Short-term Tide Prediction

Nils Hasler[1] and Klaus-Peter Hasler[2]

[1] Max-Planck-Institut für Informatik
[2] Bremischer Deichverband am rechten Weserufer

Abstract. Ever since the first fishermen ventured into the sea, tides have been the subject of intense human observation. As a result computational models and 'tide predicting machines', mechanical computers for predicting tides have been developed over 100 years ago. In this work we propose a statistical model for short-term prediction of sea levels at high tide in the tide influenced part of the Weser at Vegesack. The predictions are made based on water level measurements taken at different locations downriver and in the German Bight. The system has been integrated tightly into the decision making process at the Bremen Dike Association on the Right Bank of the Weser.

1 Introduction

The marine industries have always greatly relied on tides. The tide phase governs currents in the coastal regions and many harbours are only accessible when certain water levels are exceeded. So, in order to allow scheduling harbour utilisation and docking and sailing times it is important to predict the times and expected water levels for high and low water.

Another important application of water level forecast can be found in disaster control. Here the authorities attempt to predict storm surges and other abnormal tidal behaviour. This allows them to take actions such as closing flood barrages and flood gates, warning the population, or even to declare a national state of emergency.

Traditional tide calendars and tide tables provide a good starting point for many naval applications. These calendars are computed assuming that water levels are solely a product of the superposition of influences from different celestial bodies. Water level prediction systems normally focus on forecasting the deviation from the astronomically predicted tide called tide surge. Tide surge is assumed to be primarily a result of the influence of wind on the water. Yet, besides wind speed and direction, water salinity, temperature, and atmospheric pressure have been found to have an influence on tide surge [4]. As we are concerned with very short-term predictions only, we accept that these additional properties can be assumed to be constant and their influence is implicitly contained in the water level measurements we incorporate into the prediction.

In this paper we introduce a statistical approach to very short-term tide prediction. Namely, we focus on the interval 3 to 5 hours before the expected high tide because this is the time frame in which the Bremischer Deichverband am rechten Weserufer[3] (BDVR) needs to come to its decisions concerning closing flood barrages and flood gates. While these decisions are frequently made, it may also become necessary to take further action such as alarming the emergency committee.

Our approach employs a statistical model using water level measurements at five different locations downriver and in the German Bight to predict the water level at high tide at Vegesack which is a downriver suburb of Bremen, situated in the tide influenced region of the Weser. Additionally, wind direction and speed are incorporated to improve the prediction. Figure 1 shows the locations of the tide gauges on a map of the German Bight.

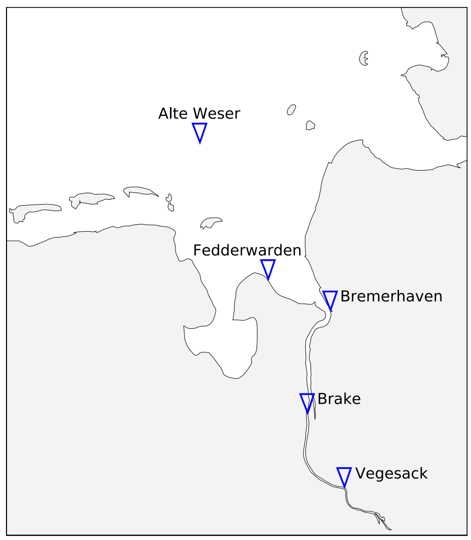

Fig. 1. Tide gauges used for predicting the water level at Vegesack

Water level prediction for the German Bight is also done by the Bundesamt für Seeschifffahrt und Hydrographie[4] (BSH) but their forecasts are focussed on a different timespan. Since they are primarily targeting the marine industries water levels are predicted semi-automatically 6 to 72 hours in advance. Due to this different target time we are able to produce more accurate forecasts although their prediction provides a valuable second opinion on the matter.

The rest of the paper is structured in the following way. Section 2 introduces previous approaches to water level prediction. Section 3 details our approach. Section 4 provides validation of prediction results and a summary is provided in Section 5.

[3] Bremen Dike Association on the Right Bank of the Weser
[4] German Federal Maritime and Hydrographic Agency

2 Previous Work

Traditionally, only astronomical influences were included into tide forecasts. These were later enhanced by a number of shallow-water constituents to account for different topologies of the sea floor. The parameters are computed using the so-called Harmonic Analysis which is described in more detail below. Despite their simplicity these models are still widely in use since they provide the only means to make long-term predictions. This allows the computation of tide tables and calendars. As these forecasts do not incorporate up-to-date weather information, they are unable to achieve the accuracy of more advanced approaches.

A simple, yet effective statistical system, is used by the BSH. It computes the tide surge as the linear combination of a limited number of carefully chosen parameters. The weights for the linear combination are computed by minimising the prediction error on historical data [4].

Other approaches have included the use of neural networks [5], Kalman filters [8] and chaos theoretical approaches [6].

The most sophisticated systems to date employ three-dimensional simulations of relevant parts of the ocean, incorporating wind, temperature, humidity, and atmospheric pressure data which are provided as input from weather forecast simulations [4].

2.1 Harmonic Analysis

The classical Harmonic Analysis, which was originally developed by Sir William Thomson (the later Lord Kelvin) around 1867, assumes that water levels can be modelled as the superposition of sinusoidal influences of the sun and the moon on the sea [7]. Fitting sines whose frequencies are the linear combination of multiples of the rotation and precession speeds of the celestial bodies relative to the earth, to historical tide data leads to a number of constituents as explained by Foreman [1]. Which and how many frequencies to choose depends on the location the predictor is to be generated for and has to be decided manually. In the following v_i represents the ith component's speed and φ_i its phase. A_i is the amplitude of the constituent and $y(t)$ the water level measured at time t. Unfortunately, as a result of increased friction on the ocean floor, the influence of some harmonics lag behind the astronomically predicted phase. Thus, the phases have to be estimated as well as their amplitudes instead of being derived from more accurate astronomical observations.

Commonly, the following equation, derived from the above assumptions is minimised with respect to the error E to obtain an optimal fit to historical water level measurements $y(t)$.

$$E = \sum_t (A_0 + \sum_{i=1}^{N} A_i \cos(v_i t + \varphi_i) - y(t))^2$$

In [7] Thomson and Tait describe a mechanical device that integrates these constituents to predict the water level at a given location. Similar tide predicting machines were in use in various countries until the 1960s.

3 Approach

In a nutshell, we try to predict the water level at high tide at Vegesack on the Weser using a statistical approach that relies on time series of water level measurements taken downriver and in the German Bight up to the time the prediction is made. Additionally, the inclusion of wind speed and direction measurements proved worthwile.

As the prediction primarily relies on water level measurements it is interesting to investigate what the best encoding of these water levels is. The two considered encodings include using the raw measurement values directly and employing the classical tide surge encoding. Considering the standard deviations of measured tide levels $\mathbf{y}$ and the difference between the astronomically predicted tide levels $\mathbf{p}$ and the measured levels called tide surges or surge water levels $\mathbf{s} = \mathbf{y} - \mathbf{p}$ it may seem advantageous to predict tide surges, since the tide surge is not scattered as widely around its mean as the raw water levels.

$$\mathrm{std}(\mathbf{s_{high}}) = 36.3\,cm < \mathrm{std}(\mathbf{y_{high}}) = 42.7\,cm$$

Yet, experimentation showed that predicting actual water levels produces slightly more accurate results.

Of the four possible encodings only two need to be investigated further. Predicting tide surges from water levels or vice versa is not worth considering because the information about the expected water level is not available to one side of the equation. The achievable accuracies are consequently limited.

Thus only the two remaining variants were compared. While the cross-validation accuracy of the tide surge based prediction reached 24.8 cm, the accuracy achieved by predicting water levels from water levels could be estimated to be 6.6 cm. The better result could probably be achieved because the prediction relies on water levels measured only hours before the predicted event. Since astronomical influences change in the order of days, they can be assumed to be constant during the prediction time span. Their influence is already implicitly included in the water level measurements used in the forecast.

3.1 Wind Direction

In contrast to these results, it was possible to significantly increase the effectiveness of wind measurements, originally available in polar coordinates. Wind levels in this representation are, as Solomantine et al. mention in [6], not sufficiently correlated with the high tide to be of any significant use. The approach the BSH follows is to use the projections of the wind vector onto the coordinate axes as input features. We extended this approach to time dependent projection on different directions.

During the time before astronomically predicted high tide the correlation between wind speed and direction and high tide y_{high} or tide surge s_{high} are given in Figure 2. Since the maximum of these values remains below 0.4, we agree with Solomantine et al. [6] that this is not sufficient to be used directly as a feature.

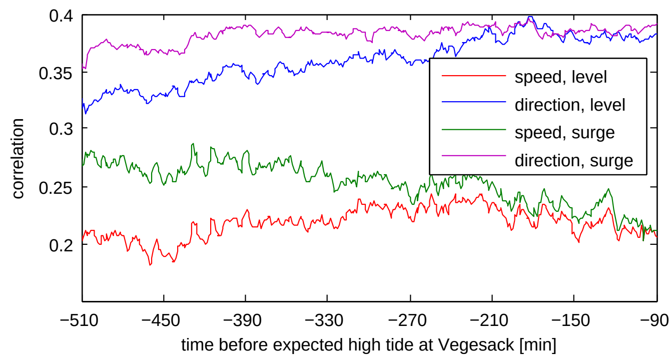

Fig. 2. Correlations between wind speed and direction with water levels and tide surges.

Fortunately, this problem can be alleviated by assuming that only the component of the wind in a certain direction contributes significantly to the tide surge. While this direction is presumed to be invariant between different tides it is allowed to change during the timespan leading up to the high tide.

To compute the most important wind direction $\omega(t)$ we projected the measured wind vectors on unit vectors sampling the unit circle in steps of $1°$. We then calculated the correlation with the observed water level y_{high} or the tide surge s_{high}. The wind direction with the highest correlation was then chosen for $\omega_y(t)$ and $\omega_s(t)$ respectively. It is interesting to note that the correlation with s_{high} is significantly higher than the correlation with y_{high} (see fig. 3). This finding is not surprising because tide surge is primarily caused by wind. However, in spite of this convincing edge, the derived wind directions are almost the same, as shown in figure 4. This explains why the difference in prediction accuracy using the two models proved to be minimal.

Interestingly, as can be seen in Figure 3 the wind loses influence on the outcome of the water level at high tide in the last two hours before high tide. To explain this phenomenon consider Figure 5 which plots wind directions and correlations into a map of the Friesian Coast. Wind directions are plotted to the positions the tidal wave is expected to be at at the corresponding time.

Obviously, in order to achieve good results, measurements have to be taken during the same phase in the tide cycle. Unfortunately, the phase cannot be determined exactly as low and high tide not only vary in level but exhibit high variance in time when compared to the expected events as predicted by the harmonic analysis.

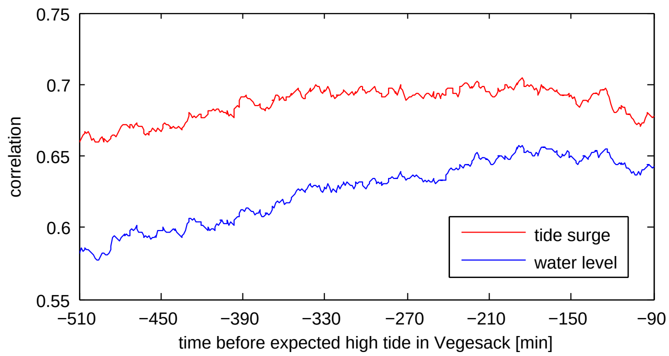

Fig. 3. Correlations between the best projected wind vector with water levels and tide surges.

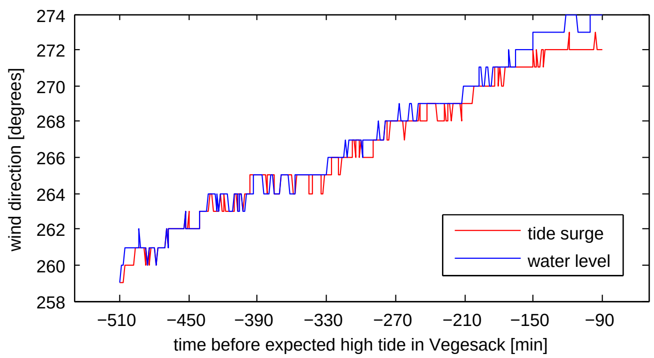

Fig. 4. The best wind direction as a function of time. The wind direction is specified in the nautical system where 0° represents north wind and the angle increases clockwise.

So again two different possibilities for anchoring the measurements in time had to be evaluated. On the one hand the tide cycle can be estimated at low or high tide. The first prediction is due approximately 30 minutes after low tide can be detected at the tide gauge Vegesack. Fortunately, since Vegesack is also the place the prediction is made for, the temporal correlation between low water and high water is quite high. Thus, anchoring the measurements at this point in time is a reasonable approach. On the other hand the astronomically predicted cycle can be used for anchoring the prediction. We could not find a great difference in prediction error with either method, so we settled for using the detected time of low tide.

These investigations result in the final formulation of the feature vector used initially for training the model and ultimately for predicting water levels. The vectors are generated by concatenating water level measurements and projected wind speeds taken every minute from three hours before low water at Vegesack

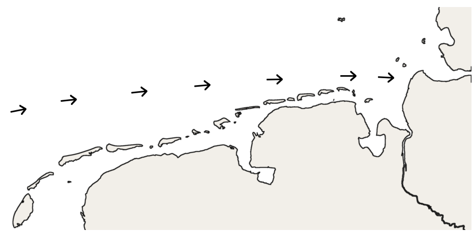

Fig. 5. The best wind direction at the position the tidal wave is situated at at the given time.

until the prediction is made, that is 30–150 minutes after low water at Vegesack or three to five hours before high tide at Vegesack.

3.2 Principal Component Analysis

The third preprocessing step, a principal component analysis (PCA) as described in [2] was used to reduce the dimensionality of the feature vector. By reducing the dimensionality of the data the training time could be reduced tremendously. In comparison with running the optimisation without reducing the dimensionality the observed loss in accuracy could not be found to deteriorate the results. On the contrary, the prediction error could be minimised by choosing the right number of principal components. For the different models that were computed 6 to 56 principal components proved to be optimal.

3.3 Training

Finally, the overdetermined linear system $\mathbf{Ax} = \mathbf{y}$ was solved in a least-squares fashion to determine the coefficients $\mathbf{x}$ of the linear predictor. Here the rows of $\mathbf{A}$ are the feature vectors and the corresponding rows of $\mathbf{y}$ specify the water levels or tide surges that are to be predicted.

By combining the PCA and the linear predictor the prediction model can be reduced to a single scalar product.

$$y_{pred} = \mathbf{x} \cdot \mathbf{f_{PCA}} = \mathbf{x} \cdot (\mathbf{E}\ \mathbf{f}) = (\mathbf{x}^{\mathsf{T}}\ \mathbf{E}) \cdot \mathbf{f}$$

Here y_{pred} is the predicted water level, $\mathbf{f_{PCA}}$ the feature vector as reduced by the PCA, $\mathbf{E}$ the matrix of used eigenvectors, and $\mathbf{f}$ the original feature vector.

4 Validation

The prediction model is computed based on measurements logged by the BDVR in the period from November 1999 to October 2002. Unfortunately, since some measurements are not complete only 1902 of 2117 tide cycles were available for training. For validating the system we used k-fold cross-validation as described by Kohavi [3] employing 10 folds.

Since the system has been installed at the BDVR for several months now and the functionality has reached a stable state, we can also evaluate the accuracy of the live system for the months January and February 2007 and compare our predictions with the forecasts of the BSH. Figure 6 displays the relative frequencies of prediction errors for this period. The cases where tide gauges failed and we were subsequently unable to compute a prediction have been omitted from the evaluation. Table 1 summarises the standard deviations and biases of the predictors in Figure 6. Clearly, the more data is available the more accurate the predictor becomes.

Unfortunately, we were unable to achieve the accuracy on live data that was estimated on the test set specified in Section 3. This is probably a result of the constantly changing depth and consequently changed streaming properties of the Weser. In order to allow larger ships to enter the harbours in Bremen, Brake, and Nordenham the Weser is frequently deepened. As the training data is several years old and the Weser has since been deepened, it is possible that the model does not represent reality any longer. Apart from this, we would like to stress that as a result of the frequent storms in the winter months predictions are particularly hard to do.

	BSH	5h	4h	3h
standard deviation [cm]	30.5	15.5	13.2	8.9
bias [cm]	6.3	4.3	1.7	0.06

Table 1. standard deviations and biases of the predictor by the BSH and our system computed at different times before high tide.

4.1 Kyrill

One disadvantage of statistical prediction systems in general is that they are not particularly good at predicting extreme weather conditions because only very little training data is available for these cases. Unfortunately, these are also the conditions that a prediction system is commonly measured by. So we would now like to present an example of a recent storm that challenged our system. The storm Kyrill that reached hurricane-strength even in the Northern German Planes raged in the night from 18 January to 19 January 2007. At this point our system was already installed and Kyrill provided the first real online test.

The three predictions that were made 5h, 4h, and 3h before expected high tide forecasted 139 cm, 131 cm, and 157 cm above median high tide respectively. The final outcome was 145 cm. While deviation of 12 cm may not seem very accurate, the last prediction by the BSH was 200–250 cm above median high tide. Since the system is statistical in nature, it is hard to tell how it could have foreseen a tide so much lower than was generally expected. Yet, we presume that on the

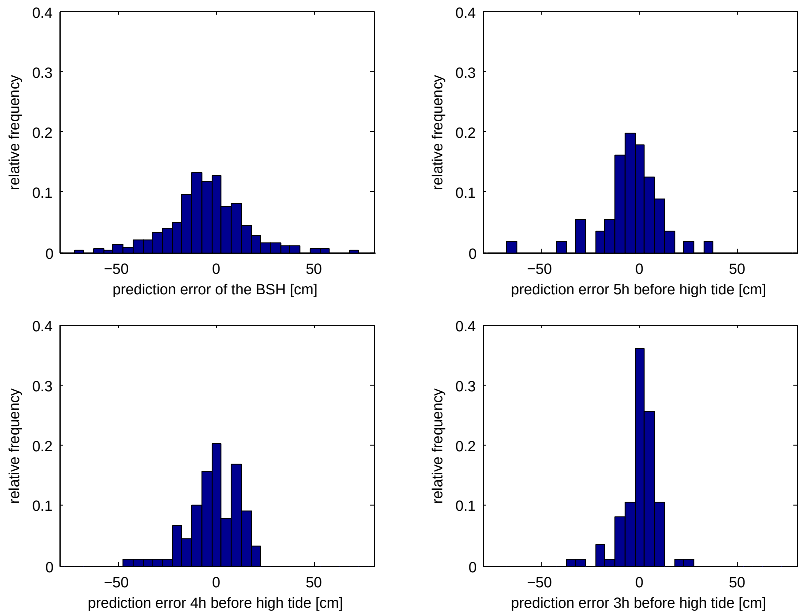

Fig. 6. Error distribution of predictions the BSH made compared to our method computed 5h, 4h, and 3h before high tide.

one hand the wind speeds our station picked up were not as high as some of the reported speeds. The highest wind speed we measured corresponds to 8 Bft and the wind direction was only westerly, not corresponding very well to the most significant wind direction. So the overall mean wind speed was registered as only 6 Bft (see fig. 7). This significantly lower wind speed probably caused the system to estimate such a low water level.

5 Summary

We have described a statistical short-term water level prediction system that has been installed at the BDVR. By integrating the system into the decision making process that has to take place twice a day a few hours before high tide, our system proved to be a valuable addition to the tools available to the human in charge. We showed that the accuracy of the approach is superior to the prediction provided by the BSH, although we target a different prediction time span.

The trial period showed, however, that it is necessary to extend the fault-tolerance of the system as frequently one or more tide-gauges fail and the prediction becomes unusable. If a tide gauge fails for a few minutes interpolating the missing measurements proved reasonable. Unfortunately, if a tide gauge becomes unavailable for an extended period of time, this approach is unfeasible. To overcome these limitations we propose to train specific models for different states of

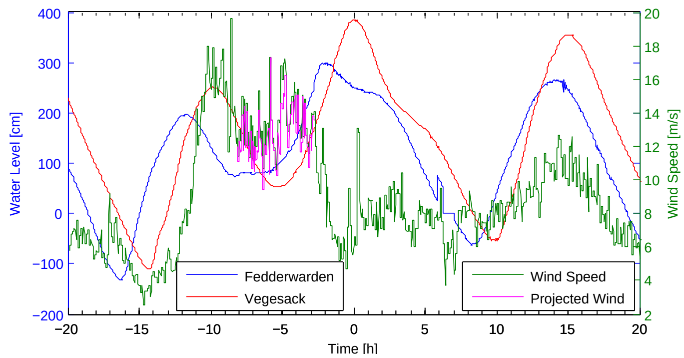

Fig. 7. Water levels and wind speeds around the storm event Kyrill

failure. For example a model using just the measurements from Vegesack could be generated. Preliminary experiments showed that this approach is practical but the accuracy of the prediction necessarily deteriorates for these models.

References

1. M. G. G. Foreman and R. F. Henry. Tidal analysis based on high and low water observations. Technical report, Institute of Ocean Sciences, Patricia Bay, Sidney, Canada, 1979.
2. I. T. Jolliffe. *Principal Component Analysis*. Springer Verlag, 2002.
3. R. Kohavi. A stdy of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1137–1145, 1995.
4. S. H. Müller-Navarra, W. Lange, S. Dick, and K. C. Soetje. über die verfahren der wasserstands- und sturmflutvorhersage. *promet*, 29(1–4):117–124, June 2003.
5. F. Röske. Sea level forecast using neural networks. *Deutsche Hydrographische Zeitung*, 49:71–99, 1997.
6. D. P. Solomatine, C. J. Rojas, S. Velickov, and J. C. Wüst. Chaos theory in predicting surge water levels in the north sea. In *Proc. 4th Int. Conference on Hydroinformatics*, Iowa, USA, July 2000.
7. W. Thomson and P. G. Tait. *Treatise on Natural Philosophy*. Oxford University Press, 1879.
8. P.-H. Yen, C.-D. Jan, Y.-P. Lee, and H.-F. Lee. Application of kalman filter to short-term tide level prediction. *Journal of Waterway, Port, Coastal and Ocean Engineering*, 122(5):226–231, September 1996.