UPDATE OF THE EUROTOP NEURAL NETWORK TOOL: IMPROVED PREDICTION OF WAVE OVERTOPPING

Barbara Zanuttigh [1], Sara Mizar Formentin[1] and Jentsje W. van der Meer [2,3]

The goal of this work is to present a synthesis of the improvements and updates developed to deliver the final version of the ANN tool adopted by the second edition of the wave overtopping manual, EurOtop, released on the internet in 2016. This tool consists of three identical but independent ANNs able to predict the main parameters representative of the wave-structure interaction processes, i.e. the mean wave overtopping discharge, the wave transmission and the wave reflection coefficients. The contribution focuses on the modifications of the ANN architecture carried out since the last ICCE conference to achieve an optimized representation of the wave overtopping, especially in case of low and extreme values of the overtopping discharge. The consistency of the ANN predictions is assessed through an artificial dataset including geometrical and climate input parameters that are varied with continuity, while the robustness of the tool is checked by applying the ANN to selected geometries excluded from the training database.

Keywords: artificial neural networks; wave overtopping; generalization; training; database; wave transmission; wave reflection

INTRODUCTION

Most coastal and harbour structures are constructed primarily to limit wave disruption, like penetration, overtopping and transmission or to prevent flooding. For conceptual design purposes, a simple and rapid approach is to use an Artificial Neural Network (ANN), which is particularly recommended in case of complicated structure geometries and variable wave conditions (EurOtop, 2007).

All the earliest examples of ANNs available from the literature for the prediction of the average overtopping discharge, q (Van Gent et al., 2007; Verhaeghe et al., 2008), including the preliminary version of the ANN tool developed by the authors (Zanuttigh et al., 2014), are affected by an overestimation bias of the low overtopping and a more or less evident scatter associated to the high values of q.

Another relevant issue is the necessity to assess the capability of generalization of such ANNs against data not used for the training. The use of the bootstrap resampling and the presentation of the results as statistical distribution instead of deterministic values constitutes a valid element to at least inform the user when the ANN is out of reach. Indeed, when the predictions of the ANN become unrealistic and unreliable, the confidence intervals widen indefinitely (up to several orders of magnitude around the average value). Yet, a proper a-priori verification of the actual generalization capability would be required before delivering an ANN tool.

In order to overcome part of these problems, the preliminary ANN tool by Zanuttigh et al. (2014) has been upgraded and modified:

- achieving an improved prediction of q in case of low and high overtopping rate by extending the training database (Zanuttigh et al., 2016);
- increasing the ANN capability of generalization without reducing the average accuracy, through a revision of the ANN architecture and training features (Formentin et al., 2017);
- introducing an innovative method to roughly assess the reliability of the predictions, besides the use of confidence bands (Zanuttigh et al., 2016; Formentin et al., 2017).

The aim of this contribution is to present a synthesis of the work done by the authors to upgrade the preliminary ANN presented during the previous ICCE (Zanuttigh et al., 2014), before delivering the final version of the ANN tool (Zanuttigh et al., 2016; Formentin et al., 2017), that has been adopted by the second edtition of the EurOtop manual (2016) and is going to be delivered on the website www.unibo.it/Overtopping-NeuralNetwork. To this purpose, the training database, the features and the performance of the final EurOtop ANN (Zanuttigh et al., 2016; Formentin et al., 2017) are presented in comparison with the preliminary ANN tool (Zanuttigh et al., 2014). Then, the paper focuses on the three main conceptual and methodological modifications introduced in the final ANN with the specific objective to achieve a greater accuracy in the representation of the overtopping and in the capability of predicting new data (i.e. data not used for the training). Afterwards, the consistency of the ANN is proved by checking the trend of its predictions as functions of selected geometrical and climate

[1] DICAM, University of Bologna, V.le Risorgimento 2, Bologna, BO, 40136, Italy
[2] Van der Meer Consulting bv, P.O. Box 11, Akkrum, 8490 AA, The Netherlands
[3] UNESCO IHE, Westvest 7, Delft, 2611 AX, The Netherlands

parameters that are made changing with continuity. An example application of the ANN tool is proposed to test its behaviour when applied to unconventional structure geometries that have been excluded from the training database and that essentially constitute the "white spots" of the field of applicability of the ANN (where the research is now focused). A final paragraph draws some conclusions of the work.

MAIN FEAUTURES OF THE EUROTOP ANN TOOL

The training database

The final database used to train the ANN tool consists of about 13,000 tests on wave overtopping, more than 7,000 tests on wave reflection and 3,500 tests on wave transmission, for a total amount of almost 18,000 tests (Zanuttigh et al., 2016; Formentin et al., 2017). This database is an update and integration of the datasets collected within the CLASH (2004) and the DELOS (Panizzo and Briganti, 2007) projects, joined with a number of further datasets, partially derived from people involved in preparation of EurOtop (2016) or from private communication. The full list of the datasets included in the database can be found in Formentin et al. (2017).

The information contained in the original datasets has been re-ordered, updated and re-organized in the new database to describe the tested wave conditions and the geometry of the structures by means of the same set of physical parameters. The database adopts the same schematization already proposed by the CLASH project (Van der Meer et al., 2009), but includes some (7) new parameters (see Tab. 1). Furthermore, new formulations are introduced for consistently evaluating the average slopes of the structures in the run-up/down area including or excluding the presence of the berm ($cot\alpha_{incl}$ and $cot\alpha_{excl}$, respectively), the average structure roughness factor γ_f and the average size of units in or on the slope D based on their values in the run-up and run-down areas, respectively. The formulations can be found in (Zanuttigh et al., 2016; Formentin et al., 2017).

In total, each test in the new database is described by 41 parameters, of which:

- 11 hydraulic parameters for the characterization of the hydraulic conditions;
- 23 structural parameters describing the geometry of the structure cross-section;
- the test name/ID, the reliability and complexity factors, RF and CF and a label indicating whether the test has been included or not in the training database of the ANN tool (4 general parameters);
- the 3 output parameters, i.e. the average wave overtopping discharge (q, m^3/s per m), the wave reflection and the wave transmission coefficients (K_r and K_t).

The main hydraulic and structural parameters characterising the database are shown in the schematization of Figure 1. The full description of the parameters and the variety of the wave conditions and the structure cross-sections collected in the database are provided in EurOtop (2016), Zanuttigh et al. (2016), Formentin et al. (2017).

The database can be downloaded for free – upon registration – from the website of the EurOtop ANN tool, www.unibo.it/overtopping-neuralnetwork.

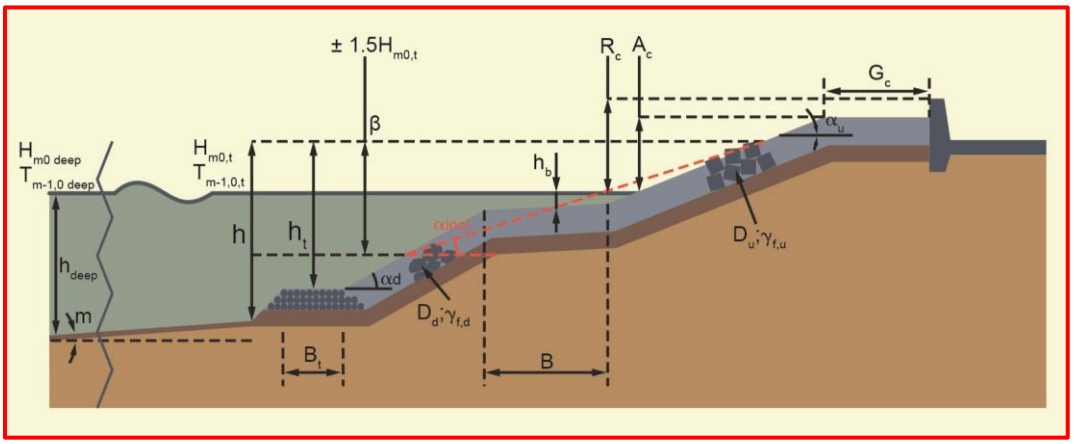

Figure 1. Schematization of the structure based on CLASH, including some of the most relevant geometrical and hydraulic parameters (from EurOtop, 2016).

The Input Parameters

Based on an in-depth sensitivity analysis, the original configuration of 13 dimensionless input parameters identified in Zanuttigh et al. (2013, 2014) has been updated with an optimal set of 15 dimensionless input parameters in the final version of the ANN (Formentin et al., 2017). The full list of the input parameters of the preliminary ANN and the final one is given in Table 1, where the CLASH symbols and terminology are adopted (and here referred in Fig. 1). It is worth to remind that the choice of adopting dimensionless input parameters allows to avoid the use of a scaling rule (typically, the Froude law, as in the original CLASH ANN, by Van Gent et al., 2007) and to get scale-independent predictions. It was already proved in the preliminary ANN (Zanuttigh et al., 2014) that an improved performance is achieved with such a choice of physically-based dimensionless input parameters.

With reference to Table 1, the main differences and novelties characterizing the configuration of the input parameters of the final ANN (Formentin et al., 2017), in comparison with the "preliminary" ANN by Zanuttigh et al. (2014) are resumed in the following.

- The parameter $h_t/L_{m-1,0,t}$ of the preliminary ANN has been replaced by the two parameters $h_t/H_{m0,t}$ and $h/L_{m-1,0,t}$; this change allows to contemporarily account for the effects of the shoaling ($h/L_{m-1,0,t}$) and "homogenize" the use of the significant wave height $H_{m0,t}$ as parameter to non-dimensionalize all the structures heights. Nevertheless, it has been verified that the use of the two parameters leads to an improved performance.
- The new parameter $A_c/H_{m0,t}$ has been introduced to better represent the geometry of the structures with crown walls or promenade, for which the crest emergence A_c differs from the maximum structure emergence comprehensive of the crown wall, R_c (see Fig. 1).
- In the final ANN, the wave length $L_{m-1,0,t}$ – used as scale parameter for the structure widths – is computed on the basis of the wave period at the structure toe ($L_{m-1,0,t} = (g \cdot T_{m-1,0,t}^{\ 2})/2\pi$) instead of using the local wave length reconstructed with specific procedures based on h and $T_{m-1,0,t}$.

Table 1. Input parameters of the preliminary ANN (Zanuttigh et al., 2014) and the final ANN adopted by the EurOtop manual (Zanuttigh et al., 2016). The different parameters are highlighted in green in the final ANN.

#	Preliminary ANN (Zanuttigh et al., 2014)	Final ANN (Formentin et al., 2017)
1	$H_{m,0,t}/L_{m-1,0,t}$	$H_{m0,t}/L_{m-1,0,t}$
2	β [rad]	β [°]
3	$h_t/L_{m-1,0,t}$	$h_t/H_{m0,t}$ t
4	$B_t/L_{m-1,0,t}$	$B_t/L_{m-1,0,t}$
5	$h_b/H_{m0,t}$	$h_b/H_{m0,t}$
6	$B/L_{m-1,0,t}$	$B/L_{m-1,0,t}$
7	$R_c/H_{m0,t}$	$R_c/H_{m0,t}$
8	$D/H_{m0,t}$	$D/H_{m0,t}$
9	$G_c/L_{m-1,0,t}$	$G_c/L_{m-1,0,t}$
10	m	m
11	$cot\alpha_d$	$cot\alpha_d$
12	$cot\alpha_{incl}$	$cot\alpha_{incl}$
13	γ_f	γ_f
14		$A_c/H_{m0,t}$
15		$h/L_{m-1,0,t}$

The ANN architecture

The conceptual layout of the architecture of the final ANN is portrayed in Figure 2. The layout shows the three conceptual layers characterizing the ANN (the input, the hidden and the output layer) and the number of parameters or neuron contained in each layer (15 input parameters, 20 hidden neurons and one output neuron, respectively).

Most of the features characterizing the ANN architecture in its final version are the same of the preliminary version of the ANN (Zanuttigh et al., 2014). Table 2 reports and compares the main features of the two ANNs. The description of the ANN working principle, the meaning and the "role" played by the several elements contained in Table are fully described in Formentin et al. (2017).

From Table 2, it can be observed that, besides the number of the input parameters, the novelties introduced in the final ANN are twofold: the number of the hidden neurons, that has dropped from 40 to 20, and the choice to adopt Weight Factors (WF) during the training phase to "drive" the selection of the data for the training towards those data that are considered "more reliable".

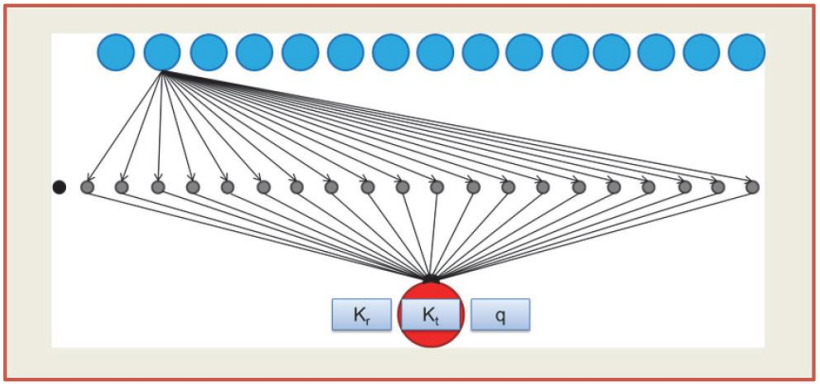

Figure 2. Schematization of the ANN architecture, composed by 15 dimensionless input parameters (blue), 20 hidden neurons (grey) and one output neuron that may be alternatively, q, K_r and K_t (from EurOtop, 2016).

Table 2. Main features of the architecture characterizing the preliminary ANN (Zanuttigh et al., 2014) and the final ANN adopted by the EurOtop manual (Zanuttigh et al., 2016).

Feature	Preliminary ANN (Zanuttigh et al., 2014)	Final ANN (Formentin et al., 2017)
Type of network	multilayer network, based on a "feed-forward back-propagation" learning algorithm	
Number and type of input parameters	13, dimensionless	15, dimensionless
Hidden layer	1 hidden layer with 40 hidden neurons	1 hidden layer with 20 hidden neurons
Output layer	1 output neuron, that alternatively corresponds either to *q*, *Kt* and K_r;	
Training algorithm	Levenberg – Marquardt (Marquardt, 1963; Hagan and Menhaj, 1994);	
Hidden neurons transfer function	Hyperbolic tangent sigmoid function	
Output neuron transfer function	Linear transfer function	
Technique to improve the generalization capability	Use of the "commitment" of networks and bootstrap resampling of the database	
Use of weight factors to drive the selection of data for the training	No	Yes (definition based on Zanuttigh et al., 2016)
Method to deliver the predicted outputs	Statistical distribution (average values, standard deviations and quantiles)	

OPTIMIZATION OF THE PREDICTION OF THE WAVE OVERTOPPING DISCHARGE

This Section aims to provide a synthesis of the modifications introduced in the final ANN for an improved prediction of the wave overtopping discharge. The first and most relevant modification is represented by the update of the training database of *q*, which has been extended to contain all the non-zero available values. Secondly, the number of the hidden neurons has been reduced from 40 to 20 to guarantee a greater capability of generalization, without significant loss of performance in the prediction of the training data. Finally, a new definition of Weight Factors has been proposed and adopted.

Extension to all non-zero values

Following the existing ANNs developed for the prediction of the wave overtopping (Van Gent et al., 2007; Verhaeghe et al., 2008), in the preliminary ANN (Zanuttigh et al., 2014) all the values of $q \leq 10^{-6}$ m^3/s/m were excluded from the training database, according to the assumption that $q=10^{-6}$ m^3/s/m represents a threshold value to distinguish between "negligible" and "significant" overtopping, i.e., $q \leq 10^{-6}$ and $q > 10^{-6}$ m^3/s/m respectively. Such choice was based on the hypothesis that the low overtopping data are likely to be affected by greater measurement errors. However, it was observed that all the above-mentioned ANNs tended to overestimate the low values of q ($q < 10^{-5}$ m^3/s/m), which is logical as the small values were not included in the ANN training. The ANNs predicted an asymptote at the lowest values, resulting in an overestimation bias of $q \leq 10^{-6}$ m^3/s/m.

Therefore, all the non-zero values of *q* have been now included in the training database, determining an increase of the number of the available tests from just over 8,000 up to more than 9,000

tests. The 1,100 additional tests represent a non-negligible part of the final training database, being more than the 13% of the total amount of tests.

Table 3 presents the quantitative results of the final ANN in comparison with the original ANN by Zanuttigh et al. (2014). The results are provided in terms of the root mean square error (*rmse*), the Willmott Index (*WI*) and the coefficient of determination R^2, as described more in detail in the previous work by the authors (Formentin et al., 2017). Through these indices, Table 3 compares the performance of the two ANN s when applied to the prediction of the "original" training database with values of $q>10^{-6}$ m^3/s/m only and to the prediction of a narrower dataset including only data with $q>10^{-3}$ m^3/s/m. Therefore, the preliminary ANN is here applied in the first case to predict the same database used for training ($q>10^{-6}$ m^3/s/m), while the final ANN to predict in both cases two datasets extracted from the whole training database.

This analysis has been performed and here reported to stress that the introduction of the "lower" values of q in the training database: i) significantly reduces the bias, ii) improves the average performance of the network, iii) extends the field of validity of the ANN itself (i.e. the variety of tested conditions which the ANN is able to deal with), iv) allows to achieve an improved predictions of the higher values of overtopping. Indeed, for both the datasets, all the performance indices associated to the final ANN are respectively higher (in case of *WI* and R^2) and lower (in case of *rmse*) than the corresponding indices values associated to the preliminary ANN.

Table 3. Synthesis of the performance of the final ANN and the previous ANN trained with values of $q\geq10^{-6}$ m^3/s/m only. The values of the indices are averaged from 500 bootstrap resamplings.

	Final ANN (trained with q>0)			Preliminary ANN (trained with $q_s\geq10^{-6}$ m^3/s/m only)		
Predicted dataset	*Rmse* [m^3/s/m]	*WI* [-]	R^2 [-]	*Rmse* [m^3/s/m]	*WI* [-]	R^2 [-]
$q_s\geq10^{-6}$ m^3/s/m	0.042±0.002	0.975±0.004	0.91±0.01	0.052±0.005	0.975±0.006	0.90±0.02
$q_s\geq10^{-3}$ m^3/s/m	0.026±0.004	0.96±0.02	0.90±0.1	0.038±0.007	0.95±0.04	0.88±0.08

The quantitative results are highlighted by the diagrams of Figure 3, that compare the experimental and predicted values of q derived from the two ANNs for the two datasets (respectively, dataset with $q_s\geq10^{-6}$ to the left and $q_s\geq10^{-3}$ to the right). From both the diagrams it can be concluded that the 90% confidence bands associated to the final ANN (red) are narrower. Moreover, the diagram to the left shows a significant reduction of the bias when the final ANN is used (blue), while in the diagram to the right the scatter associated to the preliminary ANN (magenta) is higher, especially in case of extreme overtopping values ($q_s\geq0.01$ m^3/s/m). The predictions of the extreme q derived from the final ANN are closer to the confidence bands but still tend to underestimate a few values of q_s around 0.05-0.1 m^3/s/m.

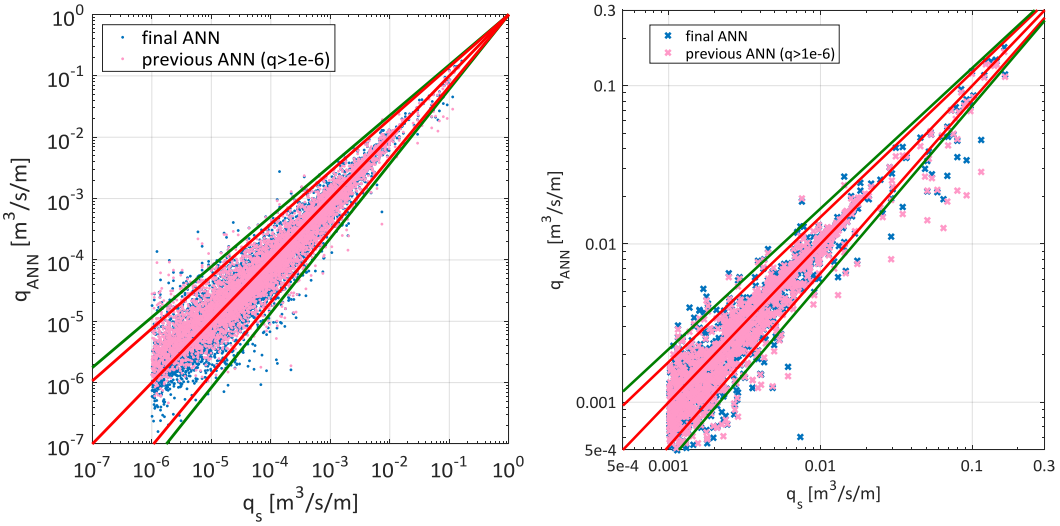

Figure 3. Comparison among predicted values of q (denoted by the 'ANN' subscript) and corresponding measurements ('s' subscript). Predictions derived from the final ANN trained with all the non-zero values (blue) and from the previous ANN trained with values of $q\geq10^{-6}$ m^3/s/m only (pink). The 90% confidence bands are shown for the final (red) and the previous ANN (green). To the left: ANNs applied to the prediction of $q_s\geq10^{-6}$ m^3/s/m; to the right: ANNs applied to the prediction of $q_s\geq10^{-3}$ m^3/s/m.

It is worth to remind that the zero values cannot be included in the training database and that the ANN would never be able to provide a prediction of q identically equal to zero (the discussion about the representation of the zero values is given in Zanuttigh et al., 2016). Therefore, it is up to the ANN tool user to decide whether a predicted value is or is not negligible, i.e. whether it can be considered as "zero" or "non-zero".

The number of the hidden neurons and the overfitting

Once defined the preliminary layout of the ANN tool (Zanuttigh et al., 2014), a series of analysis has been performed to verify its "capability of generalization" (Zanuttigh et al., 2016; Formentin et al., 2017). Note that such analysis was performed after the optimization of the input parameters from 13 to 15. Since all the available datasets were included in the training database to provide the ANN tool with the maximum possible information, and no new data were available, the ANN was re-trained on modified databases from which some selected datasets were eliminated (one per time). The following criteria were adopted to identify these datasets: i) a dataset was removed if at least another dataset of "similar" tests (i.e. similar wave conditions and structure type) was available and left in the training database; ii) the dataset/s to be removed should represent a significant sample of the different varieties of structure types included in the database.

The predictions of these datasets by means of the re-trained ANN were then checked against the experimental values, showing significantly higher standard deviations and lower performance with respect to the average performance of the ANN predictions of the training data. This result suggested that the ANN might be over-trained, i.e. trained with either too many input parameters or hidden neurons (more details about overfitting and overtraining can be found in Formentin et al., 2017). Since the input set was already optimized, the revision focused then on the number of the hidden neurons. Specifically, two attempts have been made to retrain the the ANN with 20 and 30 (instead of 40) hidden neurons.

Finally, it was found that the ANN trained with 20 hidden neurons provided the best compromise between accuracy in the prediction of the training data and capability of generalization. Indeed, the ANN with 20 hidden neurons provided the most accurate and least biased predictions of the datasets excluded from the training with respect to the others. This can be appreciated in Figure 4, which compares the ANN predictions and the measurements for a dataset of rocks not used in training. Quantitatively, the value of the *rmse* equals 0.041 and 0.028 in case of 40 and 20 hidden neurons respectively. The predictions of the ANN with 40 hidden neurons is also characterized by wider confidence bands, i.e. wider standard deviations.

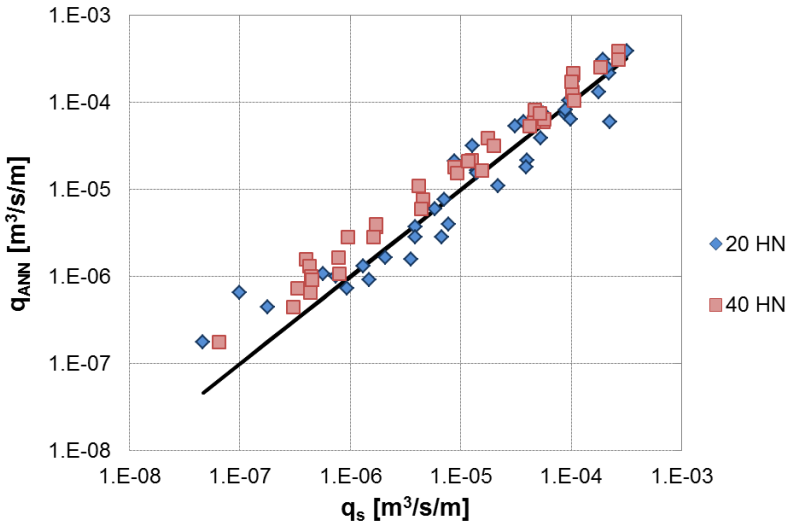

Figure 4. Comparison among ANN predictions (q_{ANN}) and measurements (q_s) for a dataset of rocks not used in training. The predictions are derived from the ANN trained with 20 and 40 hidden neurons.

As for the performance on the training database, the comparison of the *rmse* values as a functions of the number of hidden neurons is provided in Formentin et al. (2017). From this work (more precisely, in Figure 5), it can be observed that the *rmse* increases from 0.045 to 0.051, when the number of the hidden neurons drops from 40 to 20. However, the standard deviation associated to the *rmse* decreases from 0.005 to 0.003 with the decreasing hidden neurons. The average value of *rmse* for 20 hidden neurons still falls in one standard deviation interval around the average *rmse* associated to the case of 40 hidden neurons.

For this reason, in the final version of the ANN, only 20 hidden neurons are included. Definitely, it is strongly recommended to perform such kind of analysis – i.e. to test of the actual ANN capability of predicting data different from the training database – besides following the common methodology to assess the optimal number of hidden neurons on the basis of the trend of the decreasing *rmse* at the increasing number of hidden neurons proposed by Van Gent et al. (2007) and by Panizzo and Briganti (2007).

The new Weight Factors

The first ANN developed for the prediction of q (Van Gent et al., 2007) was trained with Weight Factors (WF). This means that the selection of the data for the training was not made completely randomly, but was somehow driven towards the data considered more "reliable" through the introduction of a coefficient defined as:

$$WF=(4-RF)\cdot(4-CF), \qquad (1)$$

where RF and CF are the "Reliability" and the "Complexity" factors, two parameters included in the database to characterize, respectively, the degree of reliability level of the measured or estimated structural and hydraulic parameters and how well a structure geometry could be described by the geometrical parameters. The values of RF and CF vary from 1 to 4, i.e. from the more reliable or simple to the more unreliable or complex situation, where a value of 4 gives an unreliable or too complex structure, not to be used further.

Since the CLASH (2004) definition of the WF (Eq. 1) resulted in an unbalanced distribution that tended to give too much weight to the data with high WF (the most reliable data), the preliminary ANN developed by the authors (Zanuttigh et al., 2014) was trained without such WF. In order to introduce the WF in the training, but at the same time smooth their effects, a new definition of the WF has been conceived and proposed in the last version of the ANN tool (Zanuttigh et al., 2016):

$$WF=6/(RF+CF),\ WF=0,\ \text{if RF and/or CF} = 4 \qquad (2).$$

This new definition is still based on RF and CF, but realizes a different sorting and a different range of the WF values. According to Eq. (2), WF may assume 6 values, spacing between 0 and 3, where 0 is attributed to the tests which are too unreliable (RF=4) or too complex (CF=4) to be included in the training. The CLASH WF (Eq. 1) recognised instead 7 classes of values, from 0 to 9. The comparison among the distributions of the WF values throughout the database is provided in the pie-charts of Figure 5.

Based on Eq. (2) the maximum WF (3) is only three times the minimum WF (1), while with the original definition of Eq. (1), the maximum WF (9) is 9 times the minimum (1). From a probabilistic approach, this means that, according to the adopted definition, a test has 3 or 9 times more chance to be selected for the training. Being the total number of the tests available for the training equal to more than 9,000, the probability for a single test to be selected without a WF is approximately $1\cdot10^{-4}$, while it rises up to $3.2\cdot10^{-4}$ and nearly to 10^{-3} respectively with a WF = 3 and with a WF=9, i.e. it increases up to one order of magnitude.

The effects of an unbalanced weighting system induce a worsening of the ANN performance that is probably due to the under-representation of all those data with low WF (Zanuttigh et al., 2016). On the contrary, the new definition of Eq. (2) induces a slight improvement of the performance and allows to account for the degree of reliability attributed to the tests. For this reason, in the final version of the ANN tool adopted by the EurOtop manual, the trained was based on the WF as defined in Eq. (2).

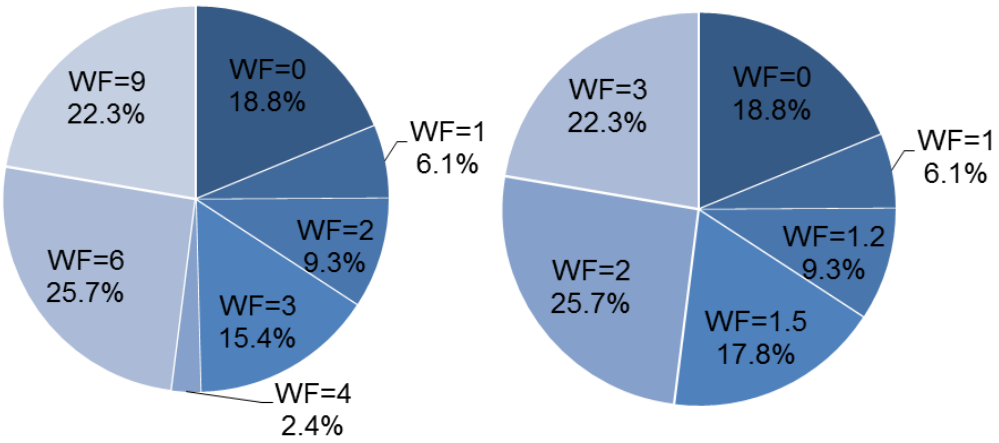

Figure 5. Distribution of the WF values throughout the overtopping database. To the left, the original definition of WF from the CLASH project (here Eq. 1); to the right, the new definition of WF by Zanuttigh et al., 2016 (here eq. 2).

ASSESSMENT OF THE ANN TOOL PERFORMANCE

Generally, the performance of an ANN is assessed based on the accuracy of the ANN predictions of the same data used for the training. This is generally due to the necessity to include in the training database all the available tests, in order to make the ANN as much "robust" as possible. This in turn leads to the lack of new data to evaluate the ANN behavior beyond its training database.

This Section proposes two methodologies useful for the assessment– or at least for a qualitative evaluation – of the performance of the ANN tool when applied to the prediction of new data.

Example application of the ANN tool

In Zanuttigh et al. (2016) and Formentin et al. (2017), for the first time the idea was proposed to eliminate some complete distinct datasets from the training database, re-train the ANN with the narrower database, and then apply the re-trained ANN to the prediction of those datasets removed from the training. This solution represented a compromise to test the ANN capability of generalization in case of the lack of new experimental data.

The reliability of the ANN predictions can then be assessed by analysing the trends of the predicted values as function of a single input parameter that varies with continuity and by comparing it with the theoretical trend that is expected from the physical process.

For this purpose, the ANN tool has been tested against the prediction of a single test (a rubble mound slope, with a toe protection and a wave wall, and a revetment of X-blocks) whose original set of geometrical features and wave attack conditions has been artificially modified in order to evaluate the response of the ANN tool at:

- the widening of the toe protection (B_t);
- the increase of the obliquity of the incoming waves direction (β).

The values of the original and modified parameters characterizing the tests are reported in Table 4, while the cross-section of the original structure is shown in Figure 6. It should be noted that the effect of the two input parameters is investigated separately, i.e. one input was made varying while keeping all the other inputs were kept constant. This allows to assess the sensitivity of the ANN to each parameter separately and to contemporarily keep the new input configuration "close" to the ANN domain of validity. Indeed, the modification of the input parameters values generates configurations that may exceed the training ranges of the ANN, while the ANN response is reliable as long as it works with input configurations that are "similar" to any of the tests in the training database.

The results of the application shown in Figure 7, where the trends of the predicted values of q are shown as functions of β and B_t, to the left and to the right graph, respectively. In both the diagrams, the average predictions (dark red) are plotted together with predicted values corresponding to the 5% and 95% percentiles, in order to show the width of the 90% confidence bands associated to the predictions.

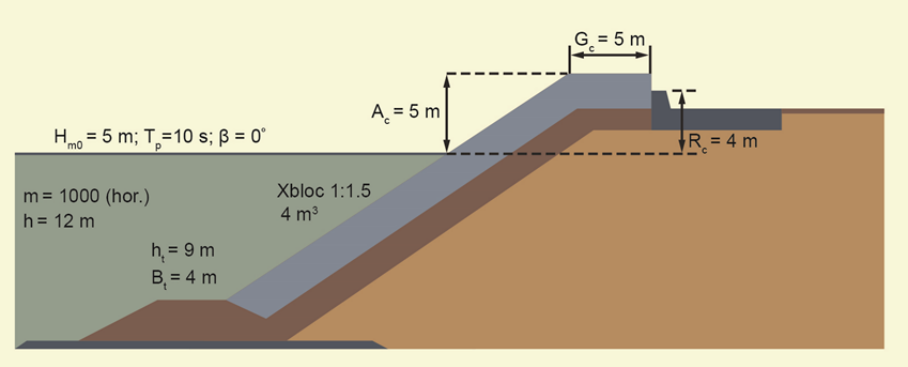

Figure 6. Cross-section of the structure used to test the continuity of the ANN tool. From EurOtop (2016).

Table 4. Values of the hydraulic and structural parameters characterizing the tests artificially built up for testing the ANN continuity. The parameters with variable values are highlighted in red.

Hydraulic and Structural Parameters	Values
m	1000
h [m]	12
$H_{m0,t}$ [m]	5
$T_{m-1,0,t}$ [s]	9.1
β [°]	[0:10:80]
h_t [m]	9
B_t [m]	[4:0.5:10]
h_b [m]	0
B [m]	[0 ; 30]
$\cot(\alpha_d)$	1.5
$\cot(\alpha_u)$	1.5
γ_{fd}	0.49
γ_{fu}	0.49
D_d [m]	1.58
D_u [m]	1.58
A_c [m]	5
R_c [m]	4
G_c [m]	5

By observing Figure 7, the following considerations can be drawn.

- The response of the ANN is monotonic in both the applications and consistent with the physical process. Indeed, the increase of β induces a significant (more than one order of magnitude) reduction of q: the greater the wave obliquity the wider the structure area exposed to waves the higher the wave energy dissipation induced by friction. On the contrary, q is almost constant with increasing B_t, as expected because the toe of the structure is deeply submerged (see Figure 5) and sensibly below the wave run-up/down area (h_t = 9 >>7.5 m, where 7.5 m = 1.5·$H_{m0,t}$ is the rough approximation of the extension of the run-down area).
- The 90% confidence band widens with increasing β and B_t: The farther from the original input configuration, the wider the confidence bands and therefore the less reliable the prediction. The widening is significantly more evident in case of β than in case of B_t. Actually, the cases of oblique wave attack represents only 14% of the whole database (see Zanuttigh et al., 2016), and just a few tens of data are characterized with values of β > 60°. The width of the confidence bands, especially if compared with the average width of the bands characterizing the ANN predictions of the training data, should be used as a powerful indicator of the reliability of the predictions themselves. In this example, the width of 90% confidence bands is less than one order of magnitude for the original input configuration (which belongs to the training database), while it increases up to more than two orders of magnitude for the configuration with β = 80°. On the contrary, the widening of B_t does not increase significantly the bands width, that are still within one order of magnitude even when B_t = 10 m Therefore, it should be concluded that the predicted values of q are more reliable in the second application than in the first one, even if the ANN response is continuous and consistent.

- The symmetry of the confidence bands around the average in Figure 8 suggests that the distribution of the predictions throughout the bootstrapping resamplings is symmetric as well. This is another indicator of the reliability of the prediction.

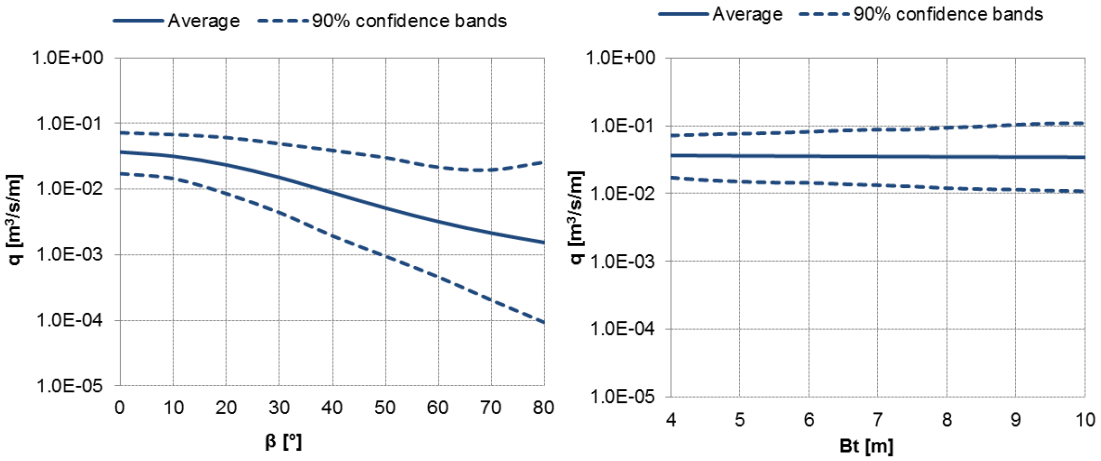

Figure 7. Trends of the predicted *q* values as functions of the wave obliquity (*β*, left) and the structure toe width (B_t, right). In both the panels, the average *q* values (dark red) are plotted together with the corresponding 90% percentiles (light red) to show the 90% confidence bands width.

Prediction of special data not used for training

The database was collected for training the ANN. However it includes also interesting and reliable tests characterised by specific features that place them out of a possible training set. That is why in the database there are core data and non-core data (Zanuttigh et al., 2016). These special "datasets" – which were given a special "label" to exclude them from the training database – consist of:

- crown walls including a small parapet (or "bull nose") on top, label BN;
- structures with a double seawall and/or a still basin in the middle, label D; note that in all these tests the first wall is submerged and the still basin is of limited width, where present,
- structures with "drainage holes" on the crest, label DH (these tests were excluded because it is impossible to quantify the effect of the reduction of *q* due to the drainage with the selected input set);
- tests performed at prototype scale, label T (further detail about scale and model effects can be found in Zanuttigh et al., 2016);
- tests performed with a wind simulator reproducing the effect of the wind with different velocities, label W;
- cases of overtopping discharge measured by means of a recirculation system (instead of through the measure of the overtopping volume), label RS.

While some of these datasets are completely out of the ANN field of validity (e.g., the W-tests, especially if considering variable wind velocities), some others could be used to verify whether the ANN response is at least in line with what expected from the physical process.

Therefore, as example application, the ANN tool is here tested against the prediction of three of these "special" datasets selected from the BN, the D and the DH-tests, respectively. The BN-tests belong to the database of Van Doorslaer et al. (2015), while the D and the DH-tests are from private communications. It is expected that the ANN overestimates the measured values of *q* for all the cases, because these elements, i.e. the bull nose, the second wall and the drainage holes all reduce the wave overtopping.

The cross-sections of the selected datasets are given in Figure 8, while the results of the application are given in Figure 9, in terms of comparison among the ANN predictions and the corresponding measurements. In Figure 9, the 90% confidence bands (dashed red lines) correspond to the average performance of the ANN when used to predict the whole training database and are included for comparison.

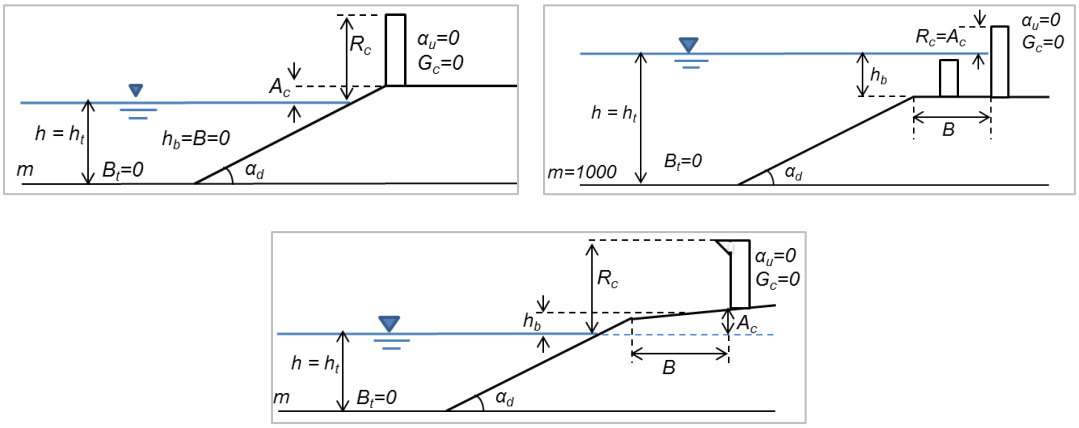

Figure 8. From top to bottom, and from left to right, cross-sections of the structures used as example application of the ANN tool for the DH, the D and the BN-tests. For each cross-section, the geometrical parameters indicate the adopted schematization.

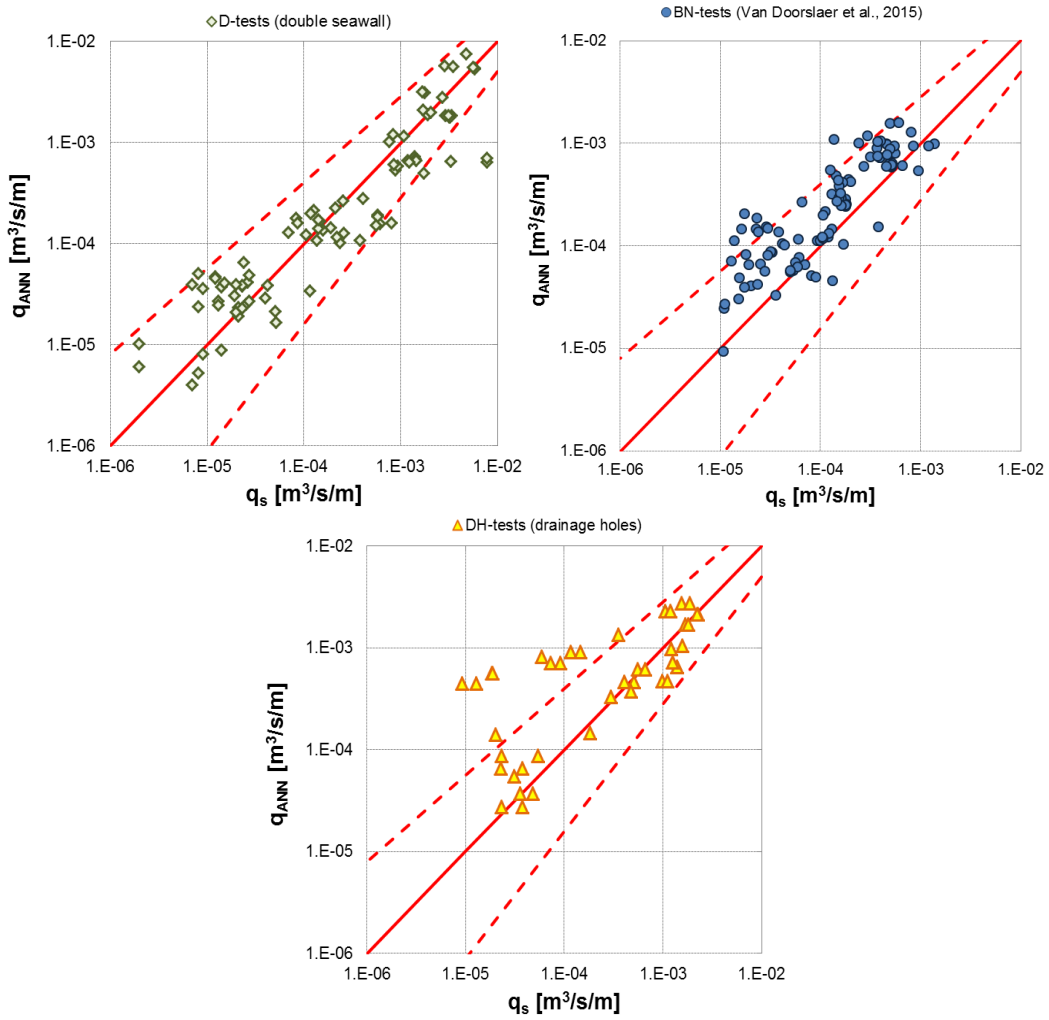

Figure 9. Comparison among ANN predictions of *q* (q_{ANN}) and corresponding measurements (q_s) of the three selected datasets for the ANN example application. The 90% confidence bands (dashed red lines) correspond to the average performance of the ANN when used to predict the training data.

For all the datasets Figure 9 shows that most of the ANN predictions fall within the average confidence bands or close to, revealing a good (D-tests) or fair (BN and DH-tests) performance even if the structure types include elements that are not schematized by the ANN input parameters.

Actually, a very good agreement among predictions and measurements is observed for the D-tests, as all the points are symmetrically aligned around the bisector line (continuous red line) with the exception of three outliers. For this application, no overestimation bias is evident, suggesting that the effect of the first (off-shore) wall (see Figure 8, top-right panel) is limited, as it could have been expected since it is working in submerged conditions. Therefore the basin on the structure crest is always filled in by water and operates full-section, inducing a very modest dissipation of wave run-up because of the limited width.

Differently, for both the BN and DH-tests (see Figure 9), the ANN predictions tend to be aligned above the bisector line, revealing that the ANN systematically overestimates the experimental values of q. The overestimation of the predictions could be taken as a rough indicator of the efficiency of the bull noses and of the drainage holes. By observing the two diagrams of Figure 9, it seems that the reduction effect induced by the bull nose is more "constant" or "linear" than the effect of the drainage holes, because the predictions of the BN-tests are biased but not scattered, while the predictions of the DH-tests do not follow a single trend.

Overall, these results suggest that the ANN response is not only meaningful from a physical point of view, but also satisfactorily accurate, even if the ANN has been applied beyond its range of training.

CONCLUSIONS

This contribution presented the main features characterizing the final version of the ANN tool recently developed by the authors for the prediction of q, a K_r and K_t (Zanuttigh et al., 2016; Formentin et al., 2017) in comparison with the preliminary ANN tool (Zanuttigh et al., 2014).

Such tool – that has been adopted by the second edition of the EurOtop (2016) manual – has achieved an improved prediction of the low and extreme values of q (Zanuttigh et al., 2016) and an improved capability of generalization (Formentin et al., 2017). These achievements have been pursued through a series of conceptual and methodological modification of the preliminary ANN architecture and its training database.

One of the most relevant issues is the extension of the training database to all the non-zero values of q (Zanuttigh et al., 2016). Differently from the previous ANN tool, that was trained with q-values $>10^{-6}$ m^3/s/m only, the new ANN is not only able to represent the cases of low overtopping ($q<10^{-6}$ m^3/s/m), but can also provide less biased and more accurate estimations of the "intermediate" (10^{-4}-10^{-6} m^3/s/m) and high/extreme values of q ($>10^{-3}$ m^3/s/m).

The introduction of a new definition of Weight Factors (Zanuttigh et al., 2016) to drive the selection of the data for the training ensures that the more "reliable" data are weighted more in balance, and therefore avoids the risk of their overestimation.

The improvement of the predictions is also due to the revision of the input parameters, that are now made dimensionless through the same criterion, i.e. that all the parameters representing the structure heights are made dimensionless with the wave height $H_{m0,t}$ (to represent the local breaking that may occur over the several parts of the structure – the foreshore, the toe, the berm), while the structure widths are scaled with the wave length $L_{m0,t}$ (to account for the induced local reflection).

Another important modification concerns the number of the hidden neurons, that have been halved from 40 to 20. It has been verified that this reduction does not cause a significant loss of accuracy in the prediction of the data used for the training and, at the same time, allows to achieve an improved capability of generalization.

Particular attention is paid to the assessment of the reliability of the predictions when the ANN is applied to new input scenarios not used for the training. It is suggested to compare the 90% confidence bands associated to each prediction with the bands characterizing the average performance of the ANN when applied to predict the same data used for the training.

Finally, the robustness and the consistency of the ANN predictions may be verified by applying the tool to a selection of available datasets not used for the training and/or to artificial cases.

ACKNOWLEDGMENTS

The support of the European Commission through Contract 244104 THESEUS ("Innovative technologies for safer European coasts in a changing climate"), FP7.2009-1 Large Integrated Project, is gratefully acknowledged.

gratefully acknowledged.

REFERENCES

CLASH, 2004. Crest Level Assessment of coastal Structures by full scale monitoring, neural network prediction and Hazard analysis on permissible wave overtopping. EC-contract EVK3-CT-2001-00058. ⟨www.clash-eu.org⟩

EurOtop (2016) Manual on Wave Overtopping of Sea Defences and Related Structures. An Overtopping Manual Largely based on European Research, but for Worldwide Application, Allsop, N. W. H., Bruce, T., DeRouck, J., Kortenhaus, A., Pullen, T., Schüttrumpf, H., Troch, P., van der Meer, J. W. & Zanuttigh, B. , www.overtopping-manual.com

Formentin S.M., Zanuttigh B. & Van der Meer J.W. (2017). A neural network for predicting wave reflection, overtopping and transmission, *Coastal Engineering Journal*, 59, No. 2 (2017), 1750006, 31 pp.

Hagan, M.T., and Menhaj M. (1994). Training feed-forward networks with the Marquardt algorithm, *IEEE Transactions on Neural Networks*, 5(6), 989-993.

Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters, *SIAM Journal on Applied Mathematics* 11 (2): 431–441. doi:10.1137/0111030.

Panizzo, A. & Briganti, R. (2007). Analysis of wave transmission behind low crested breakwaters using neural networks, *Coastal Engineering* 54, 643–656.

Van der Meer, J.W., Pullen, T., Allsop, N.W.H., Bruce, T., Schüttrumpf, H., Kortenhaus, A. (2009). In: Kim, Young C. (Ed.), Prediction of overtopping. Chapter 14 in *Handbook of Coastal and Ocean Engineering*. World Scientific, 341–382.

Van Doorslaer, K., De Rouck, J., Audenaert, S. & Duquet, V. (2015). Crest modifications to reduce wave overtopping of non-breaking waves over a smooth dike slope, Coastal Engineering, 101, 69–88, doi: 10.1016/j.coastaleng.2015.02.004.

Van Gent, M.R.A., van den Boogaard, H.F.P., Pozueta, B., Medina, J.R. (2007). Neural network modelling of wave overtopping at coastal structures. *Coastal Engineering*, 54, 586–593.

Verhaeghe, H., De Rouck, J. and Van der Meer, J.W. (2008). Combined classifier–quantifier model: a 2-phases neural model for prediction of wave overtopping at coastal structures. Coastal Engineering 55, 357–374.

Zanuttigh, B., Formentin, S.M. and Briganti R. (2013). A Neural Network for the prediction of wave reflection from coastal and harbor structures. *Coastal Engineering* 80, 49-67.

Zanuttigh B., Formentin S.M., & Van der Meer J.W. (2014). Advances in modelling wave-structure interaction through Artificial Neural Networks, Coastal Engineering Proceedings, 1(34), structures.69. doi: http://dx.doi.org/10.9753/icce.v34.structures.69.

Zanuttigh B., Formentin S.M., & Van der Meer J.W. (2016). Prediction of extreme and tolerable wave overtopping discharges through an advanced neural network, *Ocean Engineering*, 127, 7-22.