the period that should be discarded in the analysis at the beginning of the series. The model is carried out with an uniform water level (Z_0) in the whole domain equal to 0, 1 and -1m related to the NULL water level. In Figure 5.4, results for one month are presented for the cell 139x95 in the CSM model corresponding to the North-West corner of the boundary of GBM. This has been selected as a reference cell to know what is transferred to the smaller domains. Here it is evident that during the first two days, adjustments in the model are pronounced, making values totally useless. As time advances, the three models start to converge to a close value of the water level for this point.

In the Table 5.2 the time required to reach some preselected tolerances in water levels and current velocities are presented, showing that a period of at least 5 to 6 days is necessary to guarantee very small differences in the water level and currents. For further comparisons the first ten days of model results are discarded, as the differences in the modeled results are lower than the resolution of available measurements which has the least significant digit in order of one centimeter.

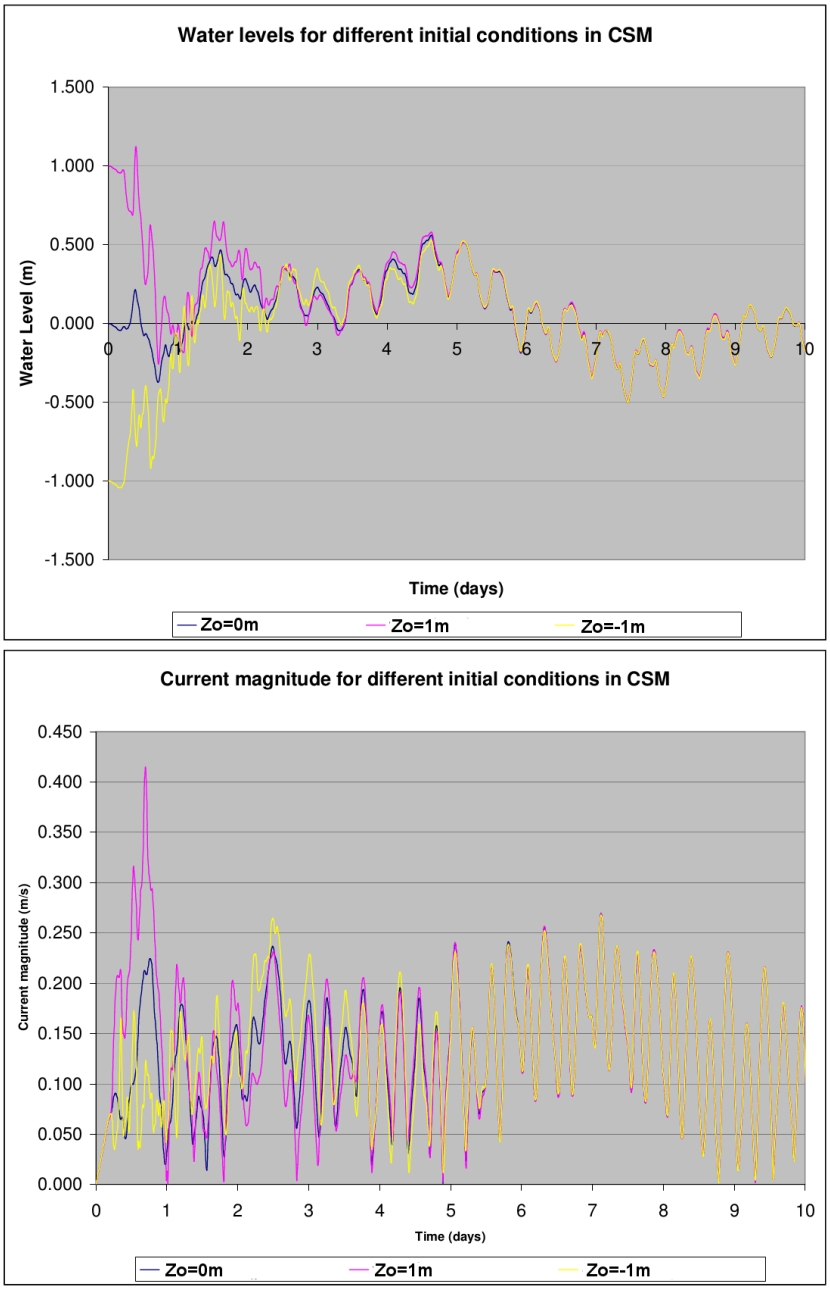


Figure 5.4: Warming period for CSM model

Table 5.2: Error magnitude and time required

Water level difference (cm)	Modeled time required (days)	Current magnitude difference (m/s)	Modeled time required (days)
$< 2,00$	5	$< 0,020$	5
$< 1,00$	10	$< 0,010$	8
$< 0,50$	15	$< 0,005$	10
$< 0,01$	30	$< 0,001$	20

5.2.2. Updating the bathymetric information

For the area of DBM model, two complete sets of bathymetry are available from previous work of Hartsuiker [1997], one corresponding to the information available until 1977 and other until 1996. A new updated bathymetry for 2009 is created based on the information of several sources as is mentioned in section 4.2.1. The differences between the 1996 and 2009 bathymetries are presented in Figure 5.5. In the flat areas there is no significant change, but a displacement in the location of the channels can be observed. This can be evidenced by the erosion and deposition accumulated in parallel along the channels. Comparisons are carried out to see the effects of the change in the bathymetry on the results. A flow model using three sets of bathymetries excluding wind is employed to observe how the system responds to the new bathymetry; the model outputs are compared with astronomical tide. The astronomical tide is obtained after an harmonic analysis of the water level measured. It does not include the effect of meteorological components. Between the results using 1977 and 2009 bathymetry, the maximum difference obtained in water levels are $2cm$ in Brunsbüttel, $12cm$ in Büsum and $11cm$ in Cuxhaven, while these differences are smaller between the results using 1996 and 2009. The average error for the astronomical tides are presented in Table

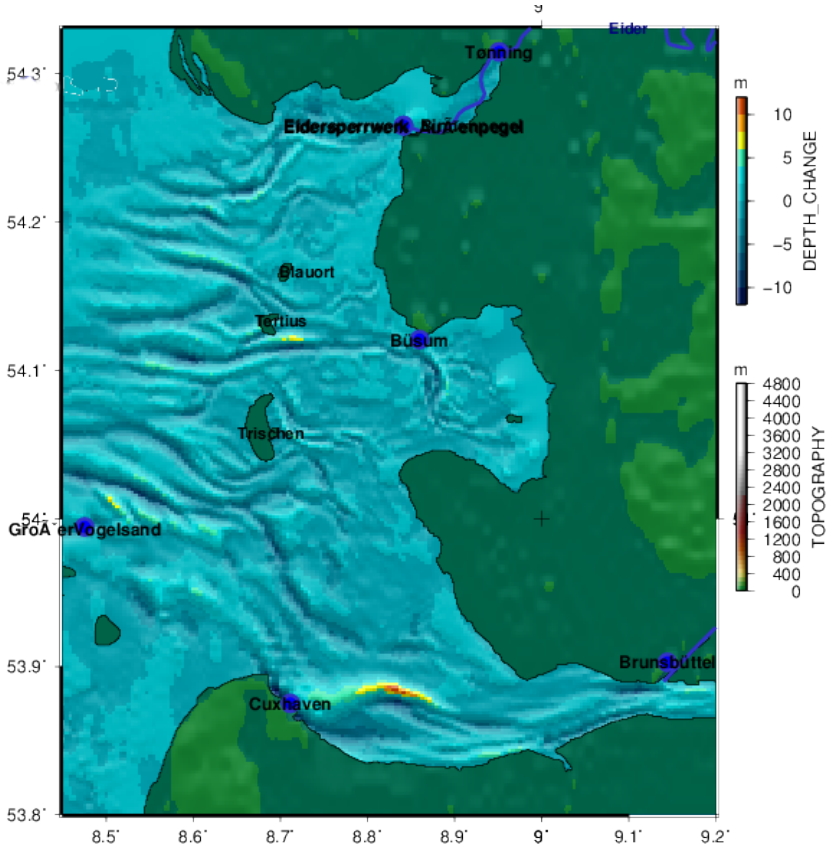


Figure 5.5: Difference in bathymetry: positive erosion, negative deposition

5.3. One positive result of using the new bathymetry is that the observed shape of the water level curve is better with the 2009 bathymetry (see Figure 5.6).

For the extreme values of the water level during the flow and the ebb tide (see Table 5.3), the averaged results for the stations are taken into account. Here the average absolute error in the water level and the time difference between the model and the astronomical tide calculated from measurements is considered. When all the samples are considered, the average error for water level is reduced to around one centimeter at Büsum and Cuxhaven, while at Brunsbüttel it is increased by the same amount. For the flow tide the error in the water level obtained with the new bathymetry is between the errors obtained from 1977 and 1996 bathymetries. In the time there is a reduction of one minute from the results obtained with the bathymetry of 1996, and three with respect to 1977. For the ebb tide the reduction in the time shift is even better, improving eight minutes with respect to 1997 and

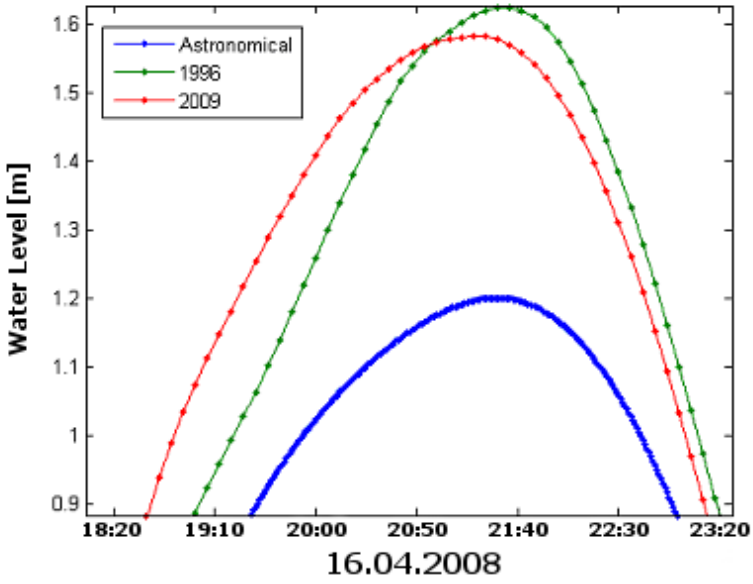


Figure 5.6: Effects of changing the bathymetry available on 1996 to an updated with information until 2009 in the shape of the flow tide, also compared with the astronomical tide obtained from a harmonic analysis of measurements at Büsum

three minutes with 1996. Errors in the water levels at the ebb tide are reduced by two centimeters. It is possible to see that in general, improvements in the water level are not more than two centimeters (with a mean tidal range of around three meter, represents that the errors are reduced from $4,0\%$ to $3,3\%$), but for the time at which the flow and ebb tides are reaching their maxima heights there is better agreement.

The inclusion of wind for the calm period in the model allows to make a comparison of water levels with measurements. For the enhancement obtained from the new bathymetry using the DWD model wind, the average of the absolute error considering every sample is calculated in Table 5.4 for Brunsbüttel, Büsum and Cuxhaven. The new bathymetry has a better performance for Büsum, which is inside the Dithmarschen Bight, and also for Cuxhaven reducing the error in the water level by one centimeter. In the case of Brunsbüttel, there is an increase in the average error but it is smaller than half centimeter.

Table 5.3: Comparison of average error in ebb and flow tide (cm) for different bathymetries comparing model results without imposing wind and astronomical tides

Test case	Time lag (min) flow tide	Water level for flow tide (cm)	Time lag (min)ebb tide	Water level for ebb tide (cm)
1977	10.47	11.96	29.77	12.13
1996	8.64	10.09	24.51	11.30
2009	7.73	11.19	21.37	09.89

Table 5.4: Comparison of average error in water level (cm) at different locations

Test case	Brunsbüttel (cm)	Büsum (cm)	Cuxhaven (cm)
1977	22.50	18.86	19.68
1996	23.53	18.16	20.80
2009	23.90	17.97	18.57

In the Tables from 5.5 to 5.13, the root mean square error (RMSE) and the absolute value of the average of the errors are aggregated for the four points used to evaluate the model performance. Such points are Brunsbüttel, Büsum, Cuxhaven and Großer Vogelsand. The estimation of errors is carried out by evaluating all the samples in the simulation time, and on only the extreme water levels at ebb and flow tides are considered(as has been described in Section 3.6 and in Figure 3.10). Also a multi-objective error is calculated as weighted averaged of the RMSE and mean values of error. The selected weights are chosen to give a higher importance to reduce the errors in the flow tide. Then, weights equal to one are applied to the mean RSME for the set of all samples in the series and for the mean RMSE of the ebb tide. A weight of two was used for the mean error of all samples, ebb and flow tides and also for the RMSE of the flow tide.

In Table 5.5, the water levels obtained by the use of the bathymetry with information from 1996 and from 2009 are compared when the low speed wind scenario is considered. For the scenario without wind the new bathymetry is providing better estimations of the water level. There is not a clear advantage of the new bathymetry, but it is important to keep in mind that the previous settings were tuned for a different wind source (PRISMA see section 5.1), that was coarser both in time and space. Because of this, the wind drag coefficient has to be improved. The activation of the tidal forces formulation of Delft3D, improves the quality of the water level results with the new bathymetry. Therefore, from this point the bathymetry from 2009 with the tidal forcing activated is the base for further calibrations.

Table 5.5: Water level averaged errors and RMSE for the four stations considered in the evaluation of DBM model with low wind speed

Test case (cm)	RMSE full (cm)	error full (cm)	RMSE ebb (cm)	RMSE flow (cm)	error ebb (cm)	error flow (cm)	multi-objective (cm)
Original b.1996	18.40	7.83	25.44	14.64	20.20	9.21	14.76
b.2009	21.33	8.35	27.16	13.36	23.92	6.56	15.29
b.2009 plus tidal forces	20.14	7.28	28.32	12.92	26.10	0.24	14.15

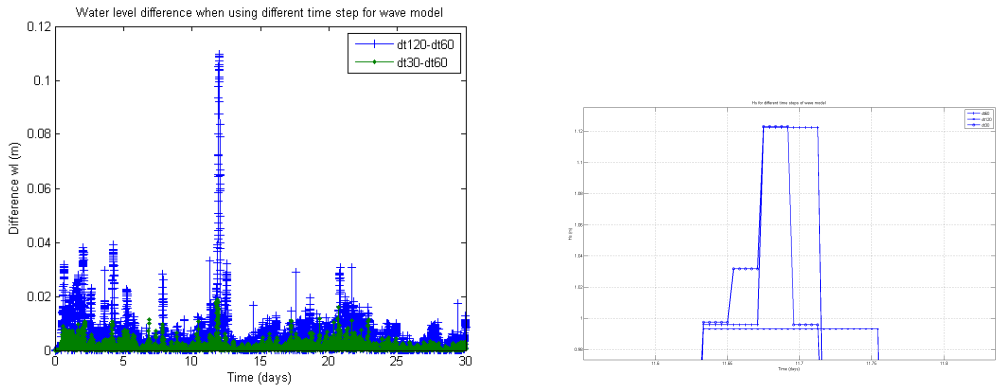


Figure 5.7: Effect of the time step for wave model, in the left the difference in the water level between using 60 and 120minutes and 60 and 30 minutes. In the right a zoom showing the significant wave height obtained by using 30,60 and 120minutes

5.2.2.1. Sensitivity of the time step for wave modelling

In the coupled flow and wave model, the wave characteristics are calculated at some predefined intervals taking into account the wind field, currents and water levels at that moment. The resulting wave field is kept constant and imposed to the flow simulation until a new wave field is calculated. The wave model demands more processor (CPU) time than the flow model; critical at this point is the search for the accurate results without involving a high amount of computational time to calculate the best results of the waves in the coupled model. A sensitivity analysis that takes time steps for the wave model of 180, 120, 90, 60, 30 and 12 minutes has been used. Here, 180 minutes corresponds to the one employed during the PROMORPH project, while 60 minutes is the temporal resolution of the available DWD forecast. An analysis of the model results shows the high influence of the time step used for the wave model on the wave results, and, a lower impact on the water levels. The wave results with a resolution coarser than 60 minutes show how they are similar to those obtained for 60 minutes. But when the peak of a storm occurs outside of this 180 minutes interval sampling, it is not captured (see Figure 5.7). However, with a resolution of 30 minutes, the wind field is not changing and only flow conditions are updated. With the resolution time of 30 minutes the obtained significant wave height H_s is an intermediate value of the one obtained by the 60 minutes interval. The difference in the coupled water level exists only in terms of millimeters. Bearing this in mind, a resolution of 60 minutes has been adopted as it is the highest wind resolution. With 60 minutes the model is able to capture the storm peaks in a more effective way than the 180 minutes available in PROMORPH. The mean absolute error between the water levels using a time step of 30 and 60 minutes is 0.14cm, which is negligible, and when the comparison is made between the time step of 60 minutes as the one with 120 minutes it is 0.50cm (with a peak of 12cm) values that are also low. The reason for the peak in the wave model is that a wind peak occurred outside of the sampled wind used

for the wave.

5.3. Calibration

After introducing a new bathymetry with the information until 2009, a recalibration of the model is done. This calibration is required to include the new wind input and the effect of coupling the hydrodynamics and wave models. To obtain the final settings of the model tuning of the roughness, the wind drag coefficient, and the inclusion of astronomical boundaries obtained from TOPEX are considered. This is an iterative process where only one parameter is slightly modified at each time, while keeping the other parameters with a constant value.

5.3.1. Wind drag coefficient with low wind speed

Delft3D permits to introduce the wind drag coefficient expressed as a piece-wise linear function of wind drag and wind speed [WL , 2003a]. In the original setting, it is defined that at $0m/s$ the coefficient is equal to 0.00063, and grows linearly until a value of 0.00723 at $100m/s$. For the low wind speed condition, the wind drag coefficient has been changed by multiplying the values provided in the original setting (0.00063 and 0.00723) by a factor. The results of using wind drag coefficients taking factors of 2.00, 1.25, 1.10, 1.00, 0.90, 0.75 and 0.50 are presented in Table 5.6. The test case with a coefficient of 1.00 is used as a reference and it is the same presented in Table 5.5 with the bathymetry of 2009 and the tidal forces activated.

Table 5.6: Averaged errors and RMSE of water levels for the four stations considered in the evaluation of DBM model. Wind drag coefficient for low wind speed

Test case factor, first and second wind drag coef.	RMSE full (cm)	error full (cm)	RMSE ebb (cm)	RMSE flow (cm)	error ebb (cm)	error flow (cm)	multi-objective (cm)
2.00 (0.00126-0.01446)	23.36	2.42	31.32	20.51	24.91	3.12	15.66
1.25 (0.00079-0.00904)	20.22	6.16	28.45	14.12	25.56	0.44	14.12
1.10 (0.00069-0.00795)	20.08	6.89	28.26	13.20	25.80	0.08	14.03
1.00 (0.00063-0.00723)	20.14	7.28	28.32	12.92	26.10	0.24	14.15
0.90 (0.00057-0.00651)	20.15	7.86	28.16	12.47	26.10	0.86	14.29
0.75 (0.00047-0.00542)	20.40	8.59	28.29	12.27	26.41	1.42	14.61
0.50 (0.00032-0.00362)	21.21	9.73	29.07	12.77	27.19	2.12	15.39

The amplification factor 1.10 (wind drag coefficients 0.000693 and 0.007953) is the only case that is able to improve the different measurements of the error when compared with the reference case (factor 1.00). An exception is the RMSE of the flow tide. The new wind drag coefficients corresponding to the coefficient of 1.10 are adopted for the evaluation of the model. For bigger changes in the wind drag coefficient, the results are all worse than the original. To understand these results, it ir required to consider that the effect of low wind conditions over the pure astronomical water levels is small, and that the model was previously calibrated for low wind speed conditions. Then effect of changing the wind data source with which the hydrodynamic model was previously calibrated should not leads to a big change in the wind drag coefficients for this scenario.

5.3.2. Roughness

For the roughness a procedure similar to that for the wind drag coefficient is employed. The original roughness map obtained from the PROMORPH project has been modified by multiplying the data at each cell by a constant value. This multiplication factor is selected in the range from 0.8 to 1.2. Then the model using the new bathymetry and the low wind condition (wind speed $< 5m/s$) is taken as the reference. In Table 5.7 five cases are presented. As it can be observed with a factor of 0.9, the RMSE and the mean error for all the measurements and ebb tide, the modeled water levels get a better agreement than the original setting. The error for the flow tide has been slightly increased, but overall the model with 0.9 times the original roughness is considered appropriate.

There is a big impact of roughness on the hydrodynamics; a way to improve the results is by taking a more accurate roughness. This can be done by doing local changes taking into account the properties of the sediment in the area. With the available information this level of detail cannot be updated in the present work.

Table 5.7: Averaged errors and RMSE for the four stations considered in the evaluation of DBM model. Roughness coefficient for low wind speed

Test case factor roughness	RMSE full (cm)	error full (cm)	RMSE ebb (cm)	RMSE flow (cm)	error ebb (cm)	error flow (cm)	multi-objective (cm)
1.2	23.37	7.53	33.17	12.82	31.40	5.38	17.08
1.1	21.34	7.45	29.95	11.68	28.00	1.53	14.86
1.0	20.14	7.28	28.32	12.92	26.10	0.24	14.15
0.9	19.33	6.87	27.03	13.05	24.49	1.40	13.80
0.8	19.14	7.12	26.56	15.88	23.80	1.89	14.31

5.3.3. Combination of upgraded model parameters and Wave activation

After getting improved values of the wind drag coefficient and roughness, they are combined to know the improvements on the model when both are used at the same time, as well as the activation of the wave model to observe the impact on the model quality.

In Table 5.8 the results from three models are presented. The first one is the combination of the new bathymetry including the low wind speed conditions and the tidal forces. This is the same used as a reference for the improvement of the wind drag coefficient and roughness in the two preceding subsections. The second includes additionally the new roughness and wind drag coefficient from sections 5.3.1 and 5.3.2. In the third case the wave model is activated. The results in the three cases are close to each other with differences of millimeters. This shows that after the positive impact of improving the bathymetry, the changes in the wind drag coefficient and the roughness of the original model is not changing too much the results for low wind speed conditions. This small effect for low wind condition is explained by the fact that the original work already tuned the model for astronomical conditions, and with low wind speed, the effect of the wind on the water levels is not significant.

Table 5.8: Averaged errors and RMSE for the four stations considered in the evaluation of DBM model. Effect of activate the wave coupling for low wind speed and using the bathymetry of 2009

Test case	RMSE full (cm)	error full (cm)	RMSE ebb (cm)	RMSE flow (cm)	error ebb (cm)	error flow (cm)	multi-objective (cm)
1*	20.14	7.28	28.32	12.92	26.10	0.24	14.15
2*	19.77	7.25	28.45	12.47	26.29	0.67	14.16
3*	19.65	6.32	28.90	15.05	26.43	2.18	14.85

1*. Old roughness and wind drag coefficient, excluding waves; 2*. new roughness and wind drag coefficient, excluding waves; 3*. new roughness and wind drag coefficient, including waves.

5.3.4. Improve the water level boundary conditions with astronomical constituents from TOPEX

Due to the large shift in time, an attempt to change astronomical boundary condition at the CSM has been made. New Boundary conditions were taken from the information coming from TOPEX/Poseidon [Egbert and Erofeeva, 2002]. For the astronomical parameters, amplitudes in both datasets were very close to each other but differences were found in phases. Those differences are equivalent to a range from -17 to 30 minutes in the phase of each component.

The analysis of this change in the boundary condition is taken out in two ways, one compares the model outputs with measurements while the other compare model outputs with the water levels obtained from a pure astronomical prediction based on the harmonics of the station. By comparison with measurements, the phase when the ebb tide occurs is increased, but the difference in the tidal range is

reduced. Then, the new constants have a better combination of phases that fit better to the shape of the tide, but are introducing a phase in time.

When comparisons are made with the astronomical reconstruction of tides at Büsum and Cuxhaven, it is observed that boundary condition imposed by TOPEX information tends to estimate the occurrence of peaks around 12 minutes earlier than the original settings. Even that all the other comparisons are made keeping original boundary conditions, a coherent shift of all the astronomical phases helps in improve the ability to predict time of peaks.

If TOPEX is included for the low wind speed (see Table 5.9) on top of the case that includes wave coupling, the new bathymetry, the improved roughness and wind drag coefficient, the RMSE for flow and ebb tide is reduced to a value between 2.64 and 3.63cm. Also there is a reduction in the ebb tide mean error, but an increase in the mean error of flow tide and in the values for the full series comparison. The last is a consequence of the time shift in the tide. It is possible to adjust the phase of each harmonic considering the angular speed of each one, keeping the shape of the tide but reducing the time shift. This correction keeps the error calculated for the water level amplitude for the flow and ebb tides, but reduce the error calculated for the entire time series.

Table 5.9: Averaged errors and RMSE for the four stations considered in the evaluation of DBM model

Test case	RMSE full (cm)	error full (cm)	RMSE ebb (cm)	RMSE flow (cm)	error ebb (cm)	error flow (cm)	multi-objective (cm)
1*	19.65	6.32	28.90	15.05	26.43	2.18	14.85
2*	22.55	8.31	25.27	12.41	22.48	4.91	14.40
3*	19.25	5.81	25.27	12.41	22.48	4.91	13.58

1*. New roughness and wind drag coefficient, including waves; 2*. new roughness, wind drag coefficient and TOPEX, including waves; 3*. new roughness, wind drag coefficient and TOPEX, including waves

5.3.5. Wind drag coefficient at high wind speed

For high wind speed, a new evaluation of wind drag coefficient is made. The evaluated factors are 1.25, 1.10, 1.00, 0.90 and 0.75. The reference set is taken from the improved version obtained from the low wind speed conditions, but calculated with the high speed winds. The reference case is shown in Table 5.10 with the coefficient equal to 1.00.

In this scenario, increasing the wind drag coefficient also increases the error in the results. The reduction of this parameter improves the results. The best results for the extreme values are obtained by a coefficient of 0.75, although that it increases the RMSE and the mean error for the full series. The coefficient 0.90 keeps the mean error lower, providing an improvement with respect to the reference case. Finally the first breakpoint coefficient is increased to 0.000693 (as the result of the product 1.1*0.00063) and the second breakpoint coefficient is reduced to a conservative value of 0.006507 (as the result of the product 0.90*0.00723).

Table 5.10: Averaged errors and RMSE for the four stations considered in the evaluation of DBM model changing the wind drag coefficient with high wind speed

Test case factor, first and second wind drag coef. (-)	RMSE full (cm)	error full (cm)	RMSE ebb (cm)	RMSE flow (cm)	error ebb (cm)	error flow (cm)	multi-objective (cm)
1.25 (0.00079-0.00904)	72.43	7.51	29.97	29.72	25.37	24.07	27.57
1.10 (0.00069-0.00795)	70.45	1.01	26.94	25.66	21.99	19.63	23.40
1.00 (0.00063-0.00723)	69.72	3.40	25.18	23.04	19.92	16.41	22.05
0.90 (0.00057-0.00651)	69.58	7.69	23.52	20.88	17.57	13.62	21.26
0.75 (0.00047-0.00542)	70.36	14.24	21.57	18.08	14.29	9.06	20.33

Table 5.11 presents a comparison of the original model, the model updated for low wind conditions, and the final version including the change in the wind drag coefficient including TOPEX boundaries and activating the tidal forces. It can be seen how the final version is able to improve all the criteria used to estimate the model quality in the four stations, except for the mean error of the full time series which is very close to the version obtained for the improved only in low wind condition.

Table 5.11: Averaged errors and RMSE for the four stations considered in the evaluation of DBM model with high wind speed

Test case	RMSE full (cm)	error full (cm)	RMSE ebb (cm)	RMSE flow (cm)	error ebb (cm)	error flow (cm)	multi-objective (cm)
Original	70.86	6.38	28.64	28.25	23.81	24.44	26.53
Imp. low wind	69.72	3.40	25.18	23.04	19.92	16.41	22.05
Final	68.75	3.46	23.90	20.79	17.11	13.70	20.28

5.4. Validation

To measure the ability of the model to evaluate the flow conditions inside the domain a new period is taken into consideration. June and July 2008 are selected due to the occurrence of different low, medium and high wind speed conditions. The nesting sequence has been evaluated once. The summarized error estimation parameter for the four stations is shown in Table 5.12. The overall performance of the model for the calibration is in the same range of the errors obtained for the calibration. This is important, as the model is able to deliver consistent outputs to other periods than the used for calibration. The discrimination of the error measurements for each station is provided in Table 5.13. The mean difference in the estimation of extreme values of tides is lower than $2cm$ for Cuxhaven; Büsum and Großer Vogelsand are in the range of $4-6cm$ which it is still a good result. Brunsbüttel on the other hand has a mixed quality, because the error of the flow is the lowest of the compared stations with only $0.66cm$. For the ebb it reaches the biggest error with $27.66cm$. The dispersion of the errors in the extreme tide is similar at the four stations, except of the ebb tide at Brunsbüttel. When all the series are considered, the mean error is around of $5cm$ to $8cm$, except for Cuxhaven where it is only $0.86cm$. The dispersion of the errors is more favorable for Brunsbüttel with a two thirds of the RMSE obtained for Cuxhaven.

Table 5.12: Averaged errors and RMSE for the four stations considered in the evaluation of DBM model

Test case	RMSE full (cm)	error full (cm)	RMSE ebb (cm)	RMSE flow (cm)	error ebb (cm)	error flow (cm)	multi-objective (cm)	correlation
Validation	28.96	5.28	15.29	13.75	8.08	3.37	10.52	0.97

In Figure 5.8 the x-axis represents the modeled water level and the y-axis the measured water levels for each one of the four stations. Brunsbüttel shows a

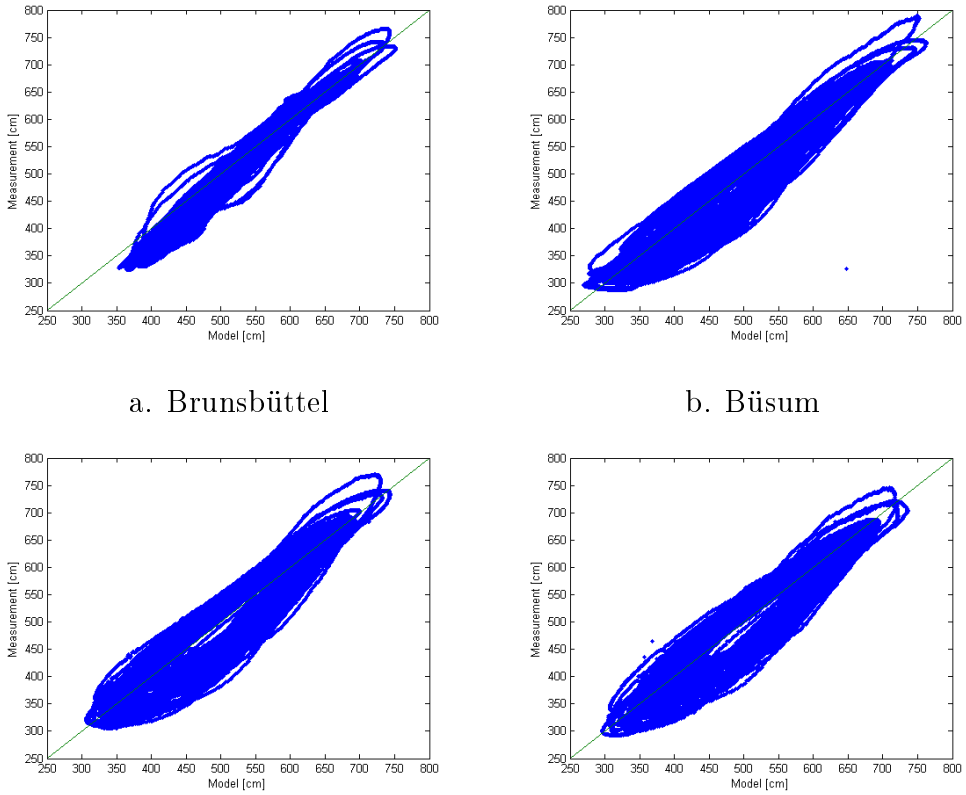
Table 5.13: Error in DBM with nesting and corrections for July 2008

Parameter	Brunsbüttel	Büsum	Cuxhaven	Goßer Vogelsand
RMSE full (cm)	20.40	28.66	33.53	33.38
Error full (cm)	6.96	5.01	0.86	8.47
RMSE ebb (cm)	29.94	12.04	10.51	8.65
RMSE flow (cm)	15.16	17.14	11.30	11.38
Error ebb (cm)	27.66	4.79	1.95	-2.09
Error flow (cm)	0.66	6.53	1.43	4.86
Correlation	0.99	0.97	0.95	0.96

slender scatter with a correlation of 0.99. The over prediction of the water levels close to the ebb tide is appreciated by the shift to the right of the 45° line in the low water level. In Cuxhaven and Großer Vogelsand and to a lower level in Büsum, it is formed an elliptical path that reduces the amplitude close to the ebb and flow tide, but presents a hole in the intermediate values. Over-prediction (shift to the right) during the ebb and under-predicting (shift to the left) during the flood is shown in the same figure. From those results, it can be said that the model works reasonably good for flow and ebb tides. However, intermediate values still need to be improved.

5.5. Summary

In this chapter an improved set of models has been calibrated and validated. A better performance in the estimation of water level (see Table 5.11) and in the occurrence time of flow and ebb tides is obtained. With this set of models it is possible to represent in an adequate form the water levels in the Dithmarschen Bight. The work in a more detailed bathymetry has been able to improve the



a. Brunsbüttel

b. Büsum

c. Cuxhaven

d. GoßerVogelsand

Figure 5.8: Model results versus Measurements during the validation period in July 2008

estimation of the ebb and flow water level, the occurrence time of it (eight minutes for the ebb and 3 minutes for the flow tide) and even the shape of the tidal series. But still the scarce availability of detailed roughness information and the restriction in the grid size (to keep fast results) limit the capacity of the model to fit better to measurements. In the next chapter data assimilation strategies are considered to improve even more the model results.

Chapter 6

Data assimilation in the Dithmarschen Bight model

In this chapter the optimization of boundary conditions for the set of nested models described in Chapter 5 is developed. First an explanation of the criteria to select the correction of boundary conditions as the data assimilation strategy to be used is given. These corrections are made by defining functions that connect one model cell with cells located at the boundary. These functions are called the local models as they consider a single location instead of a 2D or 3D domain. To build those functions, the spatial solution for every time step is considered as a boundary value problem in which it is desired to find out what boundary condition is associated with one specific value at an observation point. The relationship between the observation point and the points at boundary is analyzed in the light of dynamical systems, where as has been described in section 3.3.2, temporal dimension is removed. Also an analysis of the spatial behavior of water levels, waves and tides is done by the use of correlation maps. Through them it is possible to visualize the regions that are well represented in the model by a designed

observation cell in the domain. This is helpful to understand the differences in the patterns of one parameter inside the domain, and provides information to do a zoning of representative areas that will help in the design of local models. Three methods are considered to build such local models: Linear regression with time shift, an adaptation of autoregressive models and artificial neural networks to deal with nonlinear relationships. Local models for reconstruction of missing information (hindcast) and for doing forecast covering periods from 2 to 24 hours are made. Fuzzy logic is evaluated for easiness in the development of forecast models where wind direction is considered. A sensitivity analysis for the ANN and an ensemble of several ANN to reduce the errors is discussed. Final comparison between the validation of the process-based-model and the assimilation for the same period is presented.

6.1. Assimilation strategies used in the present work

In section 3.2 are presented different alternatives for doing the data assimilation of measurements with the numerical model. For this work, the improvement of boundary condition is selected. The modification of the boundary condition is compatible with the restrictions of the numerical solver tool (Delft3D), and at the same time as is will be illustrated in this chapter, it is able to provide a complete set of optimized results. The defined strategy comprises the optimization of the boundary and initial conditions as they are less intrusive with the model engine (solver) than the Kalman filter or variational approaches. The modification of the boundary condition is capable of providing spatial improvements in the whole domain, in contrast to model output statistics that are calculated only for locations where measurements are available. Also the modification of the boundary conditions can be used to improve studies in sediment transport or geomorphology keeping the physical formulation developed inside the numerical model. For the

proposed modifications of boundary condition, the translation of information to the boundary based on the analogy with the shooting method and two members of the family of artificial intelligence tools are in use (fuzzy logic and artificial neural network).

Ensemble methods are considered in the construction of a local model for the water level forecast based on artificial neural network ANN. The evaluation of multiple scenarios in the coupled hydrodynamic model is computationally expensive, but the fast calculation of a forecast with ANN makes it an easy task to be performed. Then with this ensembled boundary condition it is possible to run the process based model only once to obtain the improved version.

6.2. Translation of information to the boundary

In the evaluation of a numerical model, the boundary condition can be fed with information from measurements or results from other models. In the ideal case, such observations are obtained at the same places where boundary conditions are located. Normally it is not possible due to the differences in the requirements for locating the boundaries of the model and for the suitable sites for deploying measurement devices. The location of a model boundary should be placed far enough from the area of interest, to reduce the impact of the artifacts created by the imposed value on the results. Also zones with low gradients are preferred as they reduce the risk of numerical instabilities. On the other hand, the location of measurement devices is subject to some restrictions. Some stations, for example, are already placed in advance of the development of the model and cannot be moved. The selected location also must provides protection against environmental hazards and vandalism. An easy access to the station, power supply and

communication (see section 3.1.1) must also be considered.

Standard algorithms for spatial interpolation like Triangular Irregular Network (TIN) or inverse of the distance do not recognize the anisotropic behavior of the parameter being evaluated. This is the case for water levels, currents and waves as it is explained later, when building correlation maps in section 6.2.1. One of the main advantages of having a numerical application is the availability of a spatial view of the physical processes that take place in the whole domain. When a good model is obtained, it is possible to evaluate several scenarios and find a relation between two or more points from them or even all of them together. After finding this relation, it could be used for translating the values without employing standard spatial interpolation algorithms.

In section 3.2.5 it was demonstrated by an example how a boundary value problem can be solved as an initial value problem by the shooting method. In a stationary BVP, the inner solution is defined by the established boundary conditions. Once a stationary solution is obtained, it is possible to find relations among several points inside the domain and a set of boundary conditions. For the solution of a BVP the question to be solved is: once a boundary condition is provided, what is the value of one parameter in a cell in the domain? However the opposite question leads to the correction of the boundary condition. What is the boundary condition that need to be provided in order to obtain a specific value at the cell where measurements are available?

For the IBVP, a simulation period long enough to cover a significant number of scenarios can be evaluated. It is done by considering that each time step is a snap shot that captures a stationary problem. This image implicitly carries information of previous states. Then it can be created a function that connects two points in the model's domain. Considering each stationary problem as an independent

scenario, it is possible to take values from two locations and build the space phase of them. Special interest is given to the boundaries and gauges locations, where correction functions are built with the model results. In the case of water level a general function takes the form:

$$WLB_t = f\left(WLG_t\right) \tag{6.1}$$

where WLB_t is the water level imposed by nesting in the boundary condition and WLG_t is the water level calculated by model in the node corresponding to the gauge or observation point.

For a modified value for the gauge water level (WLG^*) and the corresponding water level at the boundary (WLB^*), it can be written:

$$Error_{WLB_t} = WLB_t - WLBC_t^* = f\left(WLG_t\right) - f\left(WLG_t^*\right) \tag{6.2}$$

Rearranging the terms to obtain an explicit expression for (WLB_t^*)

$$WLB_t^* = WLB_t - \left(f\left(WLG_t\right) - f\left(WLG_t^*\right)\right) \tag{6.3}$$

In this equation WLB_t and WLG_t are known from the first run of the model, using nesting; and the WLG_t^* should be obtained from measurements -hindcast- or from a local model -hindcast and forecast-. With this new value for the boundary conditions, the process based model can be evaluated to obtain an improved result.

6.2.1. Spatial correlation maps

Based on model outputs, the level of linearity among different cells can be determined. For this purpose, for given time series for each cell in the model, the correlation for several time lags is calculated between a reference point and all the others in the domain. The maximum correlation and the required time shift is stored and plotted in a map. If a high level of correlation is found, it means that this specific cell provides a good linear representation for the area where this value is reached. In case of low correlations it can be considered as an indicator of the presence of nonlinear relations or the total absence of relation between both cells.

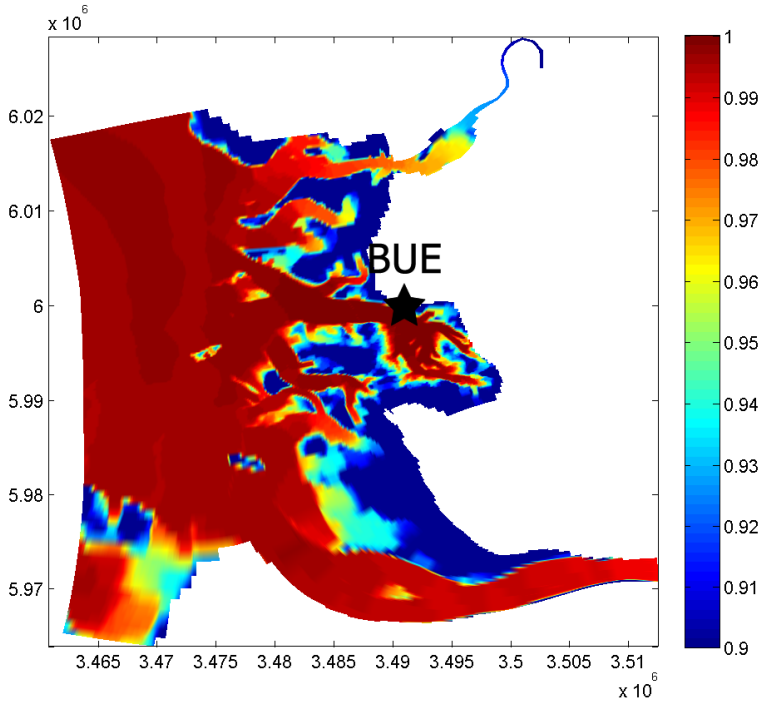
Correlation maps are created by plotting the maximal correlation obtained for a limited interval of lags. Also a time shift map is obtained by plotting the time shift when the maximum correlation is obtained. Here these maps are used to understand the variability in the linearity between the available stations and the rest of the domain. When high correlations are obtained, linear methods can be selected, but with low correlations nonlinear tools must be employed. Additionally the time shift must be considered in the analysis because even a good correlation but with a big delay will make more difficult to implement an assimilation for near real-time modeling. Maps containing the maximum correlation obtained from the previous step as well as the time shift required to obtain this correlation are plotted. The maps for stations located at Büsum, Cuxhaven, Rochelsteert and Brunsbüttel and the model values for water level, waves and currents are constructed (see Figures 6.1, 6.2, 6.4 and 6.3) where only correlations higher than 0.9 and time delays lower than 100 minutes are considered. From those maps it can be seen how the spatial linear behavior of each parameter differs around the domain and what is the potential use of nonlinear tools in case of a low linear relation.

6.2.1.1. Correlation maps for water level model outputs

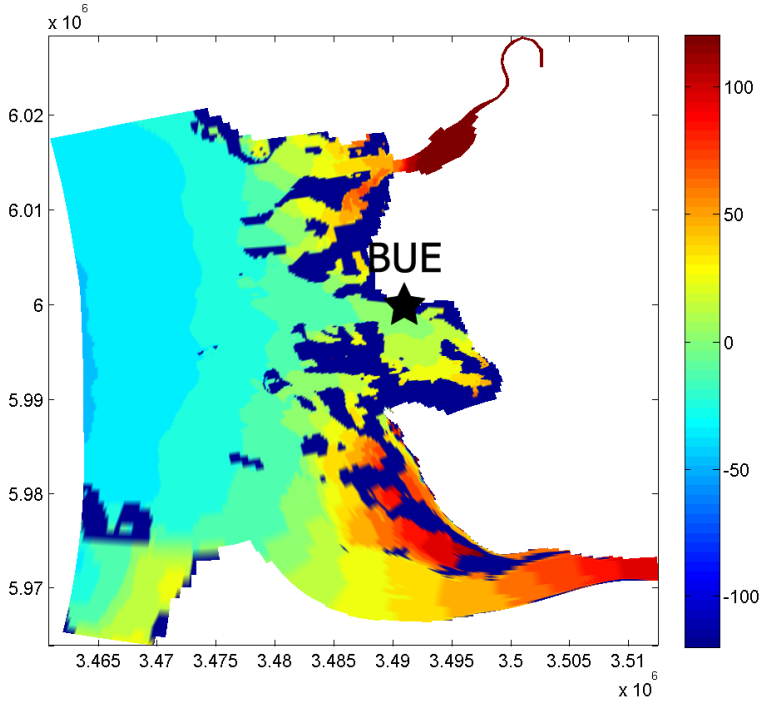
The maximum correlation between Büsum station and the rest of the domain is presented in Figure 6.1. A high correlation is obtained for points that are not necessarily in the vicinity of Büsum. Points located in the tidal flat on the other hand, have a lower correlation than the selected limit of 0.90. The Eider region does not show a good correlation either.

For water levels (see Figure 6.2) a relatively high correlation can be obtained for all the points except Brunsbüttel, which is located inside the Elbe estuary and is so more influenced by river currents. The biggest difference observed here is the time shift required to obtain this maximum correlation. The time shift plot resembles the part in the tidal cycle in which one region is contained. It means that regions with the same color are having the peak of the flow tide around the same time. The Rochelsteert pier is very representative for the water levels in the deeper areas of the Dithmarschen Bight model. Büsum has the shortest shifts in the intermediate zone where the Piep channels are, but for open sea areas it could require 50 minutes to approach to a linear relation. Brunsbüttel represents the worst case, showing that high correlations can occur only in the vicinity of this station.

In this work only corrections for water levels are presented, but to complete the analysis, waves and currents are also included.

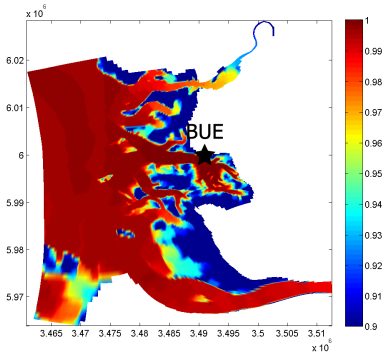


Maximum correlation at Büsum

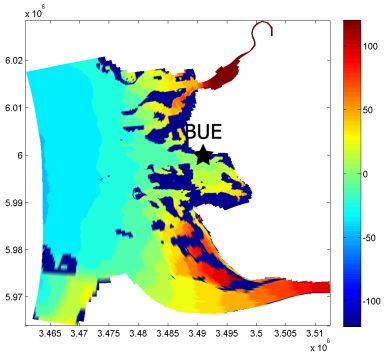


Time shift at Büsum

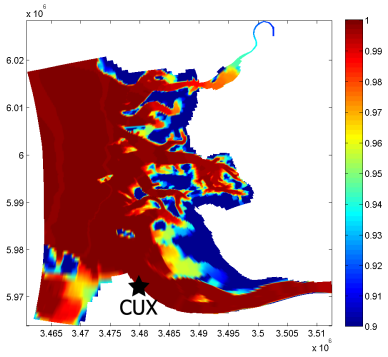
Figure 6.1: Maximum correlation coefficients and time delay in minutes for water level at Büsum



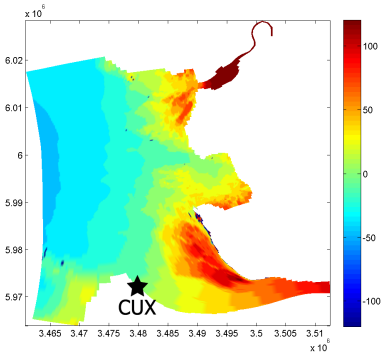
Maximum correlation at Büsum



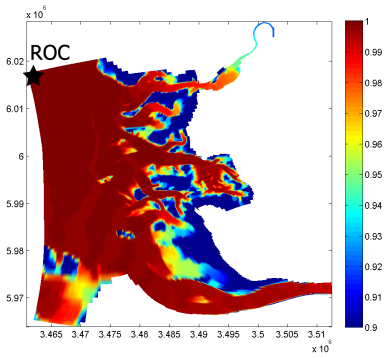
Time shift at Büsum



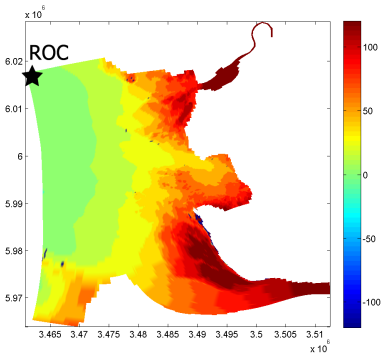
Maximum correlation at Cuxhaven



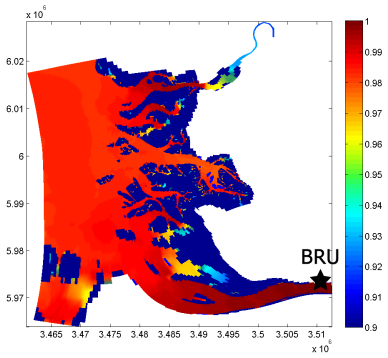
Time shift at Cuxhaven



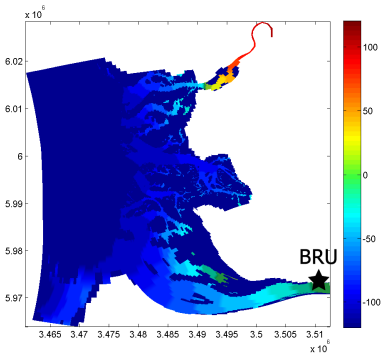
Maximum correlation at Rochelsteert



Time shift at Rochelsteert



Maximum correlation at Brunsbüttel



Time shift at Brunsbüttel

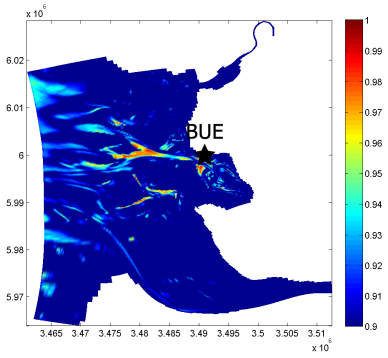
Figure 6.2: Maximum correlation coefficients and time delay in minutes for water level at Büsum, Cuxhaven, Rochelsteert and Brunsbüttel

6.2.1.2. Correlation maps for current model outputs

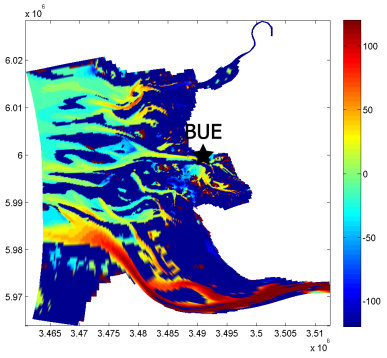
In the case of currents, the area where linear representativeness of any of the selected points is very low (see Figure 6.3). When compared with water level and waves, Cuxhaven shows a high correlation along the channel system of the Elbe estuary. This high correlation is also present when the Elbe discharge flows into the bight. Brunsbüttel has lower correlation values but shows a similar area than Cuxhaven. Büsum has a correlation higher than 0.97 with the inner part of the North and South Piep channels as well as the Piep channel and around 0.93 in some spots along the western boundary of the model. Rochelsteert has its maximum correlation in the North and South Piep Channels and in the deeper areas but not distributed uniformly. At Cuxhaven a high correlation is observed with the zone where Eider river discharge crosses the tidal flat region. But both currents are not directly connected. This high correlation is obtained by the combination of the influence of the discharge of a river plus that as stated by Toro et al. [2005], the crest of the tide arrives about the same time to those locations.

6.2.1.3. Correlation maps for significant waves height model outputs

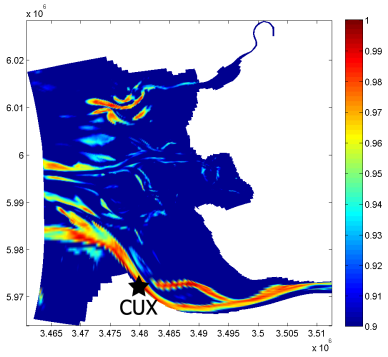
In the significant wave height correlation maps (see Fig. 6.4), Rochelsteert as in the case of water level shows a good linear relation with a large percentage of the deepest areas. But for tidal flats it decreases rapidly -due to the energy dissipation- to values lower than 0.90. Büsum is able to represent big areas on the tidal flat even if they are not well connected by the channels, since the energy coming from the sea is dissipated in a similar way in that region. A big difference is presented in Cuxhaven: for the water level, one station located there can be used to reconstruct most of the domain even if a big time shift is required. But for waves the covered



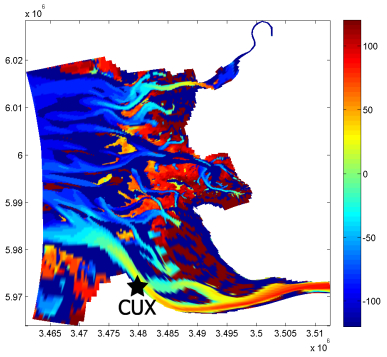
Maximum correlation at Büsum



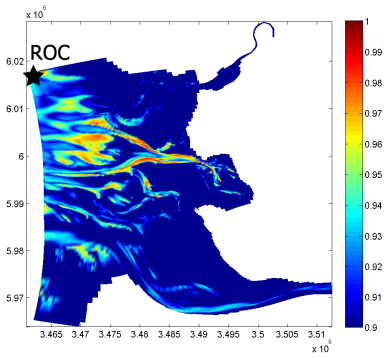
Time shift at Büsum



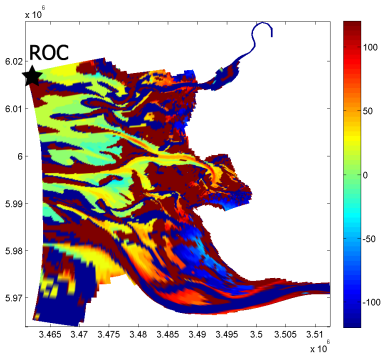
Maximum correlation at Cuxhaven



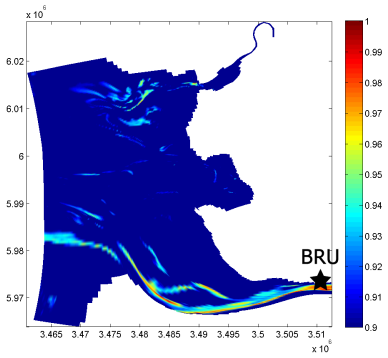
Time shift at Cuxhaven



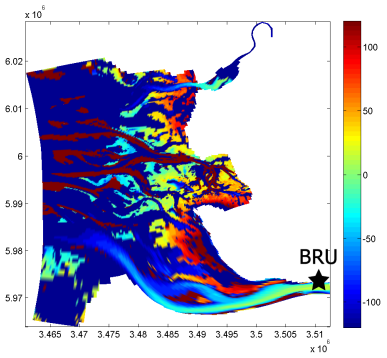
Maximum correlation at Rochelsteert



Time shift at Rochelsteert



Maximum correlation at Brunsbüttel



Time shift at Brunsbüttel

Figure 6.3: Maximum correlation coefficients and time delay in minutes for velocity at Büsum, Cuxhaven, Rochelsteert and Brunsbüttel

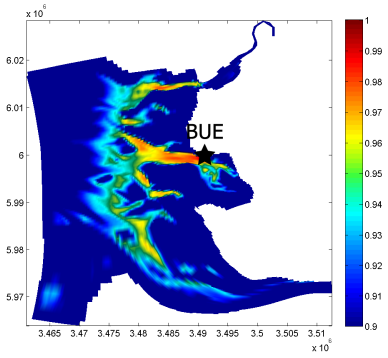
area by one device located here is limited to a small length of the Elbe estuary. For Brunsbüttel, as it has been described in the water level, the representativeness of the station is of short range. The spatial linear correlation of wave is more limited for the water levels as it reacts more to changes in bathymetry. The time shift maps allow to identify the differences in the physical characteristics of the zones. In the time shift maps, it is clear contrast between deeper areas, the braking zone, the tidal flat and the channel of the Elbe estuary.

Correlation maps can be used as an effective tool to understand the spatial distribution of linear relation of one cell with others in the model domain. Linear relations are evidenced by the appearance of high correlations in some areas. Although those high correlations are obtained by the use of a big time shift, employing nonlinear techniques reduce the required shift in the estimation of parameters.

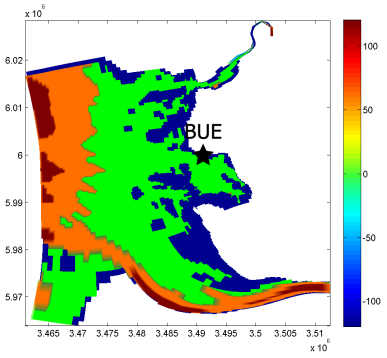
6.2.2. Treatment of corrections along the boundary

For the development of the boundary correction the characteristics of the four open boundaries are given in detail. Also depending on their particular features different approaches are required for the correction function. The cross sections for the four boundary conditions at the Dithmarschen Bight model are presented in Figure 6.5. The Figure 6.5.e shows the boundary conditions and the locations where the correction functions are calculated.

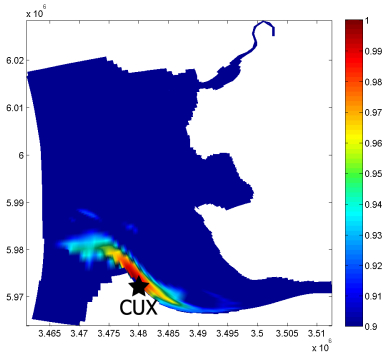
The North boundary extends West-East keeping an approximate constant latitude of 54°18'26"N. It goes from approximately $15km$ off the coast line with a depth of $13m$ where the pier at Rochelsteert is located. Then it continues with an average depth of 5m and finally becomes dry close to the coast line. When comparing the



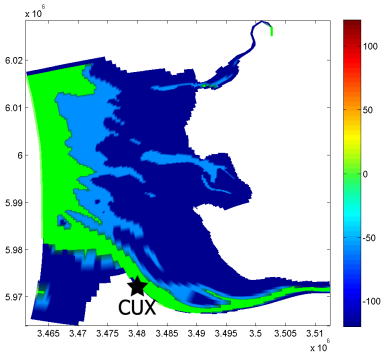
Maximum correlation at Büsum



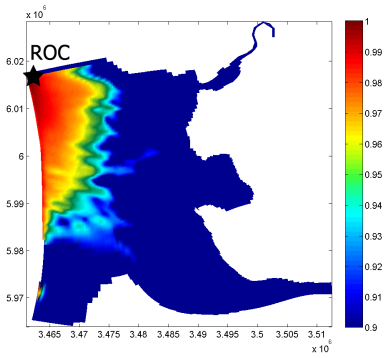
Time shift at Büsum



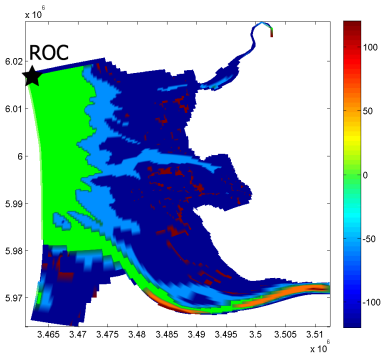
Maximum correlation at Cuxhaven



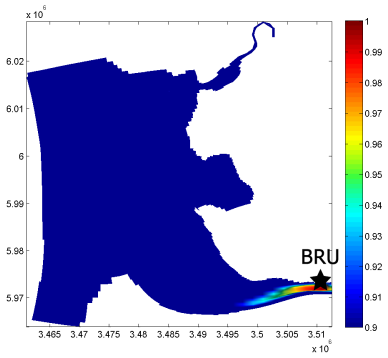
Time shift at Cuxhaven



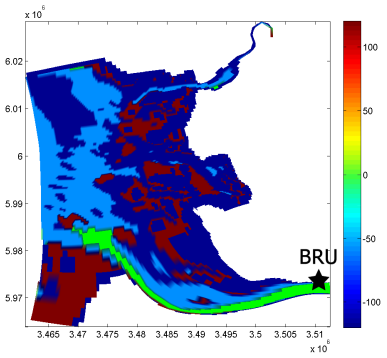
Maximum correlation at Rochelsteert



Time shift at Rochelsteert



Maximum correlation at Brunsbüttel



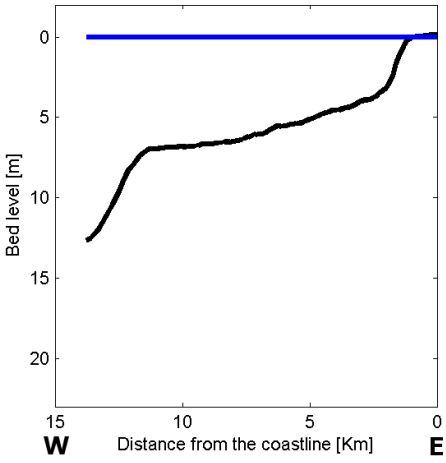
Time shift at Brunsbüttel

Figure 6.4: Maximum correlation coefficients and time delay in minutes for Hs at Büsum, Cuxhaven, Rochelsteert and Brunsbüttel

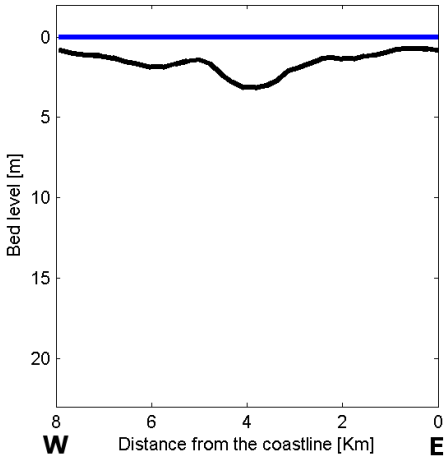
imposed water level from the German Bight model at the West and East extremes of this boundary line, they are similar, with an increase in the amplitude of $6cm$ in the point closest to the coast and a phase of around 12 minutes between them. The southern boundary extends along $8km$ also keeping a latitude of 53°49'05" the north coordinate constant, but in contrast to the northern border, shallow conditions are dominating this area, all depths are lower than 3 meters, and most of the points fall dry during the ebb tide. At the extreme of this boundary the depth is only in the order of $0.90m$. The cross section at Elbe boundary is located close to Brunsbüttel. A depth of $15m$ reduces the problems to find a correction function at the ebb tide. The water level is assumed uniformly along this boundary. Only one correction of the water level is calculated in the middle of the channel and is applied to all the points in this boundary. The western boundary goes from North to South with an approximate longitude of 8°22'26". It covers three zones that can be described separately. In the south, the tidal flat and some islands are located, depths range from 0.90 to $10m$ and an island protrudes a couple of meters over the mean water level. Next is the Elbe channel with a depth of $22m$, and in the North a submerged plain area around 11 to $13m$ depth. In this boundary currents are imposed, but as no real-time flow data is available, no modifications are made in the operational model.

6.2.3. Dynamic analysis of water level series

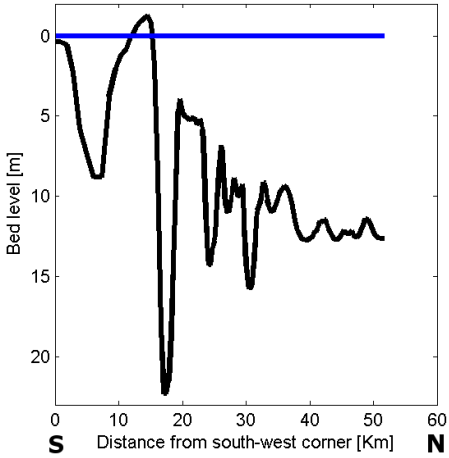
As has been described in the time series analysis (section 3.3), the conversion of a time series into phase points being represented by a phase space is possible. With the elements of dynamical analysis described in the section 3.3.2, the characteristic of the relation between water levels in two different stations is investigated. Taking a subset of the phase points describing the water levels at two different locations and plotting them in the phase space, it is possible to see how it describes a



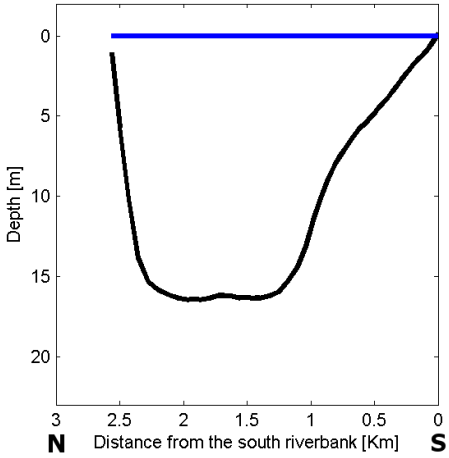
a. North



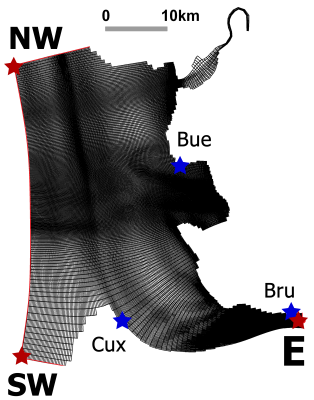
b. South



c. West



d. East -Elbe upstream-



e. location where correction functions are generated and of boundaries

Figure 6.5: Bathymetry profiles at boundary and their location

trajectory that resembles the projection in 2D of a wire wounded around a cylinder that is acting as an attractor (see Fig. 6.6). This orbital path is mainly due to the astronomical oscillation of the water levels, and the shape is given for the phase lag existing between the two locations. The position along the axis of the virtual cylinder is influenced by storm and barometric surges or seiches among other phenomena. Then it can be considered that the trajectory is a Liapunov stable (see section 3.3.2 for the definition) around the axis of a cylinder and that trajectories will be close all the time. If a time shift is applied to select the state variables, it is possible to fold the iterative map and get new shapes that can resemble a very narrow elliptical orbit, an "8" shape or triangular forms. These shapes can be observed when the iterative map is plotted for the South-West boundary with the Brunsbüttel station, North-West with Büsum and South-West with-Cuxhaven by a time shift that provides the maximum correlation (see Figure 6.7). After folding the iterative map, for the East condition with Brunsbüttel and the North boundary and Büsum, a linear relation is a reasonable solution. Even that in the case of North-West boundary the presence of two lobes can increase the RMSE of the approximation. For the South-West boundary from Cuxhaven, it is not possible to construct a single linear function. Particularly for this case a set of two linear equations -one for the flow and one for the ebb tide- can be used. But they would not explain the nonlinear shape at the extreme values.

6.2.4. Linear translation of boundary conditions

In the simplest case, for a linear relationship it is possible to write a linear function that connects the water levels at the boundary at a time t (WLB_t) with the water level at the gauge with a time shift ($WLG_{t-\mu}$) in the form:

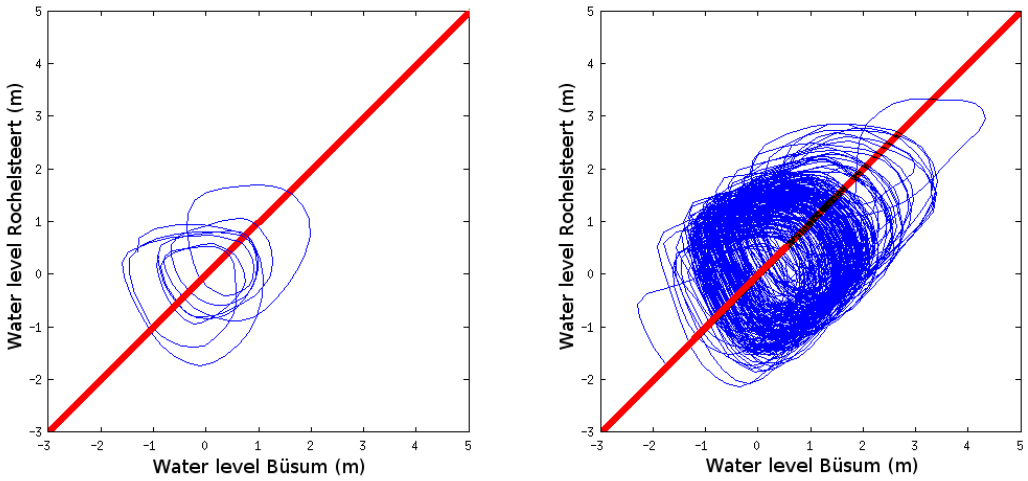


Figure 6.6: Space phase for two tidal gauge stations, left few phase points allow to see a more clear orbit, in the right side more phase points permits to observe how the orbit moves along a virtual axis -around 45° in the plot-

$$WLB_t = f\left(WLG_t\right) = a \cdot WLG_{t-\mu} + b \tag{6.4}$$

With a and b constant values obtained from a regression analysis and μ the time shift between the high water level at the gauge and the boundary.

Substituting equation 6.4 in 6.3, then:

$$WLB_t^* = WLB_t - \left(a \cdot WLG_{t-\mu} + b - a \cdot WLG_{t-\mu}^* - b\right) \tag{6.5}$$

Rearranging terms:

$$WLB_t^* = WLB_t - a \cdot \left(WLG_{t-\mu} - WLG_{t-\mu}^*\right) \tag{6.6}$$

Then the correction is proportional to the difference between the original nested model and the local value given by measurements or the local model at the gauge location. By the use of this function, the corrections are applied to the boundary conditions. Because of this nonlinearity the more complex autoregressive and artificial neural network strategies are tested.

As an extension of the linear translation and using a similar formulation that the presented for the autoregressive models, it can be constructed a regressive expression. This regressive formulation adds the contributions of more than one observation in the history of the input series at the gauge station linearly, and produces an output at the boundary location.

$$WLB_t = f\left(WLG_{t,t-1,t-2,t-3,\ldots}\right) = \sum_{i=0}^{i=n} a_i \cdot WLG_{t-\mu_i} + b \tag{6.7}$$

A fixed time step is selected and the number of samples is defined by selecting a number of parameters that minimizes the RMSE and maximizes the correlation.

6.2.5. Non linear approach

An artificial neural network ANN has the capability to take some arbitrary values in its input layer and built non linear relations among them, providing an output for a desired parameter. The ANN does not take into account the physics or dimensional correctness among variables. This means that it is the responsibility of the designer of the ANN to choose the proper inputs that carry a physical meaning to obtain a reasonable result. According to the different applications of

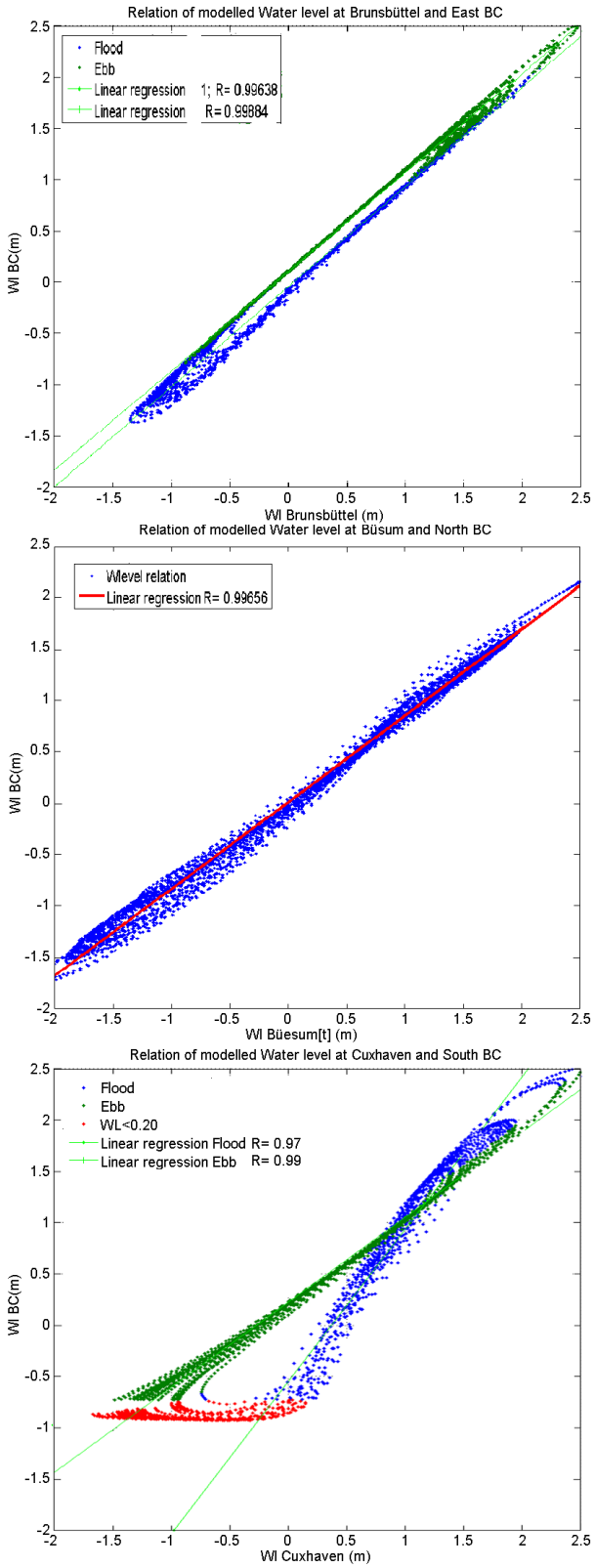


Figure 6.7: Water level correction at Brunsbüttel, Büsum and Cuxhaven

the ANN, in the present work, special assumptions are made for the inputs in each case.

6.2.5.1. Required parameters for building a neural network.

To build an ANN for the purposes of the present work, some elements of the network must be selected with special care. They have an impact on the quality of the neural model, the ability to provide a response that is valid in general cases and the easiness to obtain the model coefficients.

- The period length for the input data (the time elapsed from the oldest sample to the newest sample): This period should contain a history of measurements that can provide information about the development of the output that is estimated. The length of the period is different depending on the physics that underlies to the purpose of the neural network. A first approach to select this period can be done from a periodogram of the series. Then some iterations can be made in order to tune the final period to be used.
- The time step (Δt) when uniform sampling feeds the neural network: It should be maintained as coarse as possible to keep the number of input nodes low. It is a multiple of the time step at where the original information is provided and a resampling factor. The sampling space does not need to be uniform, but it facilitates the automation of the analysis of several models.
- The number of input nodes should be as low as possible: A big number of input nodes has effects on the difficulty of the training process and also on memorization. As the equations involved in an ANN are non linear, the increase in the number of coefficients to be estimated during the training of the network impacts in the ability to find optimal solution and the required

time to find such solution. As the number of elements in the network is increased, the memorization becomes also critical. As more parameters are available it is possible that the coefficients obtained are storing in "memory" the information provided in the training instead of providing a generalization of the problem.

- The number of samples utilized for the training: It should be the lowest possible to speed the training process, but high enough to be able to get a low error according to the Baum-Haussler rule (see section 3.4.4). The selected samples should have the highest representativeness of different scenarios. In section 6.2.5.4 it is proposed a method to reduce the number of samples, keeping the variety of input scenarios.

- The number of hidden layers: The hidden layer provides nonlinear response to the inputs. The use of two or more hidden layers allows to include a more complex nonlinear response, but at the same time restricts the optimization algorithms that can be used (i.e. finding an explicit formulation for a back-propagation error will be much more complex). The compromise in build a network with two or more layers and one with only one hidden layer with more neurons must be evaluated.

- The number of neurons per layer: It should be the lowest possible to avoid memorization. As has been pointed before, as more coefficients are involved in the design of the network, more problems related with the memorization of the scenarios are caused. Considering a set of neuronal networks with the same number of inputs in the first layer and with an error during the training of the same order, those with the less number of hidden neurons are more flexible and reduce the memorization. In the opposite case, as the number of hidden neurons is increased, it is more probable that the model cannot be applied in scenarios beyond the scope of the scenarios presented during the training.

- The number of outputs depends on the purpose of the neural network. For

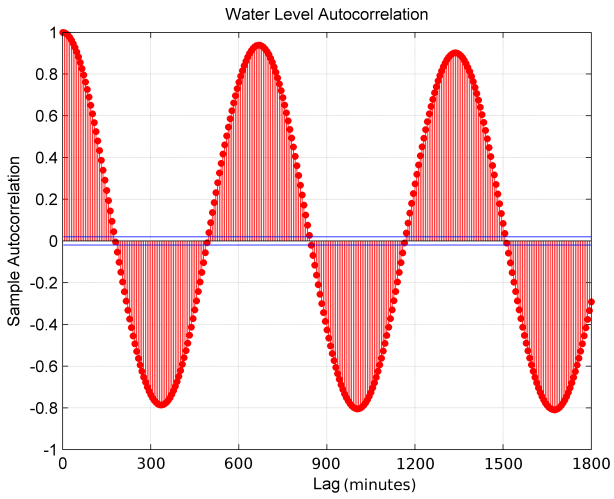


Figure 6.8: Autocorrelation Water Level at Büsum

the present work, only a single output element is applied.

6.2.5.2. Establishing the period length for the observation samples

The period of previous measurements employed in the neural network model should cover previous states of the system that can influence the state at the time desired. One mathematical tool that helps in the identification of this period is the use of correlograms. In the case of the water level, the astronomical harmonic oscillation creates correlograms that also show oscillations, obtaining the maximum peak every 12.25h that is the period of the semi-diurnal M2 (see Figure 6.8) or 24.50h that corresponds to M1 where diurnal tides are dominating. The wind action represented by surges is affecting the trajectory in the space phase as well. In the case of waves the maximum condition in a full arisen sea can be estimated by the use of a nomogram [Kamphuis, 2000]. According to this nomogram and depending on the wind speed and the fetch the maximum condition could be reached in a range of time between eight to thirty hours.

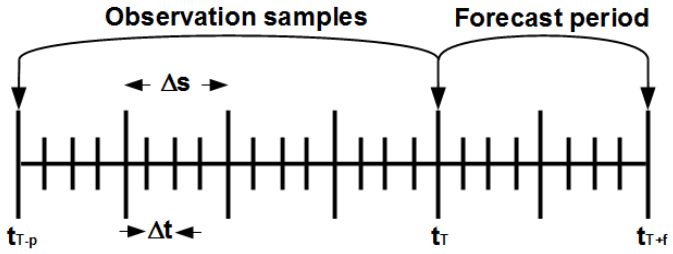


Figure 6.9: Sampling strategy for an ANN

6.2.5.3. Treatment of datasets for an artificial neural network

The available time series of wind and water levels measurements have a resolution of one minute, but along the experience gained with the neural networks, employing that fine resolution makes it harder to train the network as the input vector grows but with not providing extra information about the tidal cycle or the presence of a surge. To reduce the input layer size, a resampling has been made. Figure 6.9 shows the parameters used to create the samples for the training and later the inputs for the evaluation of the ANN. The original time step (Δt) is provided by the available time series, the resampling step (Δs) is a parameter defined when the ANN is designed and determines the resolution of the samples taken for the ANN. The time between observations that are included in the observation set is given by the product of the original time step and the resample step ($\Delta t \cdot \Delta s$). The period length for the observation is given by $P = \Delta t \cdot \Delta s\,(N_{obs} - 1)$. It is assumed that always one observation at the present time (T) is given. Only one output is delivered for the time $t = T + f$, with f the forecast period and it is provided by $f = \Delta t \cdot \Delta s\,(N_f)$ (for hindcast $f = 0$). Each element of one observation sample is assigned to one input node. If more than one kind of physical parameter is taken at these times, then they are included in to the next set of input nodes.

To facilitate the work of the optimization algorithms during the training process, the input values should be rescaled to positive values if possible. When more

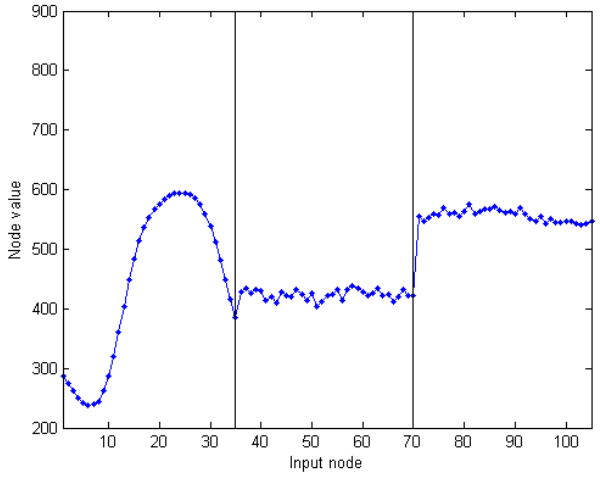


Figure 6.10: Example of the values contained in a single input set for a neural network

than one kind of physical parameters are included, the values at the input nodes should be in the same order of magnitude. This means that different scaling factors must be applied to each group of inputs. In Figure 6.10, water level and wind components u and v are employed as input nodes. Values are sampled at 35 different time steps, requiring a total of 105 input nodes. Two vertical lines are included to facilitate the visualization of the three groups of measured values. For the water level, the information is included in centimeters with the mean water level located at $500cm$ as it is originally provided by the authorities. The wind is multiplied by ten and also 500 is added to keep the same mean and a similar range of values as the water level records.

6.2.5.4. Reduction of the sample pool

To reduce redundancy and use only samples that provide additional information during the training process, a filter has been implemented. This filtering allows not only to avoid superfluous information, but also to save an important computational time in the iterative training. For the proposed method the inputs are considered

as vectors and by the use of the concept of the angle (α) between two real vectors u_i and $u_j \in \Re^n$ (with n the number of elements of the input layer). To determine the angle formed by the elements of the input set, the origin of the vector space is considered at the mean value of all the elements.

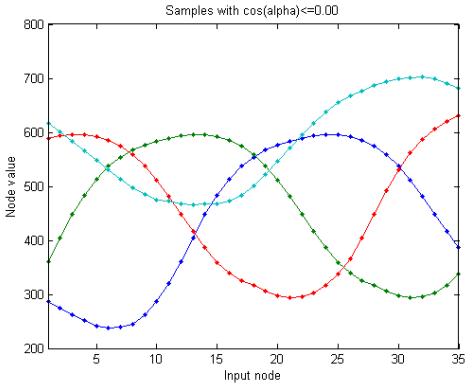
$$\cos(\alpha) = \frac{(u_i - \bar{u})(u_j - \bar{u})}{\|u_i - \bar{u}\| \, \|u_j - \bar{u}\|} \tag{6.8}$$

Whenever the angle α formed by two samples is lower than some selected reference value, a new sample is rejected from the pool employed for the training process. The reduced set of samples provides an efficient way to speed up the procurement of an ANN with no loss of significance in the results and helping to keep the generalization of the obtained model. For example if an initial sample set of $20,000$ scenarios is available, depending on the cosine of the angle that is tolerated, the number of samples can be reduced (see Table 6.1). If the cosine of one candidate sample and one that already belongs to the training pool is forced to be lower than or equal to one before its inclusion, then any sample can be accepted and the final set of data is not reduced. If this limit is lowered, increasing the angle between the new candidate and any previous member, then the number of samples to be used is going to be decreased. If the $cos(\alpha)$ is required to be -1, the number of samples can be limited to a maximum value of two, but in the practical cases it is difficult to find two samples that are exactly at opposite directions.

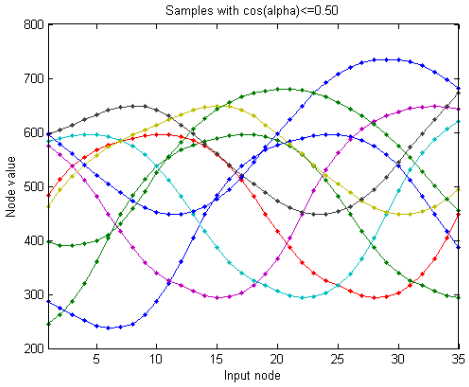
An example of the reduced number of input sets that can be obtained by applying this method are depicted in Figure 6.11. Only cases with few members are presented to keep the plot legible. Figures a and b are obtained from an input layer consisting only of water levels. It can be observed that the selected samples are covering scenarios with diverse characteristics. For $cos(\alpha) \leq 0.00$ three scenarios

Table 6.1: Reduction of sample pool for ANN training

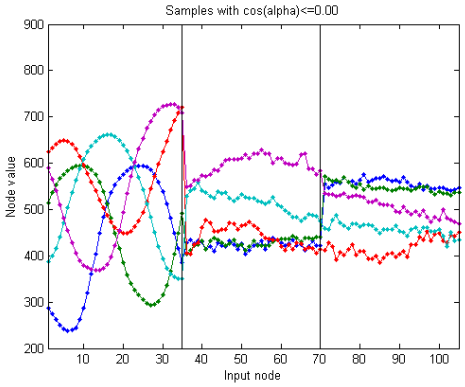
	Number of samples after filter	
$cos\,(\alpha) \leq$	Water levels	Water level, wind u and v components
	$u \in \Re^{35}$	$u \in \Re^{105}$
1.000	20,000	20,000
0.999	15,009	20,000
0.990	1,176	19,638
0.980	413	12,533
0.950	134	5,162
0.900	57	1,639
0.500	9	18
0.000	4	5
-0.500	2	2
-1.000	1	1



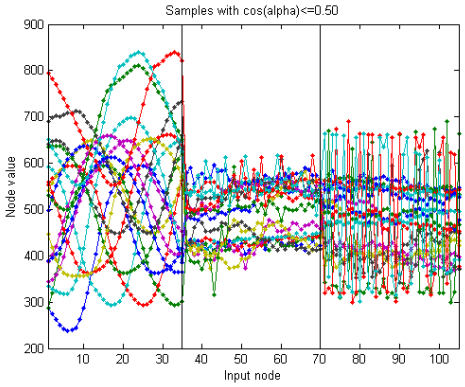
a) Water level inputs with $cos\,(\alpha) \leq 0.00$



b) Water level inputs with $cos\,(\alpha) \leq 0.50$



c) Water level and wind inputs with $cos\,(\alpha) \leq 0.00$



d) Water level and wind inputs with $cos\,(\alpha) \leq 0.50$

Figure 6.11: Filtered samples using different values of $cos\,(\alpha)$

are covering different phases in the tidal cycle, while the fourth covers a situation when the ebb tide stays in a condition of high water level due to a wind-driven surge condition. When the limit is increased to 0.50, new members are included, covering more phases and water level paths. Figure 6.11 c and d correspond to the same input layer configuration described in the last section and using the same limits imposed for a and b. Here the changes in the wind conditions are visible.

6.2.5.5. Activation function

During the development of the tools required for training the ANN, it was observed that in some cases the activation function gives a value of 0 or 1. This happens because at early stages of the training when big gradients of error are present, the weight coefficients are strongly increased by the gradient descendant method. The problem is that after such an extreme value is obtained in the activation function, the efficiency of any of the optimization algorithms described in subsection 3.4.3 is seriously affected. This is because the activation function is not sensitive to small changes in the coefficients at the input of that neuron anymore. Then the output of the activation function remains in the same stage as in the previous iteration, avoiding the evolution of the solution.

To keep the flexibility in the learning process a correction is made to the values of w_{kj} and t_k associated with the activation function (h_k) in order to keep it always in the range $[0.00001, 0.99999]$. With these values the neuron could be activated or deactivated almost completely if it is really required by the training, but keeping the ability to respond to optimization algorithms during the training of the network.

6.2.5.6. Correction size for simulating annealing

For the simulating annealing described in section 3.4.3.3, the function that defines the correction size is generally described in the literature [Eglese, 1990, Kirkpatrick et al., 1983] as a monotonic function that decreases the value as long as the iterations are advancing. The correction term is reduced until some point where its effect vanishes and the training stops. During the development of the trainer it was

difficult to find a generic constant value h_0 which shows a high speed convergence. In some cases a small correction value h allowed the neural network connections to adjust at early stages of the training to find an appropriate region in the solution space, and then a higher h allowed to move faster to an optimal area until a new small h was required for a final fine tuning. For the cooling function instead of a monotonic decreasing, it has been replaced by a sinusoidal function of the form:

$$h(i) = h_{min} + a\,|\sin(\omega \cdot i)| \tag{6.9}$$

Where: h_{min} is the lowest correction allowed, a is the amplitude that modulates the correction function, w is the frequency at which the function varies the values and i is the iteration number in the training.

This formulation permits that along the training, the network parameters are warmed and cooled by the use of a cyclical big and short steps, this helps the cooling function to adjust to the optimal step required in the optimization to speed up the solution.

6.3. Building of local models for water level

After the description of the strategies to move the information around the domain of the model, now the building of local models for water level is presented.

Local models are built as an alternative to the process based model. They are able to create time series for one single position. Basically local models are built

to reconstruct information of one measurement station that becomes off-line, for the improvements of boundary conditions by the use of a virtual station and for forecasting. Functions including linear regression with time shift, regressive or artificial neural networks are used. In the next sections, models that connects available measurement locations against selected locations at the boundary are evaluated. First an example using water levels measurements at Büsum and at Rochelsteert is presented to show the agreement that is possible to be obtained with field information. For the subsequent local models, information from model outputs is taken.

After demonstrating that one location in terms of the other for real conditions by the use of ANN can be described. then, the three methods are employed to study several combinations of locations and parameters to build hindcast water level models. Values obtained from the numerical model outputs are used to show that it is also possible to get similar results for such station to the ones obtained for the same location only based on measurements. Then the method will be extrapolated to obtain a forecast for a time frame of up to 24 hours. To calculate the parameters of the local models are used values from the first semester of 2008. For the validation of the improvements in the model result, June and July of 2008, the same period for the validation of the process-based-model is in use.

6.3.1. Water level reconstruction Rochelsteert from Büsum measurements

For the reconstruction of the water levels at Rochelsteert pier, water level measurements from Büsum and Rochelsteert are taken into account. The idea is to find a representation of the values in one station based on field measurements in

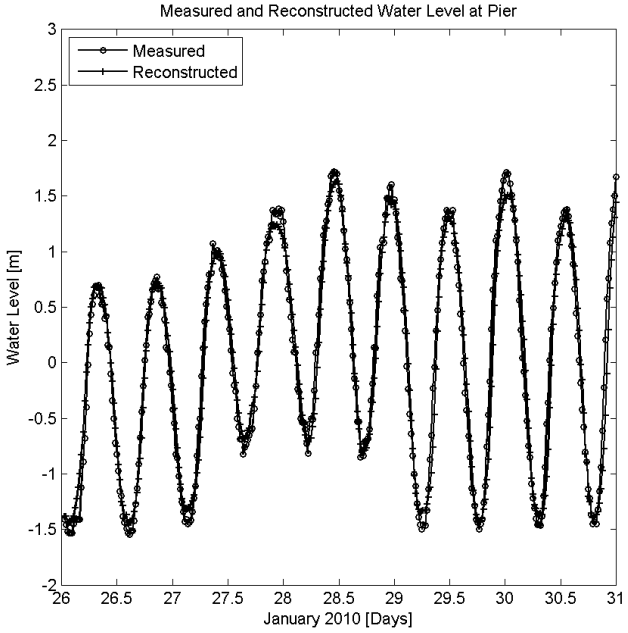


Figure 6.12: Neuronal network with 30 inputs and 12 hidden neurons W values

another place. The main utility of this approach is the possibility to reconstruct one time series when the other is available, being able to fill the gaps. A neural network using water level measurements as the input layer, and $\Delta t = 60s$, $\Delta s = 10$, $N_{obs} = 30$, $N_f = 0$ has been trained with the output layer containing the water level at Rochelsteert. A very good agreement has been obtained as Figure 6.12 shows. Where the measurements at Rochelsteert pier and the reconstruction based on measurements at Büsum are compared. An RMSE of $14.26cm$ and a correlation coefficient of 0.99 has been reached.

6.3.2. Construction of boundary conditions from translation of an alternative station

For the translation of information to the boundary of the Dithmarschen Bight model DBM, instead of real water level measurements, time series at virtual sta-

tions from the numerical model results are used. One of the virtual stations is located in the open sea boundary while the other is situated in the same place of one of the available measurement stations (see Figure 6.5).e. Here the idea is to build a function of the boundary condition using the station where later the measurements will be available. This function allows to estimate a better boundary condition to operate the DBM. In this subsection the efforts are to develop a hindcast construction of the information at the boundaries. Later the ideas presented here are extended to use a similar strategy for the forecast of the localized models.

6.3.2.1. South East boundary condition from Brunsbüttel

As it has been described in Chapter 4, the South East boundary is close to Brunsbüttel. Then water levels are measured close to the boundary condition. Also the channel has enough depth to allow an easy movement of the water with low friction. Because such conditions a linear relation is observed for the water levels without the necessity of any lag ($\mu = 0$) (see Figure 6.2). A correlation of 0.99 is reached (see Figure 6.7.a). The expression for the estimated water level at the South East boundary in terms of Brunsbüttel station is given by $Y_{BC_{SE}} = -0.0411 + 0.9754 \cdot X_{BRU}$. Regressive models or neural network models were not even considered after this good agreement.

6.3.2.2. South East boundary condition from Büsum

As it has been just shown in section 6.3.2.1, to create corrections for this boundary from Brunsbüttel a linear approach is sufficient. To show the potential of the

Table 6.2: Regressive Model for evaluation of South East boundary condition from Büsum

Order of the R model (n)	Period Covered (min)	RMSE (cm)	Correlation
27	162	43.13	0.93
16	96	45.45	0.92
12	72	48.85	0.91
3	18	55.65	0.83

Neural Network to deal with complex patterns in water level estimation the Büsum location is employed to estimate the South East boundary condition. From a linear analysis the maximum correlation (0.94) is obtained for a time shift of 162 minutes. The linear regression corresponding to this shift is $Y_{BC_{SE}} = -88.8955 + 1.0971 * X_{BUE}$ with an RMSE of $117.84cm$. This first linear approach shows that **a high correlation not necessarily provides a low error**, due to the differences in the shape of the tide in both locations (see Figure 6.7).

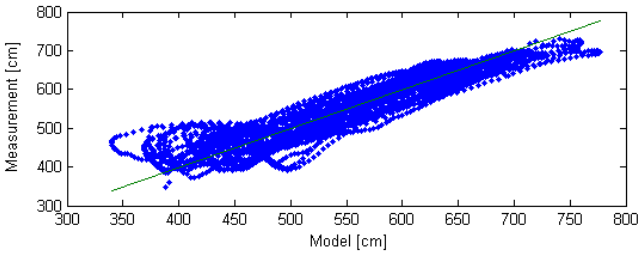
If a regressive model of the form $Y_{BC_{SE_t}} = \sum_{i=0}^{n} a_i * X_{BUE_{t-i}}$ is calculated (see Table 6.2), it is possible to decrease the RMSE error with a similar linear correlation. If the same period of information is utilized, the RMSE is reduced from $117,84cm$ to $43,13cm$. However taking a regressive model that uses information from the last 96 minutes the RMSE and the correlation is still similar to the regressive that uses 162 minutes. With this shorter series, the possible gaps in the measurement are going to have a shorter impact on the construction of the boundary conditions. 72 minutes still gives a comparable value, but the regressive model of order 3 that covers 18 minutes, even that the RMSE is better than with a simple linear adjust, correlation is the worst of all the other cases.

Table 6.3 shows the results from some trained networks to construct the south east

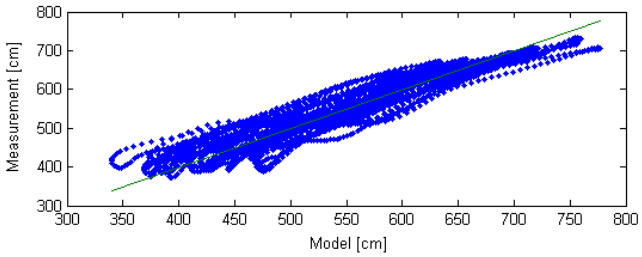
boundary condition from Büsum information. For each artificial neural network model, the columns in Table 6.3 describe the number of hidden neurons in each layer. In the present case the second hidden layer is one, meaning that there is only one hidden layer. The number of hidden neurons in layer one, the number of input nodes, the resampling step ds, the previous period covered by the samples in minutes, the RMSE and the correlation. For those models a Δt of six minutes is used. The inputs are selected in order to cover similar periods from the one utilized for regressive models showed in Table 6.2. It is evident that enhanced results that improve the correlation and the RMSE are obtained. In Table 6.3 and Figure 6.13 it is possible to see that shorter periods (models 2 and 3) have lower performance than the ones using a longer period (the remaining models). When comparing models 1, 4 and 5 that cover a similar period but with different *ds*, they are reaching a similar range of errors. Those three models are able to transform the phase space of the artificial neural network estimation versus the provided boundary conditions into a linear space (see Figure 6.13.d).

Table 6.3: Errors for neural network evaluation of South East BC from Büsum values

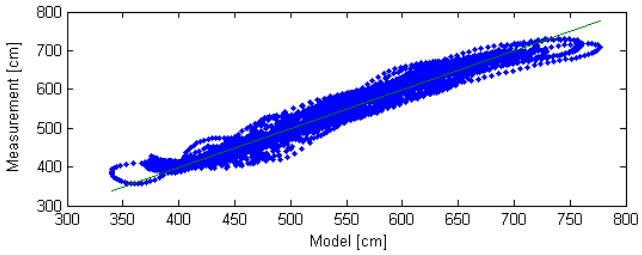
Model	N. hid 2nd layer	N. hid 1st layer	N. input nodes	dt (min)	ds	period covered	RMSE (cm)	correlation
1	1	12	27	6	1	156	22.27	0.97
2	1	12	3	6	1	12	36.12	0.93
3	1	12	12	6	1	66	30.28	0.95
4	1	12	2	6	26	156	19.54	0.98
5	1	12	10	6	3	162	20.08	0.98
6	1	12	5	6	3	72	20.69	0.97



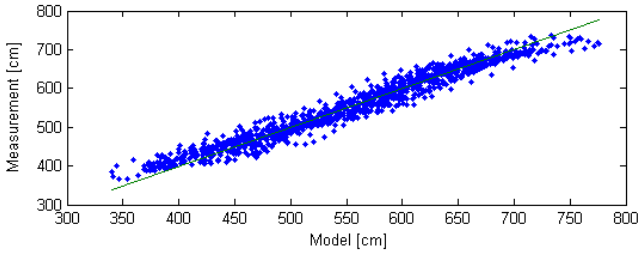
a. Model 2



b. Model 3



c. Model 1



d. Model 5

Figure 6.13: South East BC from Büsum (ordered by descending quality of the outputs).

Table 6.4: Autoregressive Model for evaluation of South West BC from Büsum

Order of the AR model (n)	Period Covered (min)	RMSE	Correlation
1	6	70.78	0.81
6	36	58.03	0.87
12	72	54.92	0.89

6.3.2.3. South West boundary condition from Büsum

The South West boundary condition imposes a challenge as it becomes dry during the tidal cycle (see Figure 6.7). It creates an additional problem that the friction in the bed is more relevant as the lowest levels are present, creating a nonlinear change in the water level. Following the same procedure of 6.3.2.2, the maximum linear correlation is 0.88 obtained for a time shift of 72 minutes, and an RMSE of 71.46cm. The regressive approach shows that it is not able to improve the correlation and even that there is a reduction in the RMSE, it is still high (see Table 6.4).

Neural networks (see Table 6.5), try to take into account the nonlinearity present in this series, when only one input is included in the artificial neural network, instead of a line, a triangle is obtained, giving a lower RMSE but reducing the correlation. As the number of inputs has increased, the shape of the triangle is more elongate. With 12 inputs this elongate triangle has a "L" shape (see Figure 6.14.a). The foot of the "L" represents the dry values at the open sea boundary. The use of a second layer with hidden neurons has been tested as well to see if it was possible to obtain a better result to deal with the values close to dry level. Several tests have been made, but no significant improvement has been obtained.

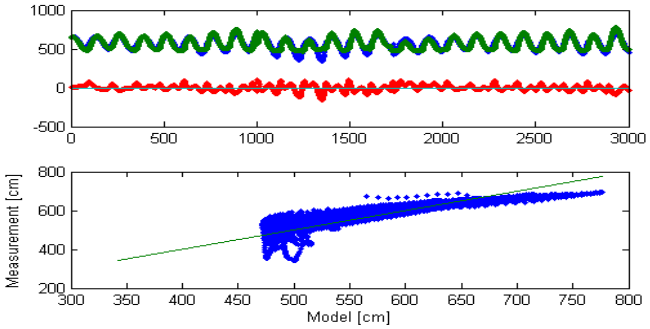
Table 6.5: Errors for neural network evaluation of South West BC from Büsum values

Model	N. hid 2nd layer	N. hid 1st layer	N. input nodes	ds	dt (min)	period covered	RMSE (cm)	correlation
1	1	12	1	1	6	0	48.35	0.78
2	1	12	6	1	6	30	48.02	0.84
3	1	12	12	1	6	66	37.42	0.89
4	12	12	12	1	6	66	36.48	0.90

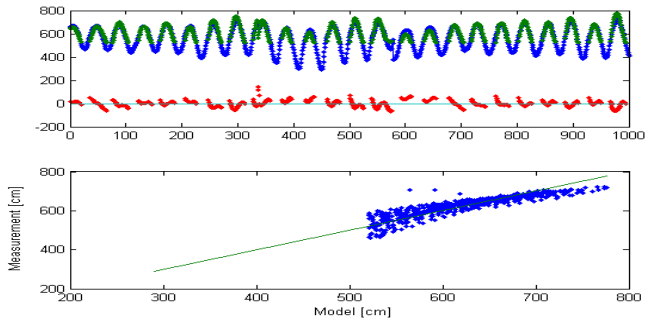
A new dataset where dry values at the open sea boundary are excluded has been applied to train the network (see Table 6.6). In this scenario the idea is to give more flexibility to the network and not force it to fold following a constant value (see Figure 6.14.b and c). The network is then allowed to create a series with values that are lower than the ground level. This ebb tide created by the artificial neural network has a realistic shape, but as they do not correspond to possible water levels, it is required a post processing to cut off those values again. The RMSE comes to a more reasonable value. Some tests to reduce the number of inputs but keeping the input period has been carried out. Taking four inputs every 18 minutes (model 5) instead of 12 every six minutes (model 1) gives an estimation of the same order but with the advantage of being more flexible as it requires to use a shorter period for inputs, and requiring less effort for training as less neurons are present.

6.3.2.4. South West boundary condition from Cuxhaven

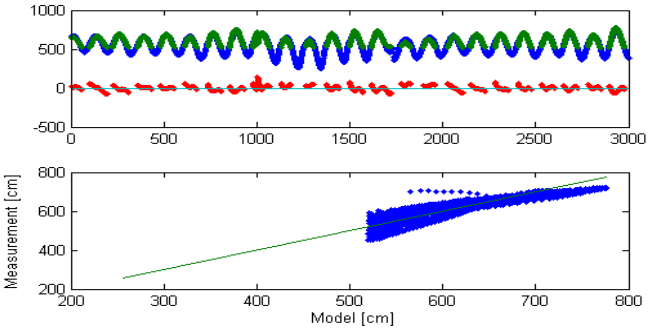
The same procedure from the previous section but using information from Cuxhaven is carried out. The highest correlation (0.89) is obtained also for a time shift of 72 minutes giving a RMSE of 58.96cm and the resulting equation is



a. Model 3 including dry conditions (Table 6.5)



b. Model 5 excluding dry conditions (Table 6.6)



c. Model 1 excluding dry conditions (Table 6.6)

Figure 6.14: South West BC from Büsum. In the time series _ represents the neural network output, _ the measurements and _ the error calculated as the difference between the measurement and the network output

Table 6.6: Errors for neural network evaluation of South West BC from Büsum values excluding samples where cell is dry

Model	N. hid 2nd layer	N. hid 1st layer	N. input nodes	dt (min)	ds	period covered	RMSE (cm)	correlation
1	1	12	12	6	1	66	27.04	0.90
2	12	12	12	6	1	66	30.97	0.87
3	12	12	12	6	1	66	32.30	0.85
4	1	12	3	6	4	48	27.96	0.89
5	1	12	4	6	3	54	27.47	0.90

Table 6.7: Autoregressive Model for evaluation of South West BC from Cuxhaven

Order of the AR model (n)	Period Covered (min)	RMSE (cm)	Correlation
1	6	58.54	0.84
6	36	48.10	0.89
12	72	45.28	0.91

$Y_{BC} = -90.1960 + 1.0970 * x_{t-72min}$. Here the travel time of the tidal wave from the boundary is similar to the one from Büsum and the correlation is more or less the same, but the RMSE drops $12cm$. This can be explained as Cuxhaven is connected with the boundary trough a wider and deeper channel formed by the continuation of the Elbe estuary having a lower effect from friction and transmitting a more linear effect. The regressive model is providing an improvement in the RMSE compared with the results from the linear estimation with a similar correlation (see Table 6.7). Neural networks are able to improve even more the correlation and the RMSE (see Table 6.8), but still dry values at the boundary make it difficult to reach a higher accuracy.

Finally, using only samples where cells at the open sea boundary are not dry, we

Table 6.8: Errors for neural network evaluation of South West BC from Cuxhaven values

Model	N. hid 2nd layer	N. hid 1st layer	N. input nodes	ds	RMSE	correlation
1	1	12	1	1	43.99	0.82
2	1	12	6	1	35.03	0.89
3	1	12	12	1	31.52	0.92

can obtain a lower RMSE (see Table 6.9). Attempts to employ more than 1 hidden neuron at the second layer, shows that is not helping to improve the results, even increasing the RMSE.

Table 6.9: Errors for neural network evaluation of South West BC from Cuxhaven values excluding samples where cell is dry

Model	N. hid 2nd layer	N. hid 1st layer	N. input nodes	ds	RMSE (cm)	correlation
1	1	12	12	1	25.42	0.91
2	12	12	12	1	26.23	0.91

6.3.2.5. North West BC from Büsum

Büsum has a good connection to the North West boundary trough a relatively deep channel that links both locations (see section 4.1). As friction is less representative, linearity can represent in a reasonable way the phenomena as can be observed with a correlation of 0.95 and an RMSE of $37.54cm$ when a linear regression is used with a time shift of 30 minutes. The equation obtained is $y_{bc} = -105.6193 + 1.1730 * xbue_{t-30min}$. The regressive model is calculated for shifts from 6 to 36 minutes and the best values for correlation and RMSE are 0.95 and $35.74cm$ that are not a significant improvement compared with the linear regression. When

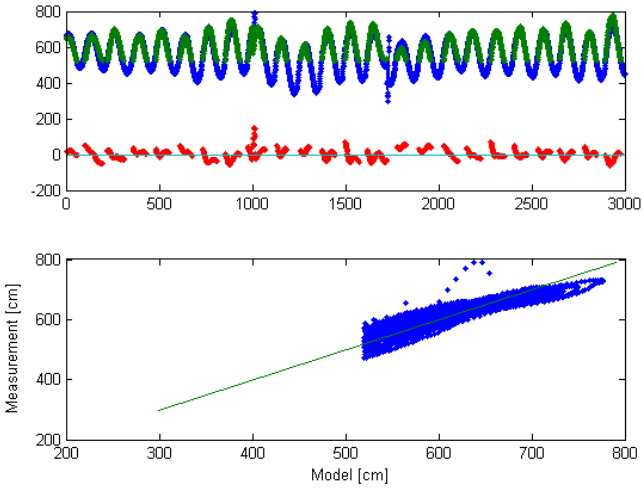


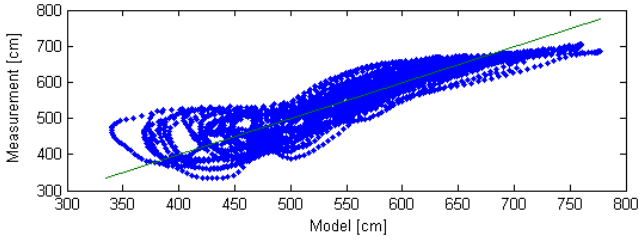
Figure 6.15: South West BC from Cuxhaven model 1 without dry state

neural networks are trained with the same input period of 30 minutes, the RMSE improves to 31.19cm. The value is better than linear or regressive approaches, but with a reduction of the correlation to 0.93. Additional tests using a period of 60 minutes make possible to reach a correlation of 0.97 and reduce the RMSE more than 38% to a value of 23.06cm (see Table 6.10 and Figure 6.16).

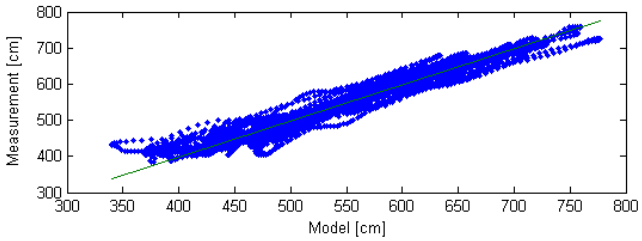
Table 6.10: Errors for neural network evaluation of North West BC from Büsum values

Model	N. hid 2nd layer	N. hid 1st layer	N. input nodes	ds	dt (min)	RMSE	correlation
1	1	12	10	1	6	23.06	0.97
2	1	12	5	1	6	31.19	0.94
3	1	12	3	1	6	34.48	0.93
4	1	12	3	3	6	38.57	0.91

This case is comparable with subsection 6.3.1, where measurements are directly used from the same locations. When outputs from the process based model are used for the training, the correlation is slightly reduced from 0.99 to 0.97, while the RMSE is increased in 8.8 cm. This shows that the procedure is able to deliver results that are in the range of the measurements. Most of the increase of the



a. Model 2



b. Model 1

Figure 6.16: North West BC from Büsum

RMSE is due to the limitations of the process based model to describe the path of the water level in the intertidal range as has been discussed in Chapter 5.

6.3.3. Forecast at the open sea boundary

After showing the good performance of the neural networks for the task of hindcast, now the procedure to obtain local models able to provide forecast values is going to be extended. First to have a reference the results obtained from the process based model at the Rochelsteert are described. Then the results from neural networks for forecast with various time windows are given. In order to improve the forecast capabilities, fuzzy logic and ensemble of model outputs are included.

6.3.3.1. Water level forecast at Rochelsteert using the nesting sequence with Delft3D model

To allow a comparison with the above methods to obtain the boundary conditions based on values taken at selected stations, the open sea boundary conditions obtained from the nesting sequence is included here (see Fig. 6.17.a). The RMSE is $48.59cm$. This value is 29% **higher** than the RMSE for the levels calculated at the North West boundary from Büsum (see section 6.3.2.5) by a linear interpolation of the measurements. Most of this difference is not coming from the estimation of the high tides, but from differences in the intermediate values. This shows that even the nesting sequence is able to provide a good estimation of the extreme values. A better representation can be made when measurements are available, for this purpose assimilation is a plausible step.

6.3.3.2. ANN for forecasting water levels at Rochelsteert

Artificial neural network models to provide a forecast at selected stations have also been developed. Table 6.11 summarizes the results for models that provide water levels at the Rochelsteert station by the use of measured water levels at Büsum and Cuxhaven and also wind information from the Eidersperrwerk. The forecasting models that exclude wind information had shown a poor capability to predict the future water level. This can be observed when comparing the forecast for 12 hours where only Büsum is used with a RMSE of $58.13cm$ and when Büsum plus wind, the RMSE is reduced significantly to $27.93cm$. This lower accuracy is obtained because in two scenarios with a similar short term history of water levels the reaction of the levels to different meteorological development evolves in a different way. Including wind makes the model consider a change in the present conditions

due to a storm. In general a historical set of data of 11 hours has been taken into account to include the fetch effect of the wind.

Table 6.11: Errors for neural network of North West BC forecast (0h forecast corresponds to a hindcast)

Forecast (h) period	Note	Büsum	Cuxhaven	Wind	RMSE (cm)	Correlation
0	Hindcast	x	x	x	17.89	0.99
0	Hindcast	x	-	x	17.30	0.99
0	Hindcast	-	x	x	18.39	0.98
0	Hindcast	x	-	-	17.81	0.99
2		x	x	x	20.71	0.98
6	Forecast	x	x	x	27.31	0.96
12		x	x	x	27.44	0.96
12		x	-	x	27.93	0.96
12		-	x	x	28.70	0.96
12		x	-	-	58.13	0.96
12	Fuzzy logic [1]	x	-	x	28.57	0.96
24		x	x	x	30.32	0.96

For the hindcast (or forecast of 0 hours) all of the four models presented have a similar performance. The best result is given by the use of the measurements at Büsum plus wind information with a RMSE of $17.30cm$, but the exclusion of wind increases the RMSE slightly to $17.81cm$. This shows that for hindcast, only water levels at Büsum are required to obtain a reasonable result. All the tested models for hindcast have a correlation close to one. For two hours forecast an increase

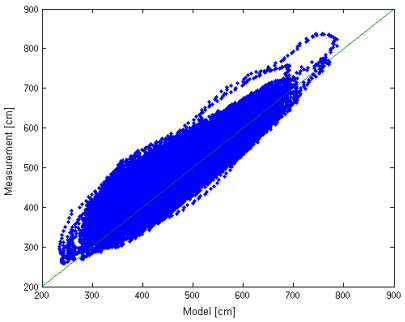
[1] Average result using a history period of 5 hours for Fuzzy logic and 11 hours for standard forecast.

x: the station is considered as part of the input.

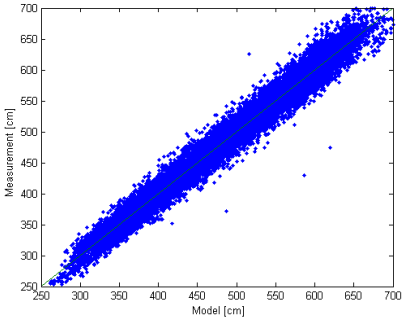
-: the station is excluded

of the RMSE and a reduction of the correlation is observed, but still in the same range of the hindcast. For six hours forecast the RMSE increase to 27.31cm and a correlation of 0.96 showing that the model starts to lose some accuracy compared with the hindcast and the two hours forecast. For 24 hours, forecast models combining the three inputs show that when wind is included, the RMSE is in the order of 27 to 28cm, which is still lower than the RMSE obtained from the nesting sequence (see section 6.3.3.1). When wind is excluded, the model obtained has a lower correlation and a higher RMSE than the results from the process based model. Then the model without wind is not considered in the prediction of the water levels. The 12 hours forecast with fuzzy logic is presented in the same table in order to be able to make comparisons with the simple ANN. The RMSE of 28.57cm is the global performance of a set of models. Details on the construction of the fuzzy logic model are provided in the next section. Here it is important only to note that the model built with fuzzy logic uses a shorter period of observations as an input than the simple models, but it is still able to provide a similar RMSE value. For 24 hours forecast, there is again a small increase in the RMSE compared with the one obtained for 12 hours.

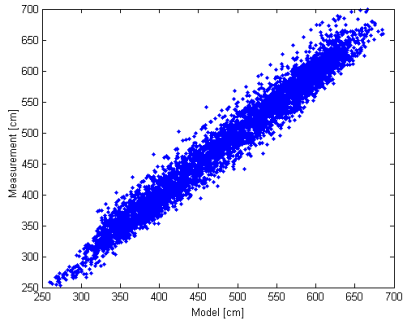
In Figure 6.17 a mosaic with selected models is presented to compare the behavior of water levels at Rochelsteert. In the horizontal are represented the water levels obtained from the model and the vertical axis contains the measured water levels. The German Bight model is able to provide a reasonable description of the extreme values but a belly is formed in the ebb zone due to an underestimation of the water level by the model. On the other hand the water levels from the ANN models are distributed uniformly around the 45° line. The dispersion is increasing as the forecast time is bigger as shown in Table 6.11.



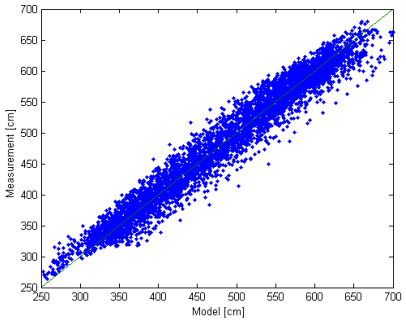
GBM results from Delft3D.



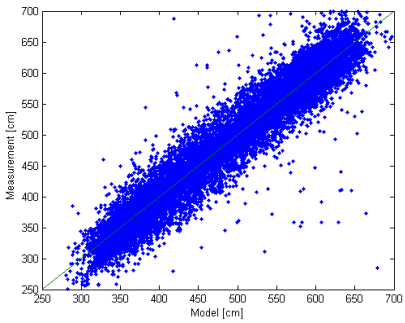
Hindcast with Cuxhaven, Büsum and wind as inputs.



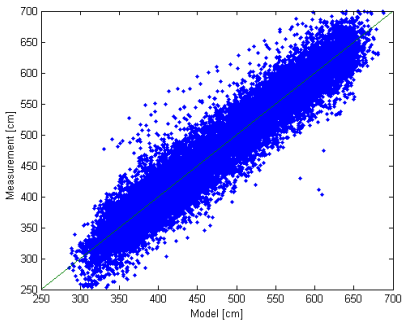
Forecast 02h with Cuxhaven, Büsum and wind as inputs.



Forecast 12h with Cuxhaven, Büsum and wind as inputs.



Forecast 12h after combining with fuzzy logic (five hours of input observations).



Forecast 24h with Cuxhaven, Büsum and wind as inputs

Figure 6.17: Comparison of water levels obtained from different model formulations with the measurements at Rochelsteert in the North-West boundary of the DBM model

6.3.3.3. ANN combined with fuzzy logic for forecasting water levels at Rochelsteert

Once the importance of including the wind information for the water level forecast is established, some attempts to improve the local model are made while reducing the historical period used in the inputs. A set of 12 ANN has been trained splitting the domain of application of each one based on the average of local wind magnitude and direction of the last 5 hours (n=6,ds=60,dt=1min) before the present time. The direction is discretized in four quadrants corresponding to NE, SE, SW and NW winds with an overlap of 10° in two contiguous domains. For the magnitude, three overlapping wind speeds groups were used. Winds from $0m/s$ to $5m/s$, $4m/s$ to $10m/s$ and higher than $8m/s$.

Neural networks can deal with circular functions with a single network as they have been used in the single ANN in the previous section or have been explained by Hsieh and Cannon [2008]. In the present section the splitting of the evaluation in several domains has the advantage of avoiding the circular functions, making the training process faster, also the possibility to differentiate between the conditions where the highest accuracy can be obtained. The Table 6.12 presents the best network obtained in each case. Independently of the wind speed it is observed a lower RMSE in the trained models for wind coming from the East and the South where land is located, while models with wind coming from the sea side have a higher RMSE. This shows that when the wind is coming from the sea there are more interaction between the sea and the air, creating more complex scenarios. When the wind blows from the coast, it has less interaction with the water to be able to produce local effects.

The problem of using sub-domains reside in the difficulty to guaranty the continuity of the global function. To reduce the gaps between two contiguous sub-

Table 6.12: Errors for neural network evaluation of North West BC forecast from Büsum values

Wind Speed	Wind Direction	RMSE	Correlation
low wind	SW	30.51	0.96
low wind	SE	22.97	0.97
low wind	NE	22.04	0.97
low wind	NW	34.04	0.95
mid wind	SW	28.97	0.96
mid wind	SE	21.36	0.98
mid wind	NE	26.64	0.97
mid wind	NW	32.77	0.95
high wind	SW	30.21	0.96
high wind	SE	28.19	0.96
high wind	NE	27.27	0.96
high wind	NW	32.59	0.95
All models combined		28.57	0.96

domains, the use of fuzzy logic permits to obtain a softer transition in the outputs. In the overlapped domain all the applicable models are evaluated independently and the output is given by the weighted average of all of them. The combined result after using the fuzzy logic is also presented at the end of Table 6.12. The RMSE of $28,57cm$ as has been mentioned in the previous section could be reached employing less inputs than the simple ANN developed in that section.

6.4. Ensemble of forecast at Rochelsteert

When several models are employed to get the forecast of one location, they can differ in the output even if they have a similar value for the RMSE. This difference may be is a small quantity but it can diverge more in some cases. To take advantage of the fast evaluation of an ANN, an ensemble of several local models is calculated to obtain an improved forecast to be imposed in flow model. Such models are combined by a weighted mean of the outputs of each model. The weight is provided by the inverse of the RMSE of the respective model. In this way models with lower RMSE have a higher influence in the final forecast.

$$y(t) = \frac{\sum_{i=1}^{n_{em}} \frac{1}{RMSE_i} \cdot y_i(t)}{\sum_{i=1}^{n_{em}} \frac{1}{RMSE_i}} \tag{6.10}$$

where n_{em} is the number of available models for the ensemble.

The results from the ensemble for the hindcast and the 12 hours forecast at Rochelsteert model described in section 6.3.3.2 are obtained. In the hindcast the models that include Büsum and Cuxhaven water levels plus the wind, Büsum water level plus wind and Cuxhaven water level plus wind are used. In this case it is reached a RMSE of 17.26cm with a correlation of 0.99. Using all the hindcast models the RMSE is reduced to 16.22cm and the correlation still is 0.99. Compared with any single model there is a reduction in the RMSE. This is because in the case of one model diverges from the other results, the combination of them keeps the ensemble in a more consistent path. This can be observed in the Figure 6.18.a. During the flow tide, one of the members is under-predicting the water level, while the other three are closer to the measurements. The final ensemble then is compensating the discrepancies in the members. The ensemble also work as a filter reducing

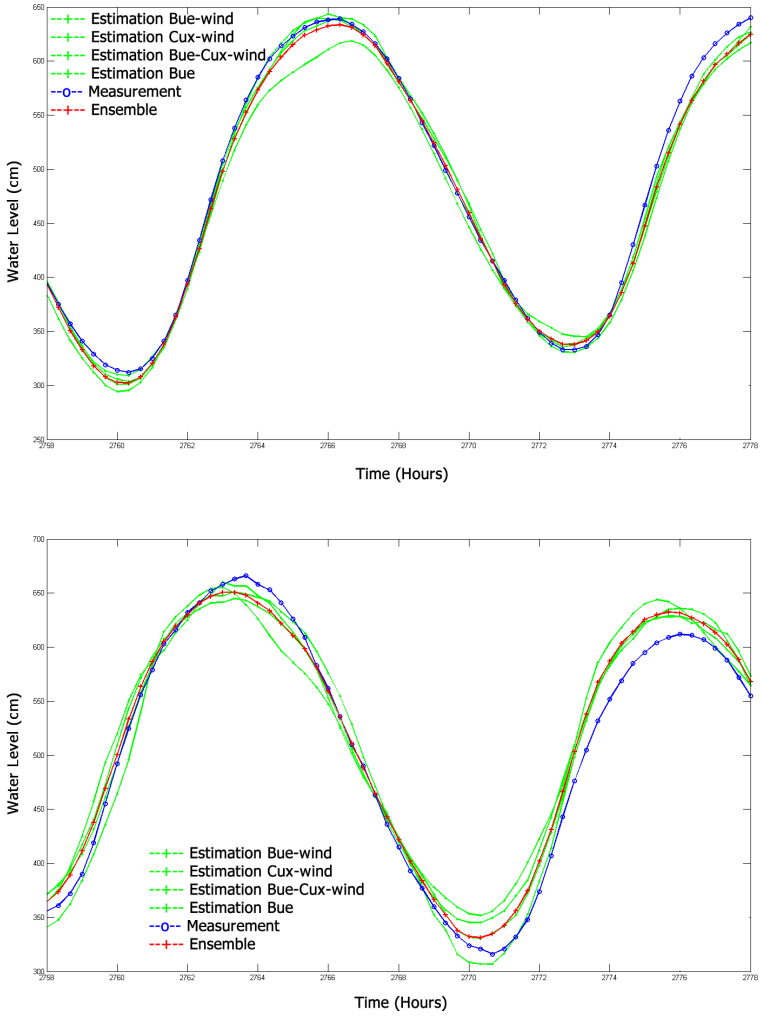


Figure 6.18: Ensemble of: a) four ANN models for hindcast and b) forecast 12 hours

small oscillations that appears in the water level.

For the 12 hours forecast (see Figure 6.18.b.), the three simple models that include wind are considered. The differences in the member results are bigger than the observed in the hindcast (see Figure 6.18 right side). As with the hindcast a reduction of the RMSE is possible with a value of $26,97cm$ and a correlation of 0.97 which is bigger than each member alone.

6.5. Sensitivity to the inputs of an ANN

To estimate how much the variations in the input arguments introduced into the ANN can affect the outputs, a sensitivity test has been done. For this purpose the original input values of water levels and wind have been modified by adding or subtracting a uniform value ($\pm\Delta$). This creates a total of three scenarios for every possible input source given: i) the original parameter value, ii) one with this value increased by Δ and iii) one reduced in the same amount. That means for a model which uses water level from Büsum, Cuxhaven, and wind direction and magnitude as inputs, there is a total of $3^4 = 81$ scenarios. Furthermore three magnitudes of Δ are used to determine the effect on the output. For water levels the Δ selected are 5, 10 and $15cm$; for wind speed 1, 2 and $5m/s$ and for wind direction 5, 10 and 15°. For wind speed, if after subtracting Δ from the measured speed the resulting value is lower than zero, then zero is taken. To measure the effect of Δ in the dispersion of the outputs, the standard deviation of the different outputs for every time step is calculated. Then the mean standard deviation and the maximum value along the whole period are tabulated.

Table 6.13 contains the results of sensitivity results for the hindcast of Rochel-

steert station using water levels at Büsum, and wind speed and direction. The combination of inputs where one or more parameters are modified (M) while the remaining are unaltered (F) is listed together with the standard deviation. As can be expected, in this table it can be seen that the standard deviation is proportional to the Δ included in the inputs. When all the parameters are altered, the mean standard deviation is generally higher, but this is not necessarily the highest case. Other combinations of parameters can bring more deviation to the outputs. The mean of the standard deviation is in general much smaller than the maximal.

The results from the sensitivity analysis show that in general there is a strong solution that is not affected by the small changes, but still at specific time steps the outputs can deviate more from the correct solution than could be desired. This is in agreement with the results discussed in Figure 6.18, showing the benefit of include the ensemble of several models to reduce the error in the forecast or hindcast output in water levels.

In terms of parameters, the wind direction alone has the lowest effect on the final result. With $\Delta = 15°$, the mean standard deviation of the water level is $2.83cm$. For the changes in the water level only, if the maximum standard deviation is multiplied by 1.96 to get the interval with 95% of confidence and then is compared with the Δ applied, they have a close value. For instance a $\Delta = 5cm$, the maximum standard deviation is $3.31cm$ and for 95% of confidence is given by $6.5cm$. This is 29% higher than the alteration introduced. However taking into account that $5cm$ is five times the maximum error expected from modern tidal gauges, the $1.5cm$ can be expected to be lower when real measurements are used. For wind direction and speed, the ratio between the mean and the maximum standard deviation is higher than for water levels. This can be explained by a strong relation between the input water level at Büsum and the output water level at Rochelsteert for hindcast. This comes from the relative vicinity of both stations, in this way they

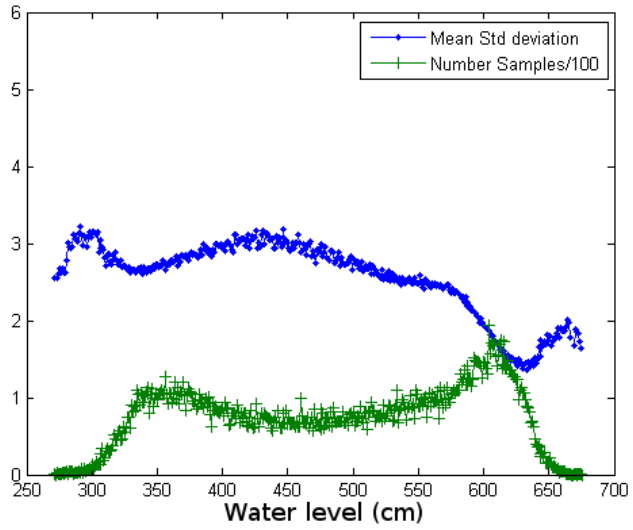
are responding to similar physical conditions.

Table 6.13: Standard deviation (cm) for the sensibility analysis of the hindcast of Rochelsteert water levels by the use of Büsum water levels and wind information

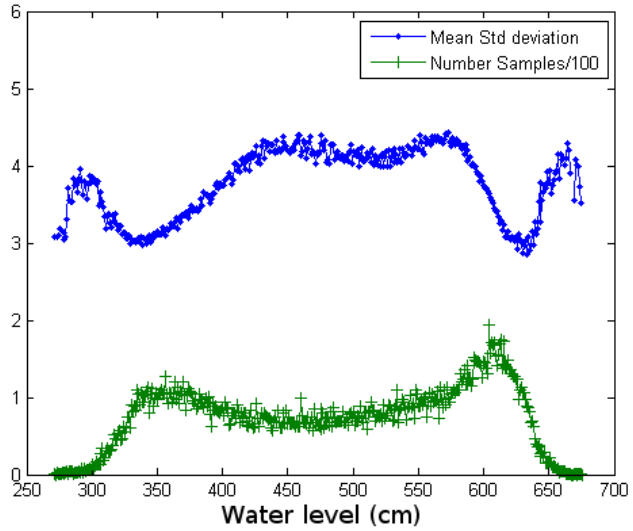
Water l.	Wind		5cm 2m/s 5°		10cm 5m/s 10°		15cm 5m/s 15°	
Bue	Mag	Dir	$\bar{\sigma}$	$max\,(\sigma)$	$\bar{\sigma}$	$max\,(\sigma)$	$\bar{\sigma}$	$max\,(\sigma)$
M	F	F	2.84	3.31	5.69	6.61	8.53	9.91
F	M	F	2.79	11.44	6.99	29.87	7.55	30.26
F	F	M	0.94	6.45	1.89	12.94	2.83	19.48
F	M	M	2.70	10.09	6.60	25.87	7.05	25.86
M	M	M	3.70	10.05	8.34	25.50	10.36	26.25

Figure 6.19.d presents the different variations obtained in more detail, which follow a parallel path. These parallel paths have a consistency behavior with the results along the rest of the time series and illustrates that the Liapunov assumption (section 6.2.3) is followed by the predictions. In figure 5.19a the distribution of the mean standard deviation along the water level is presented. In general, the mean standard deviation has a value close to 3. There are local minimums at the water level where the normal ebb and flow tides occur (around 350cm and 650cm). Beyond this range the standard deviation increases again, which is associated with a lower number of samples available for the training of the neural network that reduce the precision of the network to levels close to the ones obtained for the intertidal region.

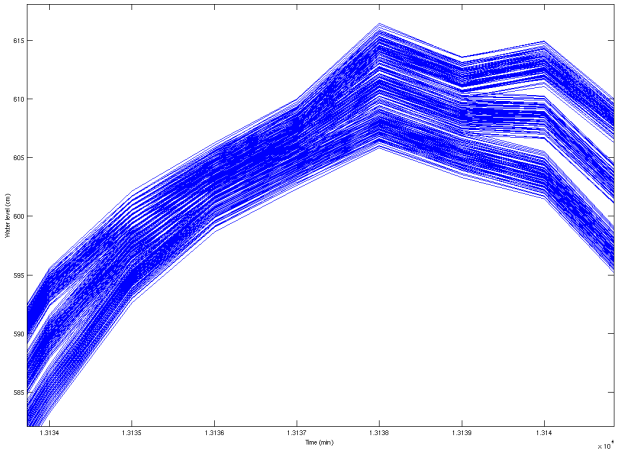
For the water level forecast of 24 hours is presented in Table 6.14 the sensitivity for the model that includes water level at Büsum and Cuxhaven, and wind magnitude and direction. Only variations of 5cm for the water level, $2m/s$ for the wind magnitude and 5° for direction are included to keep a shorter description. As in the hindcast the wind direction alone is not responsible for most of the mean variation of the outputs. On the other hand, changes of the speed of the wind has



a. Changing wl $10cm$ and wind speed constant



b. Changing wl $10cm$ and wind speed $10m/s$



c. All the single alternative member models obtained by variation of inputs

Figure 6.19: Sensitivity analysis for hindcasting

a higher effect on the outputs compared to any of the water levels from Büsum or Cuxhaven available. This is in contrast to the hindcast where variations in wind magnitude and water levels have a similar effect on the output.

This is due to for the forecast, the water level at the present time in one station has less importance by itself for the prediction 24 hours later than for the hindcast. In the case of the wind, the history carries information about the development of the atmospheric driving forces that finally are the responsible to develop the surge, meaning that for forecasting the neural network is more sensitive to the wind variation in the inputs.

Table 6.14: Standard deviation (cm) for the sensitivity analysis of the 24 hours forecast of Rochelsteert water levels by the use of Büsum and Cuxhaven levels and wind information

Water l.		Wind		5cm 2m/s 5°	
Bue	Cux	Mag	Dir	$\bar{\sigma}$	$max\,(\sigma)$
M	F	F	F	0.82	1.88
F	M	F	F	0.67	2.75
F	F	M	F	8.41	64.00
F	F	F	M	3.08	41.68
M	M	F	F	0.99	2.49
F	F	M	M	8.15	59.70
M	M	M	M	7.88	56.67

6.6. Corrected model results after assimilation

Once local models that include the assimilation of observations are built, they can be integrated into the process based model. A new evaluation of the model for

the Dithmarschen Bight model with the improved boundaries is done. Table 6.15 includes the averaged error and RMSE for the same four stations used to evaluate the model in Chapter 5. The first row contains the resume of the results from the nesting (N) -same as in Table 5.12-. In the second row are represented the results for the assimilation (H). There is an improvement in all the estimators. Reductions in the RMSE for the full series and for the ebb condition in more than 40% can be observed. The average error for the full, series is reduced to a 30%. The RMSE for flow is reduced only 9%, while the RMSE for ebb is reduced about 48%. This shows how the ANN can provide boundary conditions that can improve the final model output.

Table 6.15: Averaged errors and RMSE for the four stations considered in the evaluation of DBM model in cm

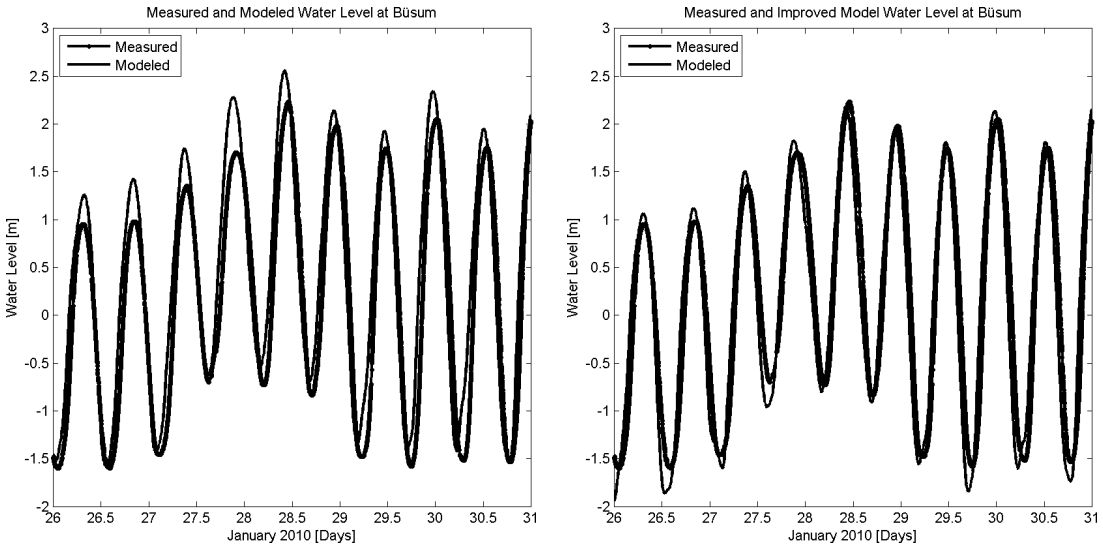
Test case	RMSE full	error full	RMSE ebb	RMSE flow	error ebb	error flow	multi-objetive	correlation
N	28.96	5.28	15.29	13.75	8.08	3.37	10.52	0.97
H	16.43	1.34	8.01	12.51	3.08	0.43	5.92	0.99

Table 6.16 provides a more detailed comparison of the RMSE and the correlation between the measurements and both, the nested process based and assimilation results for a hindcast. Even if the nesting sequence gives a high correlation, the assimilation is able to increase this to values close to the unit. For the RMSE in general the improvement is in the range of 10 to 12cm, for Brunsbüttel the RMSE is reduced in $17cm$. This represents a change close to 40%

Figure 6.20 shows the changes in the model results with and without corrections for Büsum station during a storm. Figure 6.20.a depicts the model output and the measurements, while Figure 6.20.b shows the results after assimilation is done. The nesting sequence has overestimated the storm around 50cm, while the boundaries provided by the ANN creates a series that is closer to the real conditions.

Table 6.16: Error in DBM with nesting and corrections from assimilation for July 2008

Parameter	Brunsbüttel	Büsum	Cuxhaven	Großer Vogelsand
RMSE Nesting	20.40	28.66	33.53	33.37
RMSE Corrected	3.60	18.44	20.80	22.61
Correlation Nesting	0.99	0.97	0.95	0.96
Correlation Corrected	1.00	0.99	0.98	0.98



a) measured and model without correction

b) measured and model with corrected boundary conditions

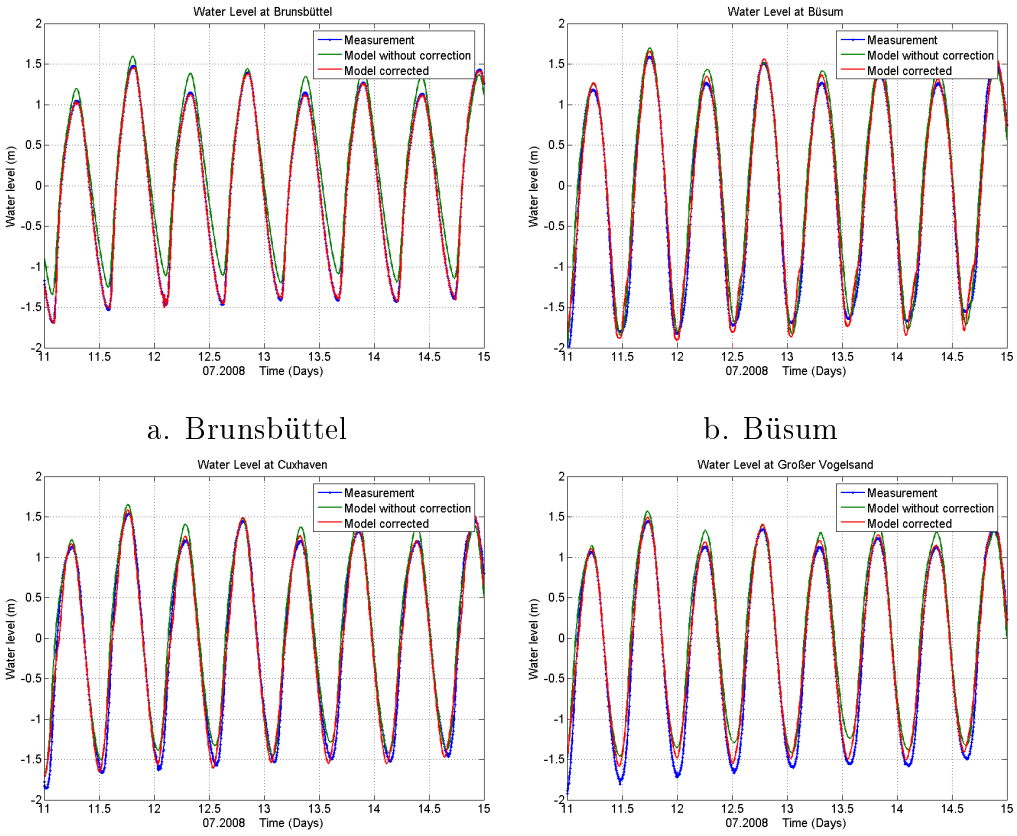
Figure 6.20: Water levels at Büsum from original and corrected boundaries, compared to the measurements

Figure 6.21 then shows for a part of the validation period of July 2008 the changes obtained by new boundary conditions at Brunsbüttel, Büsum, Cuxhaven and Großer Vogelsand. As in the previous example, there is a general reduction in the differences in all the stations. In general the results at Brunsbüttel are very good, what could be expected as this station is very close to the boundary of the model. Büsum and Cuxhaven are getting a general improvement. Großer Vogelsand that is not in the building of the local models but is taken as a verification station located in the middle of the domain, also has water level improvements.

Figure 6.22 depicts at each one of the four stations the water level obtained from the improved model against the measurements of real water levels. This figure can be compared with Figure 5.8 to see the improvements considering a bigger number of samples than in the previous figure. The reduction in the RMSE is reflected in the obtained narrower scatter plot.

6.7. Summary

In this chapter there has been described the use of correlation maps as a tool to understand the linearity among different cells in the domain. This is helpful to define a strategy to establish or to use an already existing network of sensors to be taken as source of data for a local model. From these plots it is also possible to define when linear strategies must be replaced by nonlinear tools. Three alternatives for build the local models were considered. Local models based on linear formulation are simple to construct, the regressive models provide extra degrees of freedom to the model, but both have limited quality in environments where nonlinearities are underling the process that is under consideration.



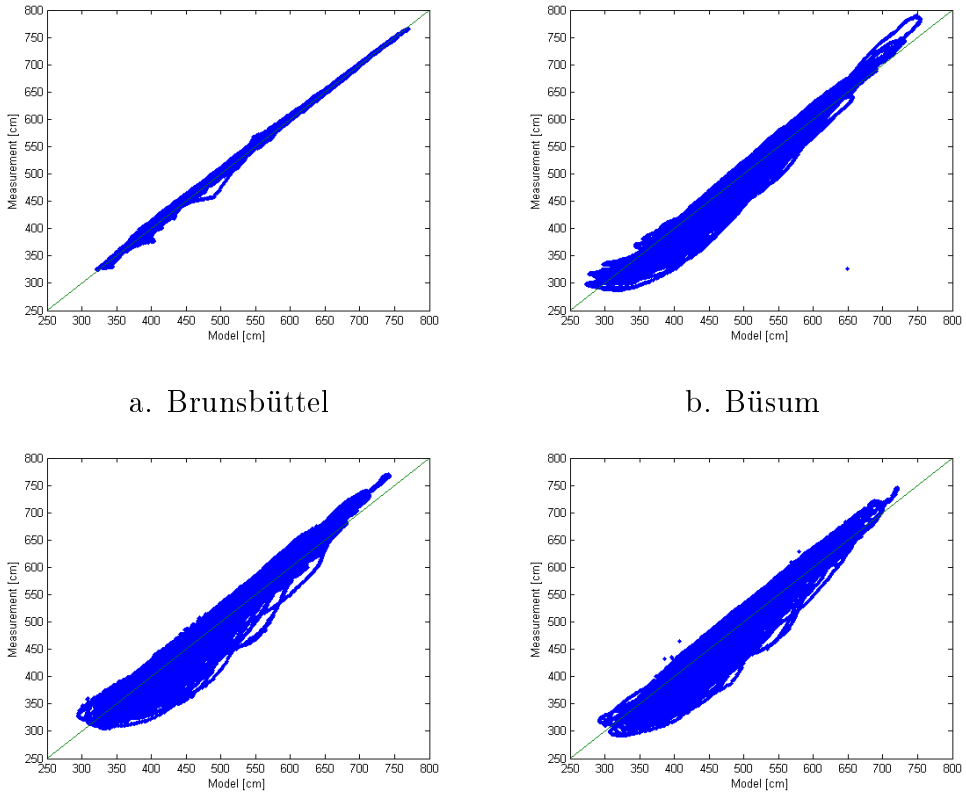
a. Brunsbüttel

b. Büsum

c. Cuxhaven

d. Großer Vogelsand

Figure 6.21: Water level correction



a. Brunsbüttel

b. Büsum

c. Cuxhaven

d. Großer Vogelsand

Figure 6.22: Improved model results versus Measurements during the validation period in July 2008, see Figure 5.8 to compare results

Neural networks are introduced to enhance the local models and they are providing a quantitative improvement in the RMSE and correlation. Fuzzy logic allows to reduce the required information to reach a similar level of errors and at the same time gives a segmentation of the errors, helping in the understanding of where the incertitudes are coming from. The ensemble reduces the sensitivity of the local model, providing robustness to the outputs.

Chapter 7

Hydrodynamical Operational Model for the Dithmarschen Bight

In the Chapter 5 a set of process based models and local models were obtained and then in Chapter 6 local models to provide improved boundary conditions were described. The present Chapter presents the integration of all the components in a system that is operational at the present time. This system is able to deliver improved forecasts of water levels for the Dithmarschen Bight. In the present chapter a description of strategies to integrate the model engine (Delft3D and SWAN solvers) into the operational model (OM) is given. Strategies to speed up the execution time of the model are discussed. With this aim, an algorithm to calculate the hydrodynamic terms of a nested sequence of models in parallel is proposed. Subsequently, an explanation of the scheme to obtain the preliminary forecast, the hindcast, and the improved forecast will follow. Also the data storage is discussed. Finally, the web application developed to display the contents is described. This system is already implemented and in current operation.

7.1. Some considerations about model engine and the operational model design

In a model for the prediction in near-real time it is desirable to have both speed of execution and accuracy of results. The speed is important to provide results as soon as possible and make it available to be integrated in the decision support within an early warning system. Accurate results are obviously valuable to reduce the uncertainty when such decisions are made. To obtain high accuracy it usually requires more computational time, which means both objectives are in conflict. In order to resolve this conflict, it is useful to combine various techniques to achieve a desired balance between them. In Chapter 5, the development of a model set-up for the hydrodynamics in the area of interest and its accuracy has been discussed. In Chapter 6 data assimilation to impose boundary conditions capable to improve the model output has been elaborated. In this section the interest lies in defining evaluation strategies to provide the outputs in a fast way. A discussion of the advantages of nesting over the domain decomposition for the assimilation of data when Delft3D is used. The weak point of nesting is the necessity of running the sequence of models in series, to improve the performance, an algorithm to reduce the computational time of the nesting sequence is proposed.

7.1.1. Domain decomposition vs. nesting in the operational model

The domain decomposition has been discarded for the implementation of the operational model. The first reason is that when the original set of models has been constructed, it was not available as a feature of Delft3D. But later, when it became

a new feature, it shows the limitations described in section 2.2.1.3 about running all the required sets together. The requisite of all the models using the same time step is not a problem due to the fact that the most expensive task is the wave model and the CSM do not have this module active. The domain decomposition also creates a problem to impose the improved boundary conditions in the area of interest. This is due to the fact that the tool does not allow the introduction of values in the region where the interchange of information in the domain decomposition is done. Then, boundary conditions can be changed only for the CSM. With nesting it is possible to run the coarser models only once with the limited early information available. When new measurements are available, a hindcast and an improved forecast for the smaller model can be done, bounding the new calculations only for the region of interest. See section 7.2 for more details.

7.1.2. Parallelization of the nesting sequence

For the calculation with the nesting sequence, the models must be calculated sequentially. This consumes valuable time when near real-time results are necessary. With this limitation in mind, an evaluation scheme for the nesting chain is proposed and implemented in the operational model.

Some sort of parallelization with the nesting of the three models (CSM, GBM and DBM) is also possible. Instead of running a long period at once, it is evaluated in smaller time windows and with the results built the boundary condition of the next model (see Figure 7.1). Then the nest model is able to start with the second time subdivision and at the same time the nested model process the first time subdivision of the calculations. Assuming that all the models require the same time to be executed (that will be depending of the number of cells in the grid and activated parameters in every model), the total time required for the simulation

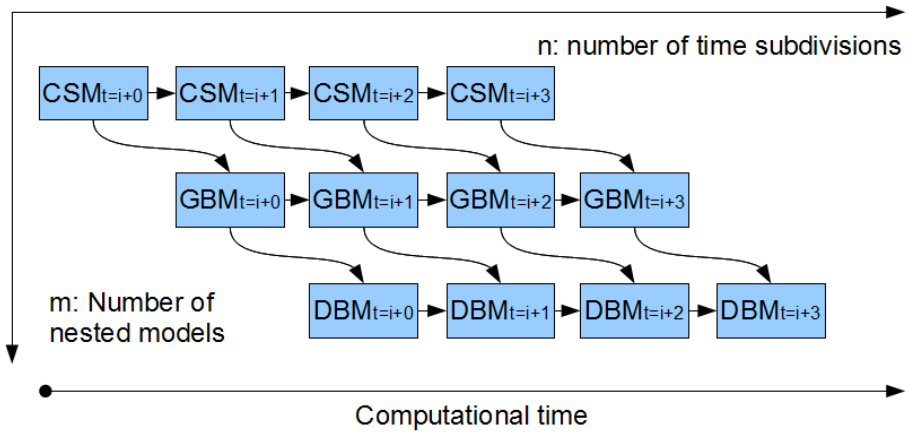


Figure 7.1: Parallelization of a nesting execution

for a nesting chain of m models and using n time subdivisions can be expressed by:

$$T_{total} = T_{total\ one\ single\ model}\left(1 + \frac{(m-1)}{n}\right) + O(n) \tag{7.1}$$

In the example provided in Figure 7.1, each model is divided into four blocks ($n = 4$), the total time is only 50% more than the required to execute one single model, much less than the three times the computational time required if a pure serial ($n = 1$) is employed. Both m and n must be at least equal to unity. The upper value to the number of time window subdivision n is theoretically unlimited, but the time span of each period must be a multiple of the model time step. For practical reasons the minimal time span used in this work will be one hour.

A test to evaluate the optimal number of subdivisions for 24 hours of simulation time has been carried. In the course of this test, the simulation using the standard nesting sequence, where each model was solved for the complete period before starting the next finer one. The total time for the evaluation of the non-parallel version has been taken as the reference time. For comparison two, three, four, five,

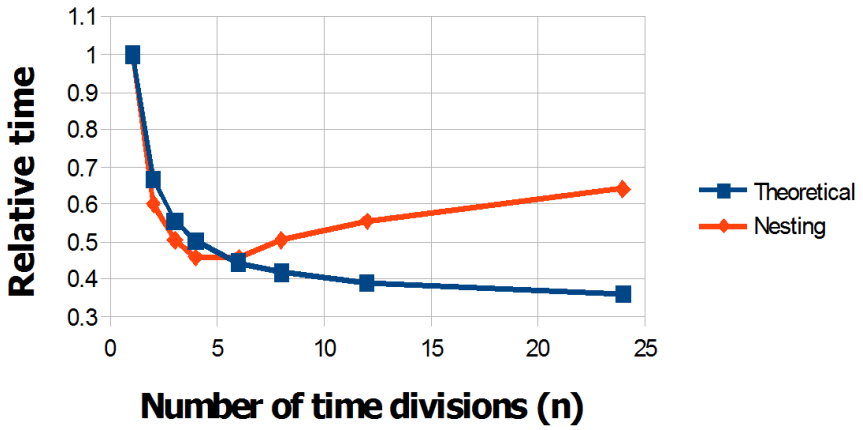


Figure 7.2: Evaluation of computational time for a parallelized nesting sequence

six, eight twelve and 24 intervals where evaluated. As it is noticeable in Figure 7.2, there is a significant reduction in the computational time when a partition is made. The best performance is obtained for four or six intervals ($6h$ or $4h$ respectively) with a required time equivalent to 47% of the time required if the nesting sequence where not split and evaluated in parallel. If more subdivisions are used an increase in the total required time for the simulation is present. This is caused by the extra effort required in the pre- and postprocessing of each model.

7.1.3. Adaptive time step

In the original set of models that employs only nesting, a time step for the flow model of three to six minutes is established as a reasonable value to get accurate results with less computational effort. For the operational model three minutes (as mentioned in Chapter 5) is defined for the flow model. This time step has worked properly for input data since December 2007. However, the model crashes in some cases with strong wind conditions when the same time step is kept. This happens when new boundary conditions based on observations are imposed for

hindcast in the operational model. This is due to the high speed currents and water level gradients that produces instabilities in the model. A way to avoid high gradients is the reduction of the time step, but depending on the flow conditions it can reach a value of $0.20s$. This extremely low value is not acceptable to be fixed, as it drastically and unnecessarily increases the simulation time. An adaptive time step is the ideal solution. But as Delft3D does not include adaptive time step as a feature, the only option is to set the time step as a parameter that can be adjusted when the input files are created before running the model.

Two approaches have been considered: i) the time step is set as a function of wind speed and ii) the other with an iterative selection. When low wind speeds are present, the solution is normally smooth and the original time step can be kept. When the wind speed rises, the time step can be lowered to avoid instabilities when strong currents occurs. With this approach it should be possible to find a reasonable time step according to the weather conditions and run the model only once with minimum error. The problem with this approach is that not only the maximum wind speed determines the gradients, but also the fetch, and the part of the tidal cycle at which it occurs. Moreover, the infrequency of these events hampers the formulation of an empirical relationship between winds and stable simulation time steps. Also the low occurrence of this kind of problems makes difficult the collection of information to perform an equation fit. For the second method to find an optimal relation in the iterative approach, the numerical model is executed once with the standard time step of 3 minutes (see Figure 7.3). If any gradient problem occurs, the time step is reduced to one fourth and the same period is evaluated again in a loop until it is able to obtain a solution or reach a minimum time step. By testing different time steps in this set, it was concluded that the lowest time step is $0.06s$. When the value of $0.06s$ is reached, instabilities in the numerical solution appear.

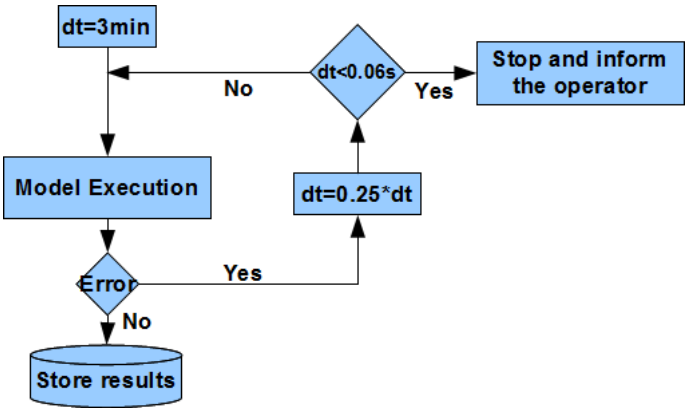


Figure 7.3: Automatic time step adjustment

7.1.4. Selection of number of cores for SWAN simulations.

SWAN is available to be run in serial or in parallel mode, taking benefit from multi-core processors. The total computational time by using p processors is given by:

$$T_{total} = \frac{T_{serial}}{p} + O(p) \tag{7.2}$$

Due to interprocess communication there is an overhead ($O(p)$) that produces a slight reduction in the performance. To decide the optimum number of processors dedicated to SWAN, the DBM model has been evaluated using both, the serial and also the parallel version with m changing from 1 to 24 -the total number of processors in the machine- (see Fig. 7.4). Every extra core utilized in the calculation reduces the global computation time; however, at a decreasing rate. The second processor reduces around 50% of the total computational time, while the fourth brings 10% and the sixth cuts down the time by only 2%. For this model four processors have been selected. In this way, during the nesting evaluation, up to nine cores are in use simultaneously (four for flow-wave in GBM and DBM, and one for flow in CSM). It is important to note that the consequence of the

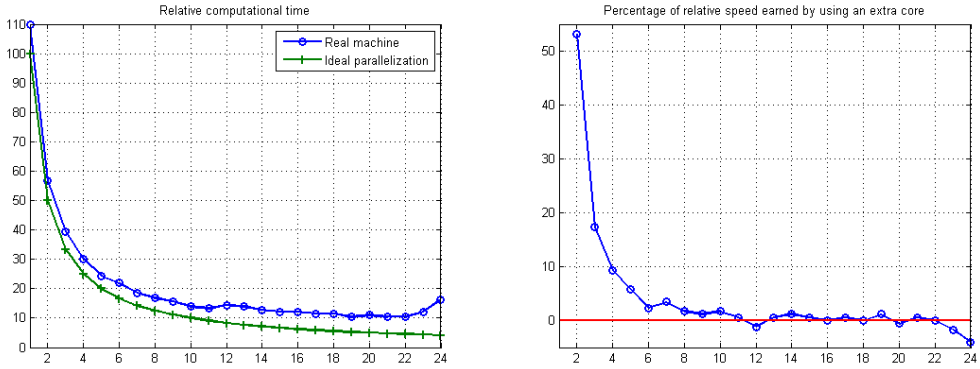


Figure 7.4: Computational time of parallel realizations of SWAN relative to the serial version (left side) and the percentage of relative speed up of the computational time when a processor is added (right side)

utilization of all cores results in a decrease in the performance.

7.2. Execution of the operational model

In section 5.1.1 the nesting sequence used in the flow model has been described. In the operational model, the same set of models is in use, with the addition of automation strategies and data assimilation (see Fig. 7.5). For the operation, three modules are defined: the nested forecast (NF), the hindcast (H) and the improved forecast (IF). The nested forecast is basically the nesting sequence that was already described in section 5.1.1 where all the models are driven by the DWD wind information and no assimilation of water levels or waves is included. The hindcast module excludes the execution of CSM and GBM. It evaluates only the DBM model and includes measurement assimilation to improve the boundary conditions. The improved forecast is also executed only for the DBM model. The boundary conditions are calculated from a combination of local models and the output from a nested forecast module. The result from each model is stored in the database and is accessible independently to other components of the system.

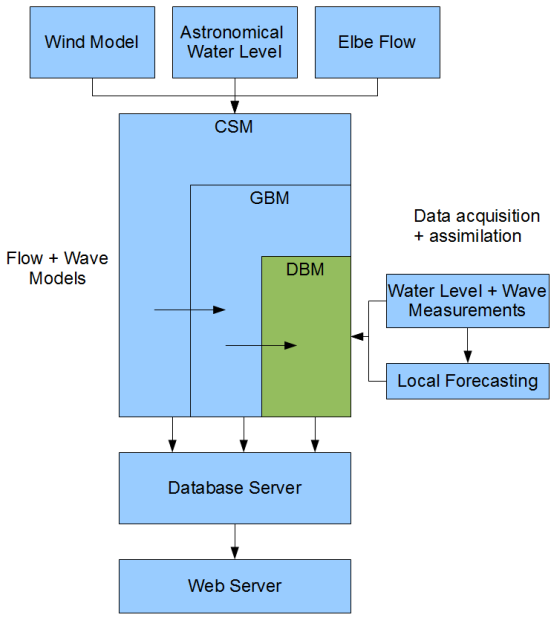


Figure 7.5: Nowcasting system general description

In Figure 7.6 a schematic representation of the execution time of the model engine to obtain outputs for a defined modeled period is depicted. At the beginning, a nested forecast for the modeled period i (NF_i) is executed. The initial conditions for every forecast period are obtained from the results of the previous execution of the (NF) module. Thus, the warming up period should be considered only when the system begins to operate without previous initial conditions. In this case the initial condition of NF_{II} is given by the final condition of NF_I. This calculation is performed every 12 hours, as soon as the wind forecast is accessible (around 6:00 and 18:00 every day). Then the results from NF are available always previously to the day of interest. The reference time is defined as the starting time of each simulation to avoid problems with the shift in the tidal boundary condition. The hindcast module at a modeled period i (H_i) employs only the DBM model with the boundary conditions imposed with the corrections calculated from the comparison of results obtained in (NF_i) and measurements for that period. The initial condition for H_i is provided by the final condition in H_{i-1}. This module can be executed only at the end of the modeled period. The minimum modeling period has been set to one hour due to the time step of the wave model.

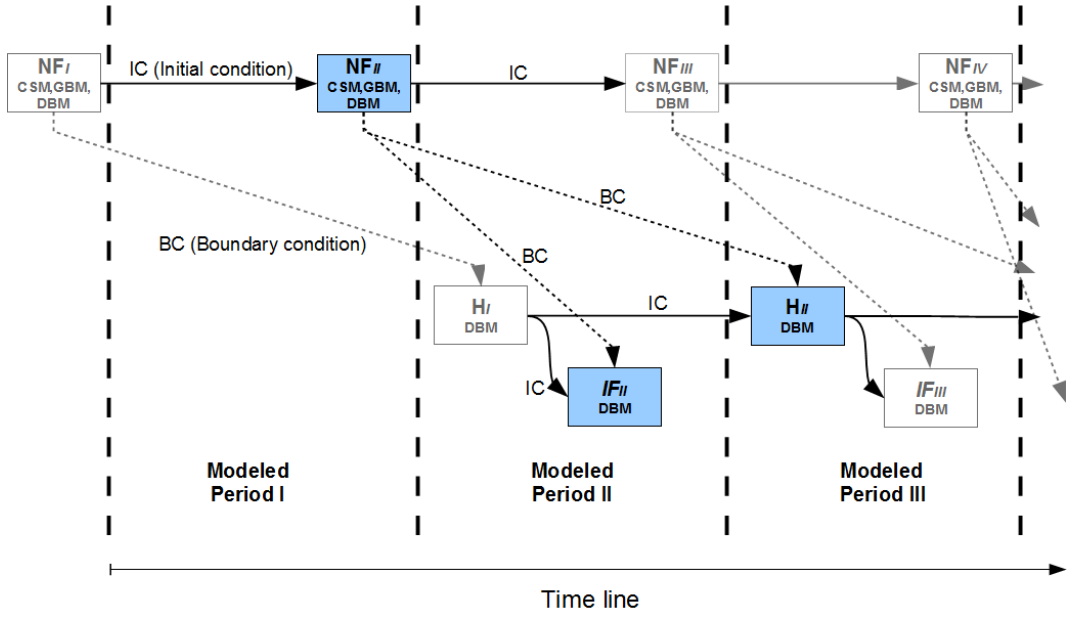


Figure 7.6: Sequence of model running showing the original forecast (NF), the hindcast (H) and the improved forecast (IF).

The improved forecast (IF_i) is evaluated to get enhanced results in the predictions. The initial conditions are obtained from the results of the preceding hindcast period H_{i-1}. The boundary conditions are built by the use of information from the nested forecast and local models based on measurements in the immediate previous time. Continuity of the boundary conditions must be guaranteed when coupling the initial conditions coming from H_{i-1} and the local model. This module is executed once the hindcast module is finished. The forecast period is defined by the local model constructed for the boundary conditions. Local models for 2, 6, 12 and 24 hours forecast are described in section 6.3.3.

7.3. Dissemination of the information

Data center A data center to gather the information from several sources has been implemented. Those sources include the real time measurements described

in section 4.2, wind models provided by the *Deutscher Wetterdienst* and the hydrodynamic model outputs. The data center is very important as it is responsible to keep the data safe and is the core of the interchange of information among the different components of the system. There are basically three types of data being stored: time series for individual observation points, 2D maps, both for static information (for example grid coordinates) or time series (currents in the domain) and metadata of the devices or other components of the system.

For the time series, one table for each parameter to be stored is created. The information of each station is stored in one column and the time of the measurement is taken as the primary key. The primary key uniquely identifies each record in the table to avoid duplication of the information. In this kind of tables the model observation points for water levels and measurements of water level, wind and wave are stored. For spatial information, the matrix containing the 2D-data is transformed into a column vector and saved in binary format. The column is defined as a Binary Large Object (BLOB). The original dimension of the matrix and the associated grid of coordinates are stored in an additional table. The primary key is a combination of the time and the stored parameter, being possible to store only information for one pair of time-parameter at the database. This kind of storage does not allow to use directly the structured query language (SQL) to extract time series for a specific location. However, storing only one vector saves a large amount of space. Whenever time series are needed, they can be stored directly as local time series described previously. Finally, metadata information for the different elements of the system is stored. For gauges this includes station name, coordinates and type of measurements. For grids the metadata information includes the number of cells in each dimension and coordinate projection.

Data delivery To deliver the information to final users, a web interface has been developed. The use of web tools facilitate the access to the public without

the necessity of installing specialized tools on the user's computer.

The interface is based in the combination of several web technologies and tools:

1. HTML
2. JavaScript
3. PHP
4. Google Earth API
5. Python
6. mapscript
7. wsgi

The HTML is the framework technology that allows to put together all the other components of the interface. In the main area of the screen a map is available to display the information (see Figure 7.7). The map is based on Google Earth API which permits the combination of geo-referenced information including spatial maps, vectors and points with a 3D map of the Earth. This is suitable to display maps with model results, localization of the fixed measurement stations or even mobile stations.

From the hydrodynamic model there are maps available for water level (see Figure 7.9), current velocity and direction, significant wave height and wave period, wave dissipation and wave direction. The wind model provides wind speed (see Figures 7.12 and 7.11), pressure and temperature. The display of this information is

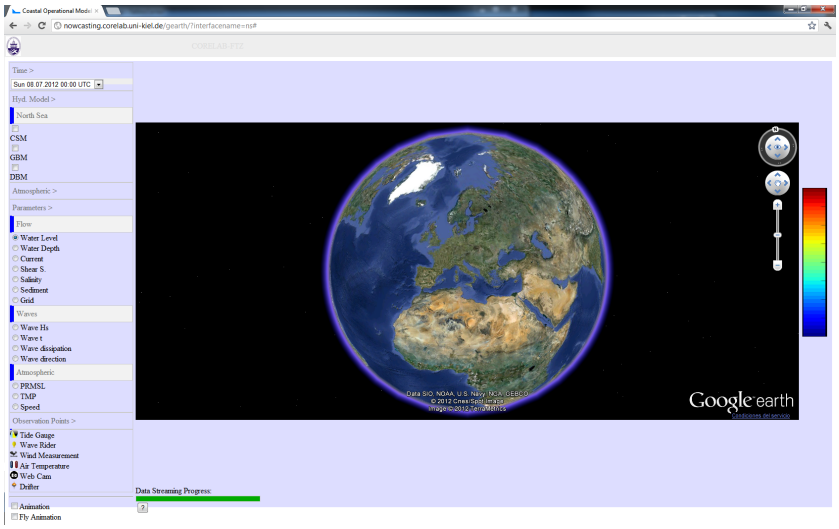


Figure 7.7: Main window of the screen

controlled in the left menu (see Figure 7.8). Here it is possible to select the active maps and the time to be displayed.

Time series from all the available measurements are also presented. Those series are displayed as pop-up windows inside the map area (see Figure 7.10) and can be seen in more detail in an additional browser tab. Other georeferenced information can be included in the system, for example cameras installed along the coast line (see Figure 7.13) permit a visual inspection of the actual state of the sea.

7.4. Summary

The structure and functionality of the operational model able to deliver optimized hind- and forecast has been presented. An algorithm to speed up the computation of a nesting sequence was also proposed. The total time required to evaluate the nesting sequence for modeling 24 hours has been reduced to only 47% of the total time required to evaluate the same period without the proposed method. Also

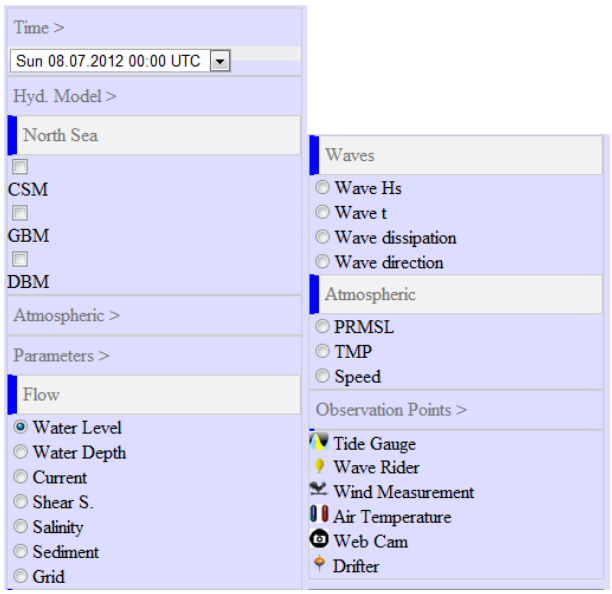


Figure 7.8: Details of the available options in the left menu

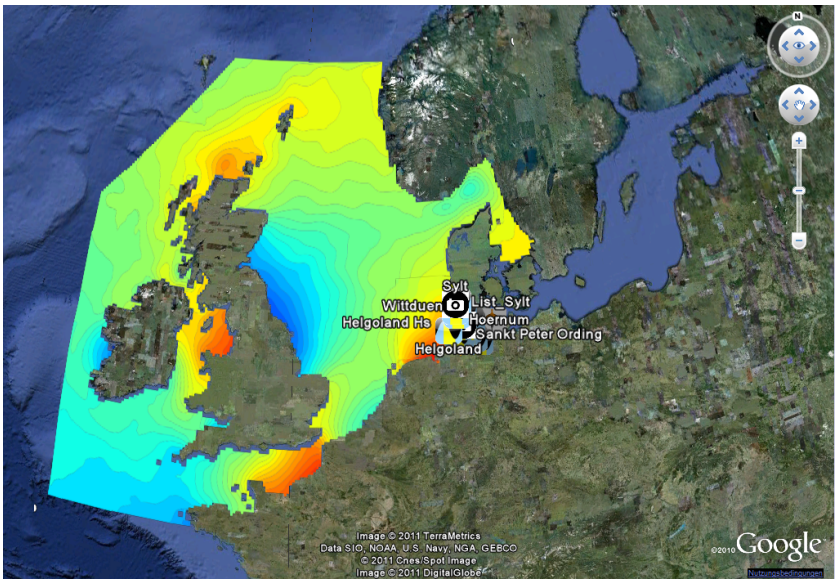


Figure 7.9: Zoom to the North Sea, showing water level map

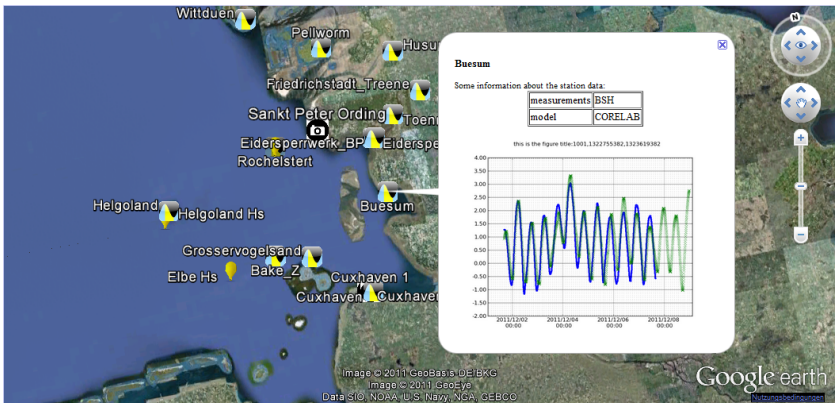


Figure 7.10: Water level time series at Büsum

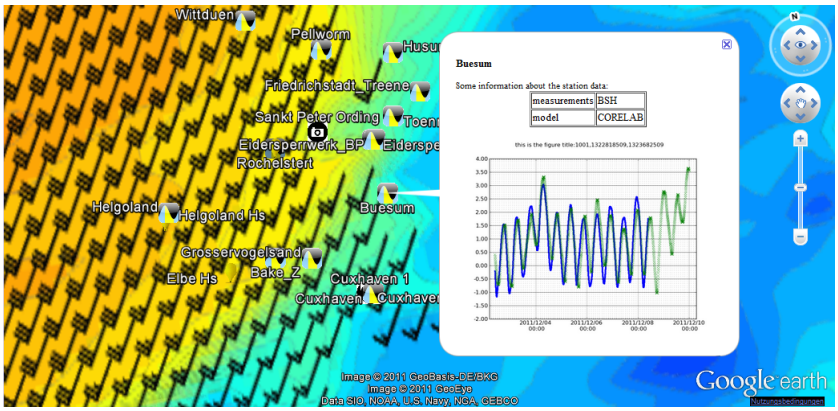


Figure 7.11: Wind maps in combination with water level time series at Büsum

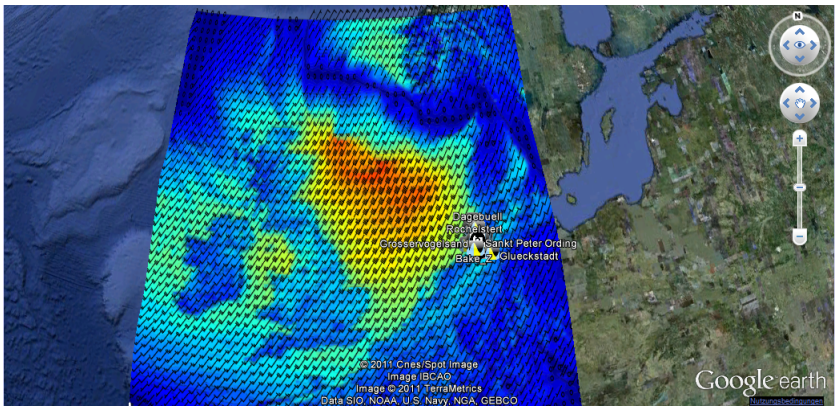


Figure 7.12: Wind fields at the North sea

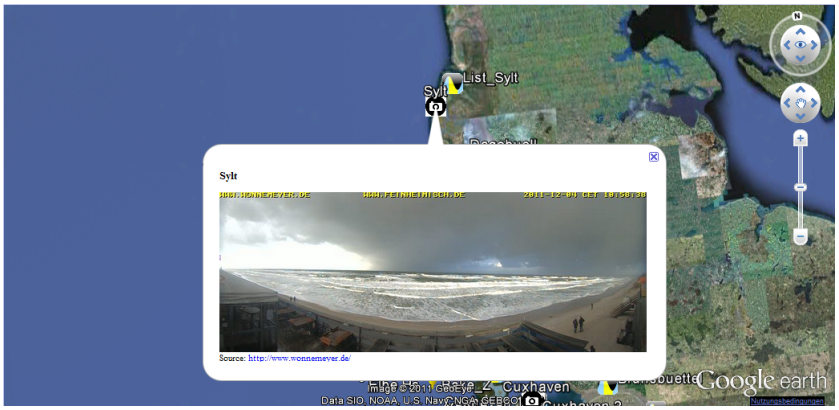


Figure 7.13: Camera attached to the system (Operated by X-H2O)

an algorithm to reduce the time step of the model has been included to make the operation of the system more robust and tolerant to fails when some extreme conditions are reached. Finally, the storage of information and the graphical interface is described. This provides a simple interface that allows the access of the most relevant information. It can be extended and adapted easily to new parameters, new devices that become available in the area or even models in other areas of the world.

Chapter 8

Conclusions and recommendations for future work

In the present work, a data assimilation technique for setting-up operational models for tidally dominated coastal area has been developed. The assimilation leads to an improvement in the model results. In addition, an operational model for the Dithmarschen Bight area has been designed incorporating the proposed data assimilation. The work has been developed in three main parts: i) improvement of the existing models in the area of interest, ii) analysis and development of functions that assimilates the measurements from tidal gauges along the coast line into the prescribed boundary conditions and iii) design and implementation of the operational model based on the established data assimilation.

8.1. Conclusions

A well calibrated and validated hydrodynamic process-based-model provides the modeler with the possibility to analyze different scenarios under conditions beyond the availability of continuous measurements. It also describes the whole domain, providing the possibility to understand the general process even in locations where no measurements are available. Process-based-models have a good performance when they are provided with accurate information. The final quality of the model is restricted by the lack of sufficient information for the entire domain. Although the bathymetry can be obtained with some level of accuracy with modern technology, for large areas like the North Sea, it is not feasible from an economical point of view. It is even harder, for example, to obtain a full knowledge of sediment information to establish the roughness properly in such big areas.

The set of depth-integrated (2DH) models for the entire continental shelf of the North Sea, the German and the Dithmarschen Bights have been improved successfully. For this, a newer bathymetry from measurements covering all the Dithmarschen Bight, parts of the German Bight and continental shelf has been added. In the previous versions of the Dithmarschen model, it was covered by one single cell of the PRISMA wind model, and it is necessary to use uniform wind. A newer source of wind information from DWD with a higher spatial and temporal resolution permits to use spatial distributed wind field over the Dithmarschen model. Additionally to the higher resolution of the wind information, the on-line coupling of wave and flow during the calculation of the model required a new calibration process for the wind drag coefficient and bottom friction. This process allows during calm periods an improvement on the shape of the tides, in addition during storm conditions, to a reduction in the multi-objective error for the test stations of 23%.

To improve the quality of the boundary conditions obtained from the nesting sequence, a new methodology to use the process-based-model outputs to build correction functions has been implemented. This method takes time series from two locations within the model domain and creates a function that can connect both points to reproduce one in terms of the other. Three types of functions have been tested, linear with a time shift, regressive and artificial neural networks. The linear functions provide a good estimation as long linearity in the model can be kept. In this case, linear functions connecting the water level of the stations at Büsum and the North West boundary at Rochelsteert and at Brunsbüttel with the East boundary provide a correlation of 0.95 and 0.99 respectively. The regressive functions are able to increase the accuracy of the correction by a reduction of the RMSE, but for the tested cases, it was not improving the correlation compared with the linear functions. Stations located in shallow areas are not able to be represented by a linear function, since the nonlinear friction terms play a more important role.

nonlinear relationships have been established by the study of the time series as state variables removing the temporal dimension. From the perspective of the dynamical analysis, the water levels at two stations are represented by an orbit around a line. This line is acting as the attractor of a Liapunov stable orbit for the water level samples. Since an artificial neural network (ANN) is able to consider nonlinear relationships among state variables, they have been used to create functions to translate water level information.

To test the performance of the ANN avoiding the effect of using model outputs as the inputs for the ANN, the reconstruction of water level measurements for one station from measurements available at an alternative tidal gauge has been done. Particular results from calculating water levels at Rochelsteert pier based on measurements at Büsum have been made. For this reconstruction, the ANN

is able to provide very good results with a correlation of 0.99 and an RMSE of $14.26cm$. To have an idea of the improvement obtained using the Rochelsteert pier and Büsum water levels, the RMSE from ANN is the half of the RMSE at Büsum during the validation of the Hindcast with the process-based-model.

Several scenarios for hindcast have been tested in combination with ANN by using water level outputs from process-based-model. With this, it was possible to reduce the errors obtained from linear or regressive functions with RMSE in the same order of magnitude as the network obtained by the use of measurements. For the forecast, with a prediction window of only two hours, only the use of water levels as input for the ANN starts to show limitations in its performance. The inclusion of hourly averaged local wind information was required to differentiate present sea states that can lead to different future sea conditions.

Model outputs are limited by the capacity of the process based model to describe the relation of the water level as some simplifications are made. If there are measurements, then it will be preferable to perform the training (the process of providing samples and adjusting the parameters of the ANN) of the ANN with them. However, the lack of measurements along the domain makes the use of the model outputs to create ANN a powerful tool to improve model results.

One limitation of ANN is related to find the optimal solution for the cost function during the training . Due to the large number of involved parameters and the nonlinearity of the terms, the training of the network can be computationally expensive. In this case, the convergence of the ANN to global optimal solution is not guaranteed. The reason is that normally, only local optimal solutions are obtained. Then, the use of alternative ANNs for the same data settings helps in the evaluation of the quality of the estimation, when different networks are able to provide a similar level of error in the forecast they can be accepted. For the

training, a combination of deterministic and stochastic optimization algorithms was found as a good way to obtain the ANN settings. Back propagation of the error generates a rapid convergence to local minimums by reducing the error at each iteration, and at the same time the simulated annealing allows the solution to escape from local minimums by allowing slight increments of the error that can bring the solution in to the vicinity of a better solution.

To limit the effort during the training, a reduction of the number of available samples has been proposed by removing the redundant input states. After calculating the angle between each element candidate to belong to the training pool, those with a small angle are discarded. The reduction for ANN with only water levels as inputs, a $cos(\alpha) <= .99$ reduces the number of samples from 20,000 to 1,176. When wind is also included, there is a higher variability of the components of the input vector. In order to reduce the number of samples, a wider range of angles should be excluded. One additional benefit of the removal of redundant samples is the reduction of memorization (When the neural network loses the ability of doing generalizations, and is only able to reproduce previous inputs).

The sensitivity for ANN shows that the variation of water level outputs for hindcast is in the same order of magnitude as the differences in the inputs. This means that as long as the level measurements are provided with good accuracy, this error will not grow. The wind during hindcast did not show relevance in the propagation of the error since hindcast is strongly dominated by water level inputs. For the forecast the inverse situation is presented, the errors in previous water levels do not show a large influence in the outputs, but the wind plays a big role. This means that good quality wind measurements are crucial to avoid introducing noise to the forecasts.

In general using ANN provides a good approach to the estimation of water levels

at the station, but in some cases, discrepancies are present. Also after knowing the parameters for an ANN, its evaluation is a fast process. With this in mind, an ensemble of several ANN was tested. The weight of each member of the ensemble (the output from each of the models used) has been selected by giving more importance to the single ANN models with a lower RMSE. This ensemble shows reduction on the RMSE when compared with each one of the members. It can be expected that an increase on the number of members will lead into a much better robustness to the ensemble.

The proposed operational system consists of five modules: the data storage, the numerical model, data acquisition, data assimilation and the graphical user interface. Within the system, the data storage module is in communication with all the other components. This is the responsible to store the information generated by all the components and to provide the information required for their functionality. The other modules can be replaced providing new features without affecting the rest of the elements. The numerical model can be extended to provide support to other available numerical solvers such as Mike or ROMS. This module also can be complemented by adding new model sets for other regions in the world. The graphical interface based in a set of web technologies provides access to the system from any place where Internet is available, being possible to display several sources of data. Model outputs and measurements are also available via the database in the form of spatial maps and local time series.

The operational model provides forecasts in two steps. At the beginning, forecast based only in astronomical boundary conditions and wind driven forces is delivered. This provides a first approach to the future conditions of the tides without depending on the network of devices. In a second step, the measurements are integrated with the model bringing enhanced results. With this design, the most accurate available results are displayed, while providing a robust solution tolerant

to problems in the data communication from individual stations.

Efforts in reducing the computational time of the system have been made. An algorithm to speed the execution time of the nesting sequence has been developed. It was possible to reduce the total execution by more than 50%, what is a valuable time when results are required for near real-time forecast. An adaptive time step is included to modify the resolution in the temporal calculations, during calm periods the system can use a relative coarse step, but when strong currents during storms are present, is able to reduce the time step to keep the model in operation.

8.2. Future work.

To improve the accuracy of the actual set of models, a finer grid able to include information of smaller features as coast line, channels in the tidal flat or islands could be implemented. This has the inconvenience of consuming more computational resources, reducing the performance of the operational model. To overcome this problem, the use of parallel models can speed up the calculations, but at the moment this alternative is not always available. In the case of Delft3D for the flow model and the SWAN for the wave model, they can be executed at the present time in parallel when each one is evaluated independently, but not when they are coupled.

The continental shelf model cannot consider small changes in the bathymetry due to the grid size. Also there is no detailed bathymetry for the entire area. This leads to a limited capacity to provide accurate boundary conditions to the nested models. A higher resolution grid with a wide number of measurements could enhance the results. But the improvements are only possible if better measurements

are available. For the German Bight model, an improvement in the region of the Wadden sea can be expected by using a finer grid

A surrogate model is one that acts in behalf of other model that in general is much more complex. The surrogate model is able to deliver results much faster with a lower accuracy than the detailed model, but if convenient transformation functions are built between both models, then is possible to avoid the calculation of the detailed model and still obtain an improved accuracy when compared with the coarse model alone.

Another possible solution to take benefit of a finer grid, while keeping a fast calculation, is the use of a surrogate model. The Surrogate model is a strategy where two models with different levels of complexity and detail are used in combination to obtain faster results. Relations between the possible outputs from both models are established, and once the coarse model is calculated, its results can be transformed as they were obtained from the finer model. Here, both linear and ANN functions could be used. It can be suggested that, since bathymetry measurements are available in an important area of the actual German Bight model (GBM), the GBM could be used as a surrogate model for a model of the same area but with a finer grid. With this, it is possible to keep the speed of the actual models, but enhance the results obtained with the GBM (Getting better boundary conditions for the DBM). The same can be done for the Dithmarschen Bight model, improving the ability to describe minor channels in this area. For the continental shelf model, since there is not an intensive source of information to improve the bathymetry resolution, a model with a finer grid will not provide a more accurate output.

On top of the deterministic models, the model output statistics (MOS) can be implemented to complement the system. It provides to the operational model with confidence intervals that allow the model to have an estimation of the prediction

quality.

In the area of the Dithmarschen, the tidal channels and tidal flats are in constant change. To keep the quality of the system, periodic updates from in-situ measurements of the bathymetry are recommended.

Bibliography

Richard A. Allard, James D. Dykes, Yuan-Huang L. Hsu, James M. Kaihatu, and Daniel Conley. A real-time nearshore wave and current prediction system. *Journal of Marine Systems*, 2007.

John D. Jr Anderson. *Computational fluid dynamics: the basics with applications*. Aeronautical and Aerospace Engineering. McGraw-Hill, New York, 1995.

Francis Auclair, Patrick Marsaleix, and Pierre De Mey. Space-time structure and dynamics of the forecast error in a coastal circulation model of the gulf of lions. *Dynamics of Atmospheres and Oceans*, 36(4):309–346, 2003.

Vladan Babovic, S.A. Sannasirajb, , and Eng Soon Chan. Error correction of a predictive ocean wave model using local model approximation. *Journal of Marine Systems*, 53:1–17, 2005.

G.K. Batchelor. *An Introduction to Fluid Mechanics*. Cambridge University Press, 1967.

Eric B. Baum and David Haussler. What size net gives valid generalization? *Neural Comput.*, 1(1):151–160, 1989. ISSN 0899-7667.

Bunchingiv Bazartseren, Gerald Hildebrandt, and K.-P. Holz. Short-term water level prediction using neural networks and neuro-fuzzy approach. *Neurocomputing*, 55(3-4):439–450, 2003.

Bob Beardlsley and Rick Pawlowicz. Air-sea v2.0: Compute surface wind stress

and heat flux components from buoy and shipboard atmospheric and near-surface oceanographic time series measurements. http://woodshole.er.usgs.gov/operations/sea-mat/air_sea-html/index.html, 1999.

W.G. Beeftink. The coastal salt marshes of western and northern europe: an ecological and phytosociological approach. In V.J. Chapman, editor, *Wet coastal ecosystems*, Ecosystems of the world, pages 109–155. Elsevier Scientific Pub. Co., 1977. ISBN 9780444415608.

Yuley Mildrey Cardona and Jose Manuel Fernandez Jaramillo. Análisis de mareas por el método de la descomposición en armónicos. Diplom's thesis, Facultad Nacional de Minas, Universidad Nacional de Colombia,Medellin,Colombia, 2001.

Ralph T. Cheng and Richard E. Smith. A nowcasting model for tides and tidal currents in San Francisco bay, California. In *Ocean Community Conference '98*, pages 537–543, Baltimore, November 1998. Marine Technology Society.

Sina Clorious. Seabed meets horizon. http://www.wattenmeer-nationalpark.de/flag/engl.pdf, 2008.

Peter Constantin. Some open problems and research directions in the mathematical study of fluid dynamics . *preprint*, pages 1–11, 2000.

COSMO. Operations at dwd - cosmo-eu. Technical report, COSMO, 12 2007.

R. Courant, K. Friedrichs, and H. Lewy. Über die partiellen differenzengleichungen der mathematischen physik. *Mathematische Annalen*, 100:32–74, 1928. ISSN 0025-5831.

Datawell Waverider, Reference Manual: WR-SG, DWR-MkIII, DWR-G. Datawell BV, 07 2010.

SWAN SCIENTIFIC AND TECHNICAL DOCUMENTATION. Delft University of Technology, March 2009.

K.S. Dijkema, J.H. Bossinade, P. Bouwsema, and R.J. Glopper. Salt marshes in the netherlands wadden sea: Rising high-tide levels and accretion enhancement. In JanJ. Beukema, WimJ. Wolff, and JoopJ.W.M. Brouns, editors, *Expected Effects of Climatic Change on Marine Coastal Ecosystems*, volume 57 of *Developments in Hydrobiology*, pages 173–188. Springer Netherlands, 1990. ISBN 978-94-010-7397-4. doi: 10.1007/978-94-009-2003-3_21.

M.W. Dingemans. *Water Wave Propagation Over Uneven Bottoms: Linear wave propagation*. Advanced series on ocean engineering. World Scientific Pub., 1997. ISBN 9789810239947.

Ernst Dittmer. Schichtenaufbau und entwicklungsgeschichte des dithmarscher alluviums. *Westküste*, 1(2):105–150, 1938.

Jack Dongarra et al. *A Users' Guide to PVM Parallel Virtual Machine*. Oak Ridge National Laboratory, inst-ORNL:adr, July 1991.

Jorge Echeverri. Assessment of performance of the dithmarschen bight model using the rmae parameter on cross section measurements. Master's thesis, Coastal Research Laboratory, Institute of Geosciences, Christian Albrechts Universität,Kiel,Germany, 2004.

Gary D Egbert and Svetlana Y Erofeeva. Efficient inverse modeling of barotropic ocean tides. *Journal of Atmospheric & Oceanic Technology*, 19(2), 2002.

R.W. Eglese. Simulated annealing: A tool for operational research. *European Journal of Operational Research*, 46(3):271 – 281, 1990. ISSN 0377-2217.

Carlos Alejandro Escobar Sierra. *Modelling of sediment dynamics in the Dithmarschen Bight, German North Sea Coast*. PhD thesis, University of Kiel, Kiel, Germany, 2007.

Mario Estévez Báez, Andrés Machado García, and José M. Estévez Carrera. Análisis de correlación y suavizado de espectrogramas. Univer-

sity Lecture, 2 2008. URL http://fbio.uh.cu/ginvest/mesna/vfc_docs/AnalisisDeCorrelacionySuavizadoDelPeriodograma.pdf.

Talal Etri. *Effects of storms on short and medium-term morphodynamics of a tide-dominated coastal region.* PhD thesis, University of Kiel, Kiel, Germany, 2007.

Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *J. Geophys. Res*, 99:10143–10162, 1994.

Jose Manuel Fernandez Jaramillo and Roberto Mayerle. Development of a now-casting modeling system for a study area on the german north sea coast. *Proceedings of the International Conference on Coastal Engineering*, 1(31), 2009.

K Figge. Begleitheft zur karte der sedimentverteilung in der deutschen bucht (karte nr. 2900). *Deutsches Hydrographisches Institut, Hamburg, Germany*, 1981.

Alexander Findlay. *The phase rule and its applications.* Longmans, Green, and CO., London, New York, 1911.

David Flater. A brief introduction to xtide. *Linux J.*, 1996(32es), December 1996. ISSN 1075-3583.

Theodor Friedrichs and Co. Meteorologische Geraete und Systeme GmbH. Englisch web catalog. http://th-friedrichs.de/TH_Friedrichs/PDF/Englisch_Web_Catalog_2004.pdf, 2004.

Helmholtz-Zentrum Geesthacht. Elbe abfluss (runoff) neu-darchau, bilex, water quality, January 2008. URL http://www.coast.gkss.de/staff/kappenberg/.

P. Ghelardoni, G. Gheri, and P. Marzulli. Error estimates for parallel shooting using initial or boundary value methods. *Applied Numerical Mathematics*, 18 (1-3):127–139, 1995. ISSN 0168-9274.

Mark S. Gockenbach. *Partial Differential Equations - Analytical and Numerical Methods.* SIAM, 2002. ISBN 978-0-89871-518-7.

Christopher Grant. Theory of ordinary differential equations. poincaré-bendixson theorem. University Lecture, 12 1999. URL `http://www.math.byu.edu/~grant/courses/m634/f99/lec39.pdf`.

Andrea Hahmann, Dorita Rostkier-Edelstein, Thomas Warner, Francois Vandenberghe, Yubao Liu, R Babarsky, and Scott Swerdlin. A reanalysis system for the generation of mesoscale climatographies. *Journal of Applied Meteorology and Climatology*, 49(5):954–972, 2010.

G. Hartsuiker. Deutsche bucht and dithmarschen bucht, set-up and calibration of tidal flow models. Draft H1821, Delft hydraulics, 1997.

R. Hecht-Nielsen. Theory of the backpropagation neural network. In *Neural Networks, 1989. IJCNN., International Joint Conference on*, pages 593 –605 vol.1, 0-0 1989.

Agnieszka Herman, Ralf Kaiser, and Hanz D. Niemeyer. Modelling of a medium-term dynamics in a shallow tidal sea, based on combined physical and neural network methods. *Ocean Modelling*, 17(4):277 – 299, 2007. ISSN 1463-5003.

P. L. Houtekamer, Herschel L. Mitchell, and L. Mitchell. Data assimilation using an ensemble kalman filter technique, 1998.

E.-J Houwing. Determination of the critical erosion threshold of cohesive sediments on intertidal mudflats along the dutch wadden sea coast. *Estuarine, Coastal and Shelf Science*, 49(4):545 – 555, 1999. ISSN 0272-7714.

WilliamW. Hsieh and AlexJ. Cannon. Towards robust nonlinear multivariate analysis by neural network methods. In ReikV. Donner and SusanaM. Barbosa, editors, *Nonlinear Time Series Analysis in the Geosciences*, volume 112 of *Lecture Notes in Earth Sciences*, pages 97–124. Springer Berlin Heidelberg, 2008. ISBN 978-3-540-78937-6. doi: 10.1007/978-3-540-78938-3_6.

J.W. Kamphuis. *Introduction to coastal engineering and management*. Advanced series on ocean engineering. World Scientific, 2000. ISBN 9789810244170.

H.B. Keller. *Numerical Methods for Two-point Boundary-value Problems*. A Blaisdell Book in Numerical Analysis and Computer Science. Blaisdel, 1968.

S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

D. S. Ko, P. J. Martin, C. D. Rowley, and R. H. Preller. A real-time coastal ocean prediction experiment for MREA04. *Journal of Marine Systems*, 69:17–28, January 2008.

G.J. Komen, L. Cavaleri, K. Hasselmann, S. Hasselmann, and P.A.E.M. Janssen. *Dynamics and Modelling of Ocean Waves*. Cambridge University Press, 1996. ISBN 9780521577816.

Ingo Lange. Der sturm kyrill vom 18. januar 2007, June 2012. URL `http://wettermast-hamburg.zmaw.de/Sturm20070118.htm`.

W. G. Large, J. Morzel, and G. B. Crawford. Accounting for surface wave distortion of the marine wind profile in low-level ocean storms wind measurements. *Journal Of Physical Oceanography*, 25(11):2959–2971, 1995.

P.A. Lessing, L.J. Bernard, C.B.J. Tetreault, and J.N. Chaffin. Use of the automatic identification system (ais) on autonomous weather buoys for maritime domain awareness applications. In *OCEANS 2006*, pages 1 –6, sept. 2006.

H. Luthardt. *Analyse der wassernahen Druck- und Windfelder über der Nordsee aus Routinebeobachtungen*. Hamburger Geophysikalische Einzelschriften: Reihe A, Wissenschaftliche Abhandlungen. G.M.L. Wittenborn Söhne, 1987.

P. Marzulli and G. Gheri. Estimation of the global discretization error in shooting methods for linear boundary value problems. *Journal of Computational and Applied Mathematics*, 28(0):309 – 314, 1989. ISSN 0377-0427.

Roberto Mayerle and Carlos Palacio. Assessment of open sea boundary condition approaches for coastal models. In *Proceedings of the 13th IAHR-APD*

Congress, ADVANCES IN HYDRAULICS AND WATER ENGINEERING, volume 1, pages 898–903, 8 2002.

Roberto Mayerle and Werner Zielke. Promorph - predictions of medium-scale morphodynamics: Project overview and executive summary. *Die Küste*, 69: 1–24, 2005.

C.C. Mei. *The applied dynamics of ocean surface waves*. Advanced series on ocean engineering. World Scientific, 1989. ISBN 9789971507893.

J. Meiss. Dynamical systems. *Scholarpedia*, 2(2):1629, 2007.

John W. Miles. On the generation of surface waves by shear flows. *Journal of Fluid Mechanics*, 3(2):185–204, 1957.

S Müller-Navarra, S. Maßmann, T. Brüning, I. Bork, P. Meier-Moosmann, and K. Stockmann. Hintergrundinformationen zu besonderen ereignissen seit der sturmflutsaison 2009/2010, December 2013. URL `http://www.bsh.de/de/Meeresdaten/Vorhersagen/Sturmfluten/Berichte/index.jsp`.

Sylvin Müller-Navarra and Ingrid Bork. Development of an operational elbe tidal estuary model. In *Proceedings of the International Conference on Coastal Engineering*, 2010.

Sylvin Müller-Navarra and Klaus Knüpffer. *Improvement of water level forecasts for tidal harbours by means of model output statistics (MOS)*. Number 47 in Berichte des BSH. Bundesamt für Seeschifffahrt und Hydrographie (BSH), Hamburg und Rostock, 1 edition, 2010. ISBN 0946-6010.

D.C. Montgomery and G.C. Runger. *Applied Statistics and Probability for Engineers*. John Wiley & Sons, 2010. ISBN 9780470053041.

MPI Forum. MPI: A message passing interface. In *Proceedings of Supercomputing '93*, pages 878–883, Portland, OR, November 1993. IEEE CS Press.

William Jaimes Nelson Fernández and Edmary Altamiranda. Neuro-fuzzy modeling for level prediction for the navigation sector on the magdalena river (colombia). *Journal of Hydroinformatics*, 12(1):36–50, 2010.

H.D. Niemeyer. Seegangsmessungen auf deichvorländer. In Forschungsstelle für Insel-und Küstenschutz (Norderney)., editor, *Jahresbericht 1976*, number 28 in Jahresbericht, pages 113–139. Forschungsstelle f. Insel- u. Küstenschutz, 1977.

J. Nocedal and S.J. Wright. *Numerical optimization*. Springer series in operations research. Springer, 2006. ISBN 9780387303031.

P. Novak. *Hydraulic Modelling: An Introduction : Principles, Methods and Applications*. Taylor & Francis, 2010. ISBN 9780419250104.

Wadden Sea National Park of Lower Saxony. Wadden sea national park of lower saxony, the essentials in brief. `http://www.nationalpark-wattenmeer.niedersachsen.de/master/C23766749_N5994936_L20_D0_I5912119`, 2008.

United Kingdom Met Office. *National Meteorological Library and Archive Fact sheet 6 - The Beaufort Scale*. United Kingdom Met Office, 2010.

OpenMP Architecture Review Board. Openmp application program interface. Specification, OpenMP Architecture Review Board, 2011.

Carlos Palacio, Roberto Mayerle, Mauricio Toro, and Nestor Jiménez. Modelling of flow in a tidal flat area in the south-eastern german bight. *Die Küste*, 69: 141–174, 2005.

S.K. Park and L. Xu. *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*. Springer, 2009. ISBN 9783540710554.

N. A. Phillips. A coordinate system having some special advantages for numerical forecasting. *Journal of Meteorology*, 14(2), 1957.

D. A. Poplawski, S. Pahwa, and J. M. Francioni. Models of parallel program behavior. In Anonymous, editor, *Proceedings of the Fourth Conference on Hypercubes, Concurrent Computers and Applications, 6–8 March 1989, Monterey,*

CA, USA, pages 857–860 (vol. 2), Los Altos, CA, USA, 1989. Golden Gate Enterprises.

Gatot Haryo Pramono. *The study of bedforms and equivalent roughness in the central Dithmarschen Bight.* PhD thesis, University of Kiel, Kiel, Germany, 2004.

William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing.* Cambridge University Press, New York, NY, USA, 3 edition, 2007. ISBN 0521880688, 9780521880688.

Maryam Rahbani. *Numerical modelling of the coastal processes in Dithmarschen Bight incorporating field data.* PhD thesis, University of Kiel, Kiel, Germany, 2011.

Gabriele Rampanelli, Dino Zardi, and Richard Rotunno. Mechanisms of Up-Valley Winds. *Journal of the Atmospheric Sciences*, 61(24):3097–3111, 2004.

H-C Reimers. Sedimentverteilung und benthosverbreitung in den watten der dithmarscher bucht als indikator fur morphodynamische veranderungen abschlussbericht zum forschungsvorhaben sedimorv im gkss-hochschulprogramm. *GKSS FORSCHUNGSZENTRUM GEESTHACHT GMBH-PUBLICATIONS-GKSS*, .(18), 2003.

K. Reise. High abundance of small zoobenthos around biogenic structures in tidal sediments of the wadden sea. *Helgoländer Meeresuntersuchungen*, 34(4):413–425, 1981. ISSN 0174-3597. doi: 10.1007/BF01995914.

American Radio Relay. *The ARRL Handbook for Radio Communications 2011.* Arrl Handbook for Radio Communications. Amer Radio Relay League, 2010. ISBN 9780872590953.

Martin Riedmiller. Rprop - description and implementation details. Technical

report, Institut für Logik, Komplexität und Deduktionsysteme, University of Karlsruhe, W-76128,University of Karlsruhe, January 1994.

L.C. Rijn. *Principles of sediment transport in rivers, estuaries, and coastal seas.* Aqua Publications, 1993. ISBN 9789080035621.

Yoshi Sasaki. A fundamental study of the numerical prediction based on the variational principle. *Journal of the Meteorological Society of Japan. Ser. II*, 33: 262–275, 1955.

Paul Shureman. *Manual of Harmonic Analysis and Prediction of Tides.* Number 98 in Special Publication. Government printing office, Washington, USA, 1958.

R. Stadelmann. *Den Fluten Grenzen setzen: Schleswig-Holstein: Küstenschutz Westküste und Elbe. Band 1: Nordfriesland.* Den Fluten Grenzen setzen: Schleswig-Holsteins Küstenschutz, Westküste und Elbe. Husum, 2008. ISBN 9783898763127.

S.H. Strogatz. *Nonlinear Dynamics And Chaos.* Sarat Book House, 1994. ISBN 9788187169857.

H.R. Taylor. *Data Acquisition for Sensor Systems.* Chapman & Hall, New York, NY, USA, 1 edition, 1997. ISBN 1441947299, 978-1441947291.

Brunello Tirozzi, Silvia Puca, Stefano Pittalis, Antonello Bruschi, Sara Morucci, Enrico Ferraro, and Corsini Stefano. *Neural Networks and Sea Time Series: Reconstruction and Extreme-Event Analysis.* Birkhäuser, 2006.

Fernando Toro, Roberto Mayerle, and Jort Wilkens. Patterns of hydrodynamics in a tide-dominated coastal area in the south-eastern german bight. *Die Küste*, 69:25–62, 2005.

J. S.. Turner. *Buoyancy Effects in Fluids.* Cambridge University Press, 1973.

UNESCOPRESS. Wadden sea, italy's dolomites and philippines's tubbataha reefs natural park inscribed on unesco's world heritage list. http://portal.unesco.org/en/ev.php-URL_ID=46001&URL_DO=DO_TOPIC&URL_SECTION=201.html, 2009.

L. Vandenbulcke, A. Barth, M. Rixen, A. Alvera-Azcárate, Z. Ben Bouallegue, and J. M. Beckers. Study of the combined effects of data assimilation and grid nesting in ocean models - application to the gulf of lions. *Ocean Science Discussions*, 3(3):291–318, 2006. ISSN 1812-0806.

Klaus Vanselow. http://www.uni-kiel.de/ftzwest/ag2/projekte/wetter-e.shtml, 2008.

S Vela-Diez. *Sediment Mapping of the Tidal Flat Channels off Buesum*. PhD thesis, Master Thesis. Coastal Research Laboratory, University of Kiel, Germany, 2001.

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

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

G. Verreet, J. van Goethem, W. Viaene, and et al. Relations between physico-chemical and rheological properties of fine-grained muds. In *Proceedings of the Third International Symposium on River Sedimentation*, volume 3, pages 1637–1646, 1986.

W.W.S. Wei. *Time series analysis: univariate and multivariate methods*. The advanced book program. Addison-Wesley Pub., 1990. ISBN 9780201159110.

M. Wenzel and J. Schröter. Reconstruction of regional mean sea level anomalies from tide gauges using neural networks. *J. Geophys. Res.*, 115:C08013, 2010.

G.B. Whitham. *Linear and nonlinear waves*. Pure and applied mathematics. Wiley, 1974. ISBN 9780471940906.

Peter Wieland. Untersuchung zur geomorphologischen entwicklungstendenz des aussensandes blauort. *Die Küste*, 23:122–149, 1972.

User Manual Delt3D-FLOW. WL Delft Hydraulics, March 2003a.

User Manual Delt3D-WAVE. WL Delft Hydraulics, March 2003b.

Erich Wohlenberg. Die wattenmeer-lebensgemeinschaften im königshafen von sylt. *Helgoländer wissenschaftliche Meeresuntersuchungen*, 1(1):1–92, 1937. ISSN 0017-9957. doi: 10.1007/BF02285337.