

# Reliable prediction of wave overtopping volumes using Bayesian neural networks

G.B. Kingston, D.I. Robinson, B.P. Gouldby & T. Pullen

Reproduced from:  
Flood Risk Management - Research and Practice  
Proceedings of FLOODrisk 2008  
Keble College, Oxford, UK  
30 September to 2 October 2008



## Reliable prediction of wave overtopping volumes using Bayesian neural networks

G.B. Kingston, D.I. Robinson, B.P. Gouldby & T. Pullen  
*HR Wallingford, Wallingford, United Kingdom*

**ABSTRACT:** Artificial neural networks have been used successfully for generating predictions of wave overtopping volumes, which are required to appropriately design coastal structures as well as support associated flood risk analyses. It is particularly important to assess the uncertainty associated with such predictions, given the complexity of the modelling problem and the difficulty in obtaining accurate measurements of the large number of variables needed to estimate wave overtopping. The bootstrapping method previously used to estimate the uncertainty associated with ANN overtopping predictions does not, however, fully capture the total prediction variance. In this paper, Bayesian ANN methods are used to improve the reliability and robustness of wave overtopping predictions and to provide accurate estimates of the associated prediction uncertainty.

### 1 INTRODUCTION AND BACKGROUND

Climate change and rising sea levels are increasing flood risks in coastal areas. In order to appropriately assess and protect against such risks, reliable predictions of wave overtopping volumes at coastal structures are required. However, the complexity of the physical processes involved and the highly nonlinear dependence between the wave-structure characteristics makes the development of robust and reliable overtopping prediction models a nontrivial problem, with many of the models traditionally used for this purpose being applicable only over a restricted range. Moreover, the accuracy of overtopping predictions is often further hindered by the difficulty in obtaining accurate measurements of the large number of variables required. As such, it is of utmost importance to characterise the uncertainty associated with overtopping predictions, as suppressing this information can create a false sense of security in the predictions generated. As a consequence, coastal structures may be inappropriately designed and maintained; in turn, resulting in dangerous conditions in cases of storm surges, wave attack and flooding.

In the attempt to address this problem, one of the main objectives of the European CLASH project (EVK3-CT-2001-00058) was to develop a generic method for estimating wave overtopping volumes and their associated uncertainty (De Rouck & Geeraerts, 2005; Van Gent et al. 2005). Wave overtopping data sets from research institutes and universities around the world were collated, resulting in a database containing more than 10,000 overtopping measurements

(henceforth, referred to as the CLASH database) on which to base the model. An artificial neural network (ANN) was chosen as the generic prediction tool, as these models are particularly suited to modelling complex and nonlinear problems when available data are abundant. To measure the reliability of the overtopping predictions, a bootstrapping procedure was applied, whereby 500 resamples of the original CLASH dataset were used to train (calibrate) the ANN model to provide an ensemble of estimated mean overtopping discharge rates and associated confidence limits.

There are, however, a number of limitations associated with bootstrapping procedures, which make them less than ideal for quantifying the uncertainty associated with ANN predictions. Firstly, such procedures only capture one aspect of prediction uncertainty. That is, they only measure the accuracy of the estimated model with respect to the mean of the target distribution, but do not measure the accuracy with which the actual target data (measured overtopping volumes) can be predicted (Heskes, 1997). In fact, for the CLASH ANN model, Van Gent et al. (2007) did not report on the prediction limits estimated using the bootstrapping procedure, but instead relied on an indicative prediction error band to assess the accuracy of the overtopping estimates. To capture both aspects of prediction uncertainty, and provide better coverage of the true range in which a prediction might lie, prediction limits must include an additional estimate of the noise inherent to the problem. Secondly, the first aspect of prediction uncertainty may be overestimated

due to the learning procedure becoming stuck in local minima in the error surface rather than converging to the (near) global minimum. In this case, not only are inaccuracies in the estimated model captured, but also inaccuracies in the training procedure. Finally, bootstrapping methods are computationally intensive and time consuming. For example, in the CLASH project, 500 bootstraps were used to assess uncertainty in the model, meaning that the ANN model was retrained 500 times. If new data become available and the model requires updating, this could take several days with a bootstrap method.

Whilst the importance of providing estimates of prediction uncertainty is increasingly being recognised in flood risk analysis, it is essential that the derivation of these estimates is as reliable as possible. This paper details the development of a Bayesian artificial neural network (ANN) model used to improve the reliability and robustness of wave overtopping predictions and to provide accurate estimates of the associated prediction uncertainty resulting from uncertainty in the data and the difficulty of accurately capturing the wave overtopping process. An advantage of Bayesian methods over bootstrapping for ANN uncertainty assessment is that uncertainty in the model parameters is handled explicitly by estimating weight distributions which include all weight vectors that provide a reasonable fit to the data, rather than repeatedly estimating point values that provide the best fit to resamples of data. Furthermore, the noise variance is also estimated using this procedure, enabling prediction limits to be generated that include the two aspects of prediction uncertainty described above.

## 2 BAYESIAN NEURAL NETWORKS

### 2.1 Concept

The concept behind the Bayesian modelling framework is Bayes' theorem, which states that any prior beliefs regarding an uncertain quantity are updated, based on new information, to yield a posterior probability of the unknown quantity. In terms of an ANN, Bayes' theorem can be used to estimate the posterior distribution of the network weights  $\mathbf{w} = \{w_1, \dots, w_d\}$  given a set of  $N$  target data  $\mathbf{y} = \{y_1, \dots, y_N\}$  as follows:

$$p(\mathbf{w}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{w})p(\mathbf{w})}{p(\mathbf{y})} \quad (1)$$

In this equation,  $p(\mathbf{w})$  is the *prior* distribution, which describes any knowledge of the weight values before observing the data;  $p(\mathbf{y}|\mathbf{w})$  is known as the *likelihood*

function and is obtained by comparing the observed data  $\mathbf{y}$  to the model outputs  $\hat{\mathbf{y}}$ . This is the function through which the prior knowledge of  $\mathbf{w}$  is updated by the data. The denominator  $p(\mathbf{y})$  is a normalising constant known as the marginal likelihood, or *evidence*, of the model.

### 2.2 Bayesian training

For complex models like most ANNs, (1) is analytically intractable. To overcome this problem, the Markov chain Monte Carlo approach developed by Kingston et al. (2005) is used to generate samples from the posterior weight distribution. Assuming that the residuals between the observed data and the model outputs are normally and independently distributed with zero mean and constant variance  $\sigma_y^2$ , the likelihood function is given by:

$$p(\mathbf{y}|\mathbf{w}, \sigma_y^2) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \prod_{i=1}^N \exp\left(-\frac{[y_i - f(\mathbf{x}_i, \mathbf{w})]^2}{2\sigma_y^2}\right) \quad (2)$$

where  $f(\mathbf{x}_i, \mathbf{w})$  is the ANN output given the  $i$ th input vector  $\mathbf{x}_i$ .

To define a lack of prior knowledge about the weights and to prevent overfitting of the data, a hierarchical prior distribution in the form:

$$p(\mathbf{w}) = \frac{1}{\sqrt{2\pi\sigma_w^2}} \prod_{g=1}^4 (2\pi\sigma_{w_g}^2)^{-d_g/2} \exp\left(-\frac{\sum_{i_g=1}^{d_g} w_{i_g}^2}{2\sigma_{w_g}^2}\right) \quad (3)$$

is used, which is the product of four different normal distributions, corresponding to four different weight groups (i.e.  $\mathbf{w} = \{\mathbf{w}_1, \dots, \mathbf{w}_4\}$ ): the input-hidden layer weights, the hidden layer biases, the hidden-output layer weights and the output layer biases. The parameters  $\sigma_{w_g}^2$  and  $d_g$  are the variance and dimension of the  $g$ th weight group, respectively. Both  $\sigma_y^2$  and  $\sigma_w^2 = \{\sigma_{w_1}^2, \dots, \sigma_{w_4}^2\}$  in (2) and (3), respectively, are treated as unknown hyperparameters with rather noninformative inverse chi-squared hyperprior distributions, which allows their values to be determined from the data. The MCMC approach then involves a two-step iterative procedure, where the hyperparameters are first sampled using the Gibbs sampler and then the weight vector is sampled using the adaptive Metropolis algorithm developed by Haario et al. (2001). Given sufficient *burn-in* iterations, the sampled sequences should converge to a stationary distribution. From this point onwards, it can be considered that the sampled parameters are generated from the posterior distribution.

Samples from the posterior weight distribution can be used directly to calculate an output distribution for each given input pattern. To convert these outputs to samples from the predictive distribution of the targets, Gaussian noise with zero mean and estimated variance  $\sigma_y^2$  is then added. As such, the estimated predictive distribution accounts for both aspects of prediction uncertainty described in Section 1;  $p(\mathbf{w}/\mathbf{y}, H)$  describes the accuracy of the estimated model with respect to the mean of the target distribution, while  $\sigma_y^2$  accounts for the variance with respect to the actual target data including noise.

For further details of the Bayesian training method, see Kingston et al. (2005).

### 3 WAVE OVERTOPPING MODEL

In this study, apart from the method used for ANN training and estimation of prediction uncertainty, the intention was to follow the CLASH model development procedure described in Van Gent et al. (2007) as closely as possible. Therefore, any preprocessing of the CLASH database was repeated in this study and the model inputs, outputs and ANN structure used by Van Gent et al. (2007) were replicated. These are described in more detail as follows.

#### 3.1 CLASH database

The complete CLASH database (available from [http://clash.ugent.be/results/Database\\_20050101.xls](http://clash.ugent.be/results/Database_20050101.xls)) contains over 10,000 wave overtopping test results, represented by parameters that describe the hydraulic and structural characteristics of the tests. It also contains information about the reliability of the tests and the complexity of the test structures, represented by a *Reliability Factor (RF)* and *Complexity Factor (CF)*, respectively. The *RF* ranges from 1 (very reliable) to 4 (not reliable), while the *CF* also ranges from 1 (very simple) to 4 (very complex). For development of the ANN model, these factors were combined into a single *Weight Factor (WF)*, which was then used during training to ensure that the simpler, more reliable tests were given more importance during training (see Section 3.2). The *WF* was calculated according to  $WF = (4 - RF) \times (4 - CF)$ ; however, tests for which  $q < 10^{-6} \text{ m}^3/\text{s/m}$  were assigned a  $WF = 1$ , as these tests were considered as less accurate than larger overtopping discharges. The inputs and outputs used in developing the Bayesian ANN model for wave overtopping are given in Table 1.

Preprocessing of the database involved removing those tests for which the overtopping discharge,  $q$ , was equal to 0  $\text{m}^3/\text{s/m}$ , tests for which  $WF = 0$  and any other tests showing clear inconsistencies. This resulted in a total of 8016 tests for developing the ANN model. The input and output parameters were

Table 1. Inputs and outputs used in the Bayesian ANN for overtopping.

Parameter	Description
$H_{m0}^*$	Significant wave height at the toe of the structure (m)
$T_{m-1,0}$	Mean wave period at the toe of the structure (s)
$\beta$	Direction of wave attack w.r.t. the normal of the structure ( $^\circ$ )
$h$	Water depth in front of the structure (m)
$h_t$	Water depth at the toe of the structure (m)
$B_t$	Width of the toe of the structure (m)
$\gamma_t$	Roughness/permeability of the structure (-)
$\cot \alpha_d$	Slope of the structure downward of the berm
$\cot \alpha_u$	Slope of the structure upward of the berm (-)
$B$	Width of the berm (m)
$h_b$	Water depth at the berm (m)
$\tan \alpha_b$	Slope of the berm (-)
$R_c$	Crest freeboard of the structure (m)
$A_c$	Armour crest freeboard of the structure (m)
$G_c$	Crest width of the structure (m)
$WF^{**}$	Weight Factor (-)
$q^{***}$	Overtopping discharge ( $\text{m}^3/\text{s/m}$ )

\*— $H_{m0}$  was not used as an input to the ANN, but was used to scale the input/output data.

\*\*— $WF$  is not an input to the ANN model, but is used in calculating the objective function.

\*\*\*  $q$  is the model output.

then scaled to  $H_{m0} = 1 \text{ m}$  using Froude's similarity law in order to extrapolate information from small-scale test results to prototype conditions, ensuring that the model developed would be applicable to both small- and large-scale overtopping tests. For further information about this scaling see Van Gent et al. (2005). Finally, the logarithm of the scaled overtopping volumes,  $q'$ , was taken and this became the target data for fitting the ANN model.

#### 3.2 Bayesian ANN development

The type of ANN used to model the wave overtopping volumes was a feedforward multilayer perceptron with 14 nodes in the input layer, corresponding to the 14 input variables given in Table 1, and 1 output node in the output layer, corresponding to  $\log q'$ . Van Gent et al. (2007) found that a network containing 15 hidden nodes was optimal for modelling this problem; thus an ANN with 15 hidden nodes was also used in this study. The hyperbolic tangent transfer function was used at the hidden layer nodes, while a linear transfer was applied at the output.

Prior to training the model, the available data were divided into two subsets: a training set containing 80% of the data (6104 data patterns) and an

independent validation set containing the remaining 20% (1912 data patterns). As mentioned in the previous section, the  $WF$  values associated with overtopping tests were used to give more importance to tests with greater reliability and simplicity during training. In the CLASH project, this was done by applying the weight factor to the total cost function calculated during training, as follows:

$$E = \sum_{i=1}^N WF_i [y_i - f(\mathbf{x}_i, \mathbf{w})]^2 \quad (4)$$

Therefore, the likelihood function given by (2) was also modified to include  $WF$ , as follows:

$$p(\mathbf{y} | \mathbf{w}, \sigma_y^2) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \prod_{i=1}^N \exp \left\{ -\frac{WF_i [y_i - f(\mathbf{x}_i, \mathbf{w})]^2}{2\sigma_y^2} \right\} \quad (5)$$

The MCMC algorithm described in Section 2.2 was run for 600,000 iterations, where the first 300,000 were discarded as burn-in iterations. Traces of the log (unnormalised) posterior, log likelihood and log prior values were inspected to ensure that approximate convergence was achieved within the burn-in period. Predictive distributions, from which mean predictions and 95% prediction limits were evaluated, were calculated based on 10,000 weight vectors, randomly sampled after approximate convergence had been achieved.

In order to properly examine the advantages of a Bayesian training procedure over bootstrapping for prediction uncertainty assessment, an ANN model was also developed using the bootstrapping procedure described in (Van Gent et al. 2005; Van Gent et al. 2007).

#### 4 RESULTS

Figure 1 displays the mean predictions and 95% prediction limits obtained using the Bayesian ANN, plotted against the measured overtopping volumes for the training and validation data sets. It can be seen in this figure that the mean overtopping predictions are reasonably accurate; however, the prediction limits indicate that these predictions have a reasonably high degree of associated uncertainty, as the ANN predictions are a factor 10 (or more) larger/smaller than these mean predictions (and the corresponding measured data). Nevertheless, for both the training data and the validation data, the 95% prediction limits account for greater than 98% of the measured overtopping volumes. In the CLASH ANN publications (Van Gent et al. 2005; Van Gent et al. 2007), the percentage of measured data accounted for by the prediction limits

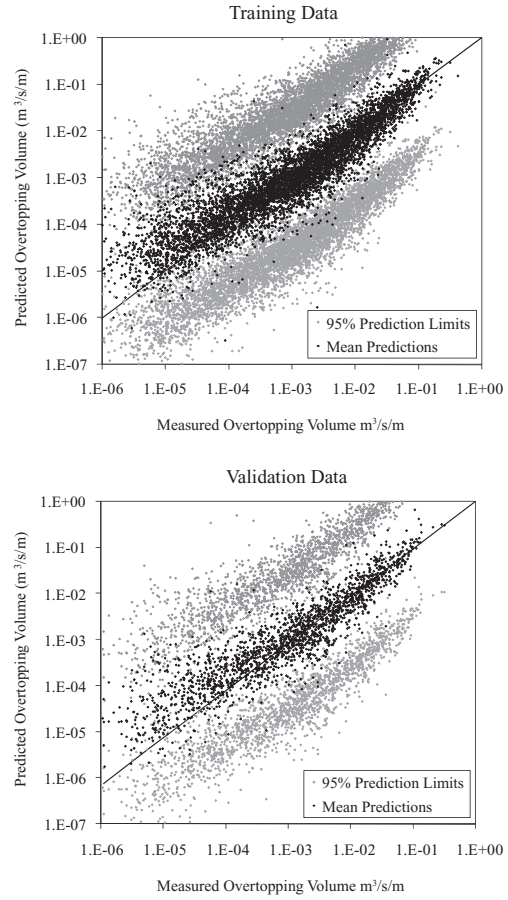


Figure 1. Mean predictions and 95% prediction limits obtained using the Bayesian ANN when applied to the training and validation data.

obtained using the bootstrapping procedure is not reported. However, for the ANN model developed using bootstrapping in this study, these limits only account for around 50% of the measured training and validation data.

Although the overtopping volumes modelled are not a time series, it is useful to compare the results obtained using Bayesian and bootstrapping methods as shown in Figure 2. In this figure, 100 of the measured validation overtopping volumes with  $WF = 9$  are plotted with the corresponding mean predictions and 95% prediction limits obtained using the Bayesian ANN. Also included in this figure are the 95% prediction limits obtained using the bootstrapping procedure. As can be easily seen in this figure, the prediction limits obtained using the Bayesian approach, which incorporate data noise, are significantly wider than the limits obtained using bootstrapping. Therefore, it



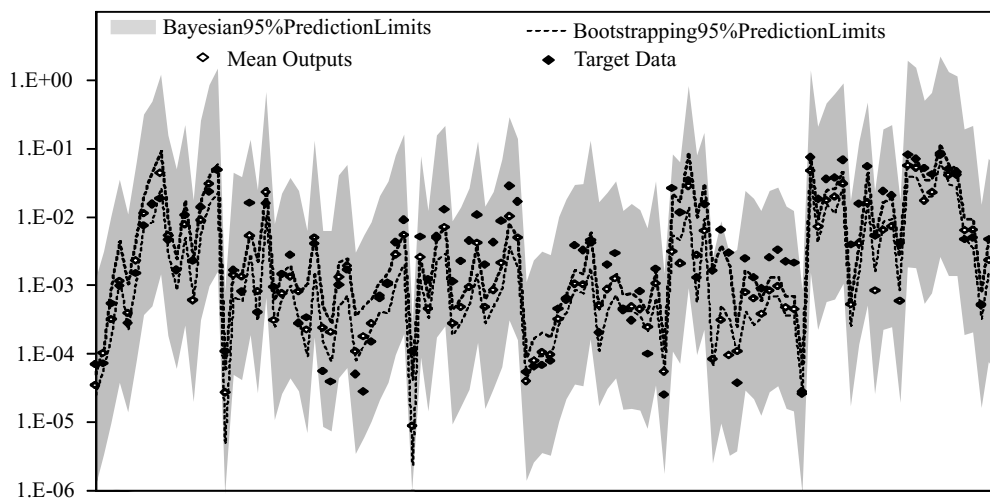


Figure 2. Comparison of 95% prediction limits obtained using the Bayesian ANN and using the bootstrapping procedure when applied to validation overtopping volumes with  $WF = 9$ .

can be concluded that noise variance is a significant component of prediction uncertainty and by not accounting for this aspect, the uncertainty associated with the bootstrapping ANN predictions is significantly underestimated.

In addition, the training time required by the Bayesian approach was 785 minutes, whereas the time taken for bootstrap training with 500 bootstraps was 7389 minutes; almost 10 times the computation time.

## 5 CONCLUSIONS

It has been demonstrated in this study that the mean predictions and 95% prediction limits obtained using a Bayesian ANNs are much better able to represent measured wave overtopping data than the prediction limits obtained using bootstrapping methods. Whilst bootstrapping methods account for *model uncertainty variance* due to limitations of the ANN model, they neglect the second component of total prediction variance; *data noise variance*. On the other hand Bayesian methods account for both of these sources of prediction variance and as a result, the prediction limits obtained provide much better coverage of the target data than those obtained using bootstrapping.

Noise is inherent to all real data and thus data noise variance is always likely to be a significant component of ANN prediction uncertainty. Therefore, in order to correctly indicate the reliability associated with ANN predictions of wave overtopping volumes, it is recommended that a Bayesian approach be considered. If bootstrapping methods are relied upon to

estimate prediction accuracy, it is likely that too much confidence could be placed in the predictions, which could have severe consequences if the predictions turn out to be incorrect.

## ACKNOWLEDGEMENTS

This study made use of much of the initial ANN development and data preprocessing approach devised by Van Gent et al. (2005; 2007). The authors would also like to acknowledge William Allsop for his input on wave overtopping and the CLASH project.

## REFERENCES

- De Rouck, J. & Geeraerts, J. 2005. CLASH Final Report. Belgium, Ghent University.
- Haario, H., Saksman, E. & Tamminen, J. 2001. An adaptive Metropolis algorithm. *Bernoulli* 7(2): 223–242.
- Heskes, T. 1997. Practical confidence and prediction intervals. *Advances in Neural Information Processing Systems*. M. Mozer, M. I. Jordan and T. Petsche. Denver, CO, USA, MIT Press. 9 (NIPS\*1996): 176–182.
- Kingston, G.B., Lambert, M.F. & Maier, H.R. 2005. Bayesian training of artificial neural networks used for water resources modeling. *Water Resources Research* 41(12): W12409.
- Van Gent, M.R.A., Pozueta, B., Van den Boogaard, H.F.P. & Medina, J.R. 2005. Prediction method; Neural network modelling of wave overtopping at coastal structures.
- 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(8): 586–593.



## Fluid thinking...smart solutions

HR Wallingford provides world-leading analysis, advice and support in engineering and environmental hydraulics, and in the management of water and the water environment. Created as the Hydraulics Research Station of the UK Government in 1947, the Company became a private entity in 1982, and has since operated as a independent, non profit distributing firm committed to building knowledge and solving problems, expertly and appropriately.

Today, HR Wallingford has a 50 year track record of achievement in applied research and consultancy, and a unique mix of know-how, assets and facilities, including state of the art physical modelling laboratories, a full range of computational modelling tools, and above all, expert staff with world-renowned skills and experience.

The Company has a pedigree of excellence and a tradition of innovation, which it sustains by re-investing profits from operations into programmes of strategic research and development designed to keep it – and its clients and partners – at the leading edge.

Headquartered in the UK, HR Wallingford reaches clients and partners globally through a network of offices, agents and alliances around the world.



**HR Wallingford Ltd**  
Howbery Park  
Wallingford  
Oxfordshire OX10 8BA  
UK

tel +44 (0)1491 835381  
fax +44 (0)1491 832233  
email [info@hrwallingford.co.uk](mailto:info@hrwallingford.co.uk)

[www.hrwallingford.co.uk](http://www.hrwallingford.co.uk)

