

Predicting Wave Run-up on Rubble-Mound Structures using M5 Model

Tree

Lisham Bonakdar and Amir Etemad-Shahidi*

School of Civil Engineering, Iran University of Science and Technology, Narmak,

Tehran, Iran, P.O. Box 16765-163, Fax: +9821 77240398.

* Corresponding Author, E-mail: etemad@iust.ac.ir

Abstract:

Prediction of run-up level is a key task in designing crest height of the coastal structures. For the design of the crest level of coastal structures, the wave run-up level with a 2% exceedance probability, $R_{u2\%}$, is most commonly used. In this study, the performance of M5 model tree for prediction of the wave run-up on rubble-mound structures was investigated. The main advantage of model trees unlike the other soft computing tools, is their easier use and more importantly their understandable mathematical rules. Experimental data set of Van der Meer and Stam was used for developing model trees. The conventional governing parameters were selected as the input variables and the obtained results were compared with Van der Meer and Stam's formula, recommended by the Coastal Engineering Manual. The predictive accuracy of the model tree approach was found to be superior to that of Van der Meer and Stam's empirical formula. Furthermore, to judge the generalization capability of the model tree method, the model developed based on laboratory data set was validated with the prototype run-up measurements on the Zeebrugge breakwater, Belgium. Results show that the model tree is more accurate than empirical formulas and TS Fuzzy approach in estimating the full-scale run-up.

Keywords: Wave run-up; M5' model tree; Rubble-mound structures; Surf similarity; Permeability; Empirical formula; Soft computing methods.

1. Introduction

Design of coastal structures such as rubble-mound breakwaters, dikes and revetments is a very important issue in coastal and ocean engineering. Breakwaters and dikes are generally used to protect harbors and coasts against wave attack. Design of coastal structures includes both hydraulic and structural aspects. Wave run-up is one of the main criteria in the design of the crest level of rock armored slopes. The wave run-up level (R_u) is described as the maximum water-surface distance measured vertically from the still-water level. Crest levels are generally designed according to the run-up level exceeded by 2% of the incident waves, $R_{u2\%}$, on the front face of coastal structure.

A large variety of researches have been conducted on wave run-up and several empirical formulae have been developed based on laboratory experiments. Losada and Gimenez-Curto (1981) developed an exponential model for the analysis of wave run-up on rough, permeable slopes under regular waves and proposed the following formula:

$$R_u = A(1 - \exp(-B\xi)) \quad (1)$$

where R_u is wave run-up height and ξ is the surf similarity parameter based on regular wave defined as:

$$\xi = \frac{\tan \alpha}{\sqrt{H/L_o}} \quad (2)$$

where α is the structure slope angle, H is the incident wave height and L_o is the deepwater wavelength.

Losada and Gimenez-Curto (1981) fit their equation to test results obtained by various researchers for different armor types and determined A and B values for various armor types. They showed that the wave run-up on rip rap slopes is higher than slopes covered with artificial armor units. Allsop et al. (1985), based on the equation of Losada and Giménez-Curto (1981), proposed the following exponential models for determining relative wave run-up ($R_{u2\%}/H_s$) on the Tetrapod and Antifer cube armored slopes for irregular waves:

For the Tetrapod armored slope:

$$\frac{R_{u2\%}}{H_s} = 1.94(1 - \exp(-0.3\xi_{op})) \quad (3)$$

And for the Antifer cube armored slope:

$$\frac{R_{u2\%}}{H_s} = 1.68(1 - \exp(-0.35\xi_{op})) \quad (4)$$

Where $R_{u2\%}$ is wave run-up value exceeded by 2% of the wave run-up events H_s is significant wave height and ξ_{op} is surf similarity parameter based on peak wave period (T_p).

Van der Meer and Stam (1992) evaluated effects of various structural and hydraulic parameters on wave run-up using a large number of laboratory tests. They analyzed the effects of the permeability, the slope angle, the spectral shape, significant wave height and mean wave period separately. They showed that the most significant factors influencing run-up phenomenon on rock armored slopes are the permeability of the structure and the surf similarity parameter. Van der Meer and Stam (1992) suggested

the following formulas for wave run-up on rock armored structures. For impermeable ($P=0.1$) armored structures they suggested:

$$\frac{R_{u2\%}}{H_s} = 0.96\xi_m \quad \text{for } 1.0 < \xi_m \leq 1.5 \quad (5)$$

$$\frac{R_{u2\%}}{H_s} = 1.17(\xi_m)^{0.46} \quad \text{for } \xi_m > 1.5 \quad (6)$$

And for permeable ($P=0.5$) and homogenous ($P=0.6$) rock armored structures they suggested:

$$\frac{R_{u2\%}}{H_s} = 0.96\xi_m \quad \text{for } 1.0 < \xi_m \leq 1.5 \quad (7)$$

$$\frac{R_{u2\%}}{H_s} = 1.17(\xi_m)^{0.46} \quad \text{for } 1.5 < \xi_m \leq 3.1 \quad (8)$$

$$\frac{R_{u2\%}}{H_s} = 1.97 \quad \text{for } 3.1 < \xi_m \leq 7.5 \quad (9)$$

where P is permeability of structure and ξ_m is surf similarity parameter based on mean wave period defined as:

$$\xi_m = \frac{\tan \alpha}{\sqrt{H_s/L_m}} \quad (10)$$

where L_m is the deepwater wave length based on the mean irregular wave period, T_m .

Equation 9 shows that the permeability of the structure is only significant for high surf similarity values. These equations are valid for relatively deep water in front of the structure and for a Rayleigh distributed wave height distribution (Van de Walle, 2003).

The Van der Meer and Stam (1992) formula has been developed based on laboratory tests mostly with a standard Pierson–Moskowitz spectrum. Hence, they did not consider the

effects of spectral shape and water depth in their formula. There still exists some scattering between measured wave run-up and predicted ones. The uncertainties in empirical formulas inevitably increase the factor of safety and the construction cost (Kim and Park, 2005). Therefore, recently, two studies have been carried out to develop more accurate models for wave run-up. Erdik and Savci (2008) proposed a new run-up model using TAKAGI-SUGENO Fuzzy approach. Erdik et al. (2009) improved the accuracy of wave run-up on rubble-mound prediction using Artificial Neural Network (ANN) method. Both of them developed their models using experimental data set of Van der Meer and Stam (1992).

In this study, a method called the M5' algorithm (Wang and Witten, 1997) is used to predict relative wave run-up. M5' model tree is a popular soft computing method first introduced by Quinlan (1992). The main advantage of the model trees is that, they provide rules that are comprehensible and transparent. The trees obtained from M5, called model trees, are binary decision trees that can have linear regression equations at the leaves. Recently, model trees have been employed successfully in modeling water level discharge relationship (Bhattacharya and Solomatine, 2005), sediment transport (Bhattacharya et al., 2007), wind estimating from waves (Daga and Deo, 2009), derivation of wave spectrum (Sakhare and Deo, 2009), prediction of significant wave height (Mahjoobi and Etemad-Shahidi, 2008, Etemad-Shahidi and Mahjoobi, 2009) and design of rubble-mound breakwaters (Etemad-Shahidi and Bonakdar, 2009). However, this method has not been applied in prediction of wave run-up yet. The model is trained and tested using 162 data of Van der Meer and Stam (1992). To verify the capability of the model tree, the predicted results are compared with those of Van der Meer and Stam's

(1992) empirical formula. The model is also applied to predict full scale run-up measurements on the Zeebrugge breakwater, Belgium.

This paper is outlined as follows: Model trees and M5' algorithm are described in section 2. Building and evaluating model trees and the results are discussed in section 3. Then, the full-scale application is given in section 4. Finally, the summary of the work and concluding remarks are presented in section 5.

2. Model Trees

The M5 algorithm is the most commonly used classifier of decisions trees family. Structurally, a model tree takes the form of a decision tree with linear regression functions instead of terminal class values at its leaves. The M5 model tree is a numerical prediction algorithm and the nodes of the tree are chosen over the attribute that maximizes the expected error reduction as a function of the standard deviation of output parameter (Zhang and Tsai, 2007). M5 Model trees were discovered and brought by Quinlan (1992) and his theory was expanded in a method called M5' by Wang and Witten (1997). Model trees have several advantages, making them a suitable regression method for performance analysis. The prediction accuracy of model trees is comparable to that of techniques such as ANNs and is known to be higher than the prediction of regression trees such as CART method (Ould-Ahmed-Vall et al. 2007). Both, the derived tree structure and the regression models at the leaves, can be used to further the knowledge of nature and severity of performance problems. Model trees are also known to efficiently handle large data sets with a high number of attributes and high dimensions.

At first, M5 model trees algorithm constructs a regression tree by recursively splitting the instance space. Figure 1 illustrates a tree structure of training procedure corresponding to a given 2-D input parameter domain of x_1 and x_2 . The splitting condition is used to minimize the intra-subset variability in the values down from the root through the branch to the node. The variability is measured by the standard deviation of the values that reach that node from the root through the branch, with calculating the expected reduction in error as a result of testing each attribute at that node. In this way, the attribute that maximizes the expected error reduction is chosen. The splitting process would be done if either the output values of all the instances that reach the node vary slightly or only a few instances remain. The standard deviation reduction (SDR) is calculated as (Quinlan, 1992):

$$SDR = sd(T) - \sum_i \frac{|T_i|}{|T|} \times sd(T_i) \quad (11)$$

where T is the set of examples that reach the node, T_i are the sets that are resulted from splitting the node according to the chosen attribute and sd is the standard deviation (Wang and Witten, 1997). After the tree has been grown, a linear multiple regression model is built for every inner node, using the data associated with that node and all the attributes that participate in tests in the sub-tree rooted at that node. Then linear regression models are simplified by dropping attributes if it results in a lower expected error on future data. After this simplification, every sub-tree is considered for pruning. Pruning occurs if the estimated error for the linear model at the root of a sub-tree is smaller or equal to the expected error for the sub-tree. After pruning, there is a possibility that the pruned tree might have discontinuities between nearby leaves. Therefore, to

compensate discontinuities among adjacent linear models in the leaves of the tree a regularization process is made. This process is started once the tree has been pruned and usually improves the prediction, especially for models based on training sets containing a small number of instances (data points) (Zhang and Tsai, 2007).

3. Model tree modeling and results

In this study, M5' algorithm was used for predicting relative wave run-up ($R_{u2\%} / H_s$). To compare the performance of model trees, two M5' models were developed. Table 1 shows the input parameters used in each model. To have a better comparison of the models, the same parameters used in Van der Meer and Stam's formula (1992) were used in the first model (MT1). The MT1 input parameters includes surf similarity and permeability of the structure. Surf similarity parameter includes the wave height, wave period and slope angle of the structure. Most of phenomena in the surf zone are described by means of the surf similarity parameter. The surf similarity parameter describes the type of wave breaking on a structure or on a beach. Permeability of structure indicates the behavior of water running up on the structure. For a permeable structure the water can penetrate into the structure decreasing the actual run-up, while in an impermeable structure all water run up on slope of the structure and do not sink into the structure (Van der Meer and Stam, 1992). The second model (MT2) was developed additionally to investigate the effect of water depth parameter (h / H_s), where h is the water depth at the toe of the breakwater, on the relative wave run-up. This parameter is included in the model to consider the possible effects of wave breaking on the foreshore.

Also, surf similarity parameter was replaced by the slope angle of the structure and wave steepness in MT2.

Both models were trained and tested using 162 data of Van der Meer and Stam (1992). Ranges of different parameters of training and testing data sets for MT1 and MT2 are presented in Tables 2 and 3, respectively. For each model, the dataset was split randomly to training and testing set. Training set includes 2/3 of the whole data (108 data points) and the rest of data was used as testing set (54 data points). Models were trained using the training set and then evaluated with the testing data. The relationship between the used input and output ($R_{u2\%} / H_s$) parameters is not necessarily linear; whereas, the model tree can only produce linear relationship between them. In order to overcome this limitation, models with $\log(\text{inputs})$ and $\log(R_{u2\%} / H_s)$, were developed to reach a nonlinear model in form of $R_{u2\%} / H_s = aP^b \xi_m^c$, (a , b and c are constants).

The following statistical parameters were used to evaluate the performance of the models for predicting the relative wave run-up: index of agreement (I_a), correlation of coefficient (CC) and root mean squared error ($RMSE$).

$$I_a = 1 - \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum (|x_i - \mu_x| + |y_i - \mu_y|)^2} \quad (12)$$

$$CC = \frac{(1/n)[(x_i - \mu_x)^T (y_i - \mu_y)]}{\sqrt{(1/n)(x_i - \mu_x)^2} \sqrt{(1/n)(y_i - \mu_y)^2}} \quad (13)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_1^n (y_i - x_i)^2} \quad (14)$$

where x_i and y_i denote the predicted and the measured values, respectively; n is the number of measurements, and T denotes the transpose of a matrix. μ_x and μ_y are the corresponding mean values of the predicted and measured parameters.

Scatter diagrams of measured and predicted relative wave run-up by MT1 are shown in Figs 3 (all data) and 4 (testing data). As seen in these figures, the data points are more concentrated on optimal line compared to figure 2 (Van der Meer and Stam's). This shows that the MT1 model is more accurate than Van der Meer and Stam's (1992) formula using the input parameters. A comparison of statistical parameters also indicates that MT1 is more precise than the empirical formula of Van der Meer and Stam (1992) in the prediction of the relative wave run-up (Table 4). Agreement index (for all data set) is 0.96 for MT1 and the root mean squared error of MT1 is 0.167.

Measured and predicted relative wave run-ups by MT2 are shown in Fig. 5 (all data set) and Fig. 6 (testing data). The MT2 results show less scatter in the data points than both of MT1 and Van der Meer and Stam's. The agreement index of MT2 for all data set has increased to 0.972 from 0.902 (by Van der Meer and Stam's empirical model) and its root mean squared error ($RMSE=0.141$) is less than those of MT1 and Van der Meer and Stam (1992). A drawback of Van der Meer and Stam's formula (figure 2) is that their prediction for a vast range of relative wave run-up (approximately 1.5 to 2.5) is constant (1.97). This has been resulted in more scattering between the measured and predicted values, while this is not seen in M5 model tree models (Figs 3 to 6). It is also interesting to note that both M5 model tree and Van der Meer and Stam's models slightly underestimate relative wave run-up larger than 2.5 that could be due to lack of enough data points in this range

In addition to higher accuracy level of the models, the other advantage of the developed M5 models is their ability to generate simple and meaningful rules that can be easily utilized in run-up calculations. The developed tree and rules for MT1 are shown in Fig. 7. The topmost splitting attribute is surf similarity, ξ_m and the next one is the permeability. It is noteworthy that $P \leq 0.22$ represents impermeable structures and $P \geq 0.22$ represents permeable structures. As discussed by Bhattacharya et al. (2007), the splitting values do not necessarily have any physical interpretation and are obtained by minimizing the prediction error. However, similar to Van der Meer and Stam's findings, figure 7 shows that the permeability of structure would be important when surf similarity parameter becomes large.

Considering $P=0.1$ for impermeable structures and $P=0.5$ for permeable structures, the given rules for MT1 are:

For impermeable structures:

$$\frac{R_{u2\%}}{H_s} = 1.0\xi_m^{0.69} \quad \xi_m \leq 2.1 \quad (15)$$

$$\frac{R_{u2\%}}{H_s} = 1.3\xi_m^{0.31} \quad 2.1 < \xi_m \leq 3.9 \quad (16)$$

$$\frac{R_{u2\%}}{H_s} = 1.5\xi_m^{0.29} \quad \xi_m > 3.9 \quad (17)$$

For permeable structures:

$$\frac{R_{u2\%}}{H_s} = 0.86\xi_m^{0.69} \quad \xi_m \leq 2.1 \quad (18)$$

$$\frac{R_{u2\%}}{H_s} = 1.16\xi_m^{0.31} \quad 2.1 < \xi_m \leq 3.9 \quad (19)$$

$$\frac{R_{u2\%}}{H_s} = 1.56\xi_m^{0.15} \quad \xi_m > 3.9 \quad (20)$$

These equations are in good agreement with engineering sense and Van der Meer and Stam's (1992) empirical formulas. For example, when the surf similarity increases, its power decreases and it becomes insignificant. Furthermore, when ξ_m is less than 3.9, the relative wave run-ups for permeable and impermeable structures are nearly the same, but when surf similarity has a large value ($\xi_m > 3.9$) the permeability becomes important. Equation 20 shows that for a permeable structure, the power of surf similarity decreases significantly for surging waves. This means that the relative wave run-up for a permeable structure with a large value of surf similarity is not so sensitive to surf similarity parameter. By contrast, for an impermeable structure the relative wave run-up is higher (equation 17) and still depends on the surf similarity parameter. This is due to the fact that the water running up on the slope of the structure can not sink into the structure.

The developed formulas have a better physical interpretation than those of previous approaches. For instance, in the range of $\xi_m < 3.1$, the Van der Meer and Stam's formulas yield the same value of relative wave run-up for permeable and impermeable structures, while the developed formulas consider the permeability of structures on the wave run-up prediction. In addition, the Van der Meer and Stam's formula is independent of the surf similarity parameter for permeable structures with $\xi_m > 3.1$, while this is not the case in the developed formulas. Complexity of the rules is a function of number of the input parameters and physics of the problem. MT2 is more complex than MT1, because it has more input parameters and rules. By adding dimensionless water depth at the toe of

the structure, the number of rules increased to seven in MT2. In this aspect, MT1 is superior to MT2 because their accuracies are nearly the same but MT1 represents a fewer number of rules that are as understandable as empirical formulas.

The major drawback of model trees is that they can only generate linear formulas at their leaves. Therefore, to achieve a more accurate and simpler model, transformations of input parameters may be required. Commonly used soft computing tools such as ANN and FL need some trial and error for model optimization. For example, a disadvantage of ANN models is that the number of hidden layers and the number of nodes per hidden layer should be found using a trial-and-error approach, while in FL models the selection of membership functions parameters and fuzzy subsets is not trivial (Kazeminezhad et al. 2005). An advantage of model tree over other soft computing approaches is that its rules are understandable and they can be easily applied by another user (Sakhare and Deo, 2009).

4. Application: Run-up on the Zeebrugge breakwater

In this Section, the M5' model tree is applied to predict full scale run-up measurements on the Zeebrugge breakwater, Belgium (Van de Walle, 2003). The prototype example includes a structure with permeable core ($P = 0.5$) and the slope of the breakwater is 1:1.5. The values of surf similarity range from 3.46 to 3.83 (table 5). Thirteen storm events have been selected out of all observed heavy weather conditions measured during the period from 1995 to 2000 along the Belgian coast. It should be noted that the wave climate during these storm events was characterized by an almost constant SWL (De Rouck et al, 2007).

The predicted relative wave run-up by MT1 is compared with the field measurements and those obtained by Van der Meer and Stam's (1992), Kingston and Murphy (1996) formula and Erdik and Savci's (2008) TS Fuzzy model. Kingston and Murphy (1996) identified the parameters of the formula of Losada and Gimenez-Curto (1981) as $A = 1.76$ and $B = -0.28$ for small scale model tests on a model of the Zeebrugge breakwater (De Rouck et al, 2007). In Table 5, M5' model tree results are given together with those of Van der Meer and Stam (1992), Kingston and Murphy (1996) and Erdik and Savci (2008) predictions. As shown in Table 5, the root mean squared error of MT1 ($RMSE = 0.10$) is less than those of Van der Meer and Stam (1992), Kingston and Murphy (1996) and Erdik and Savci's (2008) TS Fuzzy method.

The fact that MT1 predictions are more accurate than other models for the full-scale data set can be also seen from figure 8 which shows the calculated relative wave run-up by Van der Meer and Stam's formula are overestimated ($Bias = \mu_y - \mu_x = 0.245$). In contrast, Kingston and Murphy's formula has remarkably underestimated the relative wave run up ($Bias = -0.6$). This figure also indicate that, in comparison with Van der Meer and Stam's (1992) and Kingston and Murphy (1996), Erdik and Savci's model has a better performance ($Bias = 0.038$). However, the MT1 is the most accurate model among all mentioned approaches ($Bias = 0.006$). This implies that even in real cases the developed model can be used successfully.

It should be mentioned that the developed models, like previous approaches, do not consider the effects of factors such as the direction of wave attack, breaking waves and berm. They can be easily modified if enough data are available. However, all of these

factors result in decreasing of relative wave run-up. Hence it could be said the developed formulas are conservative for the cases that these effects exist

5. Summary and conclusion

In this study, the M5' model tree has been applied to predict the relative wave run-up on rock slopes and its ability was compared with that of Van der Meer and Stam's (1992) formula. The model constructed herein is based on the experimental data of Van der Meer and Stam (1992).

To compare the performance of the model trees for run-up prediction, two M5' models were developed. In the first model the same parameters used in Van der Meer and Stam's (1992) formula, were used. The MT1 input parameters includes permeability of the structure and surf similarity parameter. The second model was developed to investigate the effect of relative water depth parameter, (h / H_s) on the relative wave run-up. Statistical parameters showed that the model tree technique yields more accurate results than the Van der Meer and Stam's (1992) formula in estimating the relative wave run-up on rubble-mound structures. In addition to the higher accuracy (CC of 0.925-0.95), the other advantage of the model trees (compared to other data mining approaches such as ANN) is found to be its ability to generate simple and understandable formulas.

In addition, the developed model tree was validated with thirteen prototype run-up measurements on the Zeebrugge breakwater. The predicted relative wave run-up by MT1 was compared with those obtained by Van der Meer and Stam's (1992), Kingston and Murphy (1996) formula and Erdik and Savci's TS Fuzzy method. It was found that the

MT1 is more precise than the previous approaches in the prediction of the full-scale run-up.

Acknowledgements.

The authors would like to thank J. Van der Meer for providing his original run-up data and helpful suggestions. The authors also express thanks to B. Van de Walle for kindly providing his publications. In addition, the authors are grateful to Saman Azadi, Maral Mohammadi and Liyona Bonakdar for their help in improving the manuscript.

References

- Bhattacharya, B., Solomatine, D.P., 2005. Neural networks and M5 model trees in modeling water level–discharge relationship. *Neurocomputing* 63, 381–396.
- Bhattacharya, B., Price, R. K., Solomatine, D. P., 2007. Machine Learning Approach to Modeling Sediment Transport. *Journal of Hydraulic Engineering* 133(4), 440-450.
- CEM. Coastal Engineering Manual, 2006. Coastal Engineering Research Center, Department of Army Corps of Engineers, US, Chapter VI, Part 5.
- Daga, M., Deo, M.C., 2009. Alternative data-driven methods to estimate wind from waves by inverse modeling, *Nat Hazards* 49, 293-310.
- De Rouck, J., Van de Walle, B., Troch, P., Van der Meer, J., Van Damme, L., Medina, J.R., Willems, M. and Frigaard, P., 2007. Wave Run-Up on the Zeebrugge Rubble Mound Breakwater: Full-Scale Measurement Results. *Journal of Coastal Research* 23(3), 577–583.
- Erdik, T., Savci, M.E., 2008. TAKAGI-SUGENO Fuzzy Approach in Rock Armored Slopes for 2% Wave Runup Estimation. *Coastal Engineering Journal* 50(2), 161-177.
- Erdik, T., Savci, M.E., Sen, Z., 2009. Artificial neural networks for predicting maximum wave runup on rubble. *Expert Systems with Applications* 36, 6403-6408.
- Etemad-Shahidi, A., Bonakdar, L., 2009. Design of Rubble-Mound Breakwaters using M5' Machine Learning Method. Submitted to *Applied Ocean Research*
- Etemad-Shahidi, A., Mahjoobi, J., 2009. Comparison between M5' Model Tree and Neural Networks for Prediction of Significant Wave Height. Submitted to *Ocean Engineering*.

Kazeminezhad, M.H., Etemad-Shahidi, A., Mousavi, S.J., 2005. Application of fuzzy inference system in the prediction wave parameters. *Ocean Engineering* 32, 1709-1725.

Kim, D.H., Park, W.S., 2005. Neural network for design and reliability analysis of rubble mound breakwaters. *Ocean Engineering*. 32, 1332–1349.

Mahjoobi, J., Etemad-Shahidi, A, 2008. An alternative approach for the prediction of significant wave heights based on classification and regression trees. *Applied Ocean Research*. 30,172-177

Ould-Ahmed-Vall, E., Woodlee, J., Yount, C., Doshi, K.A and Abraham, S. 2007. Using Model Trees for Computer Architecture Performance Analysis of Software Applications, *IEEE*. 1-4244-1081-9.

Quinlan, J. R., 1992. Learning with continuous classes, *Proceedings of AI'92 (Adams and Sterling Eds)*, World Scientific. 343-348.

Sakhare, S., Deo, M.C., 2009. Derivation of wave spectrum using data driven methods, *Marine Structures*, 22, 594-609.

Van der Meer, J. W., Stam, C. J. 1992. Wave runup on smooth and rock slopes of coastal structures. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 118, 534-550, .

Van de Walle, B. 2003. Wave runup on rubble mound breakwaters. Ph.D. thesis, University of Gent, Belgium.

Wang, Y., Witten, I.H., 1997. Induction of model trees for predicting continuous lasses. In *Proceedings of the Poster Papers of the European Conference on Machine Learning*, University of Economics, Faculty of Informatics and Statistics, Prague.

Witten, I., Eibe, F., 2005. *Practical Machine Learning Tools and Techniques*. 2nd ed., Elsevier Inc.525 pp.

Zhang, D., Tsai, J.J.P., 2007. *Advances in Machine Learning Applications in Software Engineering*, Idea Group Inc. ISBN: 1-59140-941-1.

Figure Captions

Figure 1. Example of M5 model tree (Models 1–8 are linear regression models).

Figure 2. Comparison between measured and predicted relative wave run-up ($R_{u2\%} / H_s$) by the empirical formula of Van der Meer and Stam 1992, all data.

Figure 3. Comparison between measured and predicted relative wave run-up ($R_{u2\%} / H_s$) by MT1, all data.

Figure 4. Comparison between measured and predicted relative wave run-up ($R_{u2\%} / H_s$) by MT1, testing data.

Figure 5. Comparison between measured and predicted relative wave run-up ($R_{u2\%} / H_s$) by MT2, all data.

Figure 6. Comparison between measured and predicted relative wave run-up ($R_{u2\%} / H_s$) by MT2, testing data.

Figure 7. MT1 generated by M5' algorithm with logarithm of inputs and output.

Figure 8. Comparison between measured and predicted relative wave run-up ($R_{u2\%} / H_s$) by different approaches, field data.

Table Captions

Table 1. Input parameters of M5' models

Table 2. Range of input–output parameters used for the prediction of relative wave run-up in MT1

Table 3. Range of input–output parameters used for the prediction of relative wave run-up in MT2

Table 4. Statistical measures of different models

Table 5. The prediction results of different approaches

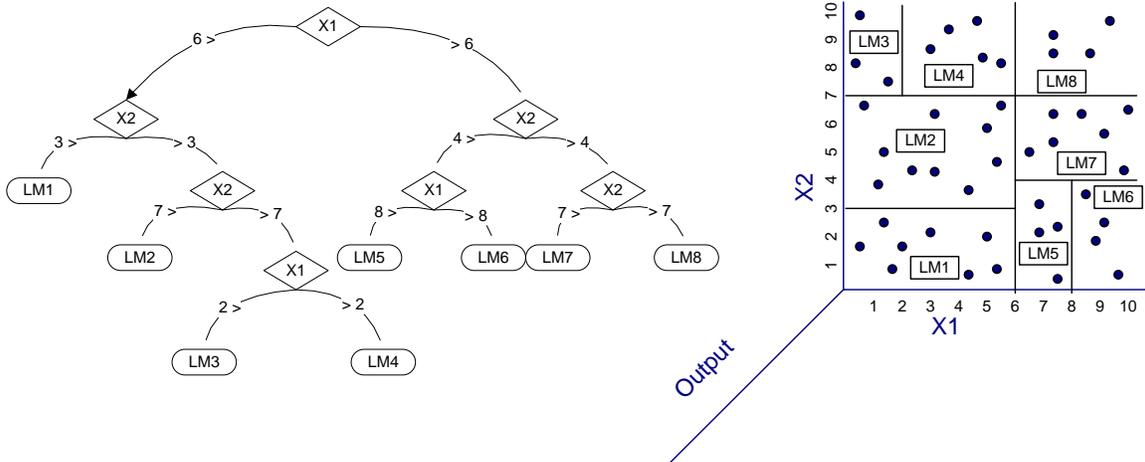


Figure 1. Example of M5 model tree (Models 1–8 are linear regression models).

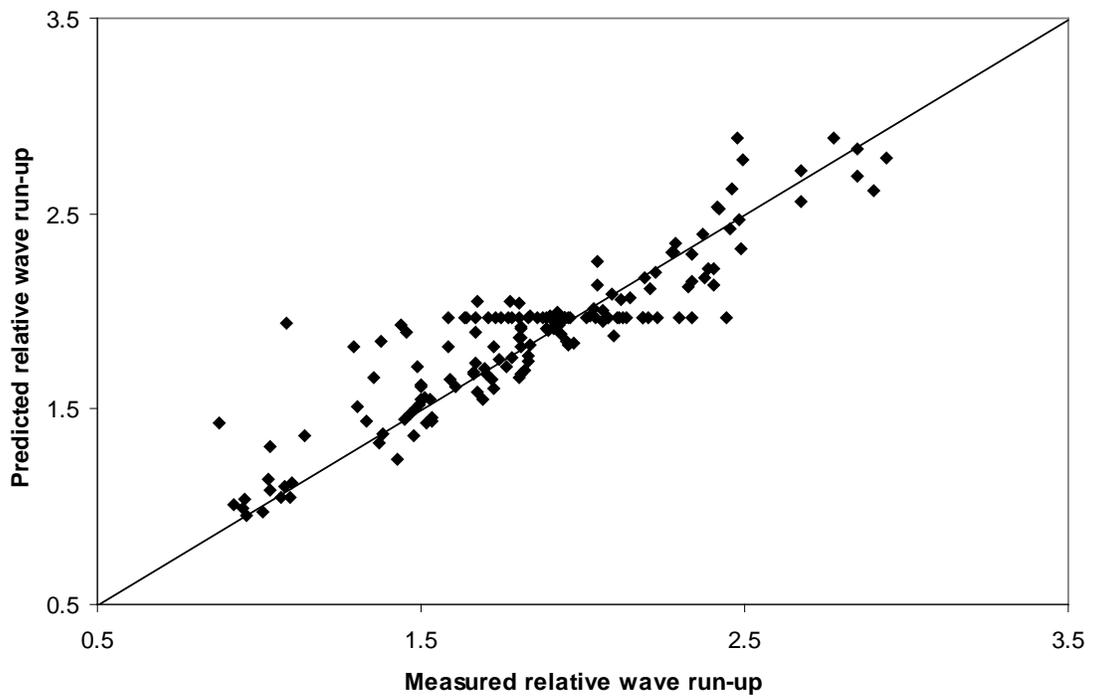


Figure 2. Comparison between measured and predicted relative wave run-up ($R_{t2\%} / H_s$) by the empirical formula of Van der Meer and Stam 1992, all data.

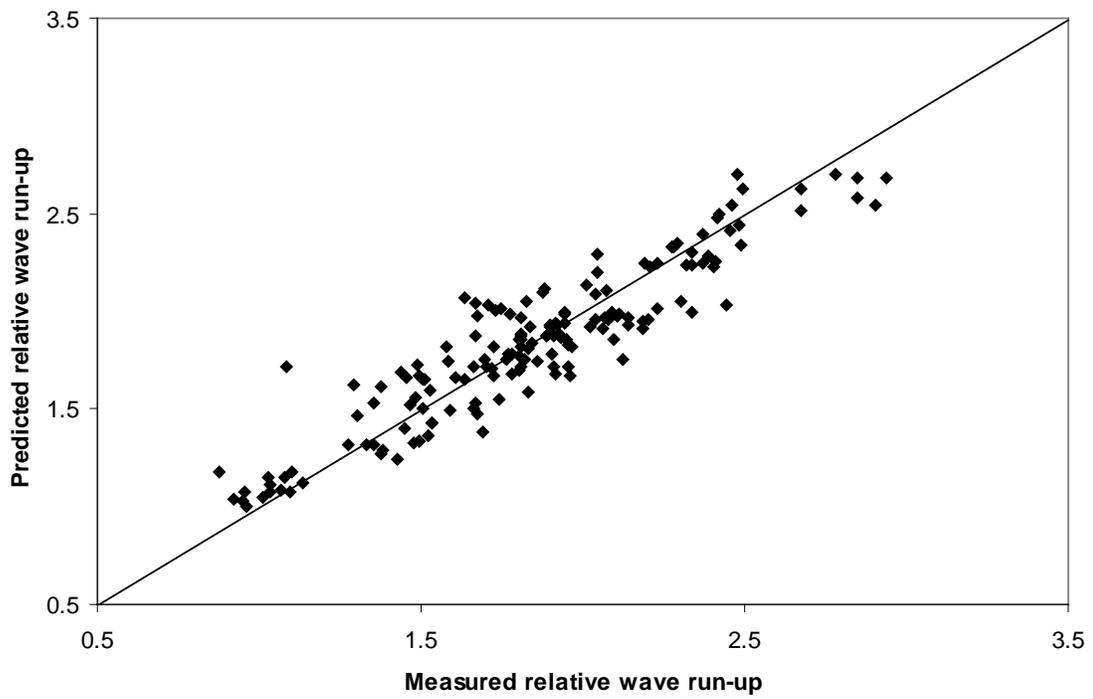


Figure 3. Comparison between measured and predicted relative wave run-up ($R_{t2\%} / H_s$) by MT1, all data.

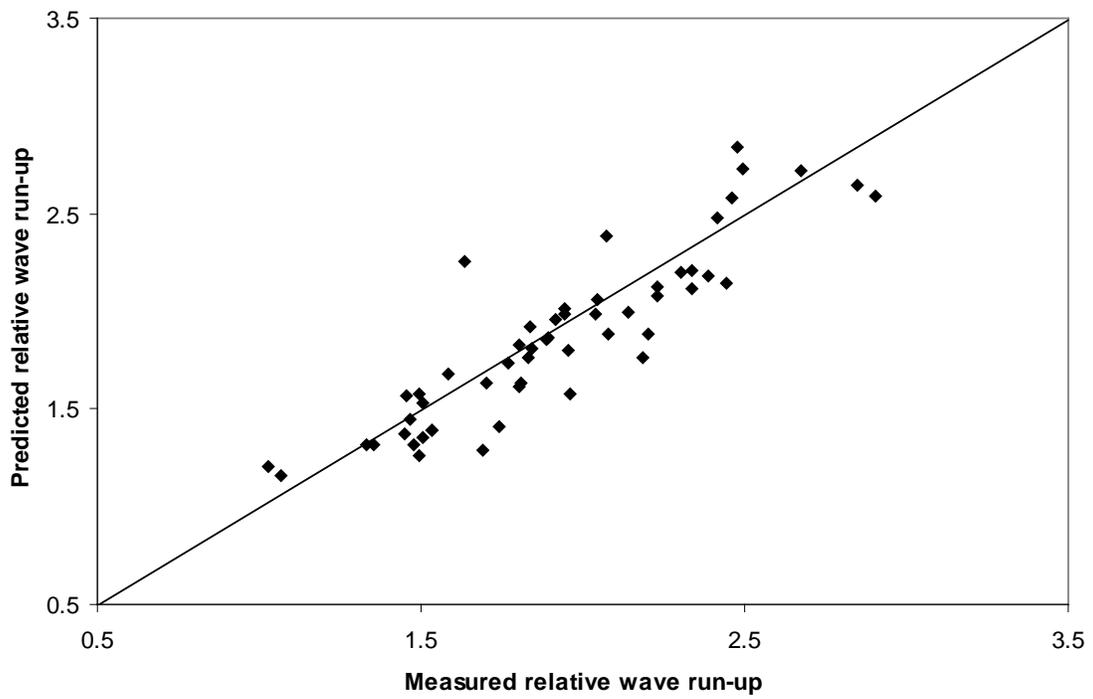


Figure 4. Comparison between measured and predicted relative wave run-up ($R_{u2\%} / H_s$) by MT1, testing data.

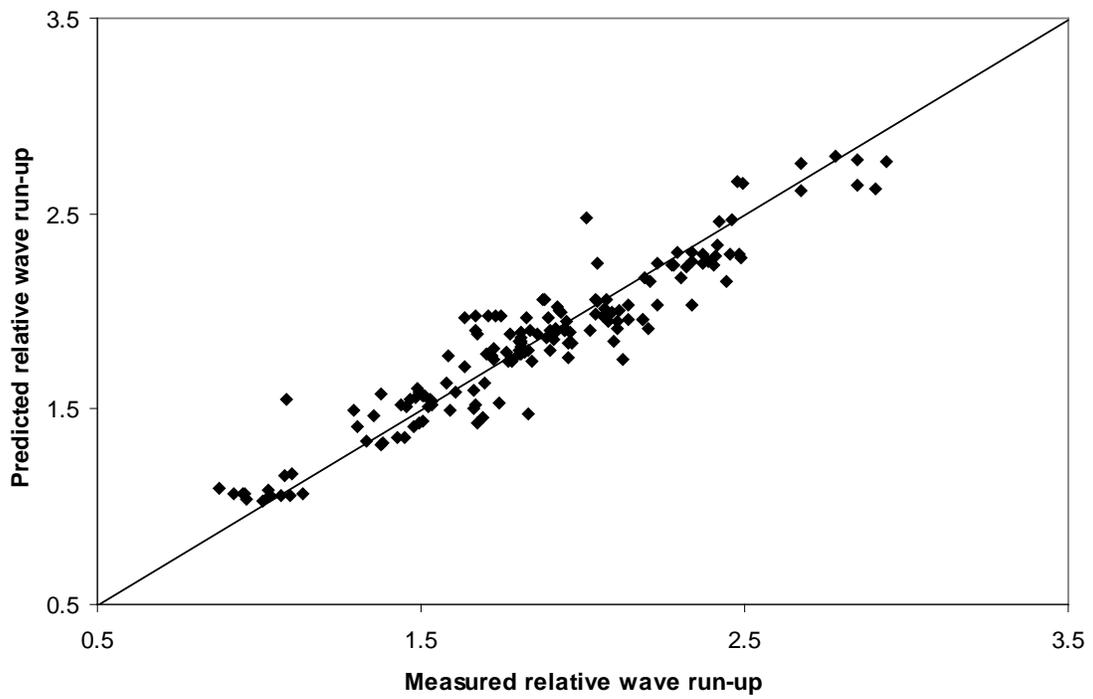


Figure 5. Comparison between measured and predicted relative wave run-up ($R_{t2\%} / H_s$) by MT2, all data.

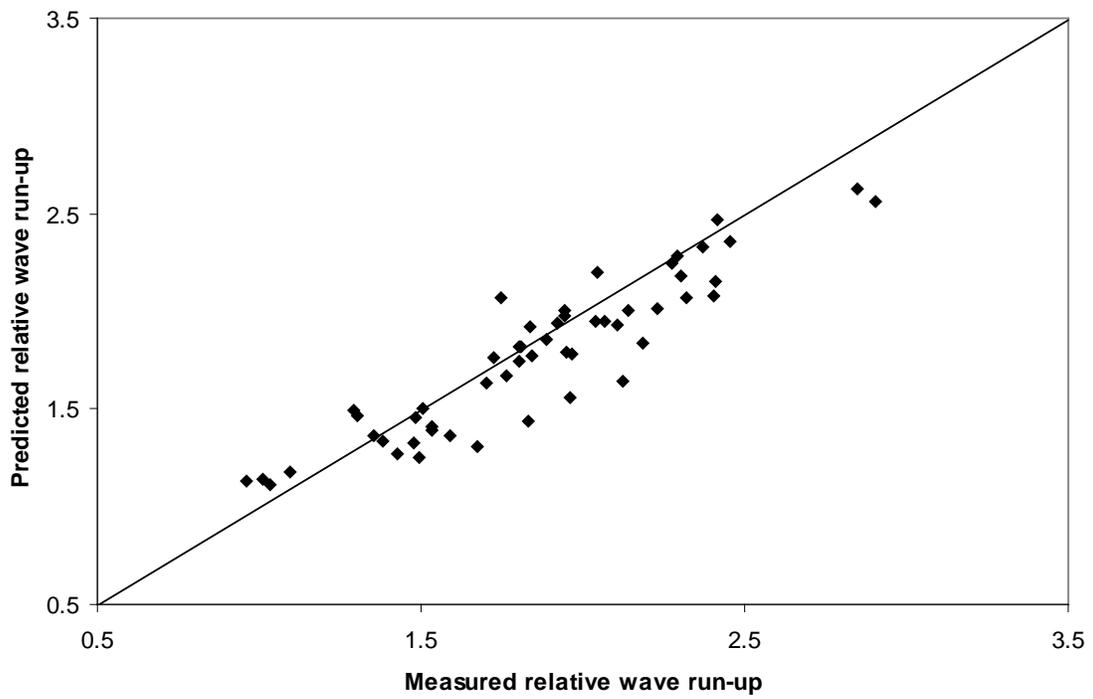


Figure 6. Comparison between measured and predicted relative wave run-up ($R_{u2\%} / H_s$) by MT2, testing data.

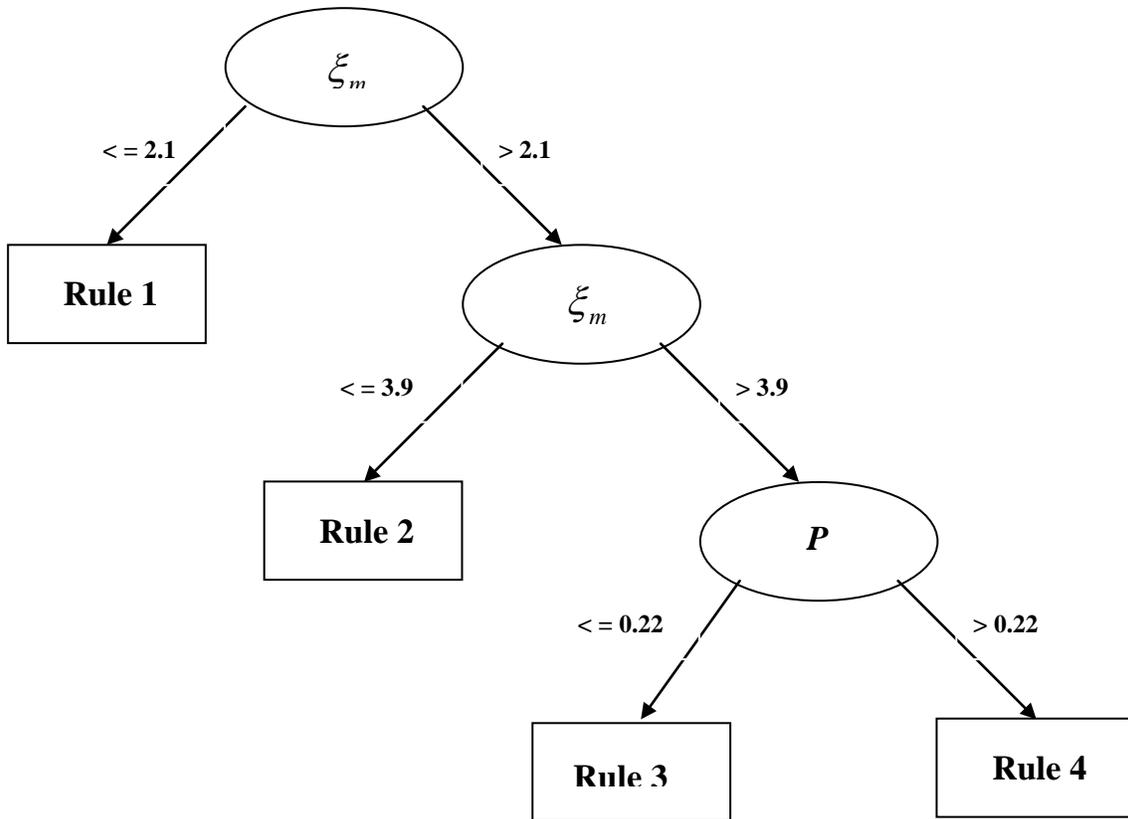


Figure 7. MT1 generated by M5' algorithm with logarithm of inputs and output.

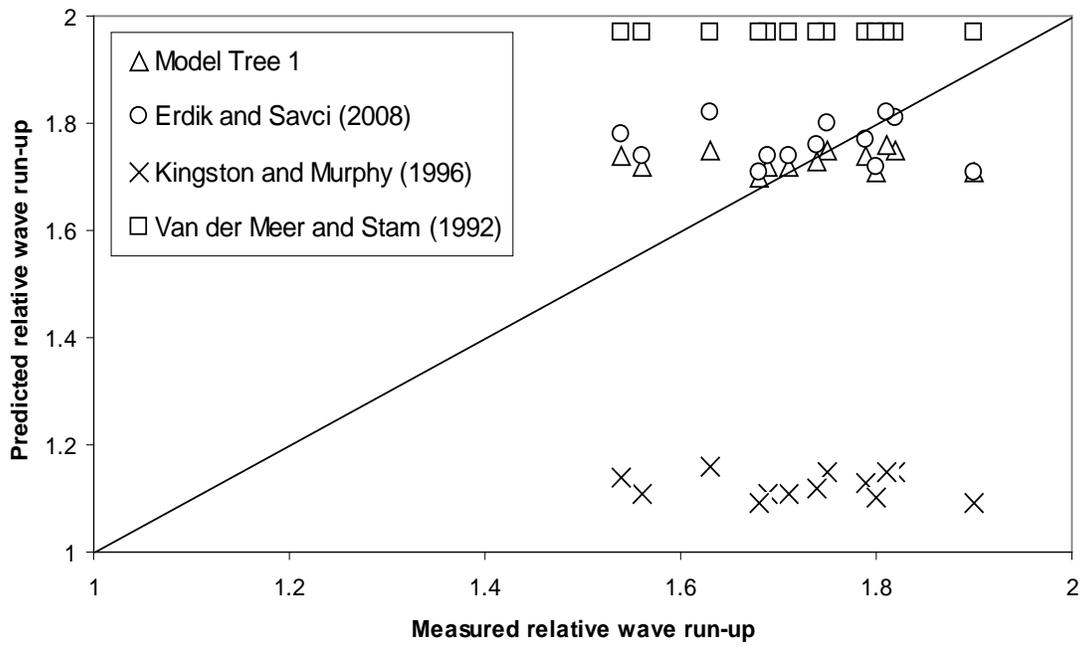


Figure 8. Comparison between measured and predicted relative wave run-up ($R_{u2\%} / H_s$) by different approaches, field data.

Table 1. Input parameters of M5' models

Model	Input parameters
MT1	$\log(P), \log(\xi_m)$
MT2	$\log(P), \log(\cot \alpha), \log(S_{om}), \log\left(\frac{h}{H_s}\right)$

Table 2. Range of input–output parameters used for the prediction of relative wave run-up in MT1

Parameter	Training data	Testing data
P	0.1 ~ 0.5	0.1 ~ 0.5
ξ_m	1 ~ 7.15	1.03 ~ 7.11
$R_{u2\%} / H_s$	0.87 ~ 2.94	1.02 ~ 2.90

Table 3. Range of input–output parameters used for the prediction of relative wave run-up in MT2

Parameter	Training data	Testing data
P	0.1 ~ 0.5	0.1 ~ 0.5
$\cot \alpha$	1.5 ~ 4	1.5 ~ 4
S_{om}	0.0046 ~ 0.0636	0.0052 ~ 0.0589
(h/H_s)	4.10 ~ 17.24	4.05 ~ 14.46
$R_{u2\%}/H_s$	0.87 ~ 2.94	0.96 ~ 2.90

Table 4. Statistical measures of different models

Model	I_a	CC	$RMSE$
Van der Meer and Stam's (1992) (all data set)	0.944	0.902	0.19
MT1 (testing data)	0.943	0.90	0.20
MT1 (all data set)	0.960	0.925	0.167
MT2 (testing data)	0.948	0.92	0.187
MT2 (all data set)	0.972	0.95	0.141

Table 5. The prediction results of different approaches

Measurements			Predictions			
Full-Scale Measurement on Zeebrugge Breakwater			Van der Meer and Stam (1992)	Kingston and Murphy (1996)	TS-Fuzzy by Erdik and Savci (2008)	Model Tree 1
Storm Number	ξ_{om}	$R_{u2\%} / H_{mo}$	$R_{u2\%} / H_{mo}$	$R_{u2\%} / H_{mo}$	$R_{u2\%} / H_{mo}$	$R_{u2\%} / H_{mo}$
1	3.58	1.56	1.97	1.11	1.74	1.72
2	3.70	1.54	1.97	1.14	1.78	1.74
3	3.77	1.75	1.97	1.15	1.80	1.75
4	3.67	1.79	1.97	1.13	1.77	1.74
5	3.57	1.71	1.97	1.11	1.74	1.72
6	3.79	1.82	1.97	1.15	1.81	1.75
7	3.57	1.69	1.97	1.11	1.74	1.72
8	3.47	1.90	1.97	1.09	1.71	1.71
9	3.81	1.81	1.97	1.15	1.82	1.76
10	3.46	1.68	1.97	1.09	1.71	1.70
11	3.51	1.80	1.97	1.10	1.72	1.71
12	3.62	1.74	1.97	1.12	1.76	1.73
13	3.83	1.63	1.97	1.16	1.82	1.75
Root Mean Squared Error (<i>RMSE</i>)			0.26	0.61	0.12	0.10