



**Wave load formulae for prediction of wave-induced forces on a slender pile within pile groups**

**Author**

Bonakdar, Lisham, Oumeraci, Hocine, Etemad-Shahidi, Amir

**Published**

2015

**Journal Title**

Coastal Engineering

**DOI**

[10.1016/j.coastaleng.2015.05.003](https://doi.org/10.1016/j.coastaleng.2015.05.003)

**Downloaded from**

<http://hdl.handle.net/10072/167707>

**Griffith Research Online**

<https://research-repository.griffith.edu.au>

# Wave load formulae for prediction of wave-induced forces on a slender pile within pile groups

Lisham Bonakdar<sup>1\*</sup>, Hocine Oumeraci<sup>1</sup> and Amir Etemad-Shahidi<sup>2</sup>

<sup>1</sup> Leichtweiss-Institute for Hydraulic Engineering and Water Resources, Technische Universität Braunschweig, Beethovenstrasse. 51a, 38106 Braunschweig, Germany

<sup>2</sup> Griffith School of Engineering, Griffith University Gold Coast Campus, QLD 4222, Australia

\* Corresponding Author, Phone: +495313913938, E-mail: l.bonakdar@tu-braunschweig.de

## Abstract

Pile-supported structures commonly found in both offshore (e.g. offshore oil and gas platforms) and coastal environments (e.g. sea bridges, piers and jetties) are generally built by means of a group of piles in different arrangements. The correct prediction of the wave loading of closely-spaced piles of these structures is vital for both safety and economical viewpoints. Unlike single isolated piles, where a large number of studies are available together with the well-known Morison equation which is still widely applied for the calculation of wave-induced force, less research studies have been made on wave-pile group interactions. In fact, no reliable wave load formula is yet available for the prediction of wave-induced forces on a slender pile, for which the pile diameter ( $D$ ) is generally less than about 0.15 times the wave length ( $L$ ), within a pile group.

In this study, new wave load formulae for the prediction of wave-induced force on a slender pile in pile groups with different arrangements are developed using a series of laboratory data obtained from systematic model tests conducted in the 2 m-wide wave flume of Leichtweiss-Institute for Hydraulic Engineering and Water Resources (LWI) in Braunschweig, Germany. For the analysis of the laboratory data and the development of the new prediction formulae, an artificial intelligence (AI) - based computational tool, named “*hybrid M5MT-GP model*”, is implemented. The new hybrid model and the new wave load formulae allow us to systematically assess the pile group effect ( $K_G$ ) as a function of the flow regime ( $KC$  number) and the relative spacing ( $S_G/D$ ) for each tested pile group arrangement. The accuracy of the new formulae in predicting pile group effect  $K_G$  is confirmed by the statistical indicators using agreement index  $I_a$ , correlation coefficient  $CC$  and scatter index  $SI$ .

**Keywords**

Pile group effect, Pile group arrangement, Relative spacing, KC number, Hybrid M5MT-GP model, Wave load formulae

## 1. Introduction

There might be a common assumption that two or more piles in a flow should have a similar behaviour to that of a single isolated pile, but this assumption is correct only when they are adequately apart (Zdravkovich, 1977). For closely-spaced piles in groups exposed to waves, the interference effects between piles may significantly change the flow around the piles, and thus the wave load as compared to that on a single isolated pile. In such structures, wave load on a single slender pile is significantly affected by the neighbouring piles and can thus not be estimated by the commonly applied formulae for a single isolated pile which are generally based on the concept of Morison et al. (1950).

According to the angle of the connecting line of the piles centres relative to the wave direction, pile groups are commonly categorized into three basic arrangements. These three arrangements include (i) tandem, where the angle of the centre connection line of the cylinders relative to the wave direction is  $0^\circ$ , (ii) side by side, where the incident wave direction is orthogonal to the connecting line of the piles located next to each other, and (iii) staggered in which the angle is between  $0^\circ$  and  $90^\circ$  (Fig. 1). In the case of slender piles where both drag and inertia forces induced by highly complex turbulent flow are important, an analytical solution is hardly feasible. Given the high complexity of the interaction between waves and pile groups in different arrangements, laboratory experiments still represent the most reliable alternative. A number of laboratory studies have been carried out to investigate the interference effects of neighbouring piles. The methods commonly used in laboratory studies to determine wave loads on a pile group may be classified in two main categories: “wave force coefficient approach” and “wave force approach”.

In the former approach, the inertia and drag coefficients ( $C_M$  and  $C_D$ ) are determined based on the knowledge of both flow velocity and acceleration by applying for instance the least square fit. This approach was used for instance by Chakrabarti (1981, 1982), Haritos and Smith (1995), Smith and Haritos (1996, 1997). Using the calculated drag and inertia coefficients, Chakrabarti (1981, 1982) computed maximum wave forces and found a relatively good agreement with measured forces. Smith and Haritos (1996, 1997) reported that drag and inertia coefficients are dependent on the Keulegan–Carpenter ( $KC$ ) number ( $KC = u_{max}T/D$ ) and relative spacing  $S_G/D$  where  $u_{max}$  is the maximum horizontal wave-induced flow velocity,  $T$  is wave period,  $D$  is pile diameter and  $S_G$  is the gap between the surfaces of two neighbouring piles in a group of piles. Due to the lack of any reliable measured velocities, and based on appropriate wave theories, the maximum horizontal wave-induced flow velocity  $u_{max}$  was calculated by Chakrabarti (1981, 1982) below still water level at the elevation of the instrumented section of the pile on which the

local wave-induced force was measured while  $u_{max}$  was computed by Haritos and Smith (1995) as well as Smith and Haritos (1996, 1997) at the water surface elevation as the total wave force on the piles was considered. Drag and inertia coefficients were usually plotted versus  $KC$  number for different relative spacing  $S_G/D$  in these studies. However, the proposed  $C_D$  and  $C_M$  values were noticeably scattered demonstrating that different  $C_D$  and  $C_M$  values can be obtained for a given  $KC$  number.

In the latter approach, the ratio of wave force on a pile within a group to that on a single isolated pile is determined. This method was applied by Apelt and Piorewicz (1986), Mindao et al. (1987) and Li et al. (1993). Li et al. (1993) stated that the wave-induced force on a slender pile within a group of piles depends on the  $KC$  number and relative spacing  $S_G/D$ . Mindao et al. (1987) introduced two parameters named *interference coefficient*  $K_g$  and *shelter coefficient*  $K_z$  for side by side arrangement and tandem arrangement, respectively. Both  $K_g$  and  $K_z$  coefficients are representative for the force ratio ( $F_{Group}/F_{Single}$ ) where  $F_{Group}$  is the wave force on a slender pile within further neighbouring piles and  $F_{Single}$  is the wave force on a single isolated pile. They stated that  $S_G/D$  is the most significant parameter and proposed the two following formulae for the estimation of interference coefficient  $K_g$  and shelter coefficient  $K_z$  for side by side arrangement and tandem arrangement, respectively:

$$K_g = 1.265 - 0.225 \ln(S_G/D) \quad \text{for side by side arrangement} \quad (1)$$

$$K_z = 0.836 + 0.141 \ln(S_G/D) \quad \text{for tandem arrangement} \quad (2)$$

In the proposed formulae (Eqs.1 and 2), the wave conditions (i.e. wave height, period, steepness and etc.) have no influence on interference coefficient  $K_g$  and shelter coefficient  $K_z$ . For a given pile group arrangement, both coefficients only depend on relative spacing parameter  $S_G/D$  which was varied from 0.5 to 3 in the laboratory tests. It was also stated by Mindao et al. (1987) that the interference coefficient  $K_g$  and shelter coefficient  $K_z$  proposed in Eqs.1 and 2 are the average of those obtained for the side and middle piles in the pile group arrangements. Li et al. (1993) introduced *significant pile group effect*  $K_{GI/3}$  for piles in side by side arrangement exposed to irregular waves. He found out that the maximum  $K_{GI/3}$  occurs when  $KC$  number is between 15 and 20 for the case of pure waves. They also showed that, for a given pile group configuration, the combination of wave and current results in smaller grouping effect compared to wave action only.

The interaction of waves and slender piles in different pile group arrangements was also studied by means of extensive large-scale laboratory tests performed in the Large Wave Flume (GWK).

A single isolated pile and 14 pile group configurations including side by side, tandem and staggered arrangements with gaps of up to three times the pile diameter ( $1 \leq S_G/D \leq 3$ ) were tested. The results were analysed by Sparboom et al. (2006), Sparboom and Oumeraci (2006), Hildebrandt et al. (2008), Bonakdar and Oumeraci (2012, 2014) and Bonakdar (2014). Some of the general conclusions drawn from these analyses are:

- (i) Pile group effect increases by decreasing the gap between the piles in side by side arrangement,
- (ii) The amplification of the wave load on the middle pile in side by side arrangement is more noticeable than the side pile due to the influence of two neighbouring piles from both sides,
- (iii) For the tested regular waves ( $5 < KC < 38$ ), the resulting wave load on the middle pile in side by side arrangement increases up to 60 % in comparison with that on the single isolated pile. For this pile group arrangement, pile group effect becomes negligible for  $S_G/D=3$  and all piles behave like a single isolated pile in terms of the wave load,
- (iv) For tandem arrangement with  $S_G/D=1$ , which is the smallest relative spacing tested in GWK, no significant sheltering effect was observed for the tested regular waves ( $5 < KC < 38$ ).

In addition to the aforementioned general outcomes, some of the main limitations of the GWK tests which were identified may be summarized as follows:

- (i) Pile group configurations with smaller relative spacing of  $S_G/D < 1$ , where higher amplification and reduction of wave loads on piles are expected, were not tested,
- (ii) The tested wave conditions only cover a small range of relative water depth  $h/L$  located in the transition zone ( $h/L=0.084 - 0.197$ ). Therefore, the shallow and depth water conditions were not investigated. Considering the tested  $KC$  number values ( $5 < KC < 38$ ), moreover, the dominant drag regime and the dominant inertia regime were not fully covered,
- (iii) Values of the  $KC$  and  $Re$  parameters change from one section of the pile to another as a result of the variation of the flow velocity with depth. Only the total wave-induced moment on the instrumented pile was measured meaning that the flow velocity was averaged over the water depth,

- (iv) Cantilever piles (truncated with lower end far from the bottom of the flume) were used in the GWK model set-up. This might result in unrealistic flow behaviour around the group of pile due to the flow separation at the lower end (a more detailed discussion is given in Bonakdar, 2014).

Overall, the following knowledge gaps and limitations of the previous studies were identified: (i) the lack of deeper understanding of the processes associated with wave-pile group interaction, (ii) the lack of reliable wave load formulae for the prediction of wave-induced forces on a slender pile within other neighbouring piles in different arrangements and (iii) the limitations of the GWK tests. Hence, further research is needed towards improving the understanding of the involved processes, including a more precise and systematic identification of the most relevant influencing hydrodynamic and structural parameters. Based on this improved understanding, simple generic formulae for the prediction of wave loads on the piles have to be derived.

Therefore, new laboratory tests were designed and performed in 2 m-wide wave flume of Leichtweiss-Institute for Hydraulic Engineering and Water Resources (LWI) in Braunschweig, Germany. The main objectives of this experimental research study were (i) to specify more precisely and systematically the most relevant influencing wave and structural parameters on the wave loading of a pile in a pile group with different arrangements and, consequently, to improve the understanding of the processes associated with the wave-pile group interaction, and (ii) to generate a comprehensive data set which covers a range of wave and structural conditions broad enough to achieve a substantially improved insight into the hydrodynamic processes involved and, consequently, to develop wave load formulae for the prediction of wave loads on a slender pile within a group of piles in different arrangements. The former objective was addressed by Bonakdar and Oumeraci (2015b) while the latter objective, namely the development of wave load formulae for prediction of wave loads on a slender pile within pile groups in different arrangements, is the focus of this study. For the systematic analysis of the generated data resulting in the development of new generic wave load formulae, an artificial intelligence (AI) - based computational tool, a combination of M5 tree (M5MT) and genetic programming (GP) named “*hybrid M5MT-GP model*”, is implemented. This paper is outlined as follows: The laboratory data is briefly described in Section 2. Next, the hybrid M5MT-GP model is described. In Section 4, the implementation of the hybrid M5MT-GP model for the analysis of the laboratory data as well as for the development of prediction formulae are provided and the obtained results are discussed. Finally, the summary of the key results and concluding remarks are drawn in Section 5.

## 2. Laboratory experiments (LWI tests)

A large number of small scale laboratory tests were carried out in the LWI wave flume, called hereafter LWI tests. The details of the model set-up, measuring technics and test programme are provided in Bonakdar (2014) and Bonakdar and Oumeraci (2015b). Therefore, only a very brief description of the LWI tests is provided in this section. Different pile arrangements including single, side by side, tandem,  $2 \times 2$  and staggered arrangements were performed and relative spacing  $S_G/D$  was varied from 0.5 to 5.0 as depicted in Fig. 2. Regular non-breaking waves with 24 different combinations of wave heights and periods were tested pile to cover a broad range of hydrodynamic conditions. Wave steepness varies from 0.008 to 0.073 which was the maximum possible wave steepness without having incipient breaking. Relative water depth  $h/L$  varies from 0.042 to 0.64 meaning that deep, transition and shallow water conditions were considered. The  $KC$  number changes from 1.1 where the inertia regime dominates to 88 where the drag regime dominates. Reynolds number varies from  $Re=0.34 \times 10^4$  to  $Re=3.68 \times 10^4$  indicating that the LWI model is located in the subcritical zone.

## 3. Data analysis methodology

The most common method for empirical model development is the regression analysis. In the process of traditional regression analysis (e.g. simple linear, polynomial, and etc.), the functional relationship between output and input parameters (variables) is pre-defined, and the goal is only to determine a set of empirical coefficients of the input parameters. For highly complex and unknown systems, however, a predefined functional structure may not result in an accurate model.

Artificial intelligence (AI)-based methods, also called data mining or soft computing methods, are also utilized to identify hidden relationships that exist in datasets by implementing different optimization algorithms (Kazeminezhad et al. , 2010). AI-based methods can be implemented for predicting processes in complex systems, reconstructing highly nonlinear functions, classifying data and developing rule-based models (Solomatine and Ostfeld, 2008). The general steps for solving a problem using an AI-based method are (i) investigating the physical problem to be modelled, (ii) generating required data by performing experimental or numerical tests or collecting data from the available recourses, (iii) selecting, building and optimizing the AI-based method for the specific problem, and (iv) validating the developed model using a new set of data called test data (Solomatine and Ostfeld, 2008).

The most frequently used data mining techniques are artificial neural networks (ANNs), fuzzy systems, support vector machine (SVM), M5 model tree (M5MT), genetic algorithm (GA) and genetic programming (GP). A detailed review of the aforementioned AI-based methods was performed to identify the most feasible approach for the development of wave load formulae. For this purpose, two main criteria were considered in developing the prospective prediction formulae and choosing the most suitable approach. These two main criteria were accuracy and simplicity of the prediction formulae to be developed by applying the prospective AI-based model. It was concluded that a combination of M5MT and GP named hybrid M5MT-GP model is the best solution among different possibilities. A more detailed discussion is provided by Bonakdar (2014).

### 3.1. M5 model tree (M5MT)

The M5 model tree was introduced by Quinlan (1992) and represents one of the most recent computational tools for data analysis which can be applied for prediction purposes. The concept of the model tree approach is based on dividing complex problems into smaller sub-problems and solving each sub-problem (Bhattacharya et al., 2007). The concept of M5MT is described in Fig. 3 by a simple example with two input parameters ( $X_1$  and  $X_2$ ). As shown in *window A* of Fig. 3, M5MT is similar to an inverse tree with a root node at the top and a number of leaves at the bottom. The implementation of M5MT generally includes three steps called *building*, *pruning* and *smoothing*.

*Building*: Firstly, the M5MT algorithm constructs a tree by splitting the instance space (data points). *Window A* of Fig. 3 shows the constructed tree while *window B* illustrates the classified data. 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, the expected reduction in error being calculated as a result of testing each attribute at that node. In this way, the attribute (input parameter) that maximizes the expected error reduction is chosen. The splitting process is performed only if either the output values of all the instances that reach the node, called leaf, vary slightly or a few instances remain. The standard deviation reduction (SDR) is calculated as (Quinlan 1992):

$$SDR = sd(T) - \sum_i \frac{|T_i|}{|T|} \times sd(T_i) \quad (3)$$

where  $T$  is the set of examples that reach the node,  $T_i$  are the sets that result from splitting the node according to the chosen attribute and  $sd$  is the standard deviation (Wang and Witten 1997).

After the initial tree has been grown, the linear regression models are generated, using the data associated with that leaf. *Window C* of Fig. 3 shows the possible linear regression models for the given example.

*Pruning:* In the second step, all sub-trees are 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. In this way, the sub-trees which cannot improve the accuracy of the model are pruned. After pruning, there is a possibility that the pruned tree might have discontinuities between nearby leaves.

*Smoothing:* Therefore, to compensate discontinuities among adjacent linear models in the leaves of the tree a regularization process is performed, which is called *smoothing process*. In this process, the estimated value of the leaf model is filtered along the path back to the root. At each node, that value is combined with the value predicted by the linear model (LM) for that node as follows:

$$P' = \frac{np + kq}{n + k} \quad (4)$$

where  $P'$  is the prediction passed up to the next higher node,  $p$  is the prediction passed to this node from the below,  $q$  is the value predicted by the model at this node,  $n$  is the number of training instances that reach the node below, and  $k$  is a constant (Wang and Witten, 1997). M5MT has a unique algorithm meaning that for a given data set of input and output variables, the model provides a unique solution for any number of simulations. Recently, M5MT has been successfully employed for water level discharge relationship (Bhattacharya and Solomatine, 2005), sediment transport (Bhattacharya et al., 2007), stability of rubble-mound breakwaters (Etemad-Shahidi and Bonakdar, 2009 and Etemad-Shahidi and Bali, 2011), prediction of wave run-up on rubble-mound breakwaters (Bonakdar and Etemad-Shahidi, 2011), prediction wave-induced scour around pile groups (Etemad-Shahidi and Ghaemi, 2011), prediction of scour depth under submarine pipeline (Etemad-Shahidi et al., 2011), sand wave overtopping at rubble-mound structures (Jafari and Etemad-Shahidi, 2012).

The main advantages of the model trees are that they are easily applied and yield simple and transparent formulae. In addition, unlike other existing soft computing methods such as artificial neural networks (ANNs), the M5 model tree is quite transparent and does not need optimization of network geometry and internal parameters. While the traditional regression method fits a single function to the whole data set, M5 model tree splits the data points into homogeneous subsets (leaves) and fit a linear function for each leaf. Sorting the whole data point into

homogeneous sub-sets can result in a more accurate model which cannot be achieved by a common regression method. The major limitation of this method is that it can provide only a linear relationship between input and output parameters at each leaf, while the relationships between the output and input parameters are not necessarily linear.

### ***3.2. Genetic programming (GP)***

Genetic Programming (GP) is an evolutionary symbolic regression method where, unlike traditional regression methods, the functional structure between output and input parameters is not pre-defined and is a result of the search process. GP was firstly introduced by Koza (1992) as a powerful tool for solving complex problems and is similar to more widely known genetic algorithms (GA). However, unlike GA, where a set of numbers is the solution of the given problem, a formula also called a computer programme is generated as the solution. GP creates an initial population of functional forms from user-specified building blocks stored as *function* and *terminal* sets. These building blocks can consist of a range of operators, including addition, subtraction, multiplication, division, etc. Using a tree-based representation, the genotype is arranged such that the top and middle of the tree is created from members of the function set, and the leaves consist of members of the terminal set. Once the initial population has been created, the so called *reproduction*, *mutation* and *crossover* are used to generate *offspring*. The best offspring (equation) resulting from this process is the solution of the problem. These steps are individually described as follows:

First, an initial population of individuals (equations or programs) of a certain size is created by randomly picking up a set of *functions*, which consists of basic mathematical operators (e.g. addition, subtraction, multiplication, division, log, etc.) and constants, and the so called *terminals* which consists of independent variables (input parameters) and constants. The constants can either be physical constants (e.g. Earth's gravitational acceleration, density of water) or randomly generated constants. The function set and the terminal set construct the main body of GP. Hence, their appropriate identification plays a pivotal role in developing a robust GP model. The determination of the mathematical operators considered in the function set depends upon the degree of complexity of the problem to be modelled and should be carefully chosen, otherwise they might result in an extremely complex formula which carries no physical insight.

For example,  $\left[ \exp(x_1/3) - 4x_2 \right]$  can be considered as an equation (programme). The tree structure of this programme is shown in Fig. 4. A population of random trees representing the programmes is initially constructed and genetic operations are performed on these trees to generate programmes (equations) with the help of the terminal and the function sets. For the given

example shown in Fig. 4,  $\{exp, -, /, *\}$  and  $\{3, 4, x_1, x_2\}$  belong to function and terminal sets, respectively, where  $x_1$  and  $x_2$  are input variables of the given example. Except for the root of the tree which must be a function, a uniform random probability distribution is usually considered for the function and terminal selection (Koza, 1992).

Second, the fitness of each equation (programme or individual) in a population is evaluated based on the chosen criterion. For instance, the root mean squared error (*RMSE*) can be used as the criterion. In this case, a smaller value of *RMSE* shows a better fitness.

Third, at each generation, new sets of models are evolved by applying the genetic operators: crossover, mutation and reproduction (Koza, 1992). These new models are named *offspring*, and they form the basis for the next generation.

- (i) *Crossover*: Crossover is applied on an individual (equation or programme) by simply switching one of its nodes with another node from another individual in the population. With a tree-based representation, replacing a node means replacing the whole branch. Therefore, two new programmes (equations) are generated as shown in Fig. 5. The resulting individuals (equations) are inserted into the new population.
- (ii) *Mutation*: Mutation affects an individual in the population. In mutation, a sub-tree is replaced by another one, randomly, (Fig. 6) and the new sub-tree is inserted into the new population.
- (iii) *Reproduction*: Based on the fitness criterion, the best programme is selected and copied into the new population.

While the role of the crossover operator is to generate new models, which did not exist in the old population, the mutation operator guards the search against premature convergence by constantly introducing new genetic material into the population (Elshorbagy et al., 2010).

Forth, if the number of generations is equal to a certain value selected for the model, the programme is terminated. The equation (individual or programme) with the best fitness is represented as the best solution (formula) for the given problem. A Genetic Programming flowchart is illustrated in Fig. 7 where  $P$  is population size. GP has lately been applied for the prediction of wave height (Nitsure et al., 2012), flow discharge in compound channels (Azamathulla and Zahiri, 2012), scour below submerged pipeline (Azamathulla et al., 2011) and scour around piles (Güven et al., 2009).

The main strengths of GP may be summarized as follows:

- (iv) Like M5MT, GP also provides the result of modelling in a form of computer programme (equation) which can be interpreted by scientists. This mathematical representation provides a great benefit in empirical modelling of unknown phenomena where a theoretical model does not exist. This has resulted in the successful application of GP to a wide range of practical problems of various degree of complexity over the last two decades.
- (v) Unlike the traditional data analysis methods, in which users have to specify the overall functional form of the model in advance, GP evolves both functional form of the model and its numerical (empirical) coefficients.
- (vi) GP can develop non-linear relationships between the output and input parameters using different mathematical functions while M5MT can only generate linear relationship between output and input parameters. Hence, GP can provide a better approximation of the complex natural processes and more insight into the functional relationship between the input variables

The drawbacks of GP might be summarized as follows:

- (i) Like regression analysis, GP also fits a single overall function to the entire data set while M5MT splits the data set into homogeneous sub-sets (classes) and fit a linear function for each sub-set which might result in a more understandable and accurate model.
- (ii) GP is not as fast as M5MT and, depending on the complexity of the problem, the initial population and determined functions, the modelling procedure might take hours or days.

### ***3.3. Hybrid M5MT-GP model***

Considering the strengths and limitations of the M5MT and GP and making use of their respective strengths, a one-way coupling of M5MT and GP was considered for the development of final wave load formulae. By this way, all data sets obtained from laboratory experiments are classified in different classes based on the criteria of M5 model tree algorithm. Once the classification of the data is completed, GP is applied to the classified data. The GP method is used as a non-linear approach to predict wave loads on a slender pile within further neighbouring piles in different arrangements as a function of the most relevant influencing wave and structural parameters. The overview of the procedure of data analysis and the development of prediction formulae for the wave load using the hybrid M5MT-GP model, including the classification, prediction and validation processes, are drawn in Fig. 8.

The one-way hybrid M5MT-GP model is more comprehensively described in Fig. 9 by providing a simple example of a problem with two input variables of  $X_1$  and  $X_2$  and an output variable  $Y$ .

### ***Step 1: Application of M5MT***

As shown in *window A* of Fig. 9, M5MT is applied to all data. This application results in three outputs illustrated in *window B*. The first output is an inverse tree where the relevant influencing parameters together with their splitting values are located at the nodes of the inverse tree. The nodes at the end of the inverse tree are called leaves. These leaves represent, in fact, the final sub-sets of the developed tree. Based on the chosen variables and splitting values at the upper nodes of the tree, all data are grouped into leaves through the branches. The second output is the classification of the data into the generated leaves (classes). Three groups of data sets are obtained (sub-sets 1-3). The third output is a set of linear models (LMs) which are generated using the data associated with each leaf. The main limitation of the M5MT model consists in its inability to generate non-linear models as the relation between the output and input parameters are not necessarily linear. In order to overcome this crucial limitation, the GP model is used in the next step.

### ***Step 2: Application of GP***

The GP model is applied to the data associated with each leaf, individually. This process is demonstrated in *window C*. For the given example, 3 GP models are needed for the three subsets 1-3 obtained from M5MT. As mentioned before an optimisation of the GP model is required to provide the desired solution describing the relationship between output and input parameters based on physical behaviour of the system. Therefore, GP model should be optimised by changing the settings of the model including mathematical functions, population size and generations and etc. GP provides a solution for each simulation. The procedure of GP simulation is drawn in both Fig. 7 and *window D* of Fig. 9. For the example, 3 possible equations for 3 optimised GP models can be seen in *window E* of Fig. 9. These so called M5MT-GP-based equations are, in fact, the solutions obtained from three optimised GP models applied to three sub-sets classified by M5MT.

## **4. Development of wave load formulae using hybrid M5MT-GP model**

### ***4.1. Data classification using M5MT***

As described in the previous section, in the first place, M5MT needs to be applied to all data and classifies them into homogeneous sub-sets so called leaves. Bonakdar (2014) performed a comprehensive analysis on the effect of non-dimensional wave parameters including  $KC$  number, Reynolds number  $Re$ , relative water depth  $h/L$  and wave steepness  $H/L$  on pile group effect  $K_G$ . The latter represents the relative wave force ratio ( $K_G=f_{Group}/f_{Single}$ ) where  $f_{Group}$  is the maximum line force on a slender pile within a pile group in different arrangements and  $f_{Single}$  is the maximum line force on an isolated single pile. Among all these parameters,  $KC$  number was identified as the most suitable parameter to describe the effect of wave conditions on pile group interaction. It was stated that pile group effect  $K_G$  related to  $KC$  number is more appropriate than that related to other non-dimensional wave parameters. In addition,  $KC$  number is a function of both wave period and flow velocity which make it an appropriate parameter for describing wave-induced flow conditions. Therefore,  $KC$  number was favoured as a parameter describing the flow regime for the development of wave load formulae. Bonakdar and Oumeraci (2014) compared pile group effect  $K_G$  values measured in LWI tests with Froude scaled ones obtained in the 6.5 larger scale model tests (GWK), where  $Re$  number was ranged from  $1.37 \times 10^5$  to  $5.25 \times 10^5$ , and found pile group effect  $K_G$  to be similar for a given structural condition (e.g. same pile group arrangement relative spacing  $S_G/D$ ) and the same  $KC$  number in both LWI and GWK tests. Therefore, the non-dimensional pile group effect  $K_G$  values obtained in this study can be applied for other cases with higher  $Re$  values.

From the structural point of view, as can be concluded from the results of the previous studies on wave loads on pile groups (e.g. Chakrabarti, 1979, 1981, 1982; Li et al., 1993; Mindao et al., 1987), pile group arrangement and relative spacing parameter  $S_G/D$  are the most significant parameters affecting the resulting wave load on a slender pile within other neighbouring piles. Overall, it can be stated that:

$$K_G = \frac{f_{Group}}{f_{Single}} = f \left\{ KC, \frac{S_G}{D}, Pile\ group\ arrangement \right\} \quad (5)$$

Different pile group arrangements including side by side, tandem, staggered and  $2 \times 2$  were individually analysed by M5MT. Pile group effect  $K_G$  was set as the output while  $KC$  number and relative spacing  $S_G/D$  were set as the inputs of the model representing the most relevant influencing hydrodynamic and structural parameter, respectively. For all pile group arrangements,  $KC$  number varies from 1.1 to 88 meaning that it covers all wave-induced flow regimes. Relative spacing parameter  $S_G/D$ , however, differs for different pile group arrangements. Table 1 provides a comprehensive overview of the hydrodynamic and structural conditions of the data used for the development of the M5MT-GP model.

#### 4.1.1 Side by side arrangement

In side by side arrangement, seven different pile group configurations with relative spacing of  $S_G/D=0.5, 0.75, 1.0, 1.5, 2.0, 3.0$  and  $5.0$  were tested. In total, 146 regular non-breaking waves were performed for the side by side arrangement. Fig. 10 illustrates the relationship between pile group effect  $K_G$  and  $KC$  number for side by side arrangement as well as the developed tree showing splitting parameters and the corresponding splitting values at nodes and leaves (sub-sets), where data points are finally classified. The classified sub-sets are also demonstrated by manually drawn dash lines. As seen, M5MT classified all data into 5 different sub-sets based on different combinations of  $KC$  number and relative spacing  $S_G/D$ . The first splitting parameter located at the root of the inverse tree is relative spacing  $S_G/D$  and its splitting value is 1.5. As discussed by Bhattacharya et al. (2007), the splitting value does not necessarily have any physical interpretation and is obtained by minimizing the prediction error. However, this value distinguishes the so-called “*closely-spaced piles*” ( $S_G/D \leq 1.5$ ) where a greater pile group interaction is expected from “*largely-spaced piles*” ( $S_G/D > 1.5$ ) where less interaction of piles occurs due to larger gaps between piles.

For the configurations with  $S_G/D \leq 1.5$ , where the piles are closely spaced next to each other in an array,  $KC$  number becomes important. As seen on the left hand side of the tree shown in Fig. 10, data points with  $S_G/D \leq 1.5$  were grouped by  $KC$  number with the splitting value of 13. As shown by the dash-line, this is almost the value at which maximum amplification of wave load on the closely-spaced piles in side by side arrangement occurs. Data with  $S_G/D \leq 1.5$  and  $KC > 13$  was classified only in one group (sub-set 3 in Fig. 10). For data with  $S_G/D \leq 1.5$  and  $KC < 13$ , another categorization is made by  $KC$  number as shown on the down-left hand side of the tree. The values that came down from the root through the branch to this node, were classified into two other group at  $KC=6$ . As can be concluded from the depicted figure, pile group effect  $K_G$  is almost constant when  $KC$  is smaller than 6 (inertia dominated regime, sub-set 1) and shows different behaviour for the cases with  $KC > 6$  (sub-set 2 in Fig. 10).

For the configurations with  $S_G/D > 1.5$ , as can be concluded from the developed M5MT model, data points were sorted into two groups according to relative spacing parameter  $S_G/D$  with a splitting value of 2. This means that the M5MT model found the data with  $S_G/D > 1.5$  more homogeneous than others with  $S_G/D \leq 1.5$ . This is also apparent from the plotted data. As seen in Fig. 10,  $K_G$  values are more or less the same for the whole range of  $KC$  values, meaning that pile group interaction is not dependent on the hydrodynamic conditions for  $S_G/D > 1.5$ . The developed tree and relationship between  $K_G$  and  $KC$  number for different  $S_G/D$  shown in Fig. 10 indicate that M5MT is able to identify the relationship between input and output parameters and to

classify the data based on the physical behaviour of the system. The data associated with each leaf will be considered for the development of prediction formulae using GP as M5MT can only generate a linear relationship between output and input parameters.

#### **4.1.2 Tandem arrangement**

In the case of tandem arrangement, seven different pile group configurations with relative spacing of  $S_G/D=0.5, 0.75, 1.0, 1.5, 3.0, 4.0$  and  $5.0$  were tested. In total, 136 data were classified by M5MT for this arrangement. As seen in Fig. 11, the developed tree is very simple and has only one root (node) and two leaves. All 136 data points were sorted into two groups based on  $S_G/D$  with a splitting value of 3. This is also demonstrated by the manually drawn dash-line splitting data with  $S_G/D \leq 3$  from the rest of the data.

It is apparent from the developed tree that the data points were not categorized by  $KC$  number which was one of the inputs of the model meaning that M5MT discovered a similar relationship between  $K_G$  and  $KC$  number for different pile configurations with  $S_G/D \leq 3$  (sub-set 1 in Fig. 11). For the configurations with  $S_G/D > 3$ , where piles are fairly far from each other, data points were sorted into another group (sub-set 2 in Fig. 11). In this case  $K_G$  values are grouped around  $K_G=1$  for all  $KC$  values indicating that there is no interaction between piles and each pile behaves like a single isolated pile (Fig. 11).

#### **4.1.3 2×2 arrangement**

For the 2×2 arrangement, four pile group configurations with relative spacing of  $S_G/D=0.5, 0.75, 1.0$  and  $2.0$  were tested. In total, 83 regular waves were performed for this arrangement. Fig. 12 illustrates the relationship between pile group effect  $K_G$  and  $KC$  number for the 2×2 arrangement as well as the developed tree showing splitting parameters and the corresponding splitting values at nodes and leaves (sub-sets). The most interesting indication from Fig. 12 is that the pile group interaction of 2×2 arrangement is clearly a combination of both pile group interactions observed in side by side and tandem arrangement. This means that both wave load amplification seen in side by side (Fig. 10) and sheltering effect observed in tandem arrangement (Fig. 11) can be seen in the so called 2×2 arrangement.

As seen in Fig. 12, M5MT model classified all data into 4 different sub-sets based on both  $KC$  number and relative spacing  $S_G/D$ . Like for the case of side by side arrangement, the first splitting parameter shown at the root of the inverse tree is relative spacing  $S_G/D$  and its splitting value is 1.5. Next, for cases with  $S_G/D > 1.5$  and  $S_G/D \leq 1.5$ ,  $KC$  number appears as the splitting parameter. In both nodes, the corresponding splitting value is 6. This means that in both closely-spaced piles ( $S_G/D \leq 1.5$ ) and  $S_G/D > 1.5$  M5MT model found different physical behaviours in the

data for  $KC < 6$  (sub-sets 1 and 3 in Fig. 12) where the resulting wave load on the pile is primarily dominated by inertia and  $KC > 6$  (sub-sets 2 and 4 in Fig. 12) where both inertia and drag forces are important.

#### **4.1.4 Staggered arrangement**

Six different pile group configurations with relative spacing of  $S_G/D = 0.6, 0.75, 1.0, 1.5, 3.0$  and  $5.0$  were used for staggered arrangement where the angle between incident waves and the axis of pile groups is  $45^\circ$ . In total, 120 regular waves cases were used for the staggered arrangement. Fig. 13 demonstrates the relationship between pile group effect  $K_G$  and  $KC$  number for staggered arrangement with different  $S_G/D$ . As seen, no specific relationship can be seen between pile group effect  $K_G$  and  $KC$  number for the tested pile configurations with different  $S_G/D$  values and almost all of the data points for different wave and structural conditions vary between 0.9 and 1.1. Applied M5MT model did not classify data points into different sub-sets and only represented a model with only one leaf. In fact,  $K_G = 1$  was found as the best fitting line to data points for the downstream pile in staggered arrangement. This result is not, however, unexpected as the  $K_G$  values obtained for staggered arrangement ( $45^\circ$ ) are between those gained for side by side ( $90^\circ$ ) and tandem ( $0^\circ$ ) arrangements. In order to study the effect of the direction of incident waves on pile group interaction, further investigations by testing pile group arrangements with different angles ( $0^\circ - 90^\circ$ ) of the centre connection line of the cylinders relative to the wave direction are needed.

#### **4.1.5 Overall M5MT model**

An overall model can be proposed including all generated models as constituents. This overall M5MT model is drawn in Fig. 14. As seen, the complete model, which consists of all models individually developed for each pile group arrangements, has 12 sub-sets (leaves) named from A to L. The first splitting parameter located at the root of the inverse tree is relative spacing  $S_G/D$  and the splitting value is 1.5. The pile group configurations with  $S_G/D \leq 1.5$  were named “*closely-spaced piles*” where a greater pile group interaction is expected.

The second splitting criterion of M5MT model is the pile group arrangement at the second node of the inverse tree based on which an appropriate arrangement is chosen among the four tested arrangements including side by side, tandem,  $2 \times 2$  and staggered arrangements. From this node, depending on the type of pile group arrangement, further splitting parameters including  $KC$  number and relative spacing  $S_G/D$  might become important and play a role in sorting data. By this way, further categorizations of a test might be made depending on its specific wave and structural conditions and it reaches the final node called leaf through the branches.

## 4.2. Development of wave load formulae using GP

After the classification of data into sub-sets performed by means of the M5MT model, the prediction is made by applying the GP model to the data associated with each leaf. For each pile group arrangement, like for M5MT,  $KC$  number and relative spacing  $S_G/D$  are used as the inputs of GP models while pile group effect  $K_G$  is considered as the output parameter. These parameters are, indeed, the *terminal set* of the GP model. For the *function set*, however, different mathematical operators need to be tested in order to optimise the GP model and, consequently, to obtain the best solution (formula). Two main criteria were considered for selecting the best solution (formula) among a large number of possible solutions that can be developed by GP. These two main criteria were accuracy and simplicity of the possible solution.

### 4.2.1 Side by side arrangement

As shown in Fig. 10, all 146 data were classified by M5MT into 5 different sub-sets based on both  $KC$  number and relative spacing  $S_G/D$ . A number of mathematical operators were tested for the optimisation of the model and the development of appropriate formulae. The optimised M5MT-GP model and formulae are shown in Fig. 15. The fitting curves of developed M5MT-GP-based wave load formulae are also drawn in Fig. 16 for different hydrodynamic conditions and pile configurations.

As seen in Fig. 15, five wave load formulae (Eqs. SS1-SS5) were obtained for the prediction of non-breaking wave loads on a slender pile in side by side arrangement. Two of these five formulae show constant values for  $K_G$ . Pile group effect  $K_G$  is equal to 1 for configurations with  $S_G/D > 2$  meaning that there is no amplification of resulting wave loads on piles in a side by side arrangement and they can be treated as a single isolated pile. This finding is physically sound as no significant pile group interactions are expected, when they are placed fairly far from each other ( $S_G/D > 2$ ). When  $1 < S_G/D \leq 2$ , wave loads on the instrumented pile in the side by side arrangement is 10% larger those measured for a single isolated pile ( $K_G = 1.1$ ). It is also apparent from Fig. 16 that the horizontal lines  $K_G = 1.1$  and  $K_G = 1$  match to the data points from the LWI tests.

For the so called *closely-spaced piles* ( $S_G/D \leq 1.5$ ), GP developed three formulae for the sub-sets obtained by the M5MT.

For  $KC < 6$ , where the resulting wave load on the pile is primarily dominated by inertia, the developed formula (Eq. SS1 in Fig. 15), is solely dependent on the spacing between the piles. This result is in agreement with the data points plotted in Fig. 16. As seen in this figure,  $K_G$  is not dependent on  $KC$  number for  $KC < 6$ . The negative exponent of  $S_G/D$  in Eq. SS1 shows that

interaction effect decreases by increasing the spacing between the piles, which is physically sound.

For  $KC > 6$ , the generated formulae (Eq. SS2 and SS3) are multivariate functions of both relative spacing  $S_G/D$  and  $KC$  number. By increasing  $KC$  number from approximately 7 to 13, where the highest pile group interaction occurs, grouping effect parameter  $K_G$  sharply increases. The positive exponent of  $KC$  shows that Eq. SS2 can correctly represent this behaviour.

For  $KC > 13$ ,  $K_G$  values decrease for *closely-spaced piles* ( $S_G/D \leq 1.5$ ). GP developed Eq. SS3 for this condition (leaf) made by M5MT. In this case, the relationship between  $K_G$  and  $KC$  number is based on a combination of exponential and power operators. This relationship provides a smooth slope for the formula (see fitting curves in Fig. 16) by which  $K_G$  decreases until  $KC$  is about 35~40. From this point, as shown in Fig. 16,  $KC$  does not affect grouping effect anymore and  $K_G$  is solely dependent on relative spacing  $S_G/D$ . This is the case that the drag force is completely dominant and inertia component is negligible. The negative power of  $S_G/D$  in Eq. SS2 and SS3 demonstrate that interaction effect decreases as relative spacing  $S_G/D$  keeps increasing. It should be noted that the proposed formulae are valid for the tested conditions where  $KC$  number is between 1.1 and 88 and  $0.5 \leq S_G/D \leq 5$ . However, all other side by side configurations with  $S_G/D \geq 3$  can be treated like single isolated piles as there is no pile group interaction. It was also reported by Chakrabarti (1979) that pile group interaction for slender piles in side by side arrangement completely disappears at  $S_G/D=4$ .

The performance of the developed M5MT-GP-based formulae was quantitatively evaluated using statistical indicators such as agreement index  $I_a$ , correlation coefficient  $CC$ , scatter index  $SI$ , and *Bias* defined as follow:

$$I_a = 1 - \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum (|x_i - \bar{x}| + |y_i - \bar{y}|)^2} \quad (6)$$

$$CC = \frac{(1/n)[(x_i - \bar{x})^T (y_i - \bar{y})]}{\sqrt{(1/n)(x_i - \bar{x})^2} \sqrt{(1/n)(y_i - \bar{y})^2}} \quad (7)$$

$$SI = \frac{\sqrt{1/n \sum (y_i - x_i)^2}}{\bar{x}} \quad (8)$$

$$Bias = \bar{y} - \bar{x} \quad (9)$$

where  $x_i$  and  $y_i$  denote the predicted and the measured values, respectively and  $n$  is the number of measurements (data).  $\bar{x}$  and  $\bar{y}$  are the corresponding mean values of the predicted and measured parameters. For the 146 train data used for the development of the M5MT-GP model, the scatter diagram of the measured and predicted  $K_G$  values as well as the statistical values are shown in Fig. 17. As seen, the scatter between measured and predicted  $K_G$  values is very small and the data points are concentrated around the optimal line. The statistical parameters also indicate that the developed M5MT-GP model system can precisely predict non-breaking wave loads on a slender pile in side by side arrangement. As seen, the agreement index ( $I_a$ ) is 0.99 and the scatter index ( $SI$ ) is only 5.3%.

#### **4.2.2 Tandem arrangement**

As shown in Fig. 11, M5MT classified all 136 data into two groups (sub-sets) based on  $S_G/D$  with a splitting value of 3. Some GP models with different mathematical functions and settings were built for the classified data and the optimised model was obtained. The M5MT-GP model and the corresponding formulae are shown in Fig. 18.

For  $S_G/D \leq 3$ , the developed formula is a function of both  $KC$  number and relative spacing  $S_G/D$  (Eq. T1 in Fig. 18). As can be concluded from this equation, the maximum  $K_G$  value is close to 1 for either very small  $KC$ -values ( $KC \rightarrow 0$ ) or for very large  $S_G/D$  values ( $S_G/D > 3$ ). Pile group effect  $K_G$  decreases when relative spacing  $S_G/D$  decreases meaning that sheltering effect becomes more significant by decreasing the gap between the piles in tandem arrangement. By increasing  $KC$  number, pile group effect  $K_G$  decreases and reaches its minimum value at the largest  $KC$  number indicating that the highest sheltering occurs for very large  $KC$  values where the resulting wave load is solely dominated by drag.

For  $S_G/D > 3$ , the data points were sorted into another group by M5MT model. For this case,  $K_G$  was found to be 1 for all tested  $KC$  values ( $KC=1 - 88$ ). This means that no sheltering effect was observed for the cases with large spacing between the piles as expected. For this condition, non-breaking wave load on a protected pile in tandem arrangement can be calculated by Morison's formula as the sheltered pile behaves like a single isolated pile. The curves of the developed M5MT-GP-based wave load formulae are plotted together with the data points from the LWI flume tests for tandem arrangements in Fig. 19, demonstrating that the developed formulae are able to predict properly non-breaking wave load on the sheltered piles in a tandem arrangement.

Fig. 20 demonstrates the scatter between the measured and predicted  $K_G$  values for tandem arrangement for all 136 data with different tested pile configurations and hydrodynamic conditions. The values of the four statistical indicators are also shown in Fig. 20. As seen, the

data points are concentrated on the 45 degree line which represents the ideal correlation. The statistical indicators with  $I_a=0.975$  and  $SI=4.5\%$  demonstrate the high capability of the developed M5MT-GP model to reproduce the experimental data.

### 4.2.3 2×2 arrangement

For 2×2 arrangement, M5MT classified all 83 data into 4 different sub-sets based on both relative spacing  $S_G/D$  and  $KC$  number parameters as shown in Fig. 12. The developed M5MT-GP model and corresponding formulae obtained from optimised GP models are shown in Fig. 21. As mentioned before, the first splitting parameter shown at the root of the inverse tree is relative spacing with a splitting value of  $S_G/D=1.5$  which distinguishes the so called closely-spaced piles cases ( $S_G/D \leq 1.5$ ) from the largely-spaced piles cases ( $S_G/D > 1.5$ ). For both closely-spaced and largely-spaced piles, the second splitting parameter is  $KC$  number with a splitting value of  $KC=6$  which separates inertia dominant conditions ( $KC < 6$ ) from cases where both drag and inertia or only drag are important ( $KC > 6$ ).

For  $KC < 6$ , it was found that the instrumented pile in the so called 2×2 arrangement behaves like a single isolated pile and, therefore,  $K_G$  is equal to 1 (Eqs. 2×2-1 and 2×2-3 in Fig. 21) for all tested configurations with different  $S_G/D$  values (Fig. 22). The comparison of Fig. 16, Fig. 19 and Fig. 22 for the cases with  $KC < 6$  indicates that the wave load on the instrumented piles in 2×2 is somehow the average of the results measured for side by side and tandem arrangements.

For  $KC > 6$ , two formulae were developed by GP shown in Fig. 21. Eq. 2×2-2 in Fig. 21 obtained for closely-spaced piles ( $S_G/D \leq 1.5$ ) is a multivariate function of both  $KC$  and  $S_G/D$  parameters, while Eq. 2×2-4 in Fig. 21 depends only on  $KC$  number for largely-spaced piles. However, in contrast to side by side and tandem arrangements, only one pile configuration was performed for  $S_G/D > 1.5$ , namely  $S_G/D=2$ . In order to determine the threshold value of the spacing at which the grouping effect becomes insignificant for the 2×2 arrangement, further tests with larger spacing between the piles are needed. In this arrangement, therefore, the proposed formulae are only valid for the tested conditions ( $0.5 \leq S_G/D \leq 2$ ). The scatter diagram and statistical values indicate that the proposed model is accurate in predicting  $K_G$  values for the tested conditions (Fig. 23). For this arrangement, agreement index  $I_a$  is 0.933 and scatter index  $SI$  is 7.3%.

For staggered arrangement, as shown in Fig. 13 and discussed in section 4.1.4, no specific relationship was found between pile group effect  $K_G$  and all  $KC$  numbers for all tested pile configurations with different  $S_G/D$ . The range of  $K_G$  is from 0.9 to 1.1 for almost all data. As stated before, further investigations on pile group arrangements with different angles between the

piles' centre connection line and wave direction ( $0^\circ - 90^\circ$ ) are needed to study how the variation of wave direction affects resulting wave loads on piles in a group.

#### 4.2.4 Overall M5MT-GP model and formulae

The individual GP models developed and optimised for each type of the pile group arrangements are brought together to build an overall M5MT-GP model which is summarized in Fig. 24. This overall model includes (i) M5MT model classifying the entire data sets and (ii) GP-based formulae developed for the classified data (Eqs. 10 – 21).

$$K_G = \frac{f_{Group}}{f_{Single}} = 1.14 \left( \frac{S_G}{D} \right)^{-0.19} \quad \text{for side by side, } S_G/D \leq 1.5 \text{ \& } KC \leq 6 \quad (10)$$

$$K_G = \frac{f_{Group}}{f_{Single}} = 0.87 \left( \frac{S_G}{D} \right)^{-0.51} (KC)^{0.26} \quad \text{for side by side, } S_G/D \leq 1.5 \text{ \& } 6 < KC \leq 13 \quad (11)$$

$$K_G = \frac{f_{Group}}{f_{Single}} = 1.4 \left( \frac{S_G}{D} \right)^{-0.46} \exp(52.7(KC)^{-2.22}) \quad \text{for side by side, } S_G/D \leq 1.5 \text{ \& } KC > 13 \quad (12)$$

$$K_G = \frac{f_{Group}}{f_{Single}} = 1 \quad \text{for } 2 \times 2 \text{ arrangement, } S_G/D \leq 1.5 \text{ \& } KC \leq 6 \quad (13)$$

$$K_G = \frac{f_{Group}}{f_{Single}} = 1.4 - 0.136 \left( \frac{S_G}{D} \right)^{-0.32} \exp\left(\frac{KC}{56}\right) \quad \text{for } 2 \times 2 \text{ arrangement, } S_G/D \leq 1.5 \text{ \& } KC > 6 \quad (14)$$

$$K_G = \frac{f_{Group}}{f_{Single}} = 1 - 0.074 \left( \frac{S_G}{D} \right)^{-0.8} \exp\left(\frac{KC}{56}\right) \quad \text{for tandem arrangement \& } S_G/D \leq 3 \quad (15)$$

$$K_G = \frac{f_{Group}}{f_{Single}} = 1 \quad \text{for staggered arrangement} \quad (16)$$

$$K_G = \frac{f_{Group}}{f_{Single}} = 1 \quad \text{for tandem arrangement \& } S_G/D > 3 \quad (17)$$

$$K_G = \frac{f_{Group}}{f_{Single}} = 1 \quad \text{for } 2 \times 2 \text{ arrangement, } S_G/D > 1.5 \text{ \& } KC \leq 6 \quad (18)$$

$$K_G = \frac{f_{Group}}{f_{Single}} = 1.1 - 0.013 \exp\left(\frac{KC}{30}\right) \quad \text{for } 2 \times 2 \text{ arrangement, } S_G/D > 1.5 \text{ \& } KC > 6 \quad (19)$$

$$K_G = \frac{f_{Group}}{f_{Single}} = 1.1 \quad \text{for side by side arrangement \& } 1.5 < S_G/D \leq 2 \quad (20)$$

$$K_G = \frac{f_{Group}}{f_{Single}} = 1 \quad \text{for side by side arrangement \& } S_G/D > 2 \quad (21)$$

The proposed M5MT-GP model is quite simple, compact and easy to use. For the purpose of this study, the first question to be answered is about the value of relative spacing  $S_G/D$  between the piles, as shown at the root of the inverse tree. The case will be identified either as “*closely-spaced pile group*” for  $S_G/D \leq 1.5$  and “*largely spaced pile group*” for  $S_G/D > 1.5$ .

Next, the pile group arrangement should be determined, including, side by side,  $2 \times 2$ , tandem and staggered arrangements as drawn at the second node of the inverse tree.

From this stage on and depending on the type of pile group arrangement,  $KC$  number, relative spacing  $S_G/D$  or both might need to be considered for further classifications. Finally, M5MT leads to the appropriate leaf (class) based on the related hydrodynamic and structural conditions. At this step, equation number at the leaf determines which GP-based formula to be applied for the calculation of pile group effect  $K_G$  of the instrumented pile within a pile group for the specific case considered.

The scatter diagram of the measured and predicted  $K_G$  values is drawn in Fig. 25 for all 485 data points used for the development of the M5MT-GP model. As seen, the predicted and measured  $K_G$  values are in a very good agreement and the scatter between them is very small as the data points are concentrated around the optimal line. The performance of the complete M5MT-GP model as well as that of each pile group arrangement were quantitatively evaluated using statistical indicators including agreement index  $I_a$ , correlation coefficient  $CC$ , scatter index  $SI$ , and  $Bias$  and the results are given in Table 2.

Though only  $K_G=1$  was obtained for the staggered arrangement with different wave and structural conditions, values of the statistical parameters indicate that the developed M5MT-GP model can precisely reproduce the experimental results for non-breaking wave loads on a slender pile in a group of piles. As shown in Table 2, the agreement index ( $I_a$ ) and scatter index ( $SI$ ) of the model for 485 tests are, 0.987 and 5.8%, respectively.

### 4.3. Validation of the developed formulae

A set of new data measured in LWI wave flume was applied for the validation of the developed hybrid M5MT-GP model. Table 3 shows the hydrodynamic and structural conditions of the testing data. A more detail description of this data set is provided by Bonakdar (2014). As given in Table 3, relative spacing  $S_G/D$  of the test data varies from 0.5 to 3 ( $0.5 \leq S_G/D \leq 3$ ) meaning that both closely-spaced pile groups ( $S_G/D \leq 1.5$ ) and those with largely-spaced pile groups ( $S_G/D > 1.5$ ) were covered.  $KC$  number of tandem arrangement is between 6 and 86.3, while that of side by side arrangement varies from 6 to 27.9. For the  $2 \times 2$  and staggered arrangements, no further tests were conducted apart from those used for the development of the model. Therefore, these two arrangements were not considered for the validation process.

In total, 124 data were used for the validation of the developed M5MT-GP-based model. The scatter diagram of the measured and predicted  $K_G$  values is drawn in Fig. 26 for testing data where the corresponding statistical values are also provided. It can be concluded from this figure that the proposed M5-MT-GP-based model is well validated as it can precisely predict  $K_G$  values for the new testing data which were not used for the development of the model. As seen, the model's agreement index ( $I_a$ ) is 0.975 for 124 test data points which is very close to what obtained for the 485 data used for the development of the model. Likewise agreement index, the scatter index of the testing data is 10.2% which is only 4.4% larger than what calculated for data applied in developing the model.

## 5. Summary and concluding remarks

Based on the knowledge of the processes associated with the interaction of waves and pile groups, which was gained from a series of systematic laboratory investigations (e.g. Bonakdar and Oumeraci 2012, 2014, 2015a, 2015b and Bonakdar 2014), new physically-based and more generic wave load formulae were derived for the prediction of wave loads on a slender pile within a group of piles in different arrangements.

For the analysis of the laboratory data and development of the wave load formulae, an artificial intelligence (AI) - based computational tool, a combination of M5 tree (M5MT) and genetic programming (GP) named "*hybrid M5MT-GP model*", was developed. Pile group effect  $K_G$ , which was set as the output hybrid M5MT-GP model, is a function of (i)  $KC$  number and (ii) pile group arrangement and relative spacing  $S_G/D$  representing the most relevant influencing hydrodynamic and structural parameter, respectively. The developed M5MT-GP model and formulae are summarized in Fig. 27. The concluding remark drawn from the new M5MT-GP model and formulae may be summarized as follows:

- (i) The new AI-based model and formulae are simple, compact, transparent and physically-based and allow us to systematically assess pile group effect  $K_G$  depending on the flow regime ( $KC$ ) and the structural conditions (pile group arrangement,  $S_G/D$ ). As can be concluded from Fig. 27, the proposed formulae needs to be applied for the estimation of the wave-induced force on slender piles within further neighbouring piles unless  $K_G=1$  where each pile in the group can be treated like a single isolated pile.
- (ii) The high accuracy of the new formulae in predicting pile group effect  $K_G$  is confirmed by the values of agreement index  $I_a$ , correlation coefficient  $CC$  and scatter index  $SI$  of the overall M5MT-GP model obtained from the analysis of all 485 train data, which are 0.987, 0.975 and 5.8%, respectively.
- (iii) New data sets were used for the validation of the developed M5MT-GP model. Agreement index  $I_a$ , correlation coefficient  $CC$  and scatter index  $SI$  of 0.975, 0.974 and 10.2% were respectively obtained, indicating a high accuracy of the model and the prediction formulae, even for a set of new data.
- (iv) The developed formulae are valid for non-breaking waves with the range of hydrodynamic and structural conditions as summarized in Table 1.

## **Acknowledgments**

The financial support of the German Research Foundation (DFG, *Deutsche Forschungsgemeinschaft*) for this study through the WaPiGS project (Ou 1/13-1) is acknowledged. The support of the German Academic Exchange Service (DAAD, *Deutscher Akademischer Austauschdienst*) is also acknowledged. The first author is also grateful to Griffith University for providing him a visiting scholar opportunity.

## References

- Apelt, C. J., Piorewicz, J., 1986. Interference Effects on Breaking Wave Forces on Rows of Vertical Cylinders. In Proceeding of 1st Australasian Port, Harbour and Offshore Engineering Conference, Sydney, Australia.
- Azamathulla, H. Md., Guven, A., Demir, Y. K., 2011. Linear genetic programming to scour below submerged pipeline. *Ocean Engineering* 38, 995–1000
- Azamathulla, H. Md, Zahiri, A., 2012. Flow discharge prediction in compound channels using linear genetic programming. *Journal Hydrology* (454–455), 203–207.
- 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. *J. of Hydr. Eng.* 133(4), 440-450.
- Bonakdar, L., 2014. Pile group effect on the wave loading of a slender pile. PhD thesis, TU Braunschweig, Germany (ISBN 978-3-86948-383-2).
- Bonakdar, L., Etemad-Shahidi, A., 2011. Predicting wave run-up on rubble-mound structures using M5 model tree, *Ocean Engineering*, 38, 111-118
- Bonakdar, L., Oumeraci, H., 2012. Interaction of waves and pile group-supported offshore structures: A large scale model study. In proceedings of 22nd International Offshore and Polar Engineering Conference (ISOPE), Rhodes, Greece, pp. 812 – 818.
- Bonakdar, L., Oumeraci, H., 2014a. Small and large scale experimental investigations of wave loads on a slender pile within closely spaced neighbouring piles. In proceeding of 33rd International Conference on Ocean, Offshore and Arctic Engineering (OMAE), San Francisco, USA.
- Bonakdar, L., Oumeraci, H., 2015a. Pile group effect on the wave loading of a slender pile: A summary of laboratory investigations. In proceeding of Pfahl-Symposiums, Braunschweig, Germany
- Bonakdar, L., Oumeraci, H., 2015b. Pile group effect on the wave loading of a slender pile: A small-scale model study. *Ocean Engineering*, (under review).
- Chakrabarti, S. K., 1979. Wave forces on vertical array of tubes Proceeding of Civil Engineering in the Oceans, ASCE, San Francisco, USA, 241-259

- Chakrabarti, S. K., 1980. In-line forces on fixed vertical cylinders in waves. *Journal of Waterways, Port, Coastal and Ocean Div.*, ASCE, 106, 145-155.
- Chakrabarti, S. K., 1981. Hydrodynamic coefficients for a vertical tube in an array. *Applied Ocean Research*. 3, 2-12.
- Chakrabarti, S. K., 1982. Inline and transverse forces on a tube array in tandem with waves. *Applied Ocean Research*. 4, 25-32.
- Elshorbagy, A., Gorzo, G., Srinivasulu, S., Solomatine, D. P., 2010. Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 2: Application. *Journal of Hydrology and Earth System Science*, 14, 1943–1961.
- Etemad-Shahidi, A., Bali, M., 2012. Stability of Rubble-Mound Breakwaters using H50 wave height parameter. *Coastal Engineering*, 59, 38-45.
- Etemad-Shahidi, A., Bonakdar, L., 2009. Design of Rubble-Mound Breakwaters using M5' Machine Learning Method. *Applied Ocean Research*, 31, 197-201.
- Etemad-Shahidi, A., Ghaemi, N., 2011. Model tree approach for prediction of pile groups scour due to waves, *Ocean Engineering*, 38, 522-527
- Etemad-Shahidi, A., Yasa R., Kazeminezhad M. H., 2011. Prediction of wave-induced scour depth under submarine pipelines using machine learning approach. *Applied Ocean Research*, 33, 54–59.
- Güven, A., Azamathulla, H. Md., Zakaria, N. A., 2009. Linear genetic programming for prediction of circular pile scour. *Ocean Engineering*, 36, 985–991.
- Haritos, N., Smith, D., 1995. The effect of spacing transverse to the wave direction on the Morison force coefficients in two cylinder groups. In proceedings of 14th International Conference on Offshore Mechanics and Arctic Engineering (OMAE), Copenhagen, Denmark.
- Hildebrandt, A., Sparboom, U., Oumeraci, H. 2008. Wave forces on groups of slender cylinders in comparison to an isolated cylinder due to non-Breaking waves. In proceeding of 31st International Conference on Coastal Engineering (ICCE), ASCE, Hamburg, Germany.
- Jafari, E., Etemad-Shahidi, A. 2012. Derivation of a new model for prediction of wave overtopping at rubble-mound structures. *J. Waterw., Port, Coast., Ocean Eng.*, ASCE, 138 (1), 42–52.

- Kazeminezhad, M.H., Etemad-Shahidi, A. and Yeganeh-Bakhtiary, A. 2010. An alternative approach for investigation of the wave-induced scour around pipelines, *J. Hydroinformatics*, 12, 51-65.
- Koza, J. R. 1992. *Genetic programming: on the programming of computers by means of natural selection*. MIT Press, Cambridge MA
- Li, Y. Ch., Wang, F. L., Wang, H. R., 1993. Wave-current forces on vertical piles in side-by-side arrangement. In *Proceedings of 3rd International Offshore and Polar Engineering Conference (ISOPE)*, Singapore.
- Mindao, G., Lihua, H., Shaoshu, S., 1987. Experimental study for the wave forces on pile groups due to regular waves, *Proc. 2nd International Conference on Coastal and Port Engineering in Developing Countries (COPEDEC)*, China Ocean Press, Beijing, pp. 1956 – 1965.
- Morison, J.R, O'Brien, M.P, Johnson, J.W, Schaaf, S.A 1950. The force exerted by surface waves on piles. *Petroleum Transactions, AIME*, Vol. 189.
- Nitsure, S. P., Londhea, S. N., Khare, K. C., 2012. Wave forecasts using wind information and genetic programming, *Ocean Engineering*, 54, 61-69.
- Oumeraci, H., 2008. *Vertiefervorlesung Küsteningenieurwesen I*. TU Braunschweig, Germany (in German).
- Quinlan, J. R., 1992. Learning with continuous classes. In *Proceedings of AI'92 (Adams and Sterling Eds)*, World Scientific.343-348.
- Solomatine, D. P., Ostfeld, A., 2008. Data-driven modelling: some past experiences and new approaches. *Journal of Hydroinformatics*, 10 (1), 3-21.
- Smith, D., Haritos, N., 1996. The effect of in-line spacing of two cylinder groups on the Morison force coefficients. In *proceedings of 15th International Conference on Offshore Mechanics and Arctic Engineering (OMAE)*, Florence, Italy.
- Smith, D., Haritos, N., 1997. The influence of grouping on the force characteristics of pairs of vertical surface-piercing cylinders. In *proceeding of 7th International Offshore and Polar Engineering Conference (ISOPE)*, Honolulu, USA.
- Sparboom, U., Hildebrandt, A., Oumeraci, H., 2006. Group interaction effects of slender cylinders under wave attack. In *proceeding of 30th International Conference on Coastal Engineering (ICCE)*, ASCE, San Diego, USA.

- Sparboom, U., Oumeraci, H., 2006. Wave loads of slender marine cylinders depending on interaction effects of adjacent cylinders. In proceedings of 25th International Conference on Offshore Mechanics and Arctic Engineering (OMAE), Hamburg, Germany.
- 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.
- Zdravkovich, M. M., 1977. Review of flow interference effects between two circular cylinders in various arrangements, *Journal of Fluids Engineering*, ASME, 99, 618-633.

## Figure Captions

Fig. 1: The three basic pile group arrangements

Fig. 2: Pile group configurations performed in the LWI wave flume tests for non-breaking waves

Fig. 3: Example of M5 model tree (Developed tree, leaves, classified data and linear equations)

Fig. 4: Programme  $\left[ \exp(x_1/3) - 4x_2 \right]$  in the form of a tree structure

Fig. 5: Genetic operator “Crossover”

Fig. 6: Genetic operator “Mutation”

Fig. 7: Genetic Programming flowchart (after Koza, 1992)

Fig. 8: Overview of one-way coupled M5MT-GP modelling for data analysis and the development of new prediction wave load formulae

Fig. 9: Principle and organisation structure of a hybrid M5MT-GP model (exemplarily for two input variables  $X_1$  &  $X_2$  and an output  $Y$ )

Fig. 10: Developed M5MT model for side by side arrangement and relationship between pile group effect  $K_G$  and  $KC$  number for different  $S_G/D$

Fig. 11: Developed M5MT model for tandem arrangement and relationship between pile group effect  $K_G$  and  $KC$  number for different  $S_G/D$

Fig. 12: Developed M5MT model for  $2 \times 2$  arrangement and relationship between pile group effect  $K_G$  and  $KC$  number for different  $S_G/D$

Fig. 13: Relationship between pile group effect  $K_G$  and  $KC$  number for staggered arrangement

Fig. 14: Overall M5MT model for different pile group arrangements exposed to non-breaking waves

Fig. 15: M5MT-GP model and formulae for side by side arrangement

Fig. 16: M5MT-GP-based formulae for the prediction of pile group effect  $K_G$  for side by side arrangement

Fig. 17: Comparison of predicted and measured  $K_G$  for side by side arrangement

Fig. 18: M5MT-GP model and formulae for tandem arrangement

Fig. 19: M5MT-GP-based formulae for the prediction of pile group effect  $K_G$  for tandem arrangement

Fig. 20: Comparison of predicted and measured  $K_G$  for tandem arrangement

Fig. 21: M5MT-GP model and formulae for  $2 \times 2$  arrangement

Fig. 22: M5MT-GP-based formulae for the prediction of pile group effect  $K_G$  for  $2 \times 2$  arrangement

Fig. 23: Comparison of predicted and measured  $K_G$  for  $2 \times 2$  arrangement

Fig. 24: Overall M5MT-GP model system for different pile group arrangements

Fig. 25: Comparison of predicted and measured  $K_G$  for all 485 data used for the development of M5MT-GP model (train data)

Fig. 26: Validation of M5MT-GP-based model using testing data

Fig. 27: Overview of the new wave load formulae obtained from the overall M5MT-GP model for the different pile group arrangements

## **Table captions**

Table 1: Range of the data used for development of M5MT-GP model and corresponding formulae (train data)

Table 2: Performance of the developed M5MT-GP model for (i) all pile group arrangements, (ii) side by side, (iii) tandem and (iv)  $2 \times 2$  arrangements (train data)

Table 3: Range of the data used for the validation of M5MT-GP-based wave load formulae (testing data)

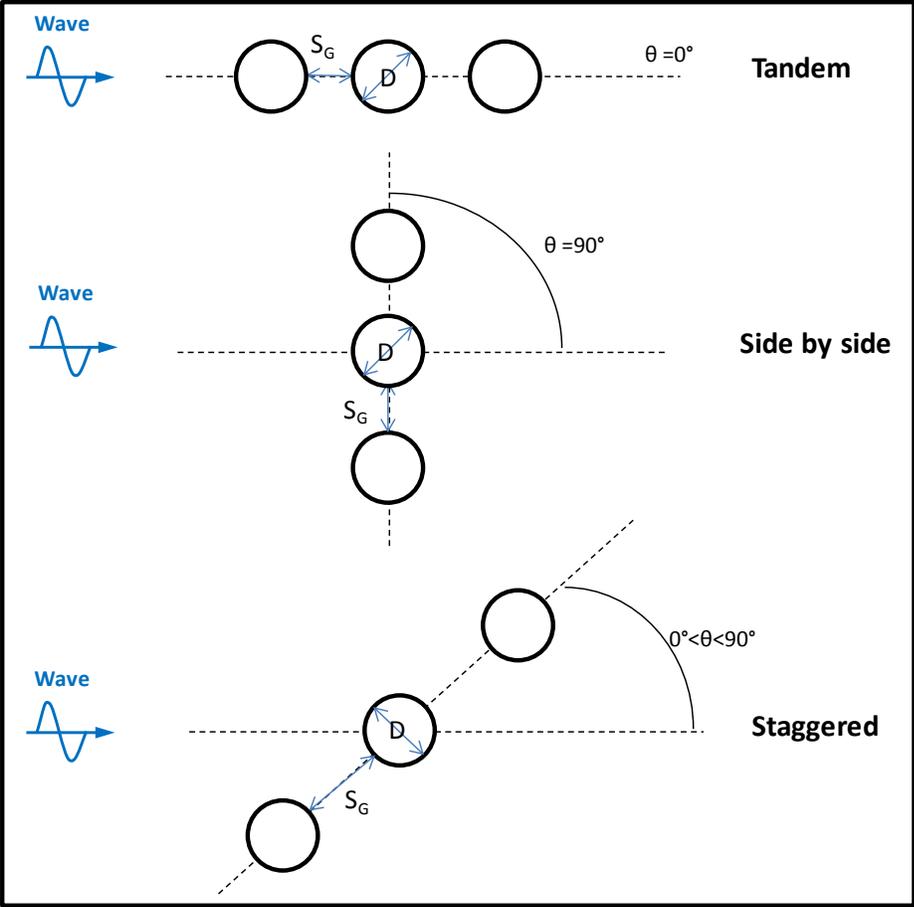


Fig. 1: The three basic pile group arrangements

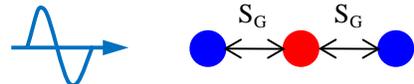
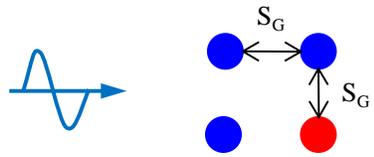
<i>Pile Arrangement</i>		<i>Relative Spacing (<math>S_G/D</math>)</i>	<i>Number of tests</i>
Single		-----	24
Tandem		0.5, 0.75, 1, 1.5, 3, 4 and 5	136
Side by side		0.5, 0.75, 1, 1.5, 2, 3 and 5	146
Staggered		0.6, 0.75, 1, 1.5, 3 and 5	120
2 × 2		0.5, 0.75, 1 and 2	83
			

Fig. 2: Pile group configurations performed in the LWI wave flume tests for non-breaking waves

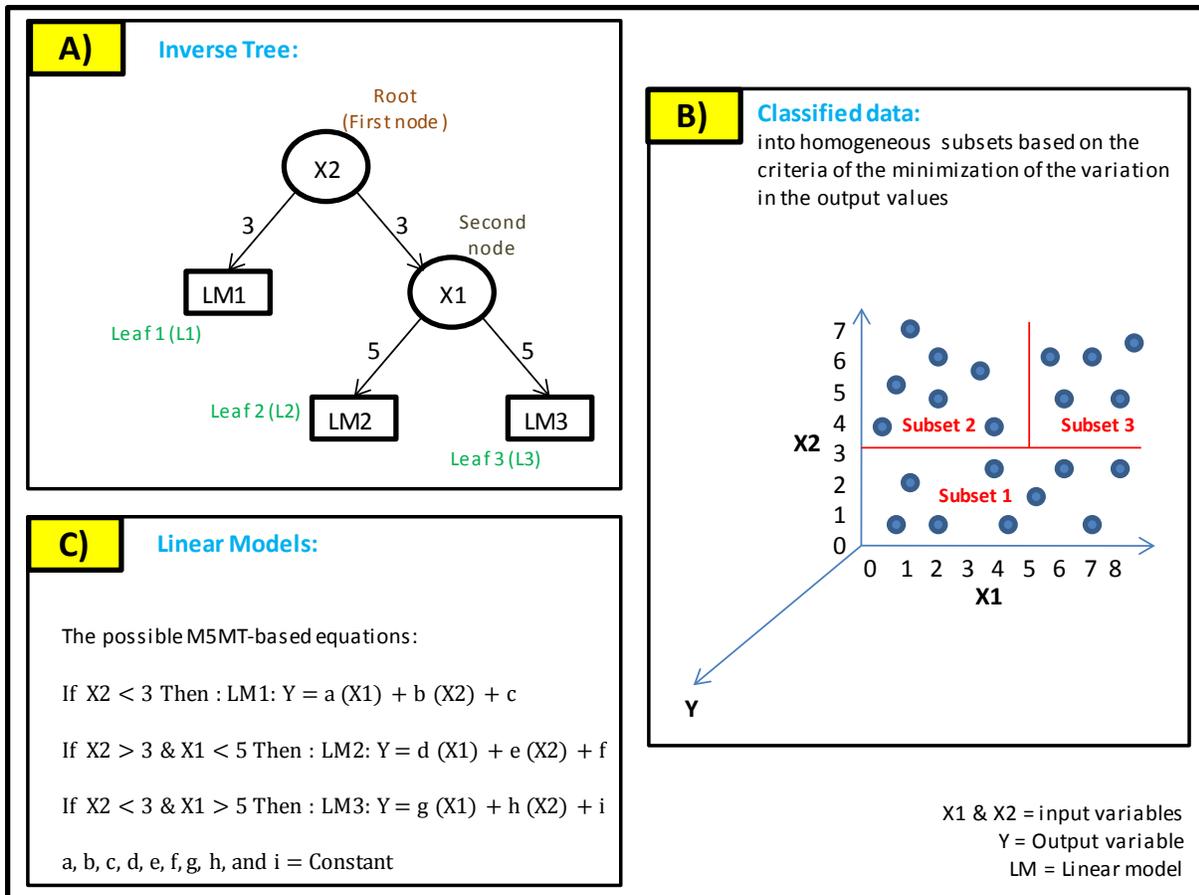


Fig. 3: Example of M5 model tree (Developed tree, leaves, classified data and linear equations)

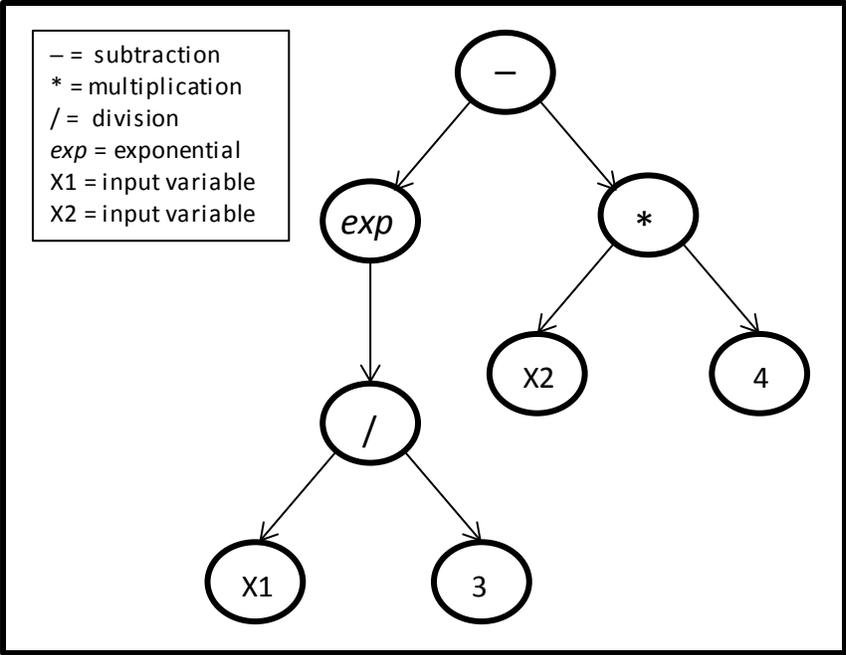


Fig. 4: Programme  $[\exp(x_1/3) - 4x_2]$  in the form of a tree structure

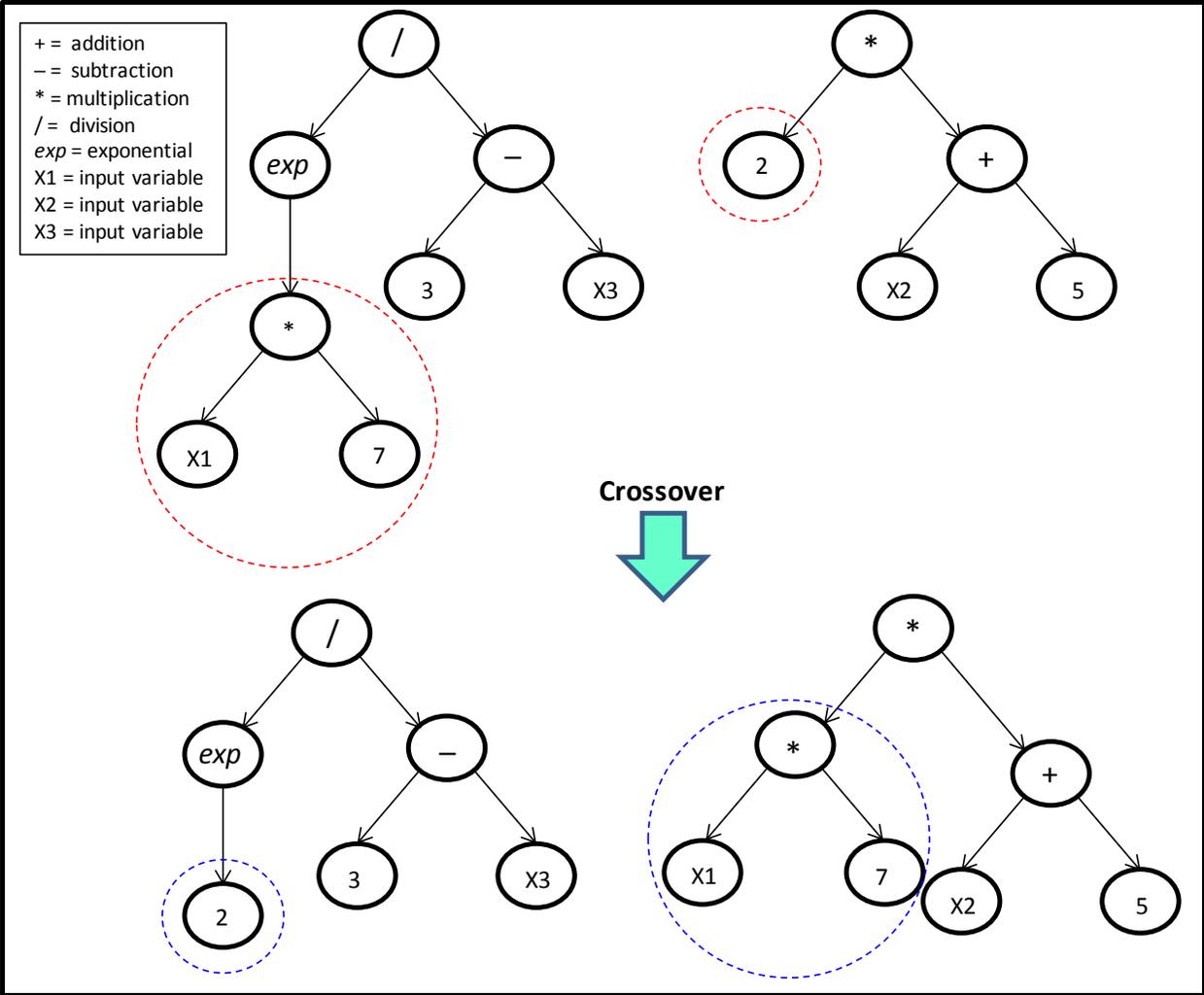


Fig. 5: Genetic operator "Crossover"

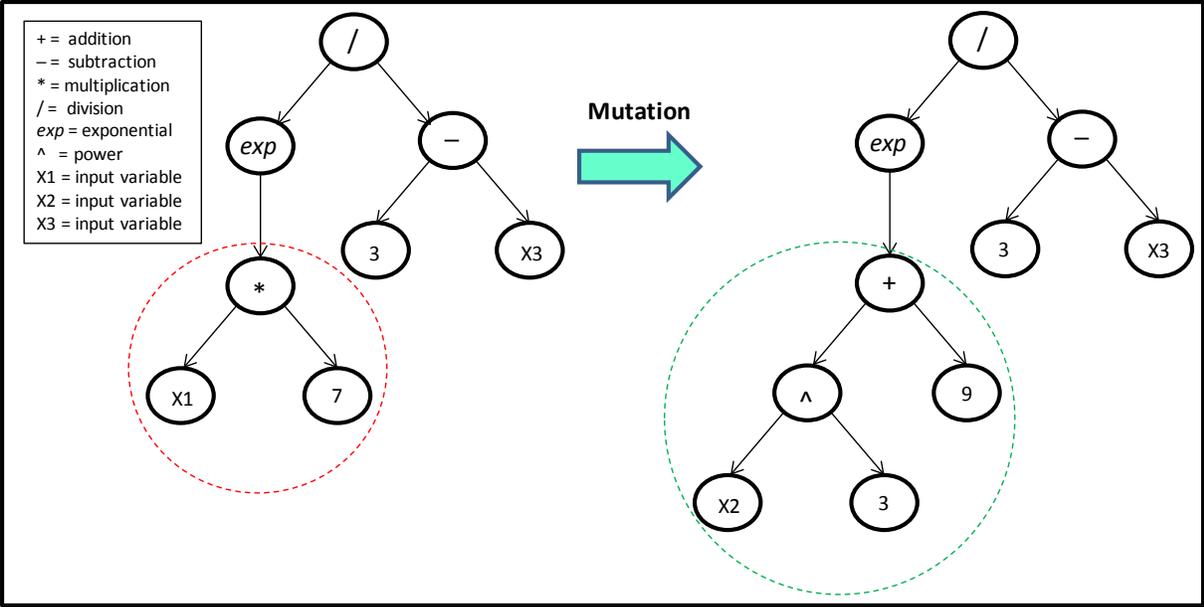


Fig. 6: Genetic operator “Mutation”

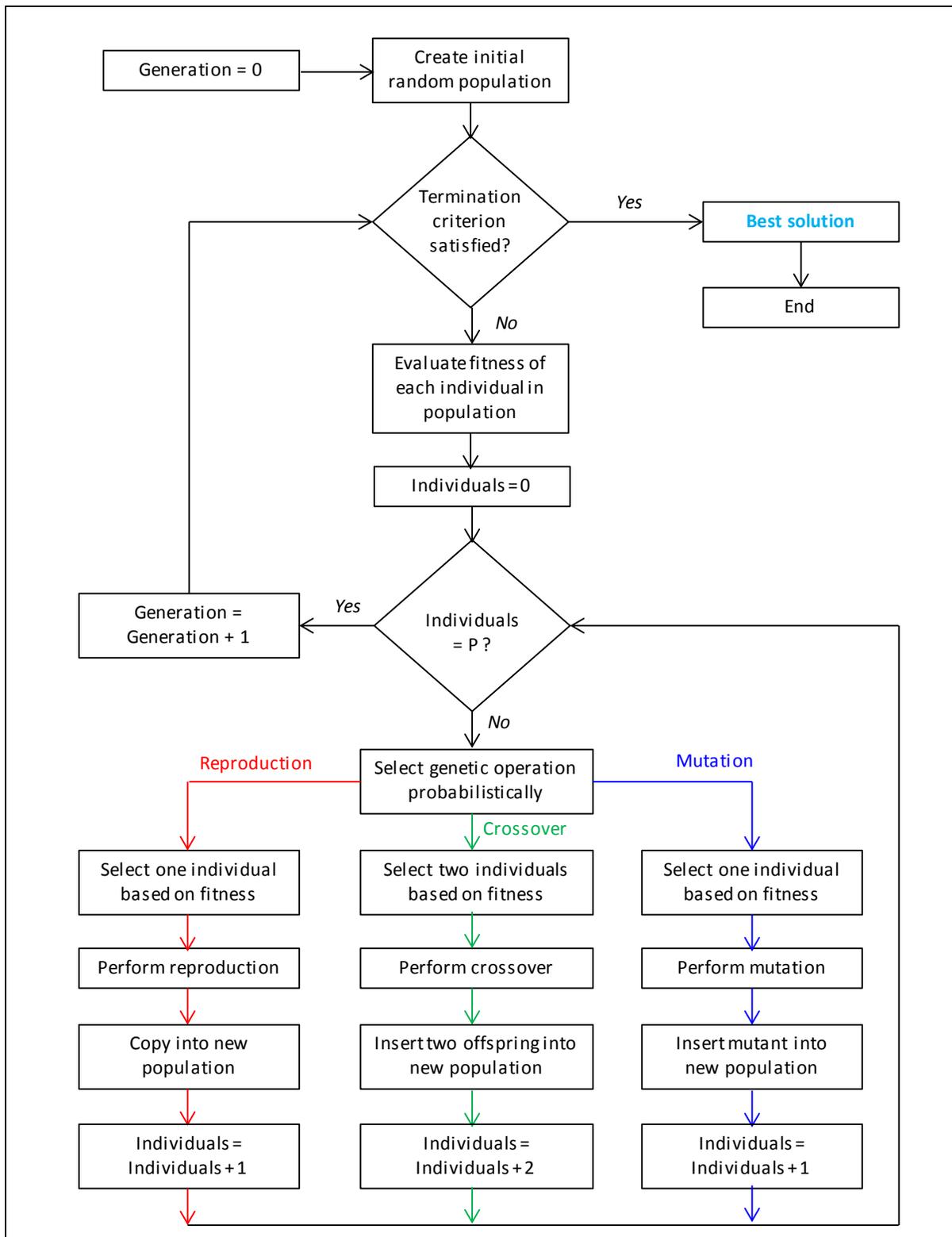


Fig. 7: Genetic Programming flowchart (modified from Koza, 1992)

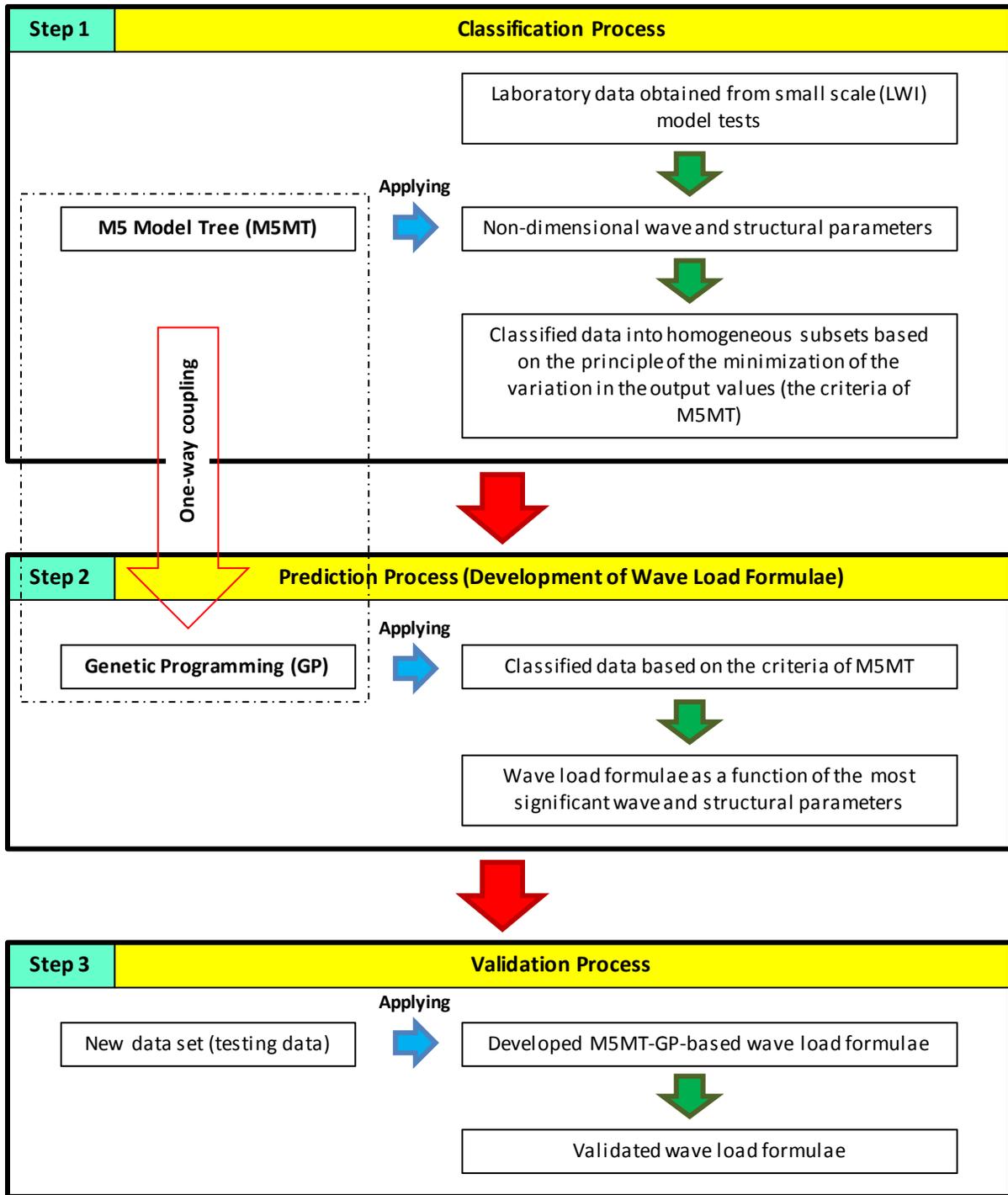


Fig. 8: Overview of one-way coupled M5MT-GP modelling for data analysis and the development of new prediction wave load formulae

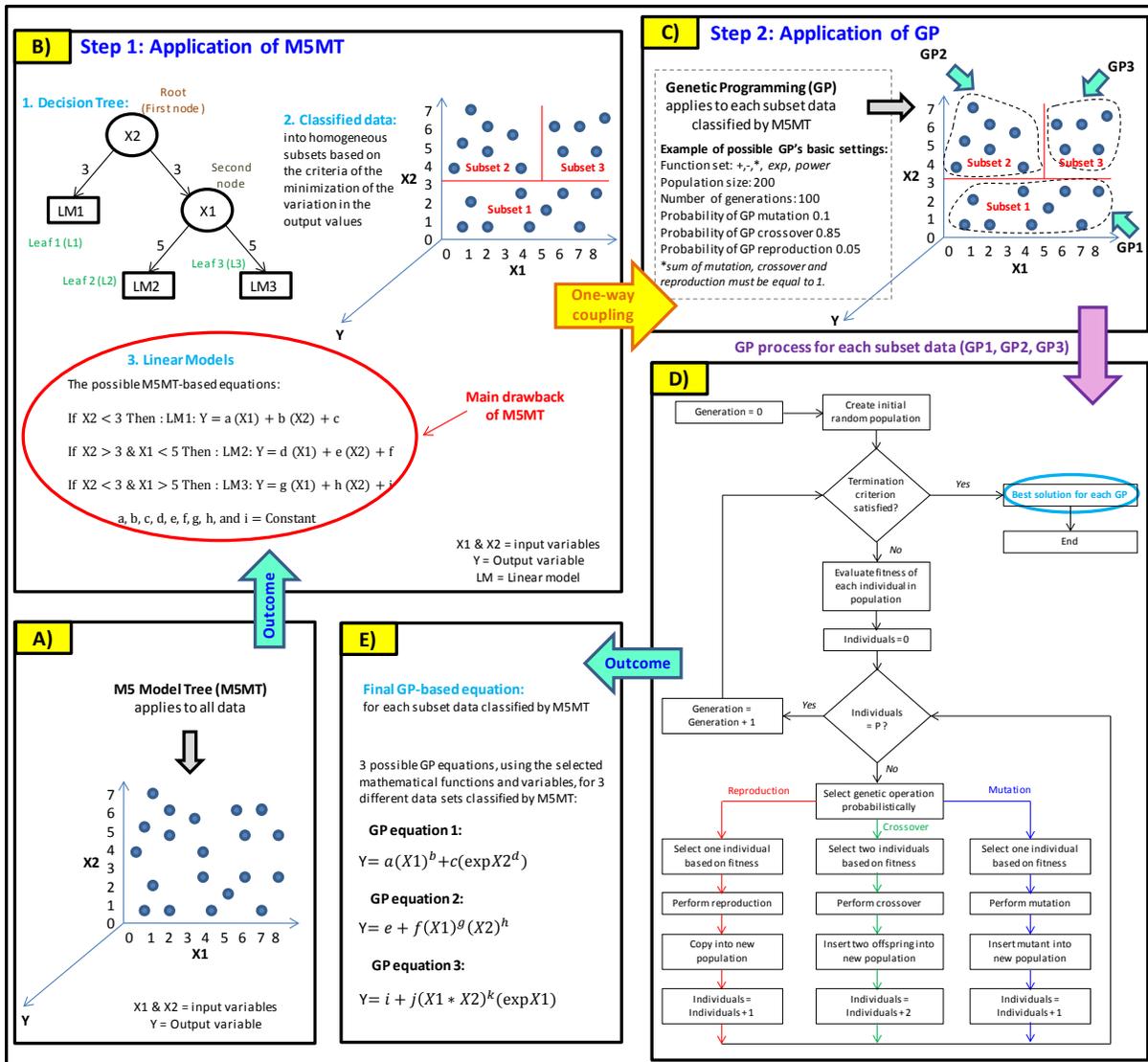


Fig. 9: Principle and organisation structure of a hybrid M5MT-GP model (exemplarily for two input variables  $X_1$  &  $X_2$  and an output  $Y$ )

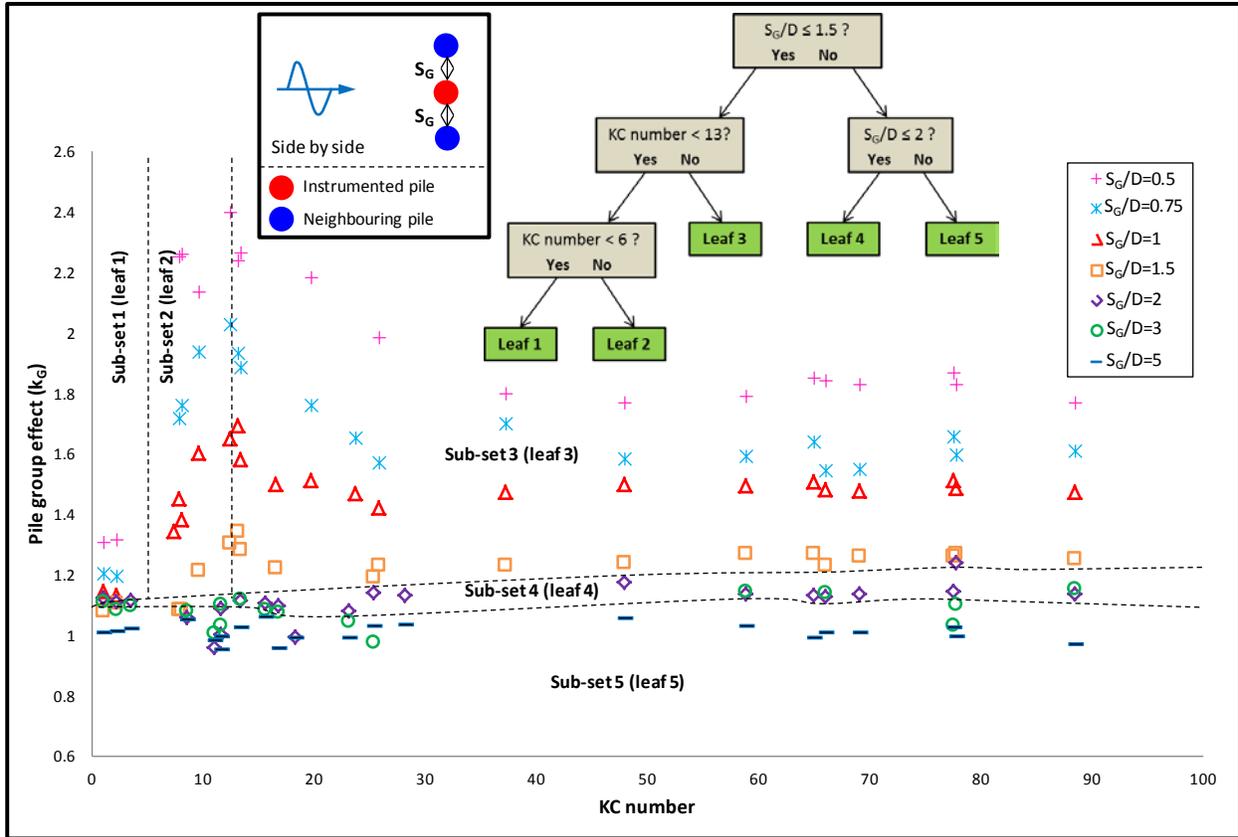


Fig. 10: Developed M5MT model for side by side arrangement and relationship between pile group effect  $K_G$  and  $KC$  number for different  $S_G/D$

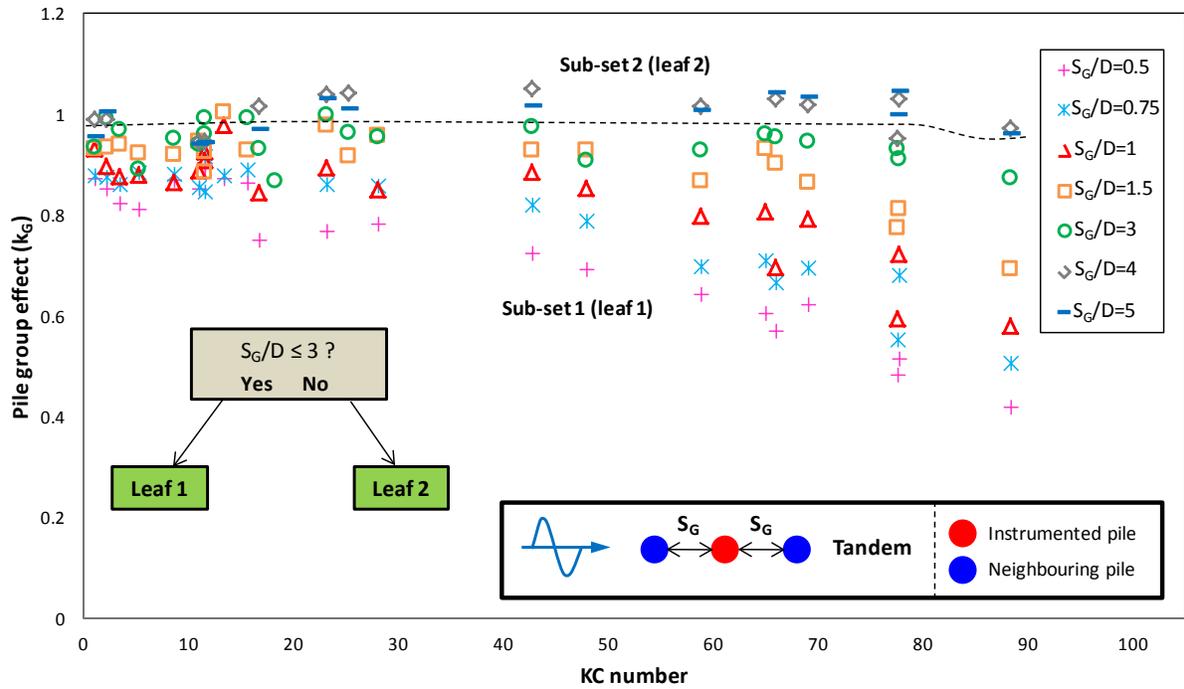


Fig. 11: Developed M5MT model for tandem arrangement and relationship between pile group effect  $K_G$  and  $KC$  number for different  $S_G/D$

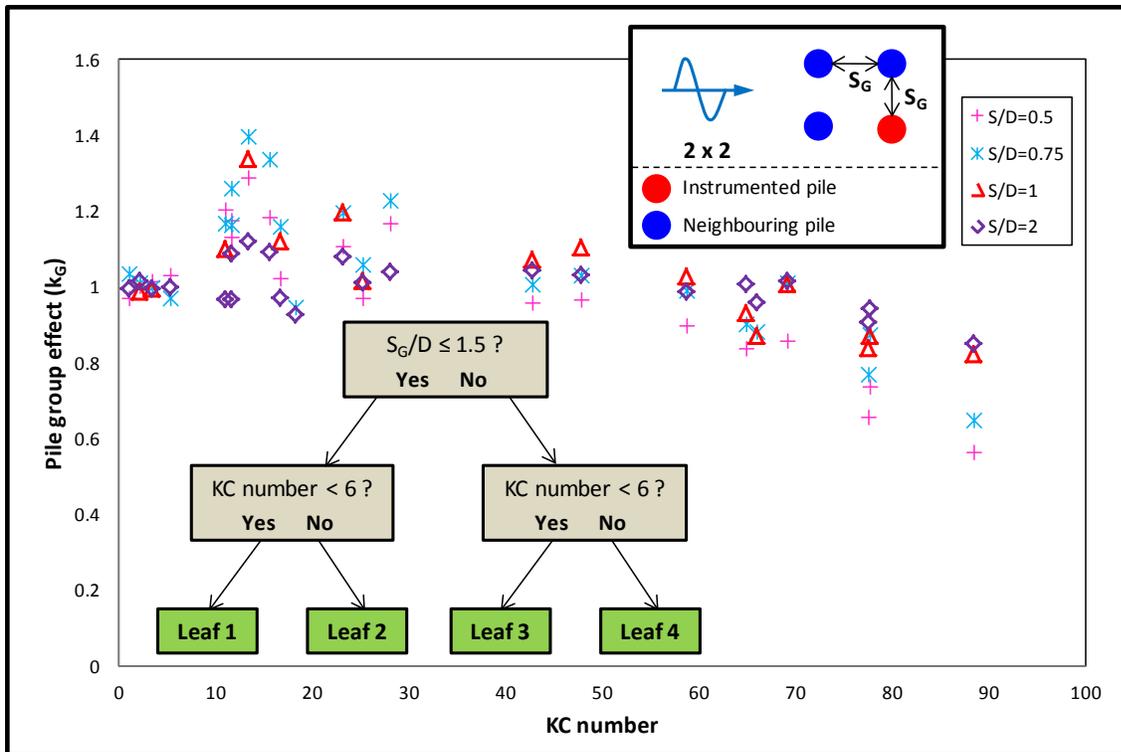


Fig. 12: Developed M5MT model for 2x2 arrangement and relationship between pile group effect  $k_G$  and KC number for different  $S_G/D$

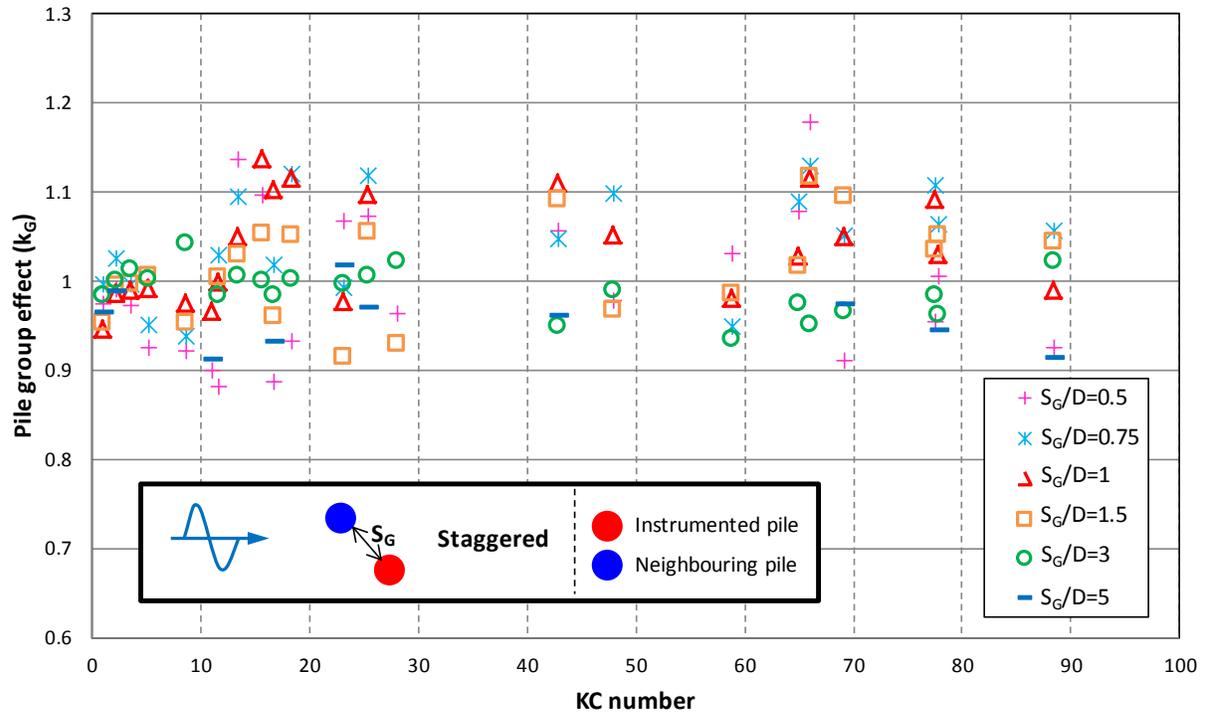


Fig. 13: Relationship between pile group effect  $K_G$  and  $KC$  number for staggered arrangement

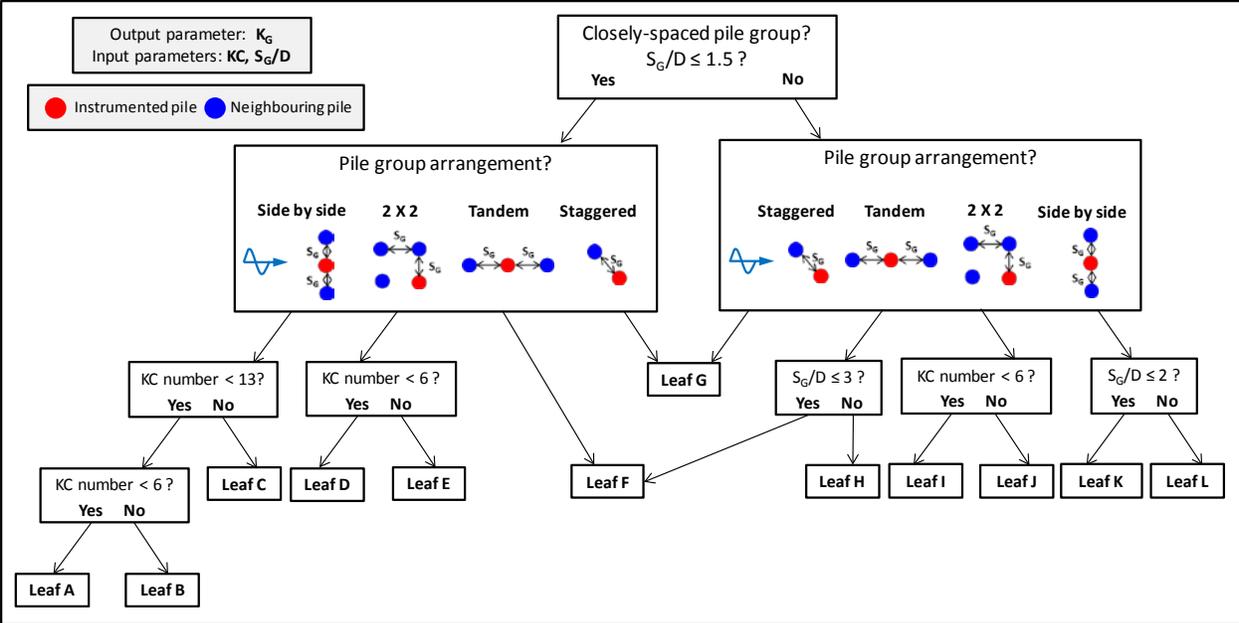


Fig. 14: Overall M5MT model for different pile group arrangements exposed to non-breaking waves

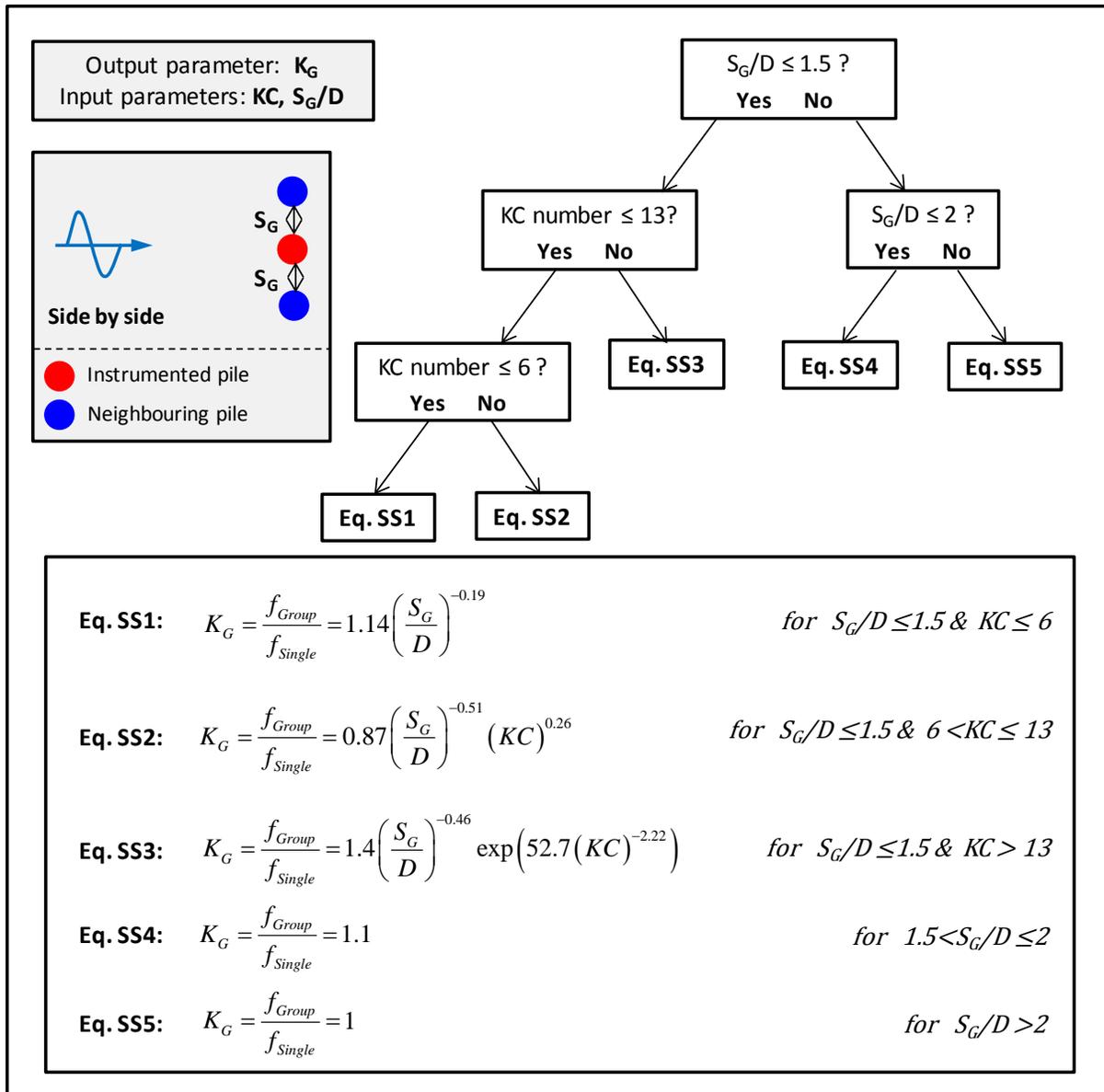


Fig. 15: M5MT-GP model and formulae for side by side arrangement

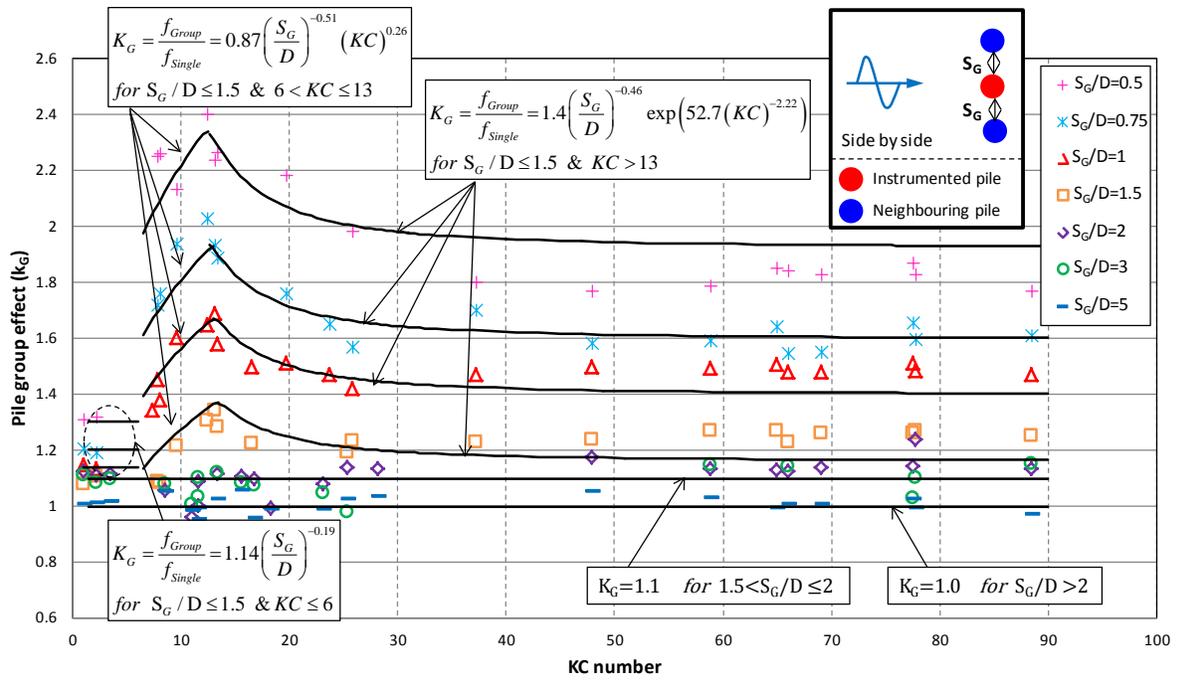


Fig. 16: M5MT-GP-based formulae for the prediction of pile group effect  $K_G$  for side by side arrangement

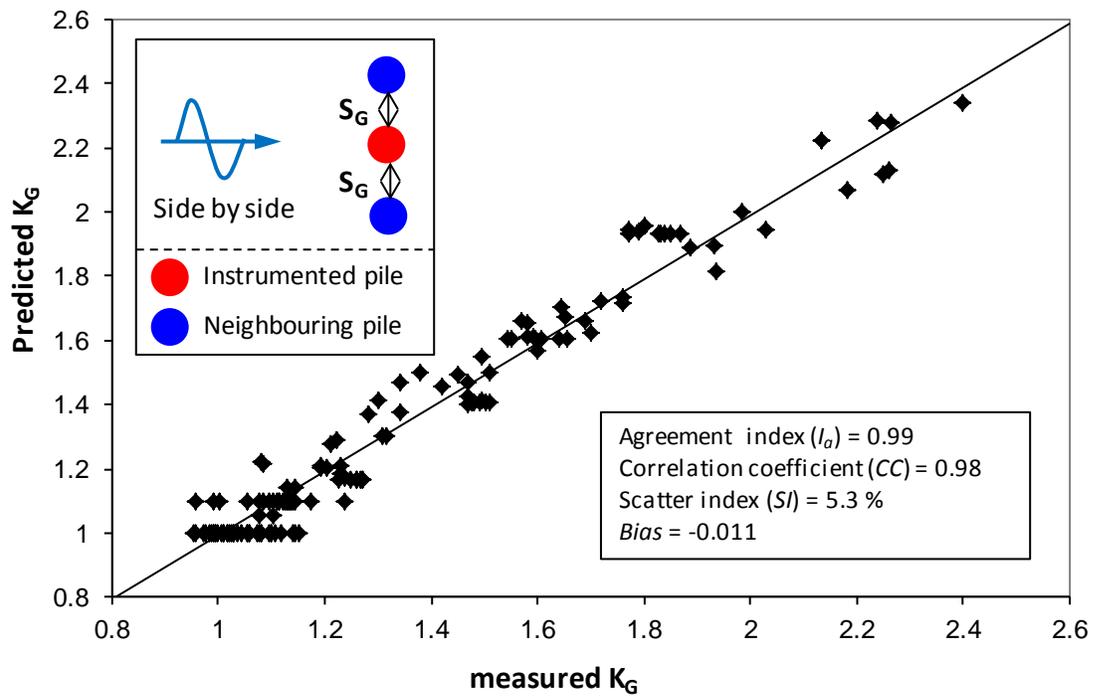


Fig. 17: Comparison of predicted and measured  $K_G$  for side by side arrangement

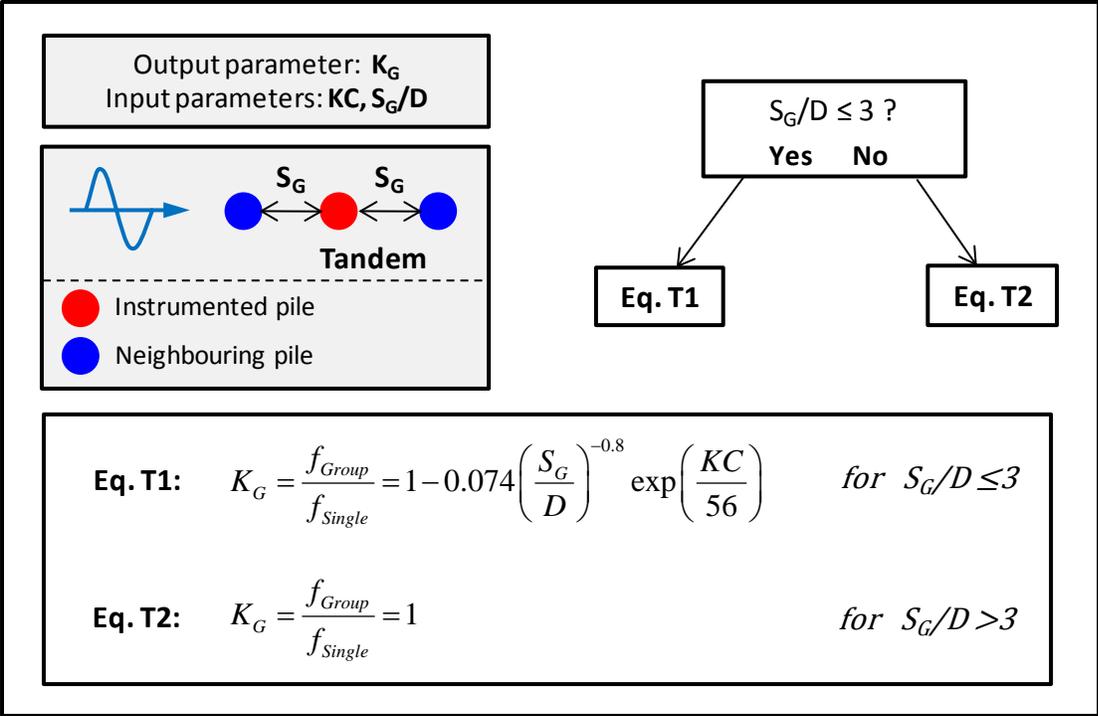


Fig. 18: M5MT-GP model and formulae for tandem arrangement

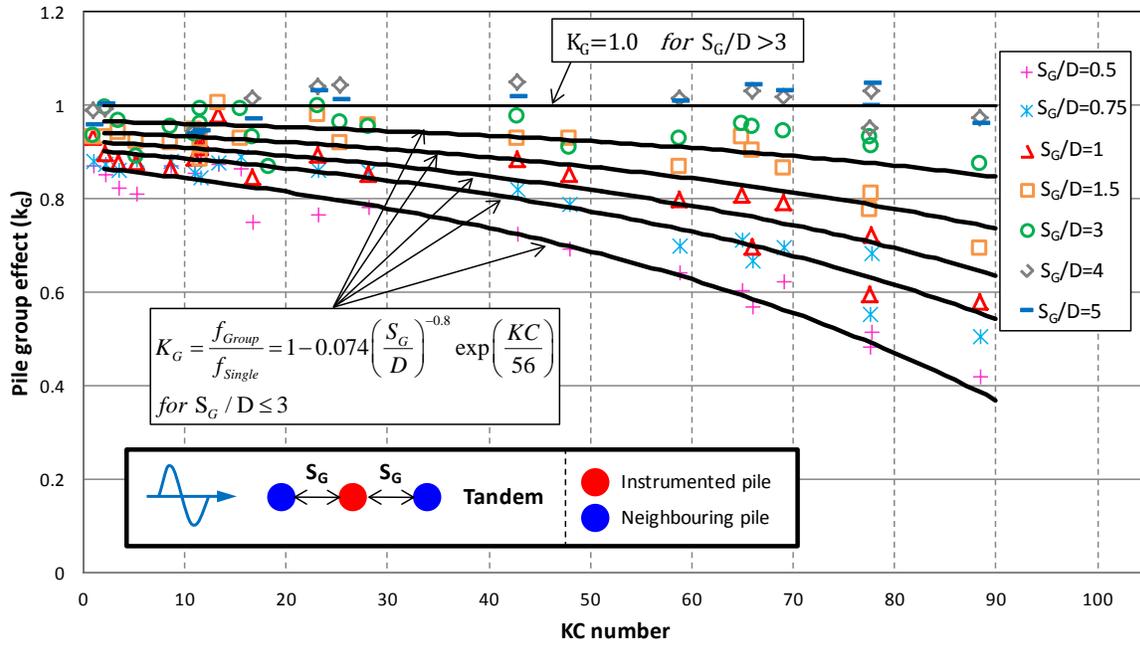


Fig. 19: M5MT-GP-based formulae for the prediction of pile group effect  $K_G$  for tandem arrangement

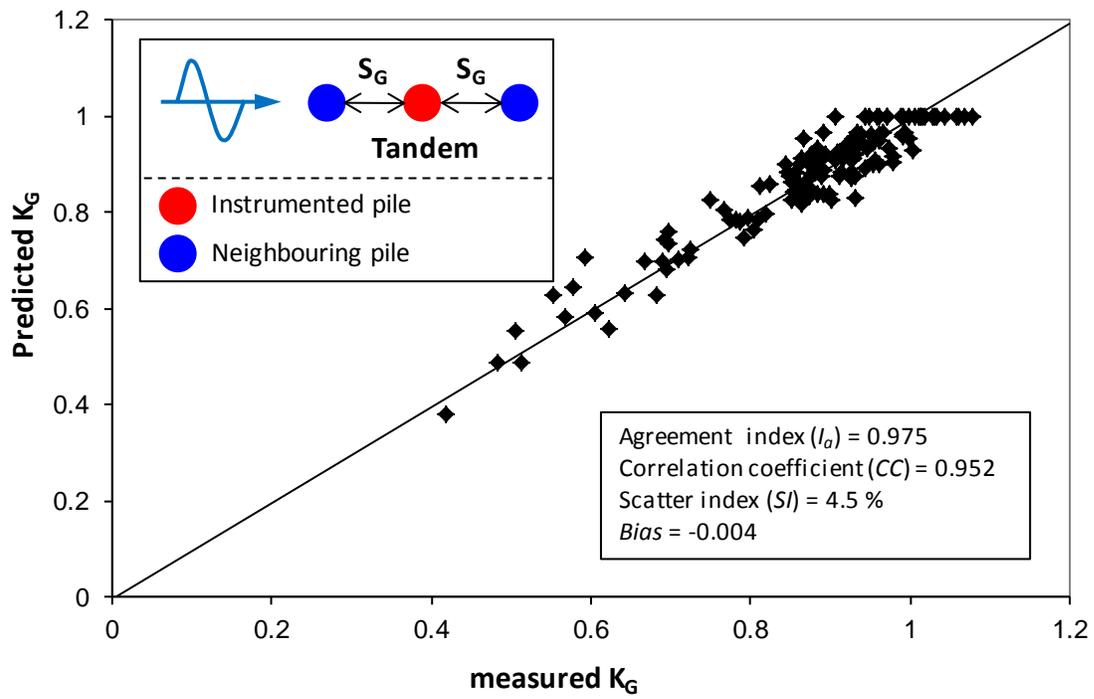
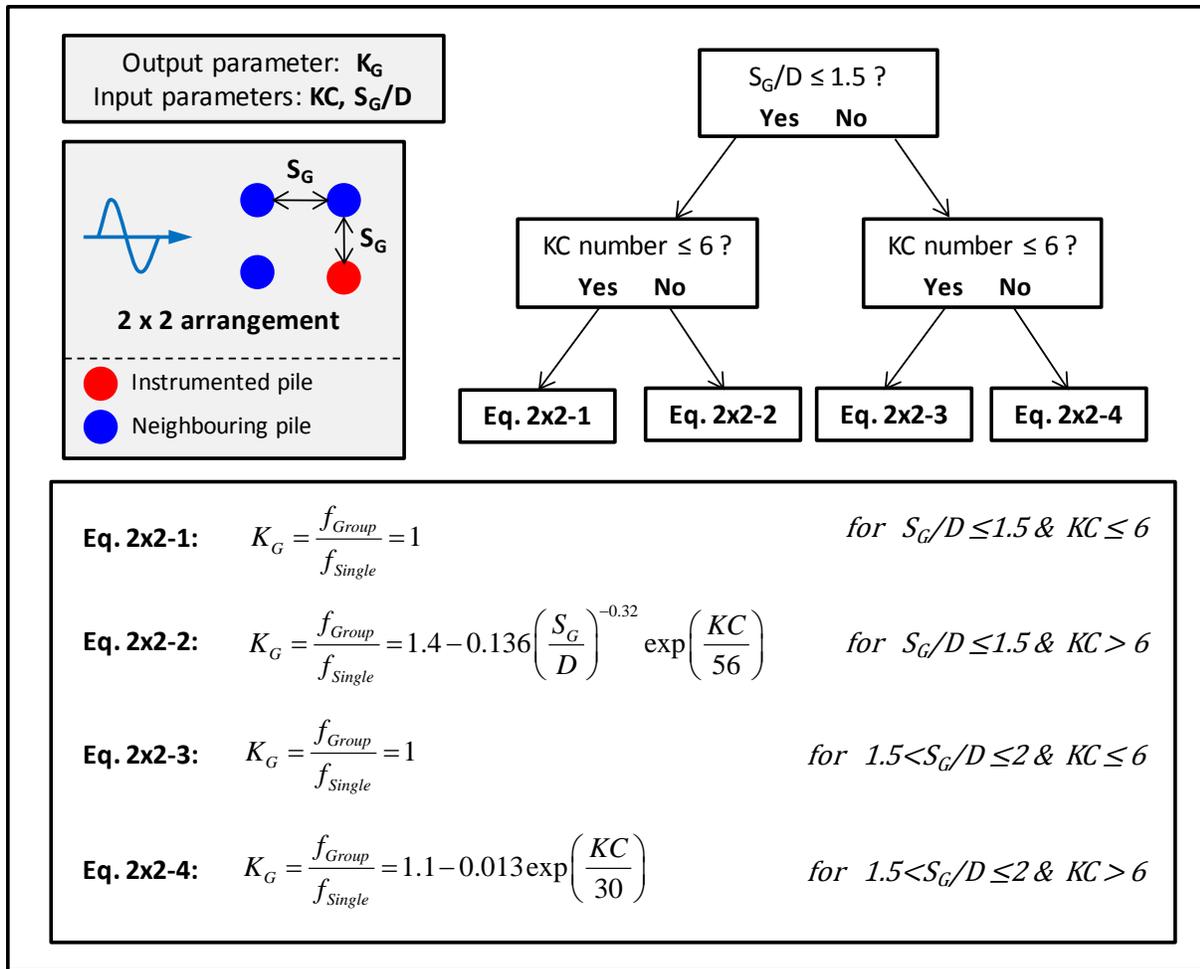


Fig. 20: Comparison of predicted and measured  $K_G$  for tandem arrangement



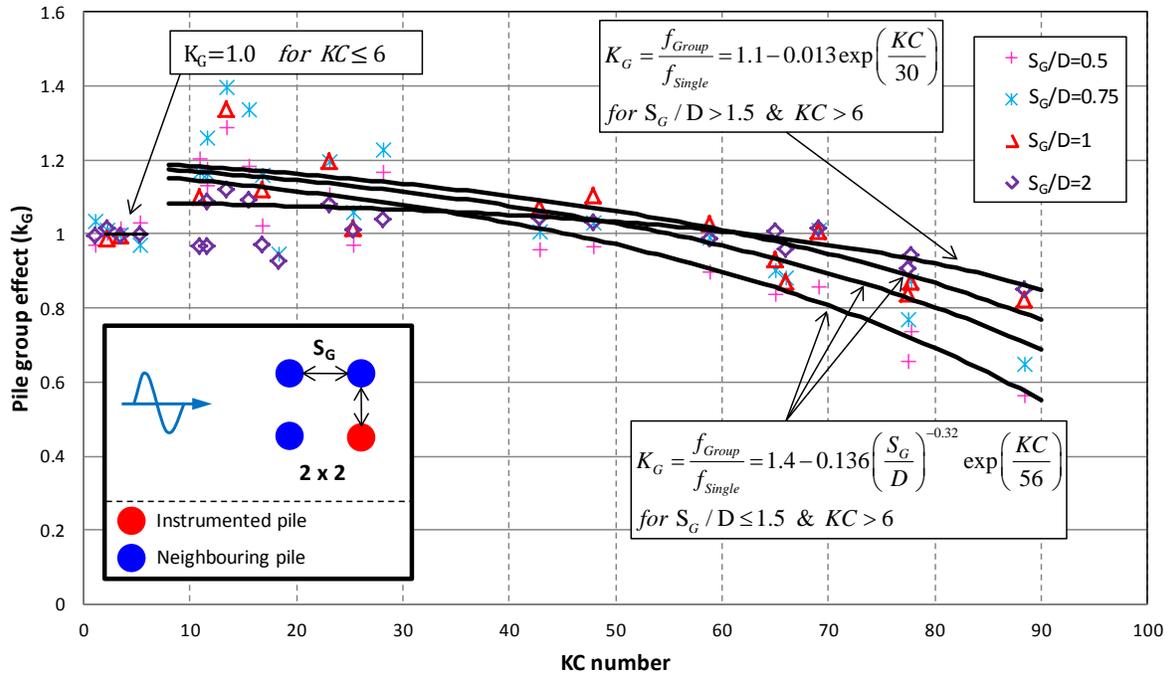


Fig. 22: M5MT-GP-based formulae for the prediction of pile group effect  $K_G$  for 2x2 arrangement

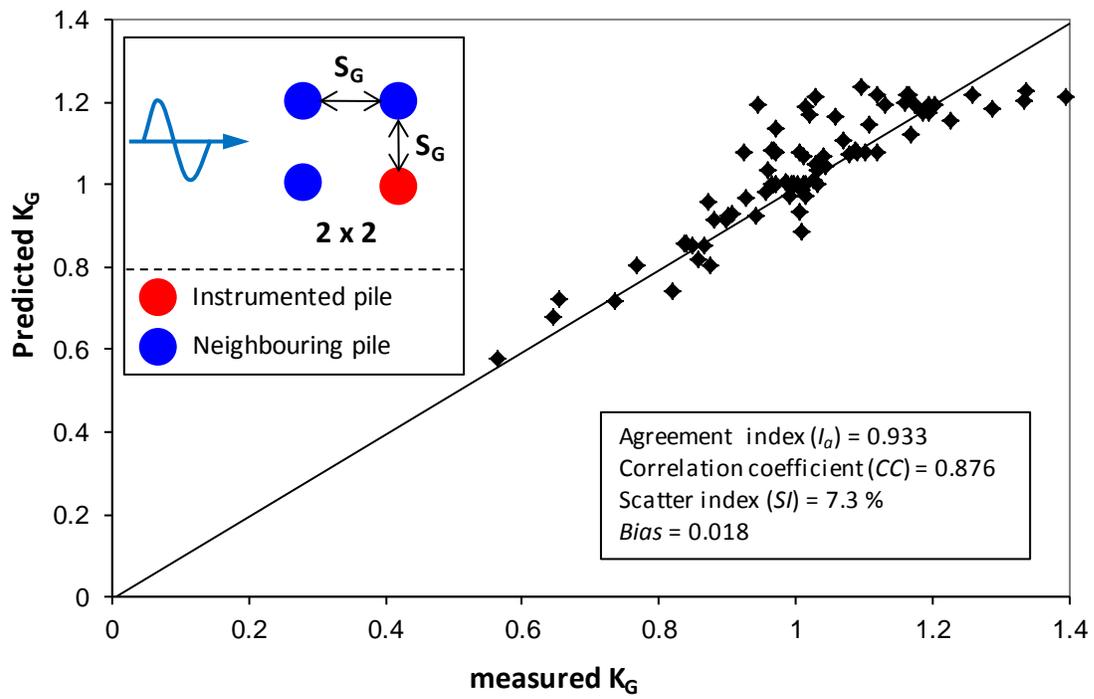


Fig. 23: Comparison of predicted and measured  $K_G$  for 2x2 arrangement

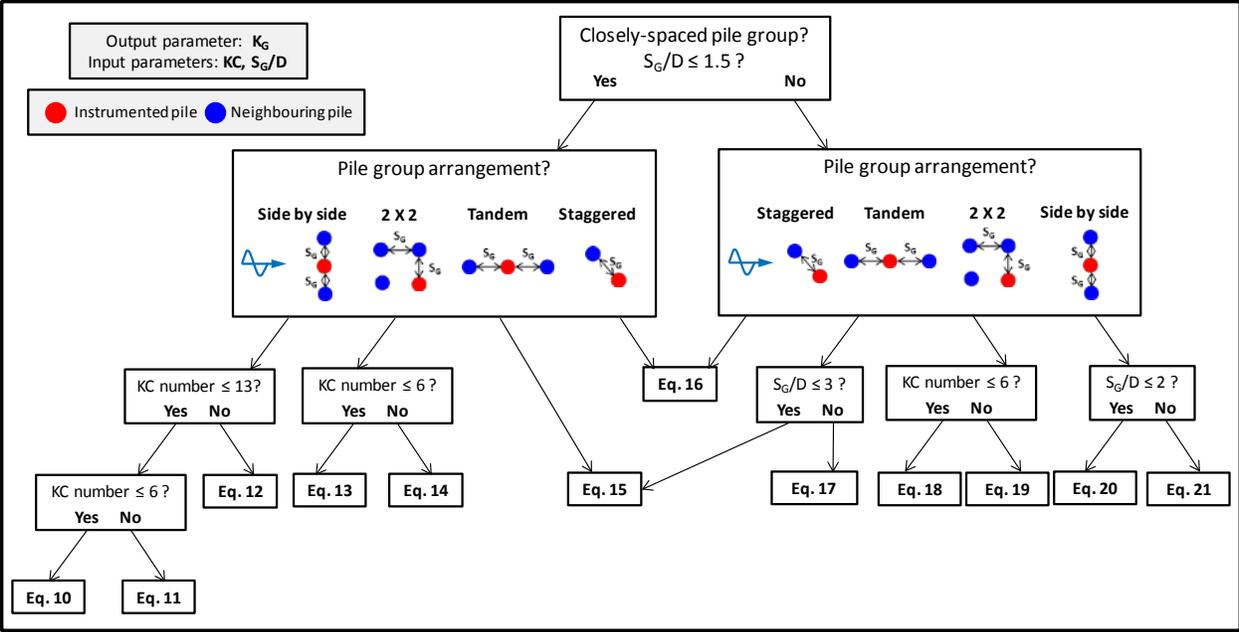


Fig. 24: Overall M5MT-GP model system for different pile group arrangements

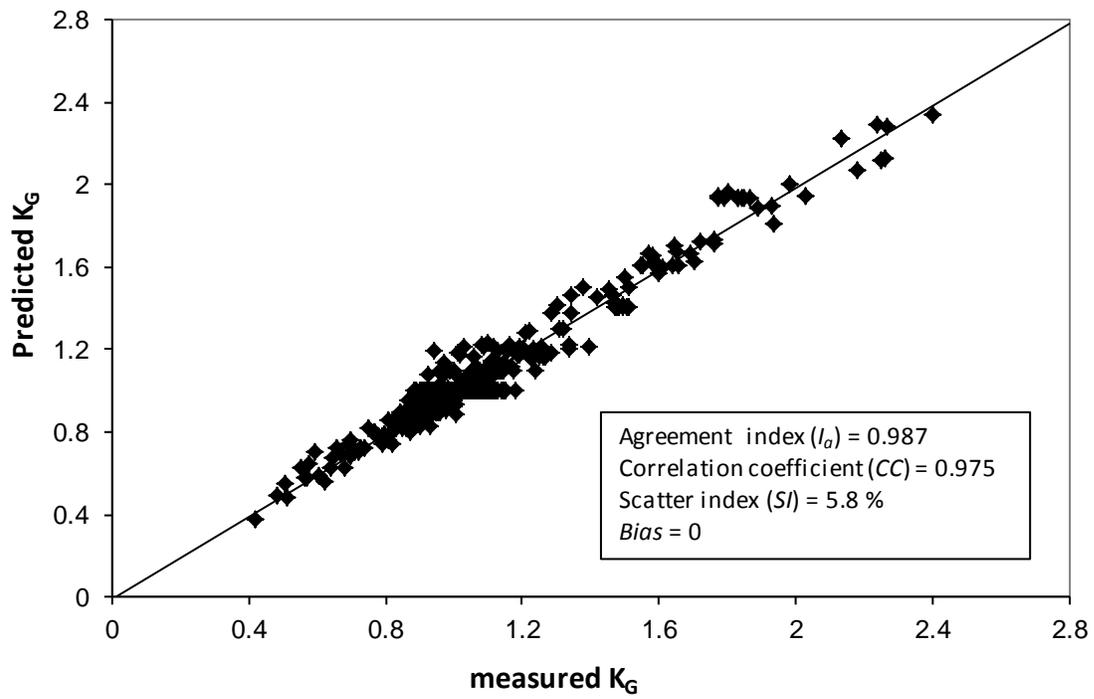


Fig. 25: Comparison of predicted and measured  $K_G$  for all 485 data used for the development of M5MT-GP model (train data)

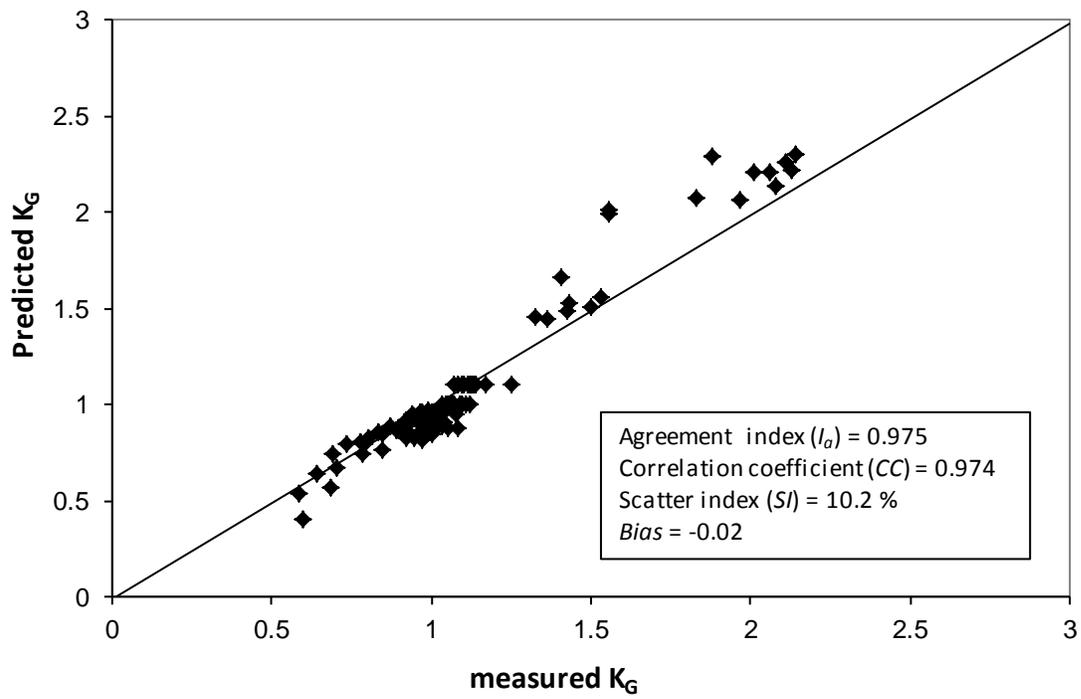


Fig. 26: Validation of M5MT-GP-based model using testing data

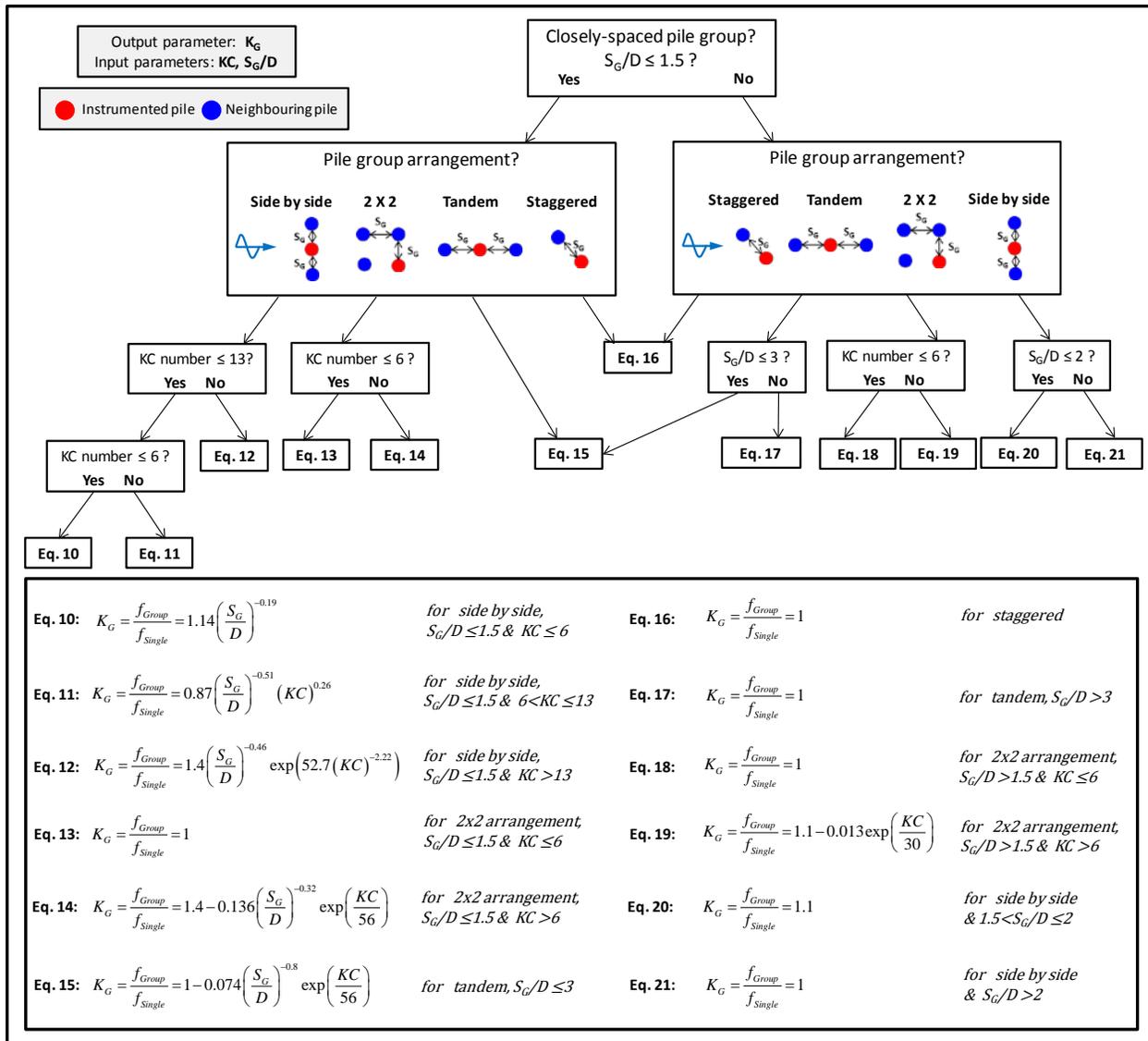


Table 1: Range of the data used for development of M5MT-GP model and corresponding formulae (train data)

<i>Arrangement</i>	<i>Wave type</i>	<i>KC number</i>	<i>Relative water depth (h/L)</i>	<i>Wave steepness (H/L)</i>	<i>Relative spacing (S<sub>G</sub>/D)</i>	<i>Nr. of data</i>
Side by Side	Non-breaking	1.1 ~ 88.5	0.042 ~ 0.64	0.008~0.073	0.5 ~ 5	146
Tandem	Non-breaking	1.1 ~ 88.5	0.042 ~ 0.64	0.008~0.073	0.5 ~ 5	136
Staggered	Non-breaking	1.1 ~ 88.5	0.042 ~ 0.64	0.008~0.073	0.6 ~ 5	120
2×2	Non-breaking	1.1 ~ 88.5	0.042 ~ 0.64	0.008~0.073	0.5 ~ 2	83

Table 2: Performance of the developed M5MT-GP model for (i) all pile group arrangements, (ii) side by side, (iii) tandem and (iv) 2×2 arrangements (train data)

<i>Hybrid M5MT-GP Model</i>	<i>Ia</i>	<i>CC</i>	<i>SI (%)</i>	<i>Bias</i>	<i>Nr. of data</i>
Overall model (all arrangements)	0.987	0.975	5.8	-0.003	485
Side by side arrangement	0.99	0.98	5.3	-0.011	146
Tandem arrangement	0.975	0.952	4.5	-0.004	136
2×2 arrangement	0.933	0.876	7.3	0.018	83

Table 3: Range of the data used for the validation of M5MT-GP-based wave load formulae (testing data)

<i>Arrangement</i>	<i>Wave type</i>	<i>KC number</i>	<i>Relative water depth (h/L)</i>	<i>Wave steepness (H/L)</i>	<i>Relative spacing (S<sub>G</sub>/D)</i>	<i>Nr. of data</i>
Side by Side	Non-breaking	6 ~ 27.9	0.083 ~ 0.2	0.016~0.073	0.5 ~ 3	50
Tandem	Non-breaking	6 ~ 86.3	0.044 ~ 0.2	0.016~0.073	0.5 ~ 3	74