UNESCO-IHE
INSTITUTE FOR WATER EDUCATION

Process-based and Surrogate Modelling of Fine Sediment Transport in the Dutch Coastal Zone

Chu Kai

MSc Thesis (WSE-HI.09-09)
April 2009

UNESCO-IHE
Institute for Water Education

Process-based and Surrogate Modelling of Fine Sediment Transport in the Dutch Coastal Zone

Master of Science Thesis
by
Chu Kai

Supervisors
Dr. B. Bhattacharya (UNESCO-IHE)
Prof. D. P. Solomatine (UNESCO-IHE)
Dr. T. Van Kessel (Deltares)

Examination committee
Prof. D. P. Solomatine (UNESCO-IHE), Chairman
Dr. B. Bhattacharya (UNESCO-IHE)
Dr. T. Van Kessel (Deltares)
Prof. Yiqing Guan (Hohai University)

This research is done for the partial fulfilment of requirements for the Master of Science degree at the UNESCO-IHE Institute for Water Education, Delft, the Netherlands

Delft
April 2009

To my mother…

Abstract

Sediments, particularly fine sediments, are of great importance to coastal engineering and environmental issues. However, sediment transport processes are usually not easy to simulate by models largely because of the limited knowledge of describing the nature behaviour in precise mathematical terms. Process-based models, such as Delft3D developed by Deltares, have been proved to be useful in solving this problem. However, due to the complex nature in the real word, simulation is often quite expensive (time consuming), which makes it inconvenient for some cases. Data-driven models (DDM) have been shown to be successful in solving sediment transport problems by many previous researches. However, model results are sometimes difficult to interpret due to the 'black-box' nature.

A current trend in modelling is to use a hybrid approach, combining both advantages of process-based and DDM, where different components complement each other. In order to improve the prediction of fine sediment processes in the Dutch coastal zone, a surrogate modelling was built in this study. Surrogate model is essentially a 'model of the model' instead of a model of a nature system. The concept of surrogate modelling in this study is using data driven techniques to approximate the process-based model and further to be used as a complement of process-based models for future prediction of SPM concentration. Artificial neural network (ANN) is applied to build the surrogate modelling with output data from the Southern North Sea model. Model results showed a strong possibility of using surrogate modelling in prediction of SPM concentration.

Parsimonious models are attractive for research purposes because they are transparent, requiring less computation time and their results are easy to interpret. The purpose of using parsimonious models in research is trying to seek the most essential character of nature processes, which is very useful for understanding the nature processes and for decision making. Linear regression method was applied to build a parsimonious model in this study. The main idea of parsimonious model is building a model with least variables. The possibility of applying parsimonious model for SPM prediction to the Southern North Sea area was explored. Model results showed the prediction of SPM for a single storm is quite well at IJmuiden but the model performance decreased at locations further offshore. The applicability of parsimonious model for predicting SPM in the Southern North Sea needs to be investigated further at more locations.

Keywords: surrogate model, parsimonious model, fine sediment transport, hybrid approach, DDM, linear regression, ANN, Southern North Sea model, Dutch coastal zones, Delft3D.

Acknowledgements

My sincere gratitude goes to my supervisors, Dr. B.Bhattacharya and Prof. D.P.Solomatine for their daily advice and thoughtful guidance without which it would have been impossible to complete this research. Some invaluable discussions with them have contributed significantly to the improvement of my work.

Part of this research was carried out in the framework of the project "Morphodynamics of the North Sea and Coastal Defence" of the Delft Cluster research programme supported by the Dutch government.

I want to thank Dr. T. Van Kessel from Delftares, for providing the model and data I used in this study and providing many valuable suggestions and recommendations to me.

Many thanks to Mick van Der Wegen and Durga Lal Shrestha for their warm hearted technical and software support.

Many thanks go to Phd researcher Ye Qinghua for sharing many experience and ideas which are of great help for my thesis going.

I want to express my sincere thanks to Hohai University and UNESCO-IHE for providing me such a unique opportunity to study in a multi-cultural and enjoyable environment.

Many thanks go to colleagues of 2007-2009 Hydroinformatics group with whom we shared fun and pain all together in the past years.

Finally I wish to thank my family for your endless support and love.

Table of Contents

Chapter 1 Introduction

1.1 Background

Coastal zones which are known as the interface between continents and oceans are vital and important to human beings because a majority of the world's population live in such zones (Nelson, 2007). Coastal systems are among the most dynamic and energetic environments on earth and they are continuously changing because of the dynamic interaction between the oceans and the land. Dronkers (2005) described coasts as multiform, infinitely complex, quasi-fractal, always changing and unpredictable. Sediment process, especially fine sediment transportation is a very complicated feature in many coastal zones as it is affected by physical dynamics, tide, wave, wind and their mutual interactions. Waves and winds along the coast are both eroding rocks and depositing sediments continuously, and the rates of erosion and deposition vary considerably from day to day. Tidal currents also have great effects on sediment transportation.

Sedimentation causes many problems in coastal systems. Fine suspended sediment affects local morphology in coastal rivers, estuaries and shelves environments. Fluid mud, a high concentration aqueous suspension of fine sediment, impedes navigation, reduces water quality and causes environmental damages (Sowed, 2008). So it is crucial and of great interests for coastal engineers and water managing authorities to improve understanding of the underlying sedimentation processes and then further to carry out plans for water management, coastal protection, channel maintenance, land reclamation and dredging of deepwater navigational channels, etc.

Along the Dutch Coast, a lot of efforts have been made to improve the prediction and understanding of sediment transport processes. Process-based models such as SOBEK and Delft3D of Deltares have been proved to be useful in simulation of 2D/3D sediment processes in the Dutch coastal areas. Delft3D solves shallow water equations and transport equations for salinity and suspended particulate matter (SPM) numerically by using a finite-difference scheme. Delft3D was used to build both large-scale and small-scale models to predict SPM concentrations and siltation rates in the Dutch coastal zones. For example, Van Kessel et al. (2007) built model of the Southern North Sea and Li (2007) built a local model focused on the mouth of River Rhine. The results from both models were satisfactory. However, simulating sediment transportations with process-based models is often quite time consuming, which restricts process-based model for widely applications. More detailed information will be introduced in Chapter 1.3.

Data-driven models (DDM) have also been used in simulation of sediment processes (Bhattacharya et al., 2006). They are based on limited knowledge of physical processes and rely on the data describing input and output characteristics. Data driven techniques are used in building models to solve mathematical equations from the analysis of concurrent input and output time series instead of the analysis of physical processes. Solomatine and Ostfeld (2008) described that the model works on the basis of connections between the system state variables (input, internal and output variables) without considering too much on assumptions about the natural processes of the system.

Artificial neural network (ANN) now has become a mainstream technology for DDM. ANN is a computing paradigm designed to imitate functioning of neurons in human's brain. ANN has been successfully used in many water engineering problems including sediment transport. A European project, H-SENSE, used ANN to predict locations of accretion rates in a harbour basin (Rosenbaum, 2000). However, because of its "black-box" nature, some limitations of DDM were reported by many researchers.

1.2 General description of the physical system

The Netherlands is situated in the deltas of the rivers Rhine, Meuse and Scheldt. Currents, waves, wind, sediment deposits from the rivers and human made structures have resulted in the present geomorphologic features of the Dutch coast. The location of the physical system is shown in Figure 1.1.

Fine suspended sediments play an important role in the morphology in the channels and coastal zones. The fine suspended sediments in the North Sea are considered to be coming mainly from the coastal erosion along the French and British cliff coasts along the Strait of Dover and the inflow from the English Channel and the Atlantic Ocean. In addition, dumped sediments from dredging, riverine inputs (e.g. Rhine River) also contribute to sources of the sediments in the North Sea.

The patterns and paths of the suspended sediment transport are largely depended on the water movement. Fine sediments are transported form south to north along French-Belgian-Dutch coastlines into the German Bight by residual currents. On the way of its transportation northward, the complex coastal hydrodynamics, consisting of gyres, divergences or convergence's of currents, mixing of the freshwater, or geological traps make the fine sediments very easy to get deposited in several areas along the Dutch coast, such as Haringvliet mouth, Maasmond, Wadden Sea, etc. Sistermans and Nieuwenhuis (2002) reported that approximately 12 million m^3 of sands are transferred annually from the North Sea to the Wadden Sea.

Sediment transportation and re-suspension of sediment in the Dutch coast are mainly governed by the dynamics of fresh water discharge and meteorological forcing such as waves, wind etc. The processes have strong stochastic characteristics due to the meteorological effects, but also deterministic components due to tides, etc. The tides, river discharge, wind and waves play a determinant role in fine sediment transportation.

The tides in the North Sea are caused by the tidal wave from the North Atlantic Ocean. Ebb and flood alternate in a cycle of 12.5 hours, which is characterized as semidiurnal tide. The tidal wave flows around Scotland and then counter-clockwise along the English coast, reaching the German Bight 12 hours after arriving in Scotland. Depth-averaged residual flow generated by the tide along the Dutch coast has a velocity ranging from 0 to 10 cm/s depending on wind and river outflow.

River discharge also has a great influence on the sediment dynamics in the North Sea. Freshwater discharge generates salinity-induced density gradients. Combined with the effects of density-driven currents and Coriolis force, a relatively narrow band of fresh-brackish water is formed along the Dutch coast. For instance, the Rhine plume with a width of 20 to 30 km stretches along the Dutch coast. Fine sediments accumulate more near the coast than the further offshore area due to the density-driven currents.

Figure 1.1 Location of physical system in the North Sea

(from http://www.worldatlas.com/aatlas/infopage/northsea.htm)

Winds along the Dutch coast are mainly come from the North Sea. The prevailing wind direction is southwest (23%), followed by west (16%), east (13%) and northwest (12%) (Stolk, 1989). The wind from southwest has an annual mean speed of around 7-9 m/s. The storm winds causing the largest wind set-up along the coast are coming from northwest (Van Rijn et al., 2002).

The waves within the North Sea can be classified as short waves ($T_s < 10$ s) and long waves or swell ($T_s > 10$ s) where Ts is wave period. The wave climate is dominated by sea waves with a mean annual significant offshore wave height of about 1.1m.

The supply of fine-grained cohesive sediments also has seasonal effects. Fine sediments are buffered in the seabed during calm weather conditions, and are mobilized during storm conditions. During winter times, the suspended sediment concentration is much higher than that in the summer values, which is mainly because more coastal erosion and re-suspension from the sea bed happens in rough weather conditions in winter. This analysis shows that along the Dutch coast, sediments are transported in suspension by tidal and wind-induced currents in the northern direction towards the Dutch coastal zone.

1.3 Problem description

Process-based and DDM can both be used in simulation of sediment processes. However, both approaches have some limitations.

One of the limitations of process-based models is long computing time, which is usually owing to long residence time of fine sediments. Furthermore, a large domain size is often needed in large-scale simulation and fine grid resolution is required to interpret local morphological changes. So it is a dilemma to compromise results accuracy with computation time. One solution is that the meteorological variability is often schematized in a very short time. It is assumed that the meteorological forcing is in the same pattern over the whole simulation period. This is done at the expense of realism and consequently it affects the accuracy of the results. In addition, although the model covers the major features of the natural processes, some processes such as water-bed exchange of sediments is still a challenge for process-based models (Ye, 2006).

The performance of DDM is satisfactory when adequate data is available for calibration. However, due to the 'black-box' nature, See et al. (2007) argued that the results are sometimes difficult to interpret because it does not consider the physical principles and mathematical reasoning too much. Besides, when the system changes (e.g. structural and bathymetry changes), it is not easy to calibrate the model with old data. Re-training the model for changed system configurations is also difficult.

1.4 General description of hybrid modelling

A current trend in modelling is combining both advantages of process-based and DDM where different components complement each other. This is denoted as hybrid modelling. Sowed (2008) used hybrid modelling to simulate sediment transport in the Maasmond area and Lake Victoria, and the results were satisfactory. Process-based models are based on the understanding of underlying mechanisms of the physical processes in a system. DDM are built on the basis of collected data. However, this does not imply that the underlying processes are not always known, but in most cases they are built when such knowledge is absent or disjointed. Solomatine (1996) noted that in many cases, there is an understanding of the modelled processes, but not very detailed to facilitate development of accurate models. Combining the two modelling approaches serves to compliment each other to produce more accurate results.

There are a couple of alternatives for the application of hybrid models to improve model performance. Abebe (2003) used ANN to forecast errors of process-based model. Sowed (2008) focused on using a DDM to generate time varying boundary conditions for process-based model. In this study a so-called surrogate modelling was built to improve predictions of sediment processes. Surrogate model is "model of the model" instead of model of natural systems; such models are also denoted as emulation or meta-models (Soon et al., 2004). The basic principle of this approach is to build a model by using data-driven techniques with the data generated from a process-based model. This surrogate model will be used as a complement of the process-based model for future predictions. Surrogate model reduces the number of simulation runs required in the process-based model considerably, thus making the computation time decreased dramatically. ANN could be applied to build a simple surrogate model and linear regression method could help build a simper parsimonious surrogate model. 'Parsimonious' refers to simplicity in statistics. Therefore a parsimonious model refers to the simplest feasible model with the fewest possible number of variables. Parsimonious modelling aims at achieving maximally simple or compact models as a result of the data analysis process. Parsimonious model is usually used as a heuristic rule to guide scientists in the development of theoretical models. On the one hand, parsimonious model decreases computation time further and parsimony makes results more understandable and interpretable on the other hand.

1.5 Objective

The main objective of this study is to improve the prediction and understanding of fine sediment processes in the Dutch coastal area by using surrogate modelling. The specific objectives are as follows:

1. Test performance of the process-based model in simulating fine sediment processes and analyze the data generated from it.

2. Examine the performance of surrogate modelling in simulating sediment processes in the Dutch coastal areas. Find out the major factor influencing sediments transportations.

3. Examine the performance of parsimonious surrogate model at one location in the North Sea. Investigate the applicability of parsimonious surrogate model to other locations in the Southern North Sea.

4. Propose improvements for future studies based on the results analysis.

1.6 Methodology

1.6.1 Philosophy of research methodology

The strength and shortcomings of both process-based models and DDM have been introduced previously. The necessity of combing the positive aspects of these two modelling approaches emerges.

The principle of this study is combining process-based model with DDM using surrogate modelling. Process-based model solves mathematical equations from the analysis of physical processes of sediment transportation. Outputs form process-based model were used as inputs for building the surrogate model with data-driven techniques. The surrogate modelling was built simply by using ANN (Chapter 3). Appropriate inputs variables were chosen by analyzing the natural processes and outputs from process-based model. Data were manipulated to expose the maximum information to data driven modelling tool. A simpler model using linear regression method which is referred to parsimonious model was also built (Chapter 4).

1.6.2 Research methodology

The specific structure of this research is shown in Figure 1.2.

1. Data collection and building process-based model

Input data such as wind, wave, SPM concentration and tide were gathered from the North Sea. Physical domain data are size, area and bathymetry of the physical system.

Delft3D developed by Deltares was used to build a process-based model which focuses on the Southern North Sea. These tasks have been done by Deltares and outputs from Southern North Sea model were provided by Deltares also.

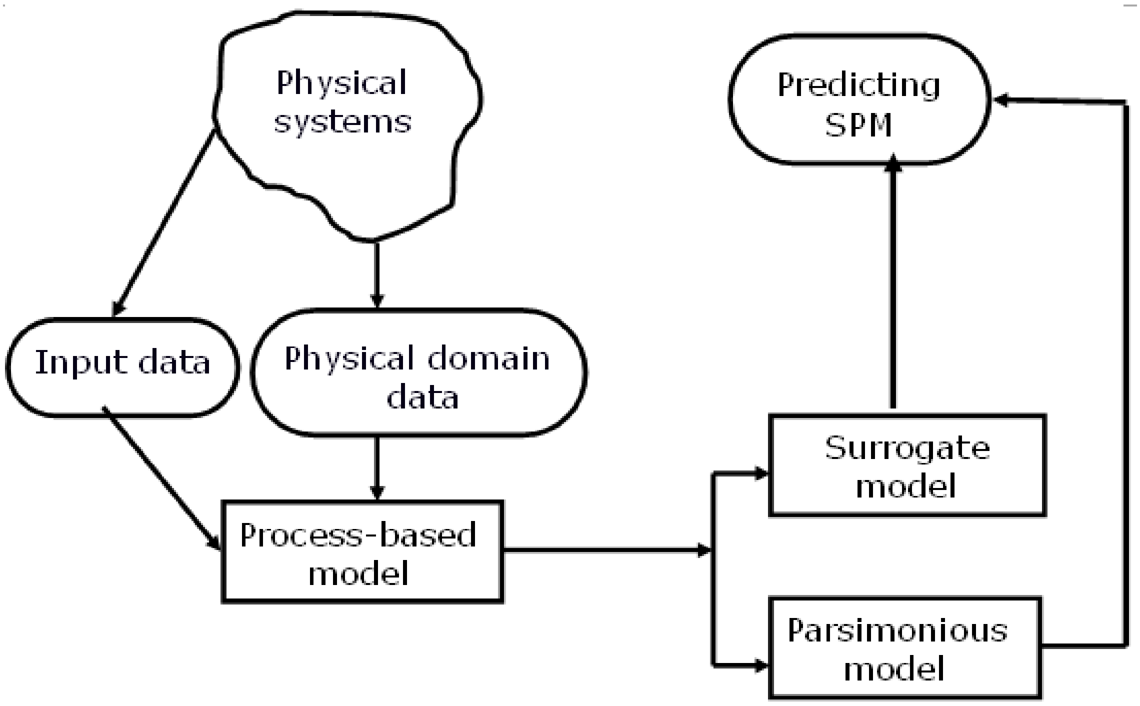

Figure 1.2 Framework of the research methodology

2. Building the surrogate model with ANN.
The output data (total bed shear stress, wave data, SPM concentration) from the Southern North Sea model were used to build a surrogate model using ANN.

3. Building the parsimonious surrogate model.

A linear regression method was used to build the parsimonious model.

4. Conclusions and recommendations.
Based on the analysis of model results, conclusions and recommendations about the fine sediment process in the Dutch coast areas can be drawn.

Chapter 2 Process-based modelling

2.1 Introduction

The importance of understanding fine sediment transportation has been emphasized in the previous chapter. The approaches used to model the sediment dynamics have been introduced and their merits and limitations have also been highlighted. This chapter describes the sediment properties and briefly introduces the process-based model.

2.2 Fine sediments properties

Sediments are classified as cohesive and non-cohesive broadly. Cohesive sediment dynamics are controlled not only by physical forces (e.g. inertia, buoyancy, drag, lift, friction) but also by electrochemical forces. The interparticle bonding forces make small particles stick together and form larger aggregates. Resistance to erosion of cohesive sediment depends on the strength of the cohesive bond binding the particles. Cohesion may far outweigh the influence of the physical characteristics of the individual particles and its behaviour is primarily dependent on the particle size, water chemistry, and sediment mineralogy (Simons and Senturk, 1992).

2.2.1 Sediment size

Sediment size is the most significant property, not only because size is the most readily measured property, but also because other properties, such as shape and specific gravity tend to vary with particle size.

Sediment size can be defined by particle diameter. Van Rijn (1993) listed several definitions of particle diameters:

- Nominal diameter: The nominal diameter is the diameter of a sphere having the same volume as the particle.

- Sieve diameter: The sieve diameter is the diameter of a sphere equal to the length of the side of a square sieve opening through which the given particle will just pass.

- Standard fall diameter: The standard fall diameter is the diameter of a sphere that has a specific gravity of 2.65 and has the same terminal settling velocity as the given particle in quiescent distilled water at a temperature of 24 °C.

According to particle size, sediment particles are classified into six general categories: Clay, Silt, Sand, Gravel, Cobbles and Boulders. Because such classifications are essentially arbitrary, many grading systems are to be found in the engineering and geologic literature. Table 2.1 shows a grade scale proposed by the subcommittee on Sediment Terminology of the American Geophysical Union (Lane, 1947).

Table 2.1 American Geophysical Union Sediment Classification System (adapted from Lane, 1947)

	Sediment Size Range		
Sediment	millimeters	microns	Inches
Very large boulders	4096 - 2048		160-80
Large cobbles	256 - 128		80-40
Medium boulders	1024 - 512		40-20
Small boulders	512 - 256		20-10
Large cobbles	256-128		10-5
Small cobbles	128-64		5-2.5
Very coarse gravel	64-32		2.5-1.3
Coarse gravel	32 - 16		1.3-0.6
Medium gravel	16 - 8		0.6-0.3
Fine gravel	8 - 4		0.3-0.16
Very fine gravel	4 - 2		0.16-0.08
Very coarse sand	2.0 - 1.0	2000-1000	
Coarse sand	1.0 - 0.5	1000-500	
Medium sand	0.5 - 0.25	500-250	
Fine sand	0.25 - 0.125	250-125	
Very fine sand	0.125 - 0.062	125-62	
Coarse silt	0.062 - 0.031	62-31	
Medium silt	0.031 - 0.016	31-16	
Fine silt	0.016 - 0.008	16-8	
Very fine silt	0.008 - 0.004	8-4	
Coarse clay	0.004 - 0.002	4-2	
Medium clay	0.002 - 0.001	2-1	
Fine clay	0.0010 - 0.0005	1.0 - 0.5	
Very fine clay	0.0005 - 0.00024	0.5 - 0.24	

The sediment with a size smaller than 62 micron is regarded as the main element of fine sediment (Ye, 2006). Because of the complicated feature of cohesive sediment, particles with a size between 4 and 62 micron, which is termed as silt will be focused in this study. Such classification is of great help to understand the sediment performance in microcosmic scales.

2.2.2 Particle density

The density of most sediment particles (< 4mm) varies between narrow limits. Since quartz is the predominant natural sediment, the average density can be assumed to be 2650 kg/m^3 (Tchouani, 2004). The specific gravity s is defined as the ratio of sediment density ρ_s and density of water ρ_w:

$$s = \frac{\rho_s}{\rho_w} = 2.65 \tag{2.1}$$

2.2.3 Particle Shape

Particle shape is the second most significant sediment property of natural sediments and can be defined by the shape factor, SF (Schulz et al., 1954).

$$SF = \frac{c}{\sqrt{ab}} \tag{2.2}$$

where a, b, and c are the lengths of the longest axis, the intermediate axis, and the shortest axis, respectively. These axes are the mutually perpendicular axes of the particle. The shape factor for a sphere would be 1.0. Natural sediment typically has a shape factor of about 0.7. Particle shape affects the fall velocity of particles.

2.2.4 Fall velocity

Fall velocity is the average terminal settling velocity of a particle falling alone in quiescent, distilled water of infinite extent. Fall velocity is the most fundamental property governing the motion of the sediment particle in a fluid. It has been shown that the bed configuration in a sand channel may change when the fall velocity of the bed material changes. Fall velocity is a function of the volume, shape and density of the particle and the viscosity and density of the fluid.

2.3 Process-based model and model analysis

Process-based modelling tool Delft3D by Deltares has been used to build both large-scale and small-scale models to simulate sediment processes in the Southern North Sea. Delft3D is a well-known 2D/3D modelling system to investigate hydrodynamics, sediment transport, morphology and water quality for fluvial, estuarine and coastal environments.

A large-scale model is focused on the whole area of the Southern North Sea, which is named as Southern North Sea model in this thesis. It is helpful to simulate the overall flow and wave pattern in macro-scale. According to scale-based theory of Vriend (1991), large-scale model can provide the boundary conditions for small-scaled models. Li (2007) studied on the Massmond area by using Maasmond model, which is a finer grid model localized in the Dutch coast. Detailed information about the Southern North Sea model and Massmond model are discussed in the following sections.

2.3.1 Southern North Sea Model

Southern North Sea model is a curvilinear, boundary fitted grid which contains 8,710 computational elements. It covers about 1,000 km × 800 km area from the south Dover Strait to the north of Scotland and most northern parts of Denmark, each cell has a size about 8 km by 6 km. In order to compromise between the covering area and computation time, the model grid is relatively coarse. The bathymetry used for the model is provided by Deltares, which is shown in Figure 2.1.

A so called “Sigma grid” is used in the vertical direction in the Southern North Sea model, which means the total water depth is separated into a certain number of layers, and each of which has a certain percentage of the water depth. Because of this definition, the same vertical resolution will be available in the entire model domain irrespective of the local water depth. Ten computational layers along the water depth are used in this study and the values used are shown in Table 2.2.

Roelvink et al. (2001) pointed out that the logarithmic distribution can provide relatively high resolution of sediment transport. This makes sure that both effects of wind near the surface and computation of sediment transport near the bed could be taken into better consideration.

Table 2.2 Distribution of layers along the vertical direction in the Southern North Sea model (adapted fromYe, 2006)

Layer number	1	2	3	4	5	6	7	8	9	10
Relative thickness of total depth (%)	4	5.9	8.7	12.7	18.7	18.7	12.7	8.7	5.9	4

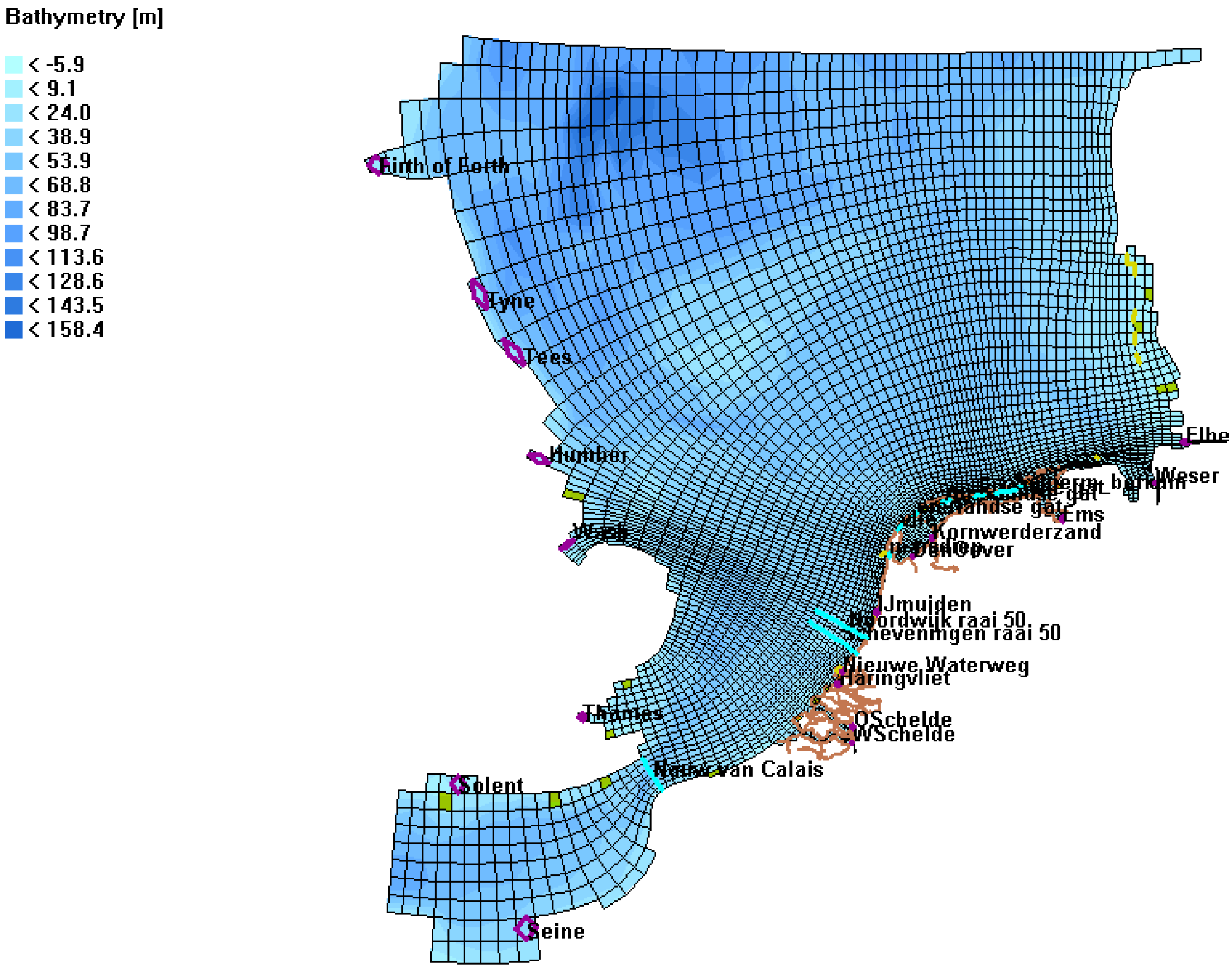

Figure 2.1 Southern North Sea model grid and bathymetry (adapted from Ye, 2006)

2.3.2 Maasmond model

The Massmond model is a finer grid model localized for the Dutch coast. The model was originally developed by Deltares and Li (2007) modified the model to facilitate modelling of sediments in the Maasmond area. The model was set up in two phases. First is the FLOW model and then the WAVE model.

The FLOW model is using finer grid (60km * 30 km) and is used to simulate the hydrodynamics of the system. The red area in Figure 2.2 represents the FLOW grids and the blue line is land boundary. The grid resolution is (*M, N, K*) = (137, 160, 13). The boundary conditions are generated from the Southern North Sea model. In order to improve accuracy of simulation of velocity and sediments, higher vertical resolution is used near to the seabed. The local depth near the bottom was divided into more layers than in the water surface along the vertical direction.

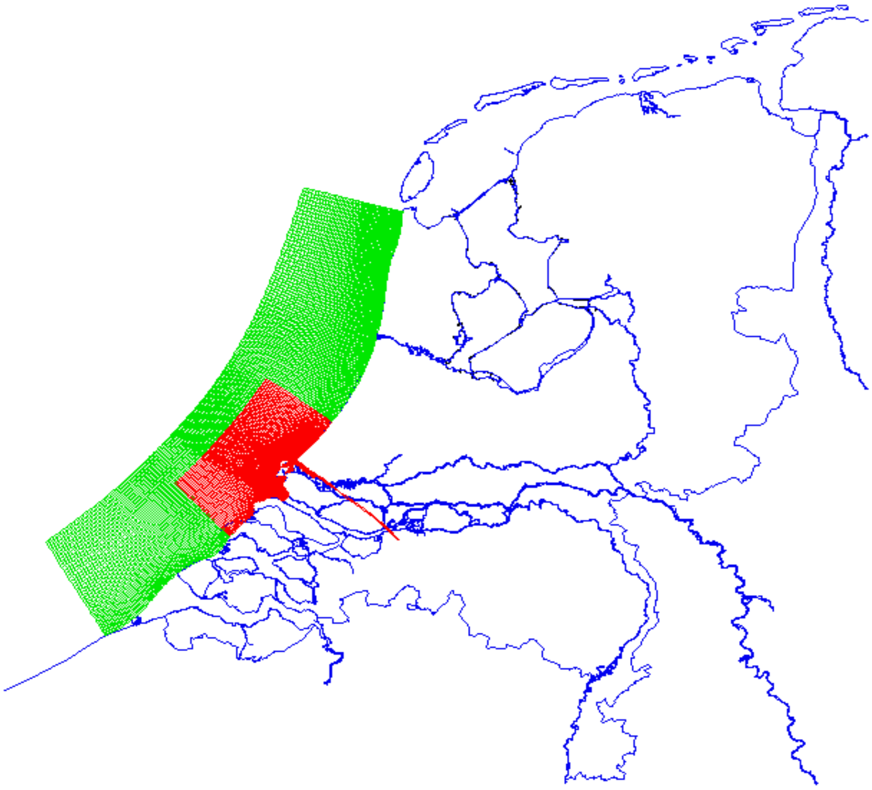

Figure 2.2 Maasmond model grid (adapted from Li, 2007)

The WAVE grid covers larger area than the flow grid and is extended westward to Euro platform. The green area in Figure 2.2 shows the WAVE grid. The original bathymetry provided by Maasmond model is applied to the area covered by FLOW model, and for the extended area, the bathymetry from the Southern North Sea model is used. Observed wave (www.golfklimaat.nl/data) and wind data (www.knmi.nl) from Euro platform during the simulated period are used as the wave boundary.

2.3.3 Flow module

The Delft3D-FLOW which is a hydrodynamic module, simulates two-dimensional (depth-averaged) and three-dimensional unsteady flow and transport phenomena resulting from tidal and meteorological forcing, including the effect of density differences due to a non-uniform temperature and salinity distribution (density-driven flow). The flow model can be used to predict the flow in shallow seas, coastal areas, estuaries, lagoons, rivers and lakes.

The depth-averaged approach is appropriate for vertically homogeneous fluid. Delft3D FLOW is able to run in two-dimensional mode (one computational layer), which corresponds to solve the depth-averaged equations. Three-dimensional modelling is of particular interest in transport problems where the horizontal flow field shows significant variation in the vertical direction. This variation may be generated by wind forcing, bed stress, Coriolis force, and bed topography or density differences.

The model provides two coordinate systems in the vertical direction: Cartesian Z co-ordinated system and sigma σ co-ordinate system. The number of vertical layers varies according to depth but the distance between the layers is fixed in Cartesian coordinated system. However, in σ co-ordinate system, the number of vertical layers decided by user is constant throughout the model domain but the distance between the layers varies with depth. This feature makes it possible for higher resolution of sediment transport near the water surface and bottom.

The governing equations are solved with the finite difference scheme in curvilinear grid system. Continuity equation, momentum equations, transport equation, together with boundary conditions are involved to solve the equations of the system. With the shallow water assumption, vertical acceleration due to buoyancy effects or sudden variations in the bottom topography is not taken into consideration. So the vertical momentum equation reduces to the hydrostatic pressure equation.

2.3.3.1 Hydrostatic pressure assumption

With the shallow water assumption, the vertical momentum equation reduces to the hydrostatic pressure equation. The resulting expression is

$$\frac{\partial P}{\partial \sigma} = -\rho g h \tag{2.3}$$

2.3.3.2 Continuity equation

The depth averaged continuity equation is given by:

$$\frac{\partial \zeta}{\partial t} + \frac{\partial (h\bar{U})}{\partial x} + \frac{\partial (h\bar{V})}{\partial y} = S \tag{2.4}$$

where:

$\bar{U}, \bar{V}$: Depth averaged velocities in the X and Y directions (m/s).

ζ : Water surface elevation above reference datum (m).

h: Total water depth (m).

S: a source or sink term per unit area (discharge, withdrawal of water, evaporation, precipitation, etc).

2.3.3.3 Horizontal momentum equations

The horizontal momentum equations in X and Y directions are:

$$\frac{\partial U}{\partial t} + U\frac{\partial U}{\partial x} + V\frac{\partial U}{\partial y} + \frac{\omega}{h}\frac{\partial U}{\partial \sigma} - fV = -\frac{1}{\rho_0}P_x + F_x + M_x + \frac{1}{h^2}\frac{\partial}{\partial \sigma}(\nu_V \frac{\partial u}{\partial \sigma}) \tag{2.5}$$

$$\frac{\partial V}{\partial t} + U\frac{\partial V}{\partial x} + V\frac{\partial V}{\partial y} + \frac{\omega}{h}\frac{\partial V}{\partial \sigma} - fU = -\frac{1}{\rho_0}P_y + F_y + M_y + \frac{1}{h^2}\frac{\partial}{\partial \sigma}(\nu_V \frac{\partial v}{\partial \sigma}) \tag{2.6}$$

P_x and P_y are horizontal pressure terms, which is given by Boussinesq approximations:

$$\frac{1}{\rho_0}P_x = g\frac{\partial \zeta}{\partial x} + g\frac{h}{\rho_0}\int_{\sigma}^{0}\left(\frac{\partial \rho}{\partial x} + \frac{\partial \sigma'}{\partial x}\frac{\partial \rho}{\partial \sigma'}\right)d\sigma' \tag{2.7}$$

$$\frac{1}{\rho_0}P_y = g\frac{\partial \zeta}{\partial y} + g\frac{h}{\rho_0}\int_{\sigma}^{0}\left(\frac{\partial \rho}{\partial y} + \frac{\partial \sigma'}{\partial y}\frac{\partial \rho}{\partial \sigma'}\right)d\sigma' \tag{2.8}$$

F_x and F_y are horizontal Reynold's stresses terms, which is determined by eddy viscosity concept.

M_x and M_y are the contributions due to external sources or sinks of momentum (by hydraulic structures, discharge or withdrawal of water, wave stress, etc).

ν_H and ν_V are horizontal and vertical kinematic viscosity coefficients (m^2/s).

2.3.3.4 Transport equation

Dissolved substances, salinity and heat in rivers, estuaries, and coastal seas are able to be delivered by flows. These processes can be simulated by the advection-diffusion equation in three co-ordinate directions. Source and sink terms are used to simulate discharges and withdrawals. First-order decay processes are taken into account as well. The transport equation reads:

$$\frac{\partial(hc)}{\partial t} + \frac{\partial(hUc)}{\partial x} + \frac{\partial(hVc)}{\partial y} + \frac{\partial(wc)}{\partial \sigma} = h[\frac{\partial}{\partial x}(D_H\frac{\partial c}{\partial x}) + \frac{\partial}{\partial y}(D_H\frac{\partial c}{\partial y})] + \frac{1}{h}\frac{\partial}{\partial \sigma}(D_V\frac{\partial c}{\partial \sigma}) + hS \tag{2.9}$$

where:

S : Source and sink terms per unit area
D_H , D_V : Horizontal and vertical eddy diffusivities

D_H is defined as:

$$D_H = D_{2D} + D_{3D} + D_H^{back} \tag{2.10}$$

Where D_{2D} is 2D turbulence associated with mixing due to horizontal motions and forcing, D_{3D} is 3D turbulence related to the turbulent eddy viscosity and D_H^{back} is the vertical background diffusion.

The vertical eddy diffusivity is defined as:

$$D_V = \frac{\upsilon_{mol}}{\sigma_{mol}} + \max(D_V^{back}, D_{3D}) \tag{2.11}$$

Where ν_{mol} is the kinematic viscosity of water and σ_{mol} is either the (molecular) Prandtlnumber for heat diffusion or the Schmidt number for diffusion of dissolved matter.

2.3.3.5 Boundary conditions

In the σ co-ordinate system, the bed and the free surface correspond with σ-planes. Therefore, the vertical velocities at these boundaries are simplified as:

$$\omega\big|_{\sigma=-1} = 0 \text{ and } \omega\big|_{\sigma=0} = 0 \tag{2.12}$$

Where σ = -1 is near the bottom and σ = 0 is near the surface.

Friction is applied at bed as follows:

$$\frac{\upsilon_V}{h}\frac{\partial u}{\partial \sigma}\bigg|_{\sigma=-1} = \frac{1}{\rho}\tau_{bx} \text{ and } \frac{\upsilon_V}{h}\frac{\partial v}{\partial \sigma}\bigg|_{\sigma=-1} = \frac{1}{\rho}\tau_{by} \tag{2.13}$$

where τ_{bx} and τ_{by} are bed stress components that include the effect of wave-current interaction.

2.3.4 Wave module

Wind induced waves have big influence on flow and consequently sediment dynamics in large open water bodies (Winterwerp, 2006). In order to simulate the evolution of wind generated wave in coastal zones, the Delft3D-WAVE module is used in which the SWAN (Simulating Waves Nearshore) third generation numerical wave model is implemented (Deltares, 2006).

The purpose of using wave model is two-fold. First of all, the wave model provides wave force for the flow model, which enables the flow model to simulate the wave-driven currents. Secondly, the wave parameters will be provided to the sediment transport model to account for the stirring effect of wave motion on the sediments.

SWAN model simulates the evolution of random, short-crested wind-generated waves in lakes, estuaries, tidal inlets, etc (Deltares, 2006). The numerical scheme for wave propagation is implicit and therefore unconditionally stable at all water depths. To model the energy dissipation in random waves due to depth-induced breaking, a spectral version of the bore-based model of Battjes and Jansen (1978) is used, and to model bottom-induced dissipation, the JONSWAP formulation (Hasselmann et al., 1973) is applied to compute bottom friction. The formulation for wave-induced bottom stress is modelled according to Fredsøe (1984).

In SWAN, the evolution of the wave spectrum is described by the spectral action balance equation given as:

$$\frac{\partial}{\partial t}N + \frac{\partial}{\partial x}c_x N + \frac{\partial}{\partial y}c_y N + \frac{\partial}{\partial \sigma}c_\sigma N + \frac{\partial}{\partial \theta}c_\theta N = \frac{S}{\sigma} \tag{2.14}$$

N represents the density spectrum with parameters σ and θ. The first term in the equation represents the local rate of change of action density in time. The second and third term represent propagation of action in geographical space (with propagation velocities c_x and c_y in x - and y -space, respectively). The fourth term represents shifting of the relative frequency due to variations in depths and currents (with propagation velocity c_σ in σ -space). The fifth term represents depth-induced and current-induced refraction (with propagation velocity c_θ in θ -space). The term S (= S (σ, θ)) at the right-hand side of the action balance equation is the source term in terms of energy density representing the effects of generation, dissipation and non-linear wave-wave interactions. In addition, wave propagation through obstacles and wave-induced set-up of the mean sea surface can be computed in SWAN as well.

2.4 Conclusions

In this chapter, the basic concepts of cohesive sediments have been reviewed. Some key features of flow module and wave module applied in Delft3D are listed as well. It shows that Delft3D can be a generic tool, which covered quite a part, but not all the known processes, to study the cohesive sediments, of which the thorough understanding of the underlying physical processes is not so clear yet.

Moreover, the main disadvantage of this approach is the requirement of long computing time, which restricts its application more widely. FLOW module and SWAN module are both used in modelling SPM concentration by Delft3D. FLOW module simulated unsteady flow and transport phenomena in an aquatic environment while SWAN module simulates wave propagation. Running both modules concurrently takes a lot of time.

In the following studies, a surrogate modelling which combines Southern North Sea model and ANN is built in Chapter 3. A simpler parsimonious model is built as well in Chapter 4.

Chapter 3 Surrogate modelling of fine sediment dynamics

3.1 Introduction

It is shown in previous chapters that both process-based models and DDM can be used in simulating sedimentation processes. Limitations of both approaches are highlighted also. It calls for a hybrid approach to combine both process-based model and data driven techniques. In this chapter, a so-called surrogate modelling is built to predict SPM concentration. Concepts of surrogate models and ANN as data driven techniques are introduced. Input data are processed to expose maximum information for surrogate model as well.

3.2 Concepts of surrogate modelling

Surrogate model which is also denoted as emulation or meta-model, is “model of the model” instead of model of natural systems. Process-based models usually require long computing time for many real world problems because of the natural processes are normally quit complicated. In order to overcome this drawback, surrogate models are introduced to model process-based models instead of the nature systems directly.

Surrogate models have been used for a quite long time (Kleijen, 1975) and are widely used by the engineering design community to reduce the time require for full simulation. Bhattaharya et al. (2003) pointed out that surrogate models are of great use in reducing the computation time. Surrogate models have been successfully applied to model a variety of water and environmental problems. For instance, Riddle et al. (2004) used surrogate models to model reconstruction and interpolation of effluent plume in estuary area. Liong et al. (1995) did calibration of urban drainage model with surrogate models. Surrogate models were also applied to predict numerical geophysical models (Tatang et al., 1997). Abebe and Price (2003) built a surrogate model with ANN to predict errors of a numerical flood routing model and subsequently predictions were updated.

The basic principle of surrogate modelling described by Soon et al. (2004) is shown as follows:

- Choose the relevant process-based model according to the research objective. Southern North Sea model is selected as the process-based model for SPM predictions in this study. Run the Southern North Sea model for a small number of runs; generate outputs and analysis the data.

- Select a surrogate model which can be used to approximate the process-based model. Surrogate model usually refers to a multivariate mathematical function. Variables are chosen based on the analysis of natural processes and outputs from process-based model. ANN is applied to build this surrogate model in this study.

- Run surrogate model, analyze the results and adjust the variables within the surrogate model to improve surrogate model performance.

- Once the adjustments are complete, the surrogate model is used to be a complement of the process-based model for future predictions.

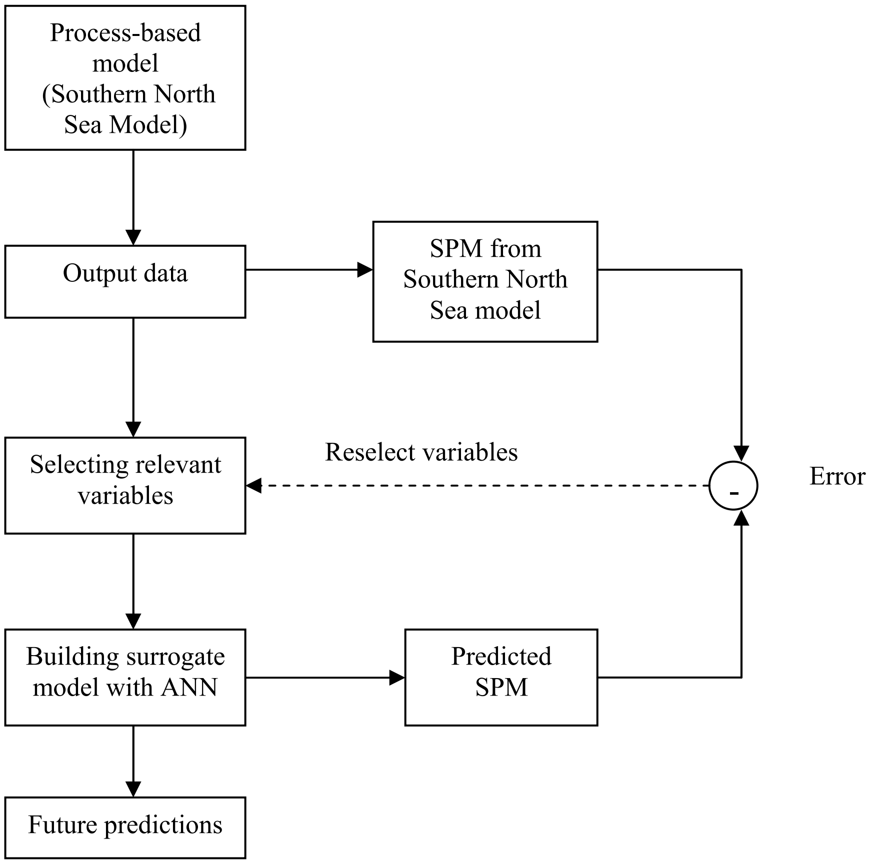

Figure 3.1 Framework of surrogate modelling.

The structure of surrogate modelling in this study is shown in Figure 3.1. Run the Southern North Sea model for a small number of runs and generate output data. Variables for building surrogate model are selected on the basis of analyzing the nature processes and output data from the Southern North Sea model. ANN is applied as data driven technique to build the surrogate model to predict SPM concentration. Compare predicted SPM from surrogate model and the Southern North Sea model. Error signals are sent back to reselect variables to improve the performance of surrogate model. The updated surrogate model will be applied as a complement of the Southern North Sea model for future predictions of SPM concentration.

3.3 Data driven models

DDM is an approach describing physical system on the basis of analyzing data characteristics. DDM works on the basis of connections between the system state variables (input, internal and output variables) without considering too much on nature processes of a system. Techniques used in DDM are borrowed from various fields such as data mining, ANN, fuzzy logic, and machine learning (ML), etc. DDM can thus be considered as an approach that focuses on using the computational intelligence (CI) methods in building models that would complement or replace the process-based models to describe physical systems.

Followed by general principles implemented in modelling, the process of building a

DDM is: study the problem– collect data – select model structure – build the model – test the model and (possibly) iterate (Solomatine, 2008). In DDM, both model parameters and model structure are often subject to optimisation.

One important part of DDM is learning. ML estimates a hitherto unknown mapping between a system's inputs and outputs from the available data (Mitchell, 1997). After the dependency of inputs and outputs is discovered, it can then be used for prediction of future system's outputs from the new input values.

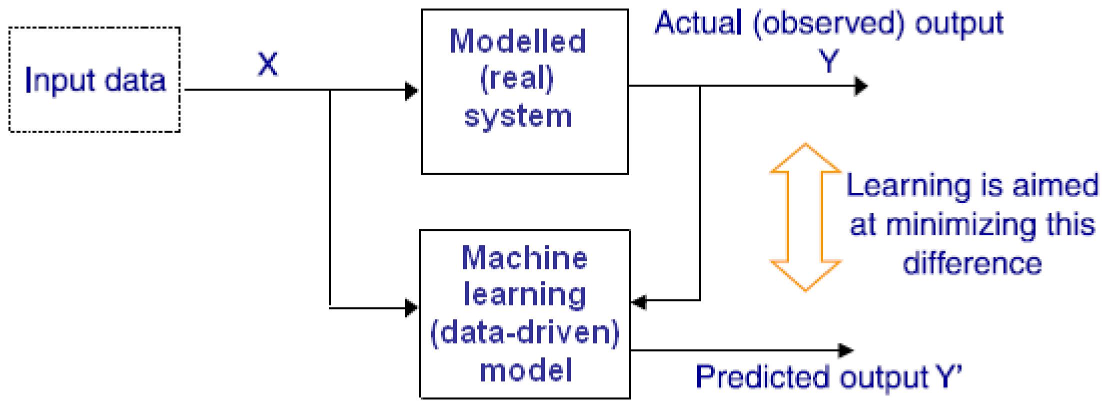

Figure 3.2 Machine learning (adapted from Solomatine, 2006).

Figure 3.2 shows the process of learning in ML. An ML model, on the basis of data samples, tries to identify ("learn") the target function $Y = f\ (X)$ describing the real system behaviors. Learning (or training) here is the process of minimizing the difference between measured data and model output. The data used for training is called training data set.

When new instances, preciously unseen by the model, are fed into the model, the error may be high. This implies that it is necessary to have a separate data set (called cross-validation set), which does not contain instances from the training set and is used to investigate the model error. During the process of training an ML model, the error on the training data will decrease, but error on cross-validation data will first decreases and then start increase, which is under the effect of over-fitting. Over-fitting is resulted from being trained too long that the model tries to follow all data points and actually miss the underlying trend of the data. So the training should be stopped when the error on cross-validation data set starts to increase.

ML model needs to be test before it is put into operation. For this purpose, another data set which is called testing set is used. It allows to see how the model would perform if new data is fed into it.

At the stage of operation, the model, being fed with the new input instances, can perform prediction that is for new input *Xi* generate an output *Yi* which value is hopefully close to what the real system would generate.

3.4 Artificial neural networks

3.4.1 Introduction

ANN is the most widely used method in ML due to its ability to emulate complex processes expressed through sets of input and output observations. ANN imitates functioning of neurons in a human's brain. It consists of an interconnected group of processing elements (artificial neurons) working in unison to solve specific problems.

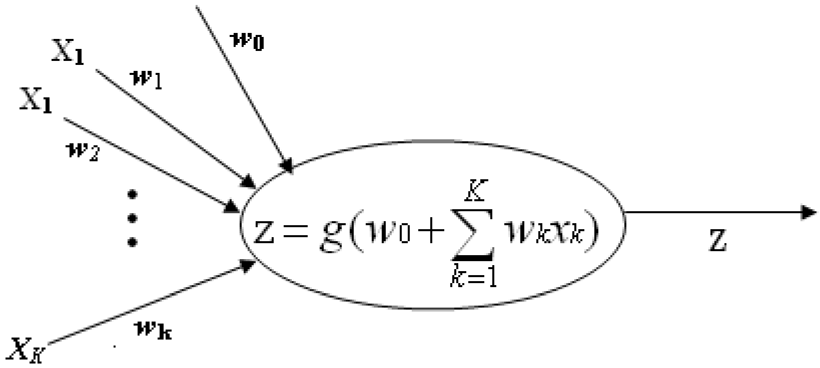

Figure 3.3 An artificial neuron.

As a simplified description of operation of a biological neuron an artificial neuron was proposed by McCulloch and Pitts (1943). Figure 3.3 shows an artificial neuron that receives a series of inputs $x = \{x_1, x_2, ..., x_k, ..., x_K\}$ with associated weights, $w = \{w_1, w_2, ..., w_k, ... w_K\}$. The weighted sum of inputs (u) to the neuron is:

$$u = w_0 + \sum_{k=1}^{K} w_k x_k \tag{3.1}$$

where w_0 is a bias term. The neuron uses a function $g(.)$ to compute the output z as

$$z = g\,(u) \tag{3.2}$$

The function $g(.)$ in Eq. 3.2 is usually known as a transfer function. Several commonly used transfer functions are tangent hyperbolic function, sigmoid function, Gaussian function, linear function, etc. These functions restrict the input signal between certain limits. The choice of a transfer function depends on the task to be learned by the neural network.

Studying from such a simplified artificial neuron, various ANN with complex architectures has appeared over time. ANN is a broad term covering a large variety of network architectures, the most common of which is a multilayer perceptron. The parameters to be found by training are the weight vectors connecting the different nodes of the input, hidden, and output layers of the network by the so-called error-back-propagation method (a specialized version of the gradient-based optimization algorithm) (Haykin, 1999).

During training the values of the parameters (weights) are varied so that the ANN output becomes similar to the measured output on the training data set. ANN has been successfully used in many water engineering problems. However, its use in sediment transport is limited. Nagy et al. (2002) used ANN to estimate the natural sediment discharge in rivers in terms of sediment concentration. Lin and Namin (2005) used ANN to estimate reference concentrations. The possibility of using ANN in building sediment transport models was reported by Bhattacharya et al (2004, 2005).

3.4.2 Supervised learning of ANN

Supervised learning incorporates an external 'teacher', so that each output unit is told what its desired response to input signals ought to be. Paradigms of supervised learning include pattern recognition, classification and regression or function approximation. An important issue concerning supervised learning is the problem of error convergence, i.e. the minimisation of error between the desired and computed unit values. The aim is to determine a set of weights which minimises the error. This is achieved by implementing an algorithm known as the back propagation.

The learning process of ANN used in this study is based on the schematisation of multi-layer perceptron (MLP), which is shown in Figure 3.4. MLP is made up of a number of interconnected nodes, arranged into three layers: input, hidden (could be more than one) and output. The lines represent weighted connections between nodes. The input layer does not perform any operation upon the input signal but simply sends x_k to the nodes in the hidden layer. A node simply multiplies input by a set of weights, and transforms the result into an output value linearly or nonlinearly. By adapting its weights, the neural network works toward an optimal solution based on a measurement of its performance. At the beginning of the learning process, the weights on the connections are assigned values randomly.

The back propagation algorithm is used in two modes: mapping and leaning. In the mapping mode, each instance is analysed and the network estimates the output vectors based on the values of the input vectors. For each instance, input node passes a value of an independent variable x_k to all hidden notes. Each node of the hidden layer computes a weighted sum of the input values based on its weights a_{kj} where j = 1, 2,…, J and J is the number of hidden nodes. The weights are determined during the learning mode. From the value of the weighted sum, hidden notes compute an output y_i using a transfer function. Each of the output nodes receives the outputs of hidden nodes y_i, computes a weighted sum of these inputs based on the weights b_{jm} where m=1, 2, … ,M and M is the number of output nodes, and finally determines the output z_m of the mth output node using a transfer function. The output of the mth output node, z_m is the estimated value of the mth dependent variable. The output from the output node is compared with the measured data and the error is used to adjust the connecting weights a and b, and this procedure is called back propagation. The weights a and b together form the parameter vector w.

For multi-layer perceptron, given the input vector $x = (x_1,x_2,\ldots,x_k,\ldots x_K)$, the output from the dinned nodes will be as follows:

$$y_j = g(a_{0j} + \sum_{k=1}^{K} a_{kj} x_k) \tag{3.3}$$

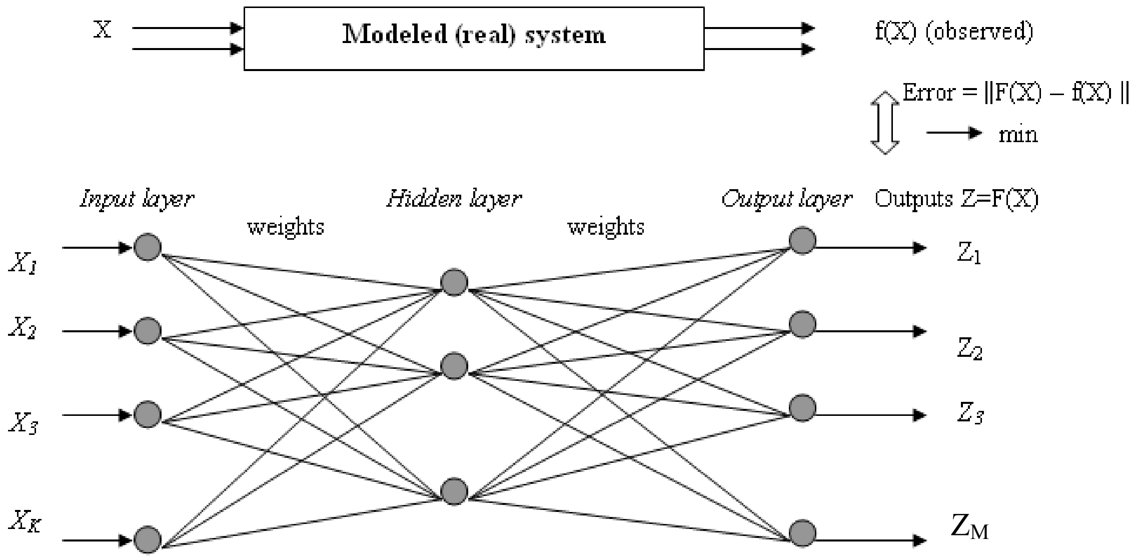

Figure 3.4 Multi-layer perceptron with one hidden layer.

where a_{kj} are the weights of the links from input node k to hidden node j and a_{0j} is bias weight of the hidden node j. The outputs from the hidden nodes would be the inputs to the output nodes. The outputs of the output nodes are calculated as follows:

$$z_m = g(b_{0m} + \sum_{j=1}^{J} b_{jm} y_j) \tag{3.4}$$

where b_{jm} are the weights on the links from hidden node j to output node m and b_{0m} is the bias weight of the output node m. The mean square error is used as a measure of the prediction accuracy and is calculated as:

$$E = \frac{\sum_{i=1}^{N} \sum_{m=1}^{M} (z_{mi} - z_{meas,\,mi})^2}{2NM} \tag{3.5}$$

where N is the number of instances in the dataset; z_{mi} is the mth output for the ith instance and $z_{meas,\,mi}$ is the mth measured output for the ith instance.

In the learning mode, an optimization problem is solved to decrease the mean square error and to find values of a and b that give minimum E. By solving the optimization problem and knowing the slope of the error surface, the weights are adjusted after each iteration. One way of adjusting the weights following the gradient descent rule is as follows:

$$\Delta w(t) = -\eta \frac{\partial E}{\partial w} + \mu \Delta w(t-1) \tag{3.6}$$

where $\Delta w(t)$ is the incremental change in weight w at iteration t; $\Delta w(t-1)$ is the incremental change in weight w at iteration $t\text{-}1$; η is the leaning rate and μ is the momentum value used to give importance to previous weight updates.

3.5 Understanding the process to be modelled

It is of great importance to understand the inter-relationship between the physical processes being modelled. It helps in selecting the best input and output variables. This section therefore reviews the interaction of variables involved in the sediment dynamics.

Suspended particulate matter (SPM) consists of sediment particles suspended in water and SPM concentration is governed by the availability of sediment and transport processes. The availability of sediment is influenced by wind, waves, tides, river discharge, etc. The sediment transport process is mainly affected by local hydro-meteorological conditions.

In large shallow water bodies, the processes of sediment dynamics usually start with wind energy being delivered to the water surface and generate waves (Jin and Ji, 2004). When the wind energy is dissipated into wave motion in the vertical direction, the energy is also transmitted from the surface to the bottom. This process will create wave orbital velocities at the sediment-water interface, which combined with current velocities together to generate bed shear stress. Before sediments are transported, sediment particles need to be firstly picked up and leave from their initial position on the bottom. This could happen only when the bed shear stress is strong enough to lift or drag sediment particles. In other words, total bed shear stress (Tbss) due to waves and currents should exceed a critical value of bed shear stress. Thus, Tbss is a significant variable which should be taken into account in modelling SPM concentration.

During storms in rough weather conditions, large wave energy due to huge waves is transmitted to the water column and produces high Tbss, and subsequently, causes high SPM concentration. However, SPM concentration is also influenced by previous storms. For example, when a storm occurred and lasted for a long time, the concentration of SPM is very high in the water column. So although the wave energy is low in the following calm days, still a high SPM concentration may be observed in the water. Figure 3.5 gives an example of SPM and corresponding significant wave height (Hs) at IJmuiden in December of 2000. The circled area in the left shows two storms happened with high Hs and SPM concentration and the storms lasted for about five days. After storms, as indicated by the area circled on the right hand, still high SPM concentration appeared although Hs decreased a lot.

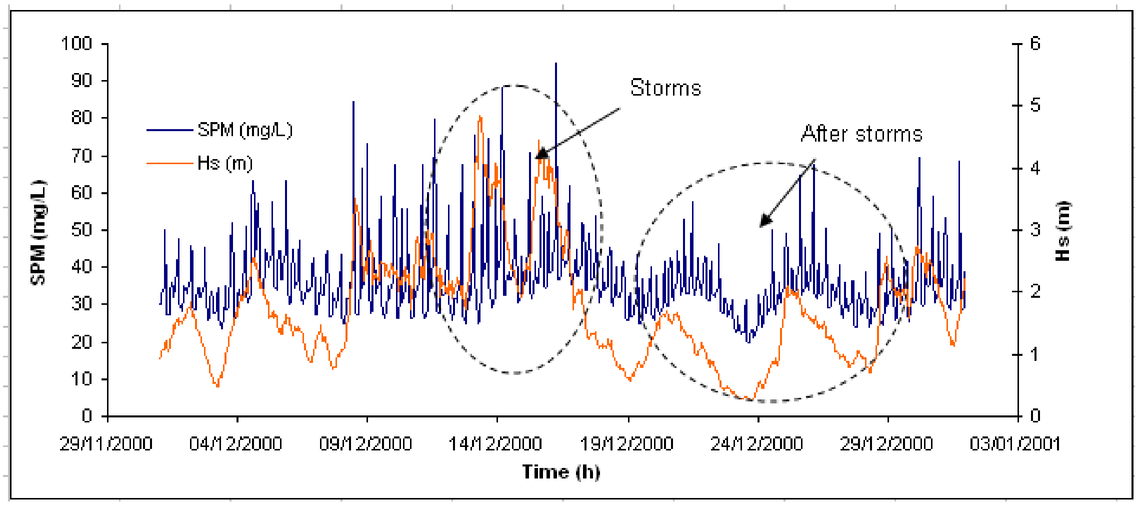

Figure 3.5 SPM and Hs at IJmuiden in December of 2000.

Thus, in order to give a better simulation of SPM concentration, the previous wave energy should be taken into account as well. In this study, it is named as wave energy past (Wep).

3.6 Building the surrogate model with ANN

3.6.1 Introduction

This study focused on the area of IJmudien, the sea port of Amsterdam. Figure 3.6 gives the exact position of IJmudien as an observation point in the Southern North Sea model grid. Water depth in this area is around 7 to 8 meters. Figure 3.7 shows the location of IJmuiden in the Dutch coastal zone. Data used for building the surrogate model are the outputs from Southern North Sea model ran for the year of 2000.

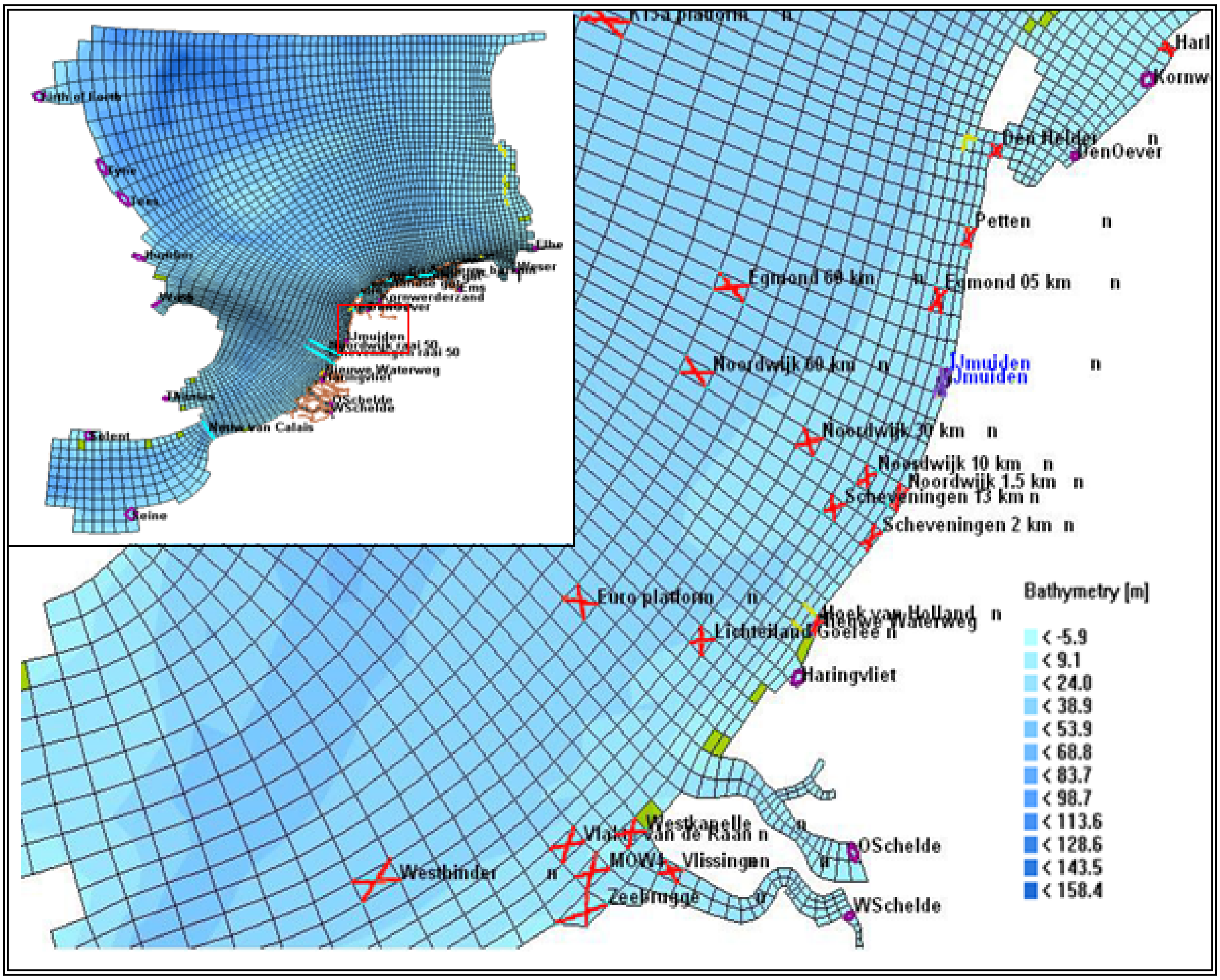

Figure 3.6 IJmuiden in the Southern North Sea model grid.

The tool used to build the ANN model is NeuroSolutions which is well-known in neural network simulation technology. The software is widely used to design, train and deploy neural network (supervised learning and unsupervised learning) models to perform a wide variety of tasks such as, classification, function approximation, multivariate regression and time-series prediction.

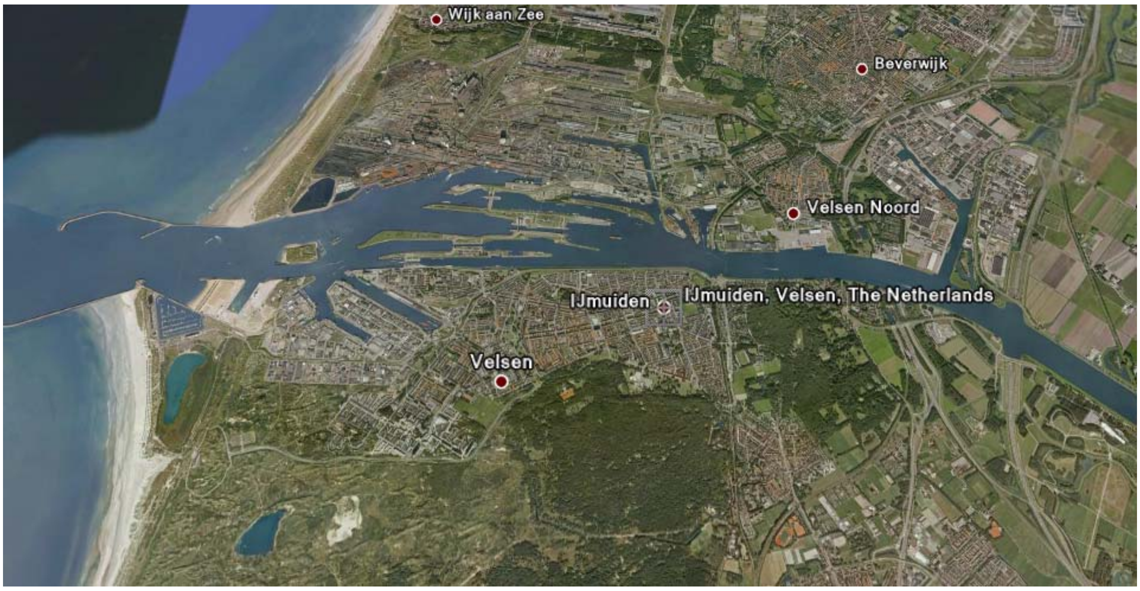

Figure 3.7 Location of IJmuiden (adapted from Google Earth).

3.6.2 Selecting of input variables

With the discussion and understanding of the major processes influencing SPM concentration in Chapter 3.5, there are many parameters could influence SPM such as wind, wave, bed shear stress, previous waves, etc. Some of them may affect SPM largely while some of them may less important. So the most relevant variables need to be selected for building the surrogate models.

Total bed shear stress is selected as the first input variable in this study. Because sediment particles can be picked up and transported only when Tbss exceeds a critical value of bed shear stress. In other words, Tbss is the major driving force for influencing sediment transport processes and changing SPM concentrations. Secondly, from the discussion of sediment processes, wind generates wave and wave energy is transmitted from water surface to the bottom and consequently generate bed shear stress. Bed shear stress can also be generated by currents. Therefore, Tbss actually contains partly information of wind, wave and currents. Moreover, Sowed (2008) used Tbss alone to predict SPM in Maasmond area and results showed the possibility of predicting SPM with Tbss. Tbss can be extracted from the Southern North Sea model directly.

In order to take the effects of previous storms to the SPM concentration, Wep is selected as the second variable for building the ANN model and Wep should be calculated according to Hs.

Wave energy is calculated firstly from the formula:

$$We = \frac{\rho_w g H_s^2}{8} \tag{3.7}$$

where We is wave energy, ρ_w is the density of sea water, Hs is significant wave height. Wep is an integration of wave energy over time. In order to calculate Wep, the first step is to explore how far of the previous wave energy has influence to the SPM concentration now. In this study, we assumed that wave energy within the last two weeks (336 hours) give effects to the current SPM concentration and no influence any more of the wave energy before two weeks ago.

Secondly, we need to find out in which way the Wep affects the SPM concentration. Wep is a weighted sum of wave energy which is shown as:

$$Wep = \sum_{i=1}^{336} We_i w_i \tag{3.8}$$

where We_i is wave energy i hours ago. w_i is the weight of We_i. Weight is higher if the wave is closer to now which means that recent waves have more influence to current SPM. Wave generated longer time ago gives lower effect of current SPM and thus has lower weight. In this study, a linear function is applied to determine the weights. And the plot of this linear function is shown in Figure 3.8.

$$w_i = -\frac{i}{336} + 1 \tag{3.9}$$

where w_i is the ith weight and i is time from 0 to 336.

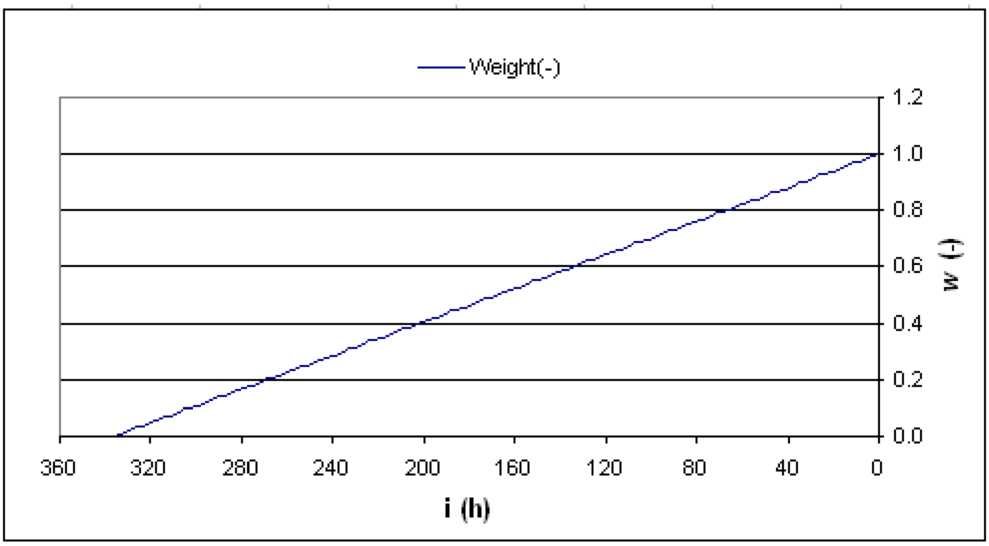

Figure 3.8 A linear function used to determine weights

Therefore, SPM can be expressed as a function of Tbss and Wep:

$$\text{SPM} = f\,(\text{Tbss, Wep}) \tag{3.10}$$

After selecting the input variables for the surrogate models, we are caring about the interdependence between Tbss and Wep and how does this influence the model performance. Figure 3.9 shows the plot of Wep and Tbss at IJmuiden in 2000. Their patterns are similar and the correlation between them is 0.73, which implies that Wep is sharing some information with Tbss. That reason is that Wep is calculated based on wave data and Tbss can also be generated from wave. So sensitivity analysis on the input variables should be done and this is introduced in detail in Chapter 3.6.9.

It needs to be mentioned that, sensitivity analysis on multiple parameters influencing SPM in addition to Tbss and Wep could be carried out. This could help to explore the most significant parameters for building the surrogate models. In this study, only Tbss and Wep are used as the input variables. More input variables could be introduced according to the model performance.

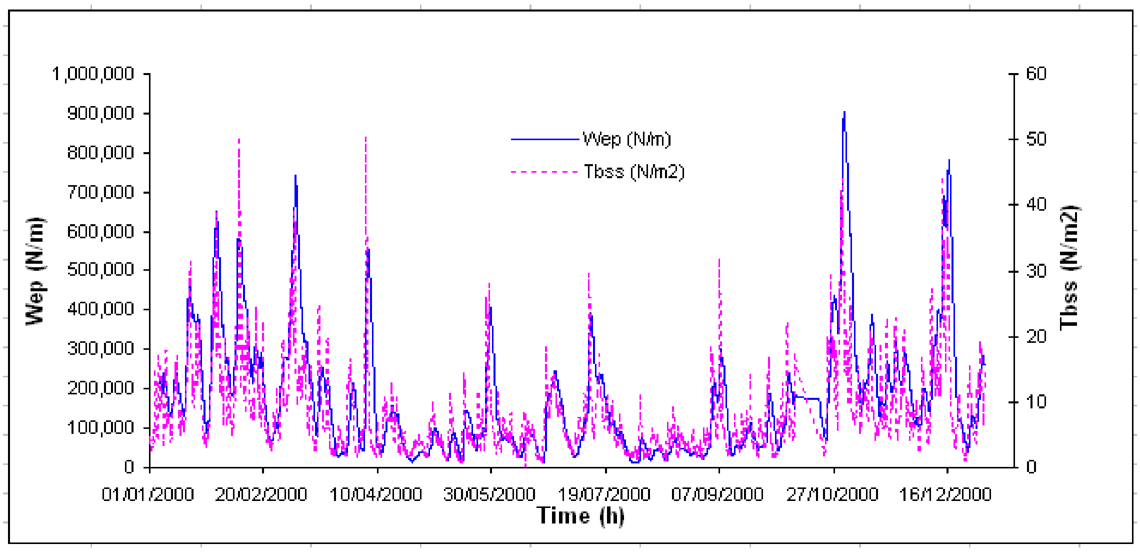

Figure 3.9 Plots of Wep and Tbss at IJmuiden in 2000.

3.6.3 Visual inspection of raw data

After defining the input variables, the next task is to process input data for building the ANN model. Inspection of raw data helps to explore the possibility of adjusting or transforming data. Further more, it establishes reasonable expectation of achieving a solution and reveals the relevance of the data to the task. This was achieved by plotting the identified input variables together with the output variables.

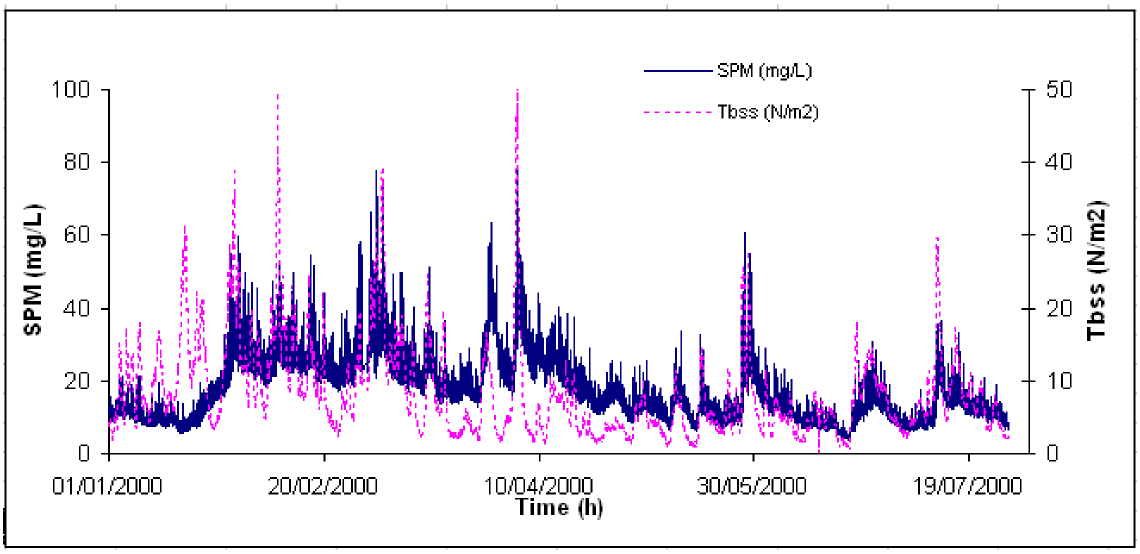

Figure 3.10 Plots of SPM and Tbss with raw data at IJmuiden in 2000.

Figure 3.10 shows the relationship between SPM and Tbss at IJmuiden while Figure 3.11 shows SPM versus Wep. The calculated correlation of SPM and Tbss is 0.43 and 0.49 for SPM and Wep. The plots show that the general pattern of SPM is related to Tbss and Wep. However, still there are many discrepancies in some instances. Tbss and Wep do not change in the same way as SPM does. Furthermore, the SPM does not rise or drop in the same proportion as Tbss and Wep do. These discrepancies could be caused by other processes in predicting SPM in addition to Tbss and Wep. SPM could be diffused along the vertical direction in the water column as well.

As shown in the plots of Figure 3.10 and Figure 3.11, values change quite frequently due to the effect of diurnal tidal movement. The frequent changes of data appear as 'noise' which is often associated with irregular, chaotic changes in the variable. Thus, data need to be smoothed to avoid these 'noise' before are fed into the model.

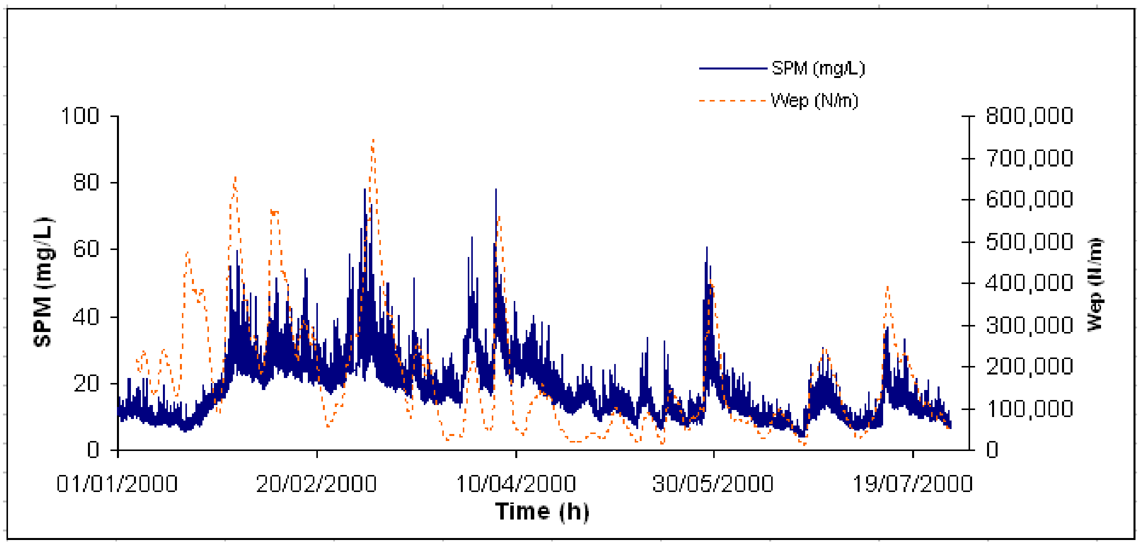

Figure 3.11 Plots of SPM and Wep with raw data at IJmuiden in 2000.

3.6.4 Data filtering

High variance in data will reduce the learning capability of DDM and consequently give unsatisfactory performance. For better performance of the DDM, smoothing data to reduce variance along the data range is necessary (Schalkoff, 1997). The purpose is to reduce the variability and/or remove the noise. By smoothing out short-term fluctuations, longer-term trends could be highlighted. Several smoothing tools can be used such as moving average, peak-valley mean, Gaussian filter etc.

Moving average was used as a smoothing tool in this study. In order to remove the effect of diurnal tidal movement, 24 hours moving average was applied. Figure 3.12 and 3.13 show the plots of SPM with Tbss and Wep after using 24 hours moving average to smooth data and high variances are removed.

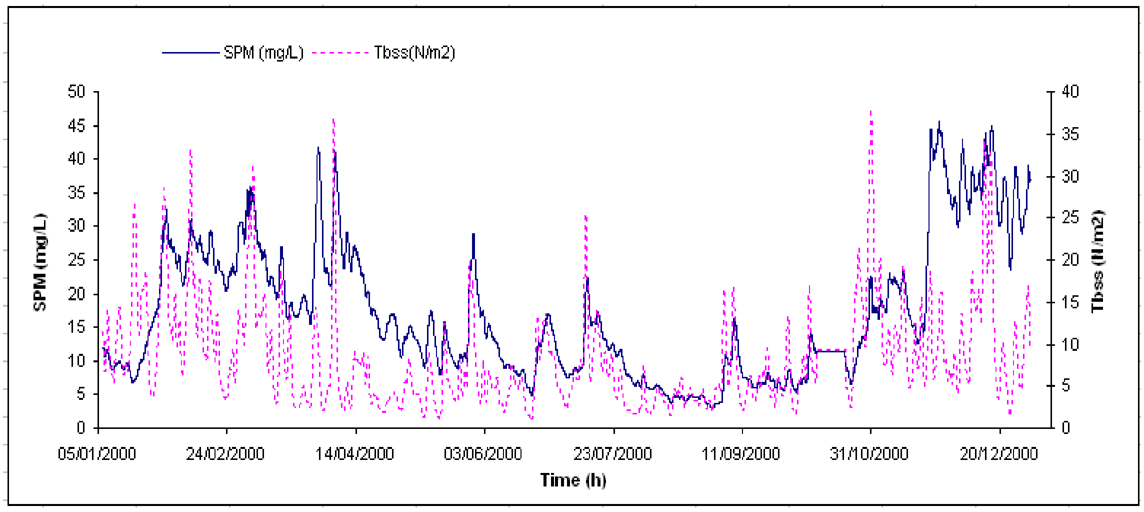

Figure 3.12 Plots of SPM and Tbss with 24 hours moving average at IJmuiden in 2000.

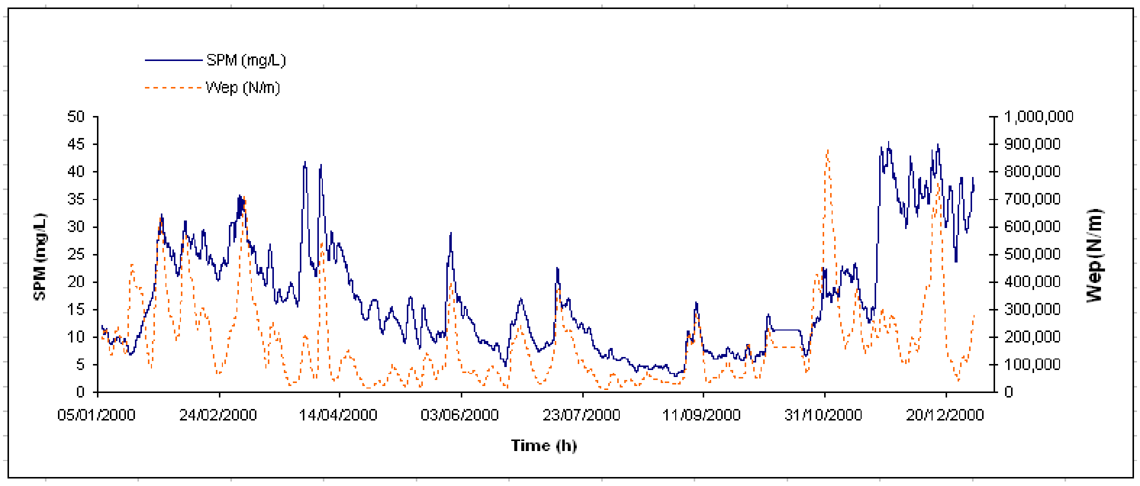

Figure 3.13 Plots of SPM and Wep with 24 hours moving average at IJmuiden in 2000.

3.6.5 Correlation and average mutual information (AMI) analysis

Analysis of correlation plays an important role in data analysis. The measure of how values of one variable changes as values of another variable change is known as correlation. Correlation measures the linear dependences between two variables.

Average mutual information (AMI) provides an elegant way of investigating dependencies and lag effects (Abarbanel, 1996). AMI is a measure of dependencies inside the time series. Formally, the mutual information of two discrete random variables X and Y can be defined as:

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \left(\frac{p(x,y)}{p_1(x)\, p_2(y)} \right) \tag{3.11}$$

where $p(x, y)$ is the joint probability distribution function of X and Y, and $p_1(x)$ and $p_2(y)$ are the marginal probability distribution functions of X and Y respectively. Intuitively, mutual information measures the information that X and Y share: it measures how much knowing one of these variables reduces our uncertainty about the other. For instance, if X and Y are independent, then $p(x,y) = p(x)\, p(y)$, and therefore the value of AMI is 0:

$$\log \left(\frac{p(x,y)}{p(x)\, p(y)} \right) = \log 1 = 0 \tag{3.12}$$

Utility toolpack (provided by Durga Lal Shrestha) is a very convenient tool to do correlation and AMI analysis. In this study, Utility toolpak was used to analyze the correlation and AMI between SPM and Tbss with different time lags. As shown in Figure 3.14, Tbss with nine hours lag (Tbss-9) gives the highest AMI (0.347) with SPM. Tbss with eleven hours lag (Tbss-11) gives the highest correlation (0.508) with SPM. By considering both correlation and AMI analysis, Tbss with lag 9, 10 and 11 hours are selected as the input variables for the ANN model in this study.

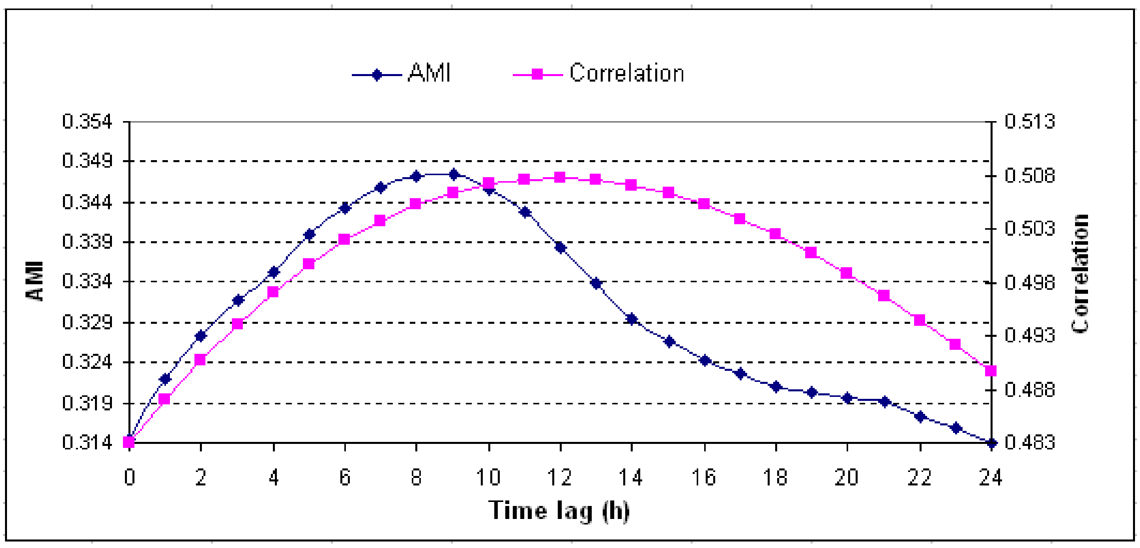

Figure 3.14 AMI and correlation of SPM and Tbss with time lags.

Figure 3.15 shows that Wep without any time lag gives the highest AMI and correlation with SPM. Thus, SPM can be expressed as Eq. 3.13:

$$SPM = f\,(Tbss\text{-}9,\ Tbss\text{-}10,\ Tbss\text{-}11,\ Wep) \qquad (3.13)$$

where Tbss-9, Tbss-10 and Tbss-11 represents Tbss with 9, 10 and 11 hours lag respectively.

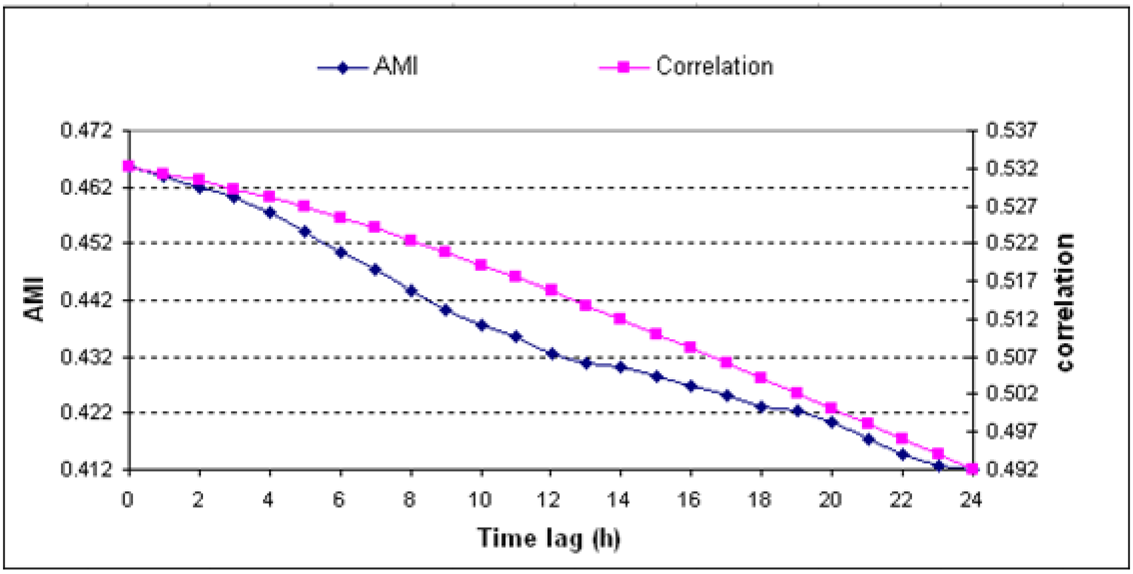

Figure 3.15 AMI and correlation of SPM and Wep with time lags.

3.6.6 Data transformation

Most ML methods perform well when data has a distribution close to normal (Pyle, 1999). However, this is often not the case in practice. It calls for data transformation to adjust the distribution of data closer to normal. Data transformation method can be done in both linear and non-linear way. Non-linear data transformation may distort the natural relationship between the nature variables because the high and low ranges of the variable are squashed during non-linear data transformation. Zero mean and unit variance is a popular and effective linear transformation method:

$$X' = \frac{X - \overline{X}}{\sigma} \tag{3.14}$$

where X is original data, X' is the transformed data, $\overline{X}$ is mean value of X and σ is the standard deviation of X. Input variables are transformed to zero mean and unit variance in this study.

3.6.7 Data partitioning and stormy, non-stormy period definition

Data need to be partitioned into training, cross-validation and testing datasets maintaining a statistical homogeneity (Bhattacharya 2005b). In other words, the mean and standard deviation of each dataset should be closed to each other. The choosing of percentage of training, cross-validation and testing datasets is somewhat arbitrary, and data are partitioned to 60%, 15% and 25% for training, cross-validation and testing data respectively in this study. Table 3.1 gives the statistics of training, cross validation and testing data of SPM. The mean value of testing data is quite far away from mean vaule of training and cross-validation data. Standard deviation of each dataset is also quite different.

As shown in Figure 3.16, SPM value is much higher during winter than in summer. So the mean value of SPM in testing set is higher than the mean of training and cross-validation sets. This phenomenon is due to the seasonal effect. Fine sediments are buffered in the seabed during calm weather conditions in summer, and are mobilized during storm conditions in winter.

Table 3.1 Statistics of training, cross validation and testing datasets

SPM	Training	Cross-validation	Testing
Mean	15.4	14.3	25.0
Std*	11.6	3.9	5.3

* Standard deviation

In order to capture the seasonal effect in the ANN model, it calls for a cut-off of stormy and non-stormy period. Then simulations will be given to stormy and non-stormy period separately. This approach will avoid high variances of mean and standard deviation in training, cross-validation and testing dataset.

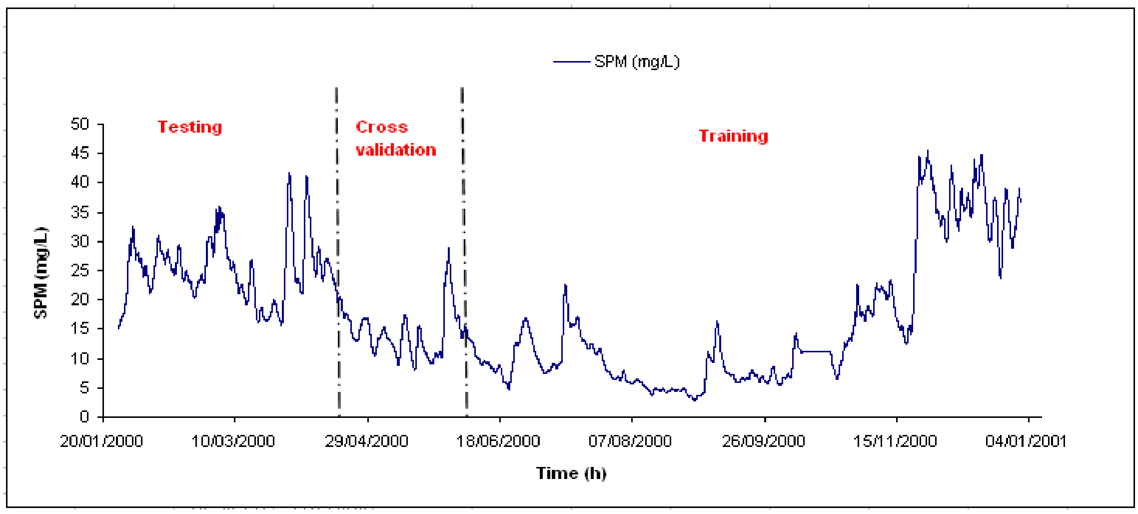

Figure 3.16 SPM for training, cross validation and testing at IJmuiden in 2000.

However, there are no clear rules to define stormy and non-stormy period. Wave height is always large when a storm happens. So the definition of stormy and non-stormy periods is based on the classification of Hs.

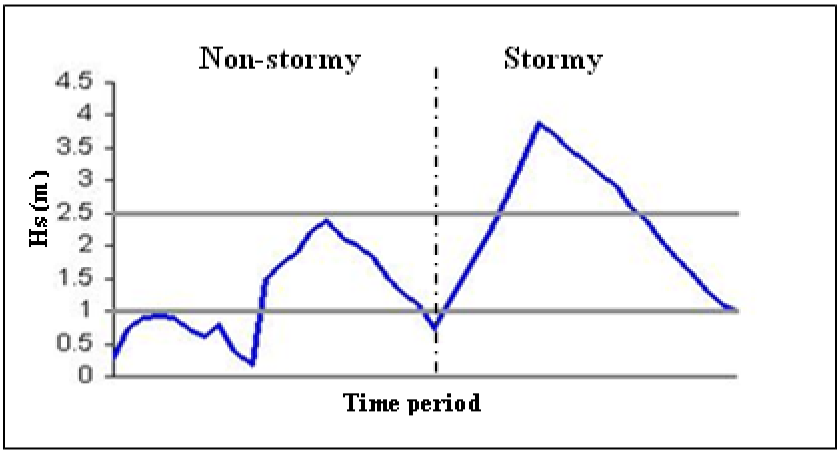

Figure 3.17 Definition of stormy and non-stormy periods

In this study, it is assumed that a storm period is a time period with a starting and ending Hs equal to 1m and peak Hs is at least 2.5m (Figure 3.17) and the remainders belong to non-stormy period. Figure 3.18 shows Hs with corresponding SPM along the original time series. We can find that in most of the instances, the SPM value is high when a defined storm happens. It needs to be pointed out that such kind of definition of storm and non-storm is arbitrary and the values of Hs used to define storm in this study (1m and 2.5m) are flexible. Sensitivity analysis of choosing storms should be investigated as well. Definitions of storms and values of Hs to be used may be different if the results differ from the reality. However, in this study, this definition of stormy and non-stormy periods help to build a more reasonable ANN models and the model performances are improved.

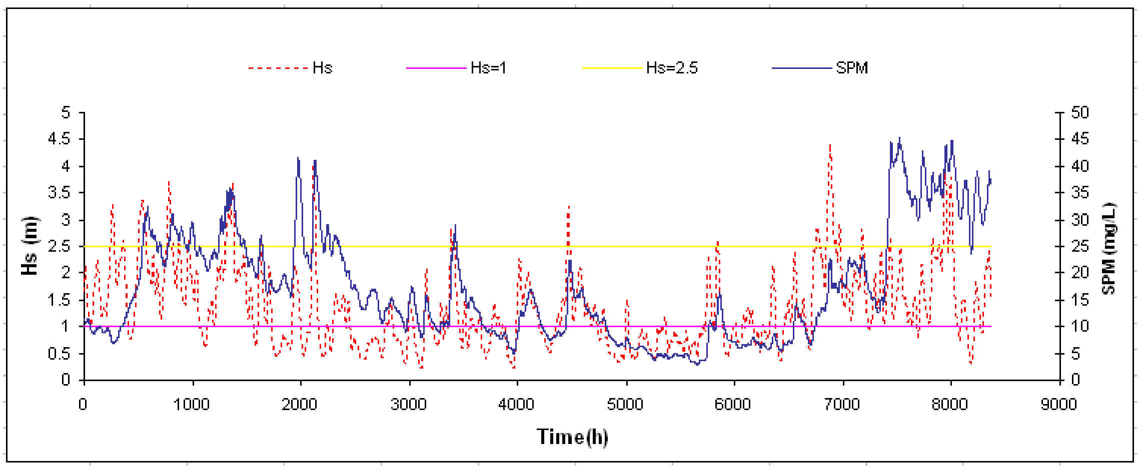

Figure 3.18 Hs and corresponding SPM along the original time series at IJmuiden in 2000.

SPM in both stormy and non-stormy periods are shown in Figure 3.19 and Figure 3.20. In order to remove the influence of pervious storms, the first storm is discarded in this study. After cutting off the storms and non-storms, the SPM value is limited into a narrower range for each model. The variance is not that high which is helpful to choose training, cross-validation and testing datasets and the model accuracy is improved as expected.

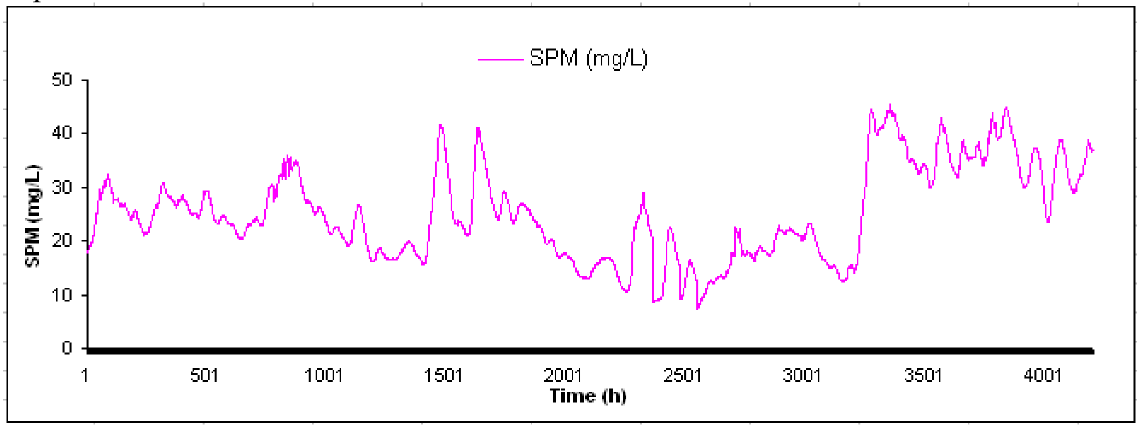

Figure 3.19 SPM in stormy periods at IJmuiden in 2000.

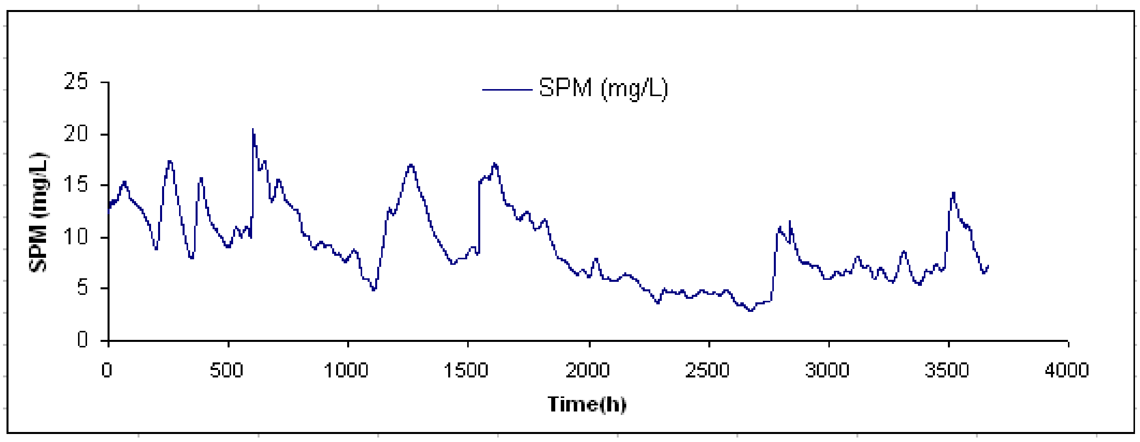

Figure 3.20 SPM in non-stormy periods at IJmuiden in 2000.

3.6.8 Building ANN models and analyzing the results

As mentioned above, input data for building ANN model are smoothed and transformed; stormy and non-stormy periods are defined. In this section, ANN models are built under different scenarios; model performances are presented and results are discussed as well.

ANN model are firstly built using data without any transformation and this model is named as M1 in this study. Then ANN model is built with data transformed to zero mean and unit variance and this model is denoted as M2 in this study. Table 3.2 shows the performance of M1 and M2. By comparing predicted SPM from both models, it shows that M2 gives better performance than M1.

Table 3.2 Comparison of model performance between M1 (ANN model without using data transformation) and M2 (ANN model with data transformed to zero mean and unit variance).

Model Performance	***M1***	***M2***
MSE [-]	62.70	46.98
NMSE [-]	2.20	1.65
MAE [mg/L]	6.71	5.50
Min Abs Error [mg/L]	0.00	0.00
Max Abs Error [mg/L]	23.30	20.80
r [-]	0.66	0.67

where MSE represents mean square error; NMSE is normalized mean square error which is expressed as MSE divided by variance of desired output; MAE is mean absolute error; Min Abs Error and Max Abs Error are the minimum and maximum of the absolute error respectively; r is the correlation coefficient between predicted SPM and desired SPM.

Figure 3.21 plots the predicted SPM from both models. Generally speaking, M2 gives better prediction of SPM than M1. However, still several discrepancies appear in the peaks. The reason could be that training, cross-validation and testing sets do not maintain statistical homogeneity well due to the seasonal effects.

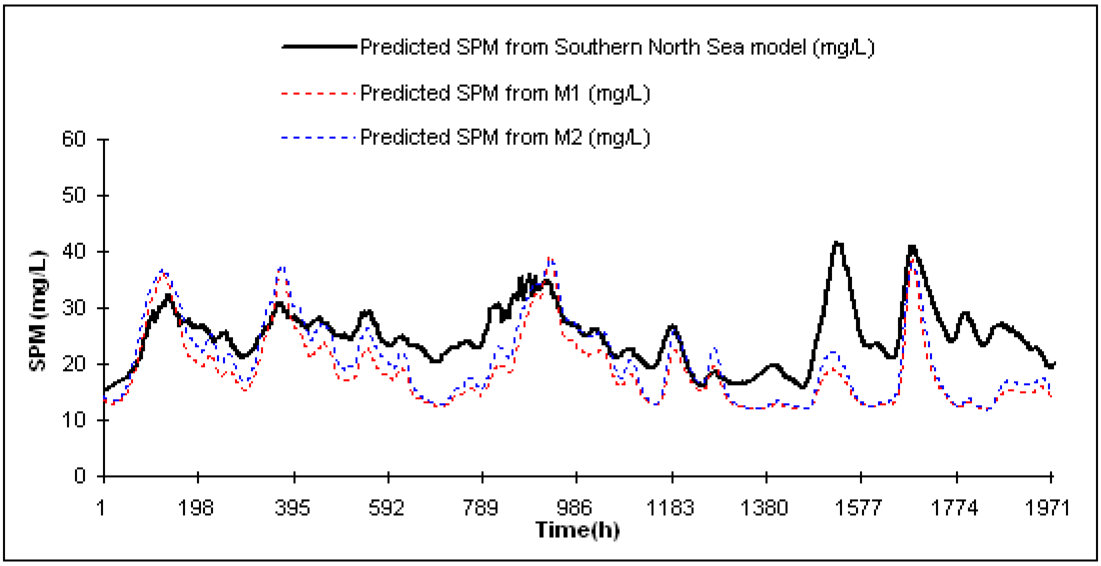

Figure 3.21 Comparison of predicted SPM from M1 (ANN model without using data transformation), M2 (ANN model with data transformed to zero mean and unit variance) and the Southern North Sea model.

ANN models are then built for stormy and non-stormy periods separately. And these two models are named as M3 and M4 respectively in this study.

Table 3.3 Model performance of M3 (ANN model for stormy periods) and M4 (ANN model for non-stormy periods).

Model Performance	***M3***	***M4***
MSE [-]	5.61	1.59
NMSE [-]	0.43	0.39
MAE [mg/L]	1.80	0.94
Min Abs Error [mg/L]	0.00	0.00
Max Abs Error [mg/L]	7.25	3.63
r [-]	0.80	0.86

Table 3.3 shows the performances of M3 and M4. MSE of both models are decreased dramatically comparing with MSE in M1 and M2. And also the correlation between SPM from the Southern North Sea model and SPM from M3 and M4 rise up to more than 0.8.

Figure 3.22 and Figure 3.23 plot the predicted SPM from the Southern North Sea model, M3 and M4. The prediction of SPM is improved after cutting off stormy and non-stormy periods. Discrepancies of SPM in M3 and M4 could be interpreted by other processes influencing SPM concentration in addition to Tbss and Wep. For example, SPM could be diffused along the vertical direction in the water column, so local water depth could affect SPM concentration as well. However, water depth in one location is almost unchanged value so it can not be taken as a variable for the ANN model. Model performance of M4 could be improved further if non-stormy periods are separated into calm periods and intermediate periods. However, this would increase the complexity of the model structure so is not taken into account in this thesis.

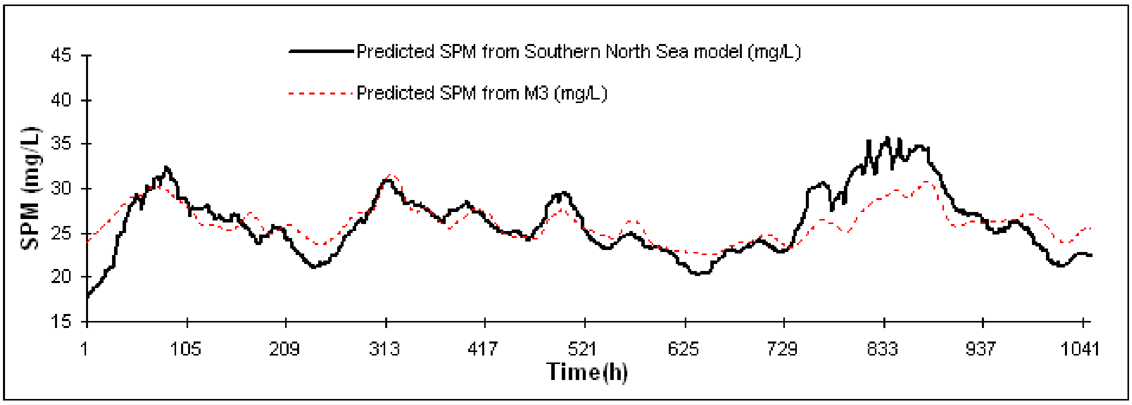

Figure 3.22 Comparison of predicted SPM from M3 (ANN model for stormy periods) and the Southern North Sea model.

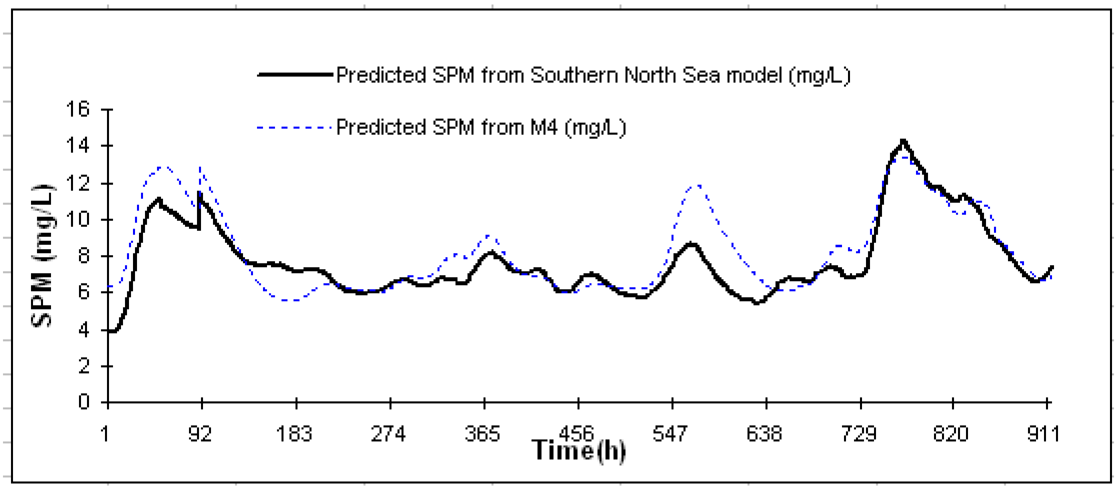

Figure 3.23 Comparison of predicted SPM from M4 (ANN model for non-stormy periods) and the Southern North Sea model.

3.6.9 Sensitivity analysis

Sensitivity analysis (SA) is the study of how the variation in the output can be apportioned, qualitatively or quantitatively, to different sources of variation in the input of a model. Sensitivity analysis is important in model results analysis. It investigates the robustness of the model and gives us confidence that the input variables used in the model are appropriate.

In this study, each input variable used in the model are examined. And this research was broken down into three blocks.

The first block explores how Wep influenced the model results. The ANN model is rebuilt without choosing Wep as input for stormy period. And this model is named as M5 in this study.

$$SPM(\textit{stormy}) = f\,(Tbss\text{-}9, Tbss\text{-}10, Tbss\text{-}11) \tag{3.15}$$

The second block explores how Tbss influenced the model results. The ANN model was rebuilt by removing Tbss-11 from input sets for stormy period and this model is denoted as M6.

$$SPM(stormy) = f\,(Tbss\text{-}9, Tbss\text{-}10, Wep) \tag{3.16}$$

In the third block, ANN model is rebuilt for stormy period using Tbss-9, ΔTbss, Wep as inputs. This model is defined as M7 in this thesis.

$$SPM(\textit{stormy}) = f\,(Tbss\text{-}9, \Delta Tbss, Wep) \tag{3.17}$$

where ΔTbss = (Tbss-9) – (Tbss-10), which is the change of Tbss.

Table 3.4 Comparison of model performances under different scenarios for stormy periods.

Model Performance	***M3***	***M5***	***M6***	***M7***
MSE [-]	5.61	9.38	6.63	7.32
NMSE [-]	0.43	0.71	0.50	0.56
MAE [mg/L]	1.80	2.36	1.94	2.08
Min Abs Error [mg/L]	0.00	0.00	0.00	0.00
Max Abs Error [mg/L]	7.25	9.57	7.23	7.19
r [-]	0.80	0.54	0.73	0.69

Table 3.4 shows the model performances under different scenarios in the stage of sensitivity analysis. For better outlook of predicted SPM from M3, M5, M6 and M7, the plots of predicted SPM from each model is split into three parts and are shown in Figure 3.24, Figure 3.25 and Figure 3.26 respectively.

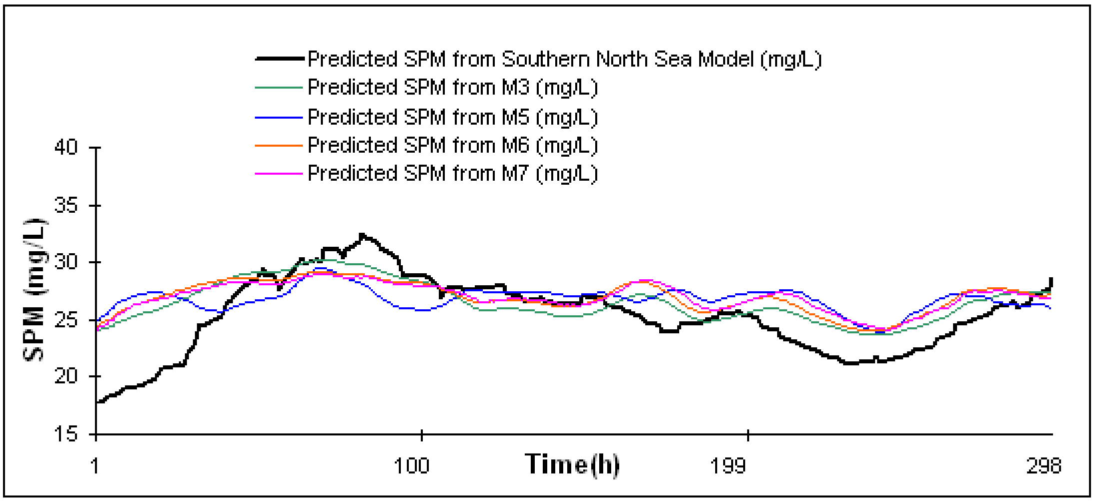

Figure 3.24 Predicted SPM from each model for stormy periods in the stage of sensitivity analysis (Part I).

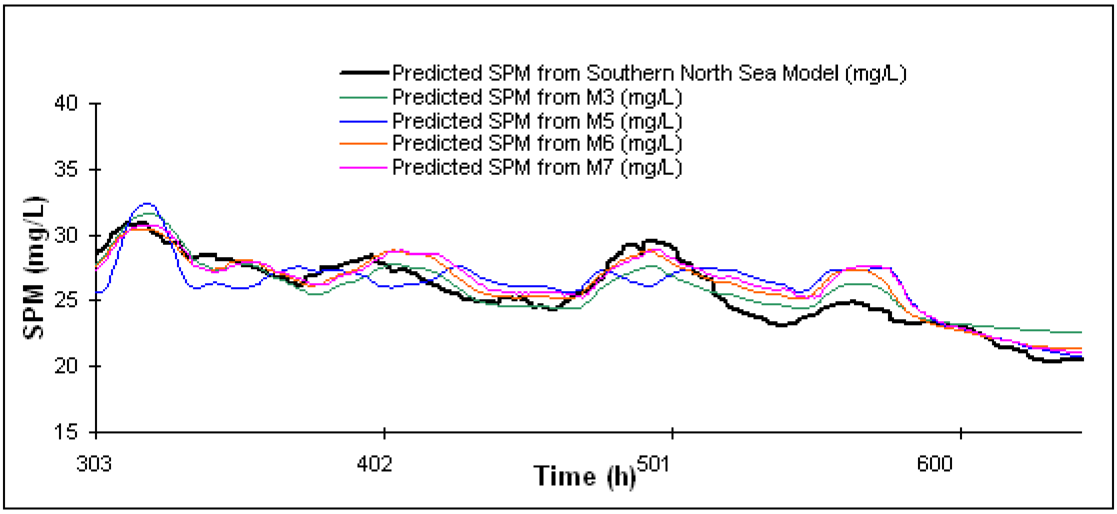

Figure 3.25 Predicted SPM from each model for stormy periods in the stage of sensitivity analysis (Part II).

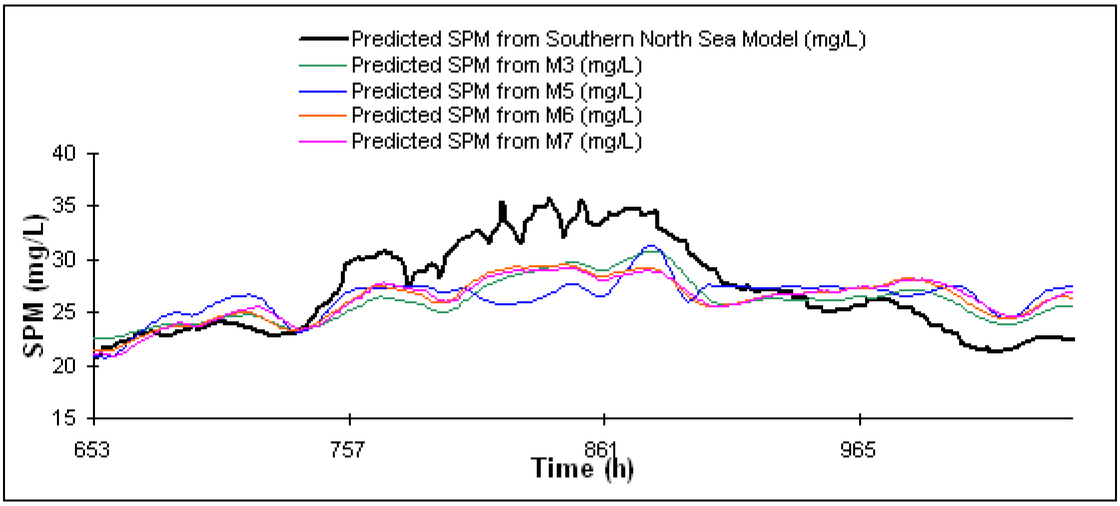

Figure 3.26 Predicted SPM from each model for stormy periods in the stage of sensitivity analysis (Part III).

M5 gives highest mean square error (9.38) and the worst prediction of SPM from the plots. This implies that Wep is a significant variable for building ANN model. Taking Wep into consideration gives better prediction of SPM in the ANN model for that Wep captures the influences of the previous waves to SPM concentration. M6 and M7 do not improve the model results. M3 gives the best performance of the ANN model which implies the input variables used in M3 is a good choice. Selecting Tbss with more time lags may increase the model accuracy further but it is at the cost of a more complex model structure so that it is not taken in this study.

Sensitivity analysis is also done for non-stormy periods. The way in investigating input variables for non-stormy periods is the same with that in stormy periods. Three models are built as below:

$$\text{SPM}(\textit{non-stormy}) = \text{f (Tbss-9, Tbss-10, Tbss-11)} \tag{3.18}$$

$$\text{SPM}(\textit{non-stormy}) = \text{f (Tbss-9, Tbss-10, Wep)} \tag{3.19}$$

$$\text{SPM}(\textit{non-stormy}) = \text{f (Tbss-9, } \Delta\text{Tbss, Wep)} \tag{3.20}$$

These three models are named as M8, M9 and M10 respectively in this thesis.

Table 3.5 Comparison of model performances under different scenarios for non-stormy periods.

Model Performance	M4	M8	M9	M10
MSE [-]	1.59	6.12	2.93	3.18
NMSE [-]	0.39	1.51	0.72	0.78
MAE [mg/L]	0.94	2.06	1.36	1.52
Min Abs Error [mg/L]	0.00	0.01	0.00	0.00
Max Abs Error [mg/L]	3.63	6.71	4.94	4.63
r [-]	0.86	0.64	0.83	0.73

Table 3.5 shows the model performances under different scenarios in the stage of sensitivity analysis. For better outlook of predicted SPM from M8, M9 and M10, the plots is split into two parts and are shown in Figure 3.27 and Figure 3.28. M8 gives highest mean square error (6.12) and the worst prediction of SPM. M4 gives the best performance similar to that in stormy periods; results show the importance of Wep for predicting SPM in non-stormy periods.

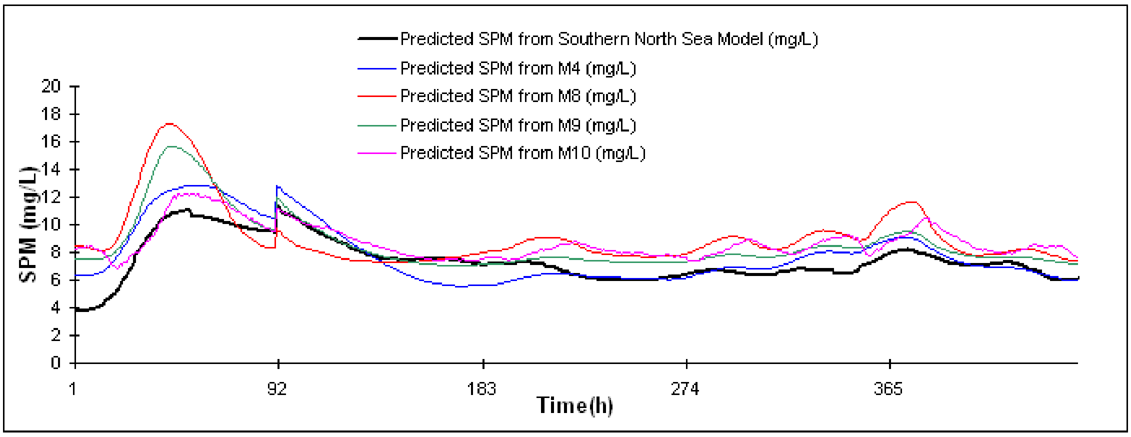

Figure 3.27 Predicted SPM from each model for non-stormy periods in the stage of sensitivity analysis (Part I).

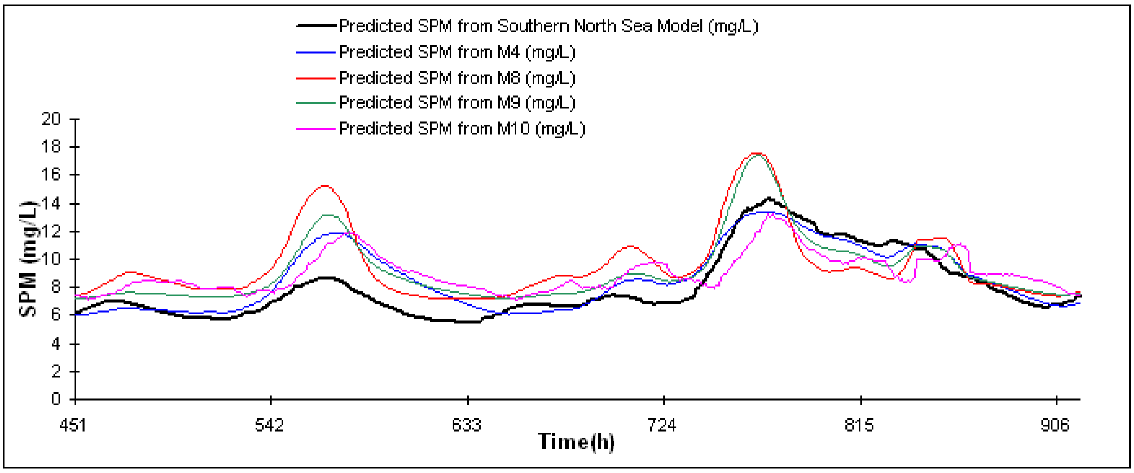

Figure 3.28 Predicted SPM from each model for non-stormy periods in the stage of sensitivity analysis (Part II).

3.7 Conclusions and Discussions

In this chapter, ANN models are built for predicting SPM concentration at IJmuiden. By analyzing the physical processes of sediment transport and analyzing outputs from the Southern North Sea model, Tbss was taken as an important input variable because it is the main driving force for influencing SPM concentration. In order to capture the influence of previous waves to SPM concentration, Wep was used as another variable and sensitivity analysis showed the importance of Wep in predicting SPM.

Data used for building the ANN model are smoothed out by 24 hours moving average to reduce variance along the data range due to diurnal tidal movement. Tbss and Wep are transformed to zero mean and unit variance to adjust the distribution of data closer to normal and transformation of data improved the model performance. In order to capture the seasonal effect, stormy and non-stormy periods were defined according to Hs and models were built separately for stormy and non-stormy periods. Consequently the predictions of SPM are improved. Sensitivity analysis was done to investigate the performance of input variables.

Model results analysis shows a strong possibility of predicting SPM by surrogate models. However, several limitations in the study should be pointed out as follow:

- When defining Wep, we assumed that waves within the last two weeks (336 hours) give effects to the current SPM concentration and no influence any more of the wave energy before two weeks ago. Furthermore, Wep is a weighted sum of wave energy. The weights used in this study are determined by a linear function. These two assumptions could be examined more carefully. Optimization approaches could be applied in defining Wep.

- The stormy and non-stormy periods were defined in an arbitrary way in this study. It needs to be mentioned that this definition may show failure at other locations or using data of other years. The critical values of Hs for defining stormy and non-stormy periods (1m and 2.5m in this study) are flexible; values could be changed if the definition is quite different from the reality.

- Surrogate model performance could be enhanced by selecting more parameters in addition to Tbss and Wep.

In the following chapter, the concept of parsimonious model is introduced and a simpler parsimonious model for predicting SPM is presented.

Chapter 4 Parsimonious modelling of fine sediment dynamics

4.1 Introduction

In statistics, ‘parsimonious’ refers to simplicity. Therefore a parsimonious model refers to the simplest feasible model with the fewest possible number of variables (From Wikipedia). Parsimonious modelling aims at achieving maximally simple or compact models. Parsimonious models are helpful for judging hypotheses and are attractive for research purposes because they are transparent, requiring less computation time and easy to interpret the results. The purpose of using parsimonious models in research is trying to seek the most essential character of physical processes, which is very useful for understanding the nature processes and for decision making.

In this chapter, a linear regression method is applied to build parsimonious models for predicting SPM concentration. The parsimonious model is firstly built at IJmuiden with Tbss and Wep as input variables. Then a series of parsimonious models with same model structure are built at other locations in the Southern North Sea area.

4.2 Building parsimonious model at IJmuiden

The concept of parsimonious model is building the feasible models with as less input variables as possible. So in this study, only Tbss-9 and Wep are selected as the inputs. Data for building the parsimonious model are smoothed out with 24 hours moving average and then transformed to zero mean and unit variance. In order to capture seasonal effects, stormy and non-stormy periods are defined according to the definition of storms in Chapter 3.6.7. Linear regression method is applied to build the parsimonious models in this study. MS-Excel provides linear regression method to analyze data, so the parsimonious models are built by Regression function of the Data Analysis toolpack in this study. The structure of the parsimonious model can be expressed as:

$$\text{SPM} = a*\text{Tbss-9} + b*\text{Wep} + c \tag{4.1}$$

where Tbss-9 is total bed shear stress with 9 hours lag, Wep is wave energy past for capturing the effect on SPM by pervious waves, *a* and *b* are the coefficients of Tbss-9 and Wep, *c* is constant.

Parsimonious models are built for stormy and non-stormy periods separately in the following sections.

4.2.1 Parsimonious model for stormy periods at IJmuiden

The premise behind building parsimonious model with linear regression methods is that SPM has a similar linear relationship with Tbss and Wep. In order to examine the feasibility of predicting SPM with Tbss and Wep in a linear way, parsimonious models are firstly built for each storm and then the possibility of building the parsimonious model for the whole stormy periods is explored.

There are totally fifteen storms in stormy periods. Figure 4.1 shows SPM, Tbss and Wep value in each storm. Parsimonious model is built for each storm and the results are shown in Table 4.1.

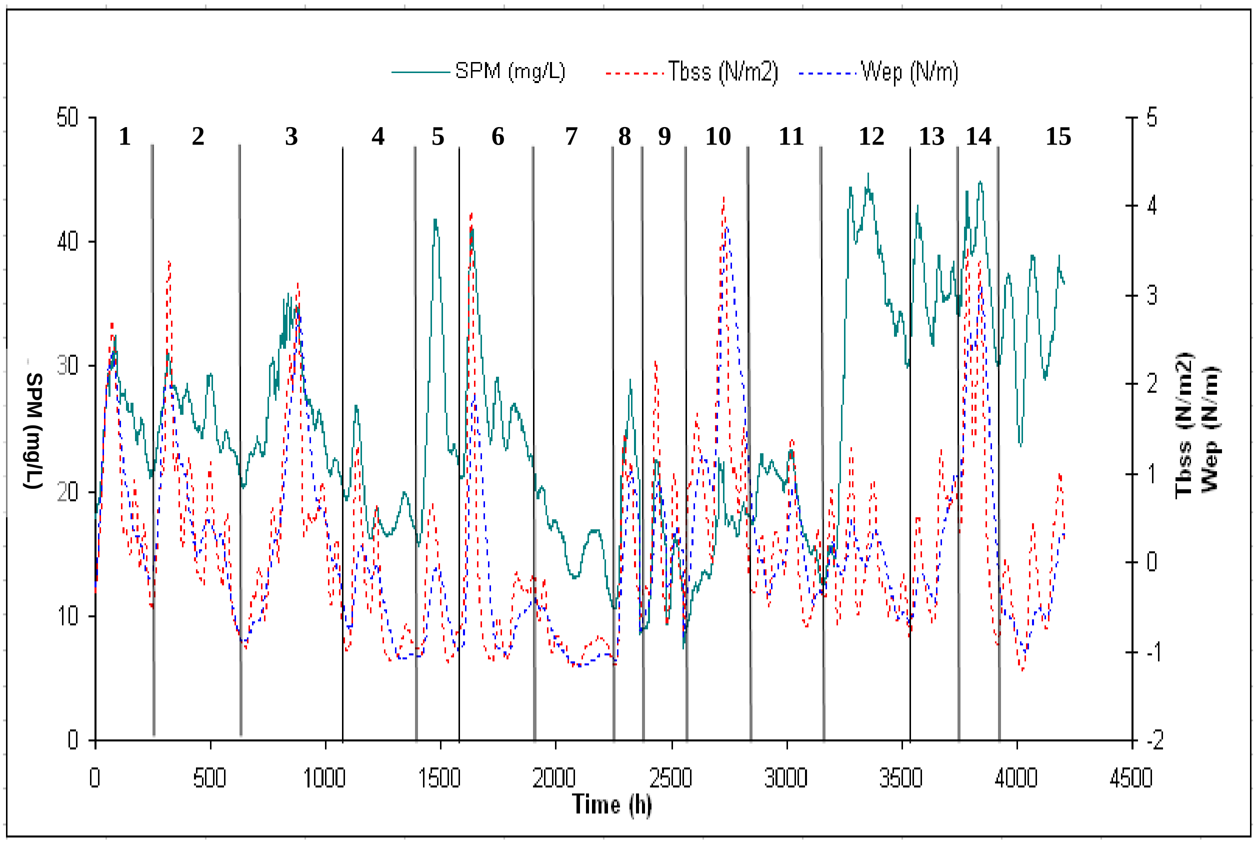

Figure 4.1 Plot of SPM between Tbss and Wep for all fifteen storms at IJmuiden

Table 4.1 Performance of parsimonious model for each storm; *a*, *b*, *c* are coefficients in Eq. 4.1.

Storm No.	***a*** **[-]**	***b*** **[-]**	***c*** **[-]**	**Correlation between predicted SPM from Parsimonious model and the Southern North Sea model [-]**
1	0.34	3.02	22.67	0.80
2	1.69	1.02	24.43	0.90
3	3.92	0.15	24.73	0.87
4	5.07	-3.55	19.25	0.86
5	2.58	17.90	39.57	0.97
6	1.65	3.56	27.84	0.84
7	8.56	1.68	24.95	0.73
8	2.63	5.21	19.44	0.98
9	2.14	5.73	13.24	0.98
10	0.07	2.18	12.13	0.76
11	2.79	1.86	18.89	0.57
12	1.14	6.21	34.23	0.26
13	4.55	-3.68	35.63	0.68
14	1.10	2.92	31.83	0.98
15	8.45	-6.59	31.79	0.87
All storms	2.69	0.22	24.67	0.34

The parsimonious model for all storms does not perform well and correlation between

SPM from the Southern North Sea model and parsimonious model is only 0.34. However, the performance is improved for most of the single storms. Parsimonious model for storm 8, 9, 14 give the best performance and correlation between SPM from the Southern North Sea model and parsimonious model is 0.98. Simulation of storm 12 gives the worst performance. Figure 4.2 and Figure 4.3 show the performance of parsimonious model for storm 14 and 12 respectively. Most parsimonious models give acceptable performances except for storm 11, 12 and 13, which means that the linear relationship between SPM and Tbss, Wep for storm 11, 12 and 13 is not clear. From Figure 4.1, we can see that SPM is not related too much with Tbss and Wep for storm 11, 12 and 13.

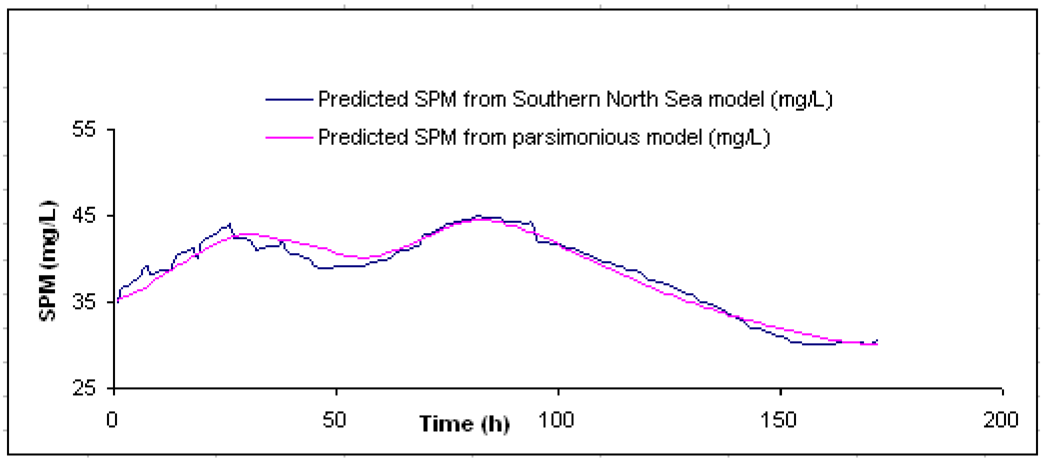

Figure 4.2 Predicted SPM from the parsimonious model compared with SPM predicted by the Southern North Sea model for storm 14 at IJmuiden.

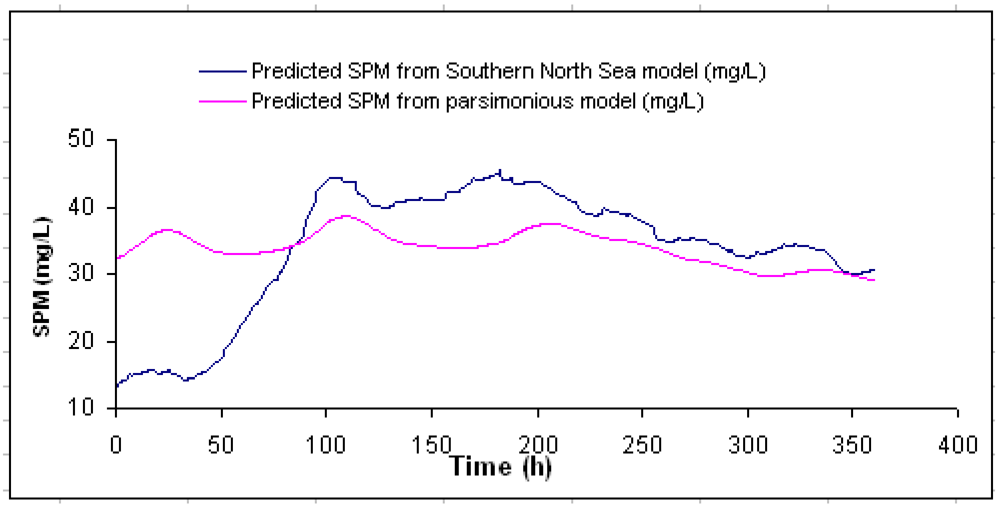

Figure 4.3 Predicted SPM from the parsimonious model compared with SPM predicted by the Southern North Sea model for storm 12 at IJmuiden.

In order to improve the prediction of SPM by parsimonious model, storm 11, 12 and 13 are removed from the dataset. Storm 1 to storm 7 are selected to build the parsimonious model and storm 8, 9, 10, 14 and 15 are choose as testing data to examine the model performance. Based on the training data, the parsimonious model for stormy periods at IJmuiden is shown as:

$$SPM = 2.62*Tbss\text{-}9 + 1.89*Wep + 19.8 \qquad (4.2)$$

This model is tested by storm 8, 9, 10, 14 and 15 respectively and the results are shown in Table 4.2. Figure 4.4 shows the testing results.

Table 4.2 Testing result of parsimonious model for each storm

Storm No.	Correlation between predicted SPM from Parsimonious model and the Southern North Sea model [-]
8	0.95
9	0.94
10	0.68
14	0.96
15	0.74

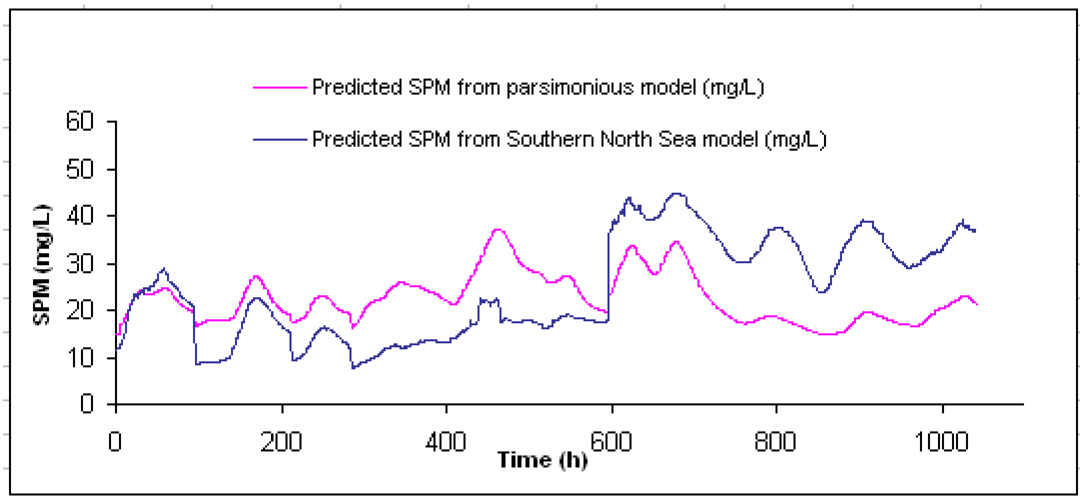

Figure 4.4 Testing results of parsimonious model for stormy periods at IJmuiden.

The parsimonious model performs well for storm8, 9 and 14 while a model result for storm 10 is not satisfactory. The reason is that in parsimonious model, SPM is predicted by Tbss and Wep in a linear way. Model performance is largely depended on the linear relationship of SPM and input variables (Tbss and Wep) and consequently depended on the values of coefficients *a*, *b* and *c*. Predictions of SPM could be satisfactory for a single storm, however, it is not easy to fix values of *a*, *b* and *c* to fulfil good predictions of SPM for all storms. This can also be interpreted by Table 4.1 also. Parsimonious models for most storms give good performance but the values of a, b, c for each storm is quite different. For example, both storm5 and storm8 give high correlations between predicted SPM and desired SPM. But *a*, *b*, *c* for storm5 is 2.58, 17.90, 39.57 while for storm8 is 2.63, 5.21 19.44. These differences are mainly due to the large variation of SPM values in the stormy periods.

4.2.2 Parsimonious model for non-stormy periods at IJmuiden

Parsimonious model is built for non-stormy periods in this section. 75% of the data are used to build the model and 25% is used for testing. Eq. 4.3 gives parsimonious model for non-stormy periods.

$$SPM = 0.79 * Tbss\text{-}9 + 2.23 * Wep + 6.41 \quad (4.3)$$

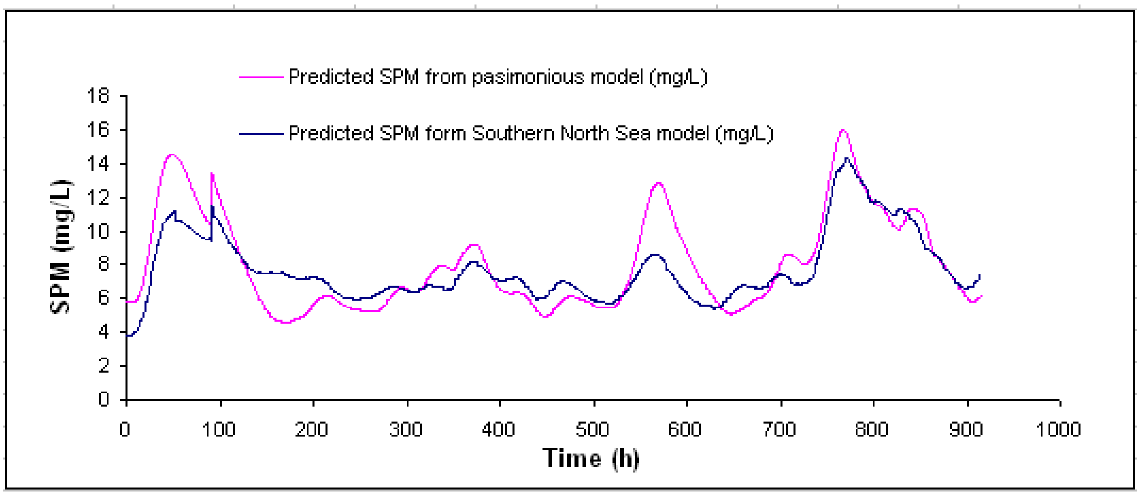

Figure 4.5 Testing results of parsimonious model for non-stormy periods at IJmuiden.

Figure 4.5 shows the model performance with testing data. The correlation between predicted SPM from parsimonious model and the Southern North Sea model is 0.86.

4.2.3 Comparison of parsimonious model with surrogate model

Both parsimonious model and surrogate model are applied to predict SPM concentration at IJmuiden. In this section, we made a comparison of these two models and explore the main differences between them.

Table 4.3 Comparison of surrogate model and parsimonious model

	Parameters for building model	Input variables	Method
Surrogate model	Tbss, Wep	Tbss-9, Tbss-10, Tbss-11, Wep	ANN
Parsimonious model	Tbss, Wep	Tbss-9, Wep	Linear regression

Table 4.3 shows the parameters, input variables and methods used for building surrogate models and parsimonious models. There are only two parameters and four input variables were used to build surrogate model, which make the surrogate model seem to be ‘parsimonious’ compared with process-based model. However, parsimonious model was built with only two inputs and simple linear regression method, for which the model is simpler.

For stormy periods, surrogate model gave an acceptable simulation of SPM and the correlation between predicted SPM and target SPM is 0.80. However, parsimonious model could predict SPM for one single storm but becoming difficult to capture the whole stormy periods.

For non-stormy periods, parsimonious model gave a similar simulation of SPM compared with surrogate model. Figure 4.6 shows the predicted SPM by parsimonious model and surrogate model.

The main advantage of parsimonious model is that SPM can be calculated very easily after knowing Tbss and Wep, which is convenient and transparent. But the premise behind the parsimonious model is linear relationship between SPM and input variables. Model may show failures if this is not the case. Besides, parsimonious model may perform poor when data has large variations.

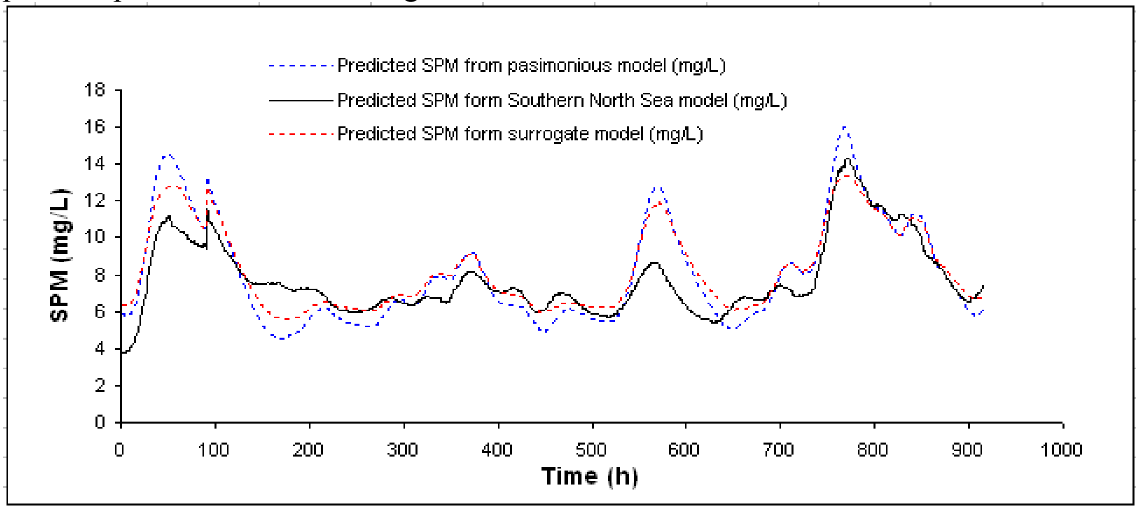

Figure 4.6 Testing results of both parsimonious model and surrogate model for non-stormy periods at IJmuiden.

4.2.4 Conclusions

In this section, the possibility of building parsimonious model to prediction SPM concentration at IJmuiden is explored. Parsimonious models are built for stormy periods and non-stormy periods separately.

In stormy periods, parsimonious model perform well for single storm but the model performance decreased after adding all storms together. By analyzing parsimonious model performance for each storm, those storms give poor predictions of SPM are removed from the stormy datasets. Then 7 storms are selected to build the parsimonious model for stormy periods and 5 storms are used as testing data. Results show erroneous of parsimonious model to predict SPM for storm 10. The reason it that the parameters of parsimonious model *a, b* and *c* are quite different for each storm, it is not easy to fix values of *a, b* and *c* to meet good predictions of SPM for all storms together. This is mainly due to the SPM value in stormy periods has a larger range from about 10 to 44 mg/L, SPM is changing dramatically from one storm to another, so that linear regression method is not robust enough to capture the trend of SPM changing for all storms. In non-stormy periods, SPM ranges from about 5 to 20mg/L, belonging to a narrower range. The results show the possibility to build parsimonious model in a linear way to predict SPM for non-stormy periods.

It seems surprising for parsimonious models of both stormy and non-stormy periods that SPM is non-zero value when Tbss and Wep equal to zero in Eq. 4.2 and Eq. 4.3. The reason could be that there are other potential processes (such as fresh water discharge, wind effect, etc) contributing to SPM concentration in addition to total bed shear stress. Furthermore, the process of sediment dynamics is quite complex in the physical systems. SPM concentration in one location could be affected by local water depth and also by SPM at other locations nearby, which are not taken into account in this study. In the following sections, parsimonious models are built at many other locations in the Southern North Sea area.

4.3 Building parsimonious models at other locations.

In order to examine the applicability of parsimonious model for perdiction SPM concentration at other locations in the Southern North Sea, parsimonious model is firstly built at a location close to IJmudien. Noordwijk 1.5 km is selected for investigating the performance of parsimonious model firstly. In order to explore influence on SPM concentration by local water depth, a series of parsimonious models are built further at Noordwijk with different distances to the coast. As it is shown in Figure 4.7, grids with a cross are observation stations in the Southern North Sea model.

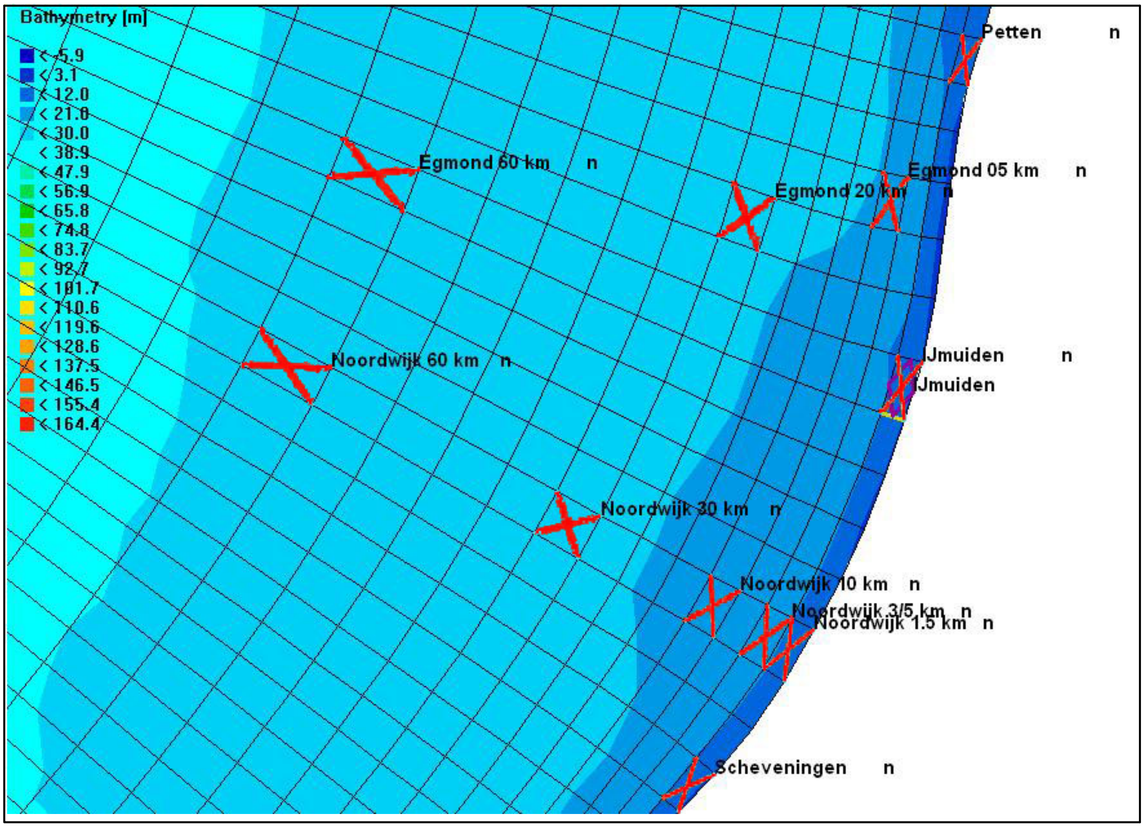

Figure 4.7 Distribution of observation stations in the Southern North Sea.

Table 4.4 shows the local water depth at IJmuiden and Noordwijk. Noordwijk 1.5 km means this location is 1.5 km far away from the coast and Nooedwijk 3/5 km mean this location is about 3 to 5 km away from the coast. It is obvious that water depth increases if the location is further offshore. Parsimonious models are built at each location in the following sections.

Table 4.4 Local water depth at IJmudien and Noordwijk

Locations	Local water depth (m)
IJmuiden	7~8
Noordwijk 1.5 km	8.5~9.5
Noordwijk 3/5 km	15.5~16.5
Noordwijk 10 km	19~19.5
Noordwijk 30 km	23~23.5
Noordwijk 60 km	27.4~27.8

4.3.1 Parsimonious model at Noordwijk 1.5km (water depth = 8.5-9.5m)

The procedure of processing data in this part is exactly the same with that at IJmuiden. Tbss-9 and Wep are selected as the inputs for building parsimonious model. The model structure is the same as shown in Eq. 4.1. Figure 4.8 shows SPM with corresponding Tbss-9 and Wep at Noordwijk 1.5 km. Parsimonious models are built for stormy and non-stormy periods separately.

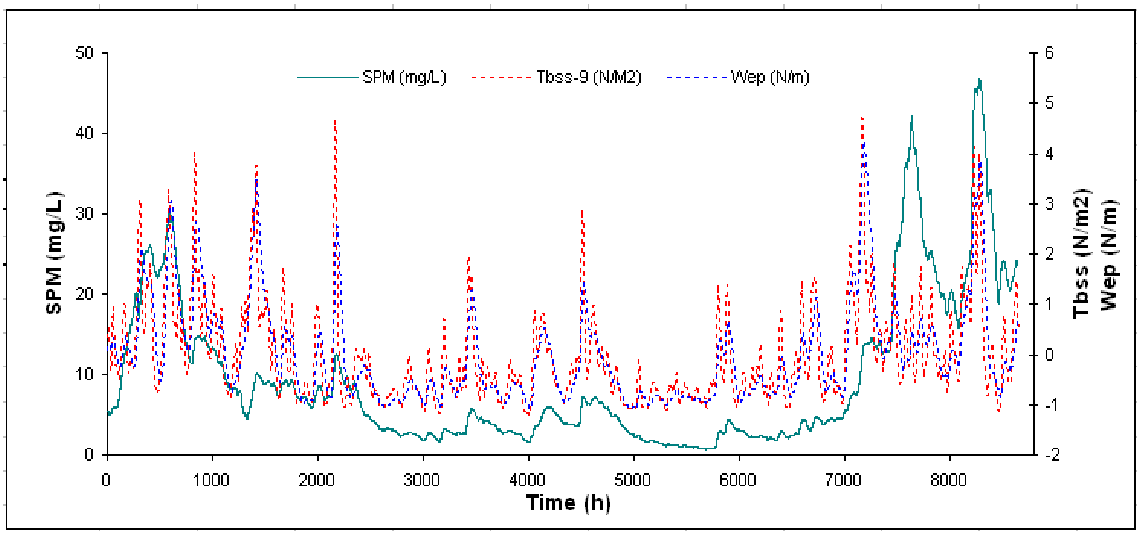

Figure 4.8 Plot of SPM between Tbss-9, and Wep at Noordwijk 1.5 km.

For stormy periods, there are totally 7 storms and parsimonious model is built for each of them. The results are shown in Table 4.5. Parsimonious model for most storms give very good simulation of SPM. Figure 4.9 gives an example of the simulation for storm 1.

Table 4.5 Performance of parsimonious model for each storm at Noordwijk 1.5 km; *a*, *b*, *c* are coefficients in Eq. 4.1.

Storm No.	*a* [-]	*b* [-]	*c* [-]	Correlation between predicted SPM from Parsimonious model and the Southern North Sea model [-]
1	0.53	1.57	23.78	0.99
2	-0.43	1.43	11.83	0.94
3	-0.94	1.68	6.60	0.72
4	0.12	1.37	8.70	0.99
5	-0.68	2.55	6.02	0.88
6	0.67	8.60	15.13	0.98
7	-0.32	4.34	31.38	0.99

Storm 3 is then removed out form the stormy periods. Storm 1, 2 and 4 are used to build the parsimonious model and storm 5, 6 and 7 are used to test the model performance. Finally, the parsimonious model for stormy periods at Noordwijk 1.5 km is shown as:

$$SPM = -1.09 * Tbss9 + 2.66 * Wep + 16.6 \tag{4.4}$$

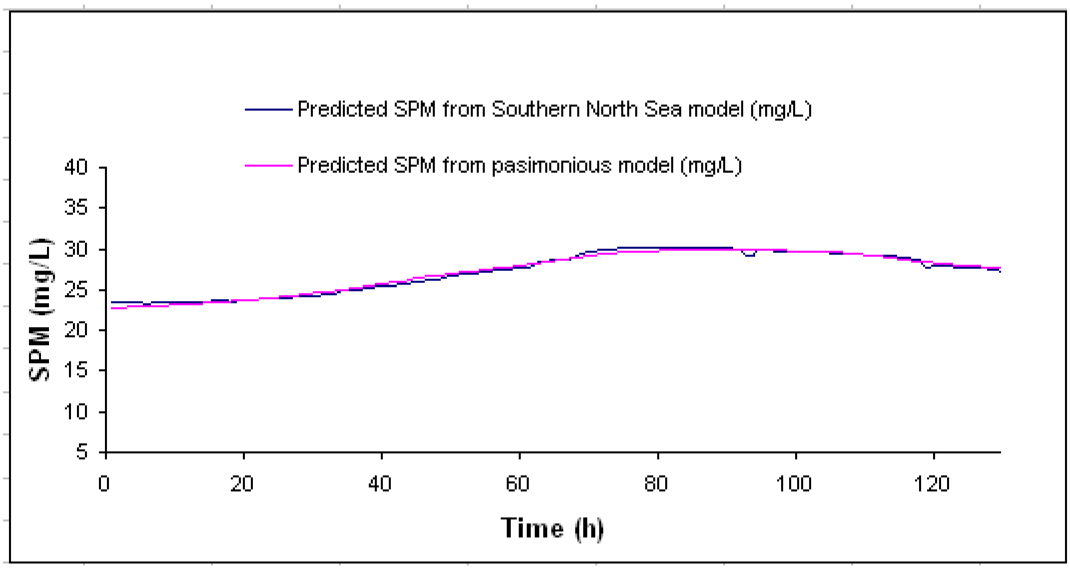

Figure 4.9 Parsimonious model performance for storm 1 at Noordwijk 1.5 km.

Table 4.6 Testing result of parsimonious model for each storm

Storm No.	Correlation between predicted SPM from Parsimonious model and the Southern North Sea model [-]
5	0.87
6	0.89
7	0.92

As it is shown in Table 4.6, the correlation between predicted SPM from parsimonious model and the Southern North Sea model is acceptable for each single storm. Figure 4.10 plots the predicted SPM from the parsimonious model for all storms.

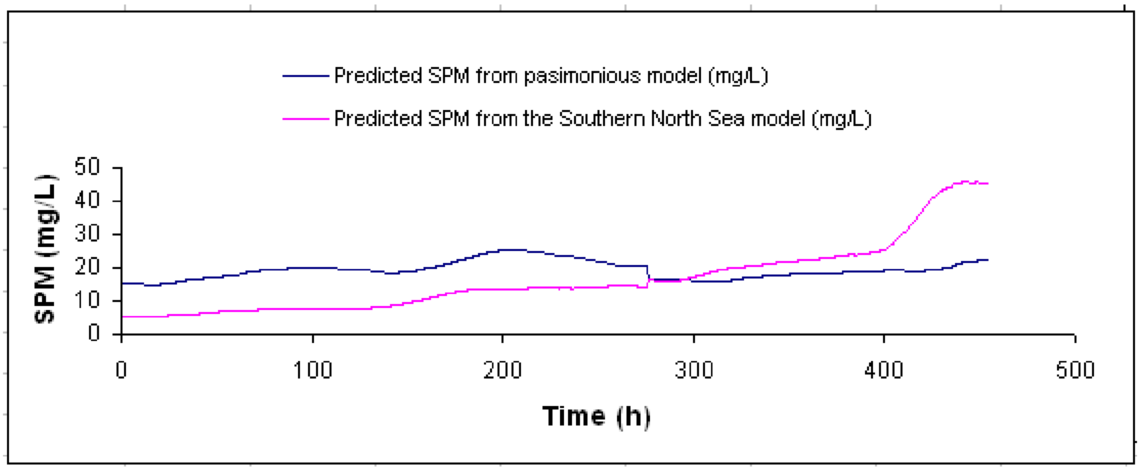

Figure 4.10 Testing results of parsimonious model for stormy periods at Noordwijk 1.5 km.

Parsimonious model for non-stormy periods at Noordwijk 1.5 km is shown as follows:

$$SPM = -0.15 * Tbss9 + 3.27 * Wep + 4.50 \qquad (4.5)$$

Figure 4.11 shows the predicted SPM from parsimonious model compared with SPM from Southern North Sea model for non-stormy periods at Noordwijk 1.5 km. The correlation between them is 0.72.

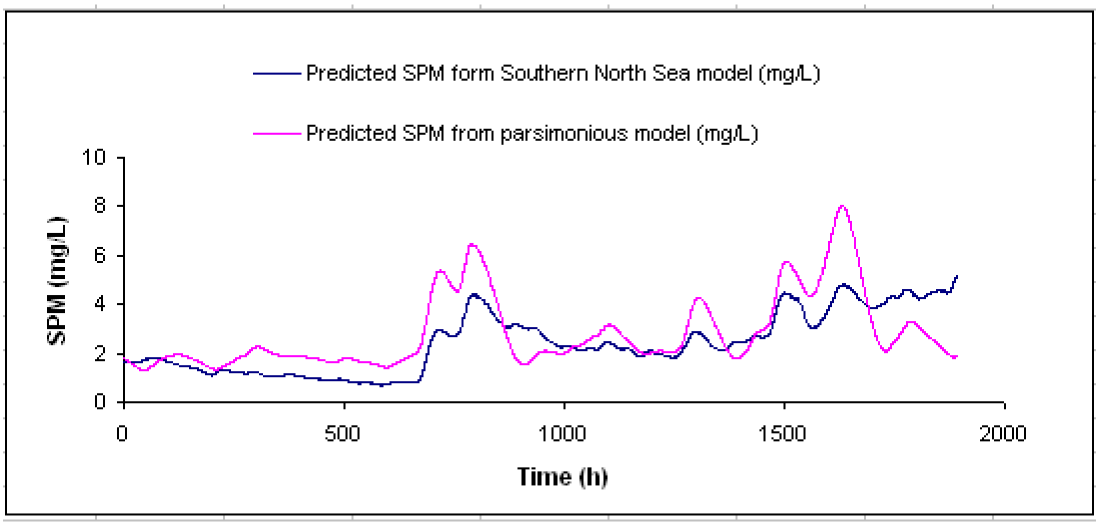

Figure 4.11 Testing results of parsimonious model for non-stormy periods at Noordwijk 1.5 km.

4.3.2 Parsimonious model at Noordwijk 3/5 km (water depth = 15.5-16.5m)

Parsimonious model is built at Noordwijk which is further offshore and parsimonious models are built for stormy and non-stormy periods separately.

There are totally 10 storms in stormy periods and parsimonious model is built for each of them. The results are shown in Table 4.7. Most of the parsimonious model performs unsatisfactory. Therefore the parsimonious model can not be applied at Noordwijk 3/5 km to predict SPM for storm periods. The reason is that the linear relationship between SPM and Tbss, Wep is quite poor so that it is impossible to predict SPM by Tbss and Wep in a linear way. This could be interpreted better from Figure 4.12.

Table 4.7 Performance of parsimonious model for each storm at Noordwijk 3/5 km; *a*, *b*, *c* are coefficients in Eq. 4.1.

Storm No.	*a* [-]	*b* [-]	*c* [-]	Correlation between predicted SPM from Parsimonious model and the Southern North Sea model [-]
1	-4.26	-1.9	21.97	0.22
2	0.28	-5.01	39.26	0.64
3	1.65	1.65	18.05	0.68
4	-1.1	1.58	12.49	0.73
5	-1.23	1.46	8.08	0.57
6	0	1.31	7.3	0.99
7	-0.06	1.56	3.85	0.99
8	-1.54	2.49	7.92	0.66
9	-1.59	5.06	28.01	0.26
10	-1.33	5.87	11.13	0.98

There is almost no relationship between SPM and Tbss, neither for SPM and Wep from the plot. The results imply that the correlation between SPM and Tbss decreased quite a lot for stormy periods in the area with deeper water depth.

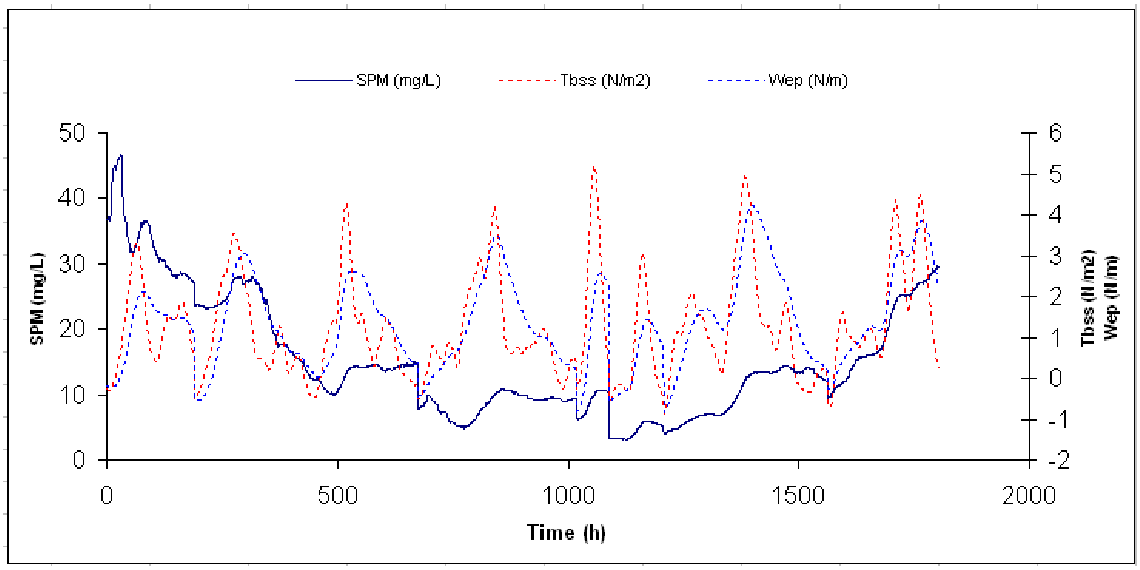

Figure 4.12 Plot of SPM between Tbss and Wep for stormy periods at Noordwijk 3/5 km.

The main reason is that wave energy generated near the water surface will be dissipated on its way transmitting to the bottom in areas with deep water depth. So the generated bed shear stress is not high enough to pick up sediment particles and consequently SPM value is low and almost unchanged. Furthermore, with the increase of water depth, horizontal exchange of sediments by advection becomes dominant compared with sediment diffusion along the vertical direction. Sediment picked up at the bottom will be both diffused vertically and advected in the horizontal direction so that the SPM value is low. The relationship between Tbss and SPM concentration thus is not that dramatic in the area with deep water depth.

Parsimonious model for non-stormy periods at Noordwijk 3/5 km is shown as follows:

$$SPM = -1.06 * Tbss9 + 5.4 * Wep + 6.9 \tag{4.6}$$

Figure 4.13 shows the predicted SPM from parsimonious model and compared with SPM from Southern North Sea model for non-stormy periods at Noordwijk 3/5 km. The correlation between them is 0.68.

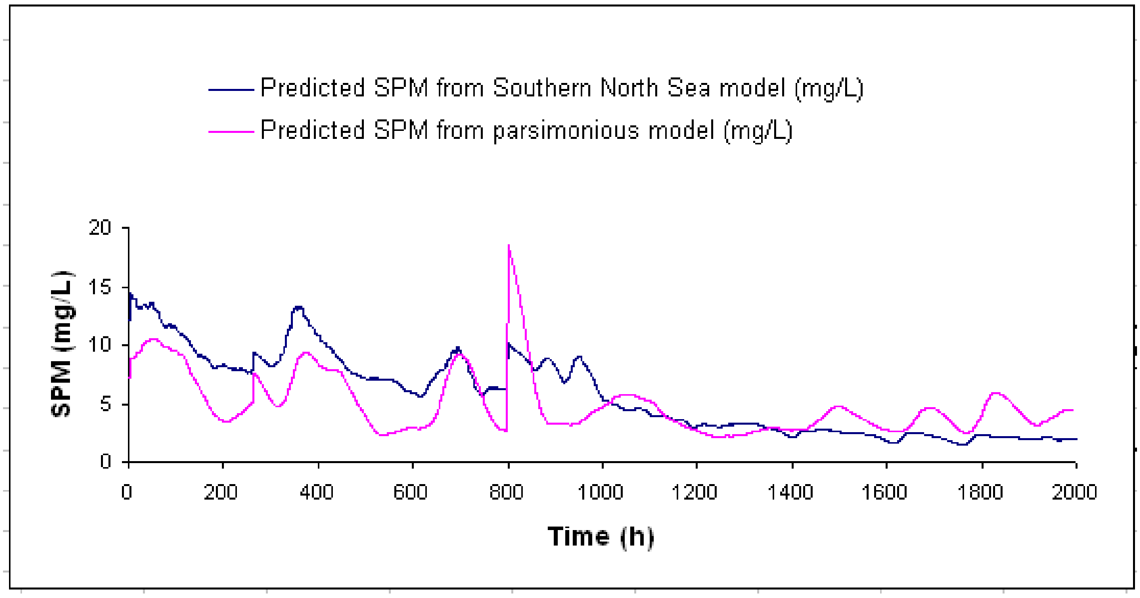

Figure 4.13 Testing results of parsimonious model for non-stormy periods at Noordwijk 3/5 km.

4.3.3 Parsimonious model at Noordwijk 10 km (water depth = 19-19.5m)

Parsimonious models are built at Noordwijk 10km for stormy and non-stormy periods separately in this section.

There are 9 storms in stormy periods and parsimonious model is built for each of them. The results are shown in Table 4.8. Similar to the results at Noordwijk 3/5km, most of the parsimonious model performs unsatisfactory. Figure 4.14 shows the plot of SPM between Tbss and Wep. The relationship of SPM between Tbss is vague and parsimonious model is not feasible for stormy period at Noordwijk 10km.

Table 4.8 Performance of parsimonious model for each storm at Noordwijk 10 km; *a*, *b*, *c* are coefficients in Eq. 4.1.

Storm No.	*a* [-]	*b* [-]	*c* [-]	Correlation between predicted SPM from Parsimonious model and the Southern North Sea model [-]
1	-4.26	-1.9	21.97	0.86
2	0.28	-5.01	39.26	0.76
3	1.65	1.65	18.05	0.61
4	-1.1	1.58	12.49	0.55
5	-1.23	1.46	8.08	0.99
6	0	1.31	7.3	0.98
7	-0.06	1.56	3.85	0.75
8	-1.54	2.49	7.92	0.55
9	-1.59	5.06	28.01	0.78

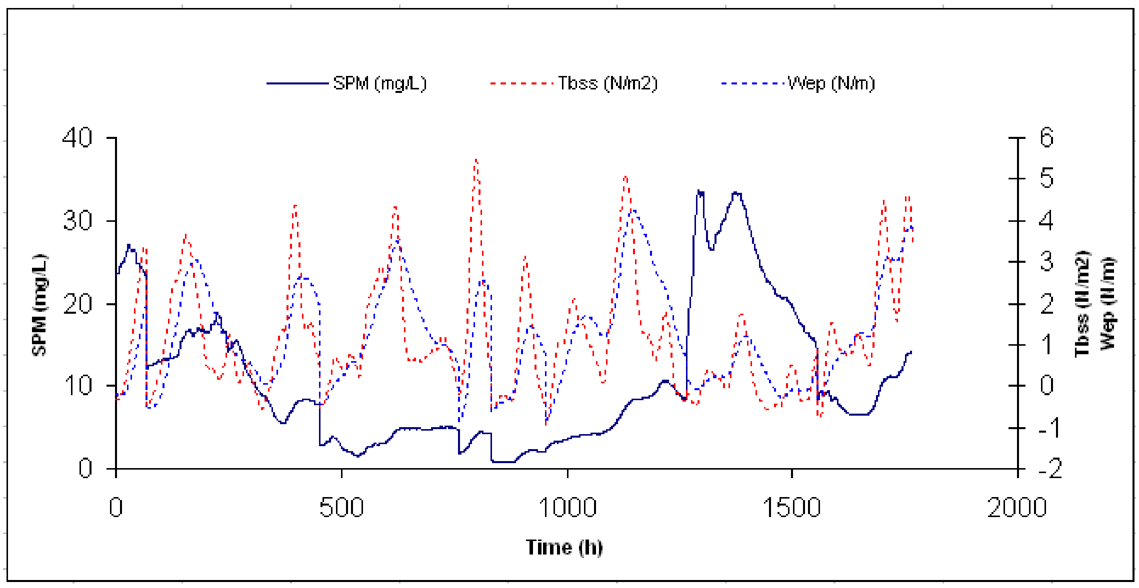

Figure 4.14 Plot of SPM between Tbss and Wep for stormy periods at Noordwijk 10 km.

Parsimonious model for non-stormy periods at Noordwijk 10 km is shown as follows:

$$SPM = 0.79 * Tbss9 + 3.15 * Wep + 3.73 \tag{4.7}$$

Figure 4.15 shows the predicted SPM from parsimonious model and compared with SPM from Southern North Sea model for non-stormy periods at Noordwijk 3/5 km. The correlation between them is 0.58.

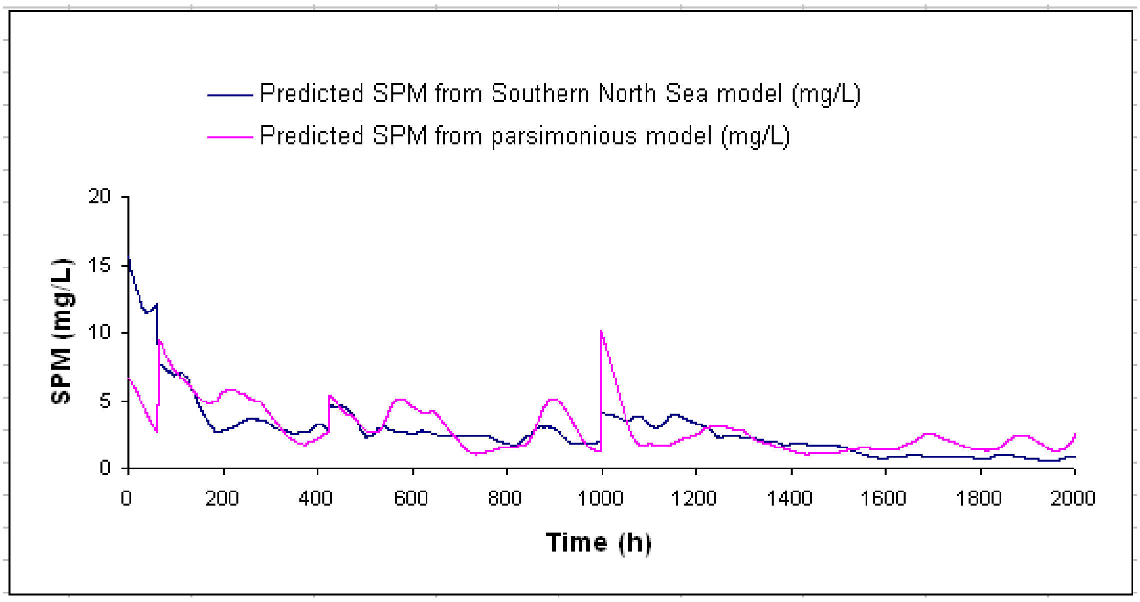

Figure 4.15 Testing results of parsimonious model for non-stormy periods at Noordwijk 10 km.

4.3.4 Parsimonious model at Noordwijk 30 km (water depth =23-23.5m)

Parsimonious models are then built at Noordwijk 30km. Figure 4.16 shows SPM, Tbss and Wep at Noordwijk 30km for all data. SPM is very low in this location mainly because of the horizontal advection of SPM.

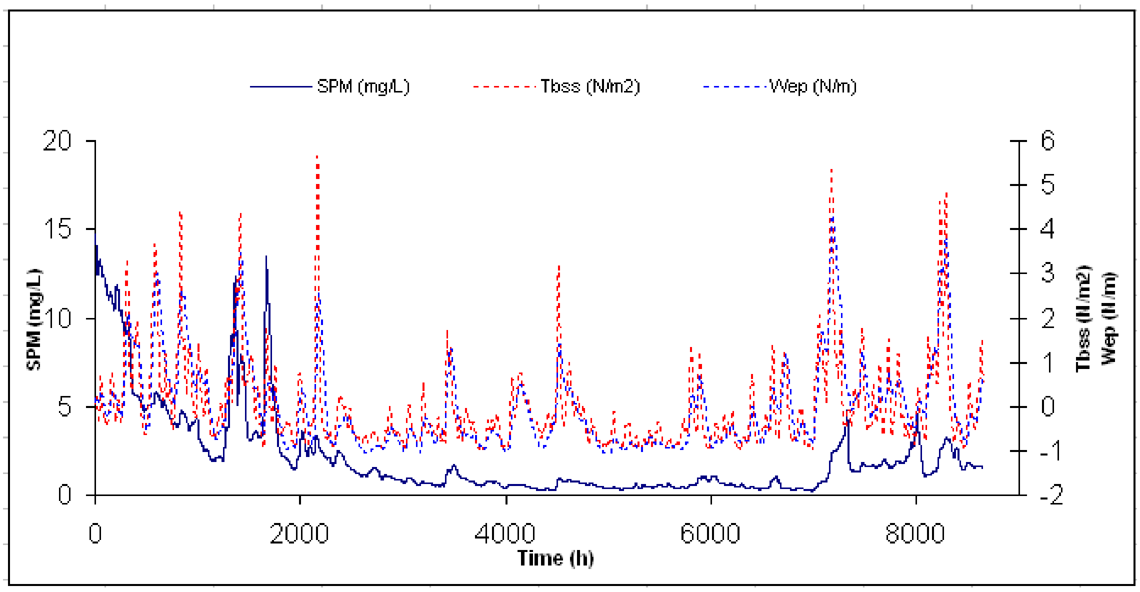

Figure 4.16 Plot of SPM between Tbss and Wep at Noordwijk 30km.

Figure 4.17 shows the plot of SPM with corresponding Wep and Tbss for stormy period. Visual inspection show that SPM is not connected well with Tbss and Wep for most of the instances. So parsimonious model is only built for non-stormy period.

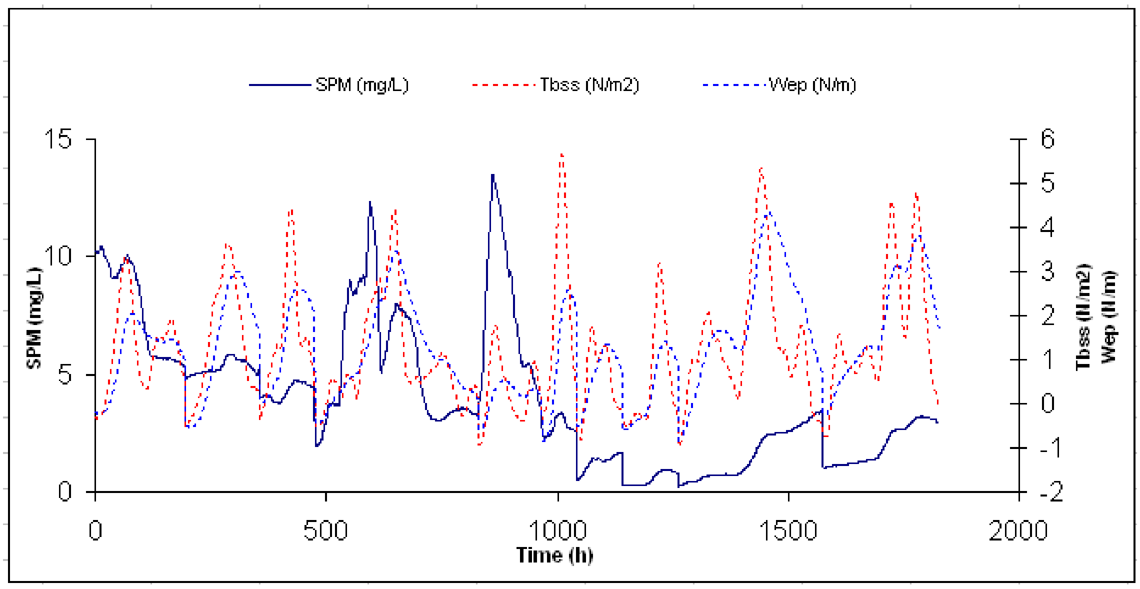

Figure 4.17 Plot of SPM between Tbss and Wep for stormy periods at Noordwijk 30 km.

Parsimonious model for non-stormy periods at Noordwijk 30 km is shown as follows:

$$SPM = -0.27 * Tbss9 + 0.82 * Wep + 1.91 \tag{4.8}$$

Figure 4.18 shows the predicted SPM from parsimonious model and compared with SPM from Southern North Sea model for non-stormy periods at Noordwijk 30 km. The correlation between them is 0.55.

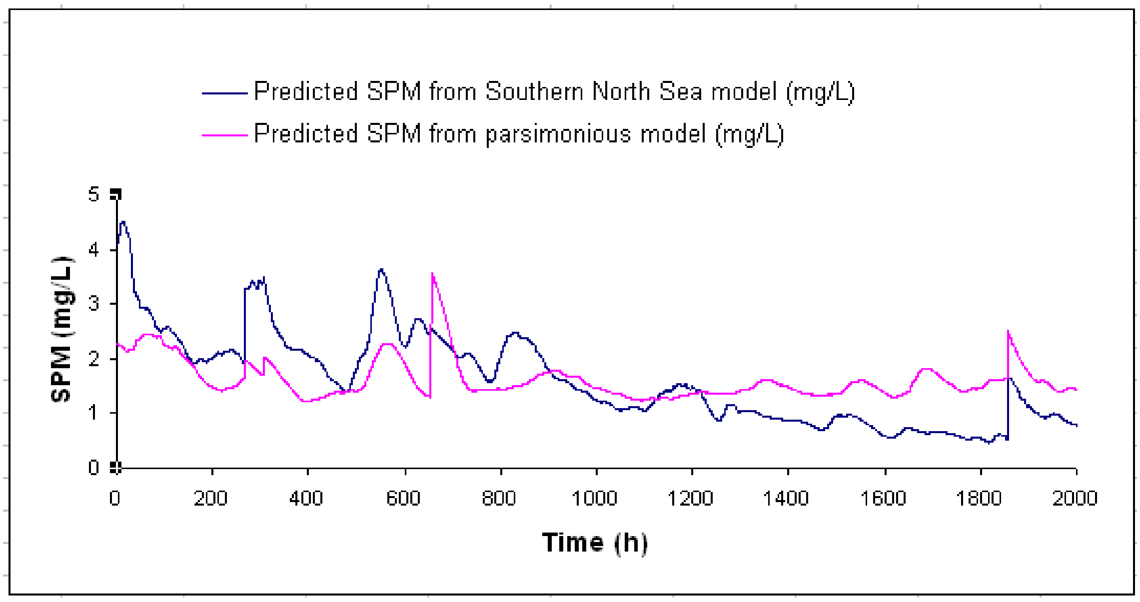

Figure 4.18 Testing results of parsimonious model for non-stormy periods at Noordwijk 30 km.

4.3.5 Parsimonious model at Noordwijk 60 km (water depth = 27.4-27.8m)

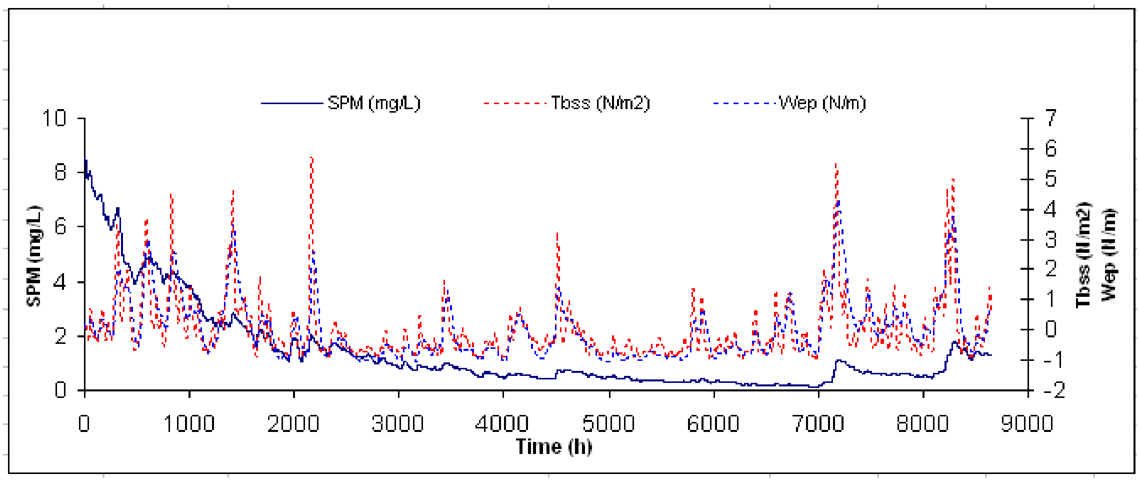

Figure 4.19 Plot of SPM between Tbss and Wep at Noordwijk 60km.

Figure 4.19 shows the plot of SPM between Tbss and Wep at Noordwijk for all data. SPM has a decreasing trend along time while Tbss and Wep have frequent changes. SPM and Tbss are unconnected for most of the instances. So it is meaningless to cut stormy and non stormy periods. It is difficult to build a parsimonious model at Noordwijk 60km.

4.4 Conclusions

In this chapter, a parsimonious model is firstly built at IJmuiden. In order to explore the influence on SPM due to local water depth, a series of parsimonious models are then built at Noordwijk 1.5km, 3/5km, 10km, 30km and 60km respectively.

For stormy periods, parsimonious models perform very well for most of the single storms at IJmuiden, which implies that SPM has a strong relationship with Tbss and Wep for a single storm. However, parsimonious model for all storms shows failure to predict SPM. The main reason is that SPM covers a large range, varying from about 10 to 44 mg/L in stormy periods. Therefore, the variation of SPM is respectively large. It is difficult for parsimonious model to capture the changing of SPM in a linear way. This is also supported by quite different values of coefficient *a, b* and *c* for each storm. In order to build a parsimonious model to predict SPM in stormy periods, storms with poor model performance are discarded. Then 7 storms are used to build the model and 5 storms are used as testing data. Testing results show that this parsimonious model can be applied for 4 of the testing storms to predict SPM. The unsatisfactory of parsimonious model performance for a single storm may due to other processes influencing SPM concentration or errors of process-based model output.

Parsimonious model is also viable at Noordwijk 1.5 km where a similar water depth with IJmuiden has. At other locations further offshore, performance of parsimonious model is decreasing, even for most of the single storm. The critical reason is that at locations with deeper water depth, SPM generated by Tbss will be advected by horizontal exchange, together with the diffusion along the vertical direction, the SPM near the bottom is lower and the changing of SPM is not very dramatic while the Tbss is still large and changing frequently. The relationship of SPM and Tbss is distorted, which leads to the failure of building parsimonious model for stormy periods at these locations.

For non-stormy periods, SPM ranges from about 5 to 20mg/L at IJmuiden, belonging to a narrower range. The results show the possibility to build parsimonious model in a linear way to predict SPM for non-stormy periods at IJmuiden. With increasing the water depth, model performance has a decreasing trend. The correlation between SPM predicted by parsimonious model and the Southern North Sea model for non-stormy periods is 0.86 at IJmuiden and drops to 0.72, 0.68, 0.58 and 0.55 at Noordwijk 1.5km, 3/5km, 10km and 30km respectively.

At Noordwijk 60km, the relationship of SPM and Tbss is quite unclear. It shows impossibility to build parsimonious model in such locations with deep water depth.

The applicability of parsimonious model for SPM prediction in the Southern North Sea area could be investigated further at more locations.

Chapter 5 Conclusions and recommendations

The main objective of this study is to improve the prediction and understanding of fine sediment processes in the Dutch coastal zone. By achieving this purpose, surrogate models were built in Chapter 3 and parsimonious models were built in Chapter 4. By analyzing the results, the following conclusions have been drawn:

- Results have shown a strong possibility of building the surrogate model with artificial neural networks to predict SPM concentration.

  In order to improve the model performance, data used for building surrogate model were smoothed out by data filter; correlation and AMI analysis were used to choose the most appropriate input variables; data are transformed to zero mean and unit variance to adjust the data distribution to normal. Model performance is enhanced after processing data. In order to capture the seasonal effect on SPM concentration, stormy and non-stormy periods were defined in this study and models were built for them separately. Predictions of SPM were improved significantly. Sensitivity analysis was applied to invest the model performance with different input datasets.

- Parsimonious surrogate modelling was built with linear regression method for predicting SPM at IJmuiden. By exploring how SPM is affected by local water depths, a series of parsimonious models were built at Noordwijk with different distances to the coast.

  For stormy periods, parsimonious models perform very well for most of the single storms at IJmuiden, which implies that SPM has a strong relationship with Tbss and Wep for a single storm. However, parsimonious model for all storms shows failure to predict SPM due to the large variation of SPM for stormy periods. Therefore, it is difficult for parsimonious model to capture the changing of SPM in a linear way. Testing results showed the possibility to predict SPM for a single storm by parsimonious model. With increasing of water depth, the parsimonious model performance was reduced. This is mainly because the SPM generated by Tbss is diffused and advected. Thus, the relationship of SPM and Tbss is distorted to illegibility.

  For non-stormy periods, parsimonious model performed acceptable at IJmuiden. The prediction of SPM became weak with increasing of the water depth due to the diffusion of SPM.

The following points are recommended in the future study:

- Wep was introduced as an input variable in this study and sensitivity analysis showed the effectiveness of Wep to the model performance. In this study, we assumed that waves within the last two weeks (336 hours) give effects to the current SPM concentration and no influence any more of the wave energy before two weeks ago. Furthermore, Wep is a weighted sum of wave energy and the weights are determined by a linear function. However, these two assumptions need to be investigated more carefully. An optimization approach could be applied for future study.

- Stormy and non-stormy periods were defined in an arbitrary way in this study. This definition may show failure at other locations or using data from other years. The critical values for defining stormy and non-stormy (1m and 2.5m in this study) are flexible; values could be changed if the definition is quite different from the reality.

- Surrogate model performance could be enhanced by selecting more parameters in addition to Tbss and Wep.

- Due to the characteristics of fine sediment, salinity could be another significant effect to the distribution of SPM concentration. This can be introduced to the surrogate model in the future.

- The applicable of parsimonious model in the Southern North Sea could be studied further in more locations.

References

Abarbanel, H. (1996). *Analysis of observed chaotic data*. Springer, New York.

Abebe, A.J. and Price, R. K. (2004). Information theory and neural networks for managing model uncertainty in flood routing. *J. Computing in Civil Eng., ASCE,* 18(4), 373-380.

Battjes, J.A., Jansen, J.P.F.M. (1978). Energy loss and set-up due to breaking of random waves. *Proc. 16th Int. Conf. on Coastal Engineering. ASCE,* New York. 570-587.

Bhattacharya, B., Lobbrecht, A.H., and Solomatine D.P. (2003). Neural networks and reinforcement learning in control of water systems, *J. of Water Resources Planning and Management,* ASCE, 129(6), 458-465.

Bhattacharya, B., Price, R. K. and Solomatine, D. P. (2004). A data mining approach to modelling sediment transport. *Proc., 6th Int. Conf. of Hydroinformatics*, World Scientific, Singapore, 1663–1670.

Bhattacharya, B., Price, R. K. and Solomatine, D. P. (2005). Datadriven modelling in the context of sediment transport. *Phys. Chem. Earth*, *Part A Solid Earth Geod*., 30(4–5), 297–302.

Bhattacharya, B. (2005b). *Learning from Data for Aquatic and Geotechnical Environment.* Taylor & Francis Group plc, London, UK.

Bhattacharya, B., Deibel, I. K., Karstens, S. A. M. and Solomatine, D. P. (2006). Neural networks in sedimentation modelling for the approach channel of the port of Rotterdam, J. Maa, L. Sanford, and D. Schoellhamer (eds.). *Estuarine and coastal fine sediment dynamics,* (Marine Sciences, Vol. 8)., Elsevier, 465-480.

Deltares (2006). *Delft3D user manual*, Delft, The Netherlands.

Dronkers, J. (2005). *Dynamics of Coastal System*. Advanced Series on Ocean Engineering. Volume 25. World Scientific.

Fredsøe, J.F. (1984). Turbulent boundary layer in wave-current interaction. *J. Hydraul. Eng* 110: 1103 - 1120.

Hasselmann, K., T.P. Barnett, E. Bouws, H. Carlson, D.E. Cartwright, K. Enke, J.A. Ewing, H. Gienapp, D.E., Hasselmann, P. Kruseman, A. Meerburg, P. Müller, D.J. Olbers, K. Richter, W. Sell and H. Walden (1973). Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP). *J. Hydrogr,* 12, A8.

Haykin, S. (1999). *Neural networks: a comprehensive foundation*. Prentice Hall.

Jin, K. K., and Ji, A.G. (2004). Case Study: Modelling of sediment transport and wind-wave impact in Lake Okeechobee. *Journal of Hydraulic Engineering*, 130 (11),1055-1067.

Kleijnen, J. P. C. (1975). A comment on blanning's meta-model for sensitivity analysis: The regression meta-model in simulation. *Interfaces*, 5(1): 21-23.

Lane, E. W. (1947). Report of the Subcommittee on Sediment Terminology, Transactions, American Geophysical Union. Volume. 28, Washington, DC. 936-938.

Li, W. (2007). Modelling of Cohesive Sediment Transport in the Maasmond Area. UNESCO-IHE Institute for Water Education, Delft.

Lin, B., and Namin, M. M. (2005). Modelling suspended sediment transport using an integrated numerical and ANNs model. *J. Hydraul. Res.,* 43(3), 302–310.

Liong, S.Y., ShreeRam. J. and Chan. W.T. (1995) Catchment calibration using fractional factorial and central composite designsbased response surface, *J. Hydr. Engrg., 121(8)*, 613-617.

McCulloch, W.S., and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115-137

Mitchell, Y.M., (1997). Machine learning, *McGraw-Hill*.

Nagy, H. M., Watanabe, B. and Hirano, M. (2002). Prediction of sediment load concentration in rivers using artificial neural network model. *J. Hydraul. Eng.,* 128(6), 588–595.

Nelson, S.A. (2007). Coastal Zones. http://www.tulane.edu/~sanelson/geol204/coastalzones.htm Cited: April, 2009.

Pyle, D. (1999). *Data preparation for data mining*. Morgan Kaufmann, California.

Riddle, A.M., Lewis, R.E. and Sharpe, A.D. (2004). Environmental quality information processor (EQUIP). *Env. Modelling & Software.,* 19(1), 57-62.

Roelvink, J. A., van der Kaajj, T. and Ruessink, B.G. (2001). Calibration and verification of large scale 2D/3D flow models - Phase 1. Deltares, Delft.

Rosenbaum, M. (2000). Harbors-silting and environmental sedimentology(H-SENSE). *Final Rep., Dept. of Civil and Structural Eng.* The Nottingham Trent Univ., Nottingham, U.K. http://hjs.geol.uib.no/Hsense Cited: April. 2009.

Schalkoff, R.J. (1997). *Artificial Neural Networks*. McGraw-Hill companies. Inc.

Schulz, E.F., R.H Wilde. and M.L. Albertson. (1954). Influence of Shape on the Fall Velocity of Sedimentary Particles. *MRD Sediment Series No, 5.*

See, L. A., Solomatine, D. P., Abrahart, R. and Toth, E. (2007). Hydroinformatics: computational intelligence and technological developments in water science applications – Editorial. *J. Hydrol. Sci.,* 52 (3), 391–396.

Simons, D. B. and Senturk. F. (1992). *Sediment Transport Technology -Water and Sediment Dynamics*. Water Resources Publications. 146-147.

Sistermans, P. and Nieuwenhuis, O. (2002). Holland Coast (The Netherlands). http://copranet.projects.eucc-d.de/files/000139_EUROSION_Holland_coast.pdf. Cited: April. 2009.

Solomatine, D. P. and Torres, L. A. (1996). Neural network approximation of a hydrodynamic model in optimizing reservoir operation. *Proc. 2nd Int. Conf. of Hydroinformatics,* Balkema, Rotterdam. 201–206.

Solomatine, D.P. (2006). Data-Driven modelling: machine learning, data mining and knowledge discovery. UNESCO-IHE Institute for Water Education, Delft.

Solomatine, D. P. and Ostfeld. A. (2008). Data-driven modelling: some past experiences and new approaches. *J. Hydrinformatics*.

Soon, T. K., Savic. D., Liu. Y. and Madsen. H. (2004). A fast Evolutionary-based Meta-Modelling Approach for the Calibration of a Rainfall-Runoff Model. http://www.iemss.org/iemss2004/pdf/evocomp/khuafas.pdf Cited: Oct. 2008.

Sower, S. (2008). Hybrid Approach to Combine Physically Based and Data-Driven Models in Simulating Sediment Transportation. UNESCO-IHE Institute for Water Education. Delft.

Stolk, A. (1989). Zansysteem kust – morfologische karakterisering. *Kustverdediging na 1990, Report GEOPRO 1989-02, Department of Physical Geography*, Utrecht University.

Tatang, M.A., Pan. W., Prinn, R.G. and McRae. G.J. (1997). An efficient method for parametric uncertainty analysis of numerical geophysical models. *J. Geophysical Res., 102*, 21925-21932.

Tchouani, N.J.M. (2004). Graded Sediment Transport Modelling. UNESCO-IHE Institute for Water Education. Delft.

Van Kessel, T., Winterwerp J.C., Van Prooijen, B.C., Van Ledden, M., and Borst, W.G., (2007). Modelling the seasonal dynamics of SPM with a simple algorithm for the buffering of fines in a sandy seabed. *Pro., INTERCOH'O7*, Brest France.

Van Rijn, L.C. (1993). Principles of Sediment transport in Rivers, Estuaries, Coastal Seas and Oceans. *Lecture Notes, HH279/93/1,* UNESCO-IHE Institute for Water Education. Delft.

Van Rijn, L.C., Ruessink, B.G., Mulder, J.P.M. (2002). The behavior of a straight sandy coast on the time scale of storms and seasons. *EC MAST Project No. MAS3-CT97-0086.*

Vriend, H.J.D. (1991). Mathematical modelling and large-scale coastal behaviour. *J. Hydraul Res.,* 29, 727-740.

Wikipedia http://en.wikipedia.org/wiki/Parsimonious Cited: April, 2009.

Winterwerp, J.C. (2006) *Fluxes of fine sediment along the Dutch coast and the impact of Maasvlakte 2: a system description*.

Ye, Q. (2006). Modelling of Cohesive Sediment Transportation, Deposition and Resuspension in the Haringvliet Mouth. UNESCO-IHE Institute for Water Education, Delft.