

Determining Necessary Storm Surge Model Scenario Runs Using Unsupervised-Supervised Artificial Neural Networks

Bernard B. Hsieh and Jay J. Ratcliff
US Army Engineer Research and Development Center
Vicksburg, MS 39180, USA

Abstract

A significantly large number of storm simulations are required in the current stochastic based modeling framework adopted by the US Army Corps of Engineers (USACE) to examine flood risk reduction measures for Southeast Louisiana. Because of the large number of numerical model runs and the detailed nature of the modeling, the adopted modeling framework requires significant amounts of computer and human resources. This paper presents an ANNs method that can adequately characterize the storm surge response, and provide a means for reliably estimating surge responses for storms not simulated with the ADvanced two-dimensional CIRCulation model (ADCIRC).

The ANNs methodology includes identifying correlations between input parameters and model output responses, quantifying the input-output relationships using supervised ANNs, development of storm parameter clustering and splitting processes (CSP) using unsupervised ANNs, checking the reliability of CSP using supervised ANNs, and comparing surge-return period estimates obtained using a reduced storm set versus those made using a larger storm set from the ongoing Louisiana Coast Protection and Restoration (LACPR) and Morganza to the Gulf projects are presented.

Keywords: Unsupervised-supervised artificial neural networks, storm surge, hurricane parameters, splitting processes.

1. Introduction

Numerical simulation is the most accurate and efficient modern technique to calculate the surge response during a storm/hurricane event. Recently, a very successful interagency team effort has been made to model the storm surge response for hurricane events along the northern coast of the Gulf of Mexico. The dependencies between surge and waves are treated through coupled models, and a probabilistic approach has been adopted for calculating inundation levels and their associated probabilities. However, numerous model runs are required to cover a wide range of possible hurricane scenarios to meet the management and project design needs. This newly established modeling framework requires a significant amount of resources including personnel and computer resources, as well as contract labor and other factors which raise project costs and completion time requirements. Good planning with an alternative technology path that can reduce costs for projects is highly desirable. Simulation using ANN techniques was examined as a possible tool for

reducing the resources required to make storm surge estimates for design purposes. Often in design, a large set of storm surge simulations must be made for each of a series of different project alternatives (such as different levee alignments). Increasing numbers of alternatives dramatically escalate the computational requirements for a detailed modeling approach.

2. Brief Overview of solution Techniques and Computational Procedures

To save time and cost for achieving project goals for this complex coastal system where numerous and very different risk reduction measures are being considered, an approach is examined with the purposes of eliminating some unnecessary runs but at the same time maintain the required high standard of results that are achieved with the detailed modeling approach applied for a large number of hurricane events. This approach employs Artificial Neural Networks (ANNs) (NeuroSolutions, 2003) and is described below.

Supervised ANNs (SANNs)

The most popular static algorithm employed in SANNs for training is a Multi-layer Feed Forward Neural Network (MLP); Time-Delayed Recurrent Neural Networks are most use for dynamic situations. In this study, the MLP is the most suitable algorithm to address the relationship between input parameters (storm characteristics) for each ADCIRC run and storm surge response (peak surge values) for each computational grid point of interest. So, a short description of MLP is provided here. The detailed theoretical development for the algorithms can be found in Principe, et al (2000), and Haykin (1994).

They are numerous works on the study and applications of MLP. The different variants of this model differ in the way the weights are updated during learning. The back-propagation training algorithm, which fully incorporates the MLP architecture, is currently the most general-purpose, commonly used neural-network paradigm. The basis principle of the back-propagation algorithm is to introduce a method of modifying the network weights by minimizing the error between a target and computed objects. The main advantage of MLP is that they are easy to use, straightforward in conceptual design, and that they can approximate any input/output map. However, key disadvantages are that they train slowly, require a large amount of training data, and classify by using static backpropagation training. Actually, the MLP model does not perform temporal processing since the vector space input encoding gives the model no hint of the temporal relationship of the inputs

Unsupervised ANNs (USANNs)

Unsupervised training means the networks learn from their own classification of the training data, without external help. It is assumed that class membership is broadly defined by the input patterns that share common features, and that the network will be able to identify those common features across the range of input patterns. Self-organizing Feature Maps (SOFM) is special kind of neural network that can be used for clustering tasks. Only one map node (winner) at a time is activated, corresponding to each input. The location of the responses in the array tends to become ordered in the learning process as if some meaningful nonlinear coordinate system for the different input features were being created over the network. This illustrates an important and attractive feature of SOFM applications, in that a multi-dimensional input ensemble is mapped into (one or) two-dimensional space, preserving the topological structure as much as possible. Hsieh and Jourdan (2005) investigated the similarity of watershed s and

hydrologic responses using supervised-unsupervised ANNs. They incorporated GIS into USANNs to quantify the similarity of watershed characteristics. The goal of this approach was to find the best match between the watershed of interest and those contained in a large knowledge base of over one thousand watersheds and to determine the reliability of using “transplant” watershed information during the clustering and classification stages.

Computational procedures and input/output parameters

A computational procedure to perform this effort is shown in Figure 1. The first step is to identify the significant storm parameters and use the resulting surge responses to build the ANNs model. The performance of this ANNs model is critical to assure the right input parameters are selected. The second step of the approach is clustering analysis, using SOFM to separate the similar storm patterns from the knowledge base (input parameters only), and to form a number of subgroups. The third step is to split each subgroup into two components: training and testing storm sets. The ensemble training component from all subgroups along with corresponding surges is the knowledge base which is assumed to represent the required ADCIRC runs, while the ensemble testing component from all subgroups along with corresponding surges are considered to be the unnecessary ADCIRC runs. More ADCIRC runs in the final ensemble testing group means a higher percentage of runs saving that can be obtained under the good performance of testing group from ANNs modeling. Although there is no particular rule to follow how to separate the training and testing group, but at least two numerical model runs from each subgroup need to selected to be the training group if this subgroup contains more than or equal to two numerical model runs. For a large number of numerical model runs, the decision is based on either second level of clustering or the variation for the most sensitive storm parameters.

They are 5 storm input parameters (CpLand (central pressure at landfall), VelAvg (forward translational speed of the storm center, which was assumed to be constant along the specified track), Rm (radius to the maximum winds), MaxWind (maximum wind speed at landfall), Distance (distance from the storm center at landfall to the location of interest), and Angle (angle between the storm center at landfall and the location of interest). The corresponding surge produced by the storm at the location of interest is the output. Since CpLand and Distance are negatively correlated to the surge response, a negative sign for both inputs is taken. The

Angle for each storm is further decomposed into cosine (Angle-x) and sine (Angle-y) components. The first 4 input parameters are considered as global storm parameters while the remaining input parameters (Distance, Angle-x, and Angle-y) are treated as local, or positional, parameters.

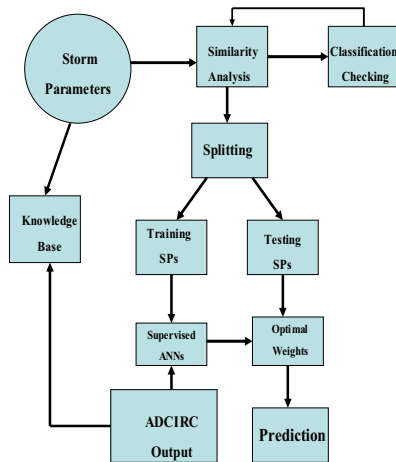


Figure 1. Computational procedure to determine necessary surge model scenario run.

3. Demonstration Example 1 – LACPR Project

This demonstration shows results for a computational point (141, circled in Figure 2) from among many considered in the LACPR study (Figure 2) to illustrate the ANN application concept. Usually, a correlation coefficient analysis is conducted between all the inputs and corresponding output in order to check the sensitivity of the system. Figure 3 shows the most significant input parameter – angle for point 141. Figure 4 illustrates the results of ANNs modeling for point 141. MLP is the training algorithm and total iterations are 5000. With high statistical significance (high correlation coefficient and low mean absolute error, for example), the ANNs modeling proves to be a satisfactory tool to quantify the relationship between inputs and output. The blue lines shows the computed peak surge using the ADCIRC model for all storms while the red line shows the computed peak surge values using ANNs modeling.



Figure 2. Surge response points from LACPR ADCIRC model near New Orleans area.

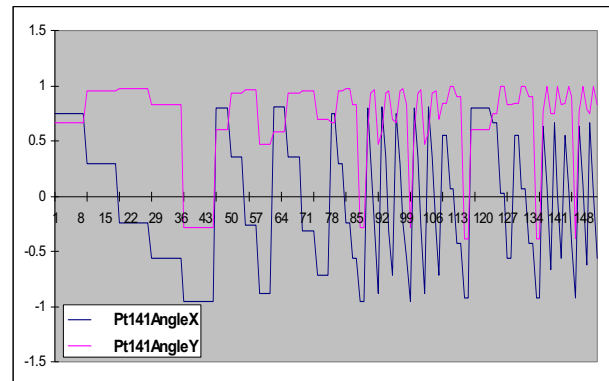


Figure 3. X and Y components for approach angle from point 141 (x-axis represents storm numbers and y-axis represents sine/cosine of the angle, between 1 and -1).

A 5x5 SOFM clustering analysis was then applied to group the dimensional 152x7 information (7 input parameter factors for each of 152 storms). The choice of the size for SOFM process is based on how detail you would like to deal with the clustering from the system. While large matrix may break 152 storms into too many pieces, small matrix may require the second level of clustering. The ratio of total storm numbers to total subgroup numbers estimates the proper matrix size is 5x5 or 4x4. The final destination for each represented storm after iteration process for position adjustment is called “Get Winner”. An

optimal pattern distribution matrix for these 152 storms all reach “WINNER” is shown in Figure 5.

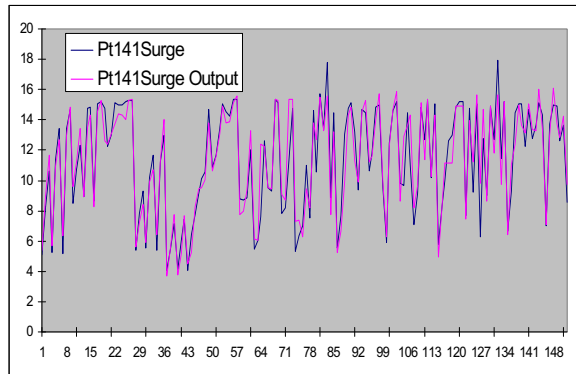


Figure 4. Comparison of results from ANNs- MLP training for point 141 and calculated surges from ADCIRC simulations of 152 storms based on 7 storm input parameters (pink represents ANNs simulation and blue shows ADCIRC results; x-axis represents storm number and y-axis represents surge response (ft)).

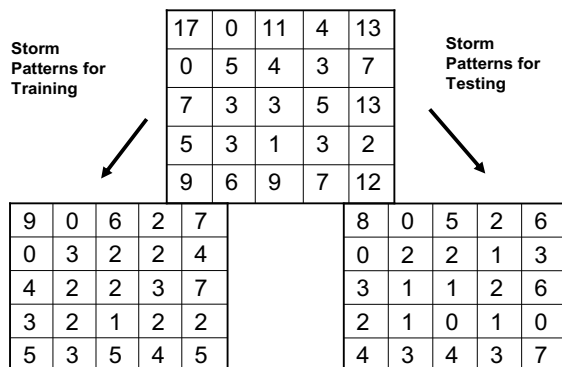


Figure 5. A 5x5 SOFM clustering analysis and its training and testing components fro152 LACPR ADCIRC runs based on 7 storm parameters (point 141).

The number for each grid cell of the matrix shows that similar patterns are found from 7 input parameters. It is noted that the “0” value for a particular grid cell in the matrix indicates that there is no storm falls that specific pattern. It usually happens when too large size of matrix is assigned or too little variation of pattern does exist. The splitting process is then applied to separate the storms within a grid cell into a training component and a testing component after an ensemble process is conducted by collecting the minimum required storm events into the training group and putting the remaining events

into the testing group. . Since the Angle-y was found to be the most significant parameter, it was used as a criterion, including extreme values and part of represented values from this parameter, to determine into which component each storm should go. The lower part of Figure 5 presents the final assignment of storms into the training and testing components. The ANNs training, using 85 selected storms with surge as output was applied to examine performance for the testing component. Figure 6 illustrates performance of the testing component, the 67 selected storms (a correlation coefficient 0.912 was achieved). From this analysis, it is possible to avoid 44 percent (67 out of 152) of ADCIRC simulations, if the storm simulations have to be repeated.

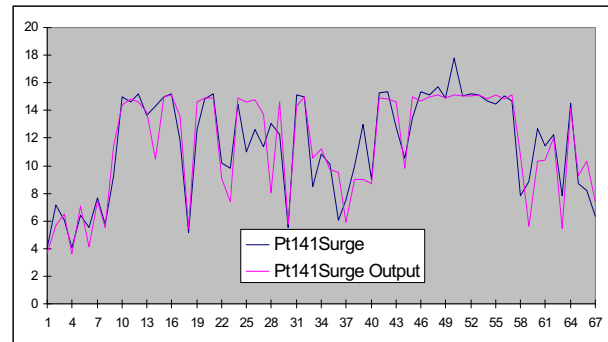


Figure 6. Performance of testing components of 152 storms for point 141 (pink represents ANNs simulation and blue shows ADCIRC results; x-axis represents storm numbers for testing component and y-axis represents surge response (ft)).

Results also suggest that it might be possible to intelligently reduce the number of storms considered in simulations to look at various alternatives. Figure 7 summarizes the surge responses from ADCIRC runs, ANNs simulation, and the combination of ADCIRC runs (training component) and ANNs simulation (testing – prediction component). Closer results are found between ADCIRC runs and the combination approach. Highly correlated relationships between top (numerical simulation) and bottom (the combined approach) figures demonstrate the accuracy of the approach. To compare the surge frequencies computed based upon these three series, the response surges for return periods up to 2000 years are computed. The maximum deviation is about 0.18 m (0.6 ft) between original ADCIRC runs and this combination approach for point 141 (Figure 8).

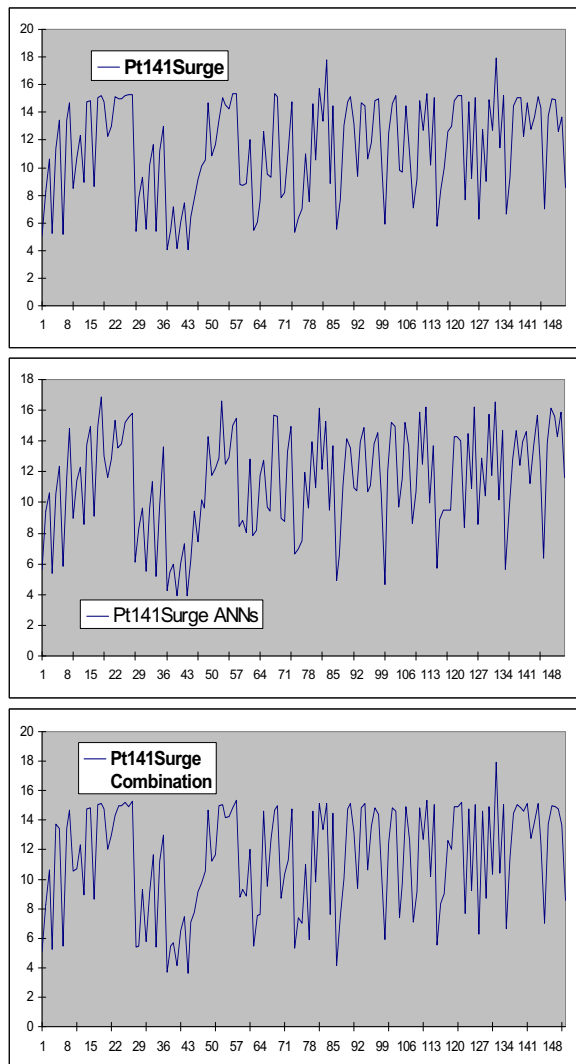


Figure 7. Surge (ft) comparison among ADCIRC simulation, ANNs simulation, and ADCIRC-ANNs combination from 152 storm model runs for point 141 (x-axis represents storm numbers and y-axis represents surge response (ft)).

4. Demonstration Example 2 – Morganza Project

A second example is from the Louisiana Morganza to the Gulf of Mexico hurricane protection project (Figure 9). The main purposes are to reduce hurricane and flood damages in an environmentally sustainable manner in the Houma area. It will protect over 150,000 people and 130 square miles of saline and fresh marshes, farmlands, heavy and light industry, residential, and other developed areas.

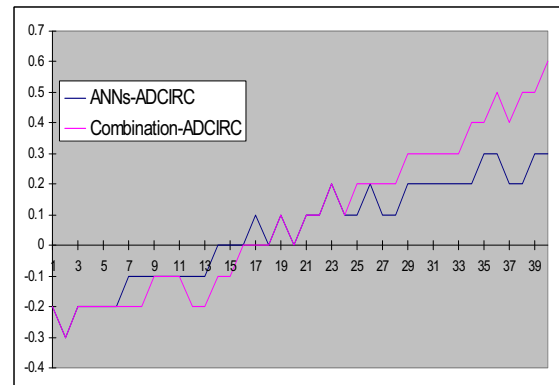


Figure 8. Deviation (in feet) of ANNs-minus-ADCIRC and combination-minus-ADCIRC (x-axis is the number of 50-yr increments in return period) for surge responses (point 141).

Protect development and the remaining fragile hurricane storm surge. The engineering related project features include lock complex, levees, environmental water control, and floodgates at road and waterways. A computational point (point 200; yellow circle – Figure 9) in front of levee is selected to demonstrate the ANNs modeling reliability based on the same type of input/output parameters as in the previous example. Although the clustering, splitting, and testing processes are the same as the previous example, instead of presenting a lattice box, the variation for two selected input parameters during the processes is provided. Figure 10 demonstrates how the input parameter Angle-x varies through clustering and the splitting processes for 359 storms. It is noted that the ordering shown for both is ascending from lower cluster to higher cluster. If the similar patterns are approximately equally distributed into training and testing components, the final patterns for the first-half and the second-half (left and right from green line) should be very close.

Results from this clustering and splitting process analysis suggest that the percentage of model simulations which can be reduced is about 50.4 percent (181 runs out of 359 runs). The final ANNs training results show a correlation coefficient of 0.961 and a testing correlation coefficient of 0.931 (shown in Figure 10). These are better results than in the previous example. The reason could be more training patterns are involved (178 patterns versus 75 patterns), or that the location of the Morganza project is an area in which the storm surge regime is less complex and less sensitive to individual storm parameters than in the previous example where we know from modeling-informed

experience that storm surge can vary considerably east and west of the Mississippi River depending upon the specific characteristics of an approaching hurricane.

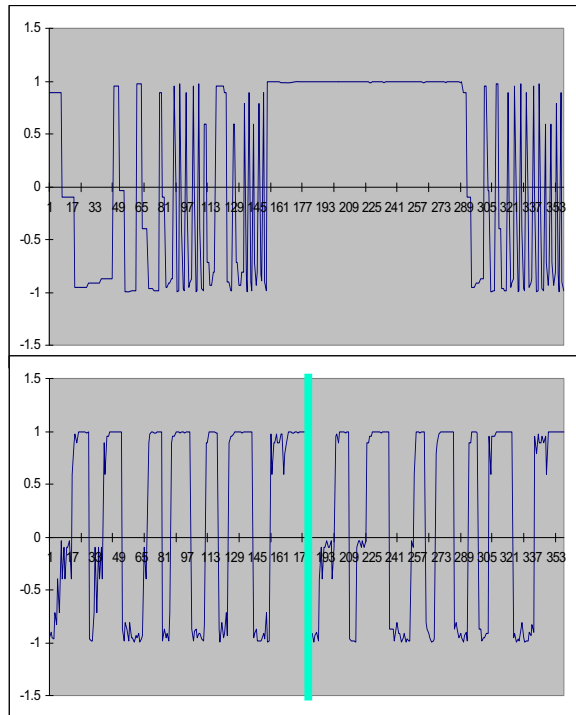


Figure 9. After clustering and splitting (bottom) processes of angle-x input from 359 storm model runs for point 200 (x-axis represents storm numbers and y-axis represents cosine function response of the storm approach angle (between 1 and -1)).

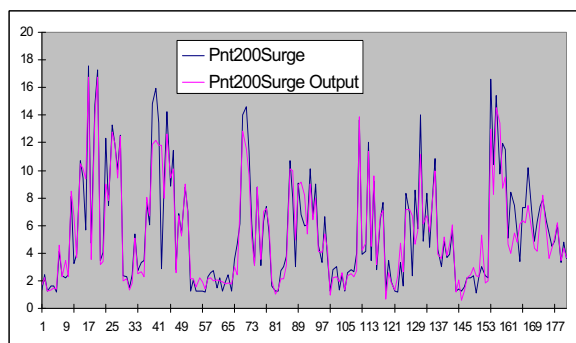


Figure 10. Performance of testing component of 359 storms for point 200 (pink represents ANNs simulation and blue shows ADCIRC results; x-axis represents the testing component of storm number and x-axis represents surge response (ft)).

5. Conclusions

This paper uses unsupervised ANNs to cluster storm patterns. This is based on four global storm parameters and three locals, or positional parameters. The angle between the location of interest and the location of storm landfall was found to be the most sensitive input parameter, due in large part to the influence of the Mississippi River delta and levee system in dictating local surge conditions in southeastern Louisiana. The splitting process is able to separate all storm patterns into training and testing components. The number of storms in the testing component equals the number of numerical runs that can be potentially be reduced by simulating surge through the ANNs model using the training components along with their corresponding surge from actual numerical simulations of storm surge using the ADCIRC model. Two demonstration projects (LACPR and Morganza), results for a single point in each case, show successful application of the developed computational procedures. While point 141 from LACPR project demonstrates reducing model runs about 40 percent of storm model runs, point 200 from the Morganza project demonstrates reduction about 50 percent of ADCIRC model runs. Results showed that the more storm patterns that are involved in the training component, the higher percentage in the reduction of numerical runs, which makes intuitive sense.

The second phase of work is to extend the analysis from single-point to multiple-point basis. It includes the analysis of statistical parameters of surge responses as well as geometry parameters in the analysis which will facilitate improved selection/omission of storms for simulations needed to evaluate multiple alternatives that might or might not involve changes which can directly influence the storm surge itself. A third area of further study is development of guidance to define an optimal storm number selection from the multiple point environments to report to the scientific community.

6. Acknowledgement

The U.S. Army Corps Engineers New Orleans District Office funded this work. Permission was granted by the Chief of Engineers to publish this information.

7. References

- Haykin, S., 1994: Neural networks: A comprehensive foundation, Macmillan Publisher
- Hsieh, B., and Jourdan, M., 2005: "Similarity analysis for military application using unsupervised-supervised ANNs", Proceedings of 25th Army Science Conference, Orlando, Florida
- NeuroSolutions, V5.0, 2003: "Developers level for window", NeuroDimensions, Inc., Gainesville, Florida

Principe, J. ., Euliano, N. R., and Lefebvre, W. C.,
2000: Neural and adaptive systems: fundamentals
through simulations, John Wiley & Sons Publisher