creative commons
COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

Master's Thesis

Derivation of Wave Overtopping Formulas for Vertical and Inclined Seawalls Using GMDH Algorithm

GMDH 알고리즘을 이용한 수직벽과 경사식 호안에 대한 월파량 공식 유도

February 2018

Graduate School of Civil and Environmental Engineering
Seoul National University

Seok Bong Lee

ABSTRACT

Since wave overtopping is a complex phenomenon and it is sensitive to hydraulic and structural parameters, hydraulic model tests have been used to estimate the wave overtopping rates at coastal structures. Efforts have also been made toward deriving empirical formulas based on a large amount of accumulated data, e.g. the EurOtop formulas based on the CLASH database. To complement the EurOtop formulas, artificial neural network (ANN) models were presented, which is one of the data models. Other data models have also been developed. On the other hand, Goda developed unified formulas for vertical and inclined seawalls with smooth, impermeable surfaces. In this study, new data-based formulas are developed by using the group method of data handling (GMDH) algorithm with the CLASH and new EurOtop datasets. The GMDH formulas are shown to be more accurate than the Goda's formulas and comparable in accuracy to other formulas developed separately for vertical and inclined seawalls. The estimation errors of 95% confidence interval and the range of 95% prediction error are also given. Sensitivity analysis of the derived formulas is also carried out.

Keywords: Wave overtopping rate, GMDH algorithm, CLASH database, Data model, Seawalls

Student Number: 2016-21261

TABLE OF CONTENTS

ABSTRACT i

TABLE OF CONTENTS ii

List of Table iv

List of Figures v

List of Symbol vii

CHAPTER 1. INTRODUCTION 1

1.1 Background 1

1.2 Objectives 3

CHAPTER 2. THEORETICAL BACKGROUDS 5

2.1 Mean wave overtopping rate 5

2.2 Overtopping rate prediction formula 6

2.2.1 EurOtop formula (2007, 2016) 6

2.2.2 Formula by Goda (2009) 11

2.2.3 Formula by Etemad-Shahidi *et al*. (2014, 2016) 12

2.2.4 ANN model (Van Gent *et al*., 2007, Zanuttigh *et al.,* 2016) 13

2.3 Group Method of Data Handling 14

2.3.1 Basic concept of GMDH algorithm 14

2.3.2 Procedure of GMDH algorithm 16

2.3.3 External criteria for self-organizing 17

2.4 Data sampling technique – bootstrap sampling 20

CHAPTER 3. METHODOLOGY 23

3.1 CLASH data for model construction 23

3.1.1 Variables applying Froude similitude law 23

3.1.2 Selected data for GMDH model....26
3.2 Training data extraction using bootstrap sampling....30
CHAPTER 4. RESULTS....34
4.1 Unified formulas for mean overtopping rate derived by GMDH algorithm ..34
4.1.1 Unified formulas by GMDH algorithm....34
4.1.2 Confidence interval of estimation error....35
4.1.3 Performance of derived formulas....38
4.1.4 Comparison of unified GMDH formulas with other overtopping formulas 44
4.1.5 Sensitivity analysis for unified GMDH formulas 51
4.2 Separate formulas for overtopping rate derived by GMDH algorithm 61
4.2.1 GMDH-based overtopping rate prediction formulas for smooth-slope, normal-incidence wave 61
4.2.2 GMDH-based overtopping rate prediction formulas for inclined seawalls considering oblique wave condition and roughness factor....67
4.2.3 GMDH-based overtopping rate prediction formulas for vertical walls considering oblique wave condition 75
CHAPTER 5. CONCLUSIONS....80
REFERENCES....82
APPENDIX A – Results of bootstrap sampling for separate formulas by GMDH..85
APPENDIX B – Distribution of estimation error....89
APPENDIX C – Performance graph of formulas....91
국문초록 93

List of Table

Table 3.1 Types of variables……..…….….24

Table 3.2 Characteristics of extracted CLASH datasets of vertical walls…....27

Table 3.3 Characteristics of extracted CLASH datasets of inclined seawalls....28

Table 3.4 New EurOtop inclined seawall data……..……29

Table 4.1 Normalized RMSE of various formulas……...……48

Table 4.2 Nash-Sutcliffe efficiency of various formulas ……..……49

Table 4.3 Index of agreement of various formulas …….……..…50

Table 4.4 Normalized RMSE of mean overtopping rate prediction model for inclined seawalls…….……..…65

Table 4.5 Nash-Sutcliffe efficiency of mean overtopping rate prediction model for inclined seawalls……...…..…..……65

Table 4.6 Index of Agreement of mean overtopping rate prediction model for inclined seawalls…….……..…65

Table 4.7 Performance comparison of wave overtopping formulas for vertical walls……..…66

Table 4.8 Characteristics of extracted CLASH datasets of inclined seawalls (oblique wave) …….……69

Table 4.9 Characteristics of extracted CLASH datasets of inclined seawalls (with roughness factor)……...……70

Table 4.10 Characteristics of datasets Characteristics of datasets (vertical walls with oblique wave)……....……76

List of Figures

Figure 2.1 Schematic diagram of ANN model for overtopping prediction (EurOtop, 2016). 14

Figure 2.2 Separation of training data and test data in the GMDH algorithm. .16

Figure 2.3 Selection of optimal GMDH model 19

Figure 3.1 Schematic diagram of vertical wall. 25

Figure 3.2 Schematic diagram of inclined seawall. 25

Figure 3.3 Results of bootstrap sampling for relative wave period. 30

Figure 3.4 Results of bootstrap sampling for relative water depth at the toe. ...31

Figure 3.5 Results of bootstrap sampling for structure slope. 31

Figure 3.6 Results of bootstrap sampling for seabed slope. 32

Figure 3.7 Results of bootstrap sampling for relative crest freeboard 32

Figure 4.1 Distribution of estimation error 37

Figure 4.2 Comparison of $q*$ between observation and estimation for training data and test data. 39

Figure 4.3 Comparison of $q*$ between observation and estimation for vertical walls and inclined seawalls and new Eurotop Manual data for inclined walls. 41

Figure 4.4 Ratio of estimated to observed wave overtopping rates versus estimated dimensionless overtopping rate. The dash-dot and dashed lines indicate 95% prediction error band. 43

Figure 4.5 Comparison of GMDH formulas with other formulas: (a) Goda (2009) formulas; (b) EurOtop (2016) formulas; (c) Etemad-Shahidi (2014, 2016) formulas. 47
Figure 4.6 Sensitivity analysis of GMDH formulas for toe depth at vertical walls 52
Figure 4.7 Sensitivity analysis of GMDH formulas for wave period at vertical walls 54
Figure 4.8 Sensitivity analysis of GMDH formulas for toe depth at inclined seawalls 56
Figure 4.9 Sensitivity analysis of GMDH formulas for wave period at inclined seawalls 58
Figure 4.10 Sensitivity analysis of GMDH formulas for structure slope of inclined seawalls 60
Figure 4.11 Comparison of $q*$ between observation and estimation with different oblique wave effect and roughness at GMDH-based inclined seawalls formula. 73
Figure 4.12 Comparison of $q*$ between GMDH-based formula and EurOtop (2016) with (a) roughness factor and (b) oblique wave factor 74
Figure 4.13 Comparison of $q*$ between GMDH-based formula and EurOtop (2016) for vertical walls at different oblique wave factor. 78
Figure 4.14 Comparison of $q*$ between GMDH-based formula and EurOtop (2016) for vertical walls with oblique wave factor. 79

List of Symbol

Latin Uppercase

H_{m0}	Significant wave height at the toe of structure
N_0	Number of total incident waves
$L_{m-1,0}$	Deepwaver wavelength with the period $T_{m-1,0}$
H_0'	Equivalent deepwater wave height
Q	Wave ovetopping rate of each wave
$T_{m-1,0}$	Mean wave period

Latin Lowercase

g	Gravitational acceleration
h^*	Dimensionless parameter for impulsive wave
h_c	Crest freeboard
h_t	Water depth at the toe of structure
q	Mean wave overtopping rate
r^2	External criteria
$s_{m-1,0}$	Wave steepness ($H_{m0} / L_{m-1,0}$)

s	Seabed slope

Greek Lowercase

α	Slope angle of structures
β	Angle of wave attack
γ_b	Influence factor of a berm
γ_f	Influence factor of roughness elements on a slope
γ_β	Influence factor of oblique wave attack
γ_v	Influence factor of wall on top of a slope
$\xi_{m-1,0}$	Surf similarity parameter

CHAPTER 1. INTRODUCTION

1.1 Background

As the sea level rises and storm intensity increases in accordance with global warming, flood risks due to wave overtopping are also increasing in coastal areas. To cope with the overtopping risks, it is common to design the coastal structures so that the wave overtopping rate does not exceed a tolerable limit. Therefore, it is important to estimate beforehand how much of wave overtopping will occur depending on the wave condition and structural characteristics.

Since the wave overtopping is a very complex phenomenon and it is sensitive to wave and structural parameters, hydraulic experiments have been widely used to estimate the wave overtopping rates. The accumulated data were then used to develop design diagrams or empirical formulas. Goda et al. (1975) proposed design diagrams to estimate the mean wave overtopping rate in terms of equivalent deepwater wave height, toe depth, crest freeboard, and bottom slope. The Shore Protection Manual (U.S. Army, 1984) presents an empirical equation that requires the estimation of two coefficients from a diagram, with very limited data given on which to base an estimate for these coefficients. Results from the use of these diagrams or equations are very approximate at best. Moreover, they can only be used for specific structure types and bottom slopes.

Recently, by collecting a large amount of data from many independent projects

and test series involving a great variety of parameters, Van der Meer et al. (2005) constructed the CLASH (Crest Level Assessment of coastal Structures by full scale monitoring, neural network prediction and Hazard analysis on permissible wave overtopping) database. The database has been used for deriving empirical formulas by the EurOtop Manual (Pullen et al., 2007), Goda (2009), Etemad-Shahidi and Jafari (2014), and Etemad-Shahidi et al. (2016). Very recently the EurOtop Manual was updated (Van der Meer et al., 2016) by including more experimental data and new information on wave overtopping. These empirical formulas are compiled in Appendix A. To complement the EurOtop Manual, artificial neural network (ANN) models were presented (Van Gent et al. 2007; Zanuttigh et al. 2016), which is one of the data models. The formulas of Etemad-Shahidi and Jafari (2014) and Etemad-Shahidi et al. (2016) were also derived using data modeling approaches.

Among the fore-mentioned empirical formulas, the Goda's (2009) unified formulas were derived to predict the mean overtopping rate at structures with smooth impermeable surfaces, either vertical or inclined. Since then, more experimental data were produced, which were used to revise the EurOtop Manual. In this study, new unified formulas are derived using the group method of data handling (GMDH) algorithm with the augmented CLASH datasets. Unlike the ANN models, the GMDH model provides a series of mathematical formulas, which can easily be used by engineers who are familiar with empirical formulas. Because of this advantage, the GMDH algorithm has been used in various fields of civil engineering (Tsai et al. 2009; Garg 2015; Shahabi et al. 2016; Tsai and Yen 2017).

The uncertainty of the developed model is given by the confidence interval of prediction error and prediction error band as done by Van Gent et al. (2007) and Goda (2009), respectively. The performance of the GMDH model is compared with previous empirical formulas in terms of several statistical indices. Finally, sensitivity analysis is made with respect to the relative crest freeboard, which is the parameter that is the most closely related to the mean wave overtopping rate.

1.2 Objectives

The main objective in this research is to propose unified wave overtopping prediction formula for vertical and inclined seawalls which are similar to Goda (2009) and EurOtop (2007, 2016) using GMDH (Group Method of Data Handling) algorithm. For EurOtop formulas, it is inconvenient to use two or more formulas depending on the condition of the wave or the form of the structure. Goda proposed an integrated formula for some structures to solve this inconvenience. Using GMDH algorithm, some unified form of formula which is similar to empirical formula can be derived, so it is easy to predict wave overtopping rate using GMDH-based formula.

The additional objective is to derive model uncertainty, such as confidence interval for prediction error and prediction error band which is similar to Van Gent *et al.* (2007) and Goda (2009). By using these methods, it is considered that the wave overtopping rate prediction will be possible considering the uncertainty of

prediction. Finally, performance of previous empirical formula and formula by GMDH model are proposed using some statistical values.

Last objective is to derive wave overtopping formula consider influence factors for oblique wave attack and roughness of slope at inclined seawalls.

CHAPTER 2. THEORETICAL BACKGROUDS

2.1 Mean wave overtopping rate

The mean wave overtopping rate is used for probability design of coastal structures near the coastal area. Goda (2010) proposed that overtopping is not a continuous process but an intermittent occurrence at times of attack of individual high waves among the storm waves. Generally, the degree of overtopping rate is measured by the amount of overtopped water onto the land area, either as the amount per wave per unit length of seawall or as the mean rate of overtopping volume per unit length during the occurrence of storm waves.

Mean overtopping rate, which is averaged overtopping rate over the duration of the waves, is denoted as *q*,

$$q = \frac{1}{t_0} \sum_{i=1}^{N_0} Q(H_i, T_i) \tag{2.1}$$

where N_0 is the total number of waves, H_i, T_i are the wave height and period of the *i*th individual wave attacking the seawall, respectively. The wave overtopping database in this study are the mean overtopping rate by laboratory experiments.

2.2 Overtopping rate prediction formula

2.2.1 EurOtop formula (2007, 2016)

Van der Meer *et al*. (2005) and other researchers have built a number of hydraulic experiment data through a large project called CLASH. Using this data, several empirical formulas for predicting mean wave overtopping rate have been proposed. EurOtop formula, which was suggested in the EurOtop manual (2007), is a representative wave overtopping rate prediction formula among them. The formula proposed in the first edition of the EurOtop manual in 2007 has been proposed in a revised form reflecting additional experimental data in the second edition of the EurOtop manual, which is newly revised in 2016.

In the manual, the formulas were proposed based on probabilistic design and deterministic design, but in this study, we propose a formula based on probabilistic design using mean wave overtopping rate data. Therefore, the formula proposed as a probabilistic design basis among the empirical formulas proposed in the previous research is set as a comparison target. The probabilistic design-based formula which first proposed in the EurOtop manual (2007) is as follows.

(2.2) EurOtop (2007) formulas

Vertical walls:

(a) $h_* > 0.3$ (non-impulsive condition)

$$\frac{q}{\sqrt{gH_{m0}^3}} = 0.04\exp\left[-2.6\frac{h_c}{H_{m0}}\right] \quad : \quad 0.1 < \frac{h_c}{H_{m0}} < 3.5 \tag{2.2a}$$

with

$$h_* = 1.35\frac{h_t}{H_{m0}}\frac{h_t}{L_{m-1,0}} \quad : \quad L_{m-1,0} = \frac{g}{2\pi}T_{m-1,0}^2$$

(b) $h_* \le 0.2$ (impulsive conditon)

$$\frac{q}{h_*^2\sqrt{gh_t^3}} = 1.5\times10^{-4}\left(h_*\frac{h_c}{H_{m0}}\right)^{-3.1} \quad : \quad 0.03 < h_*\frac{h_c}{H_{m0}} < 1.0$$

$$\frac{q}{h_*^2\sqrt{gh_t^3}} = 1.5\times10^{-4}\left(h_*\frac{h_c}{H_{m0}}\right)^{-3.1} \quad : \quad 0.02 < h_*\frac{h_c}{H_{m0}} < 0.03; \text{ unbroken waves} \tag{2.2b}$$

$$\frac{q}{h_*^2\sqrt{gh_t^3}} = 2.7\times10^{-4}\left(h_*\frac{h_c}{H_{m0}}\right)^{-2.7} \quad : \quad 0.02 < h_*\frac{h_c}{H_{m0}} < 0.03; \text{ broken waves}$$

$$\frac{q}{h_*^2\sqrt{gh_t^3}} = 2.7\times10^{-4}\left(h_*\frac{h_c}{H_{m0}}\right)^{-2.7} \quad : \quad h_*\frac{h_c}{H_{m0}} < 0.02$$

(c) $0.2 < h_* \le 0.3$ (transition condition)

$$q = \max\left\{\sqrt{gH_{m0}^3}\,0.04\exp\left[-2.6\frac{h_c}{H_{m0}}\right],\ h_*^2\sqrt{gh_t^3}\times1.5\times10^{-4}\left(\frac{h_c}{H_{m0}}\right)^{-3.1}\right\} \tag{2.2c}$$

(d) If $h_c = 0$, $\quad \dfrac{q}{\sqrt{gH_{m0}^3}} = 0.063 \pm 0.0062$ (2.2d)

Inclined seawalls:

(e) If $\xi_{m-1,0} \le 5$,

$$\frac{q}{\sqrt{gH_{m0}^3}} = \frac{0.067}{\sqrt{\tan\alpha}} \gamma_b \xi_{m-1,0} \exp\left(\frac{-4.75 h_c}{\xi_{m-1,0} H_{m0} \gamma_b \gamma_f \gamma_\beta \gamma_v} \right)$$

with a maximum of : $\frac{q}{\sqrt{gH_{m0}^3}} = 0.2 \exp\left(\frac{-2.6 h_c}{H_{m0} \gamma_f \gamma_\beta} \right)$ (2.2e)

where $\xi_{m-1,0} = \frac{\tan\alpha}{\sqrt{H_{m0} / L_{m-1,0}}}$

(f) If $\xi_{m-1,0} \ge 7$,

$$\frac{q}{\sqrt{gH_{m0}^3}} = 10^{-0.92} \exp\left(-\frac{h_c}{\gamma_f \gamma_\beta H_{m0} \left(0.33 + 0.022 \xi_{m-1,0}\right)} \right) \quad (2.2f)$$

(g) If $5 < \xi_{m-1,0} < 7$, use linear interpolation:

$$\frac{q}{\sqrt{gH_{m0}^3}} = \frac{\left(7 - \xi_{m-1,0}\right) A + \left(\xi_{m-1,0} - 5\right) B}{2} \quad (2.2g)$$

where A and B are the values calculated for $\xi_{m-1,0} = 5$ and $\xi_{m-1,0} = 7$, respectively.

(2.3) EurOtop (2016) formulas

Vertical walls:

(a) $\dfrac{h_t}{H_{m0}L_{m-1,0}} > 0.23$ (non-impulsive condition)

$$\frac{q}{\sqrt{gH_{m0}^3}} = 0.05\exp\left(-2.78\frac{h_c}{H_{m0}}\right) \tag{2.3a}$$

(b) $\dfrac{h_t}{H_{m0}L_{m-1,0}} \le 0.23$ (impulsive conditon)

$$\frac{q}{\sqrt{gH_{m0}^3}} = 0.011\left(\frac{H_{m0}}{h_t s_{m-1,0}}\right)^{0.5}\exp\left(-2.2\frac{h_c}{H_{m0}}\right) \quad : \quad 0 < \frac{h_c}{H_{m0}} < 1.35 \tag{2.3b}$$

$$\frac{q}{\sqrt{gH_{m0}^3}} = 0.0014\left(\frac{H_{m0}}{h_t s_{m-1,0}}\right)^{0.5}\left(\frac{h_c}{H_{m0}}\right)^{-3} \quad : \quad \frac{h_c}{H_{m0}} \ge 1.35$$

Inclined seawalls:

(c) If $\xi_{m-1,0} \le 5$,

$$\frac{q}{\sqrt{gH_{m0}^3}} = \frac{0.023}{\sqrt{\tan\alpha}} \gamma_b \xi_{m-1,0} \exp\left[\left(-2.7 \frac{h_c}{\xi_{m-1,0} H_{m0} \gamma_b \gamma_f \gamma_\beta \gamma_v}\right)^{1.3}\right] \tag{2.3c}$$

with a maximum of $\frac{q}{\sqrt{gH_{m0}^3}} = 0.09 \exp\left[\left(-1.5 \frac{h_c}{H_{m0} \gamma_f \gamma_\beta \gamma^*}\right)^{1.3}\right]$

(d) If $\xi_{m-1,0} \ge 7$ and $s_{m-1,0} < 0.01$,

$$\frac{q}{\sqrt{gH_{m0}^3}} = 10^{-0.79} \exp\left(-\frac{h_c}{\gamma_f \gamma_\beta H_{m0} \left(0.33 + 0.022 \xi_{m-1,0}\right)}\right) \tag{2.3d}$$

(e) If $5 < \xi_{m-1,0} < 7$, use linear interpolation:

$$\frac{q}{\sqrt{gH_{m0}^3}} = \frac{\left(7 - \xi_{m-1,0}\right) A + \left(\xi_{m-1,0} - 5\right) B}{2} \tag{2.3e}$$

where A and B are the values calculated for $\xi_{m-1,0} = 5$ and $\xi_{m-1,0} = 7$, respectively.

Among the parameters used in the EurOtop formula, γ_b, γ_f, γ_β, γ_v are the influence factor on the berm, the influence factor for the roughness of slope, the influence factor on the angle between the structure and the oblique wave and the influence factor on the presence of the upright wall of the structure, respectively. In this study, the influence factors are assumed to be 1 by using the data which have conditions that the wave propagates perpendicular to the structure in which there is no berm, no upright wall of the upper part of the structure and smooth slope.

2.2.2 Formula by Goda (2009)

Goda proposed some unified formula by using the part of CLASH database. He extracted the data of vertical wall and inclined seawall which is a simple structure among the CLASH data, and using some curve fitting method with respect to various variable, he suggested following formula.

Goda's (2009) unified formulas

$$\begin{aligned} \frac{q}{\sqrt{gH_{m0}^3}} &\equiv q^* \\ &= \exp\left\{-\left[A + B\frac{h_c}{H_{m0}}\right]\right\} \end{aligned} \tag{2.4}$$

where

$$A = A_0 \tanh\left\{\left(0.956 + 4.44s\right) \times \left[h_t / H_{m0} + 1.242 - 2.032s^{0.25}\right]\right\}$$
$$B = B_0 \tanh\left\{\left(0.822 - 2.22s\right) \times \left[h_t / H_{m0} + 0.578 + 2.22s\right]\right\}$$
$$A_0 = 3.4 - 0.734\cot\alpha + 0.239\cot^2\alpha - 0.0162\cot^3\alpha$$
$$B_0 = 2.3 - 0.50\cot\alpha + 0.15\cot^2\alpha - 0.011\cot^3\alpha$$

Goda used the exponential relationship between the wave overtopping rate and relative crest freeboard by Verhaeghe (2005), and using the formula for significant wave by Goda, the significant wave heights of some CLASH data by SWAN model was revised.

2.2.3 Formula by Etemad-Shahidi *et al*. (2014, 2016)

For inclined seawalls, Etemad-Shahidi and Jafari (2014) proposed a wave overtopping rate prediction formula by selecting data from the CLASH database. They used decision tree and nonlinear regression. The formula is as follows.

Etemad-Shahidi and Jafari's (2014) formulas for inclined seawalls

If $\frac{h_c}{H_{\mathrm{m0}}} \le 1.62$, then

$$q^* = \frac{q}{\sqrt{gH_{\mathrm{m0}}^3}} = 0.032\exp\left(-2.6\left(\frac{h_c}{H_{\mathrm{m0}}}\right)^{1.6}\left(\xi_{m-1,0}\right)^{-1.26}\right) \tag{2.5a}$$

If $\frac{h_c}{H_{m0}} > 1.62$, then

$$q^* = \frac{q}{\sqrt{gH_{m0}^3}} = 0.032 \exp\left(-5.63\left(\xi_{m-1,0}\right)^{-1.26} - 3.283\left(\frac{h_c}{H_{m0}} - 1.62\right)^{0.83} \right) \quad (2.5b)$$

Also, Etemad-Shahidi *et al.* (2016) proposed overtopping rate prediction formula for composite vertical wall using machine learning model tree, and the formula is as follows.

Etemad-Shahidi et al.'s (2016) formula for vertical walls

$$q^* = \frac{q}{\sqrt{gH_{m0}^3}} = 0.9\left(\tan\alpha\right)^{0.17} \left(\frac{L_{m-1,0}}{h_t}\right)^{0.1} \exp\left(0.67\frac{d}{h_t} - 5.77\left(\frac{h_c}{H_{m0}}\right)^{0.416} \right) \quad (2.6)$$

where d = berm depth.

2.2.4 ANN model (Van Gent *et al.*, 2007, Zanuttigh *et al.*, 2016)

Van Gent *et al*. (2007) and Zanuttigh *et al*. (2016) proposed mean wave overtopping rate prediction model using artificial neural network (ANN). In this model, 500 ANN models are constructed, and then all predicted wave overtopping rate at each ANN model are ensemble averaged. This ensemble averaged value is proposed as final predicted wave overtopping rate. Following Figure 2.1 is a conceptual diagram of ANN model included in the revised EurOtop manual (2016).

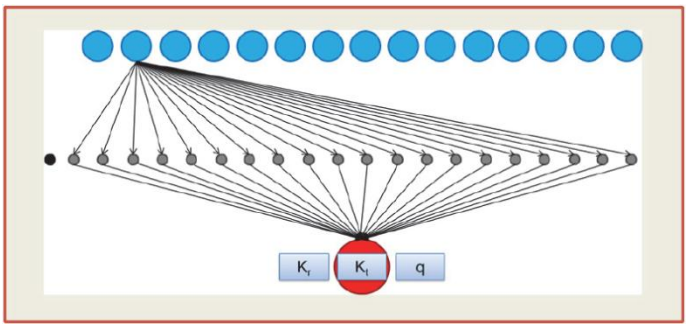


Figure 2.1 Schematic diagram of ANN model for overtopping prediction (EurOtop, 2016)

In the result of the ANN model, the confidence interval based on the bootstrap resampling is presented together for the uncertainty of the model itself. The artificial neural network model has the advantage of predicting the mean wave overtopping rate considering a wide variety of variables and showing relatively higher accuracy than the previous empirical formula. However, there is an inconvenience that 500 ANN models cannot be used as empirical formulas.

2.3 Group Method of Data Handling

2.3.1 Basic concept of GMDH algorithm

Group Method of Data Handling (GMDH) model is a computer-based data model developed by Ivakhnenko (1968). It has a structure similar to that of ANN, which is a typical example of machine learning techniques. Using GMDH model, the final result is presented in one concrete formula form and the final result can be derived using only the initial set variables. Therefore, it is easier to derive the results than

ANN.

The training and testing steps of the existing ANN are applied to the configuration of the GMDH model. First, data extraction to be used for training in the first generation is needed, and then combine two variables into two quadratic polynomials using selected two variables. In this process, we apply the least square method which is generally used in regression analysis. Each of these polynomials is called a partial model, and the first step of a partial model is called first generation (1^{st} generation). Partial models made in the first generation are applied as another input variable in the next generation. In this case, the partial models are selected in pairs in the same manner as the first generation variable and constitute another partial model of the second-order polynomial form which has same form of the first generation. The number of polynomials increases exponentially as the generations are increased. For example, if there are n variables in the first generation, the number of partial models that can be created by selecting two of n variables in the first generation is $\binom{n}{2} = \frac{n(n-1)}{2}$. In the second generation, $\binom{n(n-1)/2}{2}$ partial models can be created by using the partial model created in the first generation as a new variable. With this tendency, as the generation increases, the partial model increases exponentially.

2.3.2 Procedure of GMDH algorithm

First, we need to construct variables to form GMDH model. Figure 2.2 is a conceptual diagram that explains selection of training data and test data when constructing GMDH model.

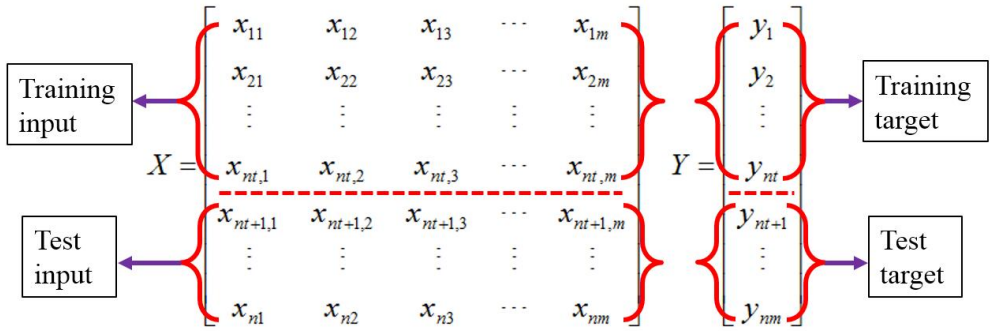


Figure 2.2 Separation of training data and test data in the GMDH algorithm

For example, assuming that two of the input variables used in training are $\mathbf{x}_1 = (x_{11}, x_{21}, \ldots, x_{nt,1})$, $\mathbf{x}_2 = (x_{12}, x_{22}, \ldots, x_{nt,2})$, we can construct the following partial model by combining the data corresponding to each variable.

$$X_{t,1} = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{11}^2 & x_{12}^2 & x_{11}x_{12} \\ 1 & x_{21} & x_{22} & x_{21}^2 & x_{22}^2 & x_{21}x_{22} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{nt,1} & x_{nt,2} & x_{nt,1}^2 & x_{nt,2}^2 & x_{nt,1}x_{nt,2} \end{bmatrix} \tag{2.7}$$

$$Y_t = \begin{bmatrix} y_1 & y_2 & \ldots & y_{nt} \end{bmatrix}^T \tag{2.8}$$

In the above partial model, the regression coefficient matrix for the partial model can be obtained by using the following least squares method.

$$
\begin{aligned}
&Y_t = X_{t,1}A \rightarrow X_{t,1}^T Y_t = X_{t,1}^T X_{t,1} A \\
&\rightarrow \left(X_{t,1}^T X_{t,1}\right)^{-1} X_{t,1}^T Y_t = \left(X_{t,1}^T X_{t,1}\right)^{-1} X_{t,1}^T X_{t,1} A = A \\
&\rightarrow A = \left(X_{t,1}^T X_{t,1}\right)^{-1} X_{t,1}^T Y
\end{aligned}
\tag{2.9}
$$

Applying the calculated regression coefficient matrix to the input variables of the test data yields the following estimator matrix.

$$
\hat{Y} = AX_{te} = [\hat{y}_{nt+1}, \hat{y}_{nt+2}, ..., \hat{y}_n]^T \tag{2.10}
$$

2.3.3 External criteria for self-organizing

If the partial model increases exponentially, unnecessary computation time will increase, which may cause inconvenience to the model construction. To prevent this, and to select the optimal combination of variables for each generation, the external criterion is introduced in the GMDH model. This is a reference value of a form in which the square error between the observed value and the estimated value is normalized to the sum of squares of the observed values, and can be expressed as follows.

$$r_j^2 = \frac{\sum_{i=nt+1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=nt+1}^{n} y_i^2} \qquad j = 1,\ 2,\ \ldots,\ \frac{m(m-1)}{2} \tag{2.11}$$

If the external criterion calculated using the above equation is smaller than the reference value set for the specific value, the partial model is adopted as a new variable in the next generation, and discarded if it is larger than the reference value. In this way, the GMDH model is subjected to a self-organizing process that removes the partial model with a relatively high external criterion and uses the remaining partial model as the input variable in the next generation, and the model is constructed. The performance criterion can be set arbitrarily, and when the external criterion of a partial model with the lowest external criterion value in a generation, as in the following Figure 2.3, shows the minimum value without any further reduction as the generation increases, the smallest external criterion will be adopted as the final GHDM model. The shape of the model is derived in the following form, commonly called the Ivakhnenko polynomial.

$$y = a_0 + \sum_{i=1}^{m} a_i x_i + \sum_{i=1}^{m}\sum_{j=1}^{m} a_{ij} x_i x_j + \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} a_{ijk} x_i x_j x_k + \ldots \tag{2.12}$$

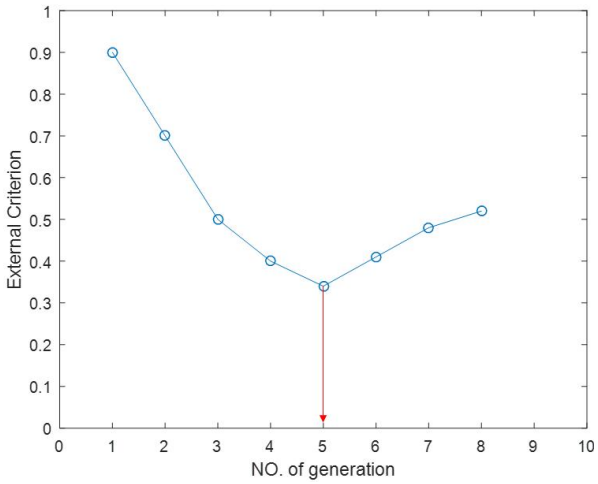


Figure 2.3 Selection of optimal GMDH model

In general, the GMDH model has to go through several generations to reach the optimal partial model, which makes the form of the model considerably complicated, making it difficult to use in the formula. Therefore, in this study, the number of generations was limited to 3 to minimize the complexity of the model. If the number of generation is reduced too much, the number of variable combinations is low, so the performance of the model may deteriorate. In this case, the minimum value of the external criterion in each generation continues to decrease do. Therefore, the partial model which shows the lowest external criterion in the third generation is presented as the final model proposed in this study. In this study, we tried to find the optimum performance of GMDH within the limited complexity by using the reference value of the external criterion including weights by Heris (2015).

$$W = \alpha r_{j.\min}^2 + (1-\alpha) r_{j.\max}^2 \tag{2.13}$$

In the above equation, W is the actual performance criterion which is used when selecting the partial model, and $r_{j.\min}^2$, $r_{j.\max}^2$ are the minimum and maximum values of the external criterion at the partial models in a certain generation, respectively. The optimal model of GMDH can be found by adjusting the ratio α. When α is increased, since the W value is shifted to the minimum external criterion, fewer partial models are selected and the number of variables that can constitute the GMDH model after the second generation is reduced, but the time for constructing the model is decreased. On the other hand, as α becomes smaller, the W value is shifted to the maximum external criterion, so that more partial models are selected and the calculation time is increased. However, the optimal partial model considering more various combinations can be found. In this study, we tried to find optimal models by considering various combinations of variables while keeping α at 0.1.

2.4 Data sampling technique – bootstrap sampling

Simliar to other machine learning models, the GMDH model divides whole data into training sets and test sets, and using the training dataset, GMDH model is constructed. The test dataset are used to evaluate model performance. This procedure also used in general regression analysis. Therefore, how to configure the training dataset and test dataset is an important factor for ensuring the model generalization. When training dataset is selected from the certain range of data, the performance of model can be biased at certain range of data, so there is a possibility

that the performance when a test dataset is applied may be extremely reduced. This is a major cause of overfitting. Therefore, in this study, the bootstrap sampling technique was applied for prevent overfitting.

Bootstrap sampling is a non-parametric estimation method that can be used to extract a sample with a similar distribution to a population without knowing the estimator of the population parameter. In general, the bootstrap sampling method is used to derive model uncertainty by constructing several models through repetitive sampling, and ensemble averaging of all model results to derive the final results. It also can present the uncertainty of the model using the confidence interval. However, in this study, bootstrap sampling was used to ensure the representative of training data, because even if we not know the distribution of all data, if similar ratio of data with total data for each variable interval can be used for training, the model will not be biased to a specific range at the test dataset and will show generalized performance.

In the bootstrap sampling method, for example, when N pieces of data exist, arbitrary data of N pieces of data is sampled with replacement N times. At this time, the probability that one specific data will be extracted becomes $1/N$, and the probability that certain data will not be extracted is $1-(1/N)$. So, if N datasets are extracted by using bootstrap sampling, the probability that specific data is not extracted in the dataset is $\left[1-(1/N)\right]^N$, and when N is very large, this converges to about 0.37. Unselected data is used as test data and selected data including duplicated data is used as training data. That is, about 37% of the total data is used

as test data and the remaining 63% is used as training data.

In this study, the data used to construct GMDH model are 2,097 (444 vertical wall data and 1,653 inclined seawall data) that are part of the CLASH dataset. Approximately 63% of these data types are selected in duplicate, so the number of training data include duplicated data is 2,097. The remaining 37% are used as test data to evaluate the performance of the partial model.

CHAPTER 3. METHODOLOGY

3.1 CLASH data for model construction

3.1.1 Variables applying Froude similitude law

The variables used in this study are the six dimensionless variables, which are the same as those of the previous study. All the variables were applied to the following Froude similitude law.

$$\left(\frac{V}{\sqrt{gl}}\right)_{\text{prototype}} = \left(\frac{V}{\sqrt{gl}}\right)_{\text{model}} \tag{3.1}$$

In general, the reason for applying the Froude similitude law when deriving the mean wave overtopping rate formula is to take into account the scale of data performed under various hydraulic experimental conditions, and the empirical formula obtained from this can be used in the field or experiment regardless of scale. Following Table 3.1 shows the dimensionless variables using the Froude similitude law. Among the following variables, the input variables are five variables excluding the dimensionless mean wave overtopping rate, and the dimensionless wave overtopping rate is used as the output value since it is a value that should ultimately be predicted using the model.

Table 3.1 Types of variables

Type	Information
h_c / H_{m0}	Relative crest freeboard
h_t / H_{m0}	Relative water depth at the toe of structure
$T_{m-1,0} / \sqrt{H_{m0}}$	Relative wave period at the toe of structure
$\cot\alpha$	Slope of structure
s	Seabed slope
$q / \sqrt{gH_{m0}^3}$	Dimensionless mean wave overtopping rate

The unit of wave period is second (s), and the unit of wave height at the toe of structure is meter (m). Also, following Figure 3.1 and Figure 3.2 show the structure of the data obtained in this study.

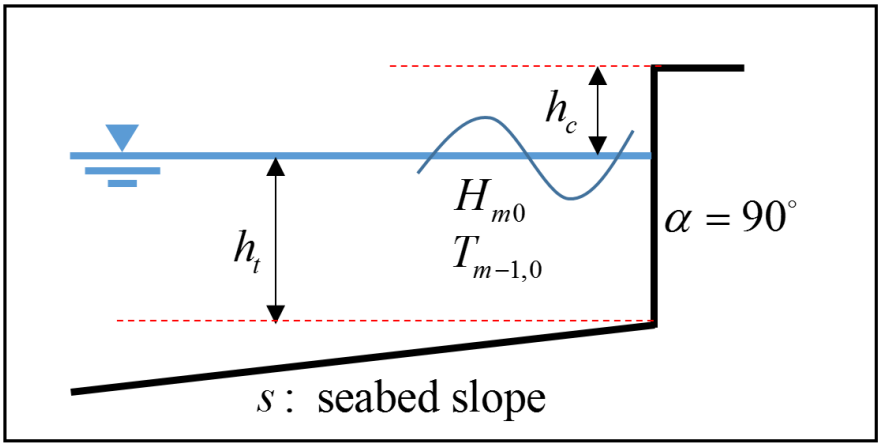


Figure 3.1 Schematic diagram of vertical wall

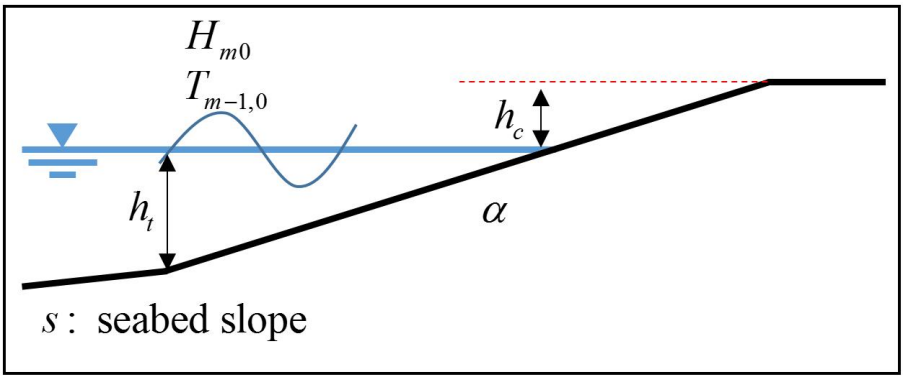


Figure 3.2 Schematic diagram of inclined seawall

3.1.2 Selected data for GMDH model

The datasets used to construct the GMDH model in this study were extracted only for structures with smooth impermeable uniform slopes (vertical or inclined) from the augmented CLASH database. The database has assigned a reliability factor *RF* and a structural complexity factor *CF* to each test. In this study, unreliable tests (*RF*=4) were excluded from the datasets. The values of *CF* were 1 or 2 for all the tests, indicating simple structures. Noting that tests with very small overtopping rates may not be accurate due to measurement errors, following Verhaeghe (2005) and Van Gent et al. (2007), the data of $q < 10^{-6}$ $m^3/s/m$ were also excluded. The datasets used in this study for vertical and inclined seawalls are listed in Tables 3.2 to 3.4, in which the ranges of the input and output variables are also given. The total number of data is 2,097; 444 for vertical walls, and 1,653 for inclined seawalls. The 606 data in Table 3.4 are the datasets included in the new EurOtop Manual (Van der Meer et al. 2016). Without these, the datasets are essentially the same as those used by Goda (2009), except that we excluded some data of low reliability. The datasets in Tables 3.2 and 3.3 were also used by Etemad-Shahidi et al. (2016) and Etemad-Shahidi and Jafari (2014), respectively, after removing some datasets and reducing the number of data in some datasets.

Table 3.2 Characteristics of extracted CLASH datasets of vertical walls

Dataset ID	Number of data	$\cot\alpha$	s	$T_{m-1,0}/\sqrt{H_{m0}}$	h_t/H_{m0}	h_c/H_{m0}	$q/\sqrt{gH_{m0}^3}$
Data_028	129	0.0	0.01, 0.033, 0.1	3.75–8.41	1.00–3.39	0.43–2.33	2.53×10^{-5}–2.69×10^{-2}
Data_106	28	0.0	0.001	2.82–5.74	2.70–14.35	0.12–1.29	5.13×10^{-4}–4.55×10^{-2}
Data_107	55	0.0	0.001	4.43–21.93	2.49–17.02	0.00–1.16	8.25×10^{-4}–7.56×10^{-2}
Data_113	28	0.0	0.001	2.82–5.74	2.70–14.35	0.10–1.29	5.13×10^{-4}–5.56×10^{-2}
Data_224	35	0.0	0.02	3.51–7.67	1.52–3.10	0.93–3.20	7.12×10^{-5}–4.39×10^{-3}
Data_225	18	0.0	0.05	3.63–5.94	1.21–3.16	1.27–2.74	3.24×10^{-4}–4.35×10^{-3}
Data_402	13	0.0	0.001	3.54–3.98	7.69–23.33	0.77–3.33	3.78×10^{-5}–7.57×10^{-3}
Data_502	47	0.0	0.02, 0.1	3.41–8.51	1.33–4.40	1.31–3.64	2.25×10^{-4}–3.72×10^{-3}
Data_802	91	0.0	0.033, 0.1	4.16–9.93	1.01–2.50	0.43–3.25	2.61×10^{-4}–2.96×10^{-2}
Total	444	0.0	0.001–0.1	2.82–21.93	1.00–23.33	0.00–3.64	2.53×10^{-5}–7.56×10^{-2}

Table 3.3 Characteristics of extracted CLASH datasets of inclined seawalls

Dataset ID	Number of data	$\cot\alpha$	s	$T_{m-1,0}/\sqrt{H_{m0}}$	h_t/H_{m0}	h_c/H_{m0}	$q/\sqrt{gH_{m0}^3}$
Data_030	139	1.0, 2.0, 4.0	0.001–0.05	3.23–5.80	1.41–5.99	0.00–3.00	2.84×10^{-5}–6.70×10^{-2}
Data_035	17	2.0	0.019	3.24–5.79	2.20–5.49	0.98–2.34	4.57×10^{-4}–1.12×10^{-2}
Data_042	215	2.0, 4.0	0.02–0.1	3.80–11.55	1.04–4.90	0.51–2.82	9.14×10^{-5}–4.47×10^{-2}
Data_101	33	6.0	0.001	3.15–22.59	3.50–10.29	0.50–1.47	3.00×10^{-3}–2.21×10^{-2}
Data_102	25	4.0	0.001	3.10–17.83	3.85–10.39	0.00–1.30	1.68×10^{-3}–1.35×10^{-1}
Data_103	15	3.0	0.001	3.39–10.29	3.55–6.72	0.00–0.92	1.19×10^{-2}–6.75×10^{-2}
Data_104	90	6.0	0.001	3.31–9.92	8.62–22.73	0.43–1.14	1.37×10^{-3}–4.06×10^{-2}
Data_109	16	6.0	0.034	3.99–5.11	1.39–1.73	1.19–1.48	2.51×10^{-4}–5.70×10^{-3}
Data_110	20	6.0	0.001	3.92–16.71	3.86–13.92	0.98–3.51	4.35×10^{-5}–9.03×10^{-3}
Data_217	13	6.0	0.001	3.86–7.82	3.04–8.04	1.25–2.50	1.29×10^{-5}–4.48×10^{-4}
Data_218	56	2.0, 4.0, 7.0	0.001	3.71–4.84	3.16–5.06	0.75–2.88	1.80×10^{-5}–4.25×10^{-3}
Data_220	20	2.5, 4.0	0.001	3.19–7.38	4.78–6.54	1.09–2.92	5.21×10^{-5}–8.56×10^{-3}
Data_221	65	3.0, 4.0	0.001, 0.01	3.42–7.87	1.64–6.00	1.47–3.27	7.07×10^{-6}–1.22×10^{-3}
Data_222	32	2.5, 4.0	0.001	3.26–7.62	4.68–7.20	0.97–2.75	1.69×10^{-5}–6.89×10^{-3}
Data_226	62	2.5, 4.0	0.004, 0.01	4.34–19.35	1.62–2.84	2.74–3.56	1.59×10^{-5}–1.07×10^{-3}
Data_227	98	3.0, 4.0, 6.0	0.001, 0.01	3.78–69.94	1.27–5.38	0.83–4.59	5.13×10^{-6}–2.33×10^{-2}
Data_703	29	1.19	0.001	2.74–6.29	3.10–10.53	0.54–1.83	2.42×10^{-3}–4.48×10^{-2}
Data_955	40	2.7	0.017–0.028	4.05–8.06	2.35–4.61	1.07–4.37	1.47×10^{-5}–3.29×10^{-2}
Data_956	38	2.72	0.020–0.021	3.83–5.39	1.27–2.84	0.65–1.77	2.31×10^{-3}–3.47×10^{-2}
Data_959	24	1.7–5.9	0.025	3.69–4.63	3.18–7.35	0.79–3.44	6.00×10^{-5}–7.43×10^{-3}
Total	1,047	1.0–7.0	0.001–0.1	2.74–69.94	1.04–22.73	0.00–4.59	5.13×10^{-6}–1.35×10^{-1}

Table 3.4 New EurOtop inclined seawall data

Dataset ID	Number of data	$\cot\alpha$	s	$T_{m-1,0} / \sqrt{H_{m0}}$	h_t / H_{m0}	h_c / H_{m0}	$q / \sqrt{gH_{m0}^3}$
Data_008	2	3.0	0.001	3.69, 5.29	3.82, 4.05	1.91, 2.03	4.67×10^{-4}–1.66×10^{-3}
Data_041	40	0.7, 1.0, 1.73	0.01	4.02–7.52	0.98–2.64	1.03–2.34	8.00×10^{-6}–8.30×10^{-3}
Data_043	366	0.36–2.75	0.001	3.64–6.51	2.59–23.43	0.10–1.69	1.78×10^{-4}–8.62×10^{-2}
Data_044	81	2.0, 3.0	0.0083, 0.009	3.24–8.57	2.34–6.94	0.84–3.15	1.39×10^{-4}–2.03×10^{-2}
Data_045	117	2.0, 3.0	0.0083, 0.009	3.51–9.51	2.23–7.33	0.59–2.63	1.15×10^{-4}–2.76×10^{-2}
Total	606	0.36–3.0	0.001–0.01	3.24–9.51	0.98–23.43	0.10–3.15	8.00×10^{-6}–8.62×10^{-2}

3.2 Training data extraction using bootstrap sampling

Figure 3.3 to Figure 3.7 shows the histogram of the training data extracted from the whole CLASH data used in the GMDH model construction using the bootstrap sampling technique. The bar with the dotted line shows the Frequency divided by the total number of data of all the data used to construct the GMDH model, and the solid line shows the frequency of the training data extracted by bootstrap sampling. The frequency divided by the total number of training data.

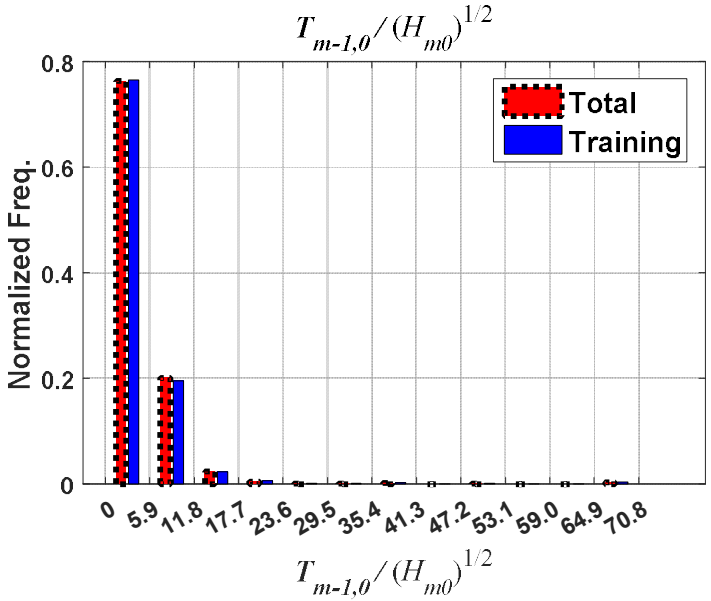


Figure 3.3 Results of bootstrap sampling for relative wave period

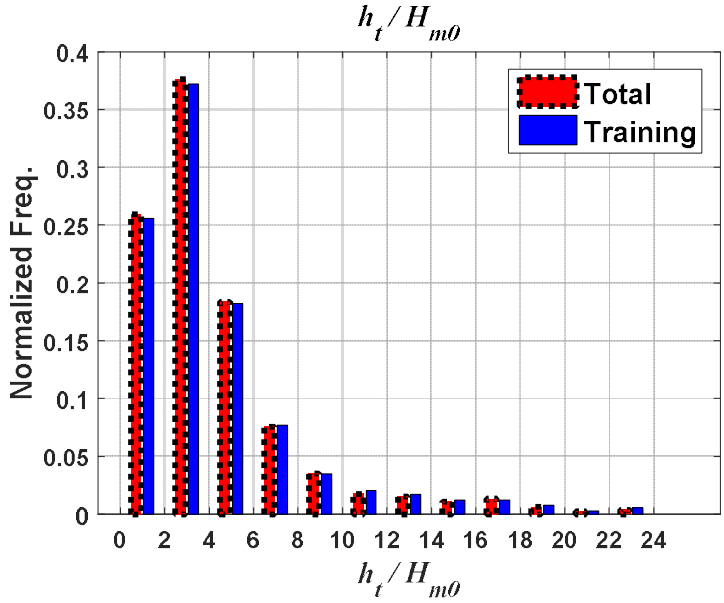


Figure 3.4 Results of bootstrap sampling for relative water depth at the toe

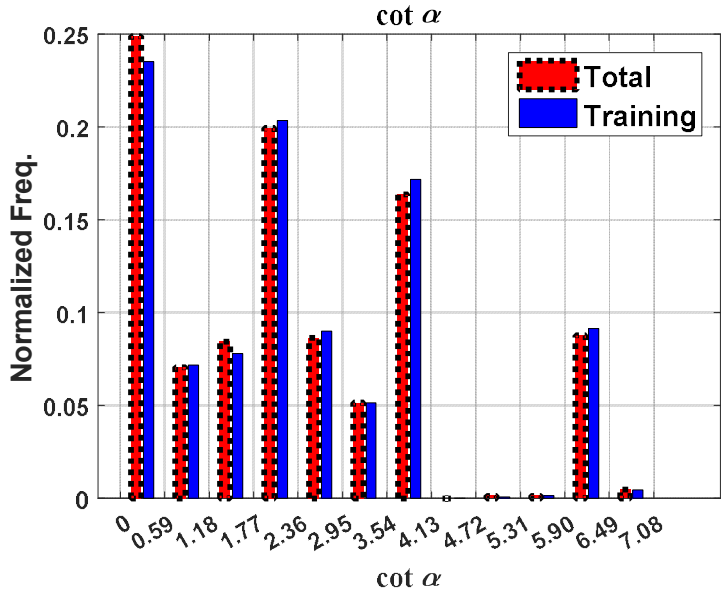


Figure 3.5 Results of bootstrap sampling for structure slope

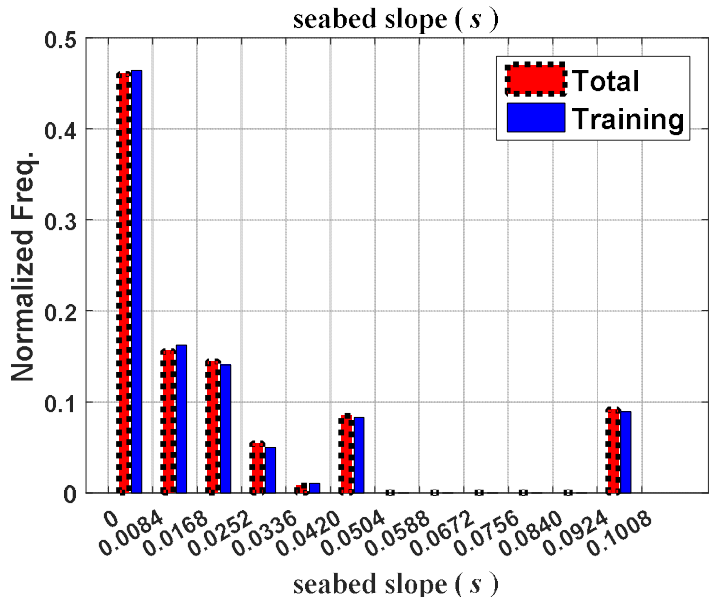


Figure 3.6 Results of bootstrap sampling for seabed slope

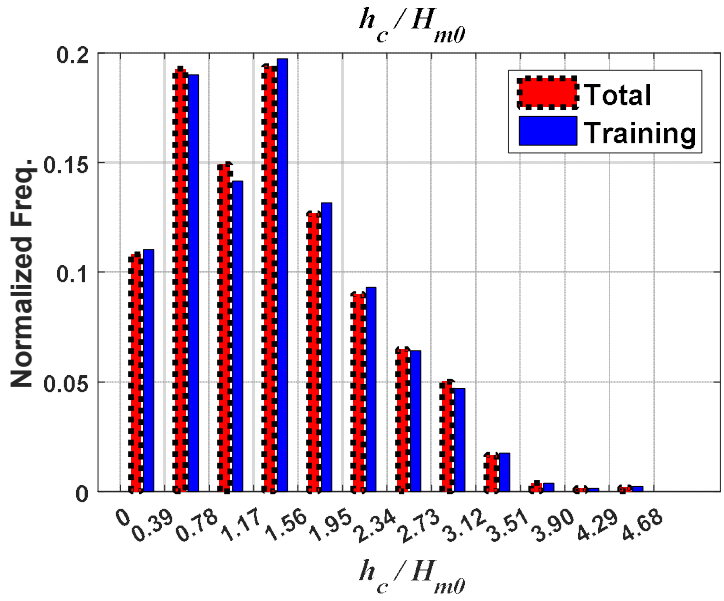


Figure 3.7 Results of bootstrap sampling for relative crest freeboard

As can be seen from the above figures, the training data with a similar ratio to the variable distribution of the whole data was extracted through bootstrap sampling in all the variables. The *K-S* test for each case showed that the hypothesis that data groups have a similar distribution is adopted with 95% significance level. In other words, the data used as training data in the model construction can be seen to reflect the tendency of the original data as a whole, so it is thought that this result can reduce the overfitting phenomenon that results are confined to specific data.

The normalized frequency is calculated by dividing the number of data in each sub-range by the total number of data. It is observed that the histograms of the training data are very similar to those of the populations for all the input variables. Without showing the result, we mention that the Kolmogorov-Smirnov test showed that the hypothesis that the training data and population have a similar distribution is adopted with 95% confidence level for each variable.

CHAPTER 4. RESULTS

4.1 Unified formulas for mean overtopping rate derived by GMDH algorithm

4.1.1 Unified formulas by GMDH algorithm

The formulas derived by the GMDH algorithm for predicting the mean rate of wave overtopping at vertical and inclined seawalls are given by following equations.

GMDH-based unified formulas for smooth, impermeable vertical and inclined seawalls.

$$\log_{10} q^* = \log_{10}\left(\frac{q}{\sqrt{gH_{m0}^3}}\right) = -0.046 + 0.314a + 0.701b + 0.881a^2 + 1.248b^2 - 2.127ab \quad (4.1a)$$

$$a = -81.079 - 2.314A - 62.69B - 0.027A^2 - 12.059B^2 - 1.296AB \quad (4.1b)$$

$$b = -0.3546 + 0.808A + 0.057C + 0.017A^2 + 0.102C^2 - 0.118AC \quad (4.1c)$$

$$A = -1.348 + 0.366\frac{T_{\text{m-1,0}}}{\sqrt{H_{\text{m0}}}} - 1.018\frac{h_c}{H_{\text{m0}}} - 0.057\left(\frac{T_{\text{m-1,0}}}{\sqrt{H_{\text{m0}}}}\right)^2 + 0.060\left(\frac{h_c}{H_{\text{m0}}}\right)^2 - 0.084\frac{h_c T_{\text{m-1,0}}}{\sqrt{H^3_{\text{m0}}}} \tag{4.1d}$$

$$B = -2.250 - 0.005\frac{T_{\text{m-1,0}}}{\sqrt{H_{\text{m0}}}} - 0.130\cot\alpha - 0.8139\left(\frac{T_{\text{m-1,0}}}{100\sqrt{H_{\text{m0}}}}\right)^2 + 0.007\cot^2\alpha + 0.002\frac{T_{\text{m-1,0}}\cot\alpha}{\sqrt{H_{\text{m0}}}} \tag{4.1e}$$

$$C = -2.549 - 0.062\frac{T_{\text{m-1,0}}}{\sqrt{H_{\text{m0}}}} + 0.049\frac{h_t}{H_{\text{m0}}} + 6.630\left(\frac{T_{\text{m-1,0}}}{100\sqrt{H_{\text{m0}}}}\right)^2 - 0.005\left(\frac{h_t}{H_{\text{m0}}}\right)^2 + 0.012\frac{h_t T_{\text{m-1,0}}}{\sqrt{H^3_{\text{m0}}}} \tag{4.1f}$$

*units: $T_{m-1,0}$(s), h_c(m), h_t(m), H_{m0}(m)

They are a little more complicated than previous empirical formulas, but they still enable one to calculate the wave overtopping rate using a desktop calculator. Note that the seabed slope was eliminated somewhere during the derivation of the GMDH formulas probably because its effect was negligible. Goda (2009) and Etemad-Shahidi et al. (2016) included the seabed slope in their formulas, but it was not included in the formulas of EurOtop Manual (2007, 2016) and Etemad-Shahidi et al. (2014).

4.1.2 Confidence interval of estimation error

Before evaluating the performance of the derived formulas, the uncertainty of the formulas is examined by the confidence interval of the estimation errors. Since the

order of magnitude of $q*$ varies greatly, the estimation error is defined as $\log q^*_{\text{est}} - \log q^*_{\text{obs}}$, where 'est' and 'obs' stand for estimation and observation, respectively, and which gives the same value for the same percent of error regardless of the magnitude of q^*_{obs}. For example, if q^*_{est} is greater than q^*_{obs} by 20% so that $q^*_{\text{est}} = 1.2 q^*_{\text{obs}}$, $\log q^*_{\text{est}} - \log q^*_{\text{obs}}$ is always 0.0792 regardless of the magnitude of q^*_{obs}. It was found that the errors calculated in this way follow a normal distribution by *t*-test. Figure 4.1 shows the histogram of the probability density of the estimation errors along with the normal distribution with the mean of -0.016 and the standard deviation of 0.321. The errors of 95% confidence interval are also given as $\log q^*_{\text{est}} - \log q^*_{\text{obs}} = -0.645$ and $\log q^*_{\text{est}} - \log q^*_{\text{obs}} = 0.613$. Therefore, if the mean wave overtopping rate is predicted by the GMDH formulas, 95% of the predicted overtopping rate is to be located in the range between 0.226 and 4.10 times the observed rate.

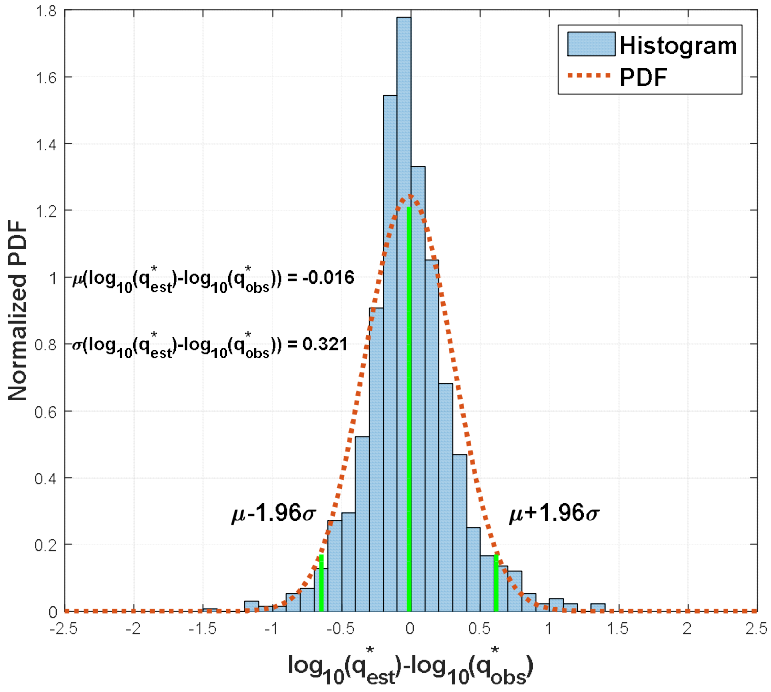


Figure 4.1 Distribution of estimation error

4.1.3 Performance of derived formulas

Figure 4.2 shows a comparison of $q*$ between observation and prediction for both training and test data along with the lines of mean and 95% confidence interval of the estimation error. The data of relatively large overtopping rates are situated within the lines of 95% confidence interval, while scattering becomes severe as the overtopping rate decreases. The derived formulas are accurate for very large overtopping rates, while they tend to underestimate and overestimate the overtopping rates for relatively large and small overtopping rates, respectively. Goda's (2009) formulas also show this tendency especially for inclined seawalls. The degree of scattering and the tendency of underestimation or overestimation are not distinguishable between training and test data.

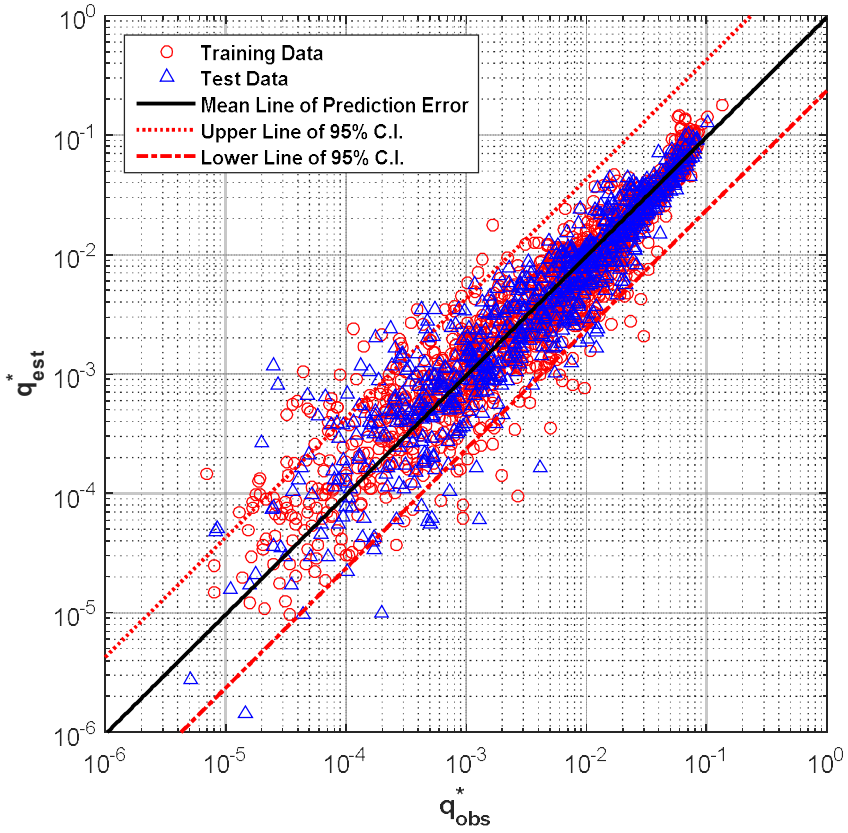


Figure 4.2 Comparison of $q*$ between observation and estimation for training data and test data

Figure 4.3 shows a similar scatter diagram as Figure 4.2 for vertical walls, inclined seawalls, and new EurOtop Manual data of inclined seawalls. Again, the degree of scattering and the tendency of underestimation or overestimation are not distinguishable between different structure types, except that slightly less scattering is observed for the new EurOtop Manual data. The reduced scattering may be because the data were produced by a small number of datasets (i.e. small number of researchers) with consistency (see Table 3.4).

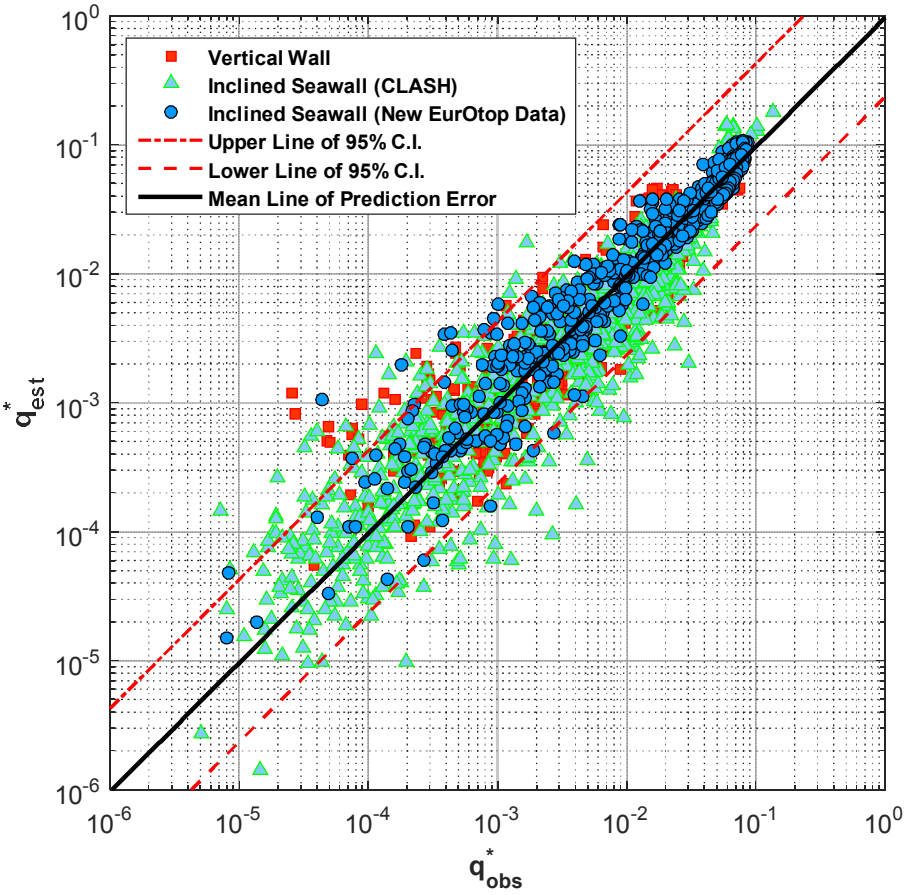


Figure 4.3 Comparison of $q*$ between observation and estimation for vertical walls and inclined seawalls and new Eurotop Manual data for inclined walls

In order to examine the tendencies of over- or under-estimation in more detail, the ratio of the predicted to the observed overtopping rates, q^*_{est} / q^*_{obs}, versus the observed dimensionless overtopping rate q^*_{obs} is plotted in Figure 4.4. Also drawn are the lines of 95% prediction error band, between which 95% of total data are located. Note that the lines of prediction error band are different from the lines of confidence interval. The lines of prediction error band are not symmetric because they were calculated separately for the upper and lower sides, respectively, of the center line. On the whole, the derived formulas tend to underestimate and overestimate the overtopping rates for relatively large and small overtopping rates, respectively. For very large overtopping rates, however, the overtopping rates are slightly underestimated and overestimated for vertical walls and inclined seawalls, respectively. In Figure 4.4, the range of 95% prediction error can be expressed as follows.

$$1.0 \times {q^*_{obs}}^{0.28} < \frac{q^*_{est}}{q^*_{obs}} < 1.0 \times {q^*_{obs}}^{-0.22} \quad : \quad q^*_{obs} < 1.0 \tag{4.2}$$

When the wave overtopping rate is calculated by the GMDH formulas, we can approximately estimate the range in which 95% of the predicted overtopping rates are located by using the observed value in the preceding equation. For example, 95% of the predicted overtopping rates are to be located in the range between 0.525 and 1.66 times the observed rate when $q^*_{obs} = 0.1$, whereas they will be located in the range between 0.275 and 2.754 times the observed rate when $q^*_{obs} = 0.01$.

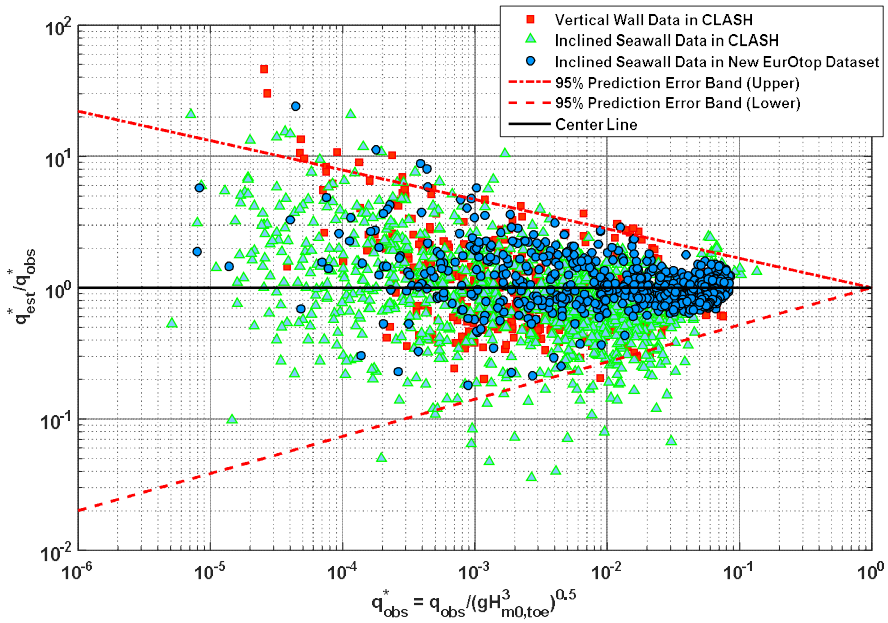


Figure 4.4 Ratio of estimated to observed wave overtopping rates versus estimated dimensionless overtopping rate. The dash-dot and dashed lines indicate 95% prediction error band

4.1.4 Comparison of unified GMDH formulas with other overtopping formulas

Tables 4.1 to 4.3 compare the performance of the GMDH formulas against previous empirical formulas in terms of several statistical parameters defined as follows:

Normalized root-mean-square error (nRMSE): $$\mathrm{nRMSE} = \sqrt{\frac{\frac{1}{N}\sum (q^*_{\mathrm{obs}} - q^*_{\mathrm{est}})^2}{\sum (q^*_{\mathrm{obs}})^2}} \quad (4.3)$$

Nash-Sutcliffe efficiency (NSE): $$\mathrm{NSE} = 1 - \frac{\sum (q^*_{\mathrm{obs}} - q^*_{\mathrm{est}})^2}{\sum (q^*_{\mathrm{obs}} - \overline{q}^*_{\mathrm{meas}})^2} \quad (4.4)$$

Index of agreement: $$I_a = 1 - \frac{\sum (q^*_{\mathrm{obs}} - q^*_{\mathrm{est}})^2}{\sum \left(\left| q^*_{\mathrm{est}} - \overline{q}^*_{\mathrm{obs}} \right| + \left| q^*_{\mathrm{meas}} - \overline{q}^*_{\mathrm{obs}} \right| \right)^2} \quad (4.5)$$

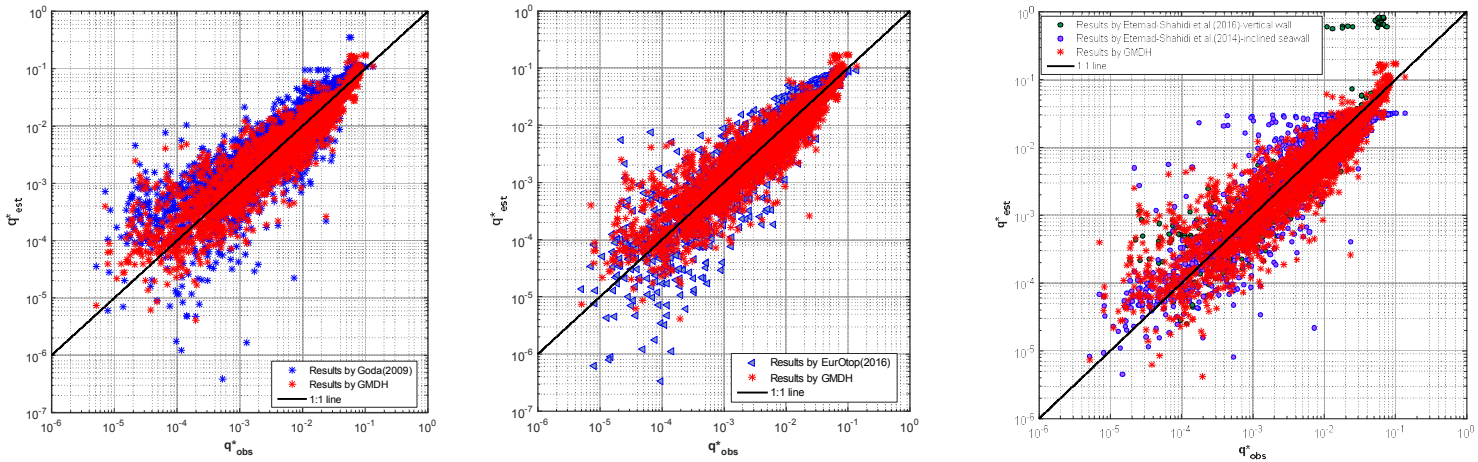
Among the above parameters, nRMSE is obtained by dividing the root-mean-square error by the square root of the sum of the squares of the measured values, which facilitates the comparison between datasets with different scales. The Nash-Sutcliffe efficiency and index of agreement were proposed by Nash and Sutcliffe (1970) and Willmott (1981), respectively, as a measure of the degree to which a model's predictions are error free but not a measure of correlation or association between the observed and predicted variates. The smaller the nRMSE, and the closer to unity the NSE and I_a, the better the performance of the formulas. In the

tables, the bold face indicates the best performance for each data group, whereas the shade indicates the best performance for the total data. For all the statistical parameters, the EurOtop (2016) formulas show the best performance for vertical walls and CLASH total data, the Etemad-Shahidi et al.'s (2014) formulas do for inclined seawalls, whereas the GMDH formulas show the best performance for the new EurOtop data. For the total data, the EurOtop (2016) formulas and the GMDH formulas are shown to equally perform best, even though slightly different performance is evaluated depending on the statistical parameters.

It is natural that the GMDH formulas outperform other formulas for the new EurOtop data because the new data were used in the derivation of the GMDH formulas. On the other hand, it is strange that the performance of the Goda's (2009) formulas is better than the EurOtop (2016) formulas for the new EurOtop data. Note that the new EurOtop data were used in the derivation of the EurOtop formulas, but they were not used in the derivation of the Goda's formulas. A distinct characteristic of the new EurOtop data is that for most of them (588 out of 606) the surf similarity parameter is larger than 2.0. For the CLASH inclined seawall data, the surf similarity parameter is larger than 2.0 for 555 out of 1047. Therefore, it can be inferred that the Goda's formulas are more accurate for the data of larger surf similarity parameter. Without showing the results, we just mention that for the CLASH inclined seawall data, the accuracy of the EurOtop formulas is constant regardless of the surf similarity parameter, whereas the Goda's formulas are more accurate for the data of surf similarity parameter larger than 2.0. These results may

be because the Goda's formulas do not include the effect of wave period, which is taken into account in the EurOtop formulas by the surf similarity parameter.

Figure 4.5 shows graphical comparison of performance between the GMDH formulas and other formulas for the total data. Just as the same as that described above based on Tables 4.1 to 4.3, the performance of the GMDH formulas is almost equal to that of the EurOtop (2016) formulas, but the GMDH formulas somewhat perform better than the Goda (2009) and Etemad-Shahidi (2014, 2016) formulas. Especially, the Etemad-Shahidi et al.'s (2016) formulas significantly over-predict the overtopping rates at vertical walls when the overtopping rates are large (see Figure 4.5(c)). It was found that the crest freeboards were zero for all these data. On the other hand, the Etemad-Shahidi and Jafari's (2014) formulas for inclined seawalls predict almost constant $q*$ between about 0.02 and 0.03 for many values of the observed $q*$ between 0.0001 and 0.1, forming a horizontal band of data (see Figure 4.5(c)). Careful examination of the formulas shows that they give the value of $q*$ at most 0.032. This is also shown in Fig. 3 in the paper of Etemad-Shahidi and Jafari (2014). These problems of the Etemad-Shahidi's (2014, 2016) formulas seem to happen because the data used in the derivation of the formulas do not include the cases of large overtopping rates at low-crested structures. In fact, the dimensionless crest freeboard, h_c / H_{m0}, was greater than 0.082 and 0.43 for vertical and inclined seawalls, respectively, in their papers. Therefore, the Etemad-Shahidi's (2014, 2016) formulas should not be used for low-crested structures.



(a) (b) (c)

Figure 4.5 Comparison of GMDH formulas with other formulas: (a) Goda (2009) formulas; (b) EurOtop (2016) formulas; (c) Etemad-Shahidi (2014, 2016) formulas

Table 4.1 Normalized RMSE of various formulas

nRMSE	Goda (2009)	EurOtop (2016)	Etemad-Shahidi (2014, 2016)	GMDH
Vertical wall	0.0077	**0.0055**	0.0070	0.0058
Inclined seawall	0.0056	0.0038	**0.0035**	0.0040
CLASH Total	0.0046	**0.0031**	0.0032	0.0034
New EurOtop data	0.0045	0.0053	0.0070	**0.0041**
Total	0.0036	**0.0027**	0.0030	**0.0027**

Table 4.2 Nash-Sutcliffe efficiency of various formulas

NSE	Goda (2009)	EurOtop (2016)	Etemad-Shahidi (2014, 2016)	GMDH
Vertical wall	0.642	**0.818**	0.702	0.799
Inclined seawall	0.678	0.852	**0.871**	0.835
CLASH Total	0.673	**0.845**	0.837	0.828
New EurOtop data	0.908	0.873	0.778	**0.923**
Total	0.761	0.866	0.840	**0.867**

Table 4.3 Index of agreement of various formulas

I_a	Goda (2009)	EurOtop (2016)	Etemad-Shahidi (2014, 2016)	GMDH
Vertical wall	0.921	**0.950**	0.928	0.943
Inclined seawall	0.896	0.962	**0.965**	0.954
CLASH Total	0.904	**0.960**	0.958	0.952
New EurOtop data	0.976	0.970	0.932	**0.979**
Total	0.934	**0.967**	0.958	0.964

4.1.5 Sensitivity analysis for unified GMDH formulas

A sensitivity analysis was carried out to investigate the effects of input variables on the wave overtopping rates calculated by the GMDH formulas. It is well known that the crest freeboard is the most influential variable for wave overtopping rates. The seabed slope was eliminated from the GMDH formulas because its effect was negligible. In the following, the dimensionless overtopping rates are calculated as a function of dimensionless crest freeboard for several different values of the remaining input variables. To measure the sensitivity of the output to an input variable, other input variables should be fixed at their baseline values, e.g. mean values. However, since the actual experimental input values are used to calculate the overtopping rates, the values within $\mu \pm 0.2\sigma$ are used as the baseline value, where μ and σ are the mean and standard deviation, respectively, of the fixed input variable.

Figure 4.6 shows $q*$ versus h_c / H_{m0} calculated by the GMDH formulas for different values of h_t / H_{m0} at vertical walls. The mean and standard deviation of $T_{m-1,0} / \sqrt{H_{m0}}$ were 5.71 and 2.69, respectively, and the number of data within $\mu \pm 0.2\sigma$ was 77. The data were grouped into three groups depending on the values of h_t / H_{m0} as shown in the legend of the figure. No distinguishable difference in $q*$ is observed depending on h_t / H_{m0}.

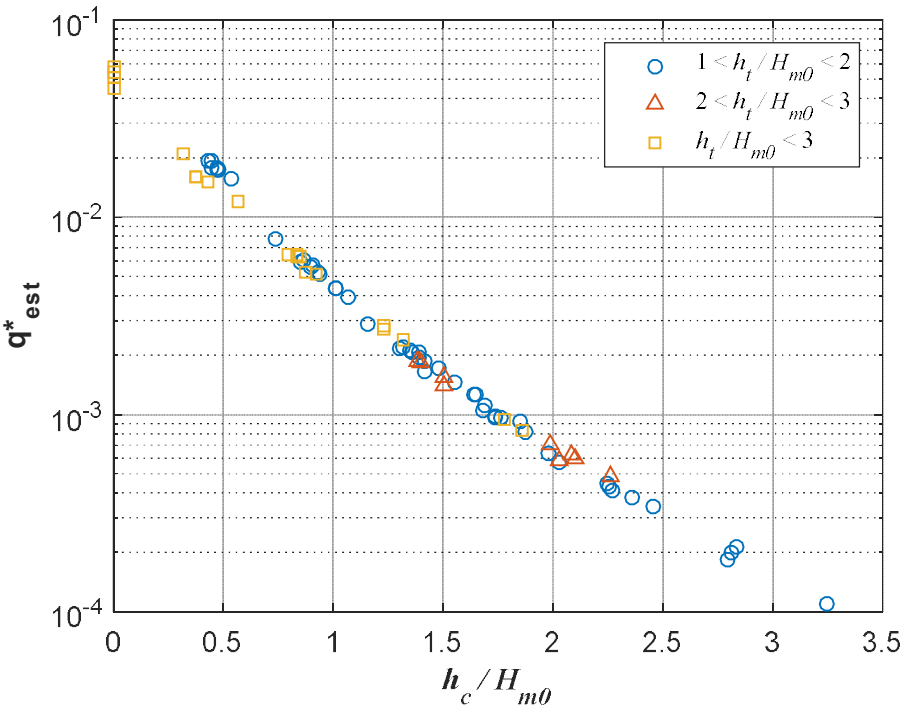


Figure 4.6 Sensitivity analysis of GMDH formulas for toe depth at vertical walls

Figure 4.7 shows $q*$ versus h_c / H_{m0} calculated by the GMDH formulas for different values of $T_{m-1,0} / \sqrt{H_{m0}}$ at vertical walls. The mean and standard deviation of h_t / H_{m0} were 3.14 and 2.85, respectively, and the number of data within $\mu \pm 0.2\sigma$ was 80. The data were grouped into three groups depending on the values of $T_{m-1,0} / \sqrt{H_{m0}}$ as shown in the legend of the figure. The wave overtopping rates increase with the wave period, showing two or three times difference at most.

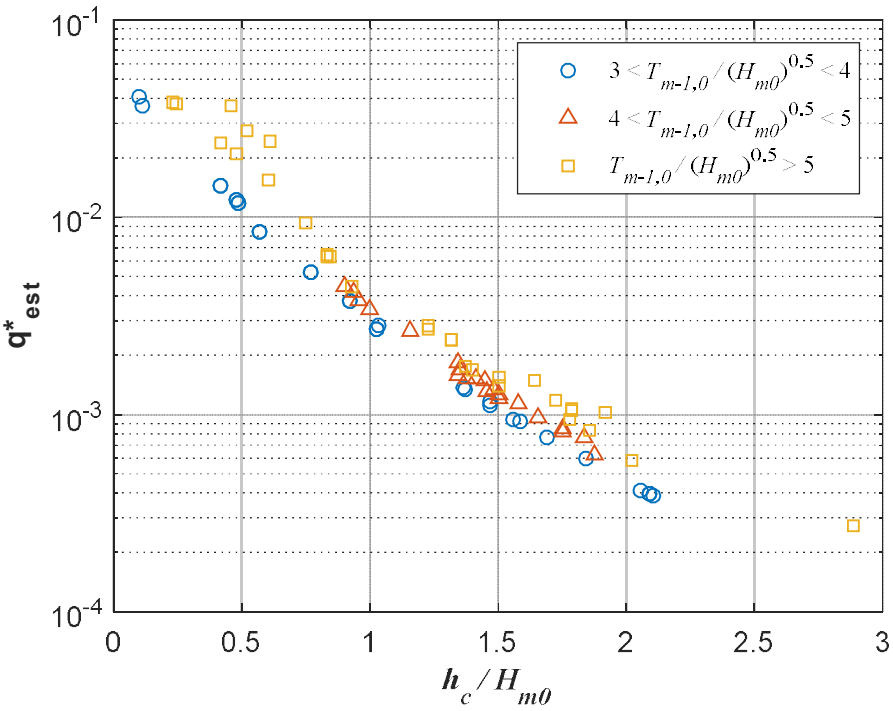


Figure 4.7 Sensitivity analysis of GMDH formulas for wave period at vertical walls

Figure 4.8 shows $q*$ versus h_c / H_{m0} calculated by the GMDH formulas for different values of h_t / H_{m0} at inclined seawalls. The data of $\cot\alpha = 2.0$ were used, the number of which is the largest as 369 among the total 1,047 data. The mean and standard deviation of $T_{m-1,0} / \sqrt{H_{m0}}$ were 4.9 and 1.33, respectively, and the number of data within $\mu \pm 0.2\sigma$ was 61. As in the vertical walls, no difference is observed in $q*$ depending on h_t / H_{m0}.

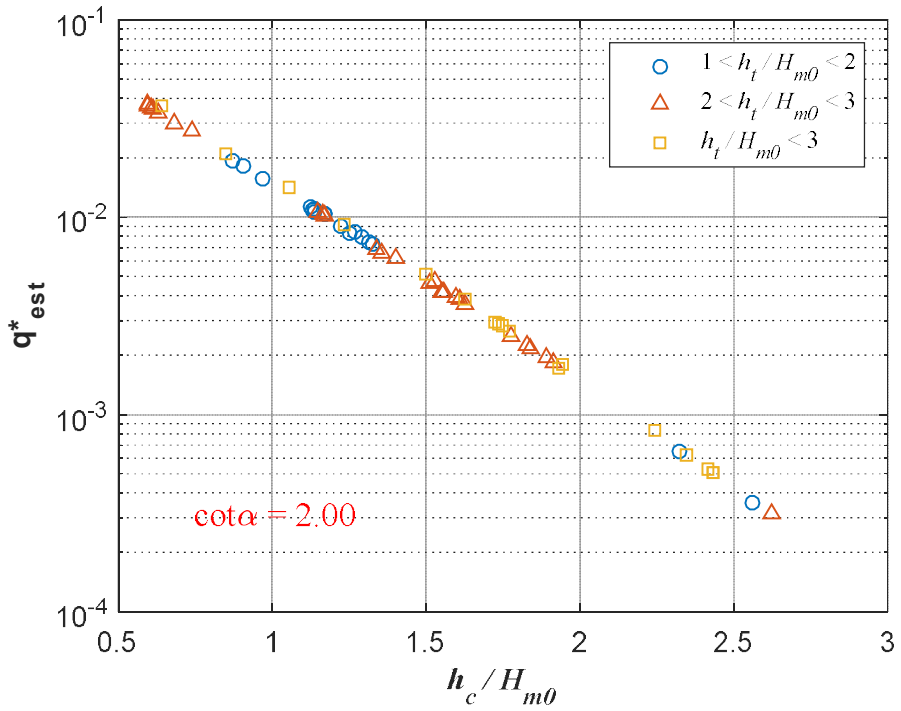


Figure 4.8 Sensitivity analysis of GMDH formulas for toe depth at inclined seawalls

Figure 4.9 shows $q*$ versus h_c / H_{m0} calculated by the GMDH formulas for different values of $T_{m-1,0} / \sqrt{H_{m0}}$ at inclined seawalls. Again the data of $\cot\alpha = 2.0$ were used. The mean and standard deviation of $T_{m-1,0} / \sqrt{H_{m0}}$ were 3.29 and 1.36, respectively, and the number of data within $\mu \pm 0.2\sigma$ was 60. As in the vertical walls, the wave overtopping rates tend to increase with the wave period, especially at high-crested structures.

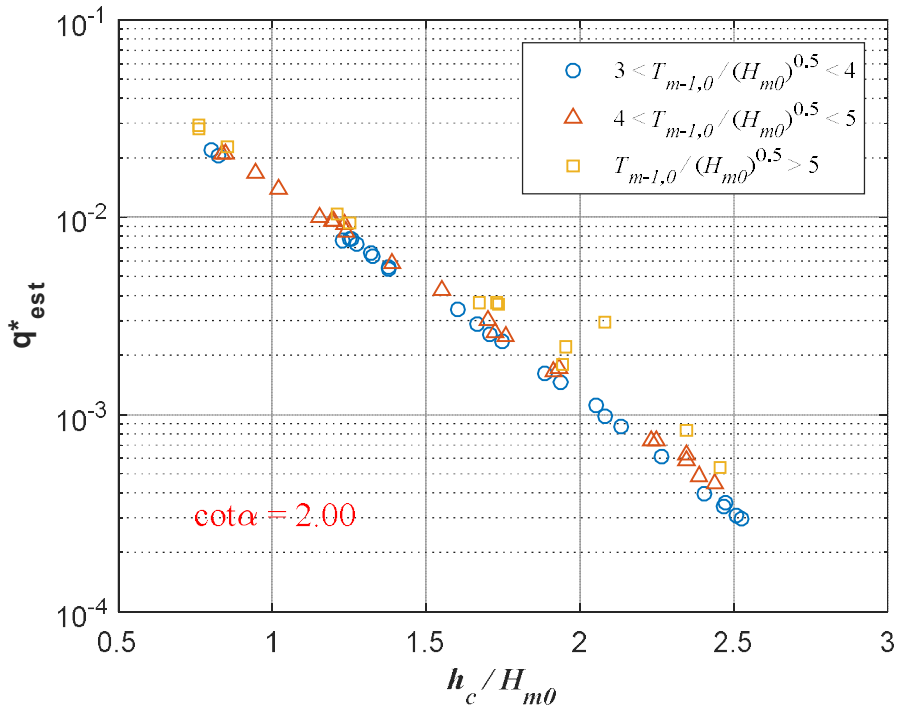


Figure 4.9 Sensitivity analysis of GMDH formulas for wave period at inclined seawalls

Figure 4.10 shows q^* versus h_c / H_{m0} calculated by the GMDH formulas for different values of $\tan\alpha$. The new EurOtop dataset_043 was used, for which the seabed slope is constant as 0.001 and the data are concentrated on low-crested structures where the effect of wave period is negligible. Even though the effects of wave period and toe depth on wave overtopping are not significant, the ranges of $T_{m-1,0} / \sqrt{H_{m0}}$ and h_t / H_{m0} were almost same for different values of $\tan\alpha$. The left panel in Figure 4.10 covers the entire range of h_c / H_{m0}, showing clear dependency of q^* on $\tan\alpha$ for small values of h_c / H_{m0}. The right panel shows a close-up of the part of small h_c / H_{m0} and the regression lines for the data of different $\tan\alpha$. As expected, the wave overtopping rate increases with decreasing structure slope.

The sensitivity analysis shows that the wave overtopping at inclined seawalls is most sensitive to the crest freeboard, second to the structure slope, and third to the wave period. It is shown that the wave overtopping is little influenced by the toe depth. A similar result is shown for vertical walls which have constant (i.e. vertical) structure slopes. However, the GMDH formulas derived by excluding the toe depth showed a little worse performance than those including the toe depth especially for vertical walls (The results are not shown here). Therefore, we recommend to use the GMDH formulas including the toe depth, although its effect on wave overtopping is shown to be insignificant by the sensitivity analysis.

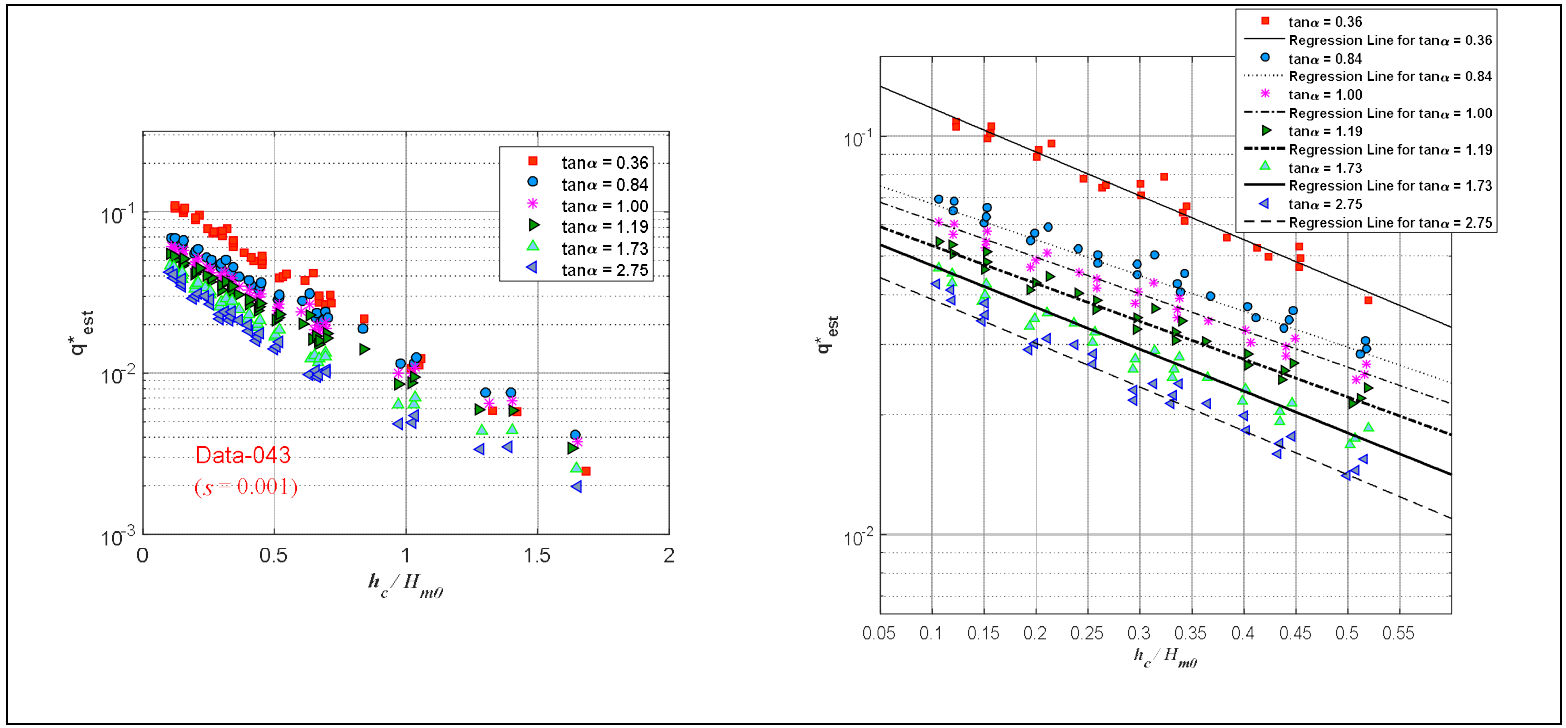


Figure 4.10 Sensitivity analysis of GMDH formulas for structure slope of inclined seawalls

4.2 Separate formulas for overtopping rate derived by GMDH algorithm

4.2.1 GMDH-based overtopping rate prediction formulas for smooth-slope, normal-incidence wave

We have been considering two types of structures simultaneously for unified formulas. However, for considering additional factor effectively which could affect wave overtopping, derivation of formula for vertical wall and inclined seawall, respectively may be needed. Thus, GMDH-based overtopping rate prediction formula for vertical wall and inclined seawall was constructed, respectively, under the smooth-slope, normal-incidence wave conditions first. Procedure of formula derivation, input and target variables, and dataset for GMDH-based formula are the same as previous unified case. Thus, types of variables are the same as Table 3.1 and the dataset are the same as Table 3.2 for vertical walls, and Table 3.3 and Table 3.4 for inclined seawalls. Results of bootstrap sampling, distribution of estimation error, and scatter plots for performance of formulas are included in Appendix A. Following formulas show the GMDH-based formula for vertical walls and inclined seawalls, respectively.

GMDH-based formulas for simple vertical walls

$$\log_{10} q^{*} = \log_{10}\left(\frac{q}{\sqrt{gH_{m0}^{3}}}\right) = 0.093 + 0.791a + 0.245b - 0.79a^{2} - 1.035b^{2} + 1.832ab \tag{4.6a}$$

$$a = 2.105 + 1.685A + 1.902B - 0.008A^{2} + 0.438B^{2} + 0.259AB \tag{4.6b}$$

$$b = 38.494 + 1.39C + 32.23D + 0.026C^{2} + 6.764D^{2} + 0.069CD \tag{4.6c}$$

$$A = -1.180 - 1.454\frac{h_c}{H_{m0}} + 0.2147\left(\frac{h_c}{H_{m0}}\right)^{2} \tag{4.6d}$$

$$B = -2.989 + 0.0505\frac{T_{m-1,0}}{\sqrt{H_{m0}}} + 0.070\frac{h_t}{H_{m0}} - 0.0017\left(\frac{T_{m-1,0}}{\sqrt{H_{m0}}}\right)^{2} - 0.0049\left(\frac{h_t}{H_{m0}}\right)^{2} + 0.007\frac{T_{m-1,0}h_t}{\sqrt{H_{m0}^{3}}} \tag{4.6e}$$

$$C = -1.358 + 0.0505\frac{T_{m-1,0}}{\sqrt{H_{m0}}} - 1.621\frac{h_c}{H_{m0}} - 0.003\left(\frac{T_{m-1,0}}{\sqrt{H_{m0}}}\right)^{2} + 0.199\left(\frac{h_c}{H_{m0}}\right)^{2} + 0.0354\frac{T_{m-1,0}h_c}{\sqrt{H_{m0}^{3}}} \tag{4.6f}$$

$$D = -2.783 + 0.115\frac{h_t}{H_{m0}} - 0.0054\left(\frac{h_t}{H_{m0}}\right)^{2} \tag{4.6g}$$

*units: $T_{m-1,0}$(s), h_c(m), h_t(m), H_{m0}(m)

GMDH-based formulas for inclined seawalls with smooth slope, normal-incidence conditions

$$\log_{10} q^{*} = \log_{10}\left(\frac{q}{\sqrt{gH_{m0}^{3}}}\right) = -0.176 - 0.872a + 1.656b - 3.902a^{2} - 3.342b^{2} + 7.21ab \tag{4.7a}$$

$$a = -4.944 + 0.187A - 4.096B - 0.043A^{2} - 0.876B^{2} - 0.239AB \tag{4.7b}$$

$$b = -0.716 + 0.910C - 0.583A + 0.663C^{2} + 0.589A^{2} - 1.402AC \tag{4.7c}$$

$$A = -1.261 + 0.068\frac{T_{m-1,0}}{\sqrt{H_{m0}}} - 1.245\frac{h_c}{H_{m0}} - 7.97\left(\frac{T_{m-1,0}}{100\sqrt{H_{m0}}}\right)^{2} + 0.059\left(\frac{h_c}{H_{m0}}\right)^{2} + 0.005\frac{T_{m-1,0}h_c}{\sqrt{H_{m0}^{3}}} \tag{4.7d}$$

$$B = -1.149 - 0.019\frac{T_{m-1,0}}{\sqrt{H_{m0}}} - 0.71\cot\alpha + 1.163\left(\frac{T_{m-1,0}}{100\sqrt{H_{m0}}}\right)^{2} + 0.076\cot^{2}\alpha + 0.002\frac{T_{m-1,0}\cot\alpha}{\sqrt{H_{m0}}} \tag{4.7e}$$

$$C = -0.903 + 0.1224\cot\alpha - 1.303\frac{h_c}{H_{m0}} - 0.033\cot^{2}\alpha + 0.12\left(\frac{h_c}{H_{m0}}\right)^{2} - 0.011\frac{h_c\cot\alpha}{H_{m0}} \tag{4.7f}$$

*units: $T_{m-1,0}$(s), h_c(m), H_{m0}(m)

As can be seen by above formulas, the number of variables which are needed at formula for vertical walls and inclined seawalls were decreased. Only three variables (Vertical walls: relative wave period, relative crest freeboard, relative toe depth; Inclined seawalls: relative wave period, relative crest freeboard, structure slope) are selected for wave overtopping rate prediction at each structures. Thus, during the process of application of GMDH algorithm for derivation of formulas with five input variables as shown in Table 3.1, two input variables at each structures (vertical walls: slope of structures, seabed slope; inclined seawalls: seabed slope, relative toe depth of the structures) were excluded.

Although it is not a unified formula, each case has the advantage of being able to estimate the overtopping rate with fewer variables in each case. Meanwhile, performance of this prediction formula is nearly same as previous empirical formula by EurOtop (2016), Etemad-Shahidi and Jafari (2014), and Etemad-Shahidi *et al*. (2016). Tables 4.4 to 4.7 shows the summary of performance values.

Table 4.4 Normalized RMSE of mean overtopping rate prediction model for inclined seawalls

nRMSE	Goda(2009)	EurOtop(2016)	Etemad-Shahidi (2014)	GMDH(Inclined)
Inclined Seawalls (CLASH)	0.0056	0.0038	**0.0035**	0.0040
New EurOtop Data	0.0045	0.0053	0.0070	**0.0044**
Total	0.0041	**0.0031**	0.0032	**0.0031**

Table 4.5 Nash-Sutcliffe efficiency of mean overtopping rate prediction model for inclined seawalls

NSE	Goda(2009)	EurOtop(2016)	Etemad-Shahidi (2014)	GMDH(Inclined)
Inclined Seawalls (CLASH)	0.678	0.852	**0.871**	0.834
New EurOtop Data	0.908	0.873	0.778	**0.913**
Total	0.779	**0.876**	0.862	**0.876**

Table 4.6 Index of Agreement of mean overtopping rate prediction model for inclined seawalls

I_a	Goda(2009)	EurOtop(2016)	Etemad-Shahidi (2014)	GMDH(Inclined)
Inclined Seawalls (CLASH)	0.896	0.962	**0.965**	0.952
New EurOtop Data	0.976	0.970	0.932	**0.977**
Total	0.936	**0.970**	0.963	0.966

Table 4.7 Performance comparison of wave overtopping formulas for vertical walls

Vertical walls	Goda(2009)	EurOtop(2016)	Etemad-Shahidi	GMDH(Vertical)
NRMSE	0.0077	0.0055	0.0070	**0.0048**
NSE	0.642	0.818	0.702	**0.862**
I_a	0.921	0.950	0.928	**0.963**

4.2.2 GMDH-based overtopping rate prediction formulas for inclined seawalls considering oblique wave condition and roughness factor

In EurOtop manual (2016), it is possible that oblique wave incidence could affect the wave overtopping. For these reasons, formula by EurOtop manual (2016) consider oblique wave effect by using γ_β in the relative crest freeboard term. In this manual, there are two types of γ_β with respect to long crested waves and short crested waves. Following formula shows the two types of γ_β.

for short-crested waves

$$\gamma_\beta = 1 - 0.0033|\beta| \quad \text{for } 0^\circ \le \beta \le 80^\circ \tag{4.8a}$$
$$\gamma_\beta = 0.736 \quad \text{for } |\beta| > 80^\circ$$

for long-crested waves

$$\gamma_\beta = \cos^2(|\beta| - 10^\circ) \quad \text{with a minimun of } \gamma_\beta = 0.6 \tag{4.8b}$$
$$\gamma_\beta = 1 \quad \text{for } |\beta| = 0^\circ - 10^\circ$$

In this research case, data conditions are long-crested waves because spreading parameter shows zero value. In this case, second formula shows a little more reliable results. Using second formula, we could derive γ_β value and modified relative crest freeboard value which relative crest freeboard is divided by γ_β.

Additionally, we can consider roughness factor which can affect the wave overtopping at inclined seawalls. In EurOtop manual, there are various roughness factor with respect to material which consists of inclined seawalls. Roughness can

dissipate wave energy during wave run-up, so if roughness is high, wave overtopping could be decreased. Thus, roughness factor can be one of the important factor when we estimate wave overtopping rate.

In EurOtop formula, roughness factor can be applied with relative crest freeboard. Relative crest freeboard is divided by roughness factor in EurOtop formula, which method is the same as the application of oblique wave effect. Thus, in this study, modified relative crest freeboard is used as a new variable for considering the effect of roughness of the slope and oblique wave condition as follows.

$$\frac{h_c}{H_{m0}}\text{(previous relative crest freeboard)} \rightarrow \frac{h_c}{H_{m0}\gamma_f\gamma_\beta}\text{(modified relative crest freeboard)}$$

All inclined seawalls dataset with smooth-slope and normal-incidence wave, which are shown in Table 3.3 to 3.4, are used together with the dataset which γ_f and γ_β are smaller than one. Table 4.8 to 4.9 shows the range of dataset which γ_f and γ_β are smaller than one.

Table 4.8 Characteristics of extracted CLASH datasets of inclined seawalls (oblique wave)

CLASH Dataset ID	Number of data	β	$\cot\alpha$	s	$T_{m-1,0}/\sqrt{H_{m0}}$	h_t / H_{m0}	h_c / H_{m0}	$q/\sqrt{gH_{m0}^3}$
Data_030	36	15–40	1.00–4.00	0.001–0.05	3.72–5.96	1.70–3.26	1.06–2.63	1.35×10^{-6}–8.88×10^{-4}
Data_222	109	5–80	2.50–4.00	0.001	3.11–7.62	5.03–6.92	1.10–2.50	1.06×10^{-6}–6.05×10^{-4}
Total	145	5–80	1.00–4.00	0.001–0.05	3.11–7.62	1.70–6.92	1.06–2.63	1.06×10^{-6}–8.88×10^{-4}

Table 4.9 Characteristics of extracted CLASH datasets of inclined seawalls (with roughness factor)

Dataset ID	Number of data	γ_f	$\cot\alpha$	s	$T_{m-1,0}/\sqrt{H_{m0}}$	h_t/H_{m0}	h_c/H_{m0}	$q/\sqrt{gH_{m0}^3}$
B_03	4	0.55	3.0	0.001	3.60–6.64	3.04–4.92	1.12–1.80	1.98×10^{-3}–1.59×10^{-2}
B_09	23	0.55	2.0, 3.5	0.001, 0.01	5.49–11.5	2.34–6.94	1.19–2.55	1.10×10^{-6}–1.31×10^{-4}
C_20	54	0.50	1.33	0.02	3.80–9.99	2.34–3.67	1.05–2.38	9.18×10^{-6}–1.84×10^{-3}
D_11	18	0.80	1.0, 2.0, 4.0	0.01	3.57–5.31	1.58–3.35	1.35–2.87	1.10×10^{-6}–1.31×10^{-4}
D_20	4	0.45	3.0	0.001	3.58–6.70	3.33–5.00	1.10–1.83	3.31×10^{-6}–3.97×10^{-5}
F_16	4	0.58–0.68	4.0	0.01	5.78–6.21	1.24–1.31	1.48–1.54	4.87×10^{-6}–1.52×10^{-5}
F_19	51	0.67–0.78	2.0, 4.0	0.001	3.51–3.78	3.16–7.20	0.55–1.56	5.16×10^{-6}–3.50×10^{-3}
Total	158	0.45–0.80	1.0–4.0	0.001–0.02	3.51–9.99	1.24–7.20	0.55–2.87	1.10×10^{-6}–1.59×10^{-2}

The GMDH-based formula considering oblique wave and roughness for inclined seawalls is as follows.

(GMDH-based formulas for inclined seawalls with oblique wave condition and slope roughness/permeability)

$$\log_{10} q^{*} = \log_{10}\left(\frac{q}{\sqrt{gH_{m0}^{3}}}\right) = -0.465 + 1.946a - 1.436b - 2.856a^{2} - 3.683b^{2} + 6.442ab \tag{4.9a}$$

$$a = -0.922 + 0.853A - 0.748B + 0.641A^{2} + 0.552B^{2} - 1.394AB \tag{4.9b}$$

$$b = -7.251 - 0.61C - 5.213A - 0.167C^{2} - 1.05A^{2} - 0.314AB \tag{4.9c}$$

$$A = -1.007 + 0.189\cot\alpha - 1.27\frac{h_c}{H_{m0}\gamma_f\gamma_\beta} - 0.032\cot^{2}\alpha + 0.175\left(\frac{h_c}{H_{m0}\gamma_f\gamma_\beta}\right)^{2} - 0.079\frac{h_c\cot\alpha}{\gamma_f\gamma_\beta H_{m0}} \tag{4.9d}$$

$$B = -1.17 - 0.082\frac{T_{m-1,0}}{\sqrt{H_{m0}}} - 1.532\frac{h_c}{H_{m0}\gamma_f\gamma_\beta} - 6.076\left(\frac{T_{m-1,0}}{100\sqrt{H_{m0}}}\right)^{2} + 0.19\left(\frac{h_c}{H_{m0}\gamma_f\gamma_\beta}\right)^{2} - 0.005\frac{T_{m-1,0}h_c}{\gamma_f\gamma_\beta\sqrt{H_{m0}^{3}}} \tag{4.9e}$$

$$C = -1.244 - 0.033\frac{T_{m-1,0}}{\sqrt{H_{m0}}} - 0.675\cot\alpha + 2.17\left(\frac{T_{m-1,0}}{100\sqrt{H_{m0}}}\right)^{2} + 0.073\cot^{2}\alpha + 0.004\frac{T_{m-1,0}\cot\alpha}{\sqrt{H_{m0}}} \tag{4.9f}$$

*units: $T_{m-1,0}$(s), h_c(m), H_{m0}(m)

Also, Figure 4.11 shows the plotting results of estimated wave overtopping rate for rough, oblique wave condition data only, and Figure 4.12 shows comparison between results by GMDH-based formula and results by EurOtop (2016) formula. As shown in Figure 4.12, the estimated results by applying the roughness and oblique wave factor to the EurOtop formula have a limitation to estimation accuracy. However, the GMDH-based formula accurately estimates the wave overtopping rate in most cases except for some data.

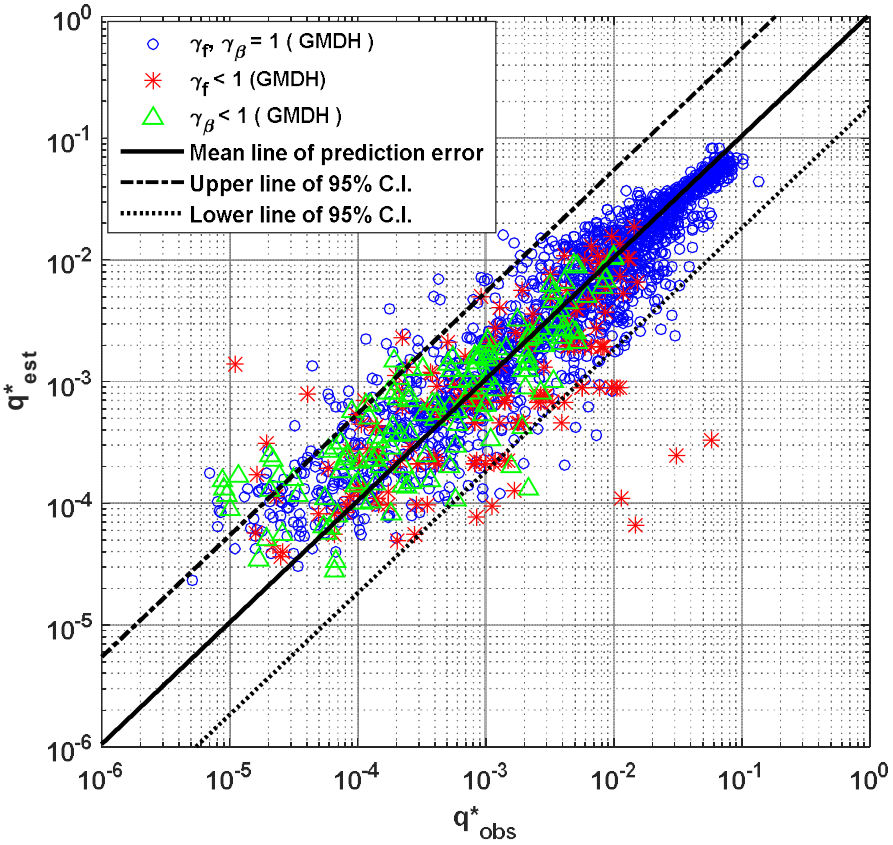


Figure 4.11 Comparison of $q*$ between observation and estimation with different oblique wave effect and roughness at GMDH-based inclined seawalls formula

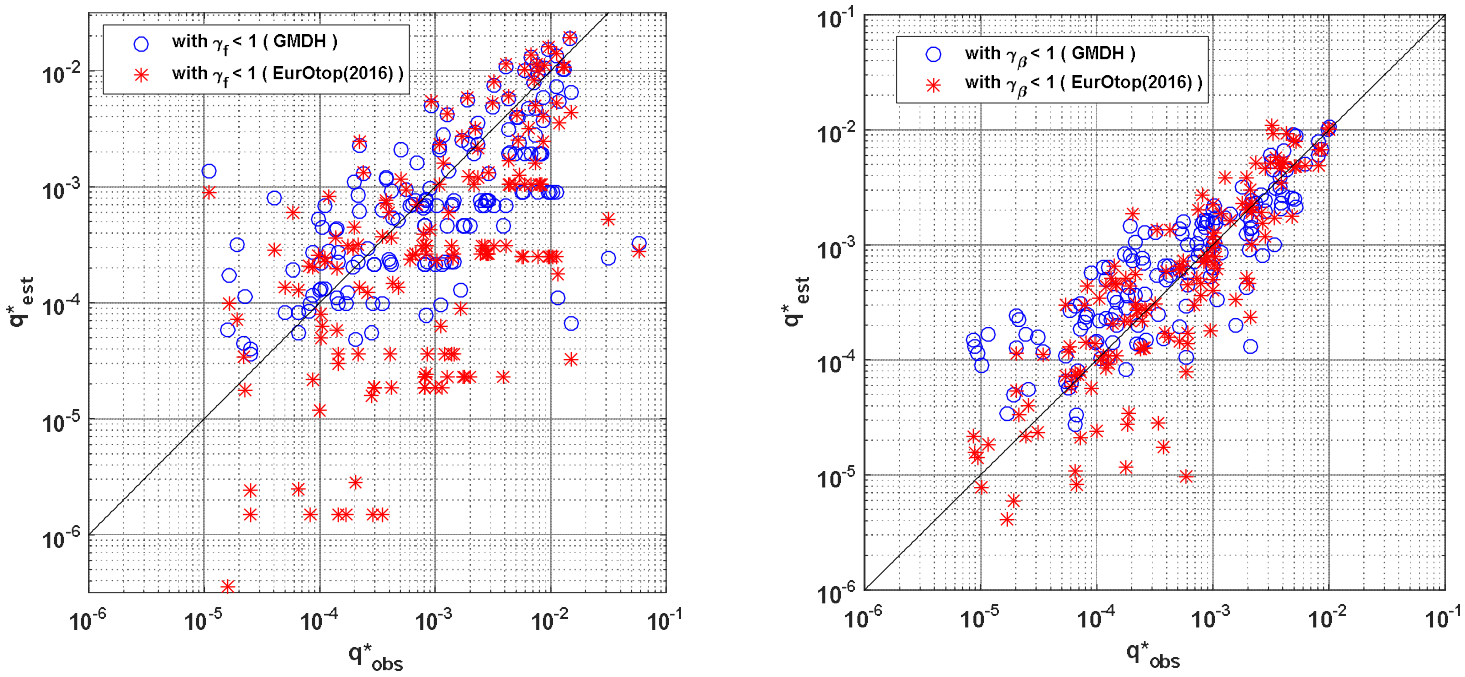


Figure 4.12 Comparison of $q*$ between GMDH-based formula and EurOtop (2016) with (a) roughness factor and (b) oblique wave factor

4.2.3 GMDH-based overtopping rate prediction formulas for vertical walls considering oblique wave condition

Considering oblique wave condition also can be possible in vertical walls case, so influence factor for oblique wave condition was applied at the relative crest freeboard, same as inclined seawalls case. The factor for oblique wave condition which is applied at inclined seawalls case also can be applied at vertical walls case. The number of data with oblique wave condition is 18, the dataset ID is G-19, and the range of variables are shown in Table 4.10. This dataset are long-crested wave condition, so oblique wave factor (4.8b) in 4.2.2 were used in this research.

Table 4.10 Characteristics of datasets (vertical walls with oblique wave)

Variable	β	$T_{m-1,0}/\sqrt{H_{m0}}$	h_t / H_{m0}	h_c / H_{m0}	$q/\sqrt{gH_{m0}^3}$
Range	20, 45	3.49–3.97	7.45–17.5	0.74–2.38	3.13×10^{-5}–7.57×10^{-3}

Using dataset which are shown in Table 4.10 (oblique wave condition) and Table 3.2 (normal-incidence wave), GMDH-based formula for vertical walls consider oblique wave condition can be derived as follows. Also, Figure 4.13 shows the comparison results between GMDH-based formula and EurOtop (2016) at vertical walls case with oblique wave condition.

(GMDH-based formulas for vertical walls with oblique wave condition)

$$\log_{10} q^{*} = \log_{10}\left(\frac{q}{\sqrt{gH_{m0}^{3}}}\right) = -0.372 + 1.169a - 0.465b + 3.477a^{2} + 3.437b^{2} - 6.978ab \tag{4.10a}$$

$$a = 0.279 + 1.110A - 0.006B - 0.045A^{2} - 0.086B^{2} + 0.134AB \tag{4.10b}$$

$$b = 0.080 + 0.916C + 0.281A - 0.013C^{2} + 0.092A^{2} - 0.017AC \tag{4.10c}$$

$$A = -0.665 - 0.168\frac{h_t}{H_{m0}} - 1.688\frac{h_c}{H_{m0}\gamma_\beta} + 0.010\left(\frac{h_t}{H_{m0}}\right)^{2} + 0.252\left(\frac{h_c}{H_{m0}\gamma_\beta}\right)^{2} - 0.013\frac{h_t h_c}{H_{m0}^{2}\gamma_\beta} \tag{4.10d}$$

$$B = -3.884 + 0.260\frac{T_{m-1,0}}{\sqrt{H_{m0}}} - 0.016\frac{h_t}{H_{m0}} - 0.017\left(\frac{T_{m-1,0}}{\sqrt{H_{m0}}}\right)^{2} - 0.008\left(\frac{h_t}{H_{m0}}\right)^{2} + 0.029\frac{T_{m-1,0}h_t}{\sqrt{H_{m0}^{3}}} \tag{4.10e}$$

$$C = -1.429 + 0.047\frac{T_{m-1,0}}{\sqrt{H_{m0}}} - 1.613\frac{h_c}{H_{m0}\gamma_\beta} - 0.002\left(\frac{T_{m-1,0}}{\sqrt{H_{m0}}}\right)^{2} + 0.170\left(\frac{h_c}{H_{m0}\gamma_\beta}\right)^{2} + 0.037\frac{T_{m-1,0}h_c}{\gamma_\beta\sqrt{H_{m0}^{3}}} \tag{4.10f}$$

*units: $T_{m-1,0}$(s), h_c(m), h_t(m), H_{m0}(m)

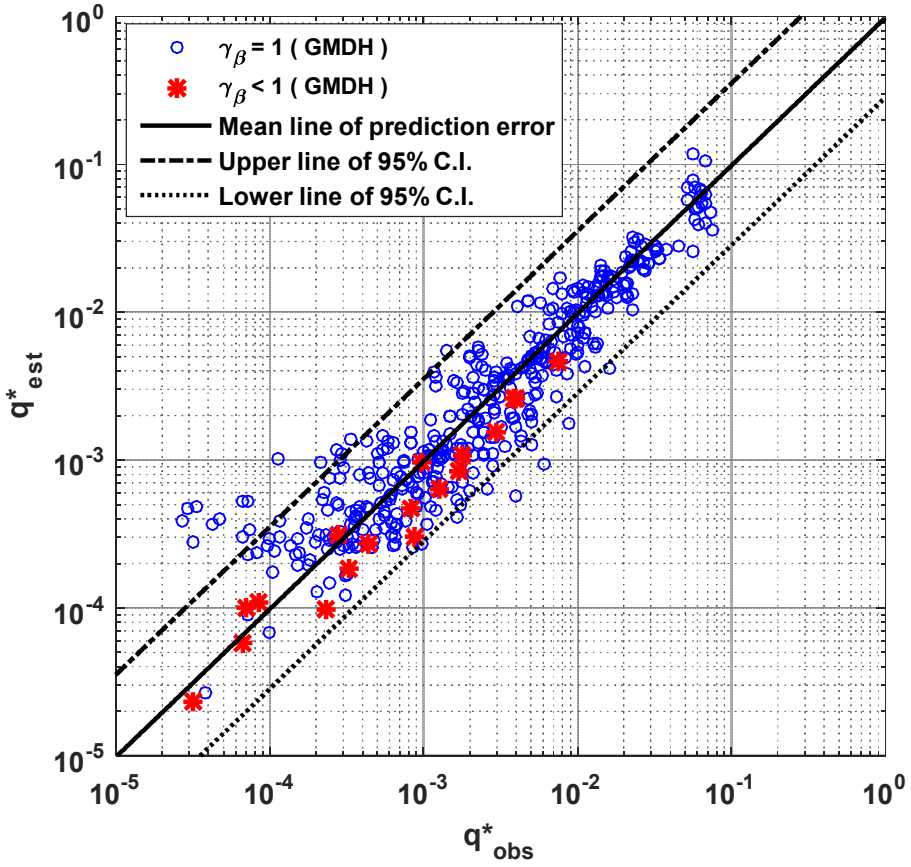


Figure 4.13 Comparison of $q*$ between GMDH-based formula and EurOtop (2016) for vertical walls at different oblique wave factor

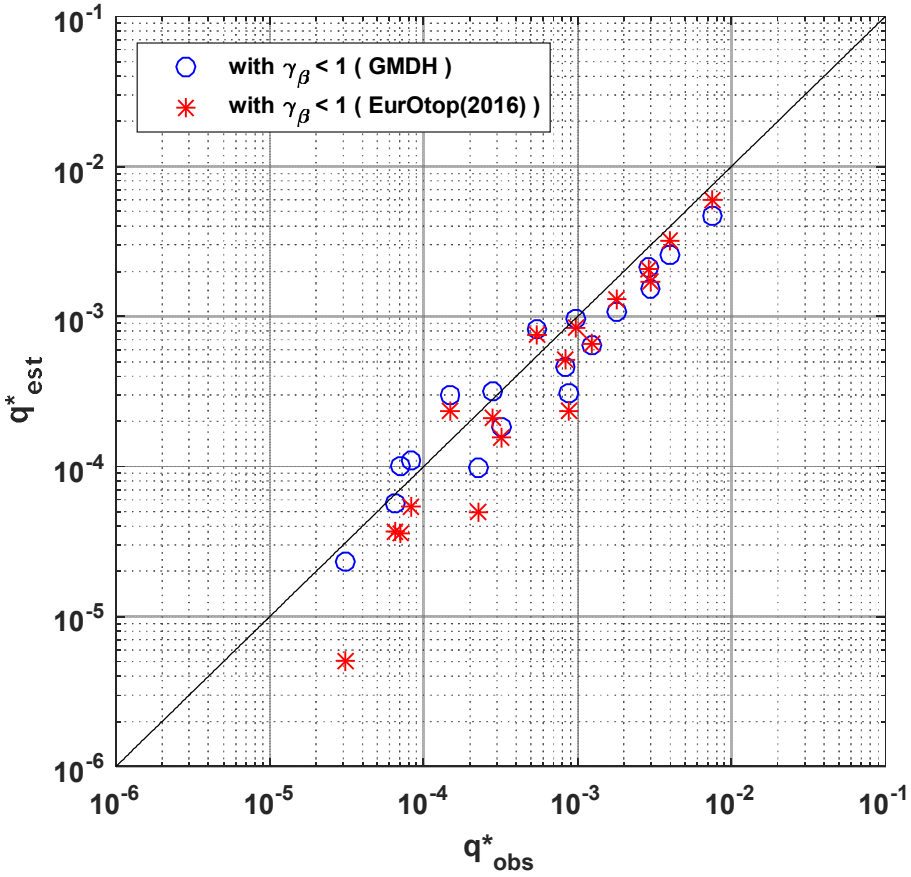


Figure 4.14 Comparison of $q*$ between GMDH-based formula and EurOtop (2016) for vertical walls with oblique wave factor

CHAPTER 5. CONCLUSIONS

In this study, unified formulas to calculate mean wave overtopping rates at vertical and inclined seawalls of smooth uniform slope were derived by applying the GMDH algorithm to the CLASH and new EurOtop datasets. The derived formulas were shown to be more accurate than the previous unified formulas of Goda (2009) with a little increase of complexity of the formulas. The accuracy of the formulas was also shown to be comparable to those of the formulas developed separately for vertical walls and inclined seawalls. The estimation errors of 95% confidence interval are given as $\log q^*_{\mathrm{est}} - \log q^*_{\mathrm{obs}} = -0.645$ and $\log q^*_{\mathrm{est}} - \log q^*_{\mathrm{obs}} = 0.613$, whereas the range of 95% prediction error is given as $1.0 \times {q^*_{\mathrm{obs}}}^{0.28} < q^*_{\mathrm{est}} / q^*_{\mathrm{obs}} < 1.0 \times {q^*_{\mathrm{obs}}}^{-0.22}$ where $q^*_{\mathrm{obs}} < 1.0$.

The crest freeboard is the most important parameter governing the wave overtopping rate. The structure slope is also important especially at low-crested structures, leading to more wave overtopping on milder slopes. The wave period also has some effects at high-crested structures, resulting in slight increase in overtopping rate for longer wave periods. The effect of toe depth on wave overtopping was shown to be insignificant by the sensitivity analysis, but the toe depth was included in the formulas because the performance of the formulas was somewhat deteriorated when it was excluded.

Meanwhile, when the formula was formed for each structure, the required variable was reduced from four to three. It is also possible to establish a GMDH-based formula that allows for the possibility of considering various conditions of structure, which can be considered to slope roughness and oblique wave condition, which can be considered more accurate than other formula.

REFERENCES

Etemad-Shahidi, A. and Jafari, E., 2014. New formulae for prediction of wave overtopping at inclined structures with smooth impermeable surface. Ocean Engineering 84, 124-132.

Etemad-Shahidi, A., Shaeri, S. and Jafari, E., 2016. Prediction of wave overtopping at vertical structures. Coastal Engineering 109, 42-52.

EurOtop, 2016. Manual on wave overtopping of sea defences and related Structures. An overtopping manual largely based on European research, but for worldwide application. Second Edition.N.W.H. Allsop, T. Bruce, J. DeRouck, A. Kortenhaus, T. Pullen, H. Schüttrumpf, P. Troch, J.W. van der Meer and B. Zanuttigh. www.overtopping-manual.com

Farlow, S. J. (1984). Self-organizing methods in modeling-GMDH type algorithms, Dekker.

Formentin, S. M., Zanuttigh, B. and van der Meer, J. W. (2017). A neural network tool for predicting wave reflection, overtopping and transmission, Coastal Engineering Journal 59(2), 31pp.

Garg, V. (2015). Inductive group method of data handling neural network approach to model basin sediment yield. Journal of Hydrology Engineering 20. Technical Note.

Goda, Y. (2009). Derivation of unified wave overtopping formulas for seawalls with

smooth, impermeable surfaces based on selected CLASH datasets. Coastal Engineering 56, 385-399.

Goda, Y., Kishira, Y. and Kamiyama, Y., 1975. Laboratory investigation on the overtopping rate of seawalls by irregular waves. Report of Port and Harbour Research Institute 14(4), 3-44 (in Japanese).

Heris, S.M.K., 2015. Implementation of group method of data handling in Matlab. Yarpiz, www.yarpiz.com.

Ivakhnenko, A. G., 1966. Group method of data handling−A rival of the method of stochastic approximation. Soviet Automatic Control 13, 43-71.

Lee, A., Geem, Z.W. and Suh, K.-D., 2016. Determination of optimal initial weights of an artificial neural network by using the harmony search algorithm: Application to breakwater armor stones. Applied Sciences 6, 164; doi:10.3390/app6060164.

Mehra, R. K. (1977). Group method of data handling (GMDH): Review and experience. IEEE Conference.

Nash, J. E. and Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part 1-A discussion of principles. Journal of Hydrology 10, 282-290.

Pullen, T., Allsop, N.W.H., Bruce, T., Kortenhaus, A., Schüttrumph, H. and Van der Meer, J.W., 2007. EurOtop – wave overtopping of sea defences and related structures: assessment manual. www.overtopping-manual.com 2007 Die Kuste, Heft 73.

Shahabi, S., Khanjani, M. and Kermani, M.H. (2016). Significant wave height

forecasting using GMDH model. International Journal of Computer Applications 133, 13-16.

Tsai, T. M., Yen, P. H. and Huang, T. J. (2009). Wave height forecast using self-organization algorithm model. Proceedings of the Nineteenth (2009) International Offshore and Polar Engineering Conference (ISOPE), Osaka, Japan.

Tsai, T.M. and Yen, P. H. (2017). GMDH algorithms applied to turbidity forecasting. Applied Water Science 7, 1151-1160.

U.S. Army, 1984. Coastal Engineering Research Center Shore Protection Manual, 4th ed., U.S. Government Printing Office, Washington, D.C., USA.

Van der Meer, J.W., Verhaeghe, H. and Steendam, G.J., 2005. Database on wave overtopping at coastal structures, CLASH WP2 database. Infram, Marknesse, The Netherlands.

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

Verhaeghe, H. (2005). Neural network prediction of wave overtopping at coastal structures. PhD thesis. University of Gent, Gent, BE.

Willmott, C. J. (1981). On the validation of models. Physical Geography 2, 184-194.

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

APPENDIX A – Results of bootstrap sampling for separate formulas by GMDH

(1) Formula for vertical walls

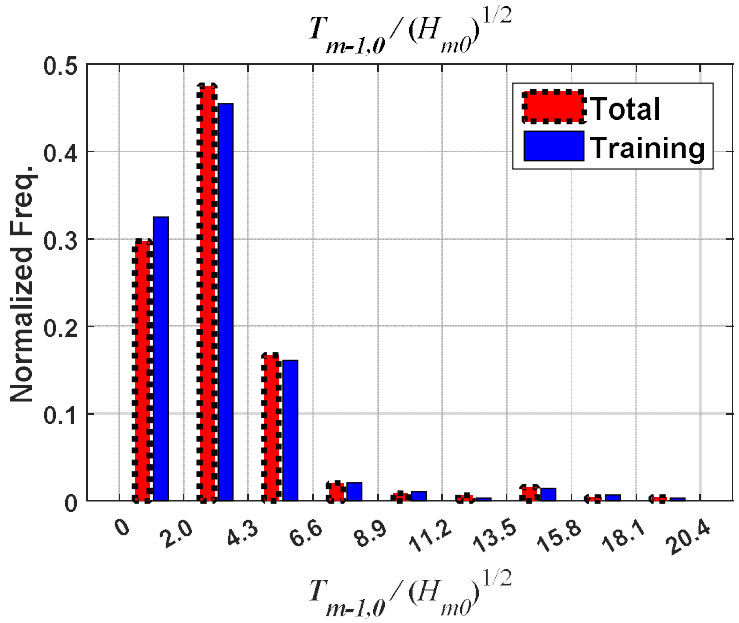


Figure A.1 Results of bootstrap sampling for relative wave period in case of vertical walls

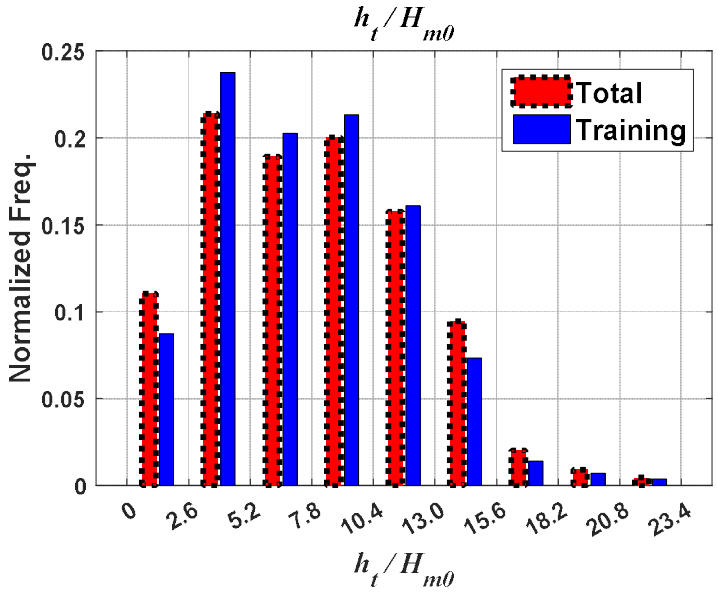


Figure A.2 Results of bootstrap sampling for relative water depth at the toe in case of vertical walls

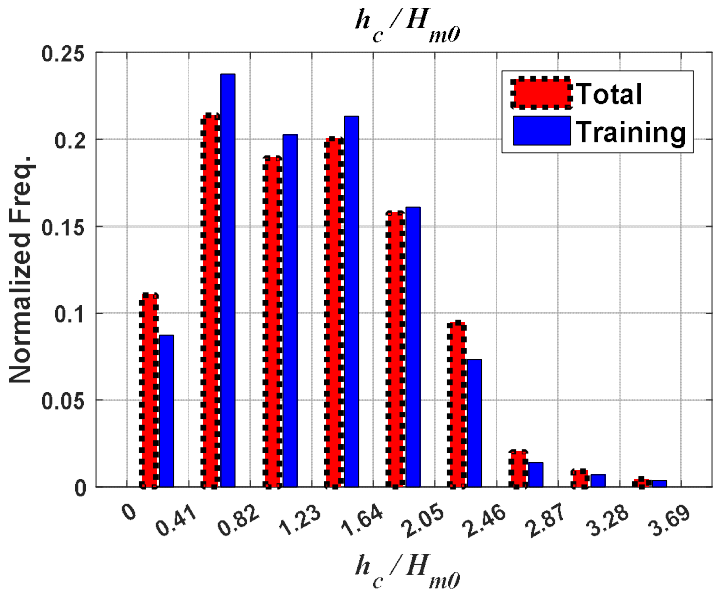


Figure A.3 Results of bootstrap sampling for relative crest freeboard in case of vertical walls

(2) Formula for inclined seawalls

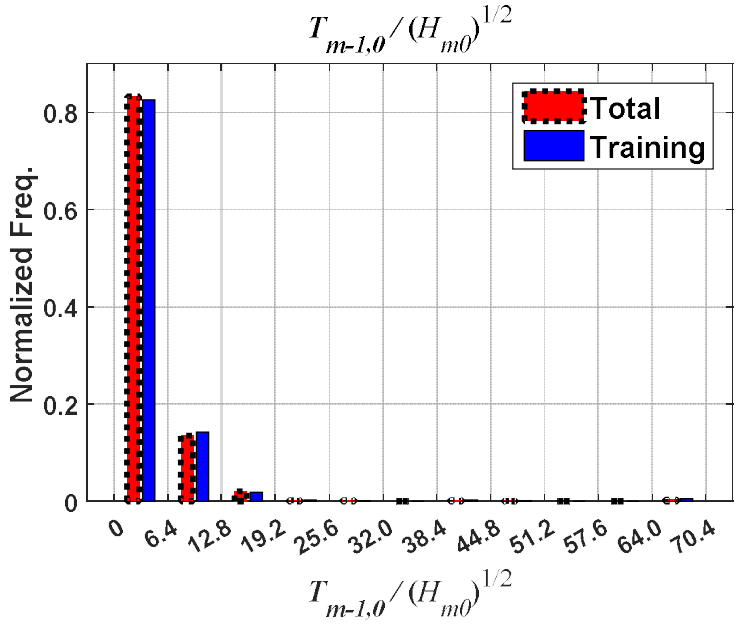


Figure A.4 Results of bootstrap sampling for relative wave period in case of inclined seawalls

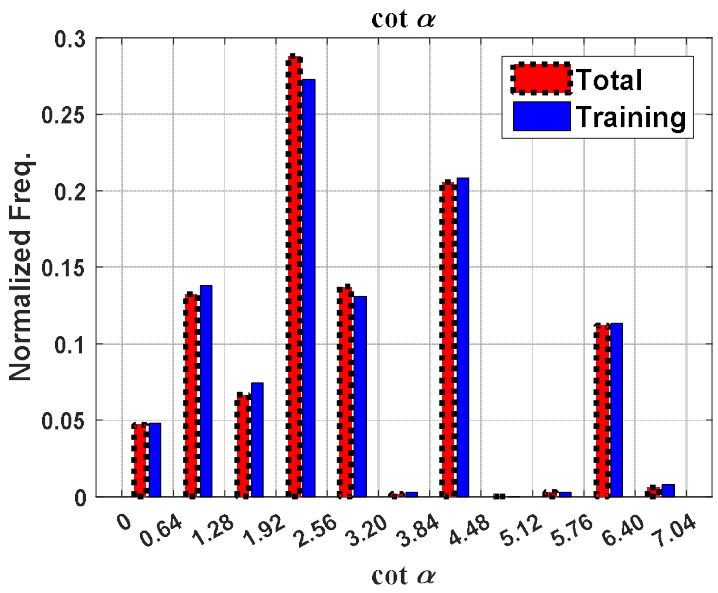


Figure A.5 Results of bootstrap sampling of structure slope in case of inclined seawalls

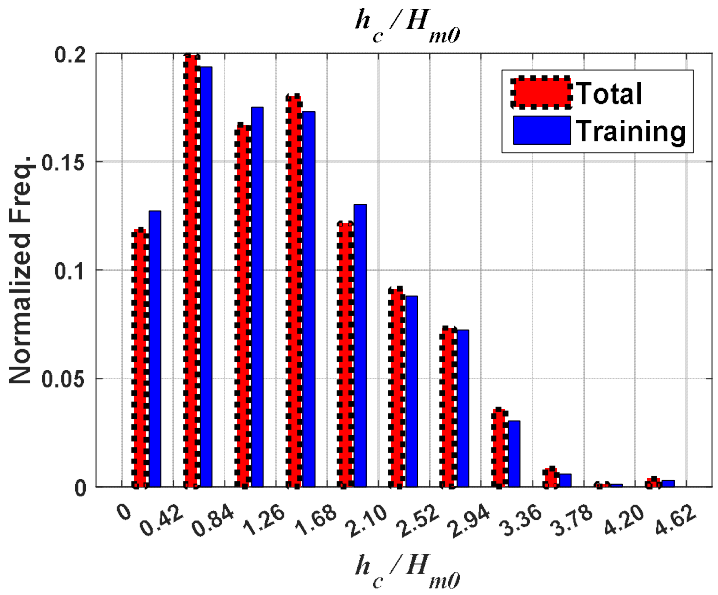


Figure A.6 Results of bootstrap sampling for relative crest freeboard in case of inclined seawalls

APPENDIX B – Distribution of estimation error

(1) Formula for vertical walls

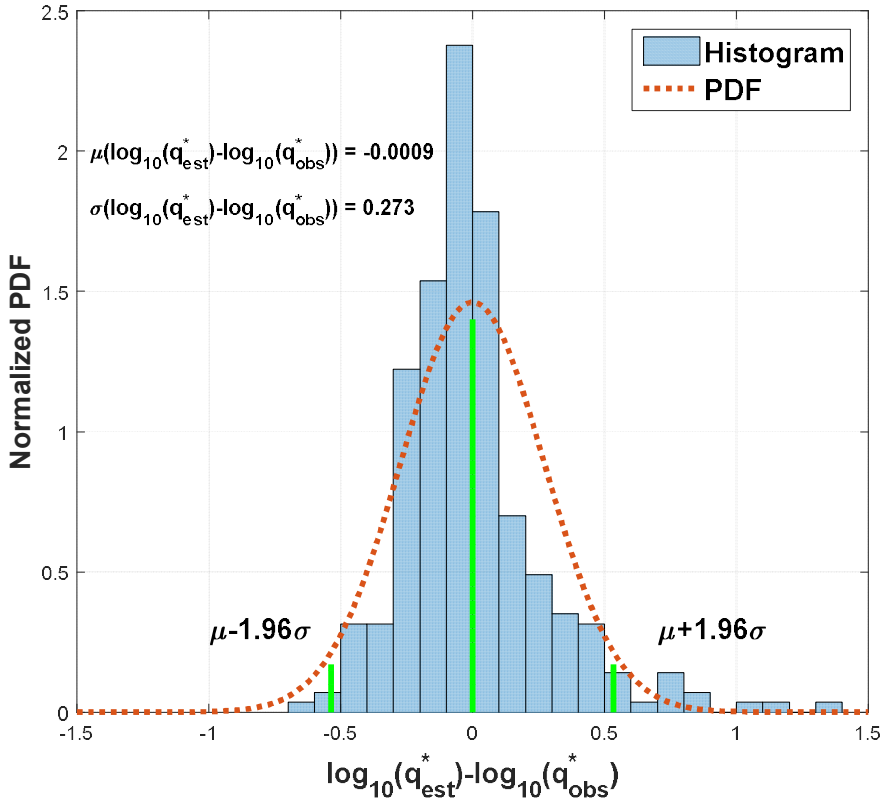


Figure B.1 Distribution of estimation error in case of vertical walls formula

(2) Formula for inclined seawalls

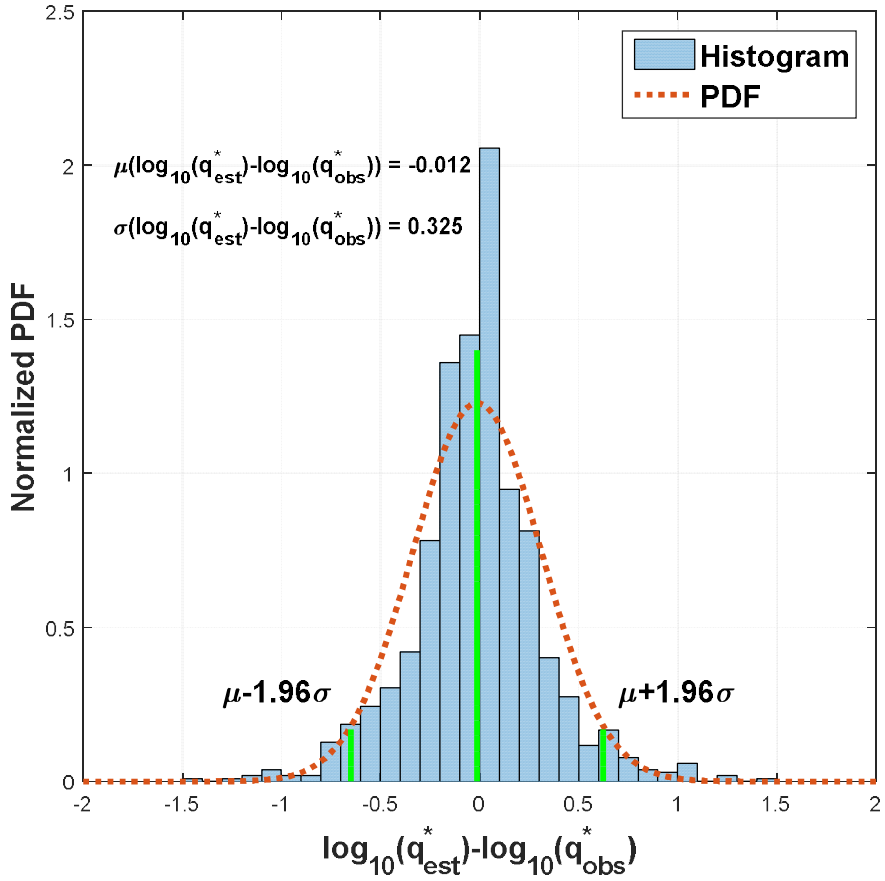


Figure B.2 Distribution of estimation error in case of inclined seawalls formula

APPENDIX C – Performance graph of formulas

(1) Formula for vertical walls

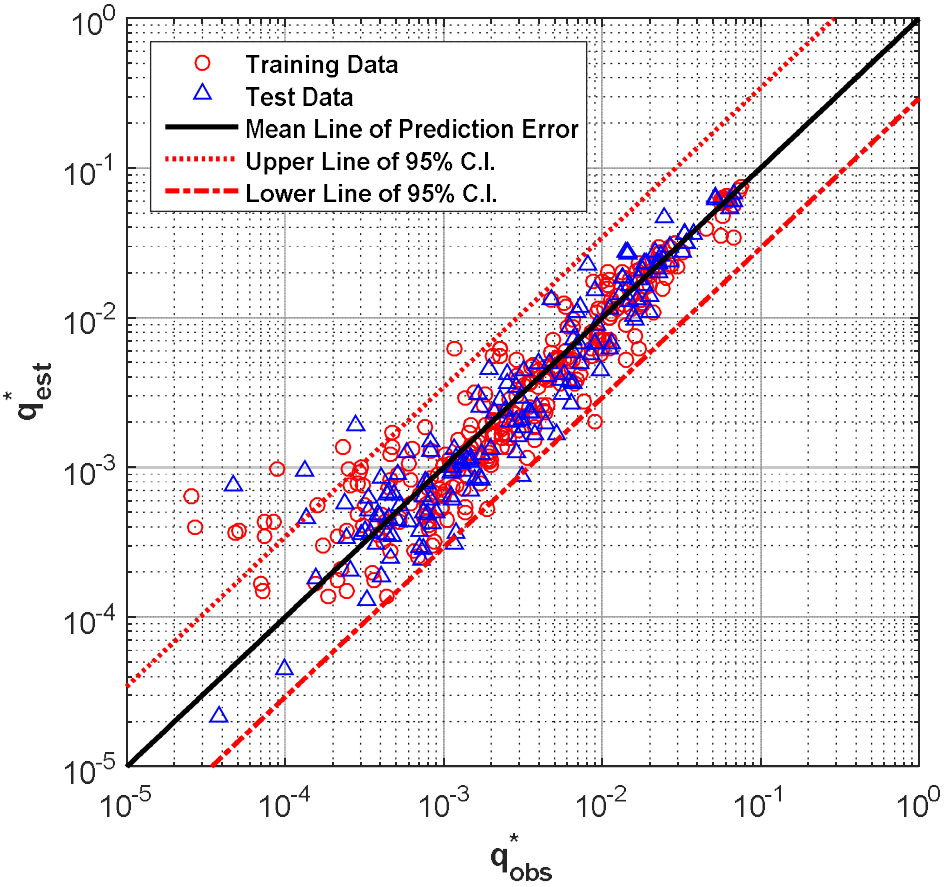


Figure C.1 Comparison of q^* between observation and estimation for training data and test data (vertical walls formula)

(2) Formula for inclined seawalls

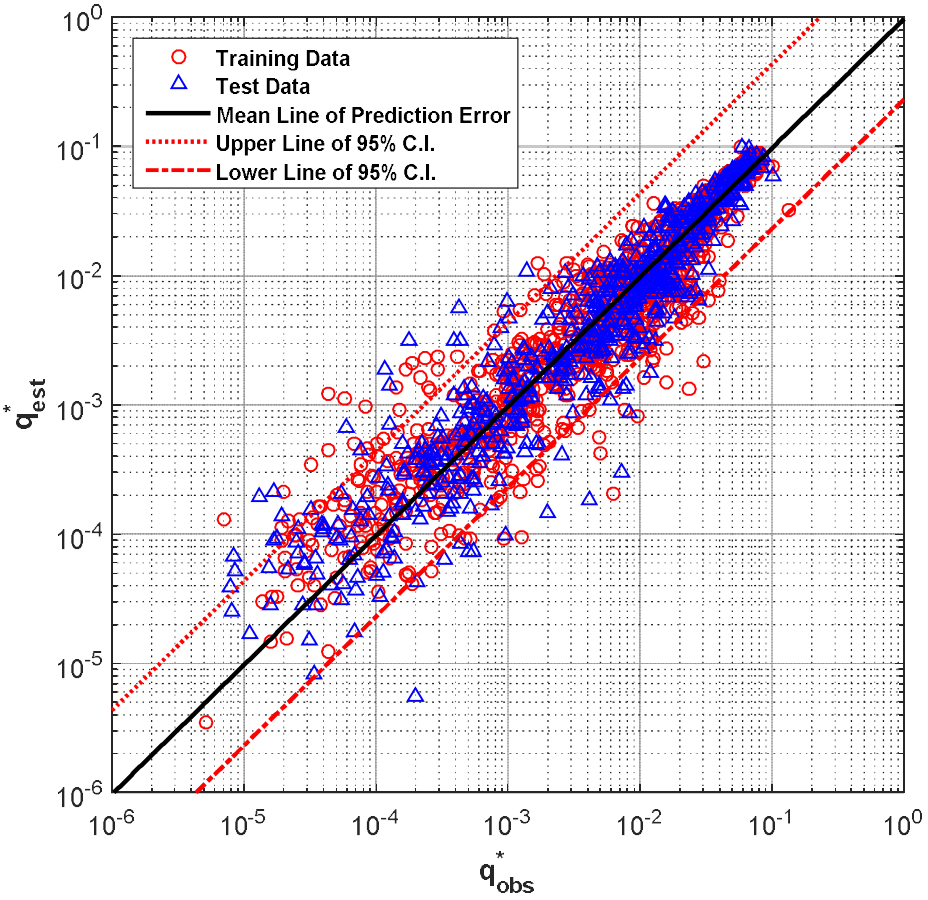


Figure C.2 Comparison of *q between observation and estimation for training data and test data (inclined seawalls formula)**

국문초록

GMDH 알고리즘을 이용한 수직벽과 경사식 호안에 대한 월파량 공식 유도

서울대학교 대학원

건설환경공학부

이 석 봉

기후변화로 인한 해수면 상승으로 인해 비정상적으로 규모가 큰 파 처오름 현상과 월파 현상이 빈번하게 발생하고 있다. 이러한 현상을 사전에 대비하기 위한 월파량 예측 모델의 개발은 필수적이다. 월파량 산정 모델에는 수많은 실험 데이터를 이용하여 각각의 변수에 따른 관계를 이용한 경험식 형태의 모델도 있지만, 컴퓨터 기반의 알고리즘을 이용한 모델도 다수 개발되어 있다. 본 연구에서는 GMDH (Group Method of Data Handling)이라는 컴퓨터 기반의 수학적 알고리즘을 적용한 월파량 공식을 제안하고자 하였다. GMDH 알고리즘은 다양한 분야에서의 예측 모델에

적용되고 있으며, 본 연구에서 예측하고자 하는 월파량 또한 본 GMDH 알고리즘을 통하여 구성된 공식을 통해 비교적 정확하게 추정될 수 있다. GMDH 모델은 기존의 machine learning 모델 (인공신경망 모델, M5 model tree 등) 과 비슷하게 Training 과정과 Test 과정을 거치는데, 다른 machine learning 모델들과는 달리 일련의 식의 형태로 모델을 제공한다는 용이성 측면에서의 장점이 있다. 따라서, 기계학습에 생소한 공학자들이 현장에서 손쉽게 월파량을 예측할 수 있는 모델을 제시할 수 있다. 본 연구에서는, 수직벽과 경사식 호안 구조물에서의 월파량을 추정할 수 있는 GMDH 알고리즘을 적용한 식을 제안하였으며, Van der Meer 등 (2004) 이 구성한 CLASH 데이터를 사용하였다. 이후 Goda (2009) 가 제안한 경험식, Etemad-Shahidi *et al*. (2014, 2016) 가 제안한 경험식, 그리고 EurOtop 매뉴얼 (2016)에서 제시된 경험식과 그 성능을 비교해본 결과, 통합형 공식의 경우 하나의 GMDH 기반 공식으로도 여러 개의 식으로 구성된 경험식들과 대등한 성능을 나타내었으며, 특히 비스듬히 입사하는 파의 조건과 경사식 호안의 경우 고려할 수 있는 구조물 경사의 거칠기를 고려한 공식을 GMDH 알고리즘을 통해 각각 구성하였을 때 기존의 공식들보다 더 정확한 월파량 추정이 가능하였다.

Keywords: 월파량, GMDH 알고리즘, 데이터모델, CLASH 데이터, 호안

학번: 2016-21261