TRANSNAV
http://www.transnav.eu
the International Journal on Marine Navigation and Safety of Sea Transportation
Volume 7 Number 2 June 2013

DOI: 10.12716/1001.07.02.09

Application of Artificial Neural Network into the Water Level Modeling and Forecast

M. Sztobryn
Institute of Meteorology and Water Management, Maritime Branch, Gdynia, Poland

ABSTRACT: The dangerous sea and river water level increase does not only destroy the human lives, but also generate the severe flooding in coastal areas. The rapidly changes in the direction and velocity of wind and associated with them sea level changes could be the severe threat for navigation, especially on the fairways of small fishery harbors located in the river mouth. There is the area of activity of two external forcing: storm surges and flood wave. The aim of the work was the description of an application of Artificial Neural Network (ANN) methodology into the water level forecast in the case study field in Swibno harbor located is located at 938.7 km of the Wisla River and at a distance of about 3 km up the mouth (Gulf of Gdansk - Baltic Sea).

1 INTRODUCTION

The dangerous sea and river water level increase does not only destroy the human lives, but also generate the severe flooding in coastal areas. The rapidly changes in the direction and velocity of wind and associated with them sea level changes could be the severe threat for navigation, especially on the fairways of small fishery harbors located in the river mouth. There is the area of activity of two external forcings: storm surges and flood wave. A possibility of predicting water level changes in a port increases safety of navigation – especially on the fairways – and makes port operation more efficient. It is extremely important for fishing ports, located in river mouths, where water level fluctuation is affected by two factors: sea level changes (forced by atmospheric pressure field) and the river level changes – forced by a flood wave. The advanced sea land water level forecasting is based on mathematical modeling. For the Polish coast of the Baltic Sea, there have been developed by the State forecast service 2 up to 4 models (depend on the localization). However only 1 model is used for the Wisla River; it is to forecast water levels from Zawichost to Tczew localities. The models to forecast water levels in the Gulf can be extrapolated to the mouth section during the storm surge (it means to predict trends of changes). On the other hand, the water level forecast's extrapolation based on the forecast for Tczew and upper part of Wisla River, essential on passing of the flood wave, would not bring positive results.

The implementation of ANN into the routine forecast service was done as the first in the Netherlands to forecast the currents in the fairway of the Ijchannel (Wust and Noort, 1994), however none of numerical models used there could satisfy the implementation conditions. Another ANN hydrological forecasting model, was constructed also in the Netherlands for the forecast for Ijsselmeer (Boogaard et al. 1998). Also in Germany were done the attempt and investigation of abilty of ANN methodology for level prediction (Röske, 1997). In Poland, the ANN method was tested for level prediction in Odra estuary and western coast

(Sztobryn M., 1999). The results of these works were very promising.

The aim of the work was the investigation of an application of Artificial Neural Network (ANN) methodology into the water level forecast (forcing by river) in the case study field in Swibno harbor located about 3 km above the entrance of Wisla River into the Gulf of Gdansk (Baltic Sea).

2 LOCALISATION AND DATA

The fishing harbor in Świbno (Fig. 1) is located at 938.7 km of the Wisla River and at a distance of about 3 km up the mouth. The harbour is situated on the western bank of the river and its structure comprises two concrete quays (160 x 90 m basin).

Figure 1. Harbour in Swibno (http://monimoni.flog.pl/wpis/1557132/swibno--port-rybacki)

Depths in the harbour range between 2 and 3 m. The harbour is used by Rybacka Spółdzielnia Rybołówstwa Morskiego „Wyzwolenie" [fishing cooperative society]. Besides, there are also located Brzegowa Stacja Ratownicza ŚWIBNO (coastal rescue station) and hydrological station (with telelimnigraph and water-gauge) of the Institute of Meteorology and Water Management –State Research Institute (IMGW-PIB). A characteristic feature of the harbour entrance is rapidly changing depth at the fairway and approaching area. It results from the harbour locating within a zone of cross sea and river penetrations influence, causing sanding up of the water area (for example during the flood wave).

The observation and measurement data used for the analysis have been produced on the stations of IMGW - PIB from April 2007 to December 2010. A frequency of data recording at the station was 10 minutes; anyhow, the data observed every hour were taken for further analyses (31082 water levels). The statistics of characteristics of the analyzed hydrological data is presented in the Table 1. (an elevation of the water gauge in Świbno is equal to – 5.08 m above sea level acc. to Kronstadt).

Table 1 The statistics of characteristics of the analyzed hydrological data

	Water level in Świbno
number of observation	31082
mean	530
median	524
min	462
max	710
lower percentile	511
upper percentile	541

The maximal observed water level, during the investigated period, was 710 cm (where the absolute maximum is equal to 767 cm from whole observation period) and the minimal was 462 cm (absolute minimum is equal to 420 cm). An average sea level value was computed to be equal to 530 and it is higher than the median by 6 cm. Standard deviation was 29,81 cm. Percentiles were calculated as equal to 562 and 500 cm respectively.

The analysis of empiric frequency distribution (Fig.2) of water levels occurrence in Świbno shows that, more than 72% of the whole population is within a range of 500 – 550 cm, it means within the mean sea water levels zone (500 cm is assumed as equal to mean sea level along the Southern Baltic coast). Within the low water level zone in the time under research there was observed not many results above 10% of population, whereas within the high levels zone 16% (including 13.7% within 550÷600 cm range and 2.3% above 600 cm).

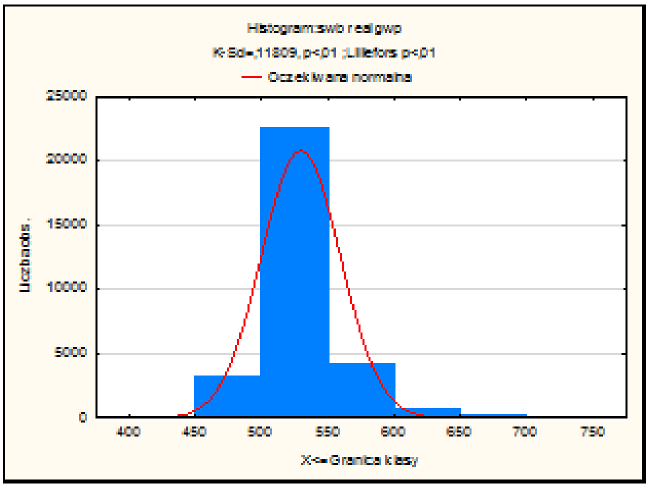

Figure 2. Frequency distribution of water levels in Swibno during the investigated period (Krzysztofik K., Kańska A., 2011)

It reveals that accepting for the model calculations the data, which are characteristic for water levels within a range of 500-550 cm would cause calibration of the model (assuming parameters) to be set only to this range. Taking it into consideration it was decided to make further analyzing only on the data which represent the highest risk, it means for the high water zone and flood wave passing one (over 555 cm). Storm surges cause not so intense sanding up as passing the flood wave. Finally the studies comprised over 2700 of cases, representing the observed in Świbno high water levels zone (i.e. above 555 cm) and for imminence threatening from the river. The imminence was represented by the input data (the levels observed/forecasted at the stations located upwards (Gdansk Głowa, Tczew, Chełmno and Toruń). An influence of sea was modelled using the sea levels recorded in Gdańsk Port Północny. The model calibration and verification was carried out taking into consideration 24 hours lead time.

3 ARTIFICIAL NEURON NETWORK METHOD

The artificial neuron network (ANN) were created as an attempt to a human brain activity performance/quality. The detailed description of the methods was included in papers worked out by Tadeusiewicz (1995/6), Sztobryn (1999, 2001, 2003), Sztobryn and Krzysztofik (2001), and on website of STATISTICA program (2010) .

The base for operation and software is the model of a singular neuron called from the neuron model authors' names the Mc Culloch Pitts model.

A structure of the ANN is composed of input data (so-called input layer), the hidden layers and the output layer. The input layer contains parameters (i.e. input vector) which, in the opinion of investigator, influence on the modelled phenomenon. It has to be emphasized that there are no limits in selection of parameters (on contrary to models of mathematical physics). Generally a list of such parameters is very long; moreover, they are often connected with each other and internally correlated. Reduction of the input parameters, called also reduction of the input data space dimension, is one of the elementary but also one of the most difficult tasks in calibration of ANN model. Modelling the hidden layer/layers through selection of suitable number of hidden neurons is a next step in calibration of the model. A number of output neurons is determined by the phenomenon character, in the analysed case it was water level with 24 hours lead time.

Building of ANN model was done by the following stages:
– reduction of the input data dimension (find the final and optimal input data vector),
– neuron network structure (i.e. decision how many hidden layers with how many neurons are included into the model) ,
– function of activation of neuron layers (i.e. the function of transformation of input vector to the output inside the individual neuron) ,
– learning method (way of comparison and correction of error, equal to the difference between modelled value and observed – for calibration the network ,

The final work was the analysis of the results in respect of quality and implementation to operational/routine work of forecast service.

Over 100 parameters, affecting water level changes in Swibno cross-section were selected. They represented one hydrology (levels, their changes and water table drops) of the Lower Wisla (from the Torun profile) and the Gulf of Gdansk (Hel and Gdañsk) as well as the meteorological conditions: current and forecasted for the Gulf of Gdansk.

Reduction of the dimension was carried out applying 3 methods: correlation, genetic algorithm and by the model sensitivity testing (it means testing the model ability to give the good simulation/forecast for input data containing and without tested parameters).

Under the literature revive and comparison of learning methods of ANN (there is the way of calculation and reduction of errors between the known i.e. observed and modelled output data values) the back propagation methods was chosen as well as mulitpreceptron structure ANN (it means that investigated structure was consists from 3 layers). The next problem was the division of whole data population into 3 series : learning and tested (for model calibration) and validation series (independent data used for model verification). 70-15-15 % division was applied it means that 70% of the population was the learning series, 15% each for the testing and validating ones. To compare the quality of performance and reliability of forecast, the statistical indicators were applied, calculated for each of the three time series separately : root mean square error (RMS) and correlation coefficient R (standard Pearson correlation coefficient with p=0.92 confidence interval).

4 RESULTS OF MODELLING

There were tested 500 network in respect of imminence from river; in case of the best 5 representations the results are presented in Table 2. There are included the characteristics of 5 best network (with changeable number of hidden neurons and one output neuron) representing the water level in Swibno with 24 hours lead time and dissection of by 70-15-15.

The first column in Table 2 specifies the grid structure. Thus MLP stands for multi-layer preceptron (Multilayer Preceptron) – one- layer in this case, it means with one layer hidden. Symbols 5-34-1 show a number of neurons.

Table 2. Characteristics of 5 best calibrated networks

network	corelation (for learning series)	correlation (for tested series)	correlation (for validation series) = model verification	MRE for learning series	MRE for tested series	MRE for validation series = model verification	number of used learning periods	activation function for hidden layer/ neurons	activation function for output layer
1	2	3	4	5	6	7	8	9	10
MLP 5-34-1	0,99	0,99	0,99	4,81	5,71	5,97	910	Tanh	Logistic
MLP 5-35-1	0,99	0,99	0,99	5,03	5,78	6,52	816	Tanh	Logistic
MLP 5-31-1	0,99	0,99	0,99	5,52	6,22	6,53	950	Tanh	Logistic
MLP 5-35-1	0,99	0,99	0,99	5,23	6,49	6,90	934	Tanh	Logistic
MLP 5-26-1	0,99	0,99	0,99	5,63	6,58	6,91	819	Tanh	Logistic

Thus MLP 5-34-1 defines a one-layer preceptron with 5 input neurons (input vector), one hidden layer containing 34 neurons and one output neuron. The values of coefficient of correlation between the real observed levels and the levels modelled (as output layer/neuron) by ANN are presented in the next three columns (2, 3 and 4). The columns are to represent series: learning, testing and validating. The learning and testing series are used for calibration of the model. The validating one is to represent the independent data (not used in the calibration process), so the model verification. The achieved correlation coefficient value of 0.99 proves very good performance/quality of the model and usability of ANN methodology in modelling/forecasting of water level. The significant performance/quality parameter is a difference between the learning and testing series errors. It displays capacity of the model to generalization of the gained knowledge (too large difference shows the neural network over-learning). In case of the network under research the differences between the errors of learning anf jdf455d testing are less than 1cm, it means the ANN – from the methodology standpoint - keeps operating properly.

Columns 5, 6 and 7 are to characterize a values of RME error (root mean error) obtained for each of 5 network and each of 3 series. According to the ANN methodology, the RME values should be the lowest in case of the learning series, the highest for the validation series (independent data). In case of all the network the values are close each other and practically they are equal to the measurement accuracy. The values of RME error obtained were less than 7 cm – what is considered to be the very good result.

The next, 8th column is to inform after how many epochs the algorithm achieves convergence. The epoch is a name for the calibration process when the ANN is to calculate output, basing on all the data of the learning and tested series. The error received from comparing of the real output value with the value modelled applying ANN is a basis for correction of weights denominated to each of neurons in the function of activation, to reduce the error in the following calculation epoch. A number of epochs oscillated between 816 and 950, what confirms a convergence of the applied algorithm.

Information put in column 9 (hyperbolical tangent) and column 10 (logistic function) represents the activation function of hidden and output layer.

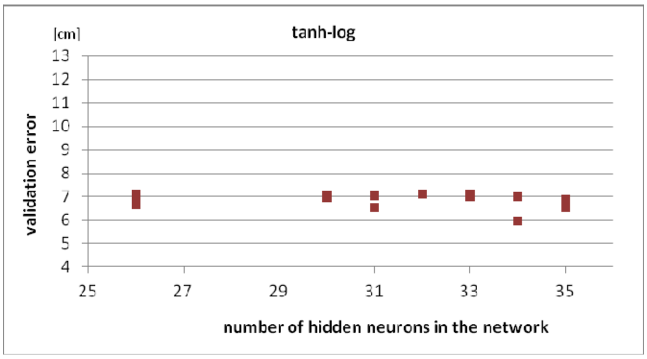

Figure 3. Validation error as the function of number of hidden neurons in network for activation functions tanh (hidden layer) and logistic in output layer (Sztobryn, Mielke 2012)

The dependence of validation error from number of hidden neurons in network for activation functions tanh and log was shown on the Fig, 3. The minimum of this function is located for 34 hidden neurons.

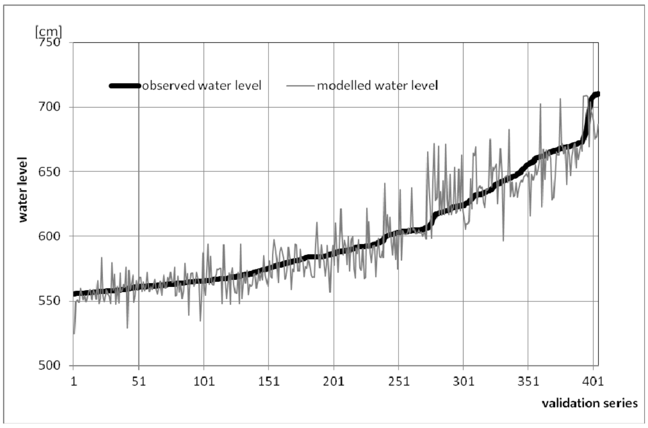

Figure 4. Comparison of observed and modelled values (Sztobryn, Mielke 2012)

The comparison of modelled and observed values are presented on the figure 4. The vertical axis stands by the water level when the horizontal : by the cases in validation series (i.e the hours of observations). In general the slope and shape of the lines thick (observed water level) and thin line (modelled) are similar. The highest values of error are recorded for the 2 extreme event : storm surge in October 2009 and flood wave in 2010 (the absolute value of water level above the Świbno were exceeded). The comparison of the values generated by ANN and other operational models show that, that these errors were generated by the meteorological factors as guest and front passing, which aren't modelled by meteorological models. It means, that ANN isn't the cause of these errors.

5 CONCLUSION

Good agreement with observed and modeled water level, especially the results gained in correlation investigation, allows to find : the presented methodology could be use for modeling and forecast of the water level.

ACKNOWLEDGEMENTS

The work was done in the frame of internal IMGW-PIB project (task DS-P1.5.1 "Opracowanie i wdrożenie metodyki sieci neuronowych do prognozowania zmian poziomów wody w ujściowym odcinku Wisły") coordinated by prof. M.Ostojski. The author wish to thank Directors prof. M.Ostojski and prof. M.Maciejewski (Institute of Meteorology and Water Management – National Research Institute) for valuable suggestions. The special acknowledgements are directed to co-workers from BPH Gdynia for participation in very useful discussion and also for their support.

REFERENCES

Röske F.,1997. *Sea Level Forecasts Using Neural Networks*. German Journal of Hydrography,vol49,no1,71-99.

Statistica NN - software documentation, USA 2010, http://www.statsoft.com

Sztobryn M., 1999. *Możliwości zastosowania sieci neuronowych w operacyjnej służbie prognoz hydrologicznych*, Inżynieria Morska: Geotechnika nr 3/1999: 107-110 . (in Polish).

Sztobryn M., Mielke M.2012 *Opracowanie i wdrożenie metodyki sieci neuronowych do prognozowania zmian poziomów wody w ujściowym odcinku Wisły*. Raport z projektu IMGW-PIB DS.-P1.5.1, 2012

Sztobryn M.2009 *Sea Ice Services in the Baltic Sea*. TransNav - International Journal on Marine Navigation and Safety of Sea Transportation, Vol. 3, No. 4, pp. 415-420, 2009

Sztobryn M.2011 *Baltic Navigation in Ice in the Twenty First Century*. TransNav - International Journal on Marine Navigation and Safety of Sea Transportation, Vol. 6, No. 4, pp. 517-523, 2011

Tadeusiewicz R., 1993. *Sieci Neuronowe*. AOW RM, Warszawa,pp.120, 1993

Wust J.C., Noort G.J.H.,1994. Neural network current prediction for shipping guidance. Proceedings Oceans 94. OSATES, Brest France 1994,. I 58 - I 63., 1994

Van den Boogaard H. F. P., Gautam D. K., Mynett A. E.,1998. *Auto-regressive Neural Networks for the Modelling of Time Series*. Preprint Hydroinformatics, 1998,

Krzysztofik K., Kańska A., 2011, Raport z Zadania 1: *Ocena aktualnego stanu zagrożenia powodziowego terenów na zachód i wschód od Wisły Śmiałej* [w:] Projekt UDA.RPPM.05.03.00-00-001/09-00: Utworzenie map terenów zalewowych zagrożonych powodzią od morza przez Wisłę Śmiałą przy wykorzystaniu niezbędnego sprzętu i oprogramowania oraz numerycznej mapy terenu.-koordynator M.Sztobryn, Raport Wewnętrzny IMGW, Gdynia/ Warszawa