# Spatio-temporal prediction of suspended sediment concentration in the coastal zone using an artificial neural network and a numerical model

B. Bhattacharya, T. van Kessel and D. P. Solomatine

## ABSTRACT

A problem of predicting suspended particulate matter (SPM) concentration on the basis of wind and wave measurements and estimates of bed shear stress done by a numerical model is considered. Data at a location at 10 km offshore from Noordwijk in the Dutch coastal area is used. The time series data have been filtered with a low pass filter to remove short-term fluctuations due to noise and tides and the resulting time series have been used to build an artificial neural network (ANN) model. The accuracy of the ANN model during both storm and calm periods was found to be high. The possibilities to apply the trained ANN model at other locations, where the model is assisted by the correctors based on the ratio of long-term average SPM values for the considered location to that for Noordwijk (for which the model was trained), have been investigated. These experiments demonstrated that the ANN model's accuracy at the other locations was acceptable, which shows the potential of the considered approach.

**B. Bhattacharya** (corresponding author)
**D. P. Solomatine**
UNESCO-IHE Institute for Water Education,
Westvest 7,
Delft,
The Netherlands
E-mail: B.Bhattacharya@unesco-ihe.org

**T. van Kessel**
Deltares,
Rotterdamseweg 185,
Delft,
The Netherlands

**D. P. Solomatine**
Water Resources Section,
Delft University of Technology,
Delft,
The Netherlands

**Key words** | ANN, data-driven modelling, Dutch coast, filtering, neural network, spatio-temporal prediction, SPM, suspended sediment

## INTRODUCTION

Sediments are important in many aquatic systems. Their transportation and deposition has significant implications in morphology, navigability and water quality. Understanding the dynamics of sediment transportation in time and space is therefore important in drawing interventions and making management decisions. This research is related to the fine sediment dynamics in the Dutch coastal zone, which is subject to human interference through constructions, fishing, navigation, sand mining, etc. These activities do affect the natural flow of sediments and sometimes lead to environmental concerns or affect the siltation rates in harbours and fairways.

Numerical models are frequently employed in order to predict suspended particulate matter (SPM) concentration and siltation rates. A number of such models have been used for the Dutch coast (see, for example, van Kessel *et al.* 2007). While these models, in general, serve the

purpose of a broad-based study of the processes involved, their performance is often not optimal. A drawback of these process-based models is the long computing time which is usually required due to the long residence time of fine sediments in the Dutch coastal area and/or due to the requirement of a fine grid. A large domain size is often needed to account for the length scale related to the area where sediment dynamics is affected by the discharge of the River Rhine.

To account for the full natural variability of the system, simulation periods of a few years would be required, leading to even longer computation times. Therefore, the meteorological variability is often schematised within a much shorter period, which is assumed to be representative for the complete period. In such a way, computation times remain within feasible limits, but at the cost of some realism. In addition, although the model formulations do include the

most prominent features of the physical processes occurring, they remain simplifications, especially regarding the water–bed exchange of sediments.

In this respect there is a scientific challenge in looking for alternative approaches to complement numerical modelling of sediment dynamics in order to provide estimates of suspended sediments over a short period of time (storm). Nagy *et al.* (2002), Lin & Namin (2005), Bhattacharya & Solomatine (2006), Bhattacharya *et al.* (2006, 2007) and others have explored the use of data-driven models to predict the temporal pattern of SPM at a location using the past data collected at the same location.

Similar research in other domains, for example in hydrological modelling, has clearly demonstrated the value of data-driven models (ASCE 2000; Solomatine & Ostfeld 2008). Nonetheless the applicability of a data-driven model depends to a large extent on the quality and quantity of the data used to build the model. As data are scarce, particularly for suspended sediments and related variables, there is a need to explore all possible sources of information, including data generated by numerical models.

An additional limitation of the numerical model employed to study the Dutch coastal area is the use of a fixed sediment boundary condition. SPM is not measured so often to be used in fine spatial grids as boundary conditions of numerical models. The influence of the use of a fixed sediment boundary condition is not clearly known. This leads to an additional challenge to address the known limitation of data-driven models to lose predictive power when conditions or location change.

In this paper a data-driven model to predict SPM concentrations is presented and a methodology to adapt the model for other locations is reported.

## DATA-DRIVEN MODELLING

Data-driven modelling (DDM) is based on the analysis of the data characterising the system under study. A model can then be defined on the basis of connections between the system state variables (input, internal and output variables) with only a limited number of assumptions about the 'physical' behaviour of the system (Solomatine & Ostfeld 2008). Examples of the most common methods used in DDM of

water systems are: statistical methods, artificial neural networks (ANN), support vector machines and fuzzy rule-based systems. The main part of DDM is, in fact, learning; it incorporates determining the, so far unknown, mappings (or dependences) between a system's inputs and its outputs from the available data (Mitchell 1997). As such a dependency (model) is discovered (induced), it can be used to predict (or deduce) the future system's outputs from the new input values.

By data we usually understand a set $K$ of examples (or data points) represented by the duple $\langle \mathbf{x}_k, \mathbf{y}_k \rangle$, where $k = 1,..., K$, vector $\mathbf{x}_k = \{x_1,..., x_n\}_k$, vector $\mathbf{y}_k = \{y_1,..., y_m\}_k$, $n =$ number of inputs and $m =$ number of outputs. The process of building a function $\mathbf{y} = f(\mathbf{x})$ is called *training*. Typically the structure of $f$ is chosen in advance so training is aimed at finding parameters of $f$ that would lead to minimum model error on the training or cross-validation set. Very often only one output is considered, so $m = 1$. Inputs and outputs are typically real numbers ($\mathbf{x}_k \in \mathfrak{R}^n$, $y_k \in \mathfrak{R}^m$), so the learning problem solved in this case is numerical prediction (regression).

The choice of model variables is an important issue. Apart from expert judgement and visual inspection, there are formal methods that help in justifying this choice, and the reader is directed to the papers by Bowden *et al.* (2005) and May *et al.* (2008) for an overview and applications of these methods. Note that the input data may require pre-processing (e.g. normalisation, filtering to remove noise, etc.) and this may increase the total number of possible inputs (and their combinations) to consider. In the case of a high number of inputs, methods of dimensionality reduction, such as principal component analysis, may help.

It is worth mentioning that DDM is sometimes used to build models of models (replicating, for example, physically based models such as 1D hydrodynamic models) rather than models of natural systems. Such models are often referred to as *surrogate*, *emulation* or *meta-models* (see e.g. Chua & Holz 2005).

ANN is the most widely used method in DDM. ANN is a broad term covering a large variety of network architectures, the most common of which is a multi-layer perceptron (MLP). Such a network is trained by the so-called error-backpropagation method, which is typically based on one of the gradient-based optimisation algorithms.

In ANN each target vector $y$ is an unknown function $f$ of the input vector $x$:

$$y = f(x). \tag{1}$$

The task of the network is to learn the nonlinear regression function $f$. The network includes a set of parameters (weight vector), the values of which are varied to modify the generated function $f'$, which is computed by the network to be as close as possible to $f$. The weight parameters are determined by training (calibrating) the ANN based on the training dataset. More details about ANNs can be found in Haykin (1999).

DDM (particularly ANN) is actively used in various water-related areas such as sediment transport modelling (Bhattacharya et al. 2007), wind–wave modelling (Zamani et al. 2008), etc. Especially popular ANN were in hydrological applications: in rainfall–runoff modelling (ASCE 2000), prediction of discharge (Muttiah et al. 1997), in modelling the stage behaviour (without considering discharge) (Thirumalaiah & Deo 1998) and in modelling stage–discharge relationships (Tawfik et al. 1997; Jain & Chalisgaonkar 2000).

Compared to the domain of hydrology the number of applications of ANN in the coastal domain is much less. Tayfur (2002) applied ANN in predicting sheet sediment transport. Cigizoglua & Kisi (2006), Cobaner et al. (2009) and Kisi (2010) used a variety of DDM techniques, including ANN, in predicting suspended sediment concentrations in rivers. Jain (2001) used ANNs to build sedimentation rating curves. Nagy et al. (1999) used ANN in predicting sediment load in alluvial rivers. Singh et al. (2008) used ANN in predicting littoral drift along a 4 km long stretch of the west coast of India.

## PREDICTING SPM AT NOORDWIJK-10

### Location and data

In this section we present building an ANN model to predict SPM at Noordwijk on the Dutch coast (Figure 1). The choice of this particular location was due to the availability of a relatively larger dataset. The data at this location were
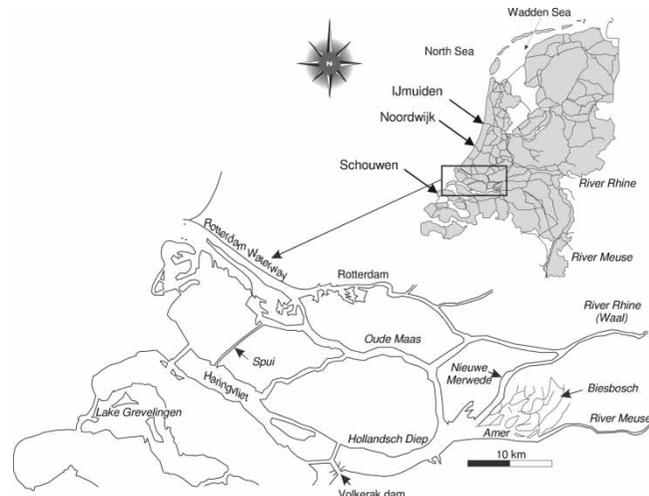


**Figure 1** │ Location map showing Noordwijk and other places.

collected under a special data measurement campaign organised by the Centre for Environment, Fisheries and Aquaculture Science (CEFAS), UK, in collaboration with the National Institute for Coastal and Marine Management (RIKZ) of The Netherlands (Mills 2002). The measuring instruments were mounted on a buoy (known as a Smart Buoy), which, among other devices, used an Optical Back Scatter (OBS) device (van de Kreeke & Hartog 2004). The OBS device emits a beam of light and the intensity of light after a short distance is measured to ascertain the amount of suspended sediment present in the water. The device needs careful calibration so that the fall in intensity of light can be reliably used in estimating the SPM concentration. Usually, the calibration results are reliable, particularly during the winter (van de Kreeke & Hartog 2004).

Using the Smart Buoy, hourly SPM data were collected about 1 m below the water surface at 10 km offshore of Noordwijk during the period 8/3/2000 to 18/9/2001 (water depth was 18 m). On some occasions the data were missing; in total for about 20% of occasions. Missing data were not estimated; rather the time slots with missing measurements were removed. Table 1 shows the available data at Noordwijk-10. In addition to the SPM data the Smart Buoy collected wave data (significant wave height, wave period and wave direction). Vos (2004), Suijlen & Duin (2001) and van de Kreeke & Hartog (2004) provided further analysis of the data measured at Noordwijk.

**Table 1** | Overview of data available at Noordwijk-10

| Data type | Frequency | Data period | Total data points | Missing data |
|---|---|---|---|---|
| Suspended particulate matter (SPM) | Hourly | 8/3/2000 to 18/9/2001 | 10,788 | 2,628 (about 20%) |
| Wave | Hourly | 8/3/2000 to 18/9/2001 | 13,416 | 0 |
| Wind | Hourly | 8/3/2000 to 18/9/2001 | 13,416 | 0 |

SPM concentration is mostly influenced by bed shear stress. Therefore, bed shear stress data were needed to build a data-driven model to predict SPM. The bed shear stress was computed using a 1DV point model (van Kessel *et al.* 2007) representative of the water depth, wave and current conditions at Noordwijk-10. The 1DV model is a closed-box model neglecting horizontal advection and diffusion. Wave-induced bed shear stress was computed from the observed significant wave height and wave period using the formulations by Swart (1974) and Soulsby *et al.* (1993). Current-induced bed shear stress was taken from an overall 3D hydrodynamic model of the southern North Sea.

## Choice of variables, data analysis and transformation

Based on the physics of the considered phenomena, we considered wave height, wind and the bed shear stress to be the most relevant variables for predicting SPM, so they were used in the initial experiments with regression analysis and ANN. However these experiments showed that without data transformation the model's performance was unacceptably low. This prompted additional data analysis.

The SPM concentration varies at different timescales due to the influences of a number of processes. The contribution of fluvial sediments coming from the River Rhine varies with the river discharge. The larger the discharge, the larger is the sediment load, which then travels north along the coast and influences sediment concentrations at Noordwijk. Biological activity can initiate flocculation, which can increase the settling velocity and thus affects the sediment dynamics. During the tranquil periods of summer, flora and fauna cause consolidation of the deposited sediment over the sea bed, raising the impermeability substantially. This causes easier erosion of sediments during the early summer than during the early winter. The vertical mixing of sediment also depends upon the spring–neap tides, leading to higher mixing during the spring tides than during the neap. At a much shorter timescale SPM concentration varies with the tidal current speed. From slack tide it increases gradually to a peak with the increase in the tidal current speed. Alongshore advection of sediments passing the measurement station also influences the concentration.

The analysis of the SPM concentrations with a classical statistical method such as spectral analysis is possible for the calm periods (Vos 2004). The variation in concentration during a storm is a transient phenomenon and, as a result, is not well suited for spectral analysis. Analysis of SPM data, particularly during the storm periods, showed lots of short-term fluctuations (Figure 2), making it imperative that building an accurate data-driven model to predict
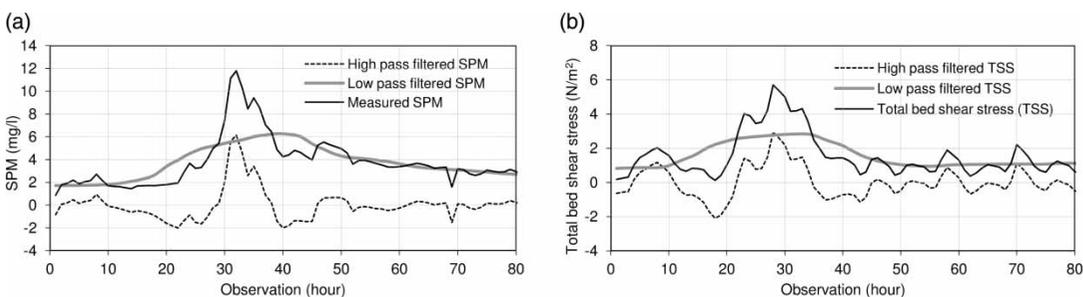


**Figure 2** | (a) Variation of suspended particulate matter (SPM) concentration at 10 km offshore of Noordwijk for a storm during 16–19 May 2001 along with the low pass and high pass filtered SPM concentration for the same storm. (b) Variation of total bed shear stress at 10 km offshore of Noordwijk for the same storm period of (a) along with the low pass and high pass filtered total bed shear stress.

short-term changes in SPM is difficult. However, from the viewpoint of the management of coastal waters it is not so important to accurately predict hourly concentrations of SPM. What is important is to be able to predict the trend, total sediment during a storm, seasonal fluctuations and average values, particularly during storms.

Due to the mentioned problems with the initial models, we tested several methods of data transformation that would allow the models to reveal the trend in the SPM data. Finally this was accomplished by applying a low pass filter to SPM as well as bed shear stress data; such a filter passes low frequency signals while it attenuates the signal with frequencies higher than the adopted cutoff frequency (we used a 24 h moving average as such a filter). The concept of low pass filtering is ubiquitous in signal and image processing. Note that the resulting filtered time series for SPM and bed shear stress is then void of short-term fluctuations, whereas the general pattern of the original time series is captured (Figure 2). While working with the filtered time series we therefore do not cater for some processes which occur at hourly timescales such as tides (except that the spring–neap effect of the tides is still present). The residue, which can be found by subtracting the low pass filtered data from the original time series, which can be called as the high pass filtered data, exhibit a higher level of fluctuations (Figure 2).

The low pass filtered data for both SPM and bed shear stress are much smoother and present a comparatively easier problem for modelling. It may be noted that, while predicting SPM over a period of 1–2 d, the low-pass filtered time series relate to more than 90% of the total sediment of the original time series. As an example, for the storm during 16 May to 19 May 2001 the average total SPM and the filtered SPM were 3.77 and 3.75 mg/l, respectively (Figure 2). The filtered SPM during the entire storm was 99.3% of the total SPM during the storm. Though the hourly variation between the filtered SPM and total SPM is significant, but over a storm period the former time series is equivalent to the latter one. Domain experts were of an opinion that prediction of the filtered SPM is adequate from the operational management point of view.

Figure 3 shows the variation of filtered SPM *vis-à-vis* the filtered bed shear stress for a storm. A high correlation (0.93) exists between the two time series. Data analysis showed the
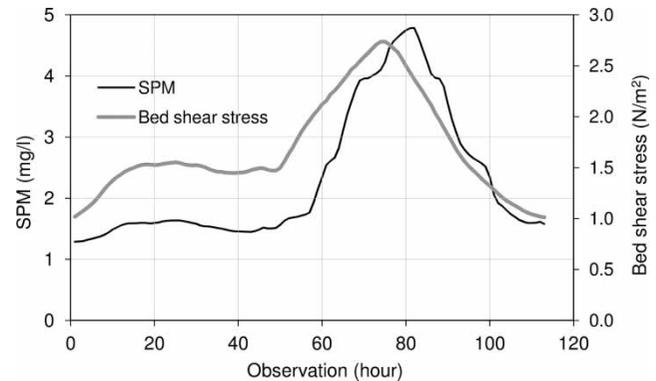


**Figure 3** | Variation of low pass filtered suspended particulate matter (SPM) concentration along with low pass filtered total bed shear stress at 10 km offshore of Noordwijk for a storm during 8–13 July 2001.

optimum lag between the filtered SPM and the filtered bed shear stress to be 10 h. Similar correlation between the filtered bed shear stress and the 10 h lagged filtered SPM was observed for other storms as well. However, when the time series of the filtered SPM and the filtered bed shear stress were constructed from several storm periods together then the correlation dropped to 0.63 (Figure 4). This implies that the response of bed shear stress can be different owing to different ambient conditions.

The sea bed consists of sand, and the fine sediments are trapped in the sand layer. When a storm comes the fine sediments are lifted from the sea bed. The release of fine sediments depends not only on the bed shear stress but also on the availability of fines in the sand layer and the consolidation of the bed. After a long tranquil period the bed
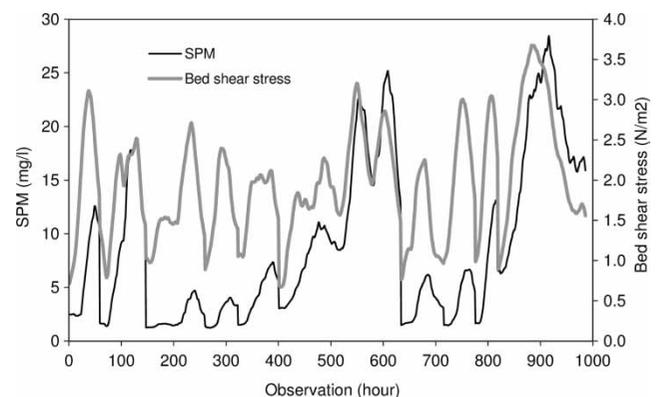


**Figure 4** | Variation of filtered SPM and filtered bed shear stress for multiple storms.

consolidates and during a storm the fine sediments are not released so easily. Also, biological activity may alter consolidation. After a storm the fines in the sand bed remain in a relatively loose state, making lifting of fines under a following storm much easier. With several successive storms the fines may be depleted as well.

Significant wave height series was also smoothed by averaging over a week. The correlation between this variable and the filtered SPM was not very high (0.66). This is not surprising as the process is highly nonlinear so the correlation coefficient (denoting a linear relationship) may not be the best statistic to express the relationship.

The wind direction and wind speed also have an influence on SPM concentrations. Stronger wind causes larger waves and higher bed shear stress. The additional effect wind may cause is in bringing parts of the sediments from the dredged spoil disposed off Noordwijk. Experience shows that south-westerly wind may bring additional sediments from dredged spoil off Noordwijk (de Kok 1994).

Based on the above discussion the following input variables were chosen for the DDM:

- Filtered bed shear stress for three (hourly) time steps: $\tau_{t-10}$, $\tau_{t-11}$, $\tau_{t-12}$ where $\tau_{t-10}$ means the filtered bed shear stress at 10 h in the past from time $t$.
- Significant wave height averaged over the last 7 days ($H_{\text{sig}, t}$).
- South-westerly wind component ($W_{\text{SW}, t}$).

The output of the model was:

- Filtered SPM ($\text{SPM}_t$).

It may be noted that, on most occasions when a data-driven model is built, the output variable of the previous time step(s) is used as an input. Modelling experiments with using $\text{SPM}_{t-1}$ as an input showed much better performance of the model. In similar cases the additional accuracy comes from the additional information from the extra input variable and the fact that output variables usually have a high autocorrelation. However, in our case one of the objectives was testing a possibility to use the trained model at different locations to predict $\text{SPM}_t$ and we may not have measured SPM values at those locations. Accordingly, the SPM of previous time steps was not used as an input.

All the input time series were converted to series with zero mean and unit variance. The output time series was not converted as we planned to use this model at other locations as well, and at those locations we may not have precise information about the mean and particularly the variance of SPM.

## Experimental set-up and results

The data were split into training (6,000 data points), testing (3,800 data points) and cross-validation (1,200 data points). In the data matrix with all the instances for all the variables an additional column with a random number was added and the data were sorted on the random number. The statistical properties of the different variables show a good match over the different datasets (Table 2).

The software NeuroSolutions (2010) was used for the ANN modelling. A multi-layered perceptron network trained by the back-propagation algorithm with the momentum rule (momentum = 0.7), a hyperbolic tangent function in the hidden layer nodes and a linear transfer function at the output layer nodes were used. A learning rate of 0.1 was used.

The number of nodes in the hidden layer was found by exhaustive search by minimising the error on a cross-validation set; it was found to be eight. Training was performed until the moment when the network error on the cross-validation set started to increase (a standard practice

**Table 2** | Statistical properties of the training, testing and cross-validation datasets

|  | Suspended particulate matter (filtered) (mg/l) | | Bed shear stress (filtered) (N/m²) | | Significant wave height (filtered) (m) | | South-westerly wind (m/s) | |
|---|---|---|---|---|---|---|---|---|
|  | Mean | Standard deviation | Mean | Standard deviation | Mean | Standard deviation | Mean | Standard deviation |
| Training | 4.60 | 4.20 | 1.02 | 0.73 | 0.98 | 0.39 | 1.83 | 5.86 |
| Testing | 4.69 | 4.23 | 1.03 | 0.73 | 0.98 | 0.38 | 1.84 | 5.83 |
| Cross-validation | 4.65 | 4.31 | 1.07 | 0.74 | 0.99 | 0.40 | 1.83 | 5.86 |

to prevent overfitting (Principe *et al*. 2000)). After determining the best number of hidden nodes the model was trained for 1,000 epochs for three times to find the best weights. The resulting root mean square error (RMSE) on the testing dataset was 3.92 mg/l whereas $R^2$ was 0.79.

From the viewpoint of operational management of ports and harbours accurate prediction of SPM concentrations is of paramount importance during stormy weather. This is due to the reason that SPM concentrations can be substantially higher during storms than during calm periods. Higher SPM concentrations may lead to settling of particles in the deep water channel leading to hindrances in navigation and the mobilisation of dredger(s) may be necessitated. Accordingly, the test data were split into two groups: belonging to storm periods and belonging to calm periods. Storm periods were defined as the time window with significant wave height starting with 1.0 m, gradually reaching a peak of at least 2.5 m and ending with 1.0 m. Other periods were considered as calm periods. This certainly is an arbitrary definition based on the visual inspection of the data. The sensitivity of this definition has not been ascertained and calls for further research. The RMSE of the model's prediction for the stormy periods (with 272 data points) was 9.77 mg/l and for the calm periods (with 3,528 data points) was 3.03 mg/l. The $R^2$ values were 0.67 and 0.64, respectively.

The error on the test data with stormy periods was considered to be quite high. In order to reduce the error it was decided to switch to modular modelling; for that, both the training and testing data were split into stormy and calm periods, and separate ANN models were built for both of them. The model with the data for stormy periods was trained with the training data (715 data points) using a similar topology mentioned above. The number of hidden nodes was again found to be eight. On the training and testing the RMSE was 2.81 and 2.08 mg/l (whereas the average SPM was 7.25 mg/l) and $R^2$ was 0.86 and 0.94. Figure 5(a) shows the prediction of the SPM concentrations *vis-à-vis* the measured concentrations during a stormy period over a week. It can be concluded that the pattern of SPM concentrations is predicted well by the ANN model.

Similarly, the ANN model for the calm period was trained with 5,285 data points and was tested with 3,528 data points. The ANN had six hidden nodes. The RMSE in training and testing was 2.83 and 3.29 mg/l whereas the $R^2$ values were 0.77 and 0.73, respectively. The ANN model trained with the storm period data was also used in predicting SPM concentrations for the calm periods. The RMSE was observed to be 3.39 mg/l ($R^2$ was 0.66), which is slightly higher than the RMSE of the ANN model for the calm periods. As accuracy of model predictions during the calm periods is not as important as for the stormy periods so we decided to use the ANN model built with the data for the storm periods also for the calm periods. This has the benefit of having just a single model at the cost of a slight reduction in accuracy. Figure 5(b) shows the prediction of the SPM concentrations together with the measured concentrations during a calm period of about two weeks. Though errors in prediction, particularly for the peak, is discernible still given the uncertainties of prediction of SPM concentrations it can be concluded that the general pattern of the concentration is predicted well.
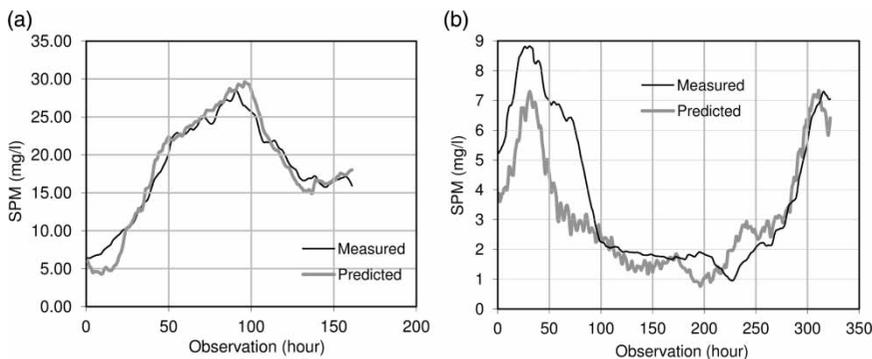


Figure 5 | Comparison of SPM concentration predicted by the ANN model with the measured data. (a) During a storm period during 7–14 September 2001 (about a week). (b) During a calm period during 10–24 February 2001 (about two weeks).

## PREDICTING SPM AT OTHER LOCATIONS

As has been mentioned in the introduction, a well-known limitation of data-driven models is in the loss of accuracy or even applicability at locations different from where the training data were collected. In the following section we present the results of testing the ANN model for Noordwijk-10 with data from other locations.

### Predicting SPM at Noordwijk-2

CEFAS and RIKZ also collected SPM and wave data at 2 km offshore of Noordwijk (Noordwijk-2) during the period 18/9/2001 to 28/12/2001. This location in terms of processes is similar to Noordwijk-10. The ANN model developed using the data of Noordwijk-10 was used to predict the SPM concentrations at Noordwijk-2. An alternative could have been to develop another ANN model with the data of Noordwijk-2. However, one research objective was to investigate the possibility of using the ANN model at locations other than the one for which it was developed. This has been shown for the locations Schouwen and IJmuiden (the following two subsections) for which there was really not enough data. Additionally, we also investigated this approach for Noordwijk-2 for which we do have enough data.

The bed shear stress data were generated using the numerical model. The wave data were taken from the measurements at Noordwijk-2. In order to account for the differences in responses between Noordwijk-10 and Noordwijk-2 the ANN model was complemented by a simple error-correction module. The variability of responses originates mainly due to the differences in depth of water, which determines the mixing process of sediments in the vertical water column. Due to this variability, even if the pattern of SPM concentrations at Noordwijk-2 predicted by the ANN is accurate, the range of its variability could be different from that of Noordwijk-10 (for which the ANN model was built). This aspect was taken care of by adopting the following correction factor. The ratio of average SPM at Noordwijk-2 and Noordwijk-10 was calculated for each month (Table 3) and was used as a multiplication factor to correct the ANN model's prediction of SPM for Noordwijk-2.

The authors realise that this simplified approach in caring for the differences of SPM concentrations under similar bed shear stress and other variables at two different locations with different depths of water needs improvement. The mixing process in the vertical water column, similar to many other natural processes, is nonlinear and thereby adopting a single coefficient to correct the ANN model's output calls for further research. Moreover, the data collection period of the two locations is not the same. However, the results as shown in Figure 6 suggest that the approach gave reasonably accurate results. The RMSE was 21.02 mg/l ($R^2$ was 0.74). The RMSE for the stormy periods was 27.71 mg/l (with 952 data points) and for the calm period was 14.93 mg/l (with 1,426 data points) whereas the $R^2$ values were 0.72 and 0.51. The average SPM at Noordwijk-2 was 28.2 mg/l.

### Predicting SPM at Schouwen

Schouwen is about 30 km south of Noordwijk and at 10 km offshore the depth of water is 16.6 m. Bi-weekly SPM data were collected there during the period 1975–2008. There were about 483 measurements of quasi-simultaneous SPM observations at Schouwen and Noordwijk during this period. At these places SPM observations were taken usually at two-weekly interval and sometimes at longer time intervals. During this period SPM data were also collected at Noordwijk-10. The ratio of measured SPM concentrations

**Table 3** | Correction factors for SPM

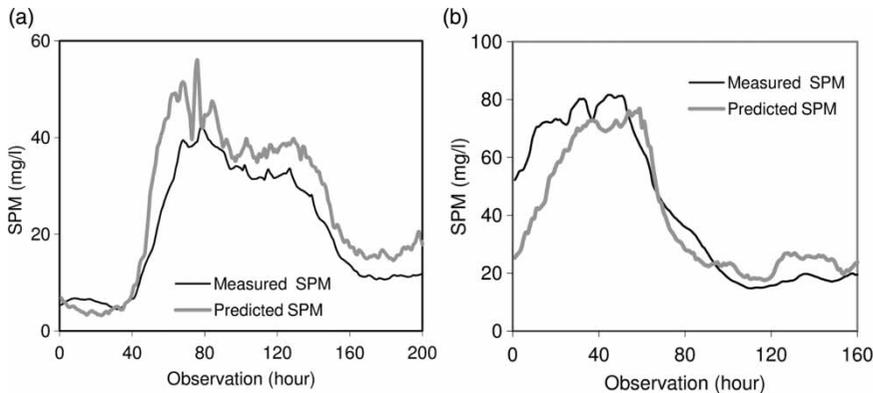|             | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|-------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Noordwijk-2 | 2.3 | 2.8 | 2.7 | 2.3 | 1.9 | 1.8 | 1.6 | 2.0 | 1.8 | 1.7 | 3.1 | 3.5 |
| IJmuiden    | 0.7 | 0.8 | 1.1 | 1.2 | 2.4 | 1.5 | 1.6 | 2.5 | 1.4 | 0.9 | 1.1 | 1.6 |
| Schouwen-10 | 3.1 | 3.3 | 4.2 | 2.1 | 2.4 | 1.3 | 1.5 | 2.2 | 2.4 | 2.6 | 4.1 | 4.7 |

**Figure 6** | Comparison of SPM concentration predicted by the ANN model with the measured data for Noordwijk-2. (a) During a storm period during 29 September 2001 to 7 October 2001 (about a week). (b) During another storm period during 20–27 November 2001 (about a week).

at Schouwen-10 and Noordwijk-10 was computed for each month (Table 3). This ratio was used as the correction factor. The correction factor applied corresponds to the ratio of the average SPM concentration at Schouwen and that of Noordwijk-10 as depicted in the Dutch Sediment Atlas (Suijlen & Duin 2002).

Similar to Noordwijk-2 the numerical model was used to generate bed shear stress data for every hour for the calendar year 1996 for Schouwen. Using this data the ANN model predicted the SPM for every hour and that was further corrected using the correction factor. Since the measured data (at Schouwen) are only available at about every two weeks, the model output can be compared to measurements only when these observations were available (after using the correction factor described above) – this is presented in Figure 7. In general, the prediction seems to be reasonable (compared with the measured data). However, correcting the ANN model's prediction by a multiplication factor is not, obviously, the only way to do it and requires further investigation. The differences in prediction of SPM with the numerical model based on the Delft3D modelling tool (WL Delft Hydraulics, Lesser *et al.* 2004) are large.

### Predicting SPM at IJmuiden

IJmuiden is about 40 km north of Noordwijk. At 2 km offshore the depth of water is about 5 m. The numerical model was used to generate hourly bed-shear stress data for the entire calendar year 1996 for IJmuiden. Additionally,
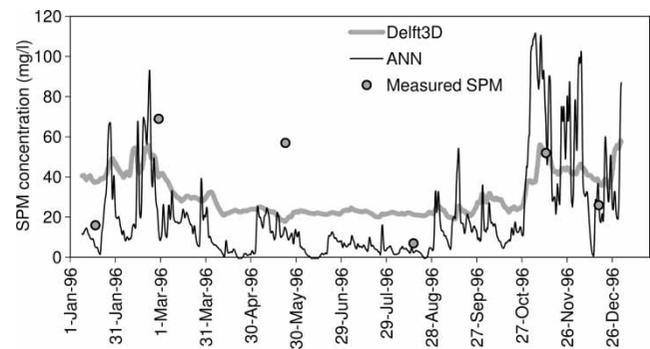


**Figure 7** | SPM concentrations predicted by the ANN model (after correction) and by the numerical model based on Delft3D together with the measured SPM concentrations at Schouwen.

the measured hourly wind data and wave heights were used to run the ANN model to predict hourly SPM values. Between 1988 and 2008 there were 342 quasi-simultaneous measurements of SPM at IJmuiden and Noordwijk-10. The average measured values of SPM for each calendar month were computed for IJmuiden and Noordwijk-10. The correction factor for each calendar month was computed as the ratio of the monthly average SPM at IJmuiden and Noordwijk-10 (Table 3). Using the correction factors the SPM values obtained from the ANN model were updated. The measured values of SPM at IJmuiden for the year 1996 was available for the period 25/3/1996 to 20/10/1996. The comparison of the ANN model's prediction (after correction) together with the measured data shows an impressive performance of the ANN model for the time moments for which the measured data are available
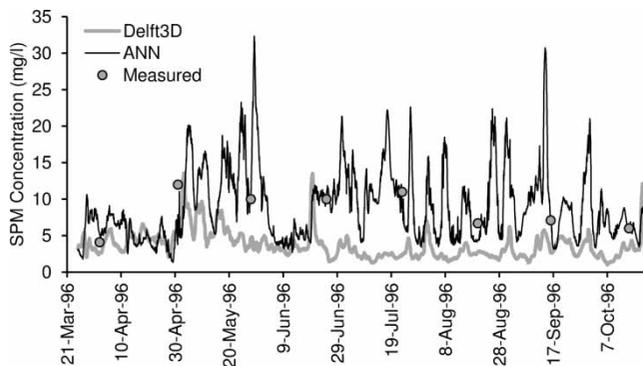
**Figure 8** | SPM concentrations predicted by the ANN model and by the numerical model based on Delft3D together with the measured SPM concentrations at Ijmuiden.

(Figure 8). SPM values predicted by the numerical model developed with the Delft3D tool are also shown in Figure 8. It may be noted that the numerical model's predictions have substantial differences with that of the ANN model.

## CONCLUSIONS

The paper presents a novel approach to modelling suspended sediment concentrations with applications to the Dutch coast. The main conclusions are as follows.

- The ANN model built with a high frequency data collected at Noordwijk showed reasonably good accuracy in predicting SPM concentrations. The measured data and data generated by a numerical model were used to build the model. The ANN model's prediction can be particularly useful for operational management.

- Filtering of data was found to be very useful. The original time series of SPM showed significant variations and this had a negative effect on the initially trained ANNs. By the use of a low pass filter it was possible to filter out the short-term fluctuations that cannot be modelled and to help the ANN model to catch the main trend.

- Three tests were undertaken to apply the ANN model trained on the data at Noordwijk-10 (and complemented by a simple error-correcting method) to predict SPM at other locations. The ANN model showed reasonable accuracy.

Future research will aim at considering other filtering methods, employing model ensembles, and developing and testing other error-correcting algorithms for improving the accuracy of SPM predictions at different locations.

## REFERENCES

ASCE 2002 Task Committee on application of artificial neural networks in hydrology. Artificial neural network in hydrology. *J. Hydrol. Eng.* **5** (2), 115–137.

Bhattacharya, B. & Solomatine, D. P. 2006 Machine learning in sedimentation modelling. *Neural Networks* **19** (2), 208–214.

Bhattacharya, B., Diebel, I. K., Karstens, S. A. M. & Solomatine, D. P. 2006 Neural networks in sedimentation modelling for the approach channel of the port of Rotterdam. In: *Estuarine and Coastal Fine Sediment Dynamics* (Marine Sciences, vol. 8) (J. Maa, L. Sanford & D. Schoellhamer, eds). Elsevier, Amsterdam, pp. 465–480.

Bhattacharya, B., Price, R. K. & Solomatine, D. P. 2007 Machine learning approach to modelling sediment transport. *J. Hydraul. Eng.* **133** (4), 440–450.

Bowden, G. J., Dandy, G. C. & Maier, H. R. 2005 Input determination for neural network models in water resources applications, Part 1 – Background and methodology. *J. Hydrol.* **301**, 75–92.

Chua, L. H. C. & Holz, K.-P. 2005 Hybrid neural network. Finite element river flow model. *J. Hydraul. Eng.* **131** (1), 52–59.

Cigizoglua, H. K. & Kisi, O. 2006 Methods to improve the neural network performance in suspended sediment estimation. *J. Hydrol.* **317**, 221–238.

Cobaner, M., Unal, B. & Kisi, O. 2009 Suspended sediment concentration estimation by an adaptive neuro-fuzzy and neural network approaches using hydro-meteorological data. *J. Hydrol.* **367** (1–2), 52–61.

de Kok, J. M. 1994 *Numerical Modelling of Transport Processes in Coastal Waters*. PhD Thesis, University of Utrecht, The Netherlands.

Haykin, S. 1999 *Neural Networks: A Comprehensive Foundation*. Prentice Hall, Englewood Cliffs, NJ.

Jain, S. K. 2001 Development of integrated sediment rating curves using ANNs. *J. Hydraul. Eng.* **127** (1), 30–37.

Jain, S. K. & Chalisgaonkar, D. 2000 Setting up stage-discharge relations using ANN. *J. Hydrol. Eng.* **5** (4), 428–433.

Kisi, O. 2010 River suspended sediment concentration modeling using a neural differential evolution approach. *J. Hydrol.* **389**, 227–235.

Lesser, G. R., Roelvink, J. A., van Kester, J. A. T. M. & Stelling, G. S. 2004 Development and validation of a three-dimensional morphological model. *Coastal Eng.* **51** (8–9), 883–915.

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

May, R. J., Dandy, G. C., Maier, H. R. & Fernando, T. M. K. G. 2008 Nonlinear variable selection for artificial neural networks using partial information. *Environ. Modell. Softw.* **23**, 1312–1326.

Mills, D. K. 2002 *Developing a Common UK–Netherlands Approach to North Sea Monitoring.* Report no. RIKZ/OS/2002.409. National Institute for Coastal and Marine Management (RIKZ), The Netherlands.

Mitchell, T. M. 1997 *Machine Learning.* McGraw-Hill, New York.

Muttiah, R. S., Srinivasan, R. & Allen, P. M. 1997 Prediction of two-year peak stream discharges using neural networks. *J. AWRA* **33** (3), 625–630.

Nagy, H. M., Watanabe, K. & Hirano, M. 1999 Prediction of total sediment load in alluvial streams: hydroinformatics modelling. In *Proc. IAHR Congress*, Graz, Austria.

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

NeuroSolutions 2010 Available from: http://www.nd.com (accessed on 15 July 2010).

Principe, J. C., Euliano, N. R. & Lefebvre, W. C. 2000 *Neural and Adaptive Systems – Fundamentals Through Simulations.* John Wiley & Sons, New York.

Singh, A. K., Deo, M. C. & Sanil Kumar, V. 2008 Prediction of littoral drift with artificial neural networks. *Hydrol. Earth Syst. Sci.* **12**, 267–275.

Solomatine, D. P. & Ostfeld, A. 2008 Data-driven modelling: some past experiences and new approaches. *J. Hydroinf.* **10** (1), 3–22.

Soulsby, R. L., Hamm, L., Klopmann, G., Myrhaug, D., Simons, R. R. & Thomas, G. P. 1993 Wave-current interaction within and outside the bottom boundary layer. *Coastal Eng.* **21** (1–3), 41–69.

Suijlen, J. M. & Duin, R. N. M. 2001 *Variability of Near-surface Total Suspended Matter Concentrations in the Dutch Coastal Zone of the North Sea: Climatological Study on the Suspended Matter Concentrations in the North Sea.* Report no. RIKZ/OS/2001.150X. National Institute for Coastal and Marine Management (RIKZ), The Netherlands.

Suijlen, J. M. & Duin, R. N. M. 2002 *Atlas of Near-surface Total Suspended Matter Concentrations in the Dutch Coastal Zone of the North Sea.* Report no. RIKZ/2002.059. Ministry of Transport, Public Works and Water Management, The Netherlands.

Swart, D. H. 1974 *Offshore Sediment Transport and Equilibrium Beach Profiles.* Delft Hydraulics Lab. Publ. 131, Delft, The Netherlands.

Tawfik, M., Ibrahim, A. & Fahmy, H. 1997 Hysteresis sensitive neural network for modeling rating curves. *J. Comput. Civil Eng.* **11** (3), 206–211.

Tayfur, G. 2002 Artificial neural networks for sheet sediment transport. *Hydrol. Sci. J.* **47** (6), 879–892.

Thirumalaiah, K. & Deo, M. C. 1998 River stage forecasting using artificial neural networks. *J. Hydrol. Eng.* **3** (1), 26–32.

van de Kreeke, J. & Hartog, J. 2004 *Analysis of Optical Back Scatter Data Observed by the Smart Buoy at the Stations Noordwijk-10, Noordwijk-5 and Noordwijk-2.* Report no. RIKZ/OS/2003.166x. National Institute for Coastal and Marine Management (RIKZ), The Netherlands.

van Kessel, T., Winterwerp, J. C., Van Prooijen, B. C., Van Ledden, M. & Borst, W. G. 2007 Modelling the seasonal dynamics of SPM with a simple algorithm for the buffering of fines in a sandy seabed. In *Proc. INTERCOH'07*, Brest, France.

Vos, R. 2004 *Analysis of Suspended Matter Data from the Minipod Deployments.* Report no. RIKZ/OS/2004/147w. National Institute for Coastal and Marine Management (RIKZ), The Netherlands.

Zamani, A., Solomatine, D. P., Azimian, A. & Heemink, A. 2008 Learning from data for wind–wave forecasting. *Ocean Eng.* **35** (10), 953–962.