NTNU

Norwegian University of
Science and Technology

Predicting scour around offshore wind turbines using soft computing techniques

Comparing Genetic Programming with
Existing Scour Prediction Methods.

Irma Hetty Centen

Wind Energy
Submission date: September 2015
Supervisor: Dag Myrhaug, IMT

Norwegian University of Science and Technology
Department of Marine Technology

Master Thesis

Predicting scour around offshore wind turbines using soft computing techniques

COMPARING GENETIC PROGRAMMING
WITH EXISTING SCOUR PREDICTION METHODS

I. Centen

04-09-2015

EWEM TU Delft DTU Carl von Ossietzky Universität Oldenburg NTNU European Commission ERASMUS MUNDUS

European Wind Energy Master - EWEM - Offshore Track

Predicting scour around offshore wind turbines using soft computing techniques

Comparing Genetic Programming with Existing Scour Prediction Methods.

Msc. Thesis

As part of obtaining the degree of Master of Science in Offshore Engineering at Delft University of Technology and in Technology-Wind Energy at Norwegian University of Science and Technology.

I. Centen

04-09-2015

European Wind Energy Master - EWEM
Deltares
Norwegian University of Science and Technology
Delft University of Technology

European Wind Energy Master - EWEM
Of
Offshore Track

The undersigned hereby certify that they have read and recommend to the European Wind Energy Master - EWEM for acceptance a thesis entitled **" Predicting scour around offshore wind turbines using soft computing techniques"** by **I. Centen** in partial fulfillment of the requirements for the degree of **Master of Science**.

Dated: 04-09-2015

Supervisor: _____
ir. T. C. Raaijmakers of Deltares

Reader: _____
prof. dr. A. Metrikine of TU Delft

Reader: _____
prof. D. Myrhaug of NTNU

Reader: _____
Dr. E. Lourens of TU Delft

Reader: _____
dr. M.C. Ong of Marintek

Reader: _____
dr. H.F.P. van den Boogaard of Deltares

Summary

When a structure is placed in a river or marine environment, the scour of bed material around the foundation can compromise the integrity of the structure. For design considerations, it is therefore important that the level of scour can be predicted accurately.

In this report a prediction method is developed for scour around monopiles, the most common support structures for offshore wind turbine to date. A soft computing technique called genetic programming (GP) is used to create a scour prediction formula that can compute scour in all offshore conditions. This means a formula that accurately describes current-induced, wave-induced and combined current- and wave-induced scour. The GP was trained with an extensive database of laboratory scour measurements from multiple sources, to ensure that a wide range of conditions was represented. Furthermore, only dimensionless parameters were used to create a formula that is also applicable for field tests.

Applying different settings and input parameters for the GP, first successfully a formula was found to describe current-only conditions followed by a formula for wave-only conditions. These formulas where combined into one equation to predict scour in all hydrodynamic conditions. The formulas where analyzed both on their mathematical and physical behavior and it was concluded that they could accurately predict scour in all conditions.

The most important parameters that describe the scour phenomena were determined with multiple types of parameter sensitivity analyses. An important finding was the paramount significance of the sediment gradation in predicting current-induced scour. This study revealed that this parameter determines the height and sharpness of the so called clear-water peak.

Furthermore it was found that for wave-only conditions a threshold value could be seen for the offset of scour, at Keulegan-Carpenter numbers equal to 4. Other interesting finds were that the maximum scour depth is limited at approximately 2.2 times the pile diameter and not by 1.3 or 1.5 as commonly suggested.

The new scour prediction method was compared to various existing scour prediction methods to observe if improvements have been made. It was seen that the formula

created in this study predicted more accurate scour depths, especially for test with larger scour depths.

This study was finalized with a comparison to a second soft computing method: the neural network. For the same database and input parameters, the scour depths were predicted with the NN. It was found that the GP is less successful in predicting the scour depth compared to the NN. However, the high accuracy of the NN could not have been achieved without the knowledge of the parameter behavior obtained by the GP.

Contents

List of Figures

List of Tables

Nomenclature

Latin Symbols

$\bar{y}$	Statistics: The mean of the observed data	–
$\hat{y}$	Statistics: Predicted data point	–
D/d_{50}	Sediment coarseness	–
D/h_w	Relative pile diameter	–
D/L	Diffraction parameter	–
d_*	Sedimentological diameter	–
d_{50}	Mean bed grain size	m
f_p	Peak wave frequency	$1/s$
f_w	Wave friction factor	–
Fr	Froude number	–
Fr_w	Froude wave number	-
h_p/h_w	Relative pile height	–
H_r	Representative wave height for irregular waves	$[m]$
H_s	Significant wave height	m
h_w	Water depth	m
K_1	Pile shape correction factor	–
K_2	Approach flow angle of attack correction factor	–
K_3	Bed form correction factor	–
K_4	Gradation and coarse fraction size correction factor	–
K_i	Threshold velocity correction factor	–
KC	Keulegan-Carpenter number	–

kr	Roughness current-related	m
L	Wave length	m
mob	Relative mobility	–
n	Statistics: The total number of observed data points	–
N_s	Sediment number	–
Re_D	Cylinder Reynolds number	–
s	Specific gravity	–
SB	Relative suspension number	–
T_p	Peak wave period	s
T_z	Wave zero crossing period	$[s]$
T_{char}	Characteristic time	s
u_w	Wave orbital velocity	m/s
$u_{c,cr}$	Critical current velocity	m/s
u_{crit}	Relative critical velocity	-
u_{ini}	Threshold velocity scour	m/s
u_{rel}	Relative velocity	–
w_s	Grain fall velocity	m/s
y_i	Statistics: Observed data point	–
${w_s}^*$	Dimensionless particle fall velocity	-
CoD	Coefficient of Determination (CoD)	–

Greek Symbols

α	Angle between current and waves	$^\circ$
Δ	Relative density	–
δ	Boundary layer thickness	m
γ	Peak enhancement factor JONSWAP	–
λ	Reproduction size GP	–
μ	Dynamic viscosity	kg/ms
μ	Population size GP	–
ν	Kinematic viscosity	m^2/s
ρ_w	Water density	kg/m^3
σ	Sediment gradation	–
σ	Sediment gradation	-
τ_c	Bed shear stress; current	N/m^2
τ_m	Bed shear stress; mean	N/m^2
τ_w	Bed shear stress; wave	N/m^2
τ_{cr}	Critical bed shear stress	N/m^2

τ_{cw}	Bed shear stress; current and wave	N/m^2
θ	Shields parameter	–
θ_{cr}	Critical Shields parameter	–

Chapter 1

Introduction

When a structure is placed in a river or marine environment, the erosion of bed material around the foundation can cause significant influence on the integrity of the structure. This particular form of erosion is defined as scour and is considered to be an important and at times even catastrophic phenomenon in the field of hydraulic engineering. That the topic is highly relevant, but not yet completely understood, is illustrated by the study of Briaud [1], which states that scour is responsible for over 60% of the bridge failures in the U.S. between 1970 and 2000. In case of marine structures the effects that cause scour are even more complex, since unlike river scour marine scour is not caused by currents alone, but by currents or waves or a combination thereof. With the offshore market growing fast, also the need to improve the understanding of marine scour is increasing.

This report deals with scour around a specific type of marine structure: the monopile. This is the most common support structures for offshore wind turbine to date. Erosion threatens the stability of the wind turbine and influences the resonance frequency, which can cause fatigue at certain wind and wave vibrations. As with other marine structures, it is therefore important that an adequate scour management strategy is developed. In the field of offshore wind it is customary to mitigate the effects of scour by installing scour protection. Scour protection can be a significant expense. According to Zaaijer and Tempel [2], it can cost as much as €350,000 per turbine. Another option of handling wind turbine scour is to allow the erosion to exist and to compensate the reduction of stability by an increase in pile length and thickness. The expenses and feasibility of this option depend highly on the depth of the scour hole.

Cost reductions are an important motivator for offshore wind. To be able to design for the occurrence of scour in the most cost effective way, one needs to improve the understanding of scour and use this to accurately predict the maximum depth of the scour hole for the prevailing conditions. Numerous studies have been executed over the past century with this goal, resulting in various empirical scour prediction models with different degrees of accuracy. However, improvements can still be made and a new model with enhanced performance will benefit the offshore wind sector.

1.1 Research problem

A scour prediction model for offshore wind turbines has to be able to deal with a wide range of hydrodynamic conditions (i.e. current-only, waves-only and combined current and waves). Additionally it needs to be able to cope with different environmental conditions to make accurate site specific estimates. Since the actual processes that govern scour are still not completely understood, most existing scour prediction formulas are based on empirical relations and often deal only with very specific situations (e.g. current-only or waves-only). Furthermore, most current formulas are only calibrated for narrow ranges of input parameters.[3]

Since scour is such an important issue in the design of the wind turbine foundation, there is need for a novel prediction model that does not have the above mentioned limitations. The solution is sought in a soft computing technique called genetic programming (GP). This is a relatively new approach towards computational modelling, which has proven itself useful in similar engineering problems, such as wave transmission behind breakwaters (Panizzo[4]) and flood routing (Abebe and Price[5]), i.e. problems that consisted of a large number of parameters and a simple fitness function (in this case; the size of the error in the scour prediction).

1.2 Research objectives

The aim of this master thesis is to use genetic programming to create a scour prediction model that can compute current-induced, wave-induced and combined current and wave induced scour. Furthermore, the important parameters in the process of scour will be established as well as the range of applicability of the model. The new scour prediction method will be compared to existing scour depth formulas to determine if improvements have been made. The overall aim of this study is to develop a scour prediction method that can be used by the offshore wind energy sector to improve the management of scour, contributing to a reduction in cost of offshore wind.

1.3 Research questions

The main research questions are:

1. Is it possible to obtain an improved scour prediction model with the aide of a genetic program compared to existing empirical equations?
2. Is it possible to obtain one model for current-only, wave-only and current and waves combined situations?
3. What parameters are necessary to accurately describe the process of scour?

And if such a model from research question 1 and 2 is found:

4. What is the validity range and the error of the model?

5. Does the formula behave according to know physical principles?

And lastly:

6. How does the GP compare to another soft computing technique: a neural network model?

1.4 Report outline

This report starts with an introduction to scour in chapter 2. In chapter 3 and 4 the available database for the genetic program is documented, divided into a data chapter and a parameter chapter. The theory behind genetic programming can be found in 5 after which the interim results of the genetic program are presented in 6. The results of the GP can be found in chapter 7, followed by a comparison with existing methods in 8. The final chapters consist of the conclusion and recommendations (chapter 9).

Chapter 2

Scour

This chapter is an introduction to the study of scour. It is a complex subject to which an avalanche of studies and books are dedicated. Much more can be said on the subject than just these few pages, but since the goal of this thesis is to find a suitable scour formula, this chapter will focus on the main effects that cause scour, the parameters commonly used to describe these effects and an overview of some of the currently available scour prediction methods.

2.1 Scour; an introduction

In short, scour is a special name for sediment erosion around a structure. The interaction between the flow and the structure causes transport of the sediment at the bed. The amount of sediment that is transported is a function of many parameters, among others the velocity of the flow and the sediment properties.

Sediment transport does not necessarily mean that there will be scouring around the structure. Only when the local sediment transport exceeds the sediment supply from upstream scour erosion will occur. This difference between supply and demand can occur due to a difference in velocity or in turbulence or in both, which generally is the case when a pile is placed in the marine environment. Figure 2.1 shows a schematic representation of the interaction between flow and cylinder. In figure (b) an example of the top view of scour pit can be seen.

The time that it takes for a substantial amount of scour to develop is called the time scale of the scour process. It can be defined in several ways, usually according to the following equation [7]:

$$\frac{S(t)}{S} = 1 - \exp\left(-\frac{t}{T_{char}}\right) \tag{2.1}$$

T_{char} is the characteristic time, i.e. the dashed line tangent to the scour depth curve at $t = 0$ in figure 2.2. After a certain amount of time the scour depth will not increase

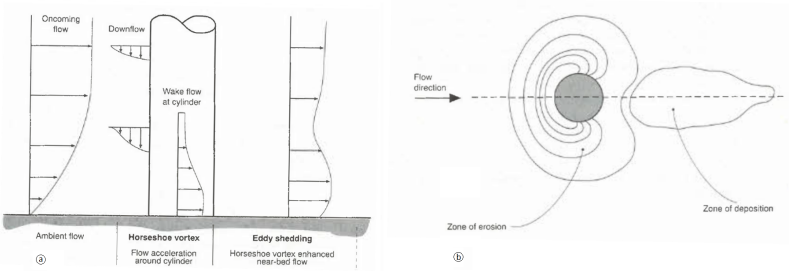


Figure 2.1: *Schematic representation scour phenomena (Whitehouse[6])*

anymore. As is visible in 2.2 for $t \to \infty$ the scour depth approaches an equilibrium value. This value is the target value for most scour prediction methods. In this report the equilibrium scour depth will be expressed as a dimensionless parameter S/D, which is the ratio of the scour depth to the pile diameter.

Theoretically, the equilibrium scour depth is never reached, since it is an asymptote. In practice, the equilibrium depth is also seldom reached since it is a dynamic parameter. For every hydraulic and structural condition a different equilibrium depth exists. Therefore, in an actual marine environment this variable will fluctuate in time.

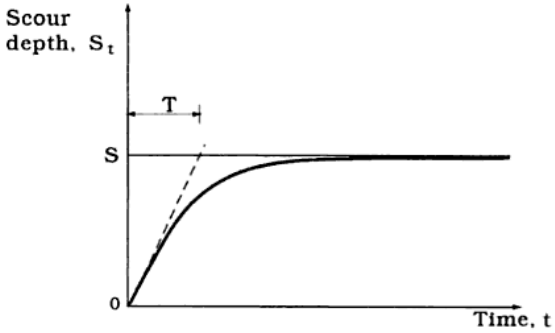


Figure 2.2: *Definition sketch time scale scour Sumer and Fredsøe[8]*

The equilibrium scour depth and time scale are both important parameters in the design of an offshore structure. The maximum scour depth defines the amount of necessary scour protection and foundation requirements. The time scale is important for managing scour. If the process is slow, the development time determines the frequency of maintenance. If the process is fast, only the equilibrium scour value is of importance.

There are basically two ways to deal with scour. According to Raaijmakers et al. [9] either scour protection is applied around the foundation (in order to guarantee a constant pile fixation level) or the foundation structure is modified by adding steel (in order to deal with a decreasing pile fixation level). The solution that is most often adopted is to install scour protection. A common scour protection is a two- or three-layer system with a top layer, consisting of large rock to resist the hydraulic loads (i.e. armour layer) and one or two layers of smaller rock to form a filter and prevent the washout of seabed sediment. However, scour development is very site- and structure specific, which means that at many locations omitting the scour protection and adding more steel could be more cost-efficient.

2.2 Detailed process of scour development

As previously mentioned, different hydraulic and structural conditions lead to different values for the equilibrium scour depth. In order to be able to predict the scour depth for any of these conditions, a better understanding is necessary of the processes that occur when a pile is placed in the marine environment. In this section a closer look is taken at the effects of flow, sediment and pile interaction.

2.2.1 Categories

To be able to get a grip on the many different hydraulic and structural conditions, scour is often categorized in separate cases. Each group has its own behavior when it comes to scour. Based on structure shape, sediment mobility and environmental conditions, the following division is made:

- Global scour vs. local scour
- Clear water scour vs. life-bed scour
- Current-only, waves-only or combined current and wave conditions

1. Global vs. local scour

Local scour is the erosion of the seabed at a single foundation. Global scour is a wider erosion around a structure consisting of multiple foundations.[10] This is clearly visible in figure 2.3. Small scour holes are present around each of the piles (local scour) as well as a scour pit around the entire structure (global scour).

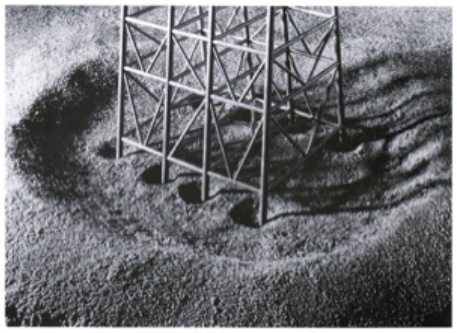


Figure 2.3: *Group scour around a jacket structure, Angus and Moore*[11]

2. Clear water scour vs. life-bed scour

This category is based on the mobility of the sediment and is used to separate tests with only local sediment movement around the structure from tests where the entire bed is mobile.

In figure 2.4, 3 situations are sketched to illustrate this phenomenon. Figure 2.4 (a) shows an example of clear-water scour. There is no sediment supply from upstream, while at

point 2 there is sediment transport downstream due to the increased flow velocity after the duct. In this situation $S_2 > S_1 = 0$, where S represents the sediment supply at point 1 and 2. Since sediment is flowing out at point 2, erosion will develop locally at this point. Clear-water scour can be caused by a lack of erodible material upstream or by insufficient transport capacity upstream. Clear-water scour stops when a depth is reached such that the velocity drops below a critical value.

Figure 2.4 (b) shows the case known as live-bed scour. In this case the flow velocity and other conditions are sufficient to mobilize the entire bed. This means that at the pile, sediment is flowing out due to the pile/flow interaction, but also sediment is flowing in from upstream due to the mobile sediment. If the transport downstream is larger than the input from upstream, so in terms of sediment supply: $S_2 > S_1 > 0$, live-bed scour will occur. Live-bed scour stops when the local eroding capacity at point 2 equals the supply from upstream.

In figure 2.4 (c) the dynamic equilibrium situation is drawn. This is the case for a live-bed without scour: sediment can picked up from the bed and settle again, but there is no net change in the position of the bottom. In this case $S_2 = S_1 > 0$.

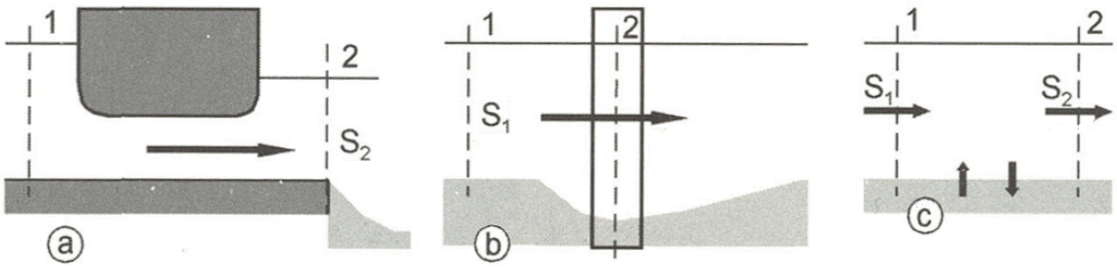


Figure 2.4: *Clear-water and live-bed scour, Schiereck*[12]

3. Current-only, waves-only or combined current and wave conditions

The three main categories in hydraulic conditions are current-induced, wave-induced and combined current- and wave-induced scour. In this report this categorization will be marked with respectively [C], [W] and [CW]. The scour in each type is dominated by a different phenomenon. For current-only situations the horseshoe vortex is the main cause of scour, while for wave-only scour the lee-wake vortices will be more important. For combined situations it is a difficult balance between enabling and counteracting each others effects. This will be explained in more detail in the next section.

2.2.2 Effects

According to Sumer and Fredsøe[8], the presence of the structure in the marine environment can result in the following list of phenomena, depending on the hydraulic conditions and structural conditions. Generally it can be said that all these effects cause an increase in sediment transport capacity, which enables the erosion of the seabed in the vicinity of the structure.

- Contraction of the flow
- Generation of turbulence
- Horseshoe vortex
- Lee-wake vortices
- Occurrence of wave reflection, diffraction and breaking
- Liquefaction
- Suspension of the sediment

Contraction of the flow

Contraction and expansion of the flow is important for scour calculations. These effects increase the flow velocity and hence the sediment transport capacity. The difficulty is to obtain the relative magnitude and location of the contraction point. This study considers one of the most elementary shapes, the cylindrical pile, but even around this basic object flow characterization can be very complicated. The flow pattern around the structure is mainly determined by the separation point. The separation itself depends on the roughness of the surface and the pile Reynolds number (i.e. $Re = \frac{uD}{\nu}$), which are difficult to predict. It will therefore be a challenge to model this effect in a formula. In case of ideal flow it would look something like 2.5 (a). In the actual case it looks more like 2.5 (b). It is clear that this pattern is not easily reproduced with a formula.

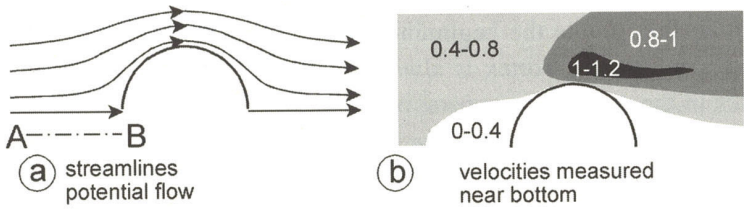


Figure 2.5: *Velocity profiles Melville [13]*

Generation of turbulence

Turbulence plays an important role in sediment transport. It is a results of velocity differences caused by the presence of the pile. In deceleration areas there will be a high amount of turbulence, opposed to acceleration areas where the turbulence is less, but the shear stress increase. Either way, the sediment transport capacity will be increased.

Figure 2.6 shows the turbulence intensity measured at the bottom of a cylinder. In the mixing layer between the wake of the cylinder and the main flow, the maximum intensity can be found. This is consistent with the location where the largest scour depth is usually found.

Horseshoe vortex

The horseshoe-vortex is caused by the rotation of the incoming flow. The sea bed boundary layer sets up a pressure gradient on the upstream side of the cylinder and because

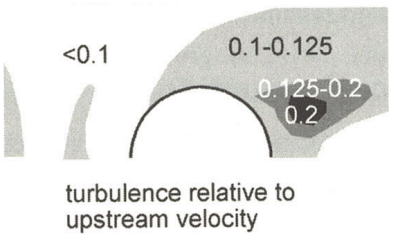


Figure 2.6: *Turbulence intensity Melville*[13]

there is low-pressure in the near-bed flow and high pressure in the flow above, the flow is driven down the face of the pile. At the sea bed, this downwards flow curls up and forms a rotating horseshoe vortex, trailing downstream.[8] [6] This is visualized in figure 2.7.

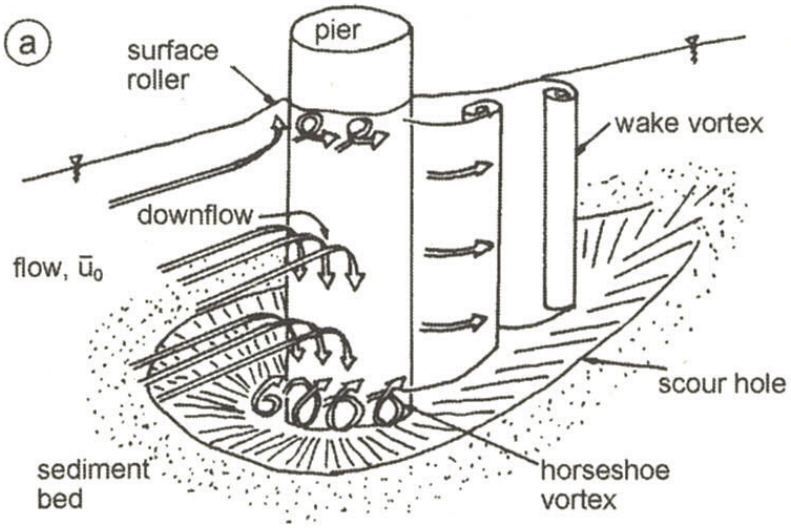


Figure 2.7: *Schematic overview horseshoe and lee-wake vortex Schiereck*[12]

For the horseshoe vortex, the separation of current [C] and wave [W] conditions becomes important. Sumer et al.[14] stated that the horseshoe vortex in [W] conditions with a small orbital wave velocity has a lower limit of existence than under [C] conditions. The horseshoe vortex can only grow for half a period. When the flow is reversed (due to the orbital motion of a wave) the horseshoe vortex is completely destroyed. Hence, according to Sumer and Fredsøe[8] the life span and effect of the horseshoe-vortex in [W] conditions will be negligible at first, but increase with increasing KC, due to larger stroke.

In the marine environment there are seldom current-only or wave-only conditions, which is why it is also important to look at the third category [CW]. In case of moderate waves superimposed to a current the amount of scour is limited, because the waves tend to break down the horseshoe vortex development created by the current due to the same effect described for [W] conditions.

Lee-wake vortices

The lee-wake vortex is formed by the rolling up and separation of the unstable shear layers generated around the structure. At a certain point eddies will be shed downstream in a periodic fashion.[6] A sketch of this phenomenon can be seen in 2.7. The shedding of vortices only occurs at specific hydraulic conditions, if the flow velocity exceeds a threshold value the periodic shedding becomes chaotic.

According to Sumer and Fredsøe[8] the lee-wake vortex in a steady current is mainly described by the pile Reynolds number and the pile geometry. For steady current the mechanisms of lee-wakes are well understood. However, it becomes more complicated when considering a system in wave-only conditions. In waves the lee-wake flow is no longer a passive flow feature, but it acts as a convection mechanism to transport the sediment away from the structure each half period of the wave motion. Therefore for predicting scour in waves the characteristics of the lee-wake vortex are essential. For current-only conditions they are of little importance.

Occurrence of wave reflection, diffraction and breaking

When a pile is placed in a marine environment there will be always be some wave reflection and diffraction, however for most piles this will have no effect on the equilibrium scour depth. Only for piles with a large diameter, wave reflection and diffraction will be an issue. According to Sumer and Fredsøe[8], the existence of the previously mention lee-wake and horseshoe vortices is limited for piles with a large body size. Commonly piles are considered large when the ratio of the pile diameter versus the wave length is $D/L > O(0.1)$.[8].

The main contribution of wave breaking to the scour depth is that it provides an additional source of turbulence in the water column, which will increase the sediment carrying capacity of the flow.

Liquefaction

A bed is in a liquefied state when it has very low or zero shear strength, i.e. the grains within the bed are unconstrained by neighboring grains. This has two effects. It removes the capacity of the bed to support a normal load and it makes the bed much more susceptible to erosion by waves and current because of the reduced intergranular friction. Steep storm waves are likely to be the most effective at causing liquefaction of the bed because they generate high pressure gradients at the bed. Especially for global scour liquefaction is a probable cause for scour according to Whitehouse[6].

Suspension of the sediment

All of the above phenomena increase the transport capacity of the flow. If the capacity is sufficient the sediment will become suspended. The difficulty is to determine the conditions for which this will occur. Suspension and hence erosion will start when the grains become unstable. To understand the stability of loose grains, it is necessary to

know which forces make a stone move, i.e. the exact balance between lift, drag, shear and friction forces.

The velocity at which the grain starts to move is called the critical velocity ($u_{c,cr}$). There have been many attempts to determine the this parameter analytically. The reason that this is not that successful is that the stability of an individual stone highly depends highly on its protrusion from the bed and protection supplied by surrounding stones. In a natural material the large variance in protrusion of grains and the differences in sizes and shapes make an analytic approach of stone stability unrealistic.[12] One of the solutions is the Shield parameter, which uses the bed shear stress to give an indication of the sediment mobility. Its definition is given in the next section by equation (2.4).

2.3 Influential scour parameters

A selection of important parameters influencing the equilibrium scour depth are described in this section. The mathematical definitions of a more extensive list of scour related parameters are given in appendix A.

Keulegan-Carpenter number [KC]

The Keulegan-Carpenter number relates the orbital velocity of a wave to the diameter of a pile. It has been proved by many studies that the KC number is the main parameter dictating the scour process in wave conditions. It is defined by equation equation (2.2):

$$KC = \frac{u_w T_p}{D} \tag{2.2}$$

Where u_w [m/s] is the orbital wave velocity, T_p [s] the peak wave period and D [m] the pile diameter.

The KC number governs the flow processes of both the aforementioned horseshoe vortex and the lee-wake vortex. In figure 2.8 the influence of KC is plotted against the equilibrium scour depth. It can be seen here that the scour depth approaches a constant value when $KC \to \infty$. This is to be expected, since the contribution of the lee-wake and horseshoe vortices approaches a constant value for large KC numbers. Lee-wake vortices have a limited lifespan of and the dependence on the horseshoe vortex disappears for infinitely large KC numbers. Interesting is that according to figure 2.8 the equilibrium scour depth approaches $1.3D$. This value is often found as the equilibrium scour depth for current-only scour. This means that according to this figure, for large KC numbers (i.e. large waves), the behavior of current conditions and wave conditions are similar.[8] Of course it can be argued that for infinitely large KC, the wave is almost a (tidal) current.

Relative mobility [mob]

The relative mobility is an important parameter in scour equations since it can be used to divide the scour data into two clear-cut regimes: clear-water or live bed scour. In the case of clear-water scour, no sediment motion takes place far from the structure, while

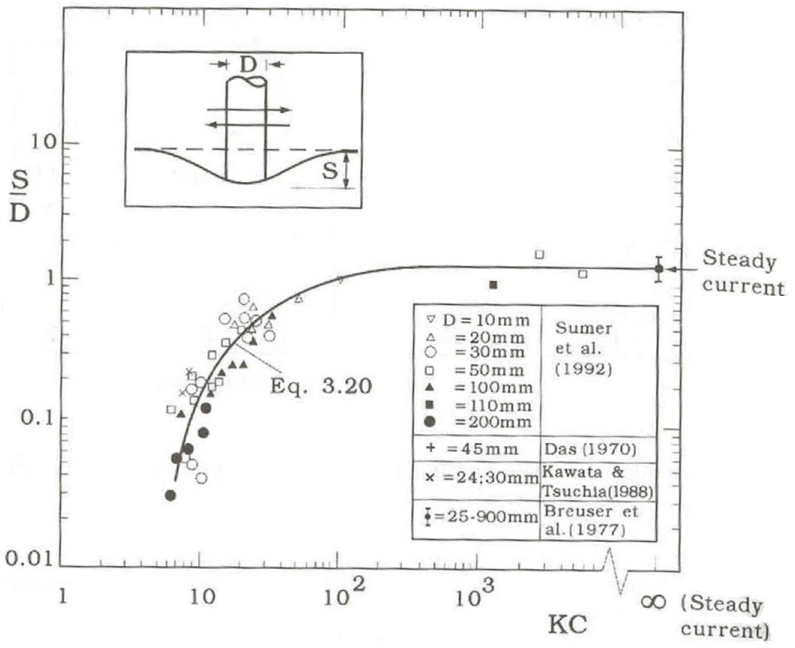


Figure 2.8: *Theoretical KC versus S/D (Sumer and Fredsøe[8])*

in case of live bed scour the sediment transport occurs over the entire bed. It is defined by equation (2.3) as the ratio of the Shields parameter and critical Shields parameter, as given by equation (2.4) and 2.5.

$$mob = \frac{\theta}{\theta_{cr}} \tag{2.3}$$

$$\theta = \frac{\tau_{cw}}{(\rho_s - \rho_w) g d_{50}} \tag{2.4}$$

$$\begin{aligned} \theta_{cr} &= \frac{0.3}{1 + 1.2 \cdot d^*} + 0.055 \cdot (1 - \exp(-0.02 \cdot d^*)) && d^* < 10 \\ \theta_{cr} &= \frac{0.24}{d^*} + 0.055 \cdot (1 - \exp(-0.02 \cdot d^*)) && d^* > 10 \end{aligned} \tag{2.5}$$

Where τ_{cw} [N/m^2] is the maximum bed shear stress, ρ_s [kg/m^3] the sediment density, ρ_w [kg/m^3] the water density, d_{50} [m] the mean grain size and d^* [-] the dimensionless sedimentological diameter.

If the Shields number is smaller than the critical Shields number (i.e. $mob < 1$), there is clear-water scour. If $mob > 1$ there is live bed scour.

As can be seen in figure 2.9 the two regimes behave quite different. In case of clear water scour the gradient of the equilibrium scour depth is large, which makes the predictions of clear-water scour sensitive to errors. On the other hand, the value of the equilibrium scour depth in the live bed regime does not alter much. The challenge in this regime is to model the peak around $mob = 1$, known as the clear-water peak (figure 2.9). This local maximum is caused by the backfilling of the scour hole the moment that the regime

transitions from clear-water to live bed. At this point the sediment in front of the pile is transported into the scour hole.

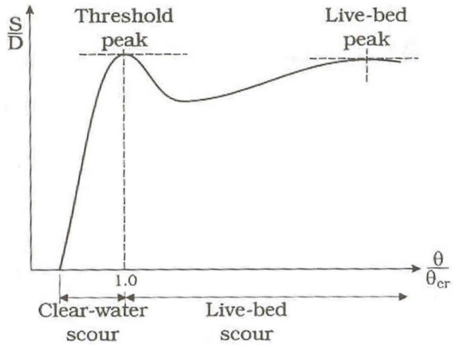


Figure 2.9: *Theoretical mob versuss S/D (Sumer and Fredsøe[8])*

Shallowness $[D/h_w]$

Shallowness is the ratio of the pile diameter over the water depth. This parameter is significant in the study of scour to account for the depth of the water relative to the water depth. In very deep waters, waves at the surface have little influence on scour at the bottom. At very shallow waters, the size of the vortices will be limited by the water depth, instead of the diameter of the pile.[12]

Boundary layer thickness ratio $[\delta/D]$

The boundary layer thickness ratio is defined as the boundary layer thickness (δ) divided by the pile diameter. The boundary layer thickness ratio is important, as can be seen in figure 2.10. The separation of the flow is delayed when δ/D is small, presumably leading to a smaller horseshoe vortex and thus also presumably leading to less scour. Or when δ/D is very small, the flow might not even separate at all and there will be no horseshoe vortex.[8] Unfortunately boundary layer thickness is not easy to measure, especially for waves, not even in confined laboratory tests, so this characteristic can not be included in the database for the GP calculations.

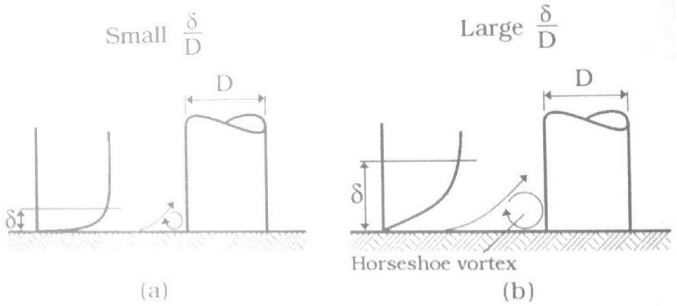


Figure 2.10: *Theoretical boundary layer sketch (Sumer and Fredsøe[8])*

Pile height [h_p]

For [C] conditions and [W] conditions with large KC numbers the pile height has a significant influence on the equilibrium scour depth. According to Sumer and Fredsøe [8] the adverse pressure gradient generated by the presence of the pile, and thus the resulting horseshoe vortex will be influenced by the pile height in the case of finite-height piles. The smaller the pile height the smaller the adverse pressure gradient, and hence also a reduction in the size of the horseshoe vortex which influences the depth of the scour.

Sedimentological diameter [d^*]

For stability and erosion, the size of a grain is an important parameter. It is a part of many other scour related parameter definitions, such as the critical Shields number. The sediment size is best represented by the sedimentological diameter. This can be seen as the ratio of the submerged weight of a grain to the viscous forces. The sedimentological diameter can be calculated with equation (2.6)

$$d_* = d_{50} \cdot \left(\frac{\Delta g}{\nu^2} \right)^{1/3} \tag{2.6}$$

Where d_{50} [m] is the mean grain size, Δ [-] the relative density between sediment and water and ν [m^2/s] the kinematic viscosity.

Sediment is difficult to scale. If extremely fine sand is used in lab tests, cohesion of the sediment might influence the results. The sedimentological diameter has the advantage over the mean grain size that it is dimensionless and therefore easier to scale between model and prototype conditions.

Relative suspension number [SB]

Particles will come into suspension if the bed shear velocity is larger than the particle fall velocity. [8] It is expected that the amount of suspended particles has some effect on scour development. The above mentioned relative mobility is a measure of the amount of total sediment transport. However, this number does not contain any information on the distribution of the total transport into bed-load and suspended load. Therefore, this SB-number is defined. It combines the current and wave related bed shear stress with the fall velocity of the sediment particles. Its equation is given by 2.7:

$$SB = \frac{1}{w_s} \frac{\sqrt{\tau_{cw}}}{\rho_w} \tag{2.7}$$

Where w_s [m/s] is the particle fall velocity, τ_{cw} [N/m^2] the maximum bed shear stress and ρ_w [kg/m^3] the water density.

Another way of incorporating the suspension of the sediment is through the dimensionless particle velocity ${w_s}^*$ as suggested by Camenen et al. [15].

$${w_s}^* = \left[\frac{(s-1))^2}{g\nu} \right]^{1/3} w_s \tag{2.8}$$

Where s [-] is the specific gravity, ν[m^2/s] the kinematic viscosity and w_s [m/s] is the particle fall velocity.

Sediment gradation [σ]

The sediment gradation is the spread of the sediment size, defined by equation (2.9).

$$\sigma = \sqrt{\frac{d_{84}}{d_{16}}} \tag{2.9}$$

Where d_{16} [-]and d_{84} [-] are the sediment diameters for which 16 and 84 percent of the sediment material is finer by weight.

The opinions about the effect of sediment gradation on the equilibrium depth are divided. According to both Molinas [16] and Guo et al. [17] the sediment size affects the scour process, but has no influence on the equilibrium scour depth. According to Whitehouse [6] there are some studies that report a correlation between the scour depth and σ in test results, but overall it was concluded that the effect of sediment gradation were negligible.

According to Sumer and Fredsøe [8] the sediment gradation does have some influence on the equilibrium scour depth due to the armouring effect (large grains preventing erosion by sheltering smaller grains). They list several studies that reveal that the scour depth decreases when σ increases. According to this source the effect is more present in case of clear-water scour, opposed to live-bed scour.

Reynolds number [Re]

The Reynolds number is given by equation (2.10).

$$Re = \frac{u \cdot D}{\nu} \tag{2.10}$$

Where u_c is the orbital velocity u_w[m/s] in case of [W] conditions and u_c[m/s] in case of [C] conditions. D[m] is the pile diameter and ν[m^2/s] the kinematic viscosity.

As mentioned the previous section, the Reynolds number is very well suited to describe turbulence. It is a ratio between destabilizing and stabilizing forces in a flow in case of an undulations. Viscous damping (ν) has a stabilizing effect, while the other two parameters increase when the turbulence increases. Also according to several sources cited in Whitehouse [6] the pile Reynolds number is an important parameter. It correlates with the scale and intensity of the horse-shoe vortex in steady turbulent flow. The only major drawback of this parameter is that it can not be scaled with the typical Froude scaling used in test settings.

Relative velocity $[u_{rel}]$

For [CW] conditions the u_{rel} is the most important parameter. It is a function that creates a smooth transition between current only conditions and wave only conditions. The formula of u_{rel} is given by equation (2.11).

$$u_{rel} = \frac{U_c}{U_c + U_w} \tag{2.11}$$

Where u_c[m/s] is the current velocity and u_w[m/s] the orbital velocity.

Its behavior according to Sumer and Fredsøe[8] can be seen in figure 2.11. The picture implies that even for small KC number a slight current superimposed on waves will cause the scour depth to increase significantly. This is due to the presence of a strong horseshoe vortex in front of the pile even in the case of a weak current.

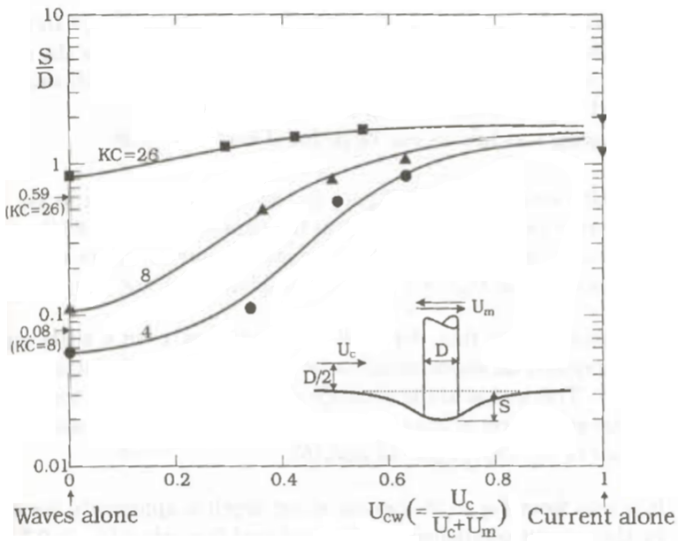


Figure 2.11: *Relative velocity versus equilibrium scour depth (Sumer and Fredsøe[8])*

2.4 Existing empirical formula

Many empirical scour prediction formulas have been proposed in the history of scour studies. In this section the most common and recent equations will be discussed. These empirical formulas will be used to validate the performance of the GP in chapter 8. A description of the equation variables can be found in appendix A.

2.4.1 Current-only formulas

Breusers et al. (1977) [18]
In 1977 Breusers et al. suggested the formula given by equation (2.12). It uses the correction factor K_i to account for site and structure specific conditions such as pier

shape, bed material gradation, group scour, etc. For a single monopile this value is assumed to be $K_i = 1$.

$$\frac{S}{D} = 1.5 K_i \tanh\left(\frac{h_w}{D}\right) \tag{2.12}$$

Richardson and Davis (1993) [19] **and Molinas (2003)** [16]
A series of studies lead by Richardson and Davis resulted in the scour prediction formula known as the CSU equation. In 2003 Molinas presented an updated version of this equation, given by equation (2.13). It incorporates two extra components with respect to the original CSU equation to include the effects of the scour initiation threshold velocity and the variance in gradation and coarse fraction size for non-uniform sediment mixtures (K_4 and K_i).

$$\frac{S}{D} = 2 K_1 K_2 K_3 K_4 K_i \left(\frac{D}{h_w}\right)^{0.35} (Fr)^{0.43} \tag{2.13}$$

With:
K_1 = Correction factor for pile shape ($K_1 = 1$ for circular piles) [20]
K_2 = Correction factor for the approach flow angle of attack ($K_2 = 1$ for direct flow) [20]
K_3 = Correction factor for bed form ($K_3 = 1.1$ for clear-water scour) [20]
K_4 = Correction factor for gradation and coarse fraction size, addition to CSU equation by Molinas [16], given by equation (2.14).
K_i = Correction factor for the threshold velocity of scour addition to CSU equation by Molinas [16], given by equation (2.15).

$$K_4 = 1.25 + 3\sqrt{\frac{d_{cfm}}{d_{50}}}\,\Psi^{0.60} \ln\left(\Psi + 0.5\right) \tag{2.14}$$

$$K_i = \left(1 - \frac{u_{ini}}{u_c}\right)^{0.45} \tag{2.15}$$

Azamathulla (2010) [21]
Azamathulla used genetic programming to formulate an equation that predicts bridge pier scour. This resulted in equation (2.16). According to the accompanying paper, this formula predicts significantly more accurate results than the CSU equation. However, when it was tried to reproduce this performance, scour depths in the order of magnitude of $\cdot 10^{83}$ were predicted. This suggests that there hides an error somewhere in this equation,

which is a good example why parsimony is important for GP functions.

$$\frac{S}{h_w} = \left(\frac{d_{50}}{h_w}\right)^{-0.5} \left[\left\{ \left(\left[-2.36 \frac{\left(1-T^2\right)^2 \left(\frac{h_p}{h_w}\right)^{0.5}}{\left(\frac{d_{50}}{h_w}\right)\sigma^2} + T^2 \right]^2 + Fr \right)^2 - 1 \right\}^2 - \frac{D}{h_w} - 1 \right]^{0.5}$$

in which:

$$T = 2 \left[\frac{\frac{F_r - 0.224}{\frac{h_p}{h_w}} + F_r - \frac{D}{h_w} - 0.739}{\frac{d_{50}}{h_w}} + (F_r - 0.224) \right]^2 \tag{2.16}$$

Sheppard [22]
One of the best formulas for current-induced scour available in literature is the equation from Sheppard and Miller. It consists of three parts that are dependent on the functions $f_1\left(\frac{h_w}{D}\right)$, $f_2\left(\frac{u_c}{u_{c,cr}}\right)$, $f_3\left(\frac{D}{d_{50}}\right)$. Each part is suited for one specific range: respectively clear-water scour (for $u_{crit} > 0.47$), live-bed scour before the live-bed peak velocity and live-bed scour above the live-bed peak velocity. Their formulas are given by equation (2.17) to (2.19). A detailed description of the calculation method of (f_1, f_2, f_3) is given in the original source Sheppard and Miller [22].

$$\frac{S}{D} = 2.5 f_1 f_2 f_3 \tag{2.17}$$

$$\frac{S}{D} = f_1 \left[2.2 \left(\frac{\frac{u_c}{u_{c,cr}} - 1}{\frac{u_{lp}}{u_{c,cr}} - 1} \right) + 2.5 f_3 \left(\frac{\frac{u_c}{u_{c,cr}} - \frac{u_{lp}}{u_{c,cr}}}{\frac{u_{lp}}{u_{c,cr}} - 1} \right) \right] \tag{2.18}$$

$$\frac{S}{D} = 2.2 f_1 \tag{2.19}$$

2.4.2 Wave-only formulas

Sumer et al. (1992) [14]
The most well known formula for wave induced scour is given by the equation of Sumer et al. in 1992. It has a high accuracy for small equilibrium scour depths, but its drawback is the small range of applicability. Equation 2.20 can not be used for KC numbers smaller than 6, which is quite common for prototype conditions. [3]

$$\frac{S}{D} = 1.3 \left(1 - \exp\left(-0.03 \left(KC - 6\right)\right)\right), \quad \text{with } KC \geq 6 \tag{2.20}$$

2.4.3 Combined current and waves formulas

Sumer et al. (2002) [8]
In 2002 Sumer et al. extended the formula from 1992 (equation (2.20)) with the relative velocity (equation (2.21)). With this update the formula is also applicable for current=only and combined current and waves conditions. Although the formula still does not cover the typical range of offshore conditions ($0 < KC < 10$) very well. In case of a wave-only conditions, $u_{rel} = 0$, equation (2.21) will turn into the same equation as equation (2.20).

$$\frac{S}{D} = 1.3\left(1 - \exp\left(-0.03\left(KC - 6\right)\right)\right), \quad \text{with } KC > B$$

$$\begin{aligned} &\text{in which:} \\ &A = 0.03 + \frac{3}{4}u_{rel}^{2.6} \\ &B = 6\exp\left(-4.7u_{rel}\right) \end{aligned} \tag{2.21}$$

Rudolph and Bos (2006) [23]
Rudolph and Bos tried to optimize the formula from Sumer et al. (2002) even further by improving the weakest part of the equation; the combined current and wave conditions. This is done by adding a term that effects only the predictions in the [CW] regime, as can be seen in equation (2.22).

$$\frac{S}{D} = 1.3\left(1 - \exp\left(-0.03\left(KC - 6\right)\right) \cdot \left(1 - u_{rel}\right)^{c}\right), \quad \text{with } KC > B$$

$$\begin{aligned} &\text{in which:} \\ &A = 0.03 + 1.5u_{rel}^{4} \\ &B = 6\exp\left(-5u_{rel}\right) \\ &C = 0.1 \end{aligned} \tag{2.22}$$

Raaijmakers (2008) [7]
The following formula suggested by Raaijmakers is based on the Breusers-formula. It is equipped with two additional correction factors to account for submerged piles (K_h) and wave action (K_w).

$$\frac{S}{D} = 1.5\tanh\left(\frac{D}{h_w}\right)K_w K_h \tag{2.23}$$

$$K_w = \left(\frac{h_p}{h_w}\right)^{0.67} \tag{2.24}$$

$$K_h = 1 - \exp\left(-0.012KC - 0.57{KC}^{1.77}{u_{rel}}^{3.76}\right) \tag{2.25}$$

Chapter 3

Data

To train the genetic program an accurate and extensive database with experimental data is necessary. This data is obtained by gathering scour test descriptions and results from a variety of sources, including journals, articles, databases, et cetera. The resulting database will be described in this chapter.

3.1 Data distribution

The complete database consists of 2512 tests, which can be subdivided into the three conditions as visible in figure 3.1. In figure 3.1 (a) the amount of data points can be found, while figure 3.1 (b) gives the amount of different sources per category.

Taking into account that some sources report tests in multiple categories, data is extracted from 75 different sources in total. A list of each source can be found in B.

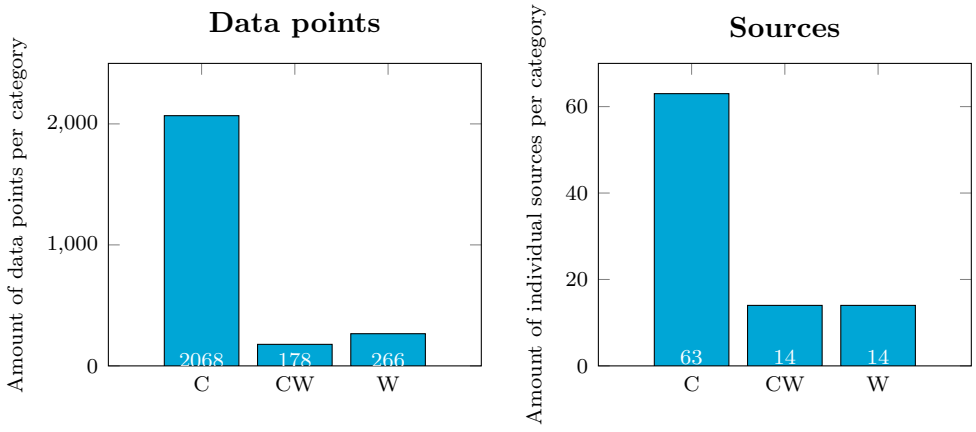


Figure 3.1: *Data analysis: number of data points and sources in the database*

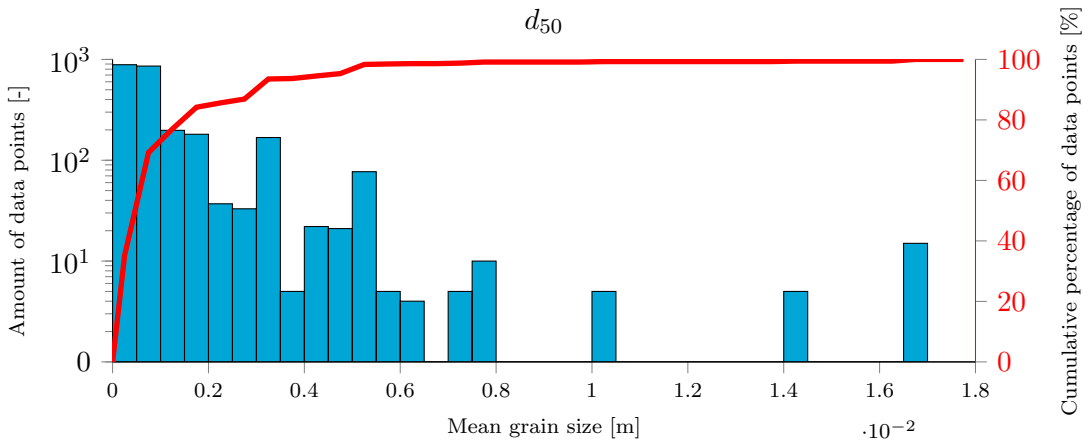


Figure 3.2: *Data analysis: range of mean grain size (d_{50}) in the database on logarithmic scale*

3.2 Data range

In appendix C a wide range of plots can be found that give an indication of the spread of the data in the database. In most cases the parameters that these graphs describe are related with one another by one simple, universally applied formula. However, some parameters have different calculation methods, leading to different results. They are based on different theories (e.g. Stokes versus linear wave theory) or deducted from different empirical relations. For the GP to work properly, the input data needs to be as coherent as possible. In this section first some of the more important graphs of straightforward parameters will be discussed, followed by parameters that have multiple calculation methods or require some fitting.

3.2.1 Universally applied parameters

Median grain size d_{50}:
Figure 3.2 shows the distribution of the grain size used in the scour tests. According to Schiereck [12] all grains with a diameter above 2 mm can be considered as gravel. The variation of equilibrium scour depth with gravel sizes is considerably different from that with sand sizes. Therefore it is not sure if the formula can fit these conditions. However, quite some tests are above this threshold. In total 398 experiments come from gravel experiments, about 15.8 %, so this can be a significant group. The largest sediment size stems from the 14 gravel experiments with d_{50} around 0.017 m from Molinas [16]. Other large grain sizes originate from the research of Raikar and Dey [24] with uniform and nonuniform gravels (fine and medium sizes, $4.10mm \leq d_{50} \leq 14.25mm$).

Pile diameter D:

In figure 3.3 the range of pile diameters can be seen. The pile diameters are quite small; about 95% is smaller or equal to $0.25m$. Although most parameters will be made dimensionless, the scour prediction model probably profits from tests that are closer to life-size. Sumer and {Fredsøe} [25] did some research on wave scour around large vertical cylinders,

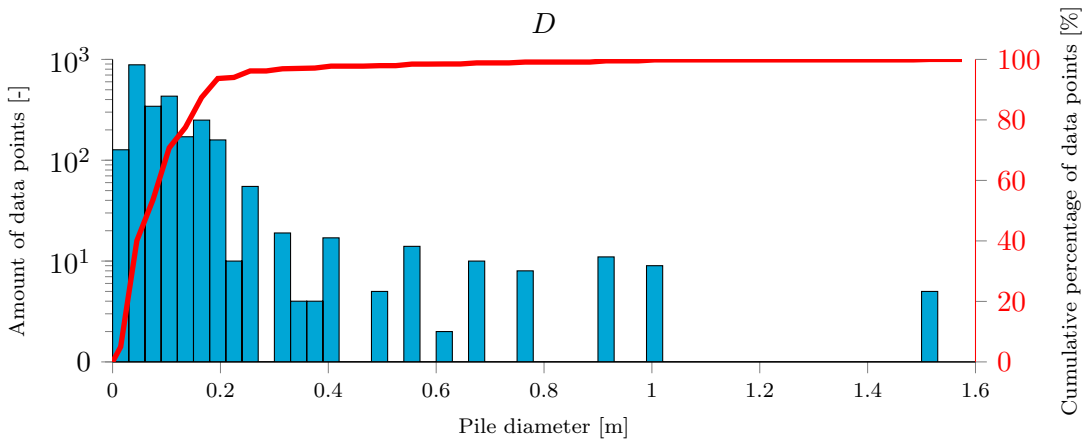


Figure 3.3: *Data analysis: range of pile diameters (D) in the database on logarithmic scale*

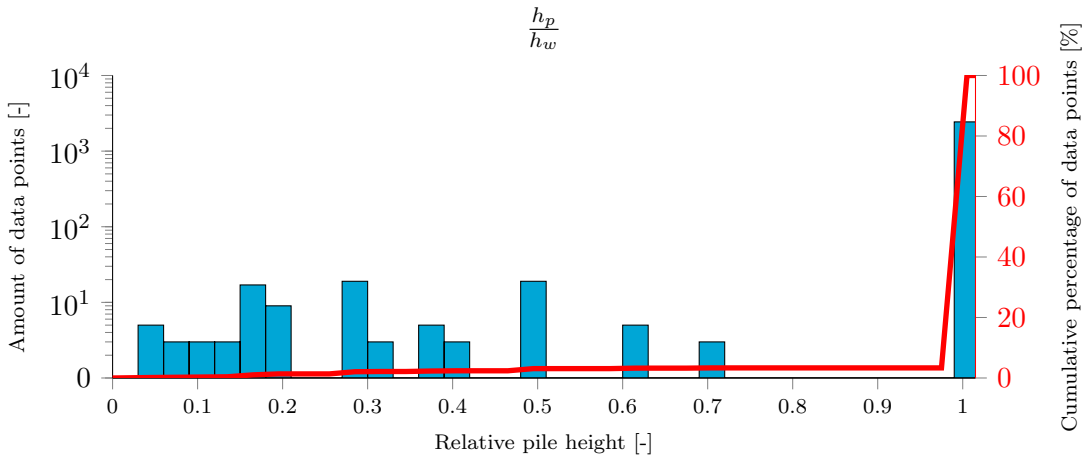


Figure 3.4: *Data analysis: range of relative pile height ($\frac{h_p}{h_w}$) in the database on logarithmic scale*

up until 1.53 m. Sheppard et al.[26] performed laboratory research on large scale cylinders in current only conditions, with a pile diameter size up to 0.915 m. In the category currents and waves there are also some large pile diameter tests performed by Zhao[27] on local scour around a large-scale cylinder due to combined currents and waves.

Relative pile height $\frac{h_p}{h_w}$:

Figure 3.4 shows the distribution of the relative pile height. 3.41% of the ratio between the pile height and the water depth is non-equal to one. These stem from only 2 sources. Inhouse source 'WL; R&D 2004' (60 points) and Zhao et al.[28] (26 points). No relative pile heights above 1 are accepted. These are interpreted as one.

Relative velocity u_{rel}:

u_{rel} is an indication of current or wave dominated flow. In figure 3.5 it can be seen that most values are of course 1 because the most datapoints are available for current-only tests. The zero values indicate the wave-only tests. In the mid regime the tests can be

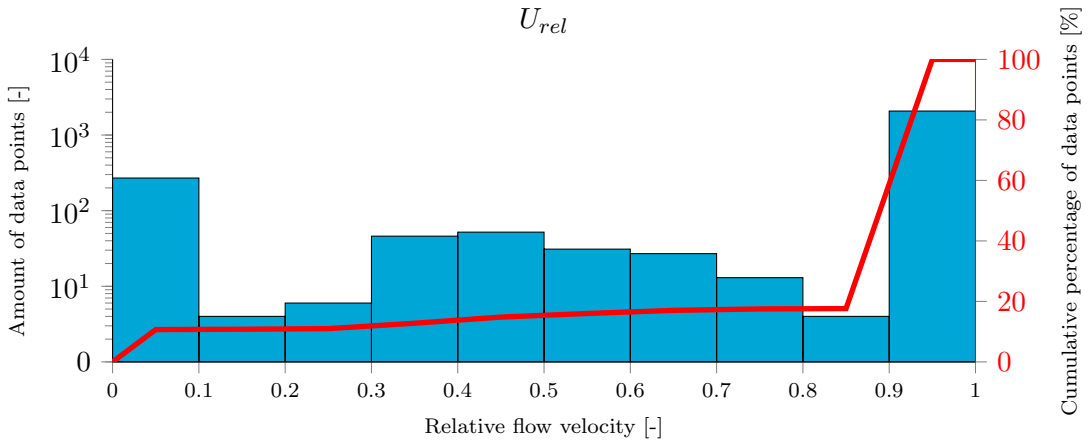


Figure 3.5: *Data analysis: range of relative flow velocity (U_{rel}) in the database on logarithmic scale*

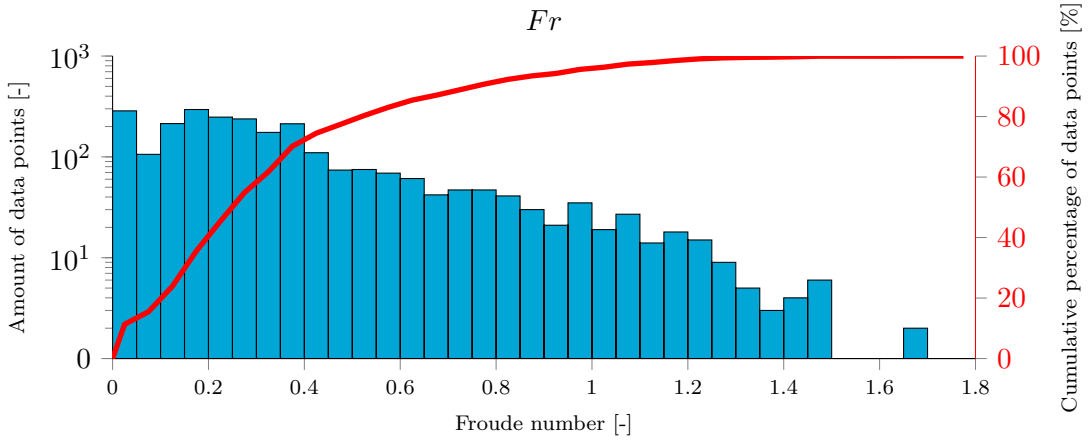


Figure 3.6: *Data analysis: range of Froude numbers (Fr) in the database on logarithmic scale*

found that were performed with combined current and wave conditions. The value of u_{rel} determines which condition was more dominant. The values of u_{rel} are nicely spread between 0 and 1. This will benefit the GP model during training, since it prevents a certain bias. There are 111 wave dominated tests (i.e. 111 test with $u_{rel} < 0.5$) and 69 current dominated tests (i.e. 69 test with $u_{rel} > 0.5$). In this graph u_{rel} is given for both regular and irregular waves, which use a different calculation method of the orbital wave velocity u_w. For regular waves this value is based on the wave height H and period T, while for irregular waves the significant wave height H_s and peak period T_p are used, as will be discussed later on. There are 39 regular combined wave and current tests (with 20 wave and 19 current dominated tests) and 141 irregular combined wave and current tests (with 99 wave and 40 current dominated tests).

Froude number Fr:

Figure 3.6 shows the range of Froude numbers. It is directly related to the flow velocity

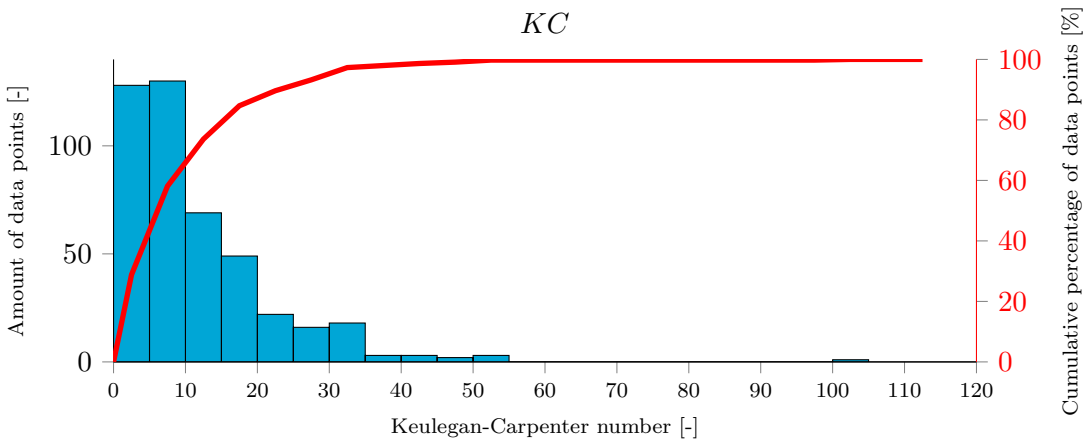


Figure 3.7: *Data analysis: range of Keulegan-Carpenter numbers (KC) in the database on linear scale. [W] and [CW] data only*

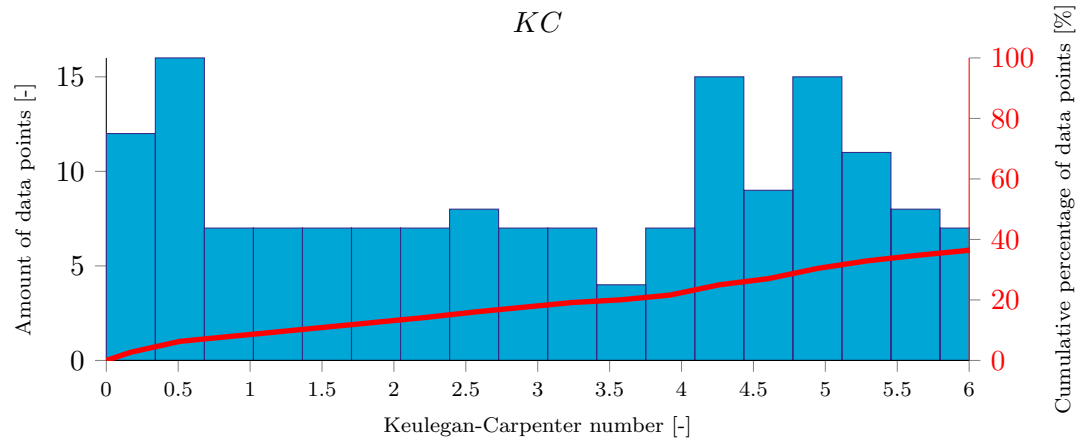


Figure 3.8: *Data analysis: range of Keulegan-Carpenter numbers (KC) in the database on linear scale. [W] and [CW] data only and* $KC \leq 6$

by equation (3.1).

$$Fr = \frac{u_c}{\sqrt{g \cdot h_w}} \tag{3.1}$$

Where u_c[m/s] is the current velocity and h_w[m] is the water depth.

The Froude number describes different flow regimes by determining the ratio of inertial and gravitational forces. If $Fr > 1$ there is supercritical flow, the water is propelled faster than it would be under just gravitational pull (for example in a jet stream). For $Fr < 1$ the water moves slower than its gravity-only counterpart, due to friction at the bed. In the database about 5% of the experiments have $Fr > 1$. At gravel experiments the highest current velocity occurs, and also the highest Fr number of 1.7 for one test. The tests with a Froude number above 1 are considered irrelevant, because the jet stream-like velocities associated with $Fr > 1$ are not a condition that will occur at offshore wind park locations.

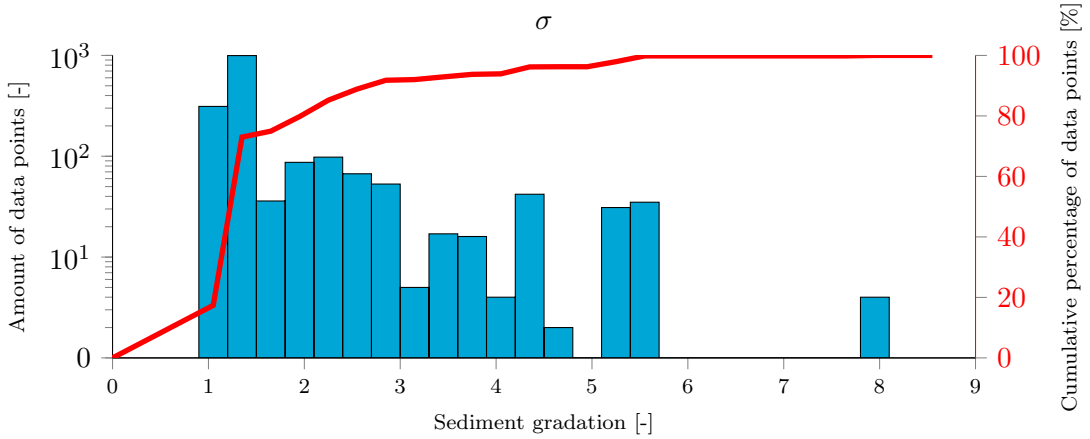


Figure 3.9: *Data analysis: range of sediment gradation (σ) in the database on logaritmic scale.*

Keulegan-Carpenter number KC:

Figure 3.7 shows the distribution of KC numbers. There is one test for the very large KC number of 101, executed by Sumer et al. [14]. To get a better perspective on the lower KC numbers, the same graph is shown in figure 3.8 for the range $0 < KC < 6$. Almost 40% of the tests have a KC number smaller than 6. This is a good distribution for modeling the test that are not valid for the Sumer formula given by equation (2.20), which can only be used for $KC \geq 6$.

Sediment gradation σ:

Figure 3.9 shows the distribution of the sediment gradation. Most tests have been done with uniform sediments (i.e. $\sigma \approx 1$. Unfortunately only 70% of all tests report the value for σ.

3.2.2 Customized parameters

Some of the parameters in the database can be calculated with different methods, for instance if multiple empirical relations exist. In this section will be discussed which equations will be used in the search for the equilibrium scour depth prediction formula.

Wave length $[L]$

Two methods are considered for the calculation of the wave length: linear wave theory and Rienecker-Fenton. The wave length according to linear wave theory can be calculated with the dispersion relation given by equation (3.2). The Rienecker-Fenton method uses a Fourier approximation method which is described in Rienecker and Fenton [29].

$$\omega^2 = \left(\frac{1}{T}\right)^2 = gk \tanh\left(kh_w\right) \qquad \text{with } L = \frac{2\pi}{k} \tag{3.2}$$

In which ω [1/s] is the frequency of the wave, T [s] the wave period, k [1/m] the wave number and L [m] the wave length.

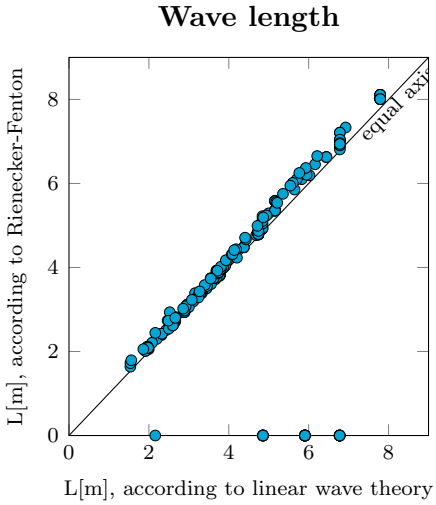


Figure 3.10: *Data analysis: Methods to calculate the wave length L*

Figure 3.10 shows the difference between the two methods. Invalid calculations outside of the applicability range of the Rienecker-Fenton method are given as zero values and the black line is equal axis. Most values are above the black line, i.e. wave length according to Rienecker-Fenton is estimated higher than with linear waves, but the effect is minimal. Since both methods give more or less the same result, it is chosen to use the simpler linear wave theory for the calculation of L.

Using linear wave theory, the orbital wave velocity is given by equation (3.3).

$$u_w = \frac{\pi H}{T \sinh\left(\frac{2\pi h_w}{L}\right)} \tag{3.3}$$

In which H [m] is the wave height, T [s] the wave period, h_w [m] the water depth and L [m] the wave length.

Irregular and regular waves:

As previously mentioned some sources in the database investigated scour under regular waves, others under irregular waves. It is well divided in the database; there are 226 regular and 218 irregular wave tests in the [W] and [CW] database combined.

This difference in test setup causes a potential problem for the input parameters. Regular tests have a constant wave height and period, while irregular wave test report the significant wave height H_s and peak wave period T_p. These parameters are used to calculated the orbital wave velocity according to the linear wave theory given by equation (3.3).

In figure 3.11 (a) is visible what the effect is on the KC values if H_s is interpreted as H and T_p as T. It shows that the curve of KC is similar, but the S/D values are slightly higher. H_s is just a representative value for an irregular wave spectrum, in principle any value could be the counterpart of the regular wave H. A calculation method for a representative irregular wave height (H_r) is sought, which has the effect that the KC-curves coincide. Furthermore, instead of the peak wave period the zero crossing period is taken. T_z can

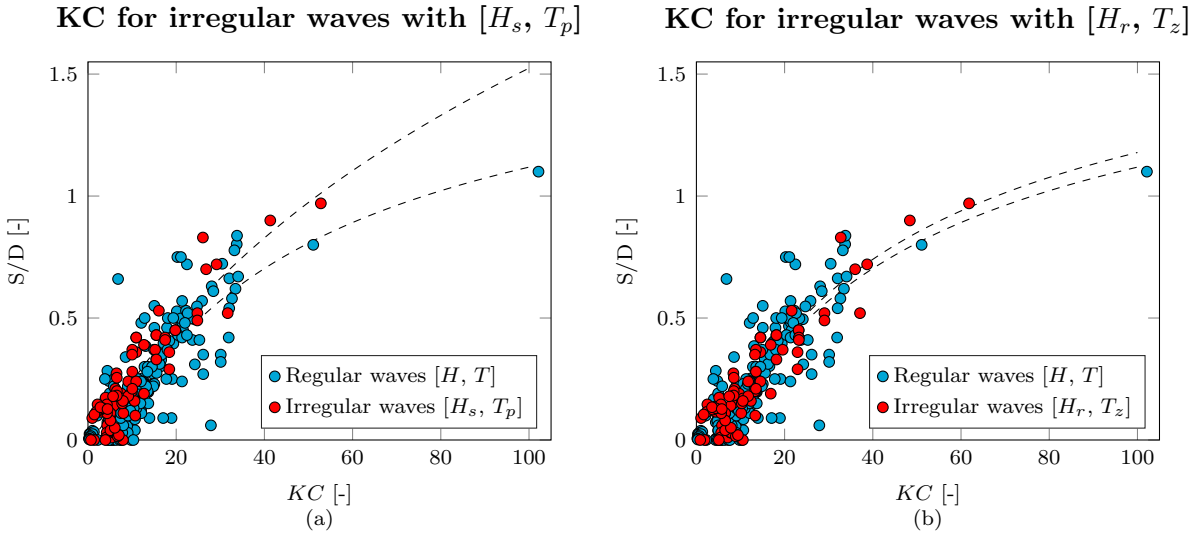


Figure 3.11: *Data analysis: range of incident wave angles (α) in the database on logarithmic scale*

be calculated with equation equation (3.4) as stated in DNV [30], in which γ is 3.3 for JONSWAP wave spectrum:

$$\frac{T_z}{T_p} = 0.6673 + 0.05037\gamma - 0.006230\gamma^2 + 0.0003341\gamma^3 \tag{3.4}$$

The wave height is multiplied by a factor loosely based on the formula from Tucker [31] that calculates the maximum wave height within a certain period. Assuming the maximum wave height within 3 hours is a good representation of the scour wave height, the equation is as follows:

$$H_r = \left(0.5\ln\left(\frac{10800}{{T_z}^4}\right)\right)^{0.5} \cdot H_s \tag{3.5}$$

In which H_r stands for representative wave height. This results in the following improvement for KC as can be seen in 3.11.

In figure 3.11 (a) the KC values lie slightly on top of the curve of the regular waves. By adjusting it with the prescribed method the curve can be shifted on top of the curve from the regular waves as visible in figure 3.11 (b).

Relative critical velocity u_{crit}

The relative critical velocity is a parameter that has the same behavior as the relative mobility. Similar to *mob* (equation (2.3)) it can be used to separate clear-water conditions from live-bed conditions. The main difference between *mob* and u_{crit}, apart from the scale, is that u_{crit} is parameter designed for current-only conditions. It is not applicable for any condition where waves influence the motion of sediment. This is unlike *mob*, which can

also calculate the mobility of the bed under [W] and [CW] conditions. The formula for u_{crit} is given by equation (3.6).

$$u_{crit} = \frac{u_c}{u_{c,cr}} \tag{3.6}$$

Where u_c [m/s] is the current velocity and $u_{c,cr}$ [m/s] is the critical current velocity at which the sediment becomes mobile.

As mentioned in the previous chapter there have been many attempts to calculate the critical velocity ($u_{c,cr}$). It is difficult to find an analytic relation for the threshold velocity, since this value differs for each grain: depending on the size, the shape, the protrusion from the bed, the amount of sheltering, et cetera. This is why many empirical relations have been created that try to capture the stability of the entire bed by a few characteristic bed parameters.

For this database 5 different calculation methods are tested, given by Whitehouse [6],Chiew [32], Melville [33], Sheppard et al. [34]and Oliveto and Hager [35]. The values of u_{crit} for the [C] database for each of the methods can be seen in figure 3.12. The method by Oliveto and Hager [35] gives the best results. This method has the least amount of variance and with the incorporation of a small factor $f_{pk} = 1.3$, the peak of S/D is exactly at 1. The set of formulas for Oliveto's $u_{c,cr}$ are given by equation (3.7). The main difference with the other methods is that these equations incorporate the sediment gradation σ in the calculations, which apparently increases the success of a $u_{c,cr}$ formula.

$$\begin{aligned}
u_{c,cr} &= \frac{2.33}{f_{pk}} d^{*-1/4} \sigma^{1/3} (9.81 \Delta d_{50})^{1/2} \left(\frac{h_w}{d_{50}}\right)^{1/6} && d^* \leq 10 \\
u_{c,cr} &= \frac{1.08}{f_{pk}} d^{*1/12} \sigma^{1/3} (9.81 \Delta d_{50})^{1/2} \left(\frac{h_w}{d_{50}}\right)^{1/6} && 10 < d^* \leq 150 \\
u_{c,cr} &= \frac{1.65}{f_{pk}} \sigma^{1/3} (9.81 \Delta d_{50})^{1/2} \left(\frac{h_w}{d_{50}}\right)^{1/6} && d^* > 150
\end{aligned} \tag{3.7}$$

Roughness length kr

The roughness length is an indication of the resistance properties of the bed. It is used in the calculations of the bed shear stress in current and in wave conditions, which means they are used in the calculation of the mobility. For the same reasons as u_{crit} this value differs for each individual grain and can therefore only be approximated with empirical relations. The paper of Camenen et al. [15] suggests 7 different methods of calculating kr which are considered for this study, as well as the commonly used suggestion from Whitehouse [6] of $kr = 2.5 d_{50}$. All kr-methods are judged on there performance in a plot of mob versus S/D. The plot that showed the least amount of variance and a small peak at $mob = 1$, was the method suggested by Camenen et al. [15] himself. This method calculates kr with equation (3.8). Compared to the rule of thumb $kr = 2.5 d_{50}$, this method is more accurate since it incorporates a term describes the mobility of the bed. The drawback of this equation is that since kr is necessary for the calculation of θ, the value can only be obtained by applying an iterative process.

$$\frac{kr}{d_{50}} = 0.6 + 2.4 \left(\frac{\theta}{\theta_{cr}}\right)^{1.7} \tag{3.8}$$

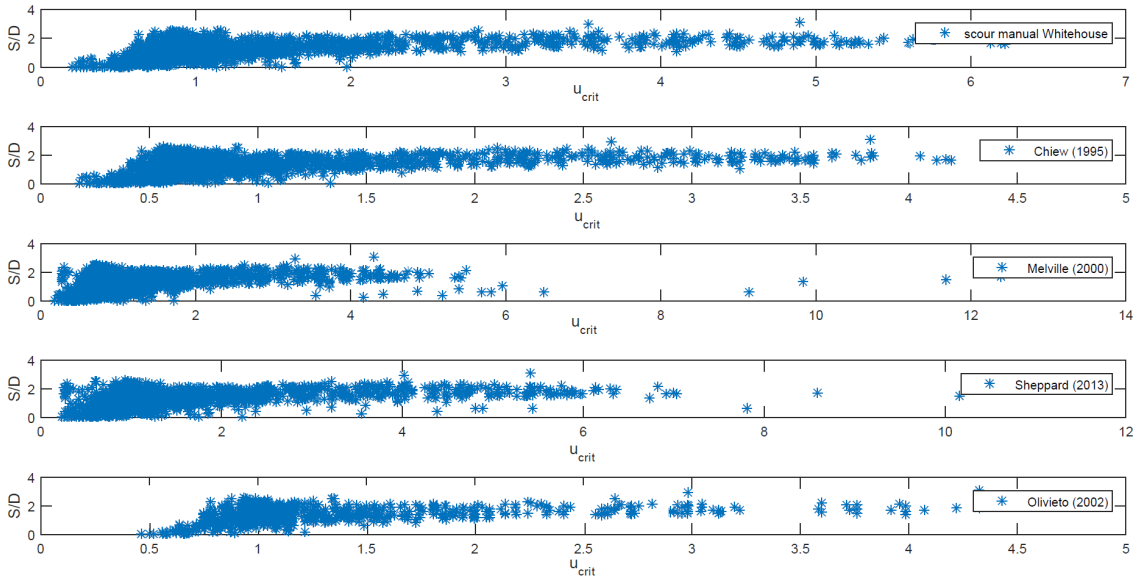


Figure 3.12: *Customized parameters: different methods to calculate* u_{crit}

3.3 Initial fit

The database for [C], [W] and [CW] are analyzed for parameters or combinations of parameters that show a clear relationship with the equilibrium scour depth. Defining how the equilibrium depth is effected by the variance of a parameter, helps to determine which parameters are important for the equilibrium scour depth prediction formula.

3.3.1 Current

It has already been suggested multiple times in the previous sections that the relative critical velocity u_{crit} and the sediment gradation σ could be important parameters in the process of scour. The relation becomes more evident when looking at figure 3.13. The mobility versus S/D by itself already shows a relation, but when a division of 4 sections is made in the data based on the size of the sediment gradation, 4 individual curves begin to form. Although there is some variance (especially in the $\sigma \leq 1.3$ region), it can be seen that red dots that represent the narrow gradated sediment with $\sigma \leq 1.3$ form the upper part of the graph. This is followed by a layer of yellow dots with values of $1.3 < \sigma \leq 2$, below that the blue dots with $2 < \sigma \leq 3$ and the lowest S/D values are for the green dotted, wide-gradated sediments with $\sigma > 4$. The same relation was found for mob and σ. This means that for a set of [C] conditions, the equilibrium scour depth can already be explained to some degree when just looking at the two parameters u_{crit} and σ or mob and σ .

It also appears that the peak that is visible for the narrow graded sediment, is smaller in size for $2 < \sigma \leq 3$ and almost non-existent for $\sigma > 4$. This is already an interesting find. This means that the clear-water peak that was discussed in section 2.3 exists only for narrow graded sediment. For larger sediment, the backfilling that occurs when transitioning from clear-water to live-bed scour might be limited due to the armouring effect. This

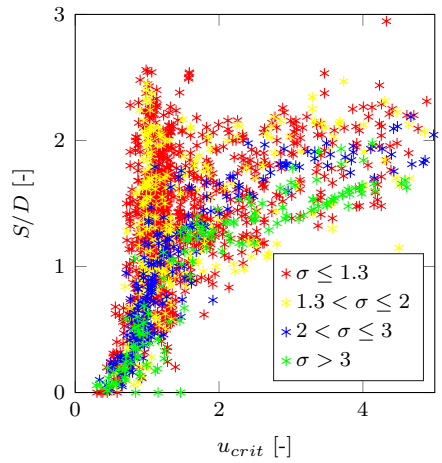


Figure 3.13: *Data fit [C]: relation between u_{crit}, σ and S/D for all parameters in the [C] database*

would explain why the clear-water peak is commonly found in laboratory studies (which often use uniform sediment), but not in field studies (where the sediment gradation is naturally wider).

3.3.2 Waves

Multiple studies have already proven that the KC number is the most important parameter in wave scour. This can be confirmed by the graph of KC versus the equilibrium scour depth shown in figure 3.14. With a simple fitting procedure in MATLAB and just the KC as input, already a scour prediction formula can be found that has a low error (low RMSE) and a low amount of unexplained variance (high CoD). The mathematical definition of these terms will be given in the chapter 5 by equation 5.1. The goal of this study will be to find additional parameters that improve the already strong relation between KC and S/D.

3.3.3 Combined currents and wave

No strong relations were thus far found for the combined current and wave conditions. However as suggested by Camenen et al. [15] a new parameter was introduced called the Froude wave number Fr_w, given by equation equation (3.9). It relates the horizontal velocity of the wave (celerity) to the gravitational forces in a similar way as the Froude number. It can therefore be a good counterpart in combined current and wave equations.

$$Fr_w = \frac{L}{T \cdot \sqrt{g h_w}} \tag{3.9}$$

Where L [m] is the wave length, T [s] the wave period and h_w [m] the water depth.

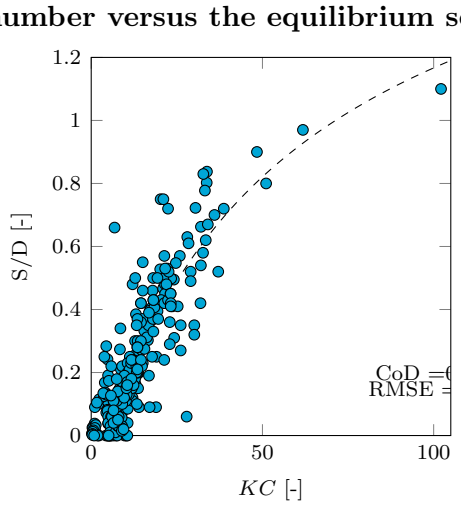


Figure 3.14: *Data fit [W]: relation between KC and S/D for all parameters in the [W] database*

3.4 Reliability tests

To incorporate the reliability and relevance of the tests, a weight factor is given to each individual entry in the database as well as a classification value for each source. Together they result in an overall weight factor per test which is used to indicate the importance of the test in the GP. The methods used to assign the weight factors are described here.

3.4.1 Individual weight factor

The individual weight factor is calculated with a point system. Each test starts out with 10 points. If a test has a quality that is considered irreliable or irrelevant, 1 or 2 points can be subtracted. If a test has a quality that is highly important, 1 or 2 points can be added.

The individual tests are judged on the following parameters:

Reliability

- Boundary influence
  If the width B of the flume or basin at the test facility is small compared to the diameter of the pile, the presence of the test section wall may cause an effect on the flow conditions at the pile. According to Chiew (1984) as cited in Molinas and Abdou [36] a blockage ratio of 8 or smaller can produce a significant wall effect.
  $$\begin{cases} 8 < \frac{B}{D} & \text{no substraction} \\ 7 < \frac{B}{D} \leq 8 & \text{-1 point} \\ \frac{B}{D} \leq 7 & \text{-2 points} \end{cases}$$

- Small pile diameters
  If the diameter of the pile is very small, the results can be affected by measurement

insecurities or inaccuracies, for example due to the effect of scaling.

$$\begin{cases} 0.05 < D & \text{no substraction} \\ 0.03 < D \leq 0.05 & \text{-1 point} \\ D \leq 0.03 & \text{-2 points} \end{cases}$$

- Small equilibrium scour depth
  If the reported equilibrium scour depth is in the same range as the ripple height of the bed (typically 1-2 cm), these tests will be considered irreliable. The reason is that for these values it is unclear if the height difference is due to the ripples or due to scour.

$$\begin{cases} 0.02 < S & \text{no substraction} \\ 0.01 < S \leq 0.02 & \text{-1 point} \\ S \leq 0.01 & \text{-2 points} \end{cases}$$

Relevance

- High Froude numbers
  As explained in section 3.2.1 if Froude is larger than 1, the flow is supercritical. This can occurs at for instance at a fast moving mountain river under a slope, but not at offshore conditions around wind turbines. Therefore the tests with high Froude numbers are considered to be less relevant.

$$\begin{cases} Fr < 0.9 & \text{no substraction} \\ 0.9 < Fr \leq 1.0 & \text{-1 point} \\ 1 \leq Fr & \text{-2 points} \end{cases}$$

- Clear-water or live-bed scour
  These tests are in a difficult regime, since they reflect an important transition in the results. These tests are therefor given an extra weight factor to make them more important.

$$\begin{cases} u_{crit} < 0.8 \quad \& \quad 1.2 < u_{crit} & \text{no addition} \\ 0.8 \leq u_{crit} < 0.9 \quad \& \quad 1.1 < u_{crit} \leq 1.2 & \text{+1 point} \\ 0.9 \leq u_{crit} \leq 1.1 & \text{+2 points} \end{cases}$$

- Diffraction.
  In case of regular waves there will be a weight factor based on the diffraction D/L. There are only a limited amount of test available in the diffraction regime. Too little to train the GP in this regime that behaves different from all the other tests. The scour in the diffraction area is considered to be beyond the scope of this study and the following weight factor is applied:

$$\begin{cases} \frac{D}{L} < 0.2 & \text{no subtraction} \\ 0.2 \leq \frac{D}{L} < 0.25 & \text{-1 point} \\ 0.25 \leq \frac{D}{L} & \text{-2 points} \end{cases}$$

- Breaking waves
  In case of waves the weight system is applied to decrease the importance of breaking waves in the database. Similar to the tests with waves in the diffraction regime, there

are too few test available with these conditions to incorporate this effect. Using the limit of $\frac{D}{L} < \frac{0.44}{KC}$ for non-breaking waves as given by xx Sumer and Fredsøe [8] and 0.8 times this limit to incorporate all test in the gray area between the transition, the following division is made:

$$\begin{cases} \frac{D}{L} < 0.8 \cdot \frac{0.44}{KC} & \text{no subtraction} \\ \frac{0.44 \cdot 0.8}{KC} \leq \frac{D}{L} < \frac{0.44}{KC} & \text{-1 point} \\ \frac{0.44}{KC} \leq \frac{D}{L} & \text{-2 points} \end{cases}$$

3.4.2 Source weight factor

The quality of the source is also important. Based on the following qualities the sources are given a value between 1 and 3. In which 1 represents a questionable source and 3 a reliable source.

Reliability

- Unknown origin
  Some datasets are part of an article that provides data from different sources. However, not for all the entries of these assembled lists the original author known. Or in two other cases the datasets were never published, which makes them more questionable as well.
- Original article
  If the original article is found this adds to the credibility of the source. The data can be verified and original articles usually contain more detail.
- Documentation
  Is there a description of the test set up a better judgment can be made.
- Age
  The age of the test is important, since over the years the measurement techniques have improved. Three categories are distinguished: less than 25 years old is considered relatively recent, between 50 and 25 years old is considered old and older than 50 years very old.
- References
  The database contains 4 large collectives, which sometimes contain identical sources. If a source is referenced multiple times, this could be an indication of its reliability.

Relevance

The relevance depends on the subject of the paper. Some research breaking waves, other different type of silts or pile group scour.

An overview of the resulting weight factor per source can be found in B.

3.4.3 Overall weight factor

The range of the above described factors are chosen in such a way that the overall weight factor is in a range between 1 and 5. The overall weight factor is calculated with equation (3.10), which results in a value ranging from 1 to 5, in which 1 is a questionable test and 5 a very reliable test. Table 3.1 gives the distribution of tests and their weight value.

$$W = round\left(\frac{W_{individual} \cdot W_{source}}{6}\right) - 1 \tag{3.10}$$

Condition	Amount of tests with:				
	$W = 1$	$W = 2$	$W = 3$	$W = 4$	$W = 5$
[C]	103	796	319	558	291
[W]	17	59	115	71	2
[CW]	0	47	26	106	1

Table 3.1: *Assigned weight factors per category: [C], [W] and [CW]*

Chapter 4

Parameters

This chapter specifies the input parameters for the genetic program (GP). It combines the knowledge of chapter 2 and 3 as well as the findings from other scour related and soft computing studies to determine the most important parameters.

4.1 Overview available parameters

In the course of this report already a number of parameters have been mentioned that could have some degree of influence on the equilibrium scour depth. A summary of all parameters and their mathematical definitions can be found in Appendix A.

Only a small part of the parameters presented in chapter 2 are independent variables. For this study, all other parameters can be estimated or derived with one or more of the following fundamental parameters: the pile diameter (D), the mean grain size (d_{50}), the pile height (h_p), the water depth (h_w), the wave height (H_s), the wave period (T_p), the current velocity (u_c) and the equilibrium scour depth (S). Each source in the database must contain a value for all of these parameters and if possible for the sediment gradation σ as well. The sediment gradation is mentioned here separately since although it is a fundamental parameter, only 70% of the sources have reported it.

To recap the literature study of chapter 2 table 4.1 is created. This gives a list of 15 important dimensionless parameters that can be used to describe the scour phenomenon.

4.2 Parameters from recent soft computing studies

A wide range of formulas have been proposed in the past century to estimate the equilibrium scour depth. A few of these empirical equations were presented in section 2.4. From these studies already some important input parameters can be derived. More recently also some studies have been done with soft computing techniques in similar areas and their results provide an additional insight in what parameters might have a significant

Effect	**Parameter**		**Indication of:**
[C]/[W]/[CW]	u_{rel}	Relative velocity	Current or wave dominated conditions
Clear-water vs. live-bed	u_{crit}	Critical velocity	Mobility of the sediment
	mob	Relative mobility	Mobility of the sediment
Turbulence	Re	Reynolds number	Turbulence intensity, location of the separation point
Horseshoe vortex	Fr	Froude number	Dimensionless current velocity, determines size pressure gradient
	Fr_w	Froude wave number	Wave counterpart of Froude number
	KC	KC number	Stroke of the wave motion, determines life span of the vortex
	$\frac{h_p}{h_w}$	Relative pile height	Pile height, limits the size of the adverse pressure gradient
	$\frac{D}{h_w}$	Shallowness	Limit of vortex size by D or h_w
Lee-wake vortex	Fr	Froude number	Dimensionless current velocity, determines vortex shedding
	$\frac{D}{h_w}$	Relative pile height	Pile diameter, determines vortex shedding
	KC	KC number	Stroke of the motion, determines sediment deposit interval
Sediment suspension	${w_s}^*$	Particle fall velocity	Determines the amount of suspended load
	d^*	Sedimentological diameter	Dimensionless size of the sediment
	Δ	Relative density	Indication of bouyancy of the sediment
	SB	Suspension number	Indication of suspended load
	σ	Sediment gradation	Indication of armouring effect.

Table 4.1: *List of dimensionless scour related parameters and their effect.*

influence on the performance of the scour prediction models. Of course the papers accompanying these studies are sparse with self-criticisms, but their choice of parameter is still interesting.

4.2.1 GP studies

Guven et al. (2009) [37]
The study of Guven et al. presents linear genetic programming (LGP) as an extension to GP, for the prediction of scour depth around a circular pile due to waves in medium dense silt and sand bed. A relatively small data set of 38 field measurements was used to develop LGP models. According to Guven the predictions were a good approximation of the measured data and performed significantly better than the regression-based equation given by Sumer et al. (1992) (equation (2.20)). The dimensional and dimensionless parameters used in the models are given by equation (4.1) and equation (4.2). However, no parameter sensitivity was performed.

$$S = f\left(d_{50}, D, H_s, T_p, U_{fm}\right) \tag{4.1}$$

$$S/D = f\left(\theta, KC, Ns, Re_d, Re, \delta\right) \tag{4.2}$$

Azamathulla (2010[21]
In 2010 Azamathulla conducted a research on genetic programming to predict bridge pier scour, using a small set of field data only. It was concluded that the developed GP model predicted scour fairly accurately and performed comparably to existing formulas and regression models. However, the presented formula in the paper of Azamathulla seems to contain an error, since for standard values the predicted scour depths are in the order of $\cdot 10^{83}$. The parameter sensitivity analysis reported by Azamathulla showed that the following input parameters were of non-negligible influence on S/D:

$$S/D = f\left(Fr, D/h_w, h_p/h_w, d_{50}, \sigma\right) \tag{4.3}$$

4.2.2 Other soft computing studies related to scour

Kambekar and Deo (2003)[38]
The study of Kambekar and Deo focuses on pile group scour in wave conditions using multiple types of neural networks (NN). The database consisted of measurements taken at a group of piles supporting a pier situated at a coastal location off Japan. The NN is used to estimate scour depth as well as the scour geometry for a group of piles supporting a pier. Kambekar and Deo's study uses two sets of input parameters that are listed below in equation (4.4) and equation (4.5), in which N_s is the so called sediment number given by equation (A.17). It was found that the type of neural network had a marginal influence on the performance and that all neural networks gave a better scour prediction that statistical curve fitting. However, the database was very small, so it is easy for the NN to be overfitted. A preference was expressed to the set that contains the raw data parameters (4.4), since it was concluded that the flexibility of the neural network suffered when input parameters were entered in a combined form, rather than individually. This can be something to consider for the GP as well.

$$S/D = f\left(H, T, h_w, D\right) \tag{4.4}$$

$$S/D = f\left(Re, KC, \theta, N_s\right) \tag{4.5}$$

$$N_s = \frac{u_c}{\sqrt{g\Delta d_{50}}} \tag{4.6}$$

Jeng et al. (2005)[39]
A very thorough study has been conducted by Jeng et al. on topic of scour depth around bridge piers in current only conditions. Multiple types of neural networks are used to estimate the equilibrium scour depth, followed by a sensitivity analyses on the input parameters for each NN. The database that was used consisted of a large amount of data from different sources. All of the above, lead to the conclusion that this paper is a good reference. The parameters that were considered most important for the dimensional and dimensionless case are:

$$S = f\left(\rho_w, \mu, U_c, U_{crit}, h_w, d_{50}, D\right) \tag{4.7}$$

$$S/D = f\left(\frac{U_c}{U_{cr}}, \frac{U_c}{\sqrt{gh_w}}, \frac{h_w}{D}, \frac{D}{d_{50}}, \frac{\rho_w U_c D}{\mu}\right) \tag{4.8}$$

Raaijmakers (2006) [3]
In 2006 Raaijmakers performed a study similar to this current study on the prediction of equilibrium scour depth around circular piles valid for all hydrodynamic conditions, using neural networks. It was found that only for scour prediction for each condition alone (i.e. current, waves, combined current and waves) the neural networks yielded reasonable to good results. Compared to the NN, existing scour prediction formulas showed larger differences between measurements and predictions. Nevertheless, also highly unrealistic values were found in case the input patterns were not covered by, or not in the vicinity of the inputs used during training (meaning when the NN was applied for extrapolation). The input parameters consisted of the following set given by equation (4.9) for dimensional NN and equation (4.10) for dimensionless NN.

$$S = f\left(h_w, h_p, D, H_s, T_p, U_c, d_{50}, \rho_s, \alpha\right) \tag{4.9}$$

$$S/D = f\left(d_*, D/h_w, h_p/h_w, \cos\alpha, U_{rel}, Fr, KC^{-1}, mob\right) \tag{4.10}$$

Ayoubloo et al. (2010) [40]
The study of Ayoubloo et al. provides a good reference for soft computing scour in wave only situations. In this study an NN was created to provide a more accurate estimation of wave scour around circular piles. Dimensional as well as dimensionless parameters were used as given by equation (4.11) and equation (4.12). A dataset was used that contain a sufficient amount of data, with scour depths ranging from 0 to 1.5 S/D. It was found that the dimensionless input set provided more accurate results. To determine the relative importance of the input parameters a sensitivity analysis was performed and it was found that the Keulegan-Carpenter number was distinctively the most important in this wave related study.

$$S/D = f\left(d_{50}, D, T, U_m, U_{fm}\right) \tag{4.11}$$

$$S/D = f\left(Re, KC, \theta, N_s\right) \tag{4.12}$$

4.3 Target parameter

Except for Kambekar and Deo [38] most studies in section 4.2 report improved results when a dimensionless input parameter set is used. Furthermore, it is easier to relate laboratory tests to actual field data with dimensionless parameters. Therefore it is chosen to use dimensionless input parameters and the dimensionless target parameter S/D, which is the equilibrium scour depth relative to the pile diameter.

4.4 Input Parameters

Based on the theoretical background of chapter 2 which can be found summarized in 4.1, and the most reliable reference studies of section 4.2, the dimensional input parameters that are presented in table 4.2 are selected for the GP. This table consists of three separate lists, one for each [C], [W] and [CW] condition. The idea behind this is that first these cases will be modeled separately with the GP, before creating one overall scour prediction formula. During the course of this study, this table will be updated until it only contains the most important parameters to describe the scour phenomenon.

[C]	[W]	[CW]	**Name**
σ	–	σ	Sediment gradation
d^*	d^*	d^*	Sedimentological diameter
Δ	Δ	Δ	Relative density
$\frac{D}{h_w}$	$\frac{D}{h_w}$	$\frac{D}{h_w}$	Shallowness
–	$\frac{D}{L}$	$\frac{D}{L}$	Diffraction
$\frac{h_p}{h_w}$	$\frac{h_p}{h_w}$	$\frac{h_p}{h_w}$	Relative pile height
Fr	–	Fr	Froude number
–	Fr_w	Fr_w	Froude wave number
–	KC	KC	KC number
mob	mob	mob	Relative mobility
Re_c	–	Re_c	Reynolds number [C]
–	Re_w	Re_w	Reynolds number [W]
SB	SB	SB	Suspendion number
u_{crit}	–	u_{crit}	Relative critical velocity
–	–	u_{rel}	Relative velocity
${w_s}^*$	${w_s}^*$	${w_s}^*$	Particle fall velocity

Table 4.2: *Dimensionless input parameters for GP*

Chapter 5

Literature study GP

Genetic programming (GP) is a collection of evolutionary computation techniques that allow computers to solve problems automatically, without requiring the user to know or specify the form or structure of the solution in advance. It is a subdivision of genetic algorithms, i.e. algorithms that use a form of the evolution theory of Darwin to obtain results. The main difference between genetic programming and other genetic algorithms is the representation of the solution. Generally genetic algorithms create an output in the form of a string of a number, while the solution of genetic programs is an optimized computer program.[41] In the case of equilibrium scour depth, the GP will generate a formula that estimates scour depth for a set of input variables. In figure 5.1 a simplified flowchart of a genetic program is given.

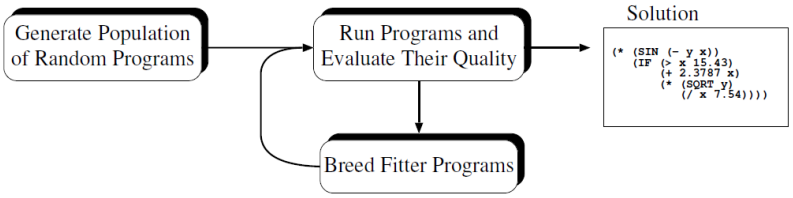


Figure 5.1: *Basic GP flowchart (Poli et al.[42])*

Evolution is embodied in GP by the survival of the most effective programs. By means of competition only the fittest programs are selected and cross-bred, resulting in convergence towards an optimal solution. This happens according to four steps described by Koza[43]:

1. Generate an initial population of random compositions of the functions and terminals of the problem (computer programs).
2. Execute each program in the population and assign it a fitness value according to how well it solves the problem.
3. Create a new population of computer programs.

i Copy the best existing programs

ii Create new computer programs by mutation.

iii Create new computer programs by crossover (sexual reproduction).

4. The best computer program that appeared in any generation, the best-so-far solution, is designated as the result of genetic programming.

Since its inception twenty-five years ago, GP has evolved rapidly, with new ideas, techniques and applications being constantly proposed.[42] Entire books can be written about this subject. However, this chapter will focus only on general genetic programming and the techniques used in the following research. First an overview of the nomenclature and architecture of GP will be given, after which steps 1 to 4 will be dealt with in section 5.2-5.5.

5.1 GP architecture

In GP, programs are usually expressed as syntax trees. There are other ways to represent a GP system, but tree-like structure is the original and most widespread type of GP.

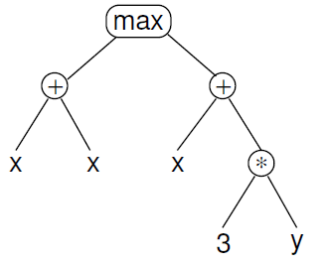


Figure 5.2: *Basic GP tree structure (Poli et al.[42])*

In figure 5.2 an example of the tree representation of the program $max(x + x, x + 3 * y)$ can be seen. In genetic programming, the variables and constants (in case of figure 5.2 x, y and 3) are called terminals, they are the leaves of the tree. The arithmetic operations ($+$, $*$ and max) are internal nodes called functions. All allowed functions and terminals together form the primitive set of a GP system. They define the ingredients that are available to GP to create computer programs.[42] The number of nodes (NoN) is defined as the total amount of functions and terminals in the tree. Again using the example of figure 5.2, the NoN of the tree is 9. This number is used to indicate the maximum program size.

Terminal set
The terminal set of the GP that will be used in this study can consist of 2 types of input:

- Input variables: such as x and y.
- Constants: pre-specified or created during the evolution process.

The terminal set of the GP system for the equilibrium scour depth estimation will be defined in chapter 4.

Function set
The type of functions in the function set depend on problem that needs to be solved. An example of a function set is visible in table 5.1. The function set of the GP for the equilibrium scour depth prediction will be defined in 6.2. Each function has a so called arity, which is the number of arguments that it needs.

Function	Arity
$-x$	1
$\exp x$	1
$\ln x$	1
abs x	1
x^2	1
$\sqrt{x}$	1
$\tanh x$	1
$(x + y)$	2
$(x - y)$	2
$(x * y)$	2
(x/y)	2
$pow(x, y)$	2

Table 5.1: *Function set GP*

The function set should be able to fulfil the following three requirements: type consistency, evaluation safety and sufficiency.[42] The first two terms are often grouped together and known as closure.[43]

Type consistency is simply the consistency of the in- and output form of each individual function, i.e. all functions must return values of the same type (e.g. only numbers or Booleans). If there is a type inconsistency, arbitrarily mixing of the nodes in subtree crossover and mutation might produce illegal type mismatches. Sometimes it is possible to cheat a little with the type consistency by introducing an automatic conversion mechanism between types. For example, by converting Booleans into numbers by treating the Boolean false as -1 and the Boolean true as 1. This could be introduced in the equilibrium scour depth equation as a way of handling specific types of hydrodynamic conditions (i.e. current-only, wave-only and combined wave and currents). Unfortunately this advanced GP setting will be beyond the scope of this study.

The other component of closure is evaluation safety. During run time it is possible that an evolved expression fails, for example, by dividing by 0. This is why evaluation safety is required. It usually consists of a protected version of a function that tests for potential problems with the input parameters before executing the function.

The last property a function set should have is sufficiency. The GP should be able to provide at least one solution to the problem at hand using the elements of the terminal and function set.

5.2 Step 1: Initialization of population

The first step of the genetic program is to create an initial population set. Although the shape and size of the initial trees can be lost within a few generations, it is still important for the success of the GP. It determines from the infinite possible programs, which ones will be explored by the genetic program.

An important way to control GP is through the size of the population. It depends on the application which parameter settings are optimal, but in general larger populations are preferred, limited only by the time the system needs to evaluate the fitness functions.

The individuals in the initial population are typically randomly created, but it is also possible to start with a given structure, as will be discussed at the end of this section. There are a number of ways to generate a random population, but in GP commonly a method known as ramped half-and-half is used. This method is actually a combination of two other methods, called the full and grow method.

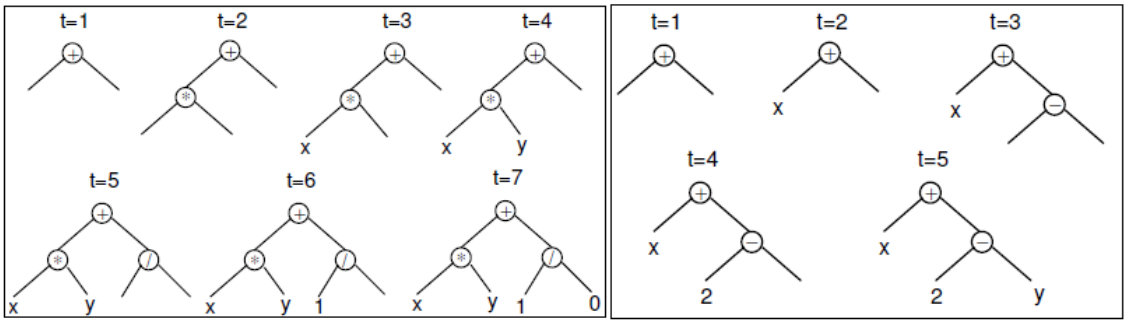


Figure 5.3: *Initial population methods; the full method (left) and the grow method (right) (Poli et al.* [42]*)*

The full and the grow method both create initial tree-structures that are limited by a user specified maximum depth. The full method generates full trees, i.e. with all the leaves at the same depth. The GP continues taking random functions from the function set until the tree is filled, after which the program switches to the terminal set to select constants and variables to fill the leaves. An example of this can be seen on the left side of figure 5.3 for the construction of a tree with depth limit 2. The only difference in shape between individuals exists when functions with different arities are chosen. This is the drawback of the full method; even with mixed-arity functions, the range of program sizes and shapes in the initial population is rather limited. [42]

The grow method, on the contrary, creates trees with a more varied size and shape. The nodes are selected from the function set as well as the terminal set, which means a branch in the tree can have a leave before it reaches the depth limit. If a branch in the tree does reach the maximum depth, only terminals can be chosen to ensure the depth limit. An example is shown in figure 5.3 on the right.

To ensure a maximum variety of sizes and shapes in the initial population Koza [43] proposed a combination of the two methods, which is now widely used, and named it ramped half-and-half. This method constructs half of the initial population with the full method and the other half with the grow method, each for a range of depth limits (hence the term

ramped).
Although this method will create more variance in the tree-structures, its disadvantage is that it is difficult to control the statistical distributions of these variations. For example, the sizes and shapes of the trees are highly influenced by the ratio between the function and the terminal set. If there are significantly more functions than terminals, the grow method will almost always chose a function, which will make it behave quite similar to the full method. On the other hand, if there are considerably more terminals than functions, the grow method will practically only create very short trees. This shows that the small change of adding a function to the function set can have significant implications for the GP.

There are other methods to generate the initial population, such as the ramped uniform initialization method. This method allows the user to specify the amount of trees that need to be created for each tree size. The advantage of this method is that it enables the possibility of trees with some leaves very close to the root. A variable close to the root of the tree will exert greater influence on the solution than a variable at the maximum depth that has to go through many function operations. This means that the ramped uniform initialization method will perform well in asymmetric problems, where one variable in the terminal set is more important than the others. On the other hand, the previously described ramped-half-and-half method will perform well for problems that have variables that are all equally important, since it mostly generates trees with leaves at an equal distance from the root. [42]

It is also possible to have a not entirely random initial population. If something is known about the likely shape or size of the desired solution, trees having these properties can be used to initiate the population. However, it is possible that evolution will cause the population to move away quickly from this initial distribution. [42]

Both the ramped uniform initialization method and the method with prescribed components, would be interesting for the prediction of scour, since some relations between the parameters and the equilibrium scour depth are already known. For instance for [W] conditions it is known that KC is the most important value. This means this can be considered as an asymmetric problem, in which case the ramped uniform initialization method could be better suited than the regular ramped-half-and-half method. Unfortunately these settings are not available in the GP-program from M. Keijzer [44] that will be used. Hence this is more an indication of the possibilities with GP predictions.

5.3 Step 2: Selection

The second step in the genetic program is to select the best programs, which will create offspring in step 3 (section 5.4). First the fitness of each solution is calculated after which a ranking system will determine which program can reproduce.

Fitness function
The task of the fitness function is to translate the problems requirements to the GP system. Fitness can have many forms and can be measured in different ways. In the GP program by M. Keijzer [44] that will be used for the equilibrium scour depth prediction, the fitness of an individual program can be tested on four objectives. The first two fitness

functions are based on the difference between the calculated value (i.e. the scour depth given by the GP generated scour formula) and the target value (i.e. the actual measured scour depth). They are the coefficient of determination (CoD equation (5.1)) and the root mean squared error (RMSE equation (5.1)). CoD is an indication of the level of variance between the estimated and measured value. A high CoD suggests that the data is well explained by the formula and therefore of importance for finding the best scour prediction formula. If RMSE is chosen as an objective the GP will evolve towards a formula with a small error.

$$CoD = \left(\frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2}\sqrt{n(\sum y^2) - (\sum y)^2}} \right)^2$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} x - y}{n}} \tag{5.1}$$

in which:

y = Observed data point

x = Predicted data point

n = Total number of observed data points

The third objective is the fitness per node; a measurement to estimate the simplicity of the expression of the individuals. Parsimonious expressions are usually preferred, since large functions tend to make it difficult to evolve correctly and are of course less practical. The fitness per node (FpN) depends on the ratio of the CoD and number of nodes (NoN). A fitter program will therefore have a larger fitness per node.

The last objective is the called the unit error, which measures the dimensional error of the equation. For dimensional equations it gives an integer of the amount of units that the estimation is off. For dimensionless in- and output parameters this error is obviously always zero. If the resulting unit of the formula does not match the unit of the output parameter, it does not necessarily mean that the formula is incorrect. For example, there is always a possibility that a certain parameter (with its dimensions) was not included in the terminal set, but was compensated for in the GP by the evolution of an extra constant, or that the GP contains a constant that has units (i.e. $g = 9.81$ [m/s^2]).

It is also possible to rank the equations on multiple objectives. For the scour prediction model, the CoD, RMSE and FpN objective will be used. Having multiple objectives means that each experiment in the GP will return three best formulas; one for each objective. However, since the three objectives are closely related to each other, for one to excel, the others will have a high performance as well.

Ranking

In GP, the most commonly used method for selecting is tournament selection. A number of individuals, called the tournament size, is chosen randomly from the population. A

tournament takes place and the fitter individual is allowed to be a parent for the next generation. In tournament selection only the determination of which individual is better is of interest, i.e. it is irrelevant how much better. This is important for the selection pressure to stay constant. A strong selection pressure favors the fit individuals too much, which reduces diversity. A weak selection pressure does not discriminate based on fitness at all, which allows unfit programs to reproduce and this slows down the evolution. If the tournament size goes up, also the selection pressure goes up, and vice versa.

5.4 Step 3: Recombination and Mutation

The third step in genetic programming is the recombination and mutation of the programs selected by Step 2 to generate a new generation of program individuals. This is done by copying the best existing programs and by creating new programs with mutation and crossover (reproduction).

Mutation
In subtree mutation the genetic program selects a random mutation point and substitutes the subtree that roots in that point with a randomly generated subtree.

Crossover
During cross over new individuals are created by the exchange of subtrees. For each of the two program-parents a randomly picked crossover point is determined. Followed by the creation of the offspring by switching the subtrees of the two parents at the given crossover points. This is illustrated in figure 5.4. It is possible for individuals to be selected to be parents multiple times and hence take part in the creation of multiple offspring programs. The crossover points are selected randomly, but not with a uniform probability. Since

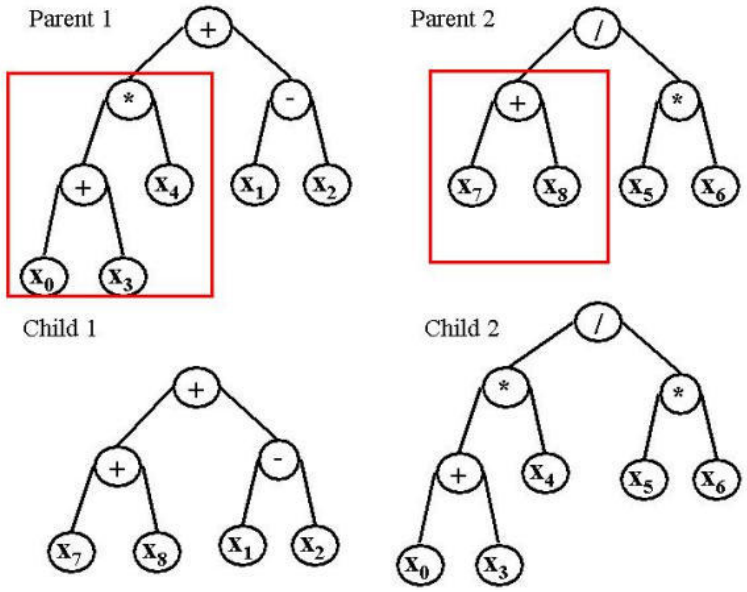


Figure 5.4: *Crossover of GP tree structure (Fernandez[41])*

generally there are more leaves than nodes, this would mean that if the probability of

crossover points was equal, most crossovers would just simply be the swapping of two leaves. As an alternative Koza [43] suggested to choose functions (i.e. nodes) 90% of the time and terminals (i.e. leaves) 10% of the time.

5.5 Step 4: Solution

Step 2 and 3 are repeated for endless generations until the program is told to stop. This can either be by prescribing a maximum amount of generations or a limited amount of time. The GP will then give the best-so-far solution over all generations.

To obtain an accurate result with the GP, step 1 to 4 need to be repeated multiple times to create formulas based on different initial populations. In this study the following terminology will be used: an experiment is each time the GP is restarted. A 'pool' contains multiple experiments with a specific set of settings (such as the function set). A 'run' consist of multiple pools with different function settings, but the same input database.

Chapter 6

GP procedure for scour experiments

Genetic programming is a very versatile modeling technique. When applied to a specific problem, it will not produce just one optimal formula. Due to the wide range of settings, each time the GP is initiated, it will evolve towards a local best formula. Only by optimizing the settings, the GP can be steered towards one overall optimal formula; a formula that has the absolute highest performance possible for the given database. This chapter presents the methodology behind the GP experiments in more detail, as well as a list of the GP settings that are used and some interim results that lead to the best performing formulas presented in the next chapter.

6.1 GP Methodology

To find the best scour prediction method the GP settings must be adjusted to the needs of the scour problem. This is done by defining the architecture of the GP. As explained in chapter 5, the main architecture is given by the terminal set, function set and general settings such as the initialization method and formula length.

The terminal set is already defined in chapter 3 and 4. These chapters discuss which data sets are reliable and relevant, and which parameters are important. Together they form the input database for the GP, called the terminal set. The next step is to adjust the function set and general settings. This is done in the following section (section 6.2) by analyzing the effect of one specific change in the general settings on the performance of the GP.

When the initial architecture of the GP is correct the experiments can start. To simplify matters for the GP, its task is first to find a suitable formula for the current-only case [C], wave-only case [W], and combined current and wave case [CW] separately. As can be read in chapter 2, the physics behind current-induced scour significantly differs from wave-induced scour, which makes it difficult for the GP to find a formula for these effects in one go. When a sufficient formula is found for the individual conditions, the three

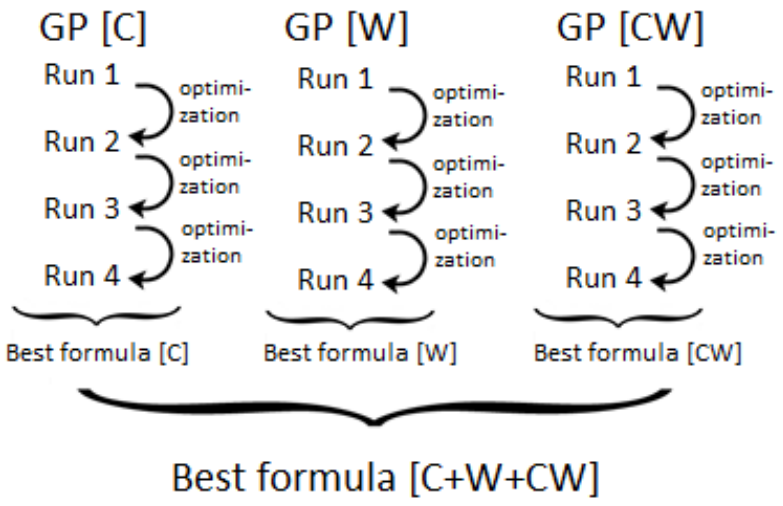


Figure 6.1: *GP procedure: Methodology*

scenarios will be combined in one formula for all conditions. This is also depicted in figure 6.1.

The methodology to find the individual formulas for [C], [W] and [CW] is based on four runs. As explained in chapter 5 a run is a GP sequence with a specific terminal set. A run can consists of multiple experiments, where the GP starts over with a new initial population, but all settings remain the same. After each run some optimizations are done, which adjust one or more of the GP characteristics.

Run 1: Uses all parameters from chapter 4 and determines with parameter sensitivity which parameters are important.

Run 2: Uses the most relevant parameters from Run 1 and adds weights to reliable and relevant tests. Another parameter sensitivity test is performed.

Run 3: Uses the most relevant parameters from Run 2 and adds runtime to give the program extra time to evolve.

Run 4: Uses known combinations of parameters as input parameter for the GP.

Between runs the results are analyzed mathematically and physically. During the mathematical analysis the formulas are judged on the coefficient of determination (CoD, equation (5.1)), the root mean square error (RMSE, equation (5.1)), the fitness per node (FpN) and number of nodes (NoN). These can be grouped together in an overall fitness term (OaF, equation (6.1)). The weights of formula 6.1 (w_1, w_2, w_3) can be adjusted to shift the importance of each coefficient.

$$OaF\,(w_1, w_2, w_3) = \frac{w_1 \cdot (1 - RMSE) + w_2 \cdot CoD + w_3 \cdot FpN}{w_1 + w_2 + w_3} \tag{6.1}$$

The physical analysis consists of an investigation of the formula behavior when key parameters are varied over a certain range.

Optimization
The last step in finding the ultimate formula is by adding constants. The GP-program by M. Keijzer [44] that is used for the experiments is incapable of optimizing constants, therefore this needs to be done by a separate fitting program.

6.2 GP settings

Chapter 5 explained in detail the workings of a general genetic program. Before tests can be done with the GP, these generic settings need to be adjusted to fit the needs of the scour problem. These settings are based on values found in literature as well as on some simple experiments with the scour database. In table 6.1 an overview of the settings for the scour GP is shown. These values are initial settings; most of the values will remain constant, but some will be optimized between runs. The meaning of all terms can be found in chapter 5.

Characteristic	Value
Objectives	$RMSE$, CoD, FpN
Experiment duration	5 minutes
Number of experiments	20
Terminal set	Chapter 4
Function set	Table 5.1
Population size μ	250
Reproduction size λ	500
Maximum initial program size	15
Maximum program size	50
Population initialization method	Ramped-half-and-half
Size of tournament selection	3
Constant probability	0.05

Table 6.1: *GP characteristics*

The most important settings from table 6.1 are discussed more extensively in the following sections.

6.2.1 Duration and number of experiments

An important setting for the quality of the results is the duration of the test runs. Obviously the longer the GP has to evolve, the better the results. However, this is not very time efficient. To find the optimal ratio of the GP between time efficiency and performance a test is done with 540 random data sets (180 from each category: [C],[W] and [CW]) and a range of experiment durations. For each duration, 10 experiments are executed with 3 pools (different function sets) and 3 objectives. This means that in total each duration is represented by 90 of its best performing formulas.

For each time unit the values of the objectives are averaged, as well as the overall fitness value from equation (6.1) with weights OaF(1,1,1). The results are shown in figure 6.2.

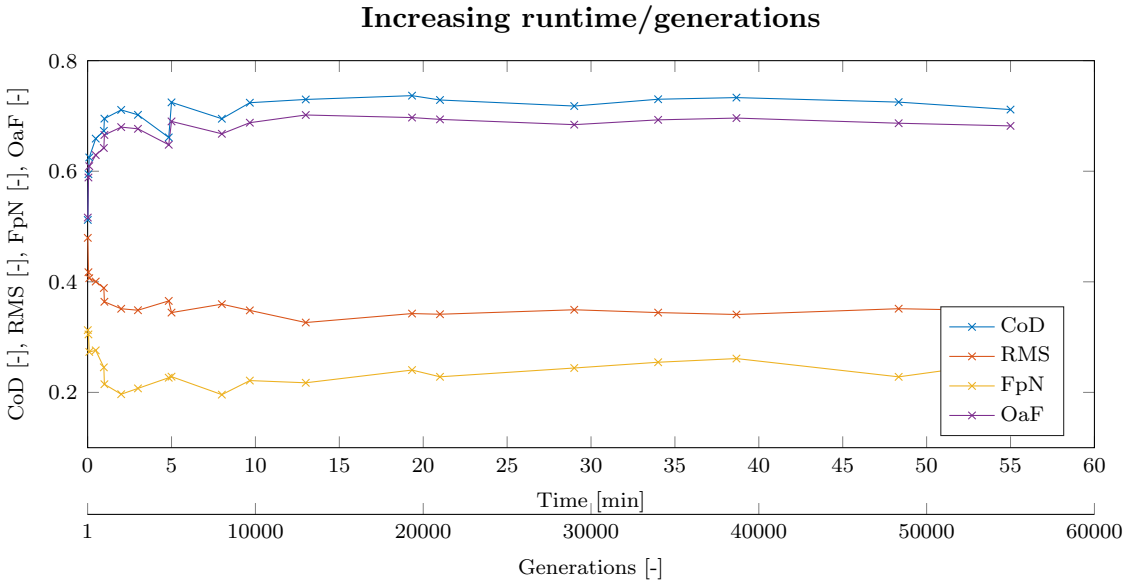


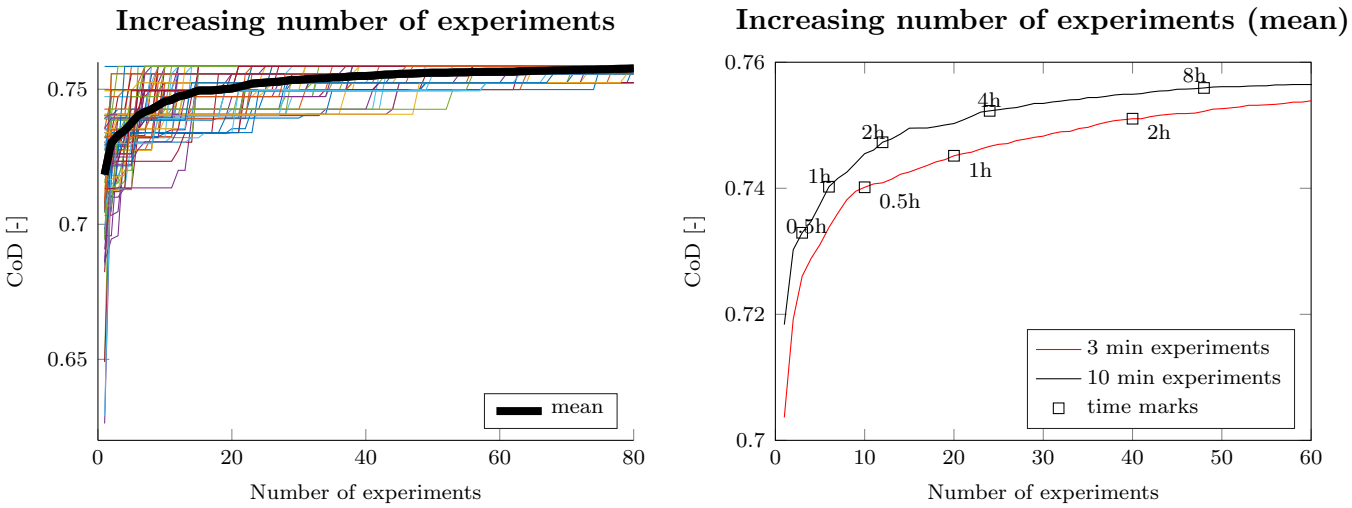
Figure 6.2: *GP setup: influence of the amount of runtime/generations versus the performance of the genetic program*

It is visible that the performance fluctuates of experiments with a short evolution time. However, already after 10 minutes the GP seems to have reached an equilibrium. The results for experiments that have run for 10 minutes do not differ from those that had 50 minutes to evolve. Therefore it would seem unnecessary to extent the GP duration of each experiment beyond 10 minutes.

As an indication also the number of generations are given. In this test 10 minutes equals about 10.000 generations, but this conversion depends highly on the size of the database, which increases the computation time.

Figure 6.3(a) shows the importance of the next setting: the number of experiments. Also for the number of experiments it is true that a higher number is better, but less time efficient. Figure 6.3(a) is created by using 540 random data sets (again 180 sets from [C], [W] and [CW]), a constant time duration of 10 minutes and a 100 experiments. These experiments are shuffled in order, creating 100 different trajectories. The graph displays the CoD of the best scour equation found in the experiments, so for 10 experiments it displays the best of the 10 CoD values. If the 11th experiment produces a better scour formula the CoD increases, if not, the value at 10 is equal to the value at 11. The bold black line in figure 6.3(a) shows the mean value of all trajectories. Of course the probability of the highest value being among 1 experiment is smaller than if you run 100 experiments, which is why on average the performance of the best scour equation improves with increasing amount of experiments.

Figure 6.3(b) shows the mean performance of figure 6.3(a) and the same test repeated for experiments with 3 minute duration. The learning curve of the GP is quite steep up to 10 experiments for both durations, therefore this is considered to be the bare minimum of required number of experiments. Interesting to see is that although the experiments of 10 minutes seem to have the upper hand, if the time performance ratio is concerned,



(a) (b)

Figure 6.3: *GP setup: influence of the amount of experiments versus the performance of the genetic program*

shorter tests are more efficient. For instance, if the total time to run the GP is 2 hours, a better performance is obtained with experiments 40 experiments of 3 minutes than 12 experiments of 10 min.

Using this information and the data from figure 6.2 it is chosen to use a compromise between efficiency and duration. The 3 minute-experiments are better in figure 6.3(b), but their performance is in the unsteady region in figure 6.2. 10 minutes is better in figure 6.2, but very time consuming, as can be seen in figure 6.3(b). The compromise is an intermediate value of 5 minutes (or 5000 generations) and a minimum amount of experiments of 20, a number well beyond the steep curve of figure 6.3(b).

6.2.2 Function table and constant probability

Another important setting is the function set. This set contains a table with all the functions the GP can combine to create a scour formula. The available functions are given by table 5.1. The weights of the function table can be adjusted to increase or decrease the probability that a function is chosen. To find the optimal settings for the weights in the function set, a test is performed with 540 random data sets (again 180 sets from [C], [W] and [CW]), a constant time duration of 10 minutes and 10 different function sets, given by table 6.2. Each function set is tested for 10 experiments and the CoD values of these 10 experiments are averaged and shown in figure 6.4.

The maximum CoD values of the different combinations are relatively similar to each other. This suggests that if the number of experiments and duration of the test is sufficient, the setting of the function set is of minor importance. Setting 6 and 9 however, have a significantly lower average than the other tests. This can be attributed to the imbalance between the weight of the functions and the weight of the parameters as was explained in the previous chapter.

Function	Weight									
	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8	Set 9	Set 10
$-x$	1	1	1	1	1	1	1	0	0	0
$\exp x$	1	1	1	1	1	3	1	0	3	5
$\ln x$	1	1	1	1	1	1	1	0	1	0
abs x	0	0	0	0	0	0	0	0	0	0
x^2	1	1	1	1	1	1	1	0	1	0
$\sqrt{x}$	1	1	1	1	1	1	1	0	0	0
$\tanh x$	1	1	2	1	3	3	1	0	3	5
$(x + y)$	1	1	2	5	5	5	2	2	3	5
$(x - y)$	1	1	2	5	5	5	2	2	3	5
$(x \cdot y)$	1	1	2	5	5	5	2	2	3	5
(x/y)	1	1	2	5	5	5	2	2	3	5
x^y	1	1	1	1	1	1	1	0	3	0
Parameter weight	1	2	2	2	2	2	2	1	4	4
Constant probability	0.1	0.1	0.1	0.6	0.7	0.1	0.1	0.1	0.1	0.6
Total weight functions	11	11	16	27	29	31	15	8	23	30
Total weight parameters	8	16	16	16	16	16	16	8	32	32

Table 6.2: *Function set GP*

Although almost all settings found at least one formula with CoD ≈ 0.75, the initial function set of the GP will be set to the settings of number 7. This setting has a good balance between parameters and functions and it is the setting for which the highest CoD was found.

It is chosen not to use the absolute function abs x. This would not contribute to the formula, since there are no negative parameters in the database. The constant probability remains set at 0.1. Some tests have been executed with increased constant probability and with constants imported through the terminal set as input parameters, but they either disappeared during evolution or had a negative effect on the performance of the formula. Therefore it is chosen to keep the constant probability rate low and add constants later with an external fitting program.

6.3 Interim results

Using the settings described in the previous section the GP is ready to find the ultimate scour formula. The interim results of the methodology given by figure 6.1 are discussed here.

6.3.1 Current-only [C]

The first step towards a good scour prediction formula, is finding a subformula that can predict scour in the special case of current only conditions. This section presents the optimization process of finding this equation based on a database containing only current-only tests.

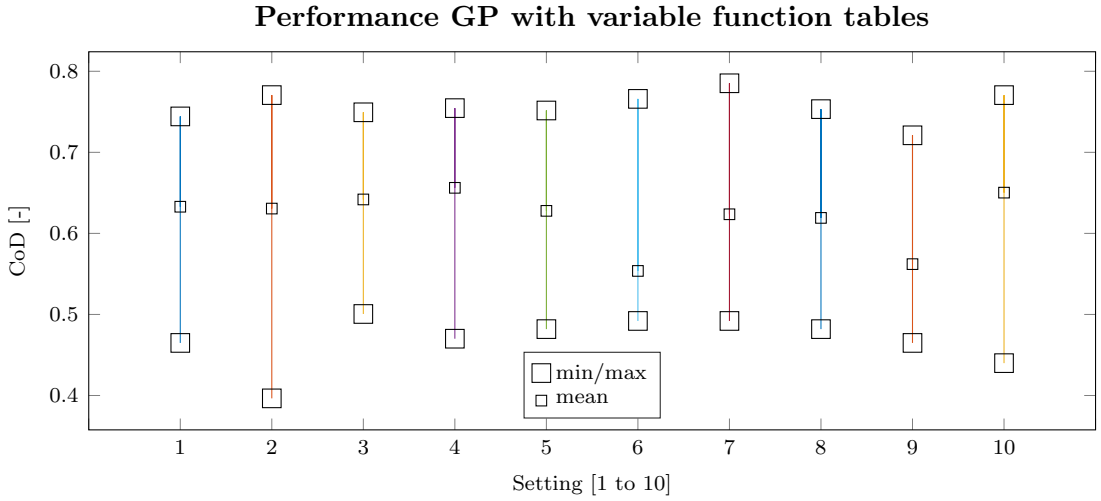


Figure 6.4: *GP setup: influence of the function table on the performance of the genetic program*

Run 1: Parameter sensitivity

In table 4.2 a large selection of input parameters is given that all could be of some importance for current-induced scour. All these parameters are used in the terminal set of the first run. This run consists of multiple pools (equal to the number of input parameters plus 1), in which in each of the pools the weight of one parameters is set to zero. In others words, all other settings remain the same, but in each pool 1 parameter is omitted. The last pool contains all parameters.

Mathematical analysis

The GP has found 60 equations for every pool (20 experiments per pool with each 3 objectives). The results are given by table 6.3. These results represent the average values of the top 10 best equations. Which equations are best depends on the value of OaF(1,1,1) equation (6.1). Based on these numbers the parameters σ, ${w_s}^*$, Fr, d_* all seem important. The performance of the other variables are too close to each other to give a definite observation of their usefulness. The lack of influence of mob and u_{crit} seems remarkable, since most scour studies assume these are very important parameters for current-only scour. However, this can be explained by the setup of this test. Since mob and u_{crit} have exactly the same behavior, if one of these parameters is omitted, the GP can simply use the other one. Further testing is required to discover their importance and which parameter is more adapt in predicting scour.

Additionally it can be deducted from table 6.3 that the formulas are quite long. Extracting from this same test 10 other equations, also the best, but now the best including the requirement that the fitness per node (FpN) needs to be higher than 0.20, leads to a decrease in CoD of 2.7% and an increase in RMSE of 3.2%, but also a reduction of number of nodes of 30%. This seems like a reasonable price to pay for a more practical formula.

Also the importance of σ from section 3.3.1 is established. Its positive that such a strong relation has been found. Unfortunately, this also means that the tests that did not document the sediment gradation cannot be used in the following runs.

Omitted Parameter	CoD	RMS	FpN	NoN	OaF
$-\sigma$	0.694	0.343	0.182	44.2	0.511
$-{w_s}^*$	0.702	0.341	0.197	34.5	0.519
$-\frac{h_p}{h_w}$	0.718	0.334	0.183	46.7	0.522
$-d_*$	0.710	0.330	0.191	41.5	0.524
$-mob$	0.723	0.333	0.183	47.8	0.525
$-Fr$	0.706	0.330	0.202	32.5	0.526
$-\Delta$	0.722	0.330	0.188	43.7	0.527
$-SB$	0.721	0.330	0.195	38.6	0.529
$-\frac{D}{h_w}$	0.732	0.326	0.188	45.8	0.531
$-Re_c$	0.724	0.328	0.202	35.2	0.533
$-u_{crit}$	0.728	0.322	0.196	39.7	0.534
All incl.	0.731	0.323	0.196	39.5	0.535

Table 6.3: *Performance Current Run 1, parameter sensitivity*

Best equation

Since the parameter sensitivity of table 6.3 was inconclusive, another check is done to get a better idea which parameters are important. For this test the formula with the highest performance is analyzed, given by equation (6.2). To keep it manageable this is the best equation with the implication that the number of nodes (NoN) is smaller than 25. Since the goal of this particular exercise is to establish important parameters for future runs, the shape of the formula is irrelevant for now.

$$\begin{aligned}\frac{S}{D} = \tanh\Bigg\{\tanh\Bigg[\tanh\left(u_{crit}\exp\left(\tanh\left(u_{crit}\right)+u_{crit}\right)-\sigma-\frac{D}{h_w}\right)\\ +\tanh\left(\frac{D}{h_w}\sigma-{w_s}^*+\frac{h_p}{h_w}\right)\Bigg]\Bigg\}+\tanh\left(Fr\right)-\tanh\left(\sigma\frac{D}{h_w}\right)+\frac{h_p}{h_w}\end{aligned} \tag{6.2}$$

This formula confirms what was expected from the results of table 6.3: Δ, SB, Re_c are not essential in the formation of a good formula. Although they are present in the terminal set, the evolution process of the GP filters them out of the best equations. In table 6.3 it was suggested that $\frac{D}{h_w}$ was not that important either, but this parameter has surfaced in this formula and even occurs multiple times. d^* on the other hand, was considered important, but does not appear. Equation (6.2) is a formula with u_{crit}, but the second and third best formula are with mob. Therefore still no conclusion can be drawn which of the two parameters performs better.

The next step is to check the influence of each parameter in the formula. This is tested by alternately replacing one type of parameter by a constant value. To preserve the behavior of the formula as much as possible, this constant will be equal to the mean value of the parameter. First all occurrences of u_{crit} will be replaced, followed by σ, $\frac{D}{h_w}$, ${w_s}^*$, $\frac{h_p}{h_w}$ and Fr. The results can be seen in figure 6.5(a).

The 2 dotted lines in figure 6.5(a) show the CoD and RMSE of equation (6.2). The blue and red dots show the CoD and RMSE if the parameter on the x-axis is replaced by a

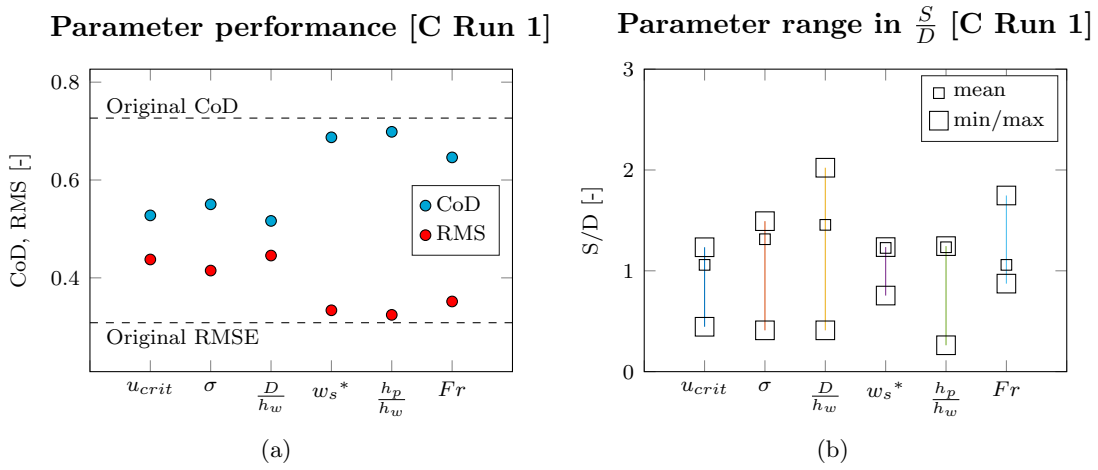


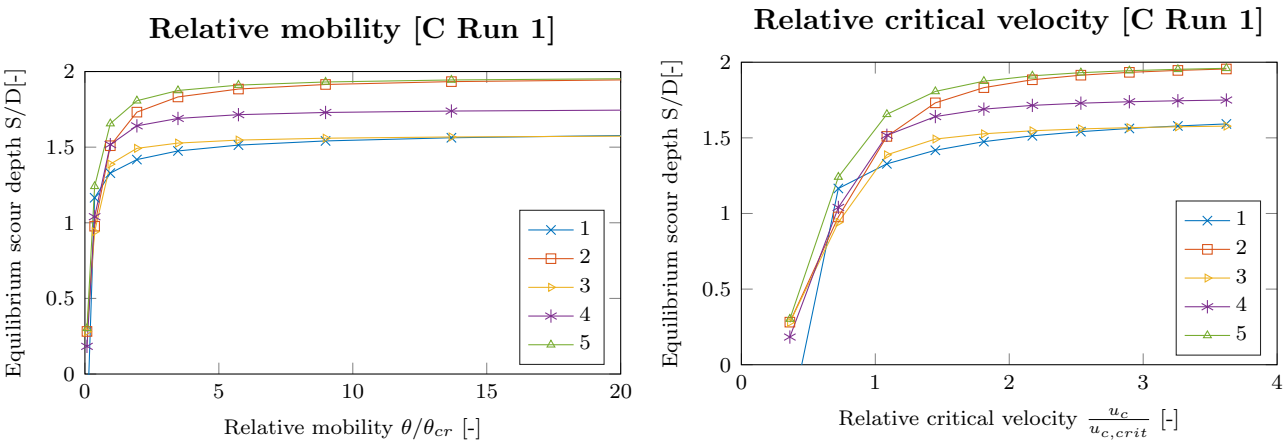
Figure 6.5: *Behaviour of individual components of the best performance formula from run 1*

constant. As can be seen in this figure, u_{crit}, σ and $\frac{D}{h_w}$ are definitely irreplaceable. When these parameters are turned to constants the CoD goes down and the RMSE goes up. This behavior is less visible with the other 3 parameters, but this can be explained with figure 6.5(b).

Figure 6.5(b) shows the reversed of figure 6.5(a). In this figure all parameters are constants (i.e. their mean value) *except* the parameter on the x-axis. This parameter is ranged from its minimum value to its maximum to see by which quantity $\frac{S}{D}$ increases of decreases. For instance, the influence of $\frac{D}{h_w}$ is much greater than that of ${w_s}^*$, since if the $\frac{D}{h_w}$ is changed, the final answer can range from $0.4\frac{S}{D}$ to $2.1\frac{S}{D}$, while if ${w_s}^*$ is ranged from its minimum to its maximum value the influence on the final answer is much smaller, it ranges only from $0.8\frac{S}{D}$ to $1.2\frac{S}{D}$.

More importantly, this figure explains why the influence of ${w_s}^*$ and $\frac{h_p}{h_w}$ is small in figure 6.5(a). This can be seen by the small squares that represent the mean value of figure 6.5(b). As can be seen the mean $\frac{S}{D}$ value of ${w_s}^*$ and $\frac{h_p}{h_w}$ are practically equal to its maximum value. This means that the spread of the variable is very small, i.e. the values are almost constant. For $\frac{h_p}{h_w}$ this is caused by the lack of spread in the database. Only a few tests have been done with submerged piles ($\frac{h_p}{h_w} < 1$), so there is not enough data for the GP to train for these cases. The small spread of ${w_s}^*$, as well as for the other parameters, can be accounted for by the frequent use of the hyperbolic tangent. The nature of the hyperbolic tangent is that there is a very small and steep increasing part and a very large constant part. Once the value of the parameter is beyond the increasing part of the curve its output will remain the same. If there are many points in the flat part of the hyperbolic tangent, the mean value will of course be close or even similar to the maximum value.

A mentioned above, there are not enough differences in $\frac{h_p}{h_w}$ to give a correct representation. Since the behavior of submerged piles is know, the $\frac{h_p}{h_w}$ parameter will be removed from the parameters for now and added separately as will be discussed in chapter 7.3.4. ${w_s}^*$ and Fr will remain in the formula.



(a) (b)

Figure 6.6: *Formula behavior: relative mobility. C Run 1, formula 1 to 5.*

Physical analysis

For the physical analysis the formula behavior of the 10 best performing equations is investigated when the key parameters are varied over a certain range. All other values are kept constant at a the standard values found in table 6.4. The most interesting cases are u_{crit} and mob. They are varied over a certain range by varying u_c from 0 to 1.5 m/s. The results can be seen in figure 6.6(a) and (b).

The figure shows that the top 10 formulas are able to make a distinction between clear-water and life-bed scour. For u_{crit} this transition is as required around a value of 1. For mobility the scale is larger so it is more difficult to see, but the curve seems also to occur around 1. Unfortunately none of the tests makes the characteristic clear-water peak. The GP probably treats the tests around this peak as outliers, since it is easier to fit a smooth curve such as the ones visible in figure 6.6(a) and (b). Adding more weight to tests with $u_{crit} \approx 1$ might help the GP to make a better fit for this complicated transitional area.

Standard test values									
category	d_{50}	D	h_p	h_w	u_c	H_s	T_p	ρ_s	σ_g
$[C]$	0.0013	0.2	0.5	0.5	0.4	0	0	2650	1.2

Table 6.4: *Standard test values [C]*

Optimization technique for Run 2

To summarize the findings of Run 1: Δ, SB and Re_c can be removed from the terminal set, because they are not essential for a good formula. $\frac{h_p}{h_w}$ is, but there are not enough submerged piles, so this will be tested separately. Therefore $\frac{h_p}{h_w}$ is also removed from the input parameters. This leaves a terminal set with: σ, ${w_s}^*$, Fr, d^*, mob and u_{crit}. The latter two need to be tested separately to see which parameter is better suited for a scour prediction formula.

Furthermore, weight needs to be added for tests around $u_{crit} \approx 1$ to force the GP to create a clear-water peak. As well as putting more weight on reliable test and less weight

on questionable sources and irrelevant tests (with situations that do not occur in offshore situations).

Lastly the maximum node length is set to 40, to reduce the lengthy equations to a more practical format.

Run 2: Weights

The optimization suggestions of Run 1 are incorporated in this run. This resulted in 2 set ups: Run 2a with u_{crit} and Run 2b with mob.

Mathematical analysis
Table 6.5 gives the results of the second run. Of the 60 equations per pool, these are the average results of the top 10. With a decrease in the formula size and in the number of parameters in the terminal set, also the performance has decreased somewhat. However, the new formulas have a length between 20 and 30 nodes, which is a more workable formula length than the equations from Run 1.

Omitted parameter	CoD	RMS	FpN	NoN	OaF
$-u_{crit}$	0.649	0.379	0.194	28.2	0.488
$-\sigma$	0.581	0.399	0.187	25.1	0.457
$-\frac{D}{h_w}$	0.553	0.413	0.163	30.7	0.434
$-Fr$	0.676	0.348	0.223	19.4	0.517
$-{w_s}^*$	0.676	0.347	0.221	20.4	0.517
$-d_*$	0.681	0.347	0.220	21.2	0.518
All incl.	0.683	0.348	0.215	22.8	0.516

Omitted parameter	CoD	RMS	FpN	NoN	OaF
$-mob$	0.641	0.381	0.195	26.0	0.485
$-\sigma$	0.471	0.442	0.179	12.3	0.403
$-\frac{D}{h_w}$	0.552	0.411	0.174	23.9	0.438
$-Fr$	0.650	0.364	0.203	23.1	0.496
$-{w_s}^*$	0.688	0.344	0.212	24.9	0.519
$-d_*$	0.689	0.352	0.211	25.1	0.516
All incl.	0.688	0.351	0.206	26.9	0.514

Table 6.5: *Performance Current Run 2 a and b*

In terms of parameter sensitivity a clear division can be seen in table 6.5. σ, $\frac{D}{h_w}$ and u_{crit}/mob are absolutely essential, while Fr, ${w_s}^*$ and d^* form the second rank. Comparing the results of Run 2a (u_{crit}) and Run 2b (mob), u_{crit} has a slight preference. Although the CoD of the set with mob is higher, the benefit of using u_{crit} is that the RMSE is lower as well as the number of nodes in the formula.

Physical analysis
Despite the decrease in performance in table 6.5, some progress has been made. By adding weights, the GP successfully produced at least two formulas in Run 2b (number 2 and number 6 of the top best performing equations), that show the characteristic clear-water peak. This can be seen in figure 6.7. This figure has been made keeping all parameters equal to the standard values found in table 6.4 except for u_c, which is ranged from 0 m/s to 1.5 m/s. In the top 10 best performing equations of Run 2a with u_{crit}, no such behavior was found.

Formula 2 and 6 differ completely in lay-out, so it is not an easy task to extract the part that causes this small bump. It requires further investigation; when more formulas with the clear-water peak are created, some recurring combinations of functions might be found.

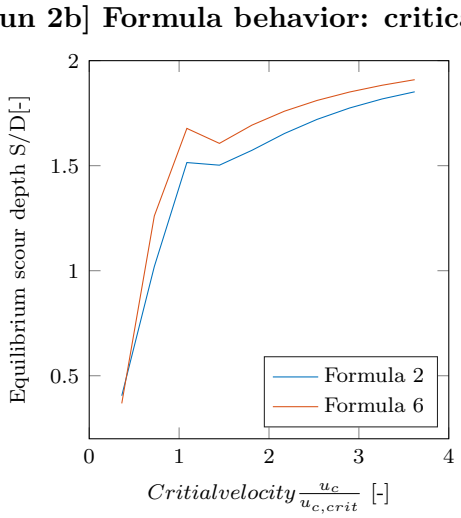


Figure 6.7: *Behavior of best formulas Run 2, Relative critical velocity*

Errors

To check if the weight system from section 3.4 works sufficiently and if the database does not contain any large errors, the performance of each individual data set from each source is evaluated. This is done by comparing the weight of a test to its mean error. The mean error for each test is calculated by extracting the 50 best equations of all runs so far and calculate the error of each data point for each of these formulas. The values are averaged per test and shown in 6.8. Also the minimum and maximum error are given in this figure.

There are over 1600 test in figure 6.8, all ranked on their mean error. As can be seen, about 10% of the tests have an error larger than 0.5 S/D and 8 tests even pass the error threshold of 1 S/D. This can be caused by two reasons: either the data is wrong or the GP is wrong. In case of the first reason, a mistake could have been made in the test set up, there could be ripples in the test section or the scour depth could be measured incorrectly because of global scour. In case of reason number 2, the GP may encounter a phenomenon that is difficult to incorporate, such as a transition between clear-water scour and life-bed or a phenomenon that has an insufficient amount of tests, so the GP is unable to train itself. This difference in reasoning is very important, since in case of incorrect data, the weight on the test must be reduced. In case of lack of training with the GP, the weight must be increased.

To take a closer look at the most erroneous tests, table 6.6 is created for tests with a mean error above 1 S/D. It can be seen here that 5 out of 8 tests have a u_{crit} value close to 1. Extra weight is given to these tests in order to help the GP with this difficult transition, but apparently there are still some tests that are difficult to model even with improved weights judging by their high errors. The extreme cases of Jain and Fisher have a lower weight, since their value of u_{crit} is far beyond regular values. The weight system worked well in this case. Their weight is low, so the GP does not have to focus on these tests. Which apparently it does not, since the error is very high.

Optimization technique for Run 3

The findings of Run 2 are as follows: adding weights and decreasing the number of nodes

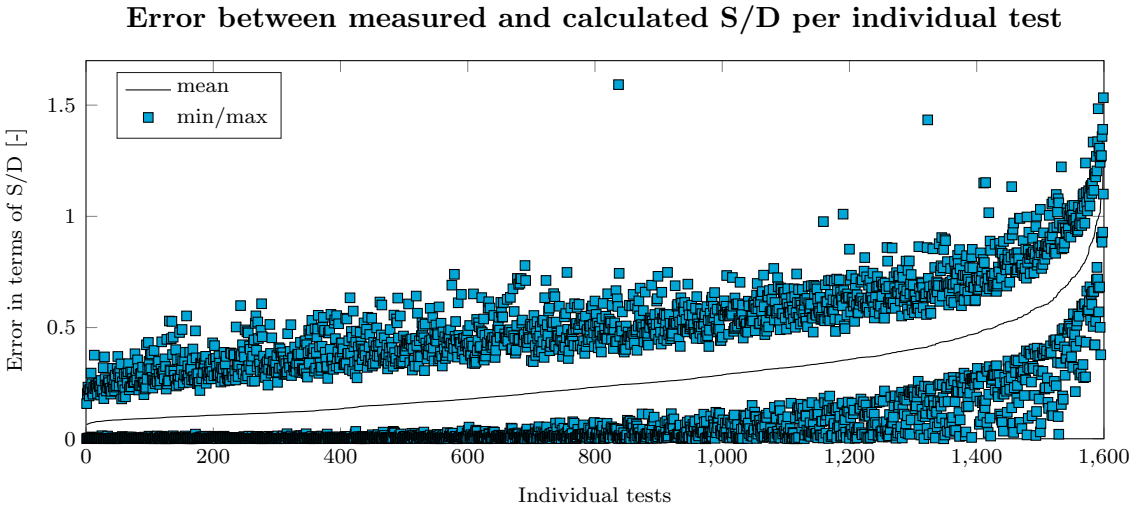


Figure 6.8: *Current Ranking Formula 2: individual mean error parameters*

Tests with dimensionless mean error $> 1 S/D$

	$W[-]$	$u_{crit}[-]$	$D[-]$	Error [S/D]	Error [m]
'Chabert and Engeldinger (1956)'	3	0.90	0.05	1.01	0.05
'Jain and Fischer (1979) '	2	6.27	0.05	1.27	0.06
'Jain and Fischer (1979) '	2	4.33	0.05	1.12	0.06
'Molinas (2003)'	5	1.02	0.18	1.00	0.18
'Molinas (2003)'	4	1.28	0.18	1.02	0.18
'Molinas (2003)'	3	1.79	0.18	1.08	0.19
'Qi (2013)'	5	0.89	0.08	1.10	0.09
'Simmaro, G. (from Lanca (2013))'	3	1.00	0.08	1.09	0.09

Table 6.6: *Error current Run 2*

had a positive effect. This will be used in future runs as well. Although by shortening the formulas the average performance went down, the workability was much better. Still no decision could be made regarding u_{crit} versus *mob*. u_{crit} had the benefit of shorter formulas, *mob* already produced a formula that could predict something resembling a clear-water peak. Since all other settings are already optimized, in Run 3 only the runtime will be increased to see which parameter evolves to the best equation.

Run 3: Additional runtime

Run 3 is executed using the optimization techniques from Run 2. The runtime per experiment is increased to 10 minutes and since all other parameters are already known, the focus is entirely on finding the best scour prediction formula with either u_{crit} or *mob*.

Mathematical analysis
The average results of the 10 best equations are shown in table 6.7. Again the results are very close to each other. Interesting to note is that according to these numbers u_{crit} has

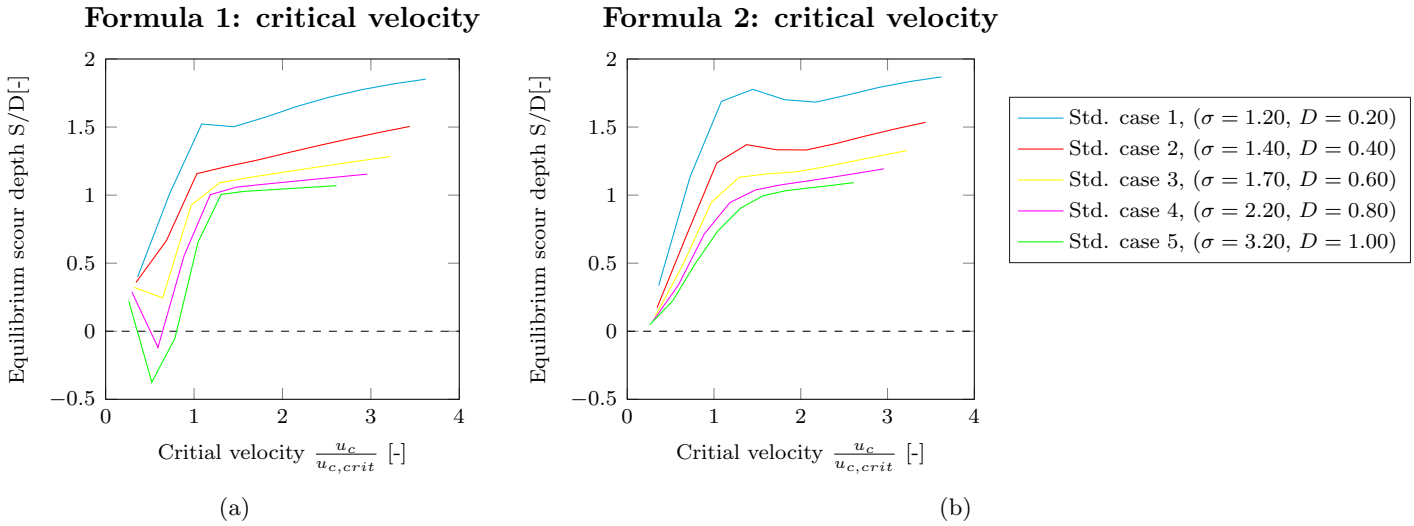


Figure 6.9: *Formula behavior: critical velocity. C Run 2. Best formula for range of input parameters.*

lost it advantage of being the shorter formula.

	CoD	RMS	FpN	NoN	OaF
Formula with *mob*	0.703	0.340	0.190	38.2	0.518
Formula with u_{crit}	0.698	0.335	0.185	40.1	0.516

Table 6.7: *Performance Run 3*

Physical analysis
The clear-water peak turns out to be the decisive factor; only formulas with *mob* are able to create this peak and still keep a high performance. Among these formulas only two formulas were found that had CoD$>$ 0.69, RMSE$<$ 0.35 and NoN$<$ 25. To make a decision between the two the behavior of the formula is plotted for u_{crit} versus S/D and σ versus S/D. Again these graphs are made with the standard values from table 6.4 unless specifically noted otherwise.

Figure 6.9a shows the behavior of the relative critical velocity versus the dimensionless equilibrium scour depth when u_c is ranged from 0 m/s to 1.5 m/s. As is visible here, the first formula (figure 6.9a (a)) already shows some strange behavior. For low values of u_{crit} and high values of σ the scour depth decreases, even becomes negative and than increases again. This is unexplanatory behavior, which would not happen in nature. Figure 6.9a (b) shows a better pattern, with a steep increasing clear-water part for $u_{crit} < 1$, the clear-water peak at $u_{crit} \approx 1$ and the life-bed part going towards an asymptotic value for $u_{crit} > 1$. This is similar to the behavior expected in theory as discussed in section 2.3.

What both graphs do show is that the clear-water peak decreases when σ increases. This could be realistic behavior, since if the sediment becomes more graded, the different sizes of sand can forms a protective armour layer as described.

Although formula 2 is already considered to be the better of the two, it is still interesting

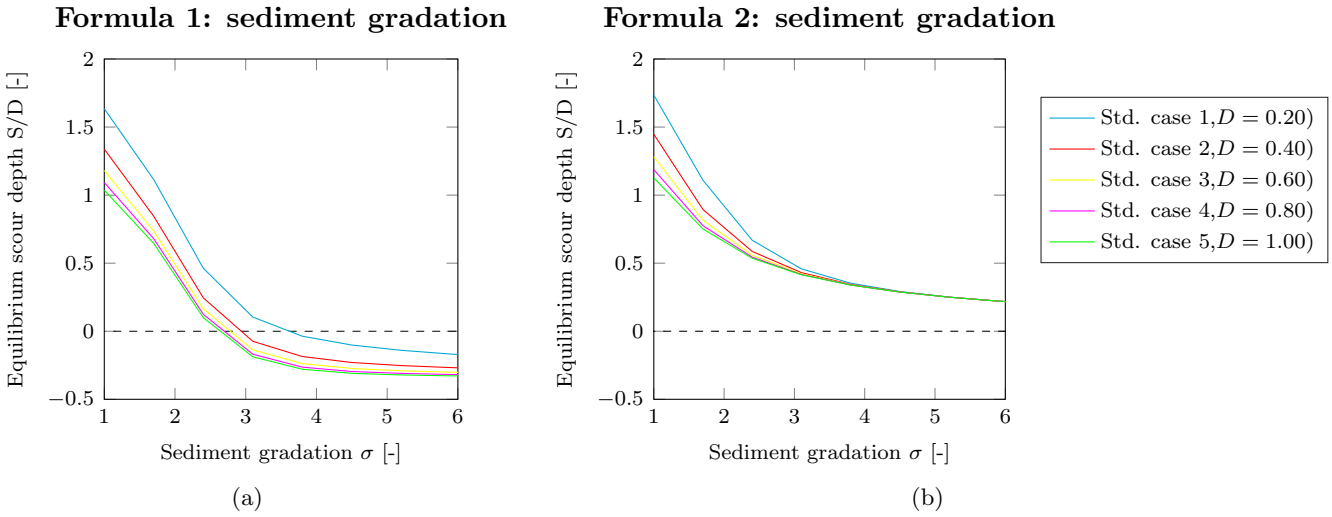


Figure 6.10: *Formula behavior: sediment gradation. C Run 2. Best formula for range of input parameters.*

to look at the difference in formula behavior if σ is ranged from 1 to 6. This can be seen in figure 6.10. The first equation finds negative scour values for $\sigma > 3$. Negative scour does not exist, so this is a large downfall of formula 1. It could be corrected by assuming that all values below zero are equal to zero, but that still would imply that when there is sediment gradation, the armouring effect is so strong that no scour occurs at all. Formula 2 shows more realistic behavior: although the armouring effect becomes higher and the scour depth goes down with increasing σ, the formula approaches an equilibrium value, so there is always some scour. This can be for instance the scour depth that is necessary for the armouring effect to settle.

Optimization technique for Run 4
To conclude, so far formula 2 from the previous section is found to be the best formula for predicting current-only scour. This formula is given by equation (6.3). As can be seen this formula can still use some optimization. It has no constants, nor does it include all parameters that were found to be important in Run 2.

$$SD = \tanh\left(\frac{mob + \tanh(mob)}{\sigma}\right) + \tanh\left(\frac{h_w}{D}\left(\frac{mob}{\exp(mob)\,\sigma} + \tanh(Fr)\right)^{\sigma}\right) \tag{6.3}$$

Run 4: Formula input

The last run uses known combinations as input parameters in the GP. In this case the entire formula 6.3 is subtracted from the measured S/D in the database. The GP is given a very short tree length to find an additional term for the remainder between the actual measured S/D and the calculated S/D from equation (6.3). This process can be repeated endlessly. Each time the new formula (equation (6.3) with the added terms) is substracted from the actual measured value. Table 6.8 shows the results.

At first the added terms contribute significantly to the performance of the equation, but after the second term, the 5 additional terms only add approximately one hundreths

	Eq. 6.3	$-\frac{d^*}{690.0}$	$+\frac{\ln(Fr)}{d^*}$	$+1.046-\tanh(w_s{}^*)$	$-\frac{D}{32.6h_w}$	$+\frac{0.2}{\exp(mob)}$	$-\frac{0.0393}{\tanh(mob)}$	$+0.0069\frac{D}{h_w}$	$+\ldots$
CoD	0.6917	0.7138	0.7282	0.7304	0.7311	0.7348	0.7401	0.7411	$+\ldots$

Table 6.8: *Performance Run 4*

of performance to the coefficient of determination. Interesting to see is that the first three equations contain the second rank rated parameters from Run 2. Apparently these parameters are adapt in fine tuning.

It is chosen to only use the first 2 additional terms. Not only because for these terms the CoD is increasing with the most significant amount, but also because during the testing these terms where found multiple times. Meaning that several different evolutions and initial starting points lead to these terms. Small variations to the first term were found 10 times in the 40 executed experiments. The exact second term was found 12 times in its 40 experiments. This type of consensus stops after the second term. It would seem that from this point on, no clear term exist that can improve equation (6.3).

Therefore the new and improved formula is given by equation 6.3.1.

$$S/D = \tanh\left(\frac{mob + \tanh(mob)}{\sigma}\right) + \tanh\left(\frac{h_w}{D}\left(\frac{mob}{\exp(mob)\,\sigma} + \tanh(Fr)\right)^{\sigma}\right) - \frac{d^*}{690.0} + \frac{\ln(Fr)}{d^*} \tag{6.4}$$

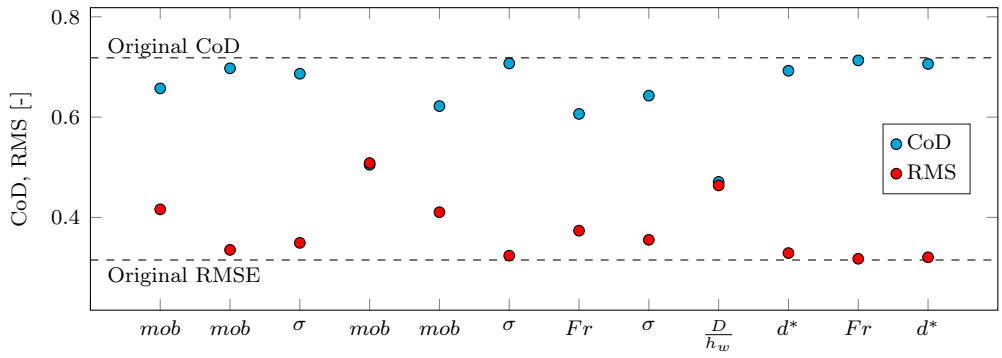
Optimization fit: Adding constants

Equation (6.3.1) is still very clean. It contains only one constant. It would seem too good to be true that the perfect formula contains only integer amounts of parameters. As explained before, although different tactics are conducted, the GP refuses to incorporate a healthy amount of constants. Therefore constants are added at strategical locations and using a fitting program in MATLAB, the optimal value of these constants are found.

Which locations are optimal is chosen with the use of figure 6.11(a). This figure is created similar to figure 6.5a, only now the same type of parameters are not grouped together but evaluated separately. The parameters on the x-axis of figure 6.11(a) are listed in the same order of appearance as in equation (6.3.1). The best equation for current-induced scour including constants is given by equation (6.5). Table 6.9 gives the optimization values of this equation since it was first found in Run 3.

$$S/D = 1.12\tanh\left(\frac{1.35mob + \tanh(mob)}{\sigma}\right) + \tanh\left(\frac{h_w}{D}\left(\frac{1.98mob}{\exp(1.52mob)\,\sigma} + \tanh(Fr)\right)^{1.22\sigma}\right) - \frac{d^*}{400} + \frac{\ln(Fr)}{0.70d^*} \tag{6.5}$$

Looking at the formula and at 6.11(a), it becomes apparent that the second mobility term does not contribute significantly to the behavior of the formula. Because it concerns a tangent inside a tangent this term can be changed to an average value without the



(a)

Figure 6.11: *Behavior of individual components of the best performance formula. [C]*

performance of the GP suffering. In figure 6.12(a) the behavior of the complete first tanh term is shown for tanh(mob) and the effect when tanh(mob) is replaced by a constant value of 1. It can be seen that depending on the sediment gradation, after approximately $mob = 2$ the effect of this term disappears. Although the mathematical implications of this change are hardly noticeable in the CoD and RMSE, as was visible in figure 6.11(a), the behavior of the formula is physically incorrect. As can be seen in figure 6.12(a) with the dashed line, when the mobility is zero, the scour is no longer zero. Therefore it is concluded that the second mobility term is redundant, but can not be replaced by a constant. To get the physical correct formula and get rid of the unnecessary extra term, the term is removed and the constant in front of the first mobility term is recalculated. Using the MATLAB fitting tool this value is found to be 2.20.

The comparison of the original equation and this shorter new equation can be seen in 6.12(b). As can be seen for narrow graded sediment nothing changes, it is a perfect fit. For wider graded sediment and at higher mobilities the difference between the two curves increases. However, the performance of this new equation is better than the original one, so this could also be a good thing. The new formula with newly fitted constants is given by equation (6.6) and its performance can be found in table 6.9 under 'Constant optimization 2'.

$$S/D = 1.12\tanh\left(\frac{2.20\cdot mob}{\sigma}\right)+\tanh\left(\frac{h_w}{D}\left(\frac{1.98\cdot mob}{\exp\left(1.52mob\right)\sigma}+\tanh\left(Fr\right)\right)^{1.23\sigma}\right)-\frac{d^*}{400}+\frac{\ln\left(Fr\right)}{0.685d^*} \tag{6.6}$$

	CoD	RMSE	FpN	NoN	OaF
Formula after Run 3	0.692	0.344	0.222	21	0.523
Formula after Run 4	0.7282	0.3186	0.2098	30	0.540
Formula after Constant optimization 1	0.7479	0.3087	0.215	30	0.553
Formula after Constant optimization 2	0.7479	0.3039	0.215	30	0.553

Table 6.9: *Performance Optimization C run*

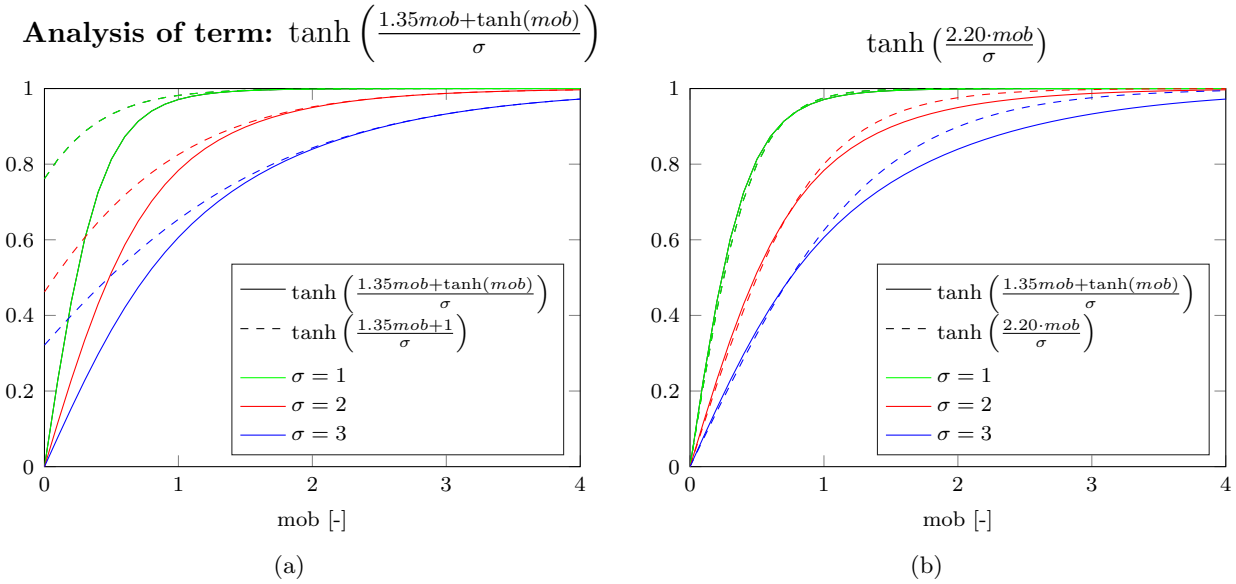


Figure 6.12: *Optimization double hyperbolic tangent term [C] formula*

Formula without sediment gradation

The past few sections have shown repeatedly the significance of the sediment gradation σ. For instance, table 6.3 and 6.5 both showed that the performance of the equations decreased drastically without σ. Unfortunately sediment gradation is not reported for over a quarter of all tests (including [W] and [CW] tests). It seems a waste not to be able to incorporate these tests, therefore Run 3 and 4 are repeated for a formula without σ. Even with the same optimizations the performance remains low and none of the equations show the clear-water peak. The best performing formula that was found even after adding terms with GP and adjusting the constants with a fit program, has a mere CoD of 0.5829 and a RMSE of 0.4103. The formula is given by equation (6.7)

$$S/D = \tanh(\frac{Fr \cdot h_w}{D}) + \tanh(7.42 \cdot mob^{2.28}) + 0.46 \tanh(\frac{1.30}{d^* - 19.0}) - 6.2 \cdot 10^{-4} {w_s}^{*2} \quad (6.7)$$

6.3.2 Wave-only [W]

Now that the formula for current-only conditions is found, the next step is to find a formula for the other extreme condition: wave-only situations. This is done according to the same methodology as used to find the formula for [C], given by figure 6.1.

Run 1: Parameter sensitivity

In the first run all potential parameters from chapter 4 are included to establish the significance of each parameter. No weights are added in this stage. The procedure is the

same as for [C]: 11 pools are created, each omitting one parameter from the terminal set. The last pool contains all parameters.

Mathematical analysis
The results can be seen in table 6.10. This are the average values of the top 10 equations found in each pool. Since the length of the formulas are all well within the limit, it is chosen to calculate the overall fitness value OaF with (1, 1, 0).

Omitted Parameter	CoD	RMS	FpN	NoN	OaF
$-mob$	0.780	0.108	0.242	28.3	0.836
$-KC$	0.771	0.112	0.239	25.2	0.829
$-{w_s}^*$	0.779	0.109	0.250	22.8	0.835
$-\frac{D}{h_w}$	0.782	0.105	0.252	21.4	0.839
$-Fr_w$	0.777	0.109	0.259	19.5	0.834
$-\frac{D}{L}$	0.771	0.112	0.266	17.9	0.830
$-d^*$	0.768	0.108	0.274	16.8	0.830
$-\Delta$	0.771	0.108	0.269	16.7	0.831
$-\frac{h_p}{h_w}$	0.766	0.113	0.272	15.7	0.827
$-Re_w$	0.766	0.109	0.278	14.3	0.829
All incl.	0.778	0.108	0.274	15.9	0.835

Table 6.10: *Performance GP for wave only tests, run 1*

Table 6.10 is not sorted on the value of OaF as the previous tables, but on the number of nodes (NoN). The reason for this is that the performance of all formulas are too close together to make a reliable analysis of parameter sensitivity. An example of this is for the case were d^* is omitted. Looking at the numbers, it seems that the performance of the GP formulas decrease when d^* is not present. However, when actually looking at all the best performing equations of this entire run, d^* is hardly present. Apparently during evolution this parameter is invaluable and can be eliminated from the formula.

By sorting table 6.10 on the NoN, this table does tell something about the importance of each parameter. It shows is that if the terms *mob* and KC are omitted from the process, the GP can still produce a reliable formula, but it needs more than 1.5 times the amount of terms, than the scenario where *mob* and KC are included.

Best equation
Another way of looking at the sensitivity is by replacing the parameters by their mean value, as done for the current only case. As an example this is shown in figure 6.13(a) for the best performing formula of the run. The best formula consists of 3 components: KC, *mob* and $\frac{D}{h_w}$.

When all the KC values are replaced by a constant (in this case by the mean value of KC), the performance of the formula drops from 0.838 to 0.235. This means this parameter is absolutely crucial for the formula. The same can be said for $\frac{D}{h_w}$. Replacing *mob* by a constant however, seems to have little impact. Still it is part of the best equation. This can be explained by looking at figure 6.13(b). This graph shows the impact on the outcome in S/D if the parameter is ranged from its minimum to its maximum. The

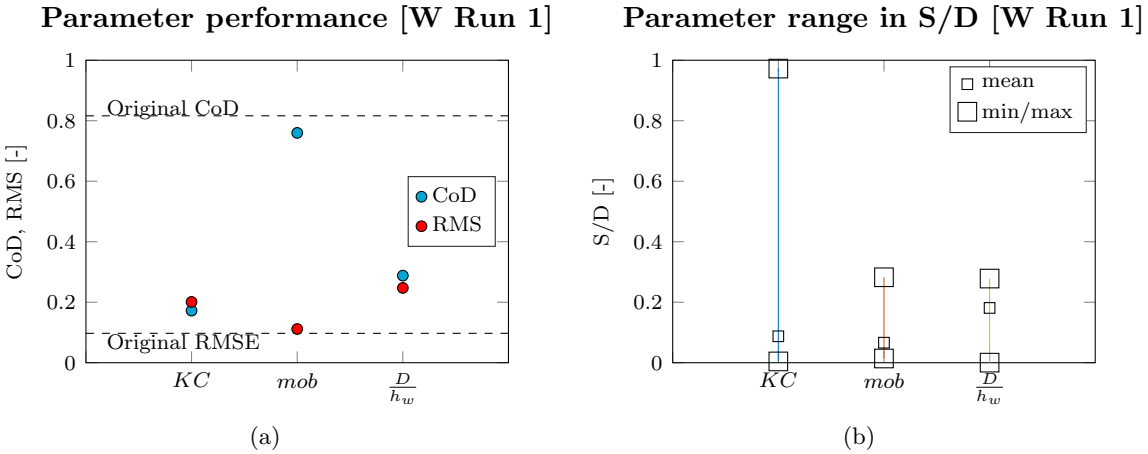


Figure 6.13: *Influence individual parameters formula*

mobility actually causes quite a change in the outcome of $\frac{S}{D}$, similar to $\frac{D}{h_w}$. The reason that $\frac{D}{h_w}$ does influence the CoD and RMSE level and *mob* does not, has to do with the average value. Most *mob* values are small, the mean value square is all the way down at one side in figure 6.13(b). There are simply not enough tests in the higher mobility region to impact the overall CoD and RMSE performance of the formula.

Using this logic and looking at the equations which parameters appear, it can be concluded that the first 3 parameters of table 6.10 (*mob*, KC and $\frac{D}{h_w}$) play an important role. The significance of Fr_w and ${w_s}^*$ is inconclusive for now, which is why they are kept in the parameter set. $\frac{h_p}{h_w}$ is removed from the terminal set for the same reason as in the current-only case: this parameter can be better fitted separately since the effect of reducing the $\frac{h_p}{h_w}$ ratio is known and there are not enough submerged pile test to fit it with the GP. Re_w seems a useful parameter, but not crucial. Therefore it is chosen not to consider this parameter for the wave scour equation, since lab experiments are usually not scaled with the Reynolds number. The influence of all other parameters is minimal and they will not be used in future runs.

Physical analysis
For the 10 best equations the behavior of the formula is analyzed to see if the formula behaves according to known physical processes. No anomalies were found, since each equation reproduced a curve similar to the one shown in figure 2.8. A more in depth physical analysis will be given of the final equation.

Optimization technique for Run 2
The obtained knowledge from this run will be used to optimize the next. The terminal set is downsized to the most important parameters and the parameters for which the parameter sensitivity was inconclusive. Furthermore weights are added to focus on the most relevant and reliable sources. The assignment of the weights is done according to chapter 3.4.

Run 2: Weights

The parameter sensitivity test of Run 1 is repeated for the second run with the reduced terminal set and the addition of weights.

Mathematical analysis
As can be seen in table 6.11, the overall average per run is increased compared to the first run. The reduction of the terminal set was therefore a good optimization. Again it turned out that KC is by far the most important parameter. If this parameter is not included, the GP needs more than double the amount of terms to create a formula that has more or less the same performance.

Omitted Parameter	CoD	RMS	FpN	NoN	OaF
$-KC$	0.781	0.109	0.208	40.2	0.836
$-\frac{D}{h_w}$	0.799	0.106	0.283	16.5	0.846
$-mob$	0.810	0.105	0.287	16.3	0.852
$-{w_s}^*$	0.812	0.103	0.292	15.1	0.855
$-Fr_w$	0.809	0.101	0.295	14.0	0.854
All incl.	0.812	0.106	0.279	17.8	0.853

Table 6.11: *Performance GP for wave only tests, run 2*

Furthermore, $\frac{D}{h_w}$ seems to be a meaningful parameter as well. This can be explained by the influence that the water depth has on wave scour. For instance in deep waters (small $\frac{D}{h_w}$), the waves at the surface will have no influence on sediment at the bottom. Therefore there will be no wave-induced scour.

The other 3 parameters do not seem to contribute, so the terminal set can be reduced to just 2 parameters. Analyzing the mathematical influence and physical behavior of these parameters gives a very similar picture as 6.13a and (b). More interesting is therefore to see how the weights are incorporated and which tests cause problems for the GP.

Errors
The errors are calculated in the same manner as for [C]: the error per test is calculated and averaged for the 50 best formulas. Unlike the [C] test, instead of the minimum and maximum values, only the mean value is given and the weight of the test. Compared to the current test the errors are much smaller. There is only one test with a mean error (over the 50 best equations), larger than 0.5 S/D and that is a silt test from Sumer et al.[45]. Other test that were above a 0.3 S/D error are the breaking waves test from Nielsen and Sumer[46].

The tests with high errors are mostly green (weight 2) and some are red (weight 3). The lower the weight, the less reliable or relevant the test is according to the rating explained in section 3.4. Therefore the weight system is working correctly. It is not a surprise or a problem that the errors of these low rated tests are high, since the reduced weight causes the GP to consider them as less important than other tests.

Optimization technique for Run 3
The performance of the equations found in Run 2 are already quite good and the errors are

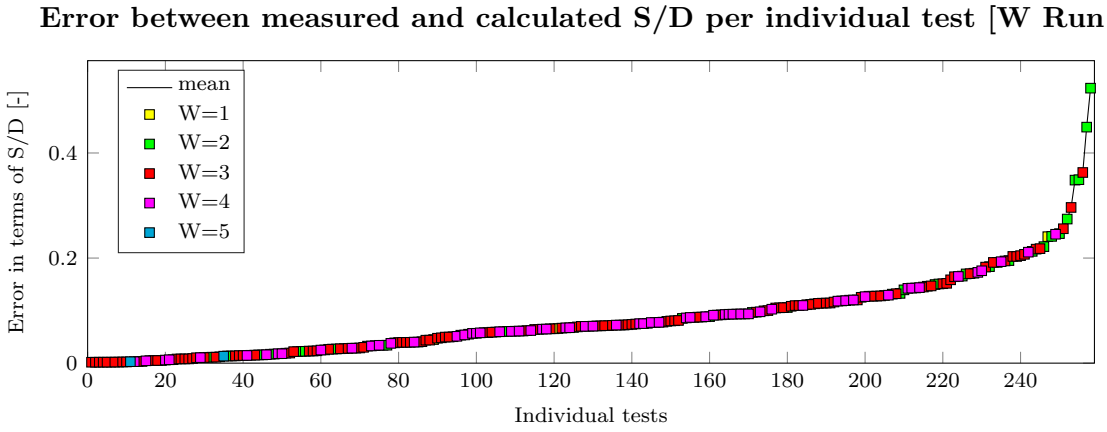


Figure 6.14: *Errors*

low. Therefore not a lot of additional optimization seems possible. However, the terminal set can be reduced to only 2 parameters and of course the runtime can be increased so the GP has more time to evolve to the best possible solution.

Run 3: Additional runtime

The third run is executed with an increased runtime of 10 minutes per experiment. This leads to a list of equations with excellent performance. Since already during Run 1 some high performance equations were found, these are added to the list. The best formula is now sought for in all [W] equations of all runs.

Mathematical analysis
Table 6.12 gives the average performance of the top 10 equations found in Run 1 to 3. Even though many equations have a high performance, one special equation stood out. This equation, or rather small variations to this equation, was found multiple times. In the top 20 best equations of all [W] runs, this equation can be found 6 times, spread over all runs.

What makes this equation stand out, is that it has equal performance to the other equation at the top, but also is surprisingly simple. It has only half the amount of nodes that the other equations have.

	CoD	RMSE	FpN	NoN	OaF
Average top 10	0.828	0.096	0.307	15.1	0.804
Best formula	0.822	0.110	0.357	8	0.805

Table 6.12: *Performance GP for wave only tests, run 3*

The equation is given below:

$$S/D = 1.1301 - \tanh(D/h_w + 14.63/KC) \tag{6.8}$$

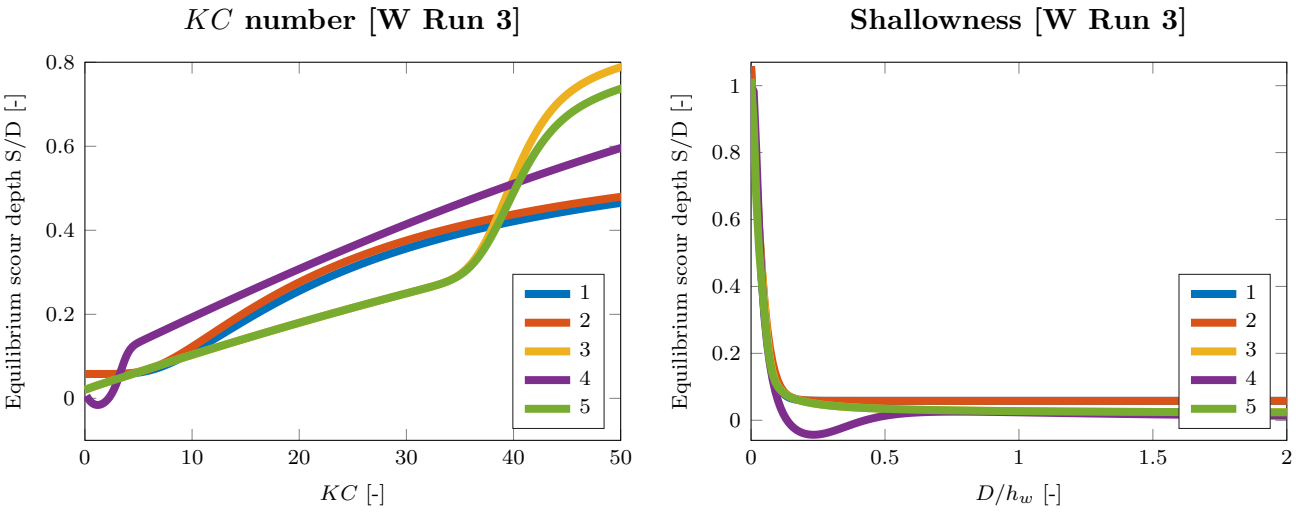


Figure 6.15: *Formula behaviour: KC number and shallowness. [W]. Formula 1 to 5.*

Physical analysis
To check whether this formula is also the best in performance the key parameters KC and $\frac{D}{h_w}$ are plotted against the equilibrium scour depth. The standard test values used for these plots can be found in table 6.13. In case of KC all parameters are kept constant except for the wave height which is varied from 0 to 5 meters. This influences the orbital wave velocity which in turn influences the KC number. To plot the effect of the shallowness on the equation, all values are kept constant at the values of table 6.13 except for the pile diameter D, which is ranged from 0 to 1 meter. The results for the top 5 equations is shown in figure 6.15(a) and (b). Formula 1 is equation (6.8). Formula 2 is a small variation on this equation, with the same functions and parameters but different constants.

Standard test values									
category	d_{50}	D	h_p	h_w	u_c	H_s	T_p	ρ_s	σ_g
[W]	0.0013	0.2	0.5	0.5	0	0.14	1.7	2650	1.2

Table 6.13: *Standard test values [W]*

It can be seen in this figure that the preferred formula from the mathematical analysis above, is also the best formula in the physical analysis. In case of the KC it shows a formula that is steadily growing, but moving towards a certain asymptotic value. The other equations, for instance formula 3 and 5, have an unexplanatory jump in S/D around $KC = 40$. No physical reason is known why the scour depth should experience this suddenly increase at these high KC numbers. Formula number 4 has interesting behavior at low KC numbers, however this formula has negative values and grows to infinite values which is not realistic scour behavior. At some point the scour pit just cannot get any deeper. For instance because the walls of the hole collapse, or the waves cannot reach the sediment.

Figure 6.15(b) shows the influence of the shallowness. For all equations in the top 5, the

curve is more or less equal. Therefore with regard to the physical analysis of finding the best equation, no conclusions can be drawn from this figure, except for the elimination of equation 4. This formula shows again the impossible behavior of negative scour.

The behavior of the curve for formula 1 will be discussed more extensively in section 7.2.1.

Optimization technique for Run 4
It can be concluded that formula 6.8 is mathematically and physically a very good wave-induced scour prediction method. Further optimization could be adjusting the constants in the equation or the addition of an extra term.

Run 4: Formula input

For the last run the same procedure as in the [C] Run 4 is performed. The entire formula is subtracted from the measured S/D value in the database and the goal of the GP is to find a term that can predict this difference and hence make the gap between measured and predicted scour smaller.

The results are clear. The best added terms are terms that are suspiciously close to zero (such as $+1.983 \cdot 10^{-4} \cdot KC$) and by adding the terms the performance increases by only 0.009. It is therefore concluded that this formula is already at its peak.

Optimizing fit: Adding constants

The final step is to see if the formula does benefit from adjusting or adding constants. The following formula was found:

$$S/D = 1.058 - \tanh(D/h_w + 14.1/KC) \tag{6.9}$$

This formula has the same CoD value of 0.822 as equation (6.8), but it reduces the RMSE from 0.110 to 0.099. A small improvement.

This was the last step in optimizing the wave scour equation. The result is a very high performance equation given by equation (6.9). Its comparison to other wave equations and further analysis of this specific equation will be presented in the chapter 7 and 8.

6.3.3 Combined current and waves ([CW] only)

This section combines the knowledge of the two previous sections to find a formula for combined current and wave conditions. The importance of this section is mostly to extract which parameters describe the combined current and wave regime, since the actual formula for combined currents and waves must be a transition between the formulas found for [C] and the formula found for [W]. To establish this transition all formulas must be combined, which will be done in the next section for the database of [C], [W] and [CW] combined.

Run 1: Parameter sensitivity

As described above, Run 1 is the most important test in this section. It performs a parameter sensitivity analysis on all possible parameters from the terminal set in table 4.2. The same method is applied as for [C] and [W], where each pool one of the parameters is omitted. The database for this run only contains tests that have both current and waves. The results of the top 10 average best equations in each pool can be seen in table 6.14.

Omitted Parameter	CoD	RMS	FpN	NoN	OaF
$-u_{rel}$	0.742	0.253	0.236	25.7	0.744
$-KC$	0.808	0.212	0.252	25.4	0.798
$-Fr_w$	0.805	0.206	0.253	23.3	0.799
$-d^*$	0.797	0.208	0.262	21.0	0.794
$-\frac{D}{h_w}$	0.799	0.207	0.270	20.0	0.796
$-mob$	0.794	0.207	0.266	19.9	0.793
$-Fr$	0.796	0.208	0.265	19.0	0.794
$-{w_s}^*$	0.800	0.207	0.268	18.9	0.796
$-\frac{D}{L}$	0.791	0.209	0.271	18.1	0.791
$-\Delta$	0.801	0.210	0.278	16.8	0.796
* All incl. *	0.798	0.210	0.271	18.7	0.794

Table 6.14: *Performance GP for combined wave and current tests, run 1*

Mathematical analysis
As expected u_{rel} is the most important parameter. This can be seen by the decrease in CoD when this parameter is omitted, as well as in the number of nodes that is necessary to produce a good formula. The significance of all other parameters is not that evident since their performance values lie close to each other. Further testing is still required.

When looking at the top 20 of best equations, some parameters do appear more often then others. Some parameters do not appear at all and hence it is concluded that they can be removed from the terminal set. The equations that do not turn up in any of the best equations are d^*, mob, ${w_s}^*$ and Δ.

Run 2: Weights

In this run the parameter sensitivity test of Run 1 is repeated with a smaller terminal set and the addition of weights according to the method described in section 3.4.

Mathematical analysis
In table 6.15 the average result of the best 10 equations per pool can be seen. Similar to Run 1, the performance of each pool is quite similar. However, a significantly better result is obtained when all parameters are included. The complete set is therefore tested with additional runtime in Run 3.

Omitted Parameter	CoD	RMS	FpN	NoN	OaF
$-Fr_w$	0.764	0.274	0.291	12.0	0.745
$-u_{rel}$	0.761	0.264	0.277	15.0	0.749
$-KC$	0.770	0.222	0.268	15.0	0.774
$-\frac{D}{L}$	0.780	0.224	0.300	12.3	0.778
$-\frac{D}{h_w}$	0.793	0.210	0.266	17.0	0.791
$-Fr$	0.803	0.203	0.262	18.0	0.800
* All incl. *	0.812	0.223	0.276	15.7	0.794

Table 6.15: *Performance GP for combined wave and current tests, run 2*

Run 3: Additional runtime

With the terminal set found in Run 2 and an increased experiment duration of 10 min, the last GP run of [CW] is executed.

Mathematical analysis
Considering only CoD and RMSE values, the overall best formula of this run is given by equation (6.10). This formula has an excellent overall fitness value with $CoD = 0.838$ and $RMSE = 0.13$. The absence of the KC number is remarkable, since it seemed to be one of the more influential parameters in table 6.14.

$$S/D = \left(\sqrt{\exp(U_{rel})} + \left(\frac{D}{h_w}\right)^{Fr}\right) \cdot \left(Fr^{\sqrt{\frac{D}{L}} \cdot Fr \cdot \sqrt{\frac{D}{h_w}}}\right) \tag{6.10}$$

Physical analysis

If the equations are adequate, the value at $u_{rel} = 0$ should be equal to the wave only equation. If the value at $u_{rel} = 1$ the value should be equal to the current only equation. Although all top ranking formula fit the data from the [CW] database very well, none of the equations is able to make a smooth transition. However, for this run, this would be a lot to expect from the GP to model, after all it has no knowledge of the [C] and [W] database. Therefore, the only way to get a correct formula for [CW] is by combining all three databases. This will be done in the next section.

Standard test values									
category	d_{50}	D	h_p	h_w	u_c	H_s	T_p	ρ_s	σ_g
[CW]	0.0013	0.2	0.5	0.5	0.3	0.14	1.7	2650	1.2

Table 6.16: *Standard test values [CW]*

6.3.4 Combined current and waves [C, W, CW]

The last tests that will be done with GP are to find a scour prediction formula that is valid for all conditions. This equation will be based on the knowledge obtained from the

best formulas for [C] and [W], and the important parameters found for [CW]. 5 different methods are used, given by table 6.17, that finally lead to one best overall equation.

Nr.	Target GP	Input parameters	Database		
			[C]	[W]	[CW]
1.	S/D	mob, u_{rel}, KC, D/h_w, Fr, Fr_w, d^*, $w_s{}^*$, $\frac{D}{L}$	All	All	All
2.	$S/D - u_{rel} \cdot f_{[C]\not{\phi}} - (1-u_{rel}) f_{[W]}$	u_{rel}, KC, D/h_w, Fr, Fr_w, $\frac{D}{L}$			
3.	$S/D - u_{rel} \cdot f_{[C]} - (1-u_{rel}) f_{[W]}$	u_{rel}, KC, D/h_w, Fr, Fr_w, $\frac{D}{L}$	Data with σ	Data with σ	Data with σ
4.	$S/D - u_{rel} \cdot f_{[C]} - (1-u_{rel}) f_{[W]}$	u_{rel}, KC, D/h_w, Fr, Fr_w, $\frac{D}{L}$	Data with σ	All	All
5.	$S/D = \frac{f_{[C]} - f_{[W]}}{1-e^{-C1(u_{rel}+C2)}} + f_{[W]}$	u_{rel}, KC, D/h_w, Fr, Fr_w, $\frac{D}{L}$			

Table 6.17: *Methods to find a scour prediction formula for all conditions*

Method 1

The first method is similar to the technique used for finding the formulas for [C], [W] and [CW]. The parameters from the reduced terminal set from each of these categories are taken as the input parameters. The exception to this is the sediment gradation σ. Unfortunately there are not enough σ values available for the [W] and [CW] section to form a full database and the GP cannot cope with empty parameters. In this method, this means that σ will be excluded from the terminal set. In methods 3 till 5 the problem will be further addressed and different solutions that do include σ as input parameter will be given.

The challenge in this section is that the GP has to find an accurate scour formula to predict the entire range of data sets. This causes a new problem, which is the imbalance of data. As discussed in chapter 3 there are more than 10 times as much [C] tests than [W] and [CW] tests. This means that if they are all combined in a GP test, the [C] tests completely overshadow all other tests. Tests like this have been done, and the effect can immediately be seen in the formulas that are created. For instance, none of these formulas contain a KC term, which is an absolute indispensable parameter for any wave formula.

To compensate for this asymmetry in the database, an equal amount of tests is taken from each condition. The GP is trained with random different combinations to ensure that all tests have an opportunity to influence the scour formula. The probability that a test is chosen for the database depends on the weight given by section 3.4.

The best equation of the entire run is found by calculating the CoD, RMSE and OaF for each formula for each complete database of [C], [W] and [CW] separately. The overall fitness from each category, respectively OaF_c, OaF_w and OaF_{cw} are averaged and ranked according to their value. The best formula obtained with this method is given by

equation (6.11) and has the meager performance given by table 6.18.

$$S/D = (0.708 + 0.76u_{rel}) \cdot \left(\tanh \left[\left(0.98 + \left(\frac{1}{KC} + Fr \right) \left(\frac{1}{KC} + 2\frac{D}{h_w} \right) \right) Fr \right] \right)^{\frac{1}{KC} + \frac{D}{h_w}} \tag{6.11}$$

Average OaF	[C]		[W]		[CW]		NoN
	CoD	RMSE	CoD	RMSE	CoD	RMSE	
0.652	0.408	0.455	0.582	0.181	0.771	0.211	27

Table 6.18: *Performance [ALL] formula, method 1*

The relationship between the predicted formula and the actual scour is not that strong without σ. In an optimized mode such as equation (6.7) the maximum CoD is 0.58, so it is not strange that this combined formula that also has to focus on other phenomena has only a CoD of 0.40. It is disappointing to see the performance of the [W] data. The GP should be able to predict the wave scour depth more accurately.

Method 2

In method 1, no knowledge of the previous sections was used. The GP had to start over from scratch and the effect was that only the [CW] data was predicted adequately. In the second method it is therefore assumed that the general scour equation will have the shape given by formula (6.12) below. $f_{[C]}$ and $f_{[W]}$ are respectively the best current and wave equation given by equation (6.6) and equation (6.9). However, since $f_{[C]}$ contains σ which is not known for all parameters, $f_{[C]\not\sigma}$ given by equation (6.7) will be used instead.

$$S/D = u_{rel} \cdot f_{[C]} + (1 - u_{rel}) \cdot f_{[W]} + *unknown* \tag{6.12}$$

The functions $f_{[C]}$ and $f_{[W]}$ are multiplied by u_{rel} to help the GP switch between the [C] and [W] conditions. When $u_{rel} = 1$ (current-only conditions), the $f_{[W]}$ term will disappear since it is multiplied by zero. When $u_{rel} = 0$ (wave-only conditions), $f_{[C]}$ will be the term that dissolves. The target parameter for the GP is to find the unknown term that connects the two extremes, in other words a term that represents the [CW] conditions when $0 < u_{rel} < 1$ and on top of that also approaches zero when u_{rel} goes to 0 or to 1.

Using the same ranking system as described in method 1, the best equation in method 2 is:

$$S/D = u_{rel} \cdot f_{[C]\not\sigma} + (1 - u_{rel}) \cdot f_{[W]} + 0.4 \left(\sqrt{\frac{D}{L}} + \tanh(u_{rel}) - 0.339 \right) \tag{6.13}$$

This formula is already much better than the formula obtained from method 1, which is clear when table 6.18 and 6.19 are compared. [C] is now at the top of its ability without σ. It even marginally improved the CoD and RMSE of equation (6.7). The performance of [W] has improved as well to a more than acceptable level.

Average OaF	[C]		[W]		[CW]		NoN term
	CoD	RMSE	CoD	RMSE	CoD	RMSE	
0.726	0.584	0.379	0.797	0.113	0.711	0.243	9

Table 6.19: *Performance [ALL] formula, method 2*

Method 3

The best way to improve the results from method 2, is if the formula for [C] including σ can be used. This is done by applying the same method as described in method 2, using the same target parameter for the GP, described by equation (6.12) and table 6.17. Only now the database is shortened to exclusively data that have a registered σ value. Again an equal amount of each [C], [W] and [CW] condition must be chosen to train the GP, so this database is as large as the smallest number of these 3. This means the database only contains 150 samples, of which all [CW] have $\sigma = 1.28$.

Nevertheless, this run is executed and the best equation is given by equation (6.14). Its performance over all tests that have reported a σ value is shown in the first row of table 6.20.

$$S/D = u_{rel} \cdot f_{[C]} + (1 - u_{rel}) \cdot f_{[W]} + \left(\tanh \left[\left(\tanh \left(\frac{u_{rel}}{{w_s}^*} \right) \right)^{1/2{w_s}^{*2}} \right] \right)^{{w_s}^*} \tag{6.14}$$

Database	Average OaF	[C]		[W]		[CW]		NoN term
		CoD	RMSE	CoD	RMSE	CoD	RMSE	
Data with σ	0.761	0.739	0.317	0.849	0.099	0.639	0.244	13
All data	0.700	0.739	0.317	0.822	0.099	0.465	0.406	13

Table 6.20: *Performance [ALL] formula, method 3*

Large improvements have been made. As expected especially the performance of [C] has increased drastically. However, because this data is just a small cut of the database, it cannot really be sure that all data is represented right. To check the solidity of the equation it is assumed for all [W] and [CW] tests with unknown sediment gradation that $\sigma = 1.3$. The performance values are recalculated with the complete database and shown in the second row of values in table 6.20. Only the [CW] performance is significantly lower. This is interesting since the [CW] database that trained the GP contains only one value of $\sigma = 1.28$, which is practically equal to the estimated $\sigma = 1.3$. Apparently the [CW] data with σ is not a good representation of the other [CW] data sets. [W] is well represented by the selection of data and experiences only a small loss in performance. [C] of course does not change, since no changes have been made to that part of the database.

Method 4

Using the formula $f_{[C]}$ with σ given by equation (6.6) showed a large improvement for predicting scour in [C] situations. However, since the test database was too small for [CW] to be accurately estimated, the performance value suffered. To solve this method 4 uses the same formula set-up as method 2 and 3, but now with a different database. A similar technique that was used in method 3 to check the solidity of the formula will now be used to train the GP. As will be explained here:

In this chapter it has been repeatedly concluded that σ is an important parameter for the current-only conditions. For [W] and [CW] this relation has not been found. This can either be because armouring has little effect during wave-induced scour, or just because there was insufficient data available for [W] and [CW] to test the relation between scour and sediment gradation.

In any case, both [W] and [CW] are able to find good prediction formulas without σ. Looking at the layout for an overall scour equation given by equation (6.12), it is not a problem for [W] predictions if $f_{[C]}$ contains a σ value, since the term containing σ will be zero due to $u_{rel} = 0$. All [W] tests can therefore have a random value for σ and the predictions would still be correct.

For [CW] this is not the case. However, after the parameter sensitivity of section 6.3.3 it can be concluded that σ is not essential for a good [CW] prediction formula. It is therefore assumed that in case of [CW] tests, σ has the constant value of $\sigma = 1.3$. This is a standard value used in sediment studies, assuming that if it is not specifically reported that graded sediment is used, probably the sand is uniformly graded.

To conclude, for [CW] a value for σ is assumed and hence all available [CW] data will be used in the training database. For [W] σ values are unnecessary, so all available [W] data can be used as well. For [C] the variance in σ is too influential to assume a value, therefore only the data of [C] with σ will be used.

Using these settings for the GP, the following formula and performance values were found:

$$S/D = u_{rel} \cdot f_{[C]} + (1 - u_{rel}) \cdot f_{[W]} + {Fr_w}^2 \cdot \sqrt{\frac{\tanh\left(Fr \cdot {w_s}^* \cdot KC^1.09\right)}{\exp\left({w_s}^*\right)}} \tag{6.15}$$

Average OaF	[C]		[W]		[CW]		NoN
	CoD	RMSE	CoD	RMSE	CoD	RMSE	
0.782	0.7479	0.3039	0.822	0.099	0.750	0.217	12

Table 6.21: *Performance [ALL] formula, method 4*

Method 5

The last method is the most simple one. Using the knowledge that if $u_{rel} = 0 \rightarrow S/D \approx f_{[W]}$ and if $u_{rel} = 1 \rightarrow S/D \approx f_{[C]}$, a mathematical operator called the Sigmoid function can be used to form the transition between the two. The standard Sigmoid function has

an 'S' like shape as can be seen in figure 6.16 (a). By adjusting the parameters c_1, c_2 and c_3 the curve can be fitted to aforementioned requirements. An example of this shape is shown in 6.16 (b). The shape of the scour prediction formula including the Sigmoid function is given by equation (6.16).

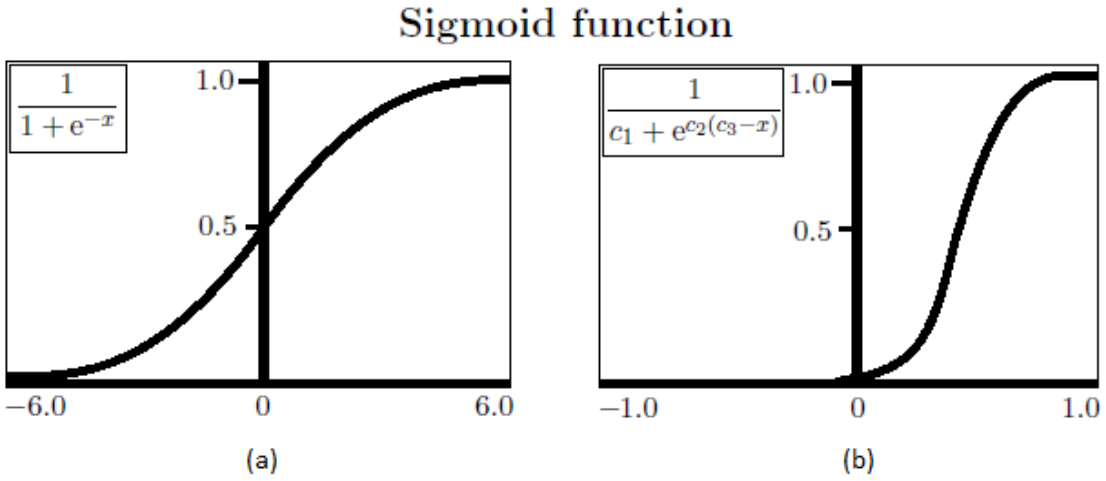


Figure 6.16: *Sigmoid function example*

$$S/D = \left(1 - \frac{1}{1 - e^{-C1(u_{rel}+C2)}}\right) \cdot f_{[W]} + \frac{1}{1 - e^{-C1(u_{rel}+C2)}} \cdot f_{[C]} = f_{[W]} + \frac{f_{[C]} - f_{[W]}}{1 - e^{-C1(u_{rel}+C2)}} \tag{6.16}$$

Using a program in MATLAB that optimizes constants, the values are found at, $c_1 = 0.996$, $c_2 = 8.5535$ and $c_3 = 0.3516$. Inserting these constants leads to equation (6.17) and the performance values as presented in table 6.22.

$$S/D = f_{[W]} + \frac{f_{[C]} - f_{[W]}}{0.996 - e^{8.5535(0.3516-u_{rel})}} \tag{6.17}$$

As can be seen this equation does indeed creates a perfect transition between [C] and [W]. Their values are at the top of what their individual value would be. The Sigmoid function is however not that adapt in modeling the [CW] data in between.

Average OaF	[C]		[W]		[CW]		NoN
	CoD	RMSE	CoD	RMSE	CoD	RMSE	
0.693	0.746	0.308	0.822	0.105	0.393	0.390	-

Table 6.22: *Performance [ALL] formula, method 5*

6.3.5 Final equation

Already during the discussion of the 5 methods it was apparent that some methods were better than others. When regarding the performance values alone, it is clear that the

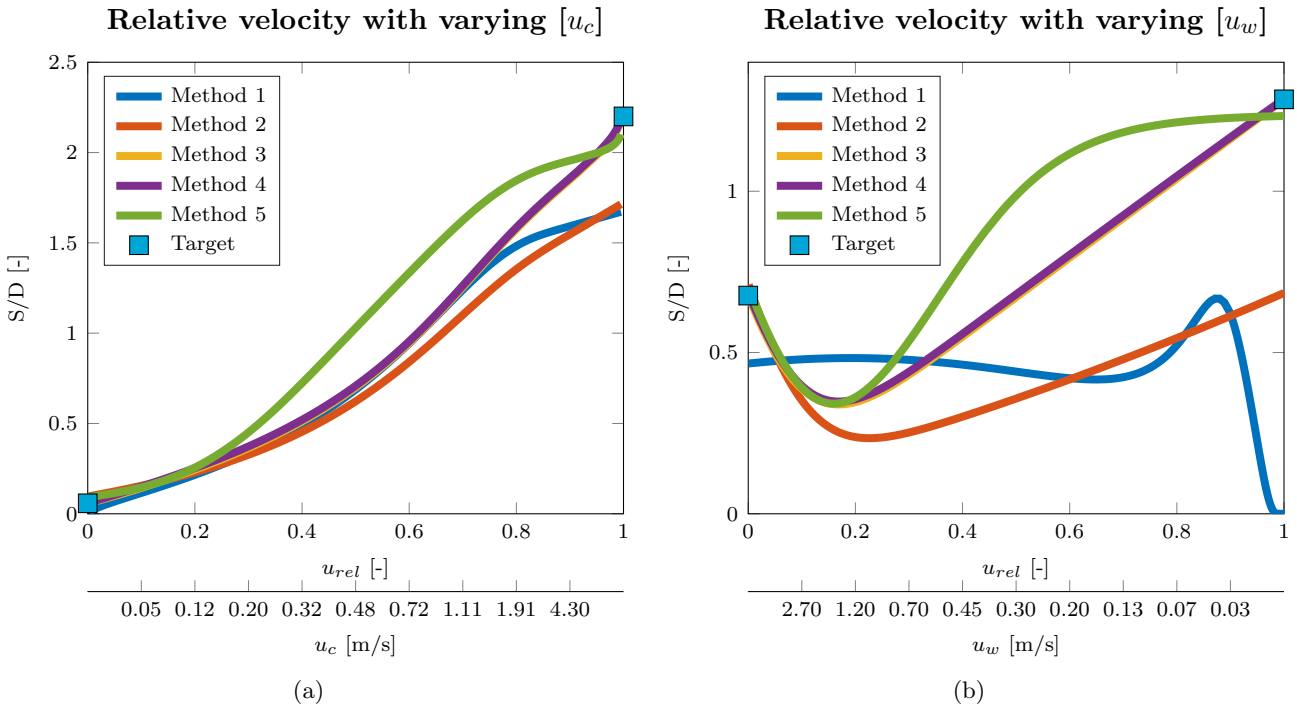


Figure 6.17: *Formula behavior: relative velocity. [CW] Formula 1 to 5.*

method 4 produced the best overall scour prediction formula. However, before discarding all other formulas, also the physical part need to be analyzed.

The best formula from each method is tested for u_{rel} in figure 6.17(a) and (b). In figure 6.17(a) the u_{rel} is plotted by increasing the current velocity u_c, while keeping u_w at 0.48 m/s. In figure 6.17(b) u_{rel} is increased by the orbital wave velocity u_w, while keeping u_c at 0.3 m/s. All other parameters are kept constant at the values given in 6.16.

The blue boxes on the boarders of the figure indicate the target values. For the figure on the left, the curve needs to go from scour in wave-only conditions at $u_{rel} = 0$, to the value for current-only scour $u_{rel} = 1$ (i.e. $u_c \to \infty$). The values are calculated with the equations for [W] and [C] and are respectively 0.058 S/D and 2.2 S/D. The curves in figure 6.17(b) need to go from 0.6765 S/D (the value at infinity for u_w) to 1.285 S/D, (the value for current-only scour with $u_c = 0.3$). Method 1 and 2 estimate the current scour with a different formula than method 3 to 5, which is why these curves move to a different point on the boundary.

Figure 6.17(a) shows a pretty straightforward phenomenon. When the current increases, the scour increases. Except for the difference in target values, the shape of curve 1 to 4 is quite similar. Method 3 and 4 are even so alike that curve 4 blocks the sighting of curve 3. The Sigmoid formula from method 5 makes a different curve than all the others, with higher values and a sudden increase at the top. This divergence from the rest and the complicated shape for a simple effect, suggest that this is probably not the most accurate solution.

In figure 6.17(b) the curves of all methods are more spread out. However, all methods

except the formula from method 1, attempt to model a curvature around $u_{rel} = 0.2$. When looking at this figure from right to left (in the direction of increasing u_w), it can be seen that this dip is in accordance with the theory discussed in chapter 2, which states that the addition of waves to a current reduces scour. Up until $u_w = 1.2$ each increase in orbital velocity causes a decrease in scour depth. This effect can be explained by the waves destroying the horseshoe vortex, the main contributor of current-induced scour. When the orbital wave velocity is larger than 1.2 m/s the waves itself start to have a negative effect on scour.

The curve for method 3 and 4 in figure 6.17(b) are again almost identical. This occurs because both the added terms of method 3 and 4 are on average very small and only for a very specific set of input parameters does this value increase to a noticeable effect.

The curve of method 2 is similar to method 3 and 4, except for the different target parameter on the boundary. It seems that method 5 attempted to have the same shape as method 3 and 4 as well, but the 'S' like structure of the function causes the predictions to be to high for the larger part of the range.

Curve 1 from method 1 shows completely different behavior than the others. It demonstrates the disadvantage of using $1/KC$. The practical application of $1/KC$ is that currents can be considered as very long waves with infinity high KC numbers. The longer the 'wave' the smaller this value gets. So for a true current with infinite length, this value will approach zero. The disadvantage is the other extreme, which occurs when KC becomes smaller. If $KC \to 0$ than the value of $1/KC$ goes to infinite. In case of formula 1, $1/KC$ is applied as a power to a term smaller than 1, which explains why the scour in figure 6.17(b) goes to zero. At zero itself formula 1 predicts 0.93 S/D, which is already a better prediction.

Looking at both figures and the mathematical analysis described for each method in the previous section, it is quite obvious that the formula from method number 4 is the best scour prediction equation. This equation will be used and discussed in the next chapters.

Chapter 7

GP Results

In the previous chapter three formulas have been found with the aide of a genetic program that predict the equilibrium scour depth. The first formula predicts current-induced scour [C], the second formula wave-induced scour [W] and the third formula predicts the equilibrium scour depth in combinations of both wave and current [CW], as well as in current-only conditions and wave-only conditions.

In this chapter the behavior of each formula will be analyzed to find its strengths and limitations.

7.1 [C] formula

First up is the equation for current only scour. To recap, the best current-only scour prediction formula of this study is given by equation (6.6):

$$S/D = 1.12 \tanh\left(\frac{2.20 \cdot mob}{\sigma}\right) + \tanh\left(\frac{h_w}{D}\left(\frac{1.98 \cdot mob}{\exp\left(1.52 mob\right)\sigma} + \tanh\left(Fr\right)\right)^{1.23\sigma}\right) - \frac{d^*}{400} + \frac{\ln\left(Fr\right)}{0.685 d^*} \tag{6.6 revisited}$$

The parameters that are necessary to compute the scour in current-only situations are listed in table 7.1. In table 7.1(a) the 6 basic parameters can be found that are necessary to compute the 5 parameters of the [C] formula. In appendix A an overview is given how to calculate the parameters in table 7.1(b) from the parameters in table 7.1(a). It is essential that the same calculation methods are used as were applied when creating this formula in GP. For instance, a different empirical relation for θ_{cr} can lead to a decrease in accuracy of the prediction.

Apart from just listing the parameters, table 7.1 also gives the range of the database that was used to formulate equation (6.6). This is in order to give an impression of the conditions that are covered by the dataset and hence the range of applicability of

formula. Actual field data will generally be outside of the range of the given dimensional parameters, which is why the formula is based on dimensionless input variables. To obtain an accurate result of a S/D calculation, it is important that these non-dimensional parameters are in the same order of magnitude as the range given by table 7.1(b).

Required input parameters			
Parameter	Range		
	min	max	mean
d_{50} [mm]	0.17	16.9	1.6
D [m]	0.015	0.92	0.11
h_w [m]	0.019	1.9	0.21
u_c [m/s]	0.11	2.5	0.58
σ [-]	1	7.8	1.79
ρ_s [kg/m^3]	1180	2690	2601

(a)

Formula parameters			
Parameter	Range		
	min	max	mean
mob [-]	0.036	479	6.38
$\frac{D}{h_w}$ [-]	0.038	19.2	0.79
Fr [-]	0.06	1.69	0.456
d^* [-]	4.3	428	39.4
σ [-]	1	7.8	1.79

(b)

Table 7.1: *List of required parameters for [C] formula and the range and mean values of the database on which the formula is based*

In figure 7.1 the performance of the [C] scour prediction formula can be seen for the current-only database described in chapter 3. This figure plots the measured scour depth (target parameter) against the calculated scour depth with equation (6.6). If the calculated value is higher than the actual scour depth, the test is considered to be in the 'safe region' since it is a conservative result. If the calculated value is below the measured scour depth, it is considered as an unsafe prediction. The dashed line represents the equal axis. If a test is on this axis, the prediction is a 100% accurate.

There are to many tests in the database to visualize the 5 weight classes from section 3.4 individually, therefore in this figure weights are grouped together in reliable ($W > 3$) and less reliable sources ($W \leq 3$). For the current test data that reported σ (and hence can be used for equation (6.6)), about 850 tests are in the first category and about 750 in the other.

The goal of adding weights was to improve the GP by focusing the training on reliable tests. It can be seen that the more reliable sources (blue dots) in figure 7.1 are indeed better predicted and have a smaller spread around the equal axis than the less reliable red dots. This is due to their rating as 'reliable' in twofold. Firstly because reliable sources are more accurate and hence easier to predict. The less reliable sources may contain a measurement error of some sort, unrelated to the scour phenomena and hence unpredictable for the GP. For instance, it is very unlikely that the two outliers around the impossibly high target value of 3 S/D are measured correctly. For these tests to be on the equal axis, the GP would not have to calculate the actual S/D, but the size of their error correctly. The other cause for the high ranked test to perform better in figure 7.1 is of course because they are rated as reliable. There weight is increased in the dataset and the hence the GP is forced to evolve towards a formula that estimates these values more

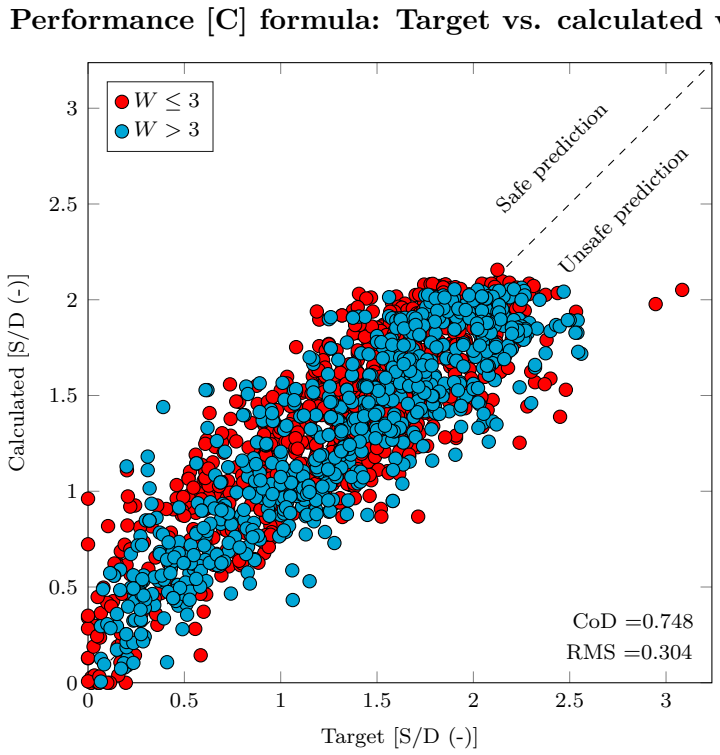


Figure 7.1: *Predicted values versus the actual values using the [C] formula, for current-only conditions*

correctly.

7.1.1 Behavior formula key parameters

Figures like 7.1 and the RMSE and CoD are usefull mathematical tools to evaluate a function. However, more importantly a function should be able to reflect the actual behavior of scour phenomena. In the previous chapter (section 6.3.1), already some plots were made to establish the formula that presented the best physical behavior. In this section some additional behavior of the key parameters for the [C] formula will be discussed.

In figure 7.2(a) the mobility is plotted versus the equilibrium scour depth. All values are kept constant, except for the current velocity. It can be seen that this plot has the same shape as u_{crit} in figure 6.9(a). The shape of the Fr number with respect to the equilibrium scour depth is not shown, but it is similarly shaped as well.

In accordance with the theory described in chapter 2 and the relation found between σ, mob and S/D found in section 3.3.1 the formula shows a peak at the clear-water to live-bed transition due to backfilling that becomes less evident with increasing sediment gradation.

Interesting to see is that the maximum equilibrium does not go to 1.3 S/D as suggested by Sumer et al. [14] and many other studies, but to a value of approximately 2 S/D. The

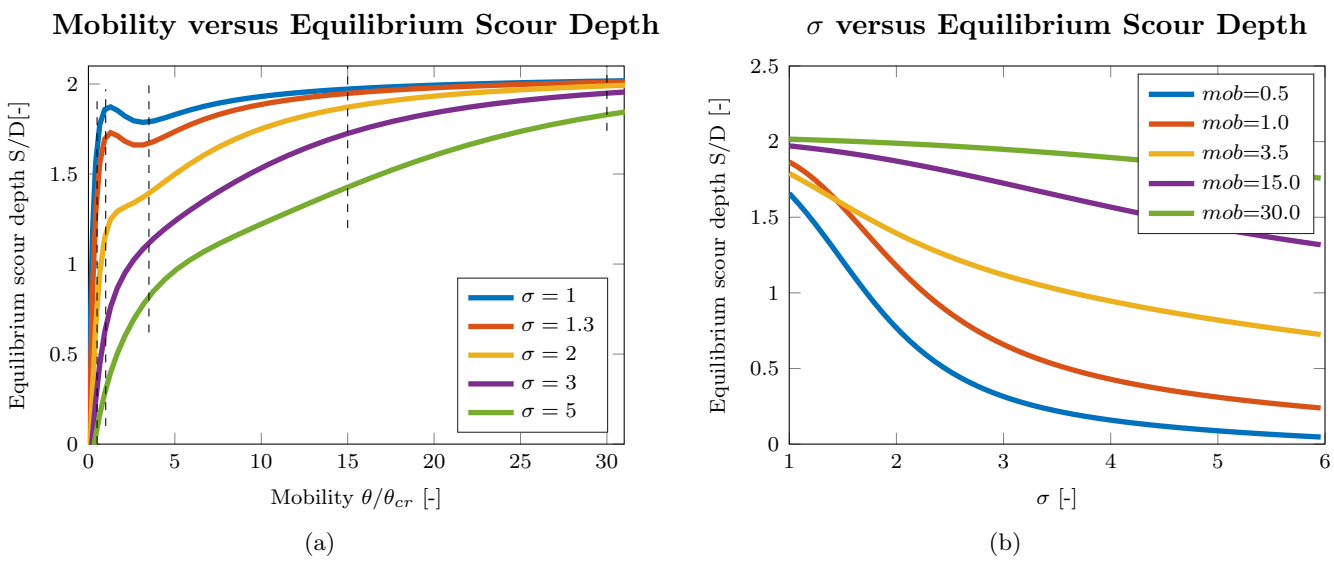


Figure 7.2: *Formula behavior [C]: scour depth prediction with increasing mobility. [d_{50}=0.0013 m, D=0.2 m, h_w=0.5 m]*

maximum value that the formula can predicts for the settings in the database is 2.2 S/D. This is a good estimate in accordance with the reported values in the database, as can be seen in 7.1, where approximately 97% of the tests is measured below 2.2 S/D.

Figure 7.2(b) shows the reversed relationship of figure 7.2(a). The effect of sediment gradation is shown for 5 values of *mob*, indicated by the dashed lines in figure 7.2(a). It can be seen that for low mobility (for example at low current velocities) the influence of sediment gradation is significant. By increasing the sediment gradation from 1 to 2, a decrease of $\approx 50\%$ can be seen in S/D. For a high mobility the entire bed is active and the armouring effect seems to disappear. This can be seen in figure 7.2(a) by the convergence of S/D to the maximum scour depth and in figure 7.2(b) by the approximately constant value of S/D independent of the size of σ.

Figure 7.3(a) is made by ranging the water depth h_w, while keeping all other values constant. It shows that when D/h_w is larger than 1, or in other words when the depth of the water is smaller than the width of the pile, the scour depth evolves towards a constant value. This is because in this scenario the size of the vortex is limited by the water depth. It seems that for large h_w, i.e. small values of D/h_w the size of the vortex at the bed is also limited, perhaps by the pile diameter.

The last figure of this analysis considers the influence of the sediment size. In figure 7.3(b) it can be seen that the equilibrium scour depth decreases when larger sediment is used. For large rocks ($d^* \to \infty$) the equilibrium scour depth will be zero, as to be expected. For graded sediments this will occur at smaller mean grain sizes than for uniform sediment. For small d^* values S/D also reduces. This suggests that this formula is even able to model the effect of cohesion of small sediments, although only non-cohesive tests were used as input data. Perhaps these non-cohesive tests, still experienced some sort of cohesion that the GP picked up.

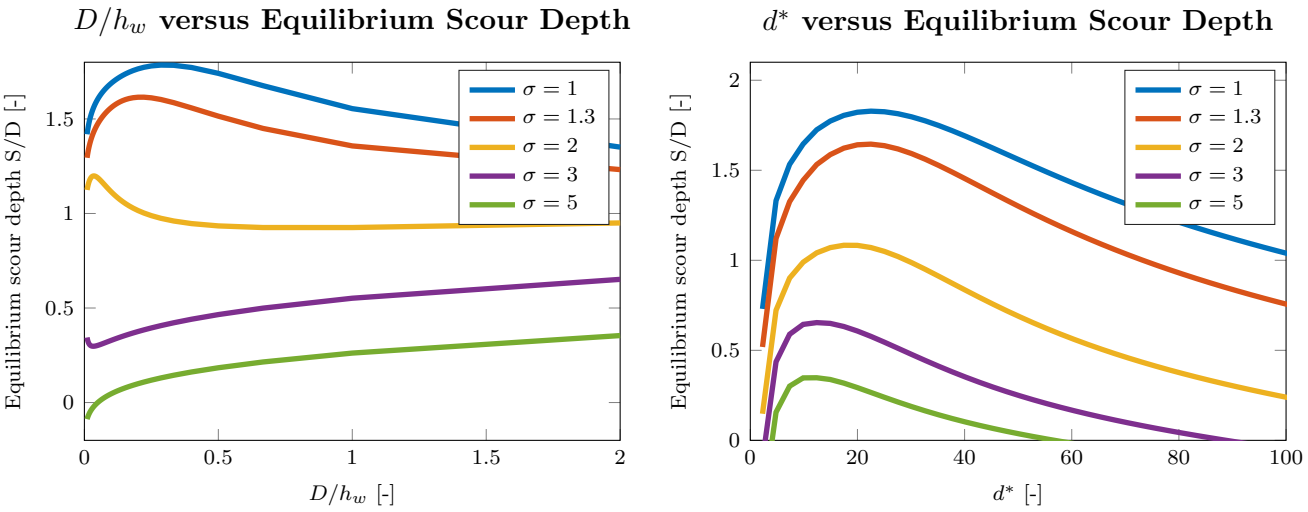


Figure 7.3: *Formula behavior [C]: scour depth prediction with increasing water depth (a) and sediment size (b). [d_{50}=0.0013 m, D=0.2 m, h_w=0.5 m, u_c=0.4 m/s]*

7.1.2 Formula parts analysis

A part of this investigation is to find new ways to express scour problems. An important part is therefore to analyze how the formulas are structured. The formula is split up in three parts as shown in equation (7.1). The behavior of the first part of the equation was already analyzed in figure 6.12b. It basically provides a term that increases towards a constant value of 1.12. For uniform sediments this will occur at $mob \approx 1.3$ (the function $\tanh(x) = 1$ for x$\geq$ 3) and at a higher sediment mobility for graded sediments.

$$S/D = \underbrace{1.12 \tanh\left(\frac{2.20 \cdot mob}{\sigma}\right)}_{\text{Part 1}} + \overbrace{\tanh\left(\frac{h_w}{D}\left(\frac{1.98 \cdot mob}{\exp\left(1.52 mob\right)\sigma} + \tanh\left(Fr\right)\right)^{1.23\sigma}\right)}^{\text{Part 2}} \underbrace{- \frac{d^*}{400} + \frac{\ln\left(Fr\right)}{0.685 d^*}}_{\text{Part 3}} \tag{7.1}$$

Part 2
Part 2 is the most interesting term of the equation. It provides the characteristic bump at the clear-water and live-bed transition. In figure 7.4 the structure of each component of this equation can be seen. Already the first combination of terms, visible in 7.4(a) creates the characteristic small peak with the maximum value depend on the sediment gradation. At this point the peak is not exactly at 1, but with the additions of part 1 and part 3, this peak will shift to its rightful place.

Continuing with the Froude number, which is added in 7.4(b), increases the entire shape by a constant amount, which is than reduced for graded sediments in 7.4(c) to account for the effect of armouring. In 7.4(d) it can be seen how the effect of the different ratios of D/h_w is incorporated in the formula. If $D/h_w < 1$ the shape of the curve is stretched,

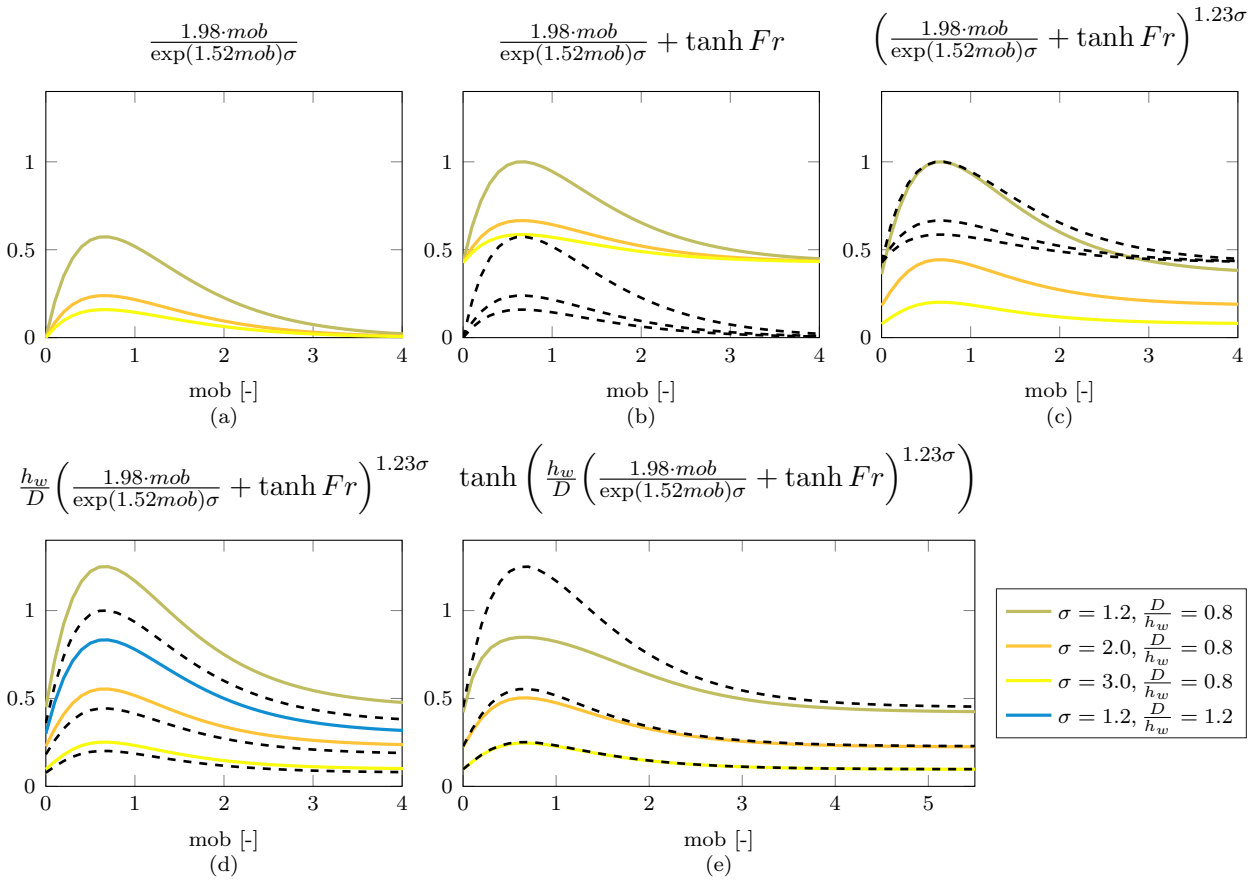


Figure 7.4: *Structure part 2 of [C] formula, [Fr=0.4]*

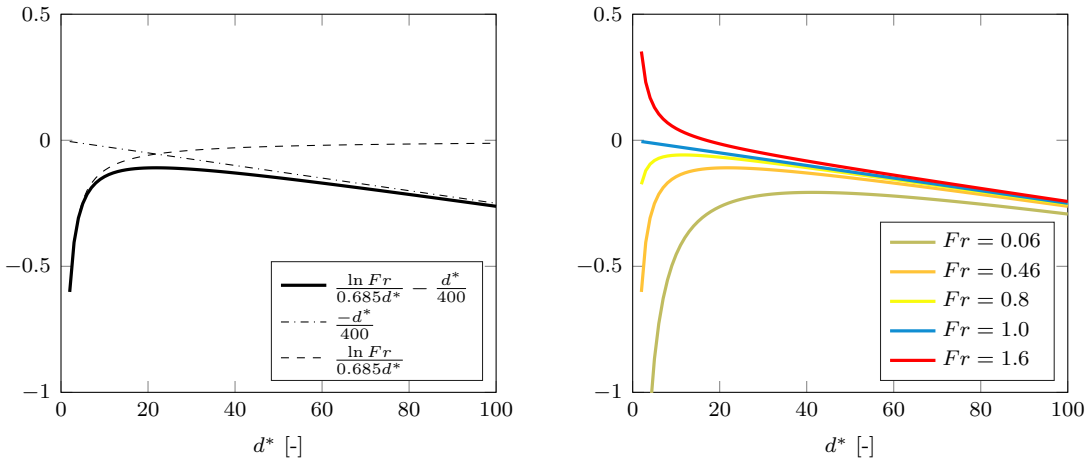


Figure 7.5: *Structure part 3 of [C] formula*

if $D/h_w > 1$ the curve decreases to account for the h_w limiting the vortex size. The last figure 7.4(e) shows how the hyperbolic tangent reduces the size of the peak to its final shape.

Two main things can be concluded from this investigation, which is that it is proven again that the sediment gradation in combination with the mobility determines the clear water peak. And secondly it can be seen that the value at $mob = 4$ remains approximately equal to a constant value of 0.5 after the addition of the Froude number in graph (b) throughout the entire process (excluding a minor influence from D/h_w at graph (d)). This means that the Fr number determines for a large part the value of the equilibrium scour depth in live-bed sediments with a high mobility.

Part 3

In figure 7.5 the effect of part 3 as defined in equation (7.1) can be seen. These figures show that the influence of part 3 on the overall equation is limited. Only for large large stone sizes ($d^* \to \infty$) the scour is decreased, as well as for fine sand at low speeds. This explains exactly the behavior of the sediment in figure 7.3(b). Interesting to see is that only if the flow is supercritical (i.e. $Fr > 1$) and the sand extremely fine, the term of part 3 is positive and this term will add to the scour depth. This can of course be correct, since at this theoretical combination of high current velocities and extremely fine sand, the entire bed could wash away, meaning the scour depth would be infinite.

7.1.3 Applicability

Range

In table 7.1 already an overview was given for the range of the input parameters in the database on which the formula is tested. To analyze the strength of this formula in more detail, each parameter is tested on its performance binning its range in the database in 5 subsections with an equal amount of tests. The results can be seen in 7.2.

Parameter		Range				
		1	2	3	4	5
mob	Step	0.0 : 0.53	0.53 : 0.74	0.74 : 1.1	1.1 : 3.7	3.7 : 480
	CoD	0.65	0.65	0.76	0.63	0.61
	RMS	0.33	0.34	0.31	0.30	0.23
σ	Step	0.0 : 1.24	1.24 : 1.30	1.30 : 1.33	1.33 : 2.24	2.24 : 7.8
	CoD	0.72	0.64	0.66	0.79	0.84
	RMS	0.31	0.33	0.31	0.33	0.23
$\frac{D}{h_w}$	Step	0.0 : 0.22	0.22 : 0.33	0.33 : 0.54	0.54 : 1.0	1.0 : 19.2
	CoD	0.56	0.64	0.68	0.74	0.58
	RMS	0.33	0.30	0.34	0.26	0.29
Fr	Step	0.0 : 0.20	0.20 : 0.28	0.28 : 0.39	0.39 : 0.63	0.63 : 1.7
	CoD	0.75	0.76	0.67	0.73	0.70
	RMS	0.34	0.33	0.32	0.27	0.25
d^*	Step	0.0 : 14	14 : 19	19 : 27	27 : 51	51 : 428
	CoD	0.71	0.84	0.78	0.79	0.60
	RMS	0.27	0.27	0.34	0.31	0.34

Table 7.2: *Performance parameters in different ranges*

Two interesting things can be deducted from this table. First of all the coefficient of determination of the mobility in the range of $0.74 < mob \leq 1.1$ is significantly higher than for the other regions of mob. This means that the formula has a limited amount of unexplained variance around the clear-water peak. Of course extra weight was added to this region to force the GP to model this part correctly, but it good to see that the GP succeeded in modeling this difficult transition area.

The second conclusion that can be drawn for table 7.2 is that the formula is more accurate for wide graded sediment, as can be seen by the increase in CoD and reduction of RMSE. This can also be explained by the shape the formula has when wide graded sediment is considered. When looking at for instance figure 7.2(a) is can be seen that the shape of the curve for high values of σ is much simpler than for uniform sediments and hence easier to predict.

Of course if a new prediction is made with this formula, its accuracy depends highly on the amount of tests that were used to train this formula with similar conditions. Therefore a method is suggested in appendix D to quantify the coverage of a new data point by the database.

If the data is outside of the input values of the GP training database, formula [C] will probably still provide an accurate result, because most parameters are bounded by either

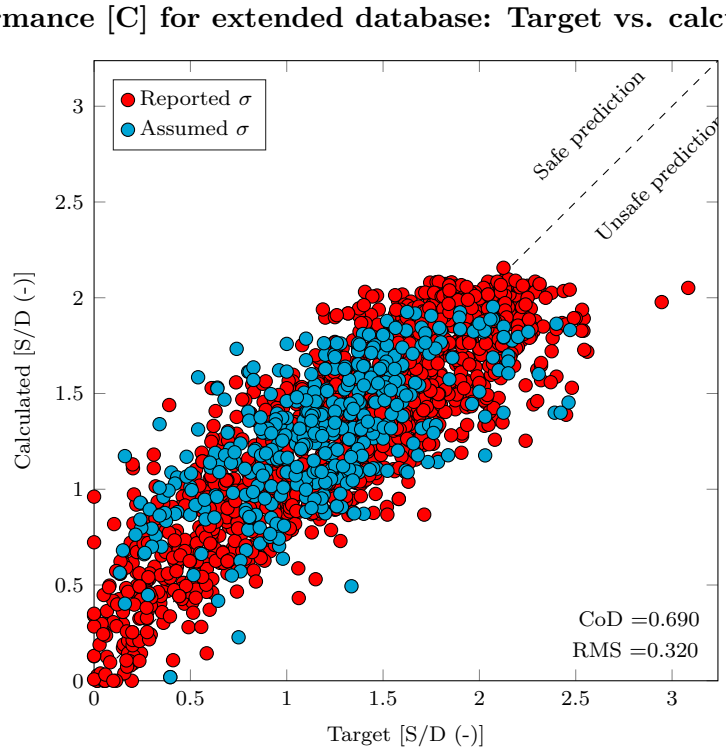


Figure 7.6: *[C] formula tested for [C] database, including tests with unknown σ*

zero or a constant value at the extremes. This is due to the use of the hyperbolic tangent, which will have a constant value if its argument is larger than 3.

Non-sigma tests

To test the quality and robustness of the [C] formula, the formula is tested with data that was not used in the training of the GP. For this end the [C] data without reported σ value was used, since these tests were omitted from the GP input database. Because the formula does need a value for σ it is assumed that the tests without reported value used standard uniform sand, which usually has a value of $\sigma = 1.3$.

In figure 7.6 the result can be seen. The additional data is predicted quite well, with the exception of a view outliers. This proves that the GP was not overtrained on the database and that the formula can be useful to predict current-induced scour even when the sediment gradation is not known.

7.1.4 Limitations and safety factor

The [C] formula of equation (6.6) is quite adapt in predicting scour. However, there is of course still a margin of error. This could be either caused by limitations of the formula or by errors in the measured scour values of the source.

A good example of this dilemma can be seen for the maximum scour depth. According to the GP formula the maximum equilibrium depth is limited at 2.2 S/D, while some

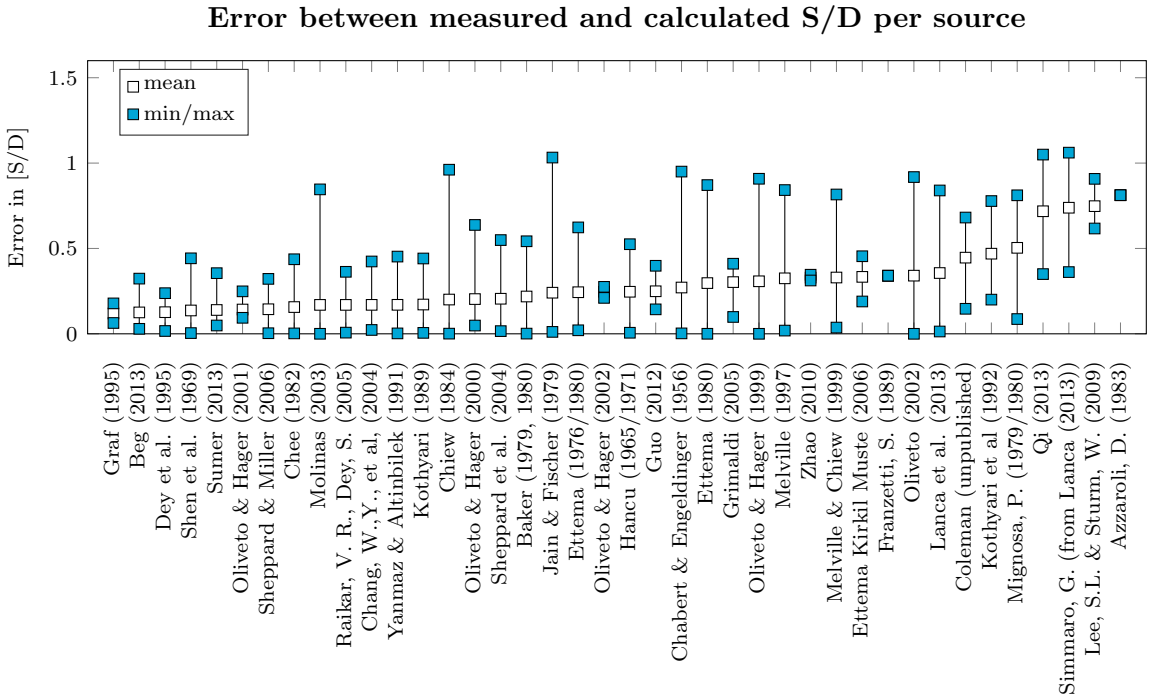


Figure 7.7: *Average error [C] formula per source*

measurements report scour values up to 2.5 S/D and in case of Jain & Fisher (1979) even values of more than 3 S/D. There is a chance that values up to $2.5S/D$ are realistic under extreme conditions and in that case the value of 2.2 S/D is a error of the formula. Values over $3S/D$ are not that realistic and therefore the large error in these tests can be assigned to an error in the measurements.

To assess the errors of the [C] formula for different data sources in more detail, figure 7.7 is created. This figure ranks the mean value of the error in S/D for each source, as well as the minimum and maximum value. With this tool the quality of the source is analyzed as well as the quality of the formula. Again a judgment is made whether an error can be assigned to a shortcoming of the source or of the GP formula.

An example of an error in the source, is the data of Qi (2013). It was unclear from the article if full equilibrium scour depth was reached, since the test duration was only 100 minutes. Therefore it seems logical that because the data from Qi (2013) was significantly underpredicted, the tests were indeed not long enough to reach the scour equilibrium state.

The 2 top ranking mean errors of Azzaroli (1983) and Lee and Sturm (2009) are maybe due to shortcomings of the [C] formula. It could be possible that their conditions are not well covered by the [C] formula. Both tests have extreme values for D/h_w. The D/h_w value of Azzaroli (1983) is small, while the D/h_w value of Lee and Sturm (2009) is relatively large. This could mean that the [C] formula does not incorporate the effects of the full range of D/h_w very well. Of course for D/h_w the extremes are also the most difficult to model, since the effect of the limit of the vortex size by either D or h_w needs to be incorporated.

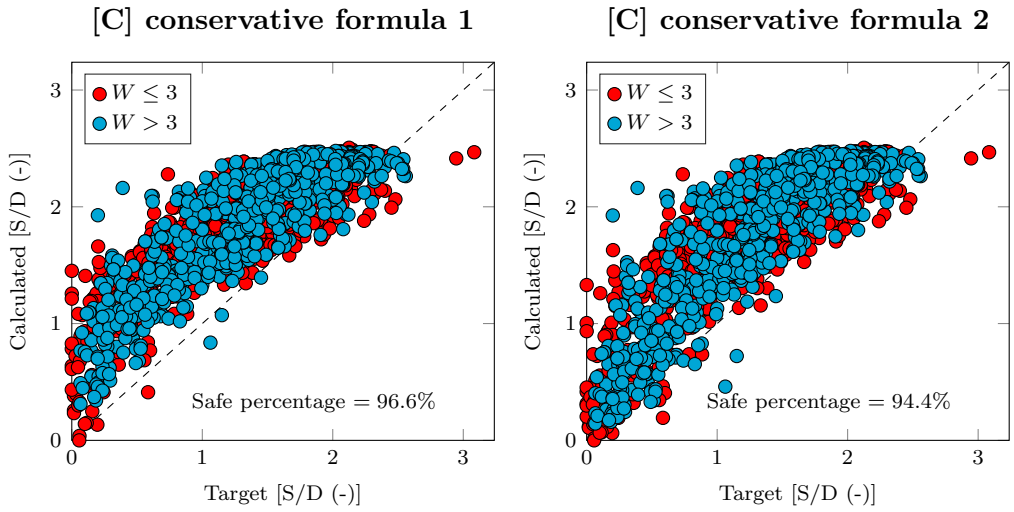


Figure 7.8: *Predicted values versus the actual values using the [C] formula including safety factor, for current-only conditions*

Safety factor
For design considerations it is important to add a safety factor. This will increase the error per test, but towards a more conservative design. The constants of formula 6.6 are adjusted in such a matter that the larger part of the data is in the safety region. This conservative formula is given by equation (7.2) and shown in figure 7.8(a).

$$S/D = 1.5 \tanh\left(\frac{2.9 \cdot mob}{\sigma}\right) + \tanh\left(0.45\frac{h_w}{D}\left(\frac{0.71 \cdot mob}{\exp\left(2.3mob\right)\sigma} + 0.49\tanh\left(Fr\right)\right)^{0.68\sigma}\right) - \frac{d^*}{590} + \frac{\ln\left(Fr\right)}{2.3d^*} \tag{7.2}$$

For 96.6% of the input conditions equation (7.2) will lead to a conservative design. Especially for the smaller values of measured S/D, the predicted scour depth will be significantly overestimated. To reduce this effect another term is added to the equation that only corrects the lower region. This can achieved by multiplying equation (7.2) with another tanh function: $\tanh\left(0.7\{equation\ (7.2)\}^4 + 0.5\right)$.

In figure 7.8(b) the effect of this term in shown. It is already a better fit than just adjusting the constants of equation (7.2), but this additional term has no ideal shape and hence some more optimization could be done here.

7.2 [W] formula

The formula for [W] conditions will be evaluated in the same manner as the [C] formula in the previous section. To revise, the formula found with the GP for wave-only conditions

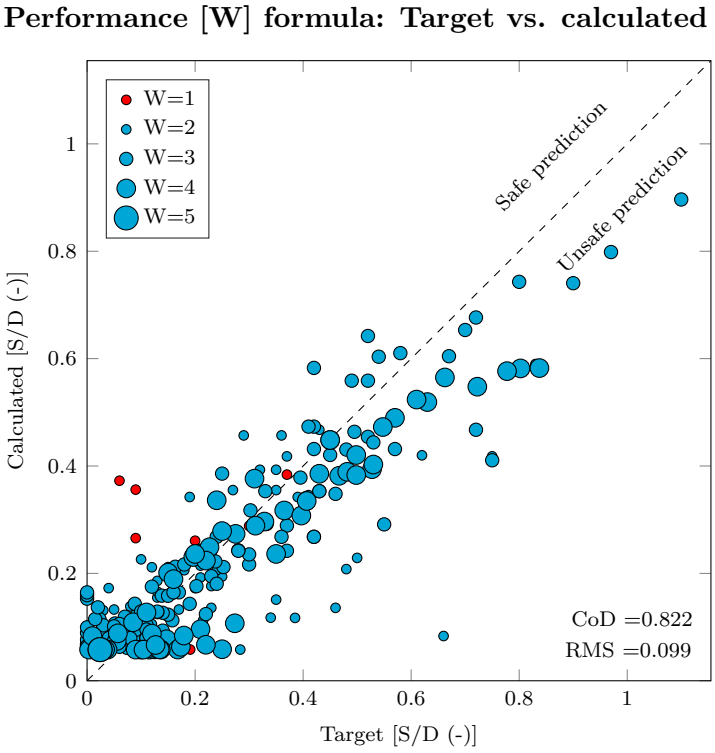


Figure 7.9: *Predicted values versus the actual values using the [W] formula, for wave-only conditions*

is given by 6.9.

$$S/D = 1.058 - \tanh(D/h_w + 14.1/KC) \qquad \text{(6.9 revisited)}$$

Figure 7.9 shows the measured values versus the predicted values, including the weights ranked from 1 to 5. It can be seen that for larger S/D values the model is quite accurate, while for small values of S/D there is a wide range of variance.

Only 2 parameters are necessary to calculate S/D in [W] conditions. The range of these parameters on which the formula is based is listed in table 7.3 as well as the range of the dimensional parameters required to calculate KC and D/h_w.

7.2.1 Behavior formula key parameters

The key parameters KC and D/h_w are plotted in figure 7.10(a) and (b). In figure 7.10(a) the behavior of the formula is shown for KC versus the equilibrium scour depth. As to be expected the equilibrium scour depth increases with increasing KC numbers. This is plotted for different values of D/h_w, which shows that the equilibrium depth decreases with increasing D/h_w ratio. For $D/h_w = 1.5$ the equilibrium value is almost constant, which suggests that at this point, the scour depth is limited by the pile diameter.

It is also visible in figure 7.10(a) that the formula has a threshold value for KC. For $KC < 4$, the equilibrium scour depth remains at its initial value. This is a lower KC

Required input parameters			
Parameter	Range		
	min	max	mean
D [m]	0.01	1.53	0.14
h_w [m]	0.15	0.5	0.39
H_s [m]	0.025	0.39	0.15
T_p [s]	1	4.5	2.4

(a)

Formula parameters			
Parameter	Range		
	min	max	mean
KC [-]	0.21	102	13.8
$\frac{D}{h_w}$ [-]	0.025	3.8	0.38

(b)

Table 7.3: *List of required parameters for [W] formula and the range and mean values of the database on which the formula is based*

threshold than at $KC = 6$ as suggested by Sumer et al. [14]. This lower threshold value of $KC = 4$ is considered to be credible, since it was seen in multiple other well performing [W] equations generated by the GP as well.

The only non-physical behavior that the [W] formula shows, is that for $KC = 0$ the value of $S/D = 0.06$. If $KC = 0$ there are no waves or any other movement, so the scour should also be zero. However since the value is very small, this will be considered as a small safety factor.

Figure 7.10(b) shows the counterpart of figure 7.10(a). In this graph D/h_w is plotted against the equilibrium scour depth for different KC values. Obviously the same conclusion can be drawn from this figure, that the scour depth decreases with increasing D/h_w values.

7.2.2 Applicability

Table 7.4 shows the performance of the [W] scour equation when the data in the database is grouped in 5 bins with an equal amount of tests for each parameter.

This table confirms what was already clear from figure 7.9: that the formula is better suited for high KC numbers. The variance at low S/D values is not at explained at all by the formula, which is indicated by the zero value of CoD. However the errors in this region are also small. It can be seen in figure 7.9 that low S/D values with large variance all have errors below 0.2 S/D. This means that the RMSE value is also small, which is confirmed by the values in table 7.4. Possibly a better formula could be found with the GP by increasing the importance of tests with $KC < 10$. However, this might influence the prediction at higher KC values, whose accuracy is more relevant from a design point of view, than those of the practically non-existent scour in the lower regions.

Similar to the [C] formula, also this formula is can be used outside the maximum value of the parameters that were present in the database, since also this formula is limited by a hyperbolic tangent, which ensures that the equilibrium depth does not vary for larger values.

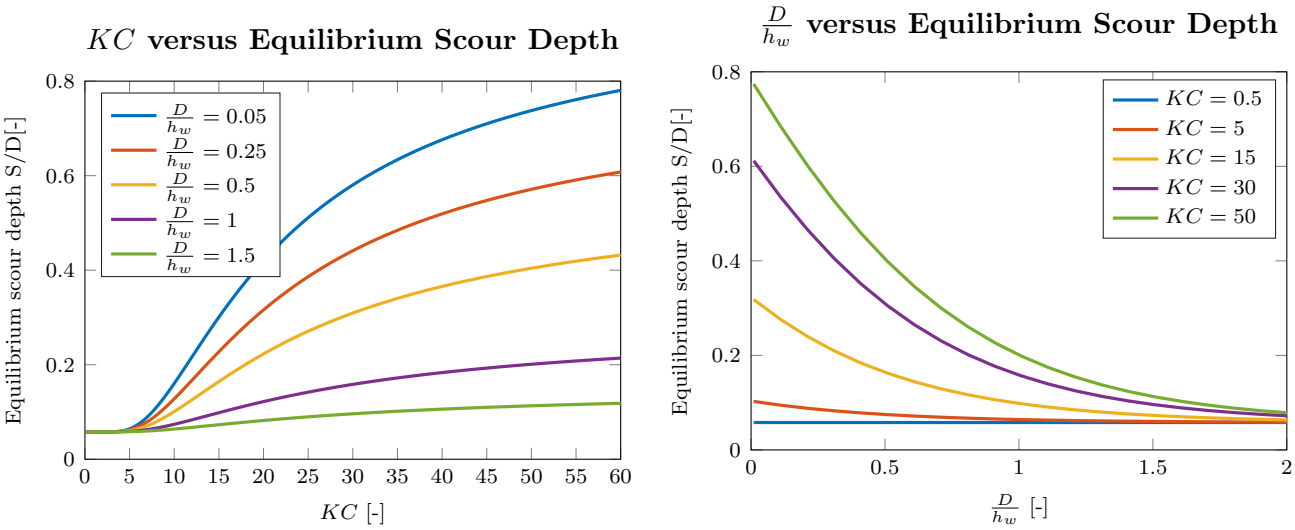


Figure 7.10: *Formula behavior [W]: scour depth prediction with increasing wave height (a) and increasing pile diameter (b). [d_{50}=0.0013 m, D=0.2 m, h_w=0.5 m, H_s=0.14 m, T_p=1.7 s]*

Parameter		Range				
		1	2	3	4	5
KC	Bin	0.0 : 5.7	5.7 : 9.2	9.2 : 13.5	13.5 : 21.3	21.3 : 102
	CoD	0.04	0.00	0.30	0.37	0.66
	RMS	0.07	0.10	0.08	0.11	0.12
$\frac{D}{h_w}$	Bin	0.0 : 0.10	0.10 : 0.12	0.12 : 0.15	0.15 : 0.34	0.34 : 3.8
	CoD	0.76	0.94	0.68	0.63	0.37
	RMS	0.13	0.09	0.10	0.09	0.08

Table 7.4: *List of performance for the [W] database for specific ranges of input parameter, for all parameters of the [W] formula*

7.2.3 Limitations and safety factor

The mean error per source for the [W] formula is given by figure 7.11. It shows that the average error is much smaller than for the [C] tests. This is not only because the predictions are more accurate, but also because the average scour value is lower in wave conditions compared to current conditions.

The highest error was found for the Sumer (2007) tests that contained different ranges of sediment density (indicated by [s]). The formula given by equation (6.9) is not equipped to implement this effect.

Safety factor
Also for the [W] formula the equation is adjusted to a conservative prediction method.

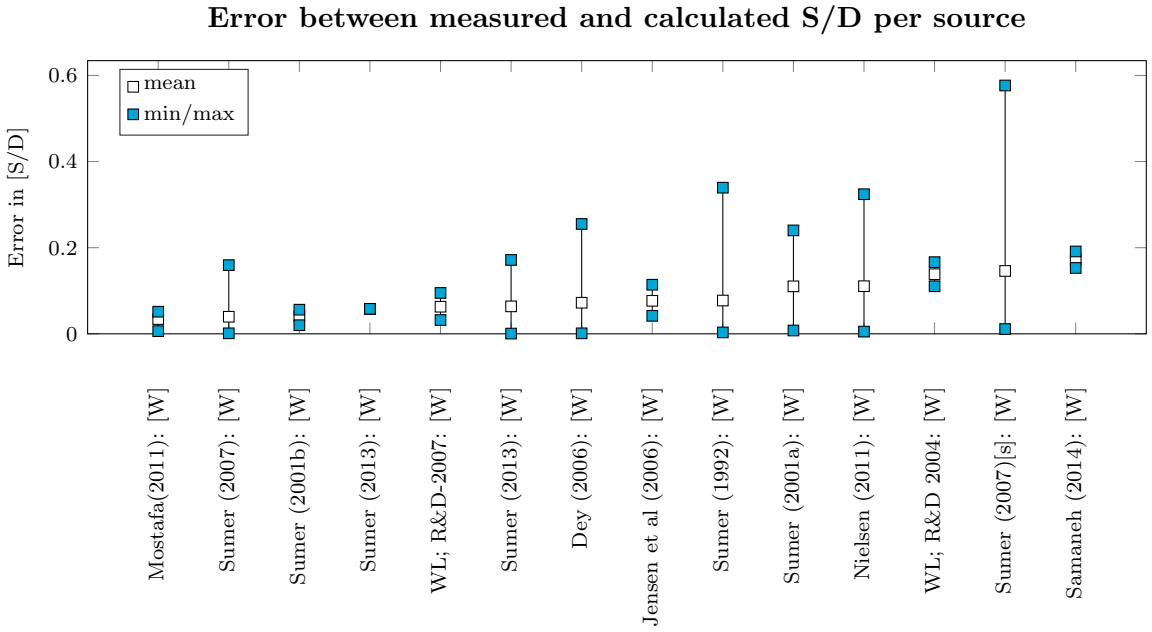


Figure 7.11: *Average error [W] formula per source*

As previously mentioned the original equation already contains a safety factor for scour below the KC threshold. To ensure that all other values are in the safe prediction regime as well, the constants of equation (6.9) are adjusted as shown in equation (7.3). The resulting calculated values are plotted against the measured S/D values in figure 7.12.

$$S/D = 1.2 - \tanh(D/h_w + 12.6/KC) \tag{7.3}$$

7.3 [CW] formula

In this section the combined equation for current-only, wave-only and combined waves and current conditions will be discussed, which was given by equation (6.15). The predicted versus the actual measured scour depth is shown in figure 7.14. This contains [CW] data only, since the formula is designed is such a way that for [C] tests and [W] tests the results are equal to the aforementioned [C] and [W] equation.

$$S/D = u_{rel} \cdot f_{[C]} + (1 - u_{rel}) \cdot f_{[W]} + {Fr_w}^2 \cdot \sqrt{\frac{\tanh\left(Fr \cdot {w_s}^* \cdot KC^{1.09}\right)}{\exp\left({w_s}^*\right)}} \quad \text{(6.15 revisited)}$$

The range of parameters that are necessary to calculate S/D are given by 7.5.

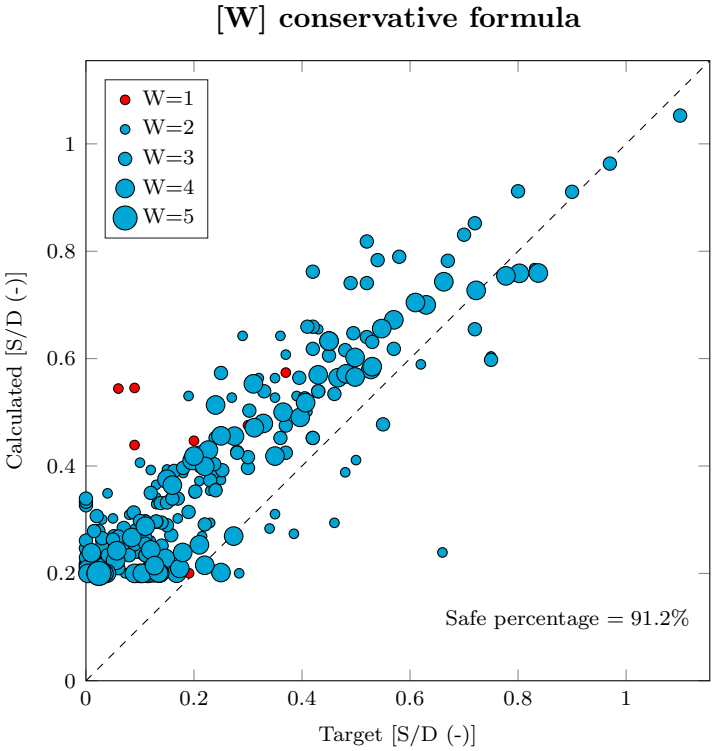


Figure 7.12: *Predicted values versus the actual values using the [W] formula including safety factor, for wave-only conditions*

7.3.1 Formula parts analysis

In this section the term that is added to $u_{rel} \cdot f_{[C]} + (1 - u_{rel}) \cdot f_{[W]}$ in equation (6.15) to represent the combined current and wave data is analyzed.

An important part of what makes this formula successful is that it goes to zero for both current-only and wave-only equations. This is ensured by the multiplication with KC and Fr. But due to the exponential function the dominant parameter in determining what value this extra term has is the dimensionless fall velocity, i.e. the rate at which the disturbed sediment settles. The behavior of the term is shown in figure 7.14. Because the values in this graph are so low, it suggests that the term $u_{rel} \cdot f_{[C]} + (1 - u_{rel}) \cdot f_{[W]}$ by itself is already quite adapt in predicting the transition regime between [C] and [W]. The extra term in equation (6.15) plotted in figure 7.14 is only necessary for low sediment numbers and high KC values.

7.3.2 Behavior formula key parameters

In figure 7.15(a) and (b) the behavior of the [CW] formula is shown for its key parameter: the relative velocity. u_{rel} is varied in two different ways: by increasing the current velocity u_c in figure (a) and the wave orbital velocity u_w in figure (b).

In figure 7.15(a) at $u_{rel} = 0$ the scour equilibrium value is equal to the value found for wave-only conditions. The value at $u_{rel} = 1$ corresponds to the equilibrium scour depth

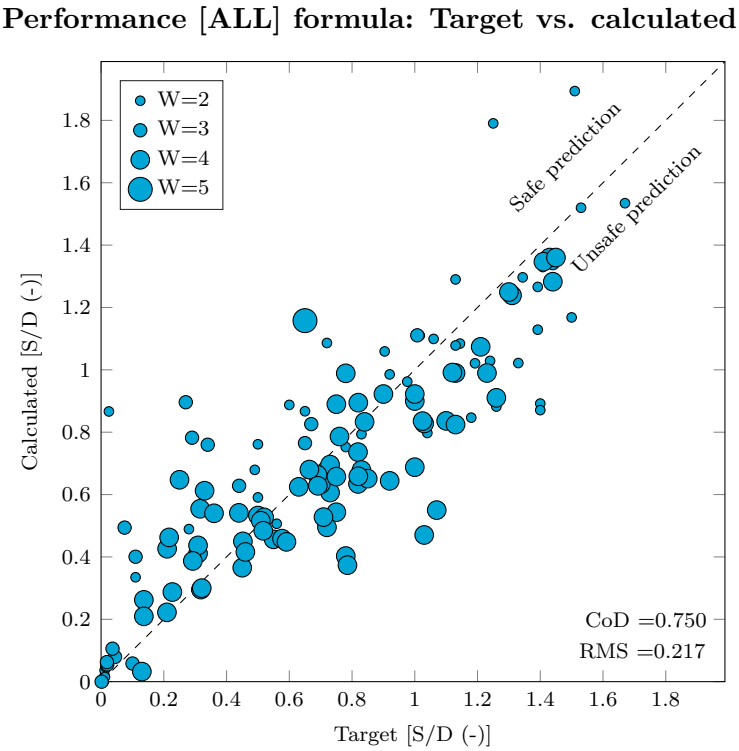


Figure 7.13: *Predicted values versus the actual values using the [CW] formula, for current-only, wave-only and combined conditions. Performance on [CW] data only*

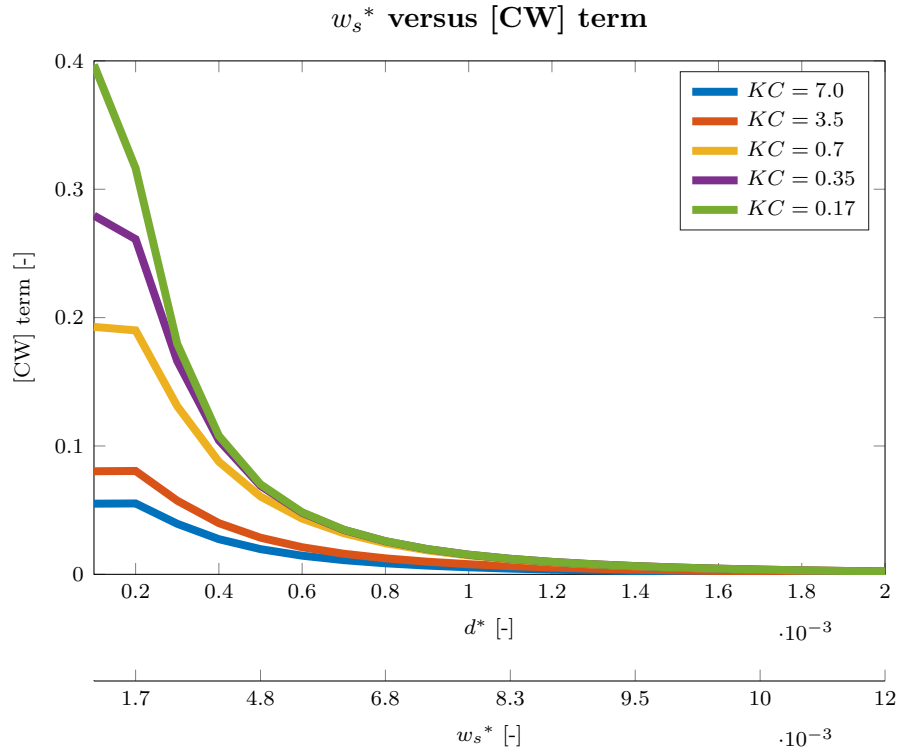


Figure 7.14: *Formula behavior [CW] additional term: [Fr=0.47]*

Required input parameters			
Parameter	Range		
	min	max	mean
d_{50} [mm]	0.1	0.5	0.19
D [m]	0.03	0.67	0.14
h_w [m]	0.15	0.75	0.39
u_c [m/s]	0.04	0.84	0.22
σ [-]	1.28	1.3	1.3
ρ_s [kg/m^3]	2450	2650	2627
H_s [m]	0.022	0.51	0.15
T_p [s]	0.84	7.8	2.4

(a)

Formula parameters			
Parameter	Range		
	min	max	mean
mob [-]	0.11	70	10.4
$\frac{D}{h_w}$ [-]	0.077	4.2	0.54
Fr [-]	0.02	0.42	0.11
d^* [-]	2.4	12.6	4.8
σ [-]	1.28	1.3	1.3
KC [-]	0.11	33	9.3
Fr_w [-]	0.68	1	0.9
u_{rel} [-]	0.12	0.86	0.45
${w_s}^*$ [-]	0.42	4.8	1.5

(b)

Table 7.5: *List of required parameters for [ALL] formula and the range and mean values of the database on which the formula is based. Ranges for [CW] values only, for the others, see previous sections table 7.1 and 7.3*

with $u_c \to \infty$. All values in the combined wave and current regime in the middle increase almost linearly with u_{rel} for uniform sediments. For wider graded sediments the slope is less steep until a certain threshold at $u_{rel} > 0.7$. At this point the current is strong enough to counteract the armouring effect, from this point on all curves converge towards the equilibrium value of current-only conditions.

In figure 7.15(b) the effect of ranging u_{rel} from 0 to 1 by increasing the orbital wave velocity is shown for a range of water depths. The value at $u_{rel} = 1$ the scour equilibrium value is equal to the value found for current-only conditions. At $u_{rel} = 0$ the value corresponds to the equilibrium scour depth with $u_w \to \infty$.

This figure can best be analyzed in 3 parts, moving from right to left: from $0.8 < u_{rel} \leq 1$, $0.2 < u_{rel} \leq 0.8$ and $0 < u_{rel} \leq 0.2$. It is generally considered that when waves are added to the system, the scour reduces, due to the waves destructing the horseshoe vortex created by the current. However, this is only true for water depths smaller than 0.2 m. According to this formula and under the conditions described in the caption of the figure, the scour depth increases when there is a combination of small waves and current. This effect will not increase above the value seen here by the green line for larger h_w. Possibly this could occur when the orbital motion of the wave resembles the characteristic rotations of the horseshoe vortex at the bed. Which would mean that the effects enhance each other and therefore the scour depth increases. However, it makes no sense that this would occur at larger water depth and that the effect remains the same, even if the $h_w \to \infty$. These small sized waves would not even be noticed at the bottom where the scouring occurs.

There is no physical explanation for this behavior, but it can be explained why the formula behaves in this matter. This is due to the constant term in the [W] equation. It

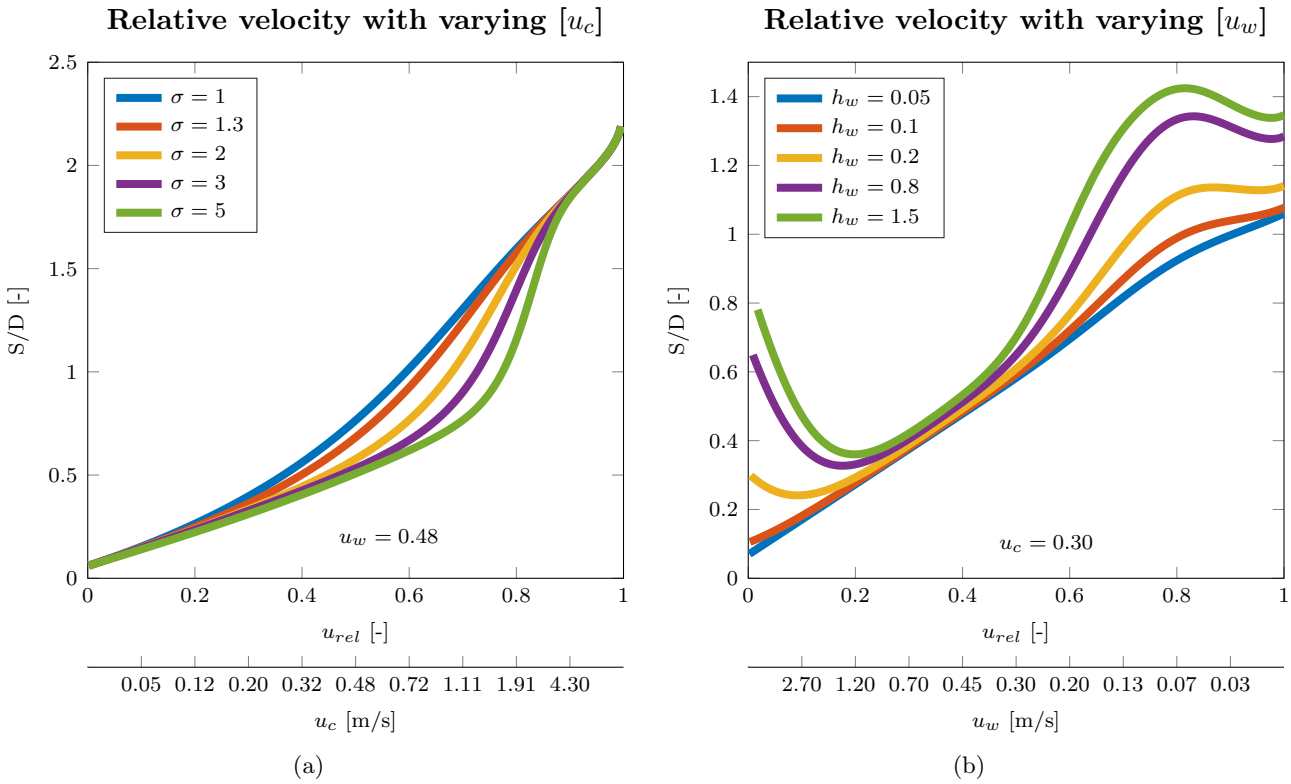


Figure 7.15: *Formula behavior [CW]: scour depth prediction with increasing current velocity (a) and increasing wave velocity (b). [d_{50}=0.0013 m, D=0.2 m, h_w=0.5 m, H_s=0.14 m, T_p=1.7 s, $\sigma = 1.3$]*

was already noted it the previous section that this leads to non-realistic behavior, since the scour is not zero when there are no waves and no current. The effect that this has on the combined equation can be seen here. Therefore, for the combination of very small waves, with a large water depth and a current, the formula will be inaccurate. Since the equilibrium value is estimated higher than expected, the prediction is this region will be conservative.

Continuing with the other 2 parts of this graph, it can be seen that between $0.2 < u_{rel} \leq 0.8$ the formula does behave according to the expected behavior and the equilibrium scour depth decreases when larger waves are added to the system. This happens until it reaches the last region $0 < u_{rel} \leq 0.2$, where the waves are large enough to generate their own scouring effect. Lee-wake vortices are created that are strong enough to act as sediment deposits. When the water depth is shallow, the waves are limited by h_w and scouring can not occur, as can be seen by the flat blue line in figure 7.15(b). For larger water depths the effect of wave scour can build up which is why for the largest water depth, there will be the largest amount of scour.

7.3.3 Applicability

7.3.4 Range

The range of the database on which the [CW] formula was tested consist of table 7.2 and 7.4 combined. The additional term to the equation does limit the validity of the equation for larger input parameters than the maximum value of the database. This is because the Fr_w term is not limited by a tangent hyperbolic or other function, which means that it can grow to infinite. This gives inaccurate results, since the scour depth is actually bounded by a certain S/D limit. Theoretically this Fr_w term can grow to infinity, however, it is observed that this occurs at such a slow pace, that even with extreme wave heights above 30 m, the predicted value is still only ≈ 1 S/D.

Submerged-piles

The [CW] formula is created for piles that are equal or larger than the water depth. However, it is quite simple to adjust the formula is such a matter that it is valid for submerged piles as well. If no additional term is added to account for the relative pile height, the predicted scour depths will be very conservative. In case of $h_p/h_w < 1$ the length over which the adverse pressure gradient is created is smaller and hence also the pressure gradient and the horseshoe vortex will be smaller than under the same conditions with $h_p/h_w < 1$.

A term like $\frac{h_p}{h_w}^{c_1}$ where c_1 is a constant smaller than 1, assures that the performance of tests with $\frac{h_p}{h_w} = 1$ is still exactly the same as was fitted before, while improving the results for submerged piles. A value of $c_1 = 3$ is found and the complete term is multiplied with equation (6.15). The results for submerged tests can be seen in figure 7.16.

7.3.5 Safety factor

The safety factor for [CW] is created by replacing $f_{[C]}$ and $f_{[W]}$ in 6.15 by the [C] and [W] formulas including safety factor as defined in previous sections. The additional term in the [CW] are refitted, to ensure that also the combined wave and current test are conservative. The result is given by equation (7.4) and figure 7.17.

$$S/D = u_{rel} \cdot f_{[C]} + (1 - u_{rel}) \cdot f_{[W]} + {Fr_w}^2 63 \cdot \sqrt{\frac{\tanh\left(Fr \cdot {w_s}^* \cdot KC^{-67}\right)}{\exp\left({w_s}^*\right)}} \tag{7.4}$$

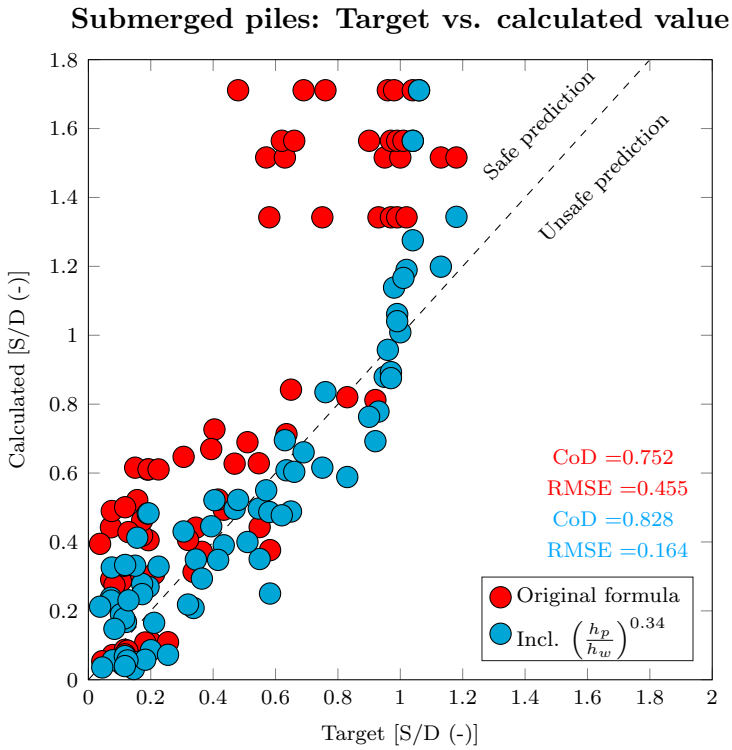


Figure 7.16: *Predicted values versus the actual values using the [ALL] formula, for current-only, wave-only and combined conditions. Performance on [CW] data only*

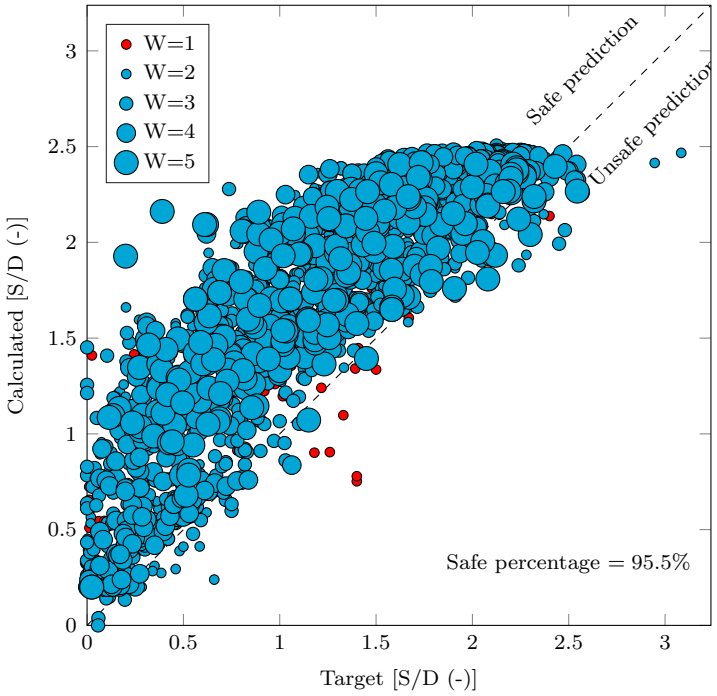


Figure 7.17: *Predicted values versus the actual values using the [CW] formula including safety factor, for all conditions*

Chapter 8

GP Performance Comparison

In this chapter the quality of the three formulas for [C], [W] and [CW] conditions will be judged in comparison with the existing scour prediction methods mentioned in chapter 2.4 and with a neural network model created with the same database. Of course it has to be kept in mind that the GP has the advantage over the other equation that it is created for this specific database.

8.1 Comparison existing scour equations

[C]

The performance of the GP is significantly better than the methods of Breuser [18] and Molinas [16]. The formula of Breuser predicts a maximum value of 1.5 S/D for almost all tests. This shows that this formula does not incorporate enough parameters to accurately predict scour. The method of Molinas overpredicts all values, which leads to an excessively conservative design. The method of Sheppard [22] has a similar performance to the GP for $S/D > 1$. For smaller equilibrium wave depths, the GP has a higher accuracy since the predictions of Sheppard are more scattered. The advantage of Sheppard's equation is that it contains a safety factor. For each test it predicts a conservative value, without overshooting it like Molinas. The exception is for tests with $u_{crit} \leq 0.47$, for these conditions the value is estimated to be zero and hence in most cases severely underestimated. The [C] formula from the GP does not have this limitation, which makes it a better allround formula.

[W]

A comparison between the [W] formula found by the GP and the formula for wave-only conditions by Sumer et al. [14] is given in figure 8.2. It can be seen here that the formula established by the GP is a good fit. In comparison with the formula of Sumer, it contains

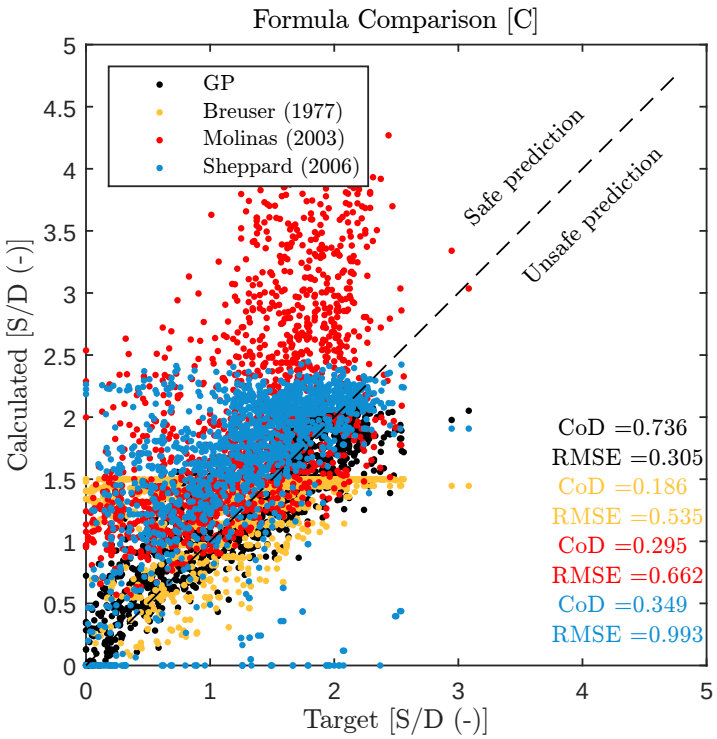


Figure 8.1: *Comparison of GP [C] formula with existing empirical formulas (Chapter 2.4) for the [C] only database*

less unexplained variance (higher CoD) and also the values on the x-axis with $KC < 6$ are better predicted.

[CW]

The performance of the combined equation was tested against the empirical formulas of Sumer et al.[14], Rudolph[23] and Raaijmakers and Rudolph[7]. The results are shown in figure 8.3. It is clear that the [CW] formula created by the GP provides the best estimations of scour depth for the combined current and wave data for the higher scour regions. At this point all other equations are limited to 1.3 or 1.5 S/D. For the lower scour values it can be seen that the method of Rudolph[23] and Raaijmakers and Rudolph[7] estimate more values in the unsafe prediction region, while the estimates of the GP are more on the conservative side.

8.2 Comparison Neural Networks

A neural network model (NN) is another form of soft computing. Similar as to genetic programming, it is a data-oriented modeling technique that can identify relations between input- and output patterns without using any knowledge of the process. In this case a Multi Layer Perceptron (MLP) neural network is used, which means that the architecture of the NN consists of multiple layers. This architecture can be seen in figure 8.4. The first

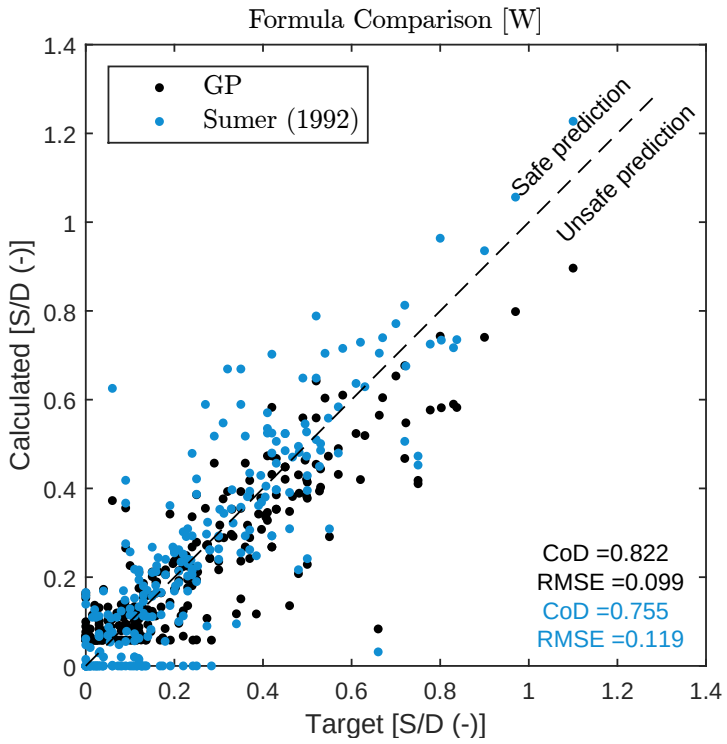


Figure 8.2: *Comparison of GP [W] formula with existing empirical formulas (Chapter 2.4) for the [W] only database*

layer contains the input parameters and the third layer the output value. These layers are connected to each other by a hidden layer with a certain number of weights. The learning process of the NN consists of adjusting the values of the weights. The program starts with a random set of weights, which are optimized in such a way that the difference between the calculated value and the target value is minimized.

A neural network is very flexible system, which can produce highly accurate predictions if the right architecture is used. The disadvantage of NN compared to a GP is that it gives no insight in the process that determined its predicted equilibrium scour value. It is basically a black box with a large set of numbers that represent the weights. Therefore it also less practical than the GP formula. The main advantage is that it gives more accurate predictions. Especially with all the knowledge obtained with the GP, it is easy to create an NN with the best database and the right set of input parameters.

[C]

A neural network is created with the same database as for the [C] formula. It contains the parameters mob, σ, D/h_w, Fr, and d^*. The prediction can be seen in figure 8.5. As is visible here the performance of both soft computing techniques are quite similar.

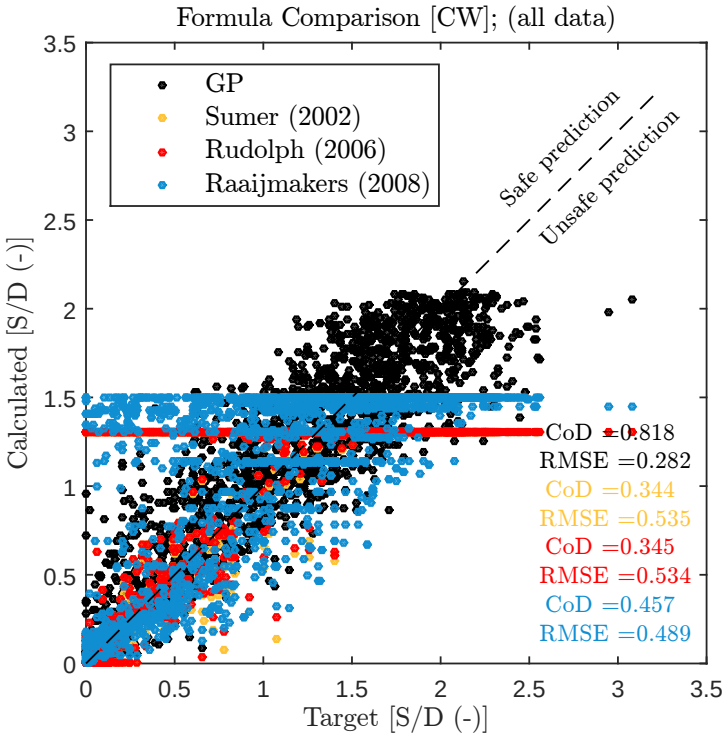


Figure 8.3: *Comparison of GP [CW] formula with existing empirical formulas (Chapter 2.4) for the all data*

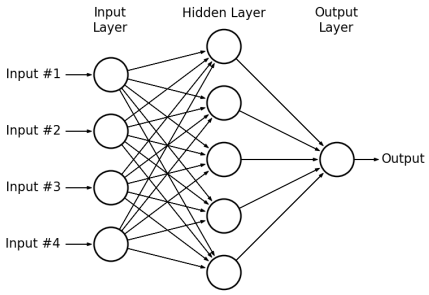


Figure 8.4: *Architecture neural network[47]*

[W]

An NN is created for the database with [W] data as well. This model is compared to the [W] formula. The NN contains the parameters KC, D/h_w, d^*, ${w_s}^*$ and Fr_w. The prediction can be seen in figure 8.6. As is visible here the predictions of the NN are more accurate than the GP, with a CoD value of 0.914 and a RMSE of only 0.065. This suggests that there is a relationship between the aforementioned parameters that can predict the scour even better than the current [W] formula from the GP. However, because the database for [W] does not contain that many tests and because the NN is a very lenient system, there is a chance that this data is overfitted.

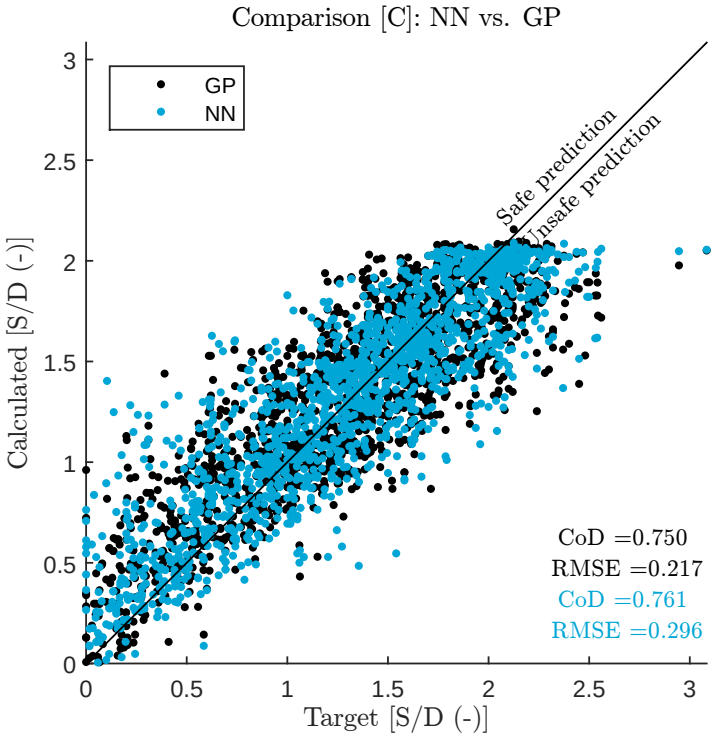


Figure 8.5: *GP [C] formula versus NN*

[CW]

Also for the complete database a Neural Network is created. It contains the parameters mob, σ, Fr, KC, D/h_w, u_{rel}, d^*, ${w_s}^*$ and Fr_w. The prediction can be seen in figure 8.7. The NN actually has a better performance than the GP. This can be seen by the reduction of variance in figure 8.6.

8.3 Comparison with top 50 GP formulas

The last method that will be presented, is a method that combines the best GP equations. As extensively explained in chapter 6, in the process of finding the best scour prediction formula, not just one equation is found. The best 50 equations of each category [C], [W] and [CW] are combined to estimate the equilibrium scour depth. The mean value of all these equations for each tests represents the predicted equilibrium scour depth. The result can be seen in figure 8.8, where the black lines represent the range between the maximum and the minimum value.

The performance of this method in terms of CoD has increased with respect to the single [CW] equation. The disadvantage is that the RMSE has also increased, and of course 50 equations are less practical than just one. The combined GP prediction is however suited for one other purpose. When predicting the value of the scour depth including safety factors, taking the maximum found value of the combined GP model including one standard deviation, gives a very slim, yet safe result for almost all tests as shown in 8.9.

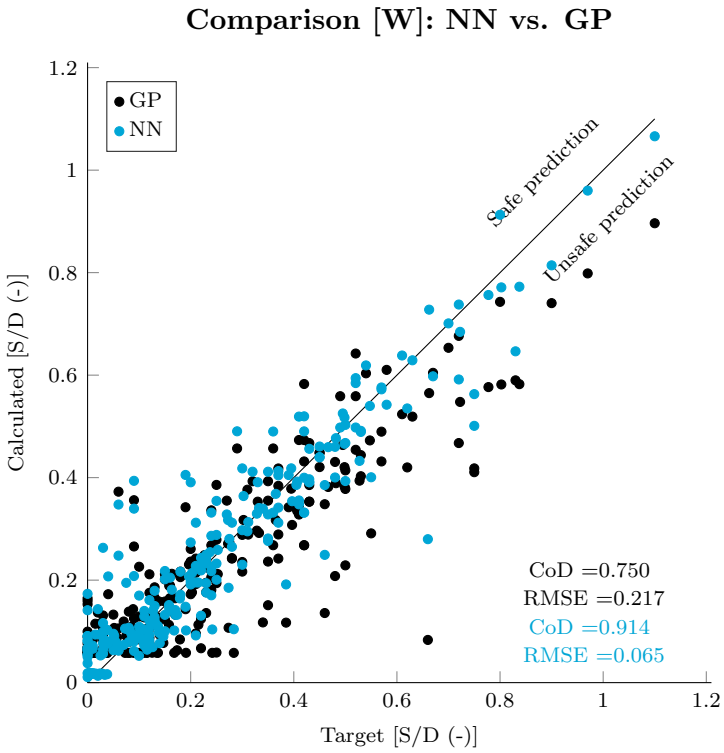


Figure 8.6: *GP [W] formula versus NN*

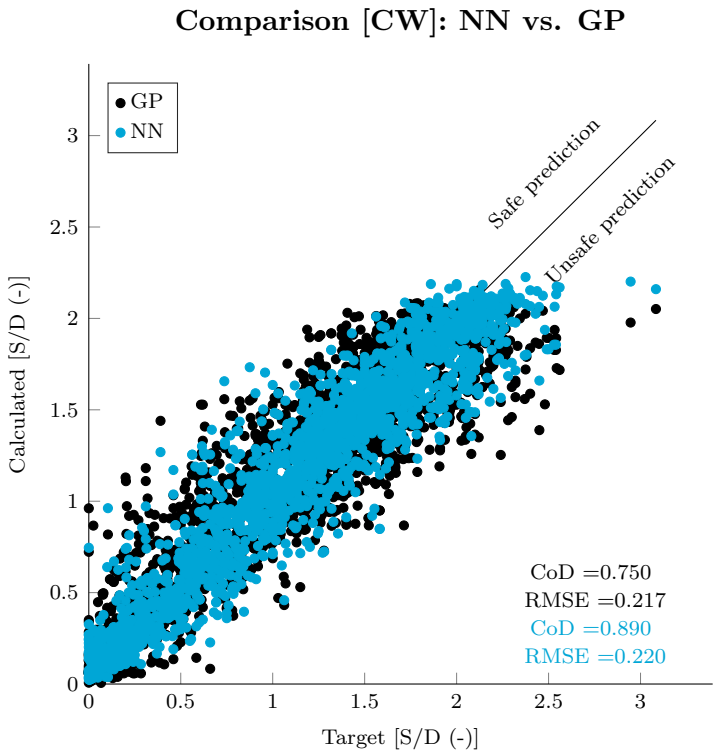


Figure 8.7: *GP [CW] formula versus NN*

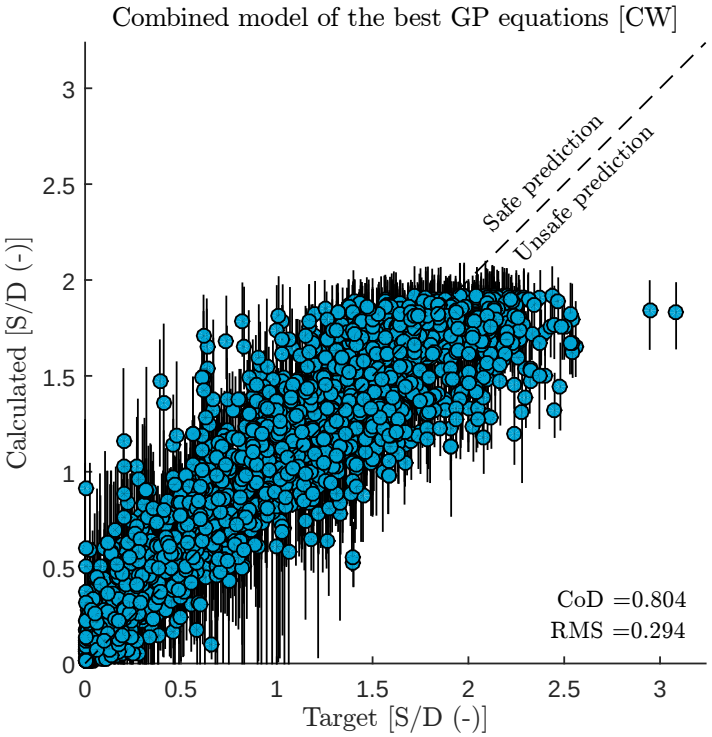


Figure 8.8: *Mean prediction value of the best GP equations*

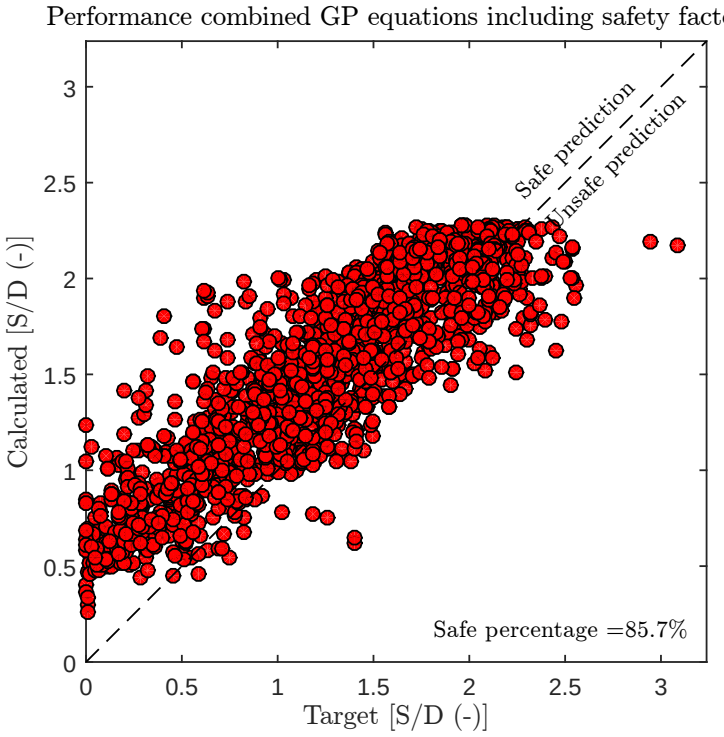


Figure 8.9: *Maximum prediction value of the best GP equations*

Chapter 9

Conclusion and recommendations

9.1 Conclusion

The aim of this master thesis was to use genetic programming to create a scour prediction model that can compute current-induced, wave-induced and combined current and wave induced scour. Applying different settings and input parameters for the GP, first successfully a formula was found to describe current-only conditions followed by a formula for wave-only conditions. These formulas where combined into one equation to predict scour in all hydrodynamic conditions. The formulas where analyzed both on their mathematical and physical behavior and it was concluded that they could accurately predict scour in all conditions.

Parameters
The most important parameters that describe the scour phenomena were determined with multiple types of parameter sensitivity analyses. For current-induced scour they were found to be the relative mobility mob, the shallowness of the water $\frac{D}{h_w}$, the Froude number Fr, the sedimentological diameter d^* and the sediment gradation σ. An important finding was the paramount significance of σ in predicting current-induced scour.

It was found that the relative critical velocity u_{crit} and the mobility both had a similar prediction accuracy, but only with the use of the mobility, the GP would create a formula that showed a clear-water peak.

The most important parameters for wave-induced scour were found to be limited to only two parameters: the KC number and the shallowness of the water $\frac{D}{h_w}$. This is in agreement with previous findings from other sources. For the combined current and wave conditions, it was found that the relative velocity u_{rel}, the Froude wave number Fr_w and the particle fall velocity ${w_s}^*$ are parameters of importance, in addition to the parameters of the individual conditions.

Physical analysis
The scour prediction formula found in this study is able to model all known physical relations. One of its strengths is that it models the clear-water peak at exactly $mob = 1$.

This study revealed that the height and sharpness of the clear-water peak is dependent on the sediment gradation. For wide graded sediments the peak is completely flatted.

Furthermore the formula is able to model a smooth transition between current-only, wave-only and combined waves and currents by the implementation of the relative velocity u_{rel}. Only for extremely small waves combined with a current in deep water, the formula will show some unwanted behavior. However, the error caused by this behavior is small and towards a more conservative design.

For the wave-only case it was found that the offset of scour started at KC values above 4. Other interesting finds were that the maximum scour depth is limited at approximately 2.2 S/D and not by 1.3 or 1.5 S/D as suggested by other sources.

Validity range
Dimensionless parameters are used to ensure that the formula is also applicable for field tests. Furthermore an extensive database with data from multiple sources was used to formulate the equilibrium scour prediction model, which means that a wide range of conditions was represented. A method is suggested to validate if a new prediction point is covered by the database.

Outside of the maximum values of the database, the prediction method is less reliable but still valid; Due to the frequent use of the hyperbolic tangent in the formula, all parameters in the equation are bounded by either zero or a constant value at the extremes.

Comparison with existing methods
The new scour prediction method was compared to various existing scour prediction methods to observe if improvements have been made. It was seen that the formula created in this study predicted more accurate scour depths, especially for test with larger scour depths.

Comparison with neural networks
This study was finalized with a comparison to a second soft computing method: the neural network. For the same database and input parameters, the scour depths were predicted with the NN. It was found that the GP is less successful in predicting the scour depth compared to the NN. However, the high accuracy of the NN could not have been achieved without the knowledge of the parameter behavior obtained by the GP.

9.2 Recommendations

The results found in this study were very satisfying, but of course there is always potential for even better results by improving the method or extending its range of application.

There is a number of parameters that were not considered during this study because not enough measurement data was available to train the GP. Examples of interesting influences could be the angle between the current and the waves, the different shapes of a pile, scouring under a slope, cohesive sediment, et cetera.

It is recommended to try more advanced settings with the GP. For example, the different initialization methods described in chapter 5. It is also advised to use a GP with the ability to incorporate constants. It is expected that this will give a significant improvement in the model.

With respect to the found formula. It is recommended that a more extensive study is conducted to optimize the formula that included the safety factor. Furthermore the formula should be tested with field data to be able to evaluate its true value.

For further studies it would be interesting to examine the effect of sediment gradation in [W] and [CW] conditions as well. This was not possible with the database of this study, since no information was available on [W] or [CW] data with wide graded sediment.

References

[1] Ting F.C.K. Chen H.C. Gudavalli R. Perugu S. Wei G. Briaud, J.-L. SRICOS: Prediction of scour rate in cohesive soils at bridge piers. *J. Geotechnical and Geoenvironmental Engineering, ASCE*, 125(4):237–246, 1999. as cited in Sumer and Fredssøe (2002).

[2] Mb Zaaijer and J Van Der Tempel. Scour protection: a necessity or a waste of money? *TU Delft*, 2004. URL http://aerospace.lr.tudelft.nl/fileadmin/Faculteit/LR/Organisatie/Afdelingen_en_Leerstoelen/Afdeling_AEWE/Wind_Energy/Research/Publications/Publications_2004/doc/Tempel_scour.pdf.

[3] Tim Raaijmakers. Modelling of equilibrium scour depth around circular piles using NN Modelling of equilibrium scour depth around circular piles using NN. Technical report, WL — Delft Hydraulics, Delft, 2006.

[4] Andrea Panizzo. Analysis of wave transmission behind low-crested breakwaters using neural networks. *Coastal Engineering*, 54(9):643–656, 2007.

[5] A. Abebe and R. Price. Information Theory and Neural Networks for Managing Uncertainty in Flood Routing. *Journal of Computing in Civil Engineering*, 18(4): 373–380, 2004.

[6] R. J. S. Whitehouse. *Scour at marine structures.* Thomas Telford Publications, London, 1998.

[7] Tim Raaijmakers and Daniel Rudolph. Time-dependent scour development under combined current and waves conditions - laboratory experiments with online monitoring technique. *Deltares*, 2008.

[8] Sumer and Fredsøe. *The mechanics of scour in the marine environment.* World Scientific Publishing Co. Pte. Ltd., 2002. doi: ISBN981-02-4930-6.

[9] T.C. Raaijmakers, T. Joon, M.L.A. Segeren, and P. Meijers. SCOUR : TO PROTECT OR NOT TO PROTECT , THAT S THE QUESTION ! Technical report, Deltares, TU Delft, Delft, 2013.

[10] Tim Raaijmakers. Scour & scour protection in the marine environment. In *Lecture "Bed, bank and shore protection"*, number december, pages 1–41, Delft, 2014. Deltares.

[11] N.M. Angus and R.L. Moore. Scour repair methods in the Southern North Sea. *Proceedings of the 14th Annual Offshore Technology Conference, Houston, Texas*, (4410):385–399, 1982. as cited in Sumer and Fredsœ(2002).

[12] Gerrit J. Schiereck. *Introduction to Bed, bank and shore protection.* VSSD, DUP Blue Print, Delft, 2001.

[13] Raudviki A.J. Melville, B.W. Flow Characteristics in local Scour at Bridge Piers. *Journal of Hydraulic Research*, 15(4):373–380, 1977. as cited in Schiereck (2001).

[14] B. M. Sumer, J. Freds{ø}e, and N. Christiansen. Scour around vertical pile in waves. *Journal of Waterway, Port , Coastal and Ocean Engineering*, 118(1):15–31, 1992.

[15] Benoît Camenen, Magnus Larson, and Atilla Bayram. Equivalent roughness height for plane bed under oscillatory flow. *Estuarine, Coastal and Shelf Science*, 81:409–422, 2009. ISSN 02727714. doi: 10.1016/j.ecss.2008.11.019.

[16] Albert Molinas. Bridge Scour in Nonuniform Sediment Mixtures and in Cohesive Materials : Synthesis report. FHWA-RD-03-083. Technical report, Office of Infrastructure Research and Development, Federal Highway Administration, 2003.

[17] J Guo, O Suaznabar, H Shan, and J Shen. Pier Scour in Clear-Water Conditions with Non-Uniform Bed Materials. *FHWA-HRT-12-022*, pages 1–62, 2012. URL http://www.fhwa.dot.gov/publications/research/infrastructure/hydraulics/12022/12022.pdf$\delimiter"026E30F$nhttp://trid.trb.org/view.aspx?id=1143991.

[18] Nicollet G. Shen H.W. Breusers, H.N.C. Local scour around cylindrical piers. *Journal of hydraulic research, IAHR*, 15(3):211–252, 1977.

[19] Harrison L.J. Richardson J.R. Richardson, E.V. and S.R. Davis. Evaluating scour at bridges, 2d ed.: U.S. Department of Transportation, Federal Highway Administration. *Hydraulic Engineering Circular 18, Publication FHWA-IP- 90-017*, page 132, 1993.

[20] L.a. Arneson, L.W. Zevenbergen, P.F. Lagasse, and P.E. Clopper. Evaluating Scour at Bridges Fifth Edition. (18):340, 2012. URL http://www.fhwa.dot.gov/engineering/hydraulics/pubs/hif12003.pdf.

[21] H. Md. Azamathulla, Aminuddin Ab Ghani, Nor Azazi Zakaria, and Aytac Guven. Genetic Programming to Predict Bridge Pier Scour. *Journal of Hydraulic Engineering*, 136(March):165–169, 2010. ISSN 0733-9429. doi: 10.1061/(ASCE)HY.1943-7900.0000133.

[22] D. Max Sheppard and William Miller. Live-Bed Local Pier Scour Experiments. *Journal of Hydraulic Engineering*, 132(7):635–642, 2006. ISSN 0733-9429. doi: 10.1061/(ASCE)0733-9429(2006)132:7(635).

[23] Bos K.J. Rudolph, D. Scour around a monopile under combined wave-current conditions and low KC-numbers. *Third International Conference on Scour and Erosion Amsterdam*, 2006.

[24] R. V. Raikar and S. Dey. Clear-water Scour at Bridge Piers in Fine and Medium Gravel Beds. *Canadian Journal of Civil Engineering*, 32(4):775–781, 2005. as cited in Jalali (2014).

[25] B. M. Sumer and J. {Fredsøe}. L.v.c.c. *JOURNAL OF WATERWAY, PORT, COASTAL, AND OCEAN ENGINEERING*, 127(May/June):125–134, 2001.

[26] D. Max Sheppard, Mufeed Odeh, and Tom Glasser. Large Scale Clear-Water Local Pier Scour Experiments. *Journal of Hydraulic Engineering*, 130(10):957–963, 2004. ISSN 0733-9429. doi: 10.1061/(ASCE)0733-9429(2004)130:10(957).

[27] Teng B. Li L.: Zhao, M. Local Scour around a Large-Scale Vertical Circular Cylinder due to Combined Wave-Current Action. *Journal of Hydrodynamics (China)*, 1(B): 7–16, 2004. as cited in Haddorp (2005).

[28] Ming Zhao, Liang Cheng, and Zhipeng Zang. Experimental and numerical investigation of local scour around a submerged vertical circular cylinder in steady currents. *Coastal Engineering*, 57(8):709–721, 2010. ISSN 03783839. doi: 10.1016/j.coastaleng.2010.03.002. URL http://dx.doi.org/10.1016/j.coastaleng.2010.03.002.

[29] M. M. Rienecker and J. D. Fenton. A Fourier approximation method for steady water waves. *Journal of Fluid Mechanics*, 104:119, 1981. ISSN 0022-1120. doi: 10.1017/S0022112081002851.

[30] DNV. Modelling and Analysis of Marine Operations, DNV-RP-H103. Technical Report APRIL 2011, Det Norske Veritas, 2011.

[31] M. J. Tucker. Waves in Ocean Engineering, 1999. URL http://www.orcina.com/SoftwareProducts/OrcaFlex/Documentation/Help/Content/html/ReferencesandLinks.htm#Tucker1991. Accessed: 01-06-2015.

[32] Yee-meng Chiew. Discussion: Mechanics of Riprap Failure at Bridge Piers. *Journal of Hydraulic Engineering*, 123(September):481–483, 1997. ISSN 0733-9429. doi: 10.1061/(ASCE)0733-9429(1997)123:5(481).

[33] Coleman S. E. Melville, B. W. *Bridge Scour*. Water Resources Publication, 2000. Jalali (2014).

[34] Dm Sheppard, B Melville, and H Demir. Evaluation of existing equations for local scour at bridge piers. *Journal of Hydraulic Engineering*, (1994):14–23, 2013. ISSN 0733-9429. doi: 10.1061/(ASCE)HY.1943-7900.0000800. URL http://ascelibrary.org/doi/abs/10.1061/(ASCE)HY.1943-7900.0000800.

[35] Giuseppe Oliveto and Willi H. Hager. Temporal Evolution of Clear-Water Pier and Abutment Scour. *Journal of Hydraulic Engineering*, 128(September):811–820, 2002. ISSN 0733-9429. doi: 10.1061/(ASCE)0733-9429(2002)128:9(811).

[36] Albert Molinas and Mohammed I. Abdou. Effect of Sediment Gradation and Cohesion on Bridge Scour, Vol. 1. Effect of Sediment Gradation and Coarse Material Fraction on Clear Water Scour Around Bridges. FHWA-RD-99-183. Technical report, Office of Infrastructure Research and Development, Federal Highway Administration, Fort Collins, Colorado, 1999.

[37] Aytac Guven, H. Md Azamathulla, and N. a. Zakaria. Linear genetic programming for prediction of circular pile scour. *Ocean Engineering*, 36(12-13):985–991, 2009. ISSN 00298018. doi: 10.1016/j.oceaneng.2009.05.010. URL http://dx.doi.org/10.1016/j.oceaneng.2009.05.010.

[38] A. R. Kambekar and M. C. Deo. Estimation of pile group scour using neural networks. *Applied Ocean Research*, 25:225–234, 2003. ISSN 01411187. doi: 10.1016/j.apor.2003.06.001.

[39] D-S Jeng, S. M. Bateni, and E Lockett. Department of Civil Engineering Sydney NSW 2006 Environmental Fluids / Wind Group Neural Network assessment for scour depth around bridge piers. (November 2005), 2006.

[40] M. K. Ayoubloo, a. Etemad-Shahidi, and J. Mahjoobi. Evaluation of regular wave scour around a circular pile using data mining approaches. *Applied Ocean Research*, 32(1):34–39, 2010. ISSN 01411187. doi: 10.1016/j.apor.2010.05.003. URL http://dx.doi.org/10.1016/j.apor.2010.05.003.

[41] J.J. Fernandez. The GP Tutorial, 2013. URL http://www.geneticprogramming.com/Tutorial/. Accessed: 05-03-2015.

[42] R. Poli, W. B. Langdon, and N. F. McPhee. *A Field Guide to Genetic Programing (Summary for Wyvern)*. Number March. Published via http://lulu.com and freely available at http://www.gp-field-guide.org.uk, 2008. (With contributions by J. R. Koza)., 2008. ISBN 9781409200734. URL http://discovery.ucl.ac.uk/1327678/.

[43] J. R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA: MIT Press, 1992.

[44] V. Babovic M. Keijzer. Genetic Programming Tool (c), 1999. Version 0.4 (final).

[45] B. Mutlu Sumer, Figen Hatipoglu, and Jø rgen Fredsø e. Wave Scour around a Pile in Sand, Medium Dense, and Dense Silt. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 133(February):14–27, 2007. ISSN 0733-950X. doi: 10.1061/(ASCE)0733-950X(2007)133:1(14).

[46] A.W. Nielsen and B.M. Sumer. Experimental Study on the Scour around a Mono Pile in Breaking Waves. *Journal of Waterway, Port ...*, (December):501–506, 2012. doi: 10.1061/(ASCE)WW.1943-5460.0000148. URL http://ascelibrary.org/doi/pdf/10.1061/(ASCE)WW.1943-5460.0000148.

[47] J. Vanderplas. AstroML, 2012. URL http://www.astroml.org/book_figures/appendix/fig_neural_network.html. Accessed: 01-03-2015.

[48] Dynamic, Absolute and Kinematic Viscosity, howpublished = http://www.engineeringtoolbox.com/dynamic-absolute-kinematic-viscosity-d_412.html, note = Accessed: 01-03-2105.

[49] M Hoffmans. Ontgrondingen rondom brugpijlers en aan de kop van kribben. *W-DWW-94-312*, 1995.

[50] S. K. Jalali. Prediction of Clear-water Local Scour at Bridge Piers. *M.Sc. Thesis in Civil Engineering for Risk Mitigation, Politecnico di Milano*, (July), 2014.

[51] Stephen T. Benedict Caldwell and Andral W. *A Pier-Scour Database: 2,427 Field and Laboratory Measurements of Pier Scour*. U.S. Geological Survey Data Series 845. Reston, Virginia, 2014.

[52] Jonsson. Wave boundary layers and friction factors. *Coastal Engineering*, 1:127–148, 1966.

Appendix A

Database Parameters

This appendix contains a list of formulas and definitions of all parameters used in the course of this research.

Structural Parameters

Relative obstruction height $[h_p/h_w]$
The relative obstruction height is defined as the ratio of the pile height and the water depth.

Shallowness $[D/h_w]$
Shallowness is the ratio between pile diameter and water depth.

Structure shape
Piles in the marine have a wide range of shapes (e.g. circular, quadrangular, triangular). A rectangular shaped structure will have more flow separation than a smooth circular structure and therefore more scour. However, the scope of this research is limited to circular shapes only (i.e. monopiles), so the structure shape parameter will be a constant.

Fluid Parameters

Dynamic viscosity $[\mu]$
Dynamic viscosity is a measure of internal resistance of a fluid. The value of the dynamic viscosity of water at 20°C is assumed for all data (i.e. $\mu = 1 \cdot 10^{-3}\ kg/ms$).

Kinematic viscosity $[\nu]$
Kinematic viscosity is the ratio of dynamic viscosity to density. Kinematic viscosity can be obtained by dividing the absolute viscosity of a fluid with the fluid mass density.[48] For water at 20°C this value is $\nu = 1 \cdot 10^{-6}\ m^2/s$.

Environmental Parameters

Bed shear stress $[\tau_c]$, $[\tau_w]$,$[\tau_{cw}]$,$[\tau_m]$
Bed shear stress is the frictional force exerted by the flow on the bed per unit area of bed. The corresponding threshold value for sediment motion is called τ_{cr}.[6] The bed shear stress for current, waves and combined waves and current are given by equation (A.1), equation (A.2) and equation (A.3). Equation (A.4) gives the resulting mean bed shear stress.[6]

$$\tau_c = \rho_w \cdot C_D \cdot {u_c}^2 \tag{A.1}$$

$$\tau_w = 0.5 \cdot f_w \cdot \rho_w \cdot {u_w}^2 \tag{A.2}$$

$$\tau_{cw} = \left({\tau_w}^2 + {\tau_m}^2 + 2\tau_w \tau_m \cos\alpha\right)^{1/2} \tag{A.3}$$

$$\tau_m = \tau_c \left(1 + 1.2 \cdot \left(\frac{\tau_w}{\tau_c + \tau_w}\right)^{3.2}\right) \tag{A.4}$$

Critical Shields parameter $[\theta_{cr}]$
The critical Shields number is dependent on the sedimentological diameter. For the following two ranges it is defined by:[6]

$d_* < 10$

$$\theta_{cr} = \frac{0.3}{1 + 1.2 \cdot d_*} + 0.055 \cdot (1 - \exp(-0.02 \cdot d_*)) \tag{A.5}$$

$d_* > 10$

$$\theta_{cr} = \frac{0.24}{d_*} + 0.055 \cdot (1 - \exp(-0.02 \cdot d_*)) \tag{A.6}$$

Fall velocity of individual grains $[w_s]$
The particle fall velocity is the rate at which suspended solids subside and are deposited. As defined by Soulsby and reported by Whitehouse.[6]

$$w_s = \frac{\nu}{d_{50}}[(10.36^2 + 1.049{d_*}^3)^{0.5} - 10.36] \tag{A.7}$$

The dimensionless particle fall velocity is given by:

$${w_s}^* = \left[\frac{(s-1))^2}{g\nu}\right]^{1/3} w_s \tag{A.8}$$

Friction factor waves $[f_w]$
The friction factor of the waves is based on the theorem of Soulsby [6]:

$$f_w = 0.27 \left(\frac{A_w}{k_r} \right)^{-0.52} \tag{A.9}$$

Mean grain size; bed $[d_{50}]$
Mean grain size is the median diameter of individual grains of sediment.

Sediment gradation $[\sigma]$

The sediment gradation is defined as by as: [16]

$$\sigma = \sqrt{\frac{d_{84}}{d_{16}}} \tag{A.10}$$

Relative critical velocity $[u_{crit}]$
The relative critical velocity is a parameter that has the same behavior as the relative mobility. The formula for u_{crit} is given by the following equation:

$$u_{crit} = \frac{u_c}{u_{c,cr}} \tag{A.11}$$

The critical velocity $u_{c,cr}$ is estimated with the formula from Oliveto and Hager [35]

$$\begin{aligned} u_{c,cr} &= \frac{2.33}{f_{pk}} d^{*-1/4} \sigma^{1/3} (9.81 \Delta d_{50})^{1/2} \left(\frac{h_w}{d_{50}} \right)^{1/6} && d^* \leq 10 \\ u_{c,cr} &= \frac{1.08}{f_{pk}} d^{*1/12} \sigma^{1/3} (9.81 \Delta d_{50})^{1/2} \left(\frac{h_w}{d_{50}} \right)^{1/6} && 10 < d^* \leq 150 \\ u_{c,cr} &= \frac{1.65}{f_{pk}} \sigma^{1/3} (9.81 \Delta d_{50})^{1/2} \left(\frac{h_w}{d_{50}} \right)^{1/6} && d^* > 150 \end{aligned} \tag{A.12}$$

Relative density $[\Delta]$
The relative density is given by:

$$\Delta = \frac{\rho_s - \rho_w}{\rho_w} = s - 1 \tag{A.13}$$

Relative mobility $[mob]$
The value that relates the Shields parameter to the critical Shields parameter is called the relative mobility. It gives the threshold of motion for the sediment.

$$mob = \frac{\theta}{\theta_{cr}} \tag{A.14}$$

Relative suspension number $[SB]$
Particles will come into suspension if the bed shear velocity is larger than the particle fall velocity. The following definition for the relative suspension number is used:

$$SB = \frac{1}{w_s}\frac{\sqrt{\tau_{cw}}}{\rho_w} \tag{A.15}$$

Roughness current-related $[kr]$
The seafloor roughness is estimated with the following equation from Camenen et al. [15]:

$$\frac{kr}{d_{50}} = 0.6 + 2.4\left(\frac{\theta}{\theta_{cr}}\right)^{1.7} \tag{A.16}$$

Sediment coarseness $[D/d_{50}]$
The sediment coarseness is a dimensionless representation of sediment size. It is defined as the ratio of the pile diameter and the mean grain size.

Sediment number $[N_s]$
The sediment number is defined with equation (A.17)

$$N_s = \frac{u_c}{\sqrt{g\Delta d_{50}}} \tag{A.17}$$

Specific gravity $[s]$
Specific gravity is the ratio of the density of a substance to the density of a given reference material, in this case the ratio of sediment density versus water density equation (A.18).

$$s = \frac{\rho_s}{\rho_w} \tag{A.18}$$

Sedimentological diameter $[d_*]$
The sedimentological diameter is a dimensionless representation of sediment size. It is defined by equation equation (A.19).

$$d_* = d_{50} \cdot \left(\frac{\Delta g}{\nu^2}\right)^{1/3} \tag{A.19}$$

Shields parameter $[\theta]$
The Shields parameter is used to determine the threshold of motion, i.e. if life-bed or clear-water scour occurs. The definition of Soulsby is used: [6]

$$\theta = \frac{\tau_{cw}}{(\rho_s - \rho_w)gd_{50}} \tag{A.20}$$

Suspension criterion; bed load or suspended load
Suspension is the state where solid and fluid are mixed and not easily separated. This criterion depends on relative suspension number.
If $SB > 1$: suspended load
If $SB < 1$: bed load

Hydrodynamic Parameters

Cylinder Reynolds number $[Re_D]$
The Reynolds number is a dimensionless quantity that is used to predict flow patterns, e.g. it can be an indication of the amount of turbulent flow that can be expected. The Reynolds number is determined for current only (equation (A.21)) and wave only (equation (A.22)) situations.

$$Re_{D,c} = \frac{u_c D}{\nu} \tag{A.21}$$

$$Re_{D,w} = \frac{u_w D}{\nu} \tag{A.22}$$

Diffraction parameter $[D/L]$
Wave diffraction occurs when a wave encounters an obstacle or slit. The diffraction parameter is defined as the ratio of the pile diameter and the wave length.

Froude number $[Fr]$
The Froude number is a dimensionless number defined as the ratio of a characteristic velocity to a gravitational wave velocity. It may equivalently be defined as the ratio of a bodies inertia to gravitational forces.

$$Fr = \frac{u_c}{\sqrt{g \cdot h_w}} \tag{A.23}$$

The wave Froude number is given by:

$$Fr_w = \frac{L}{T_p \cdot \sqrt{g h_w}} \tag{A.24}$$

Keulegan-Carpenter number $[KC]$
The Keulegan-Carpenter number relates the orbital velocity of a wave to the diameter of a pile. For steady currents a value of KC='999' is assumed.

$$KC = \frac{u_w T_p}{D} \tag{A.25}$$

Peak wave frequency $[f_p]$
The peak wave frequency is given by the following equation:

$$f_p = \frac{1}{T_p} \tag{A.26}$$

Peak wave period $[T_p]$
The peak wave period is the wave period with the highest energy. As a rule of thumb the following relation can be used: $Tp \approx 5.3{H_s}^{1/2}$. However, in all wave related entries in the database, this value is given by the source of the experiment.

Relative velocity $[u_{rel}]$
The relative velocity is a dimensionless number relating the depth-averaged flow velocity of the current with the wave orbital velocity.

$$u_{rel} = \frac{u_c}{u_c + u_w} \tag{A.27}$$

Significant wave height $[H_s]$
The mean of the highest third of the waves or four times the standard deviation of the surface elevation.

Wave orbital velocity $[u_w]$
The near-bed orbital velocity according to linear wave theory is:

$$u_w = \frac{\pi H_s}{T_p \sinh\left(\frac{2\pi h_w}{L}\right)} \tag{A.28}$$

Wave spectrum
Three types of wave spectra are considered: JONSWAP, monochromatic and no waves. In case of JONSWAP there is a distribution in the wave spectrum, in case of monochromatic regular waves there is one single peak in the spectrum, in case of no waves there is no wave spectrum.

Appendix B

Database sources

This appendix gives an overview of the sources in the database.

- 'catg' is the type of condition under which the tests were performed, i.e. current-only [C], wave-only [W] or combined waves and current [CW]. Some sources have tested multiple conditions.
- 'Source' gives the location where the data was found. Some tests were part of a large collective and of others the original paper was found. The larger collectives are indicated with [1] to [4], in which [1] is gathered by Hoffmans [49], [2] by Jeng et al. [39], [3] by Jalali [50] and [4] by Caldwell and W. [51].
- 'Nr' is the amount of tests per source
- 'ϵ' the error predicted by the scour prediction formula created in this study
- 'W' the weight of the source, in which 1 is a very good source and 3 a more questionable one

catg	Author	Source	Nr.	ϵ	W	Remarks
C	'Azzaroli, D. (1983)'	[3]	1	0.81	3	Small sediment
C	Baker (1979, 1980)'	[1]	115	0.22	2	
C	'Batuca and Dargahi (1986)'	[1]	50	0.92	2	
C	'Beg (2013)'	Original	7	0.13	1	Long duration, contains comparison with 14 other prediction methods
C	'Bonousandas (1973)'	[1]	74	0.95	2	Unsure if equilibrium scour depth is reached
C	'Breusers (1971)'	[1]	21	1.34	2	
C	'Chabert and Engeldinger (1956)'	[2], [1], [3], [4]	261	0.27	2	Small channel width
C	'Chang, W.,Y., et al, (2004)'	Original, [3]	10	0.17	1	Measures influence sediment gradation
C	'Chee (1982)'	[2], [1], [4]	84	0.16	2	
C	'Chen (1980)'	[1]	39	1.35	2	
C	'Chiew (1984)'	[2], [1], [3], [4]	248	0.20	1	
C	'Coleman (unpublished)'	[4]	6	0.45	3	
CW, C	'DeSonneville et all (2010)'	Original	3	0.15	1	Scour reduction by collars around offshore monopiles
W	'Dey (2006)'	Original	41	0.07	1	Wave spitter plate and cable around pile
C	'Dey and others (1995)'	[3], [4]	18	0.13	2	Very shallow waters
CW, C	'Eadie (1986)'	Original	20	0.14	2	Combined current and wave test
C	'Ettema (1976/1980)'	[1], [4]	43	1.10	2	
C	'Ettema (1980)'	[2], [3], [4]	106	0.30	1	
C	'Ettema Kirkil Muste (2006)'	[3], [4]	6	0.33	1	
C	'Franzetti, S. (1989)'	[3]	1	0.34	2	
C	'Graf (1995)'	[2], [3], [4]	3	0.12	1	
C	'Grimaldi (2005)'	Lanca (2013)	3	0.30	2	
C	Guo (2012)'	Original	10	0.25	1	Measures influence sediment gradation
C	'Hancu (1965/1971)'	[2], [1]	62	0.25	2	
C	'Jain and Fischer (1979)'	[2], [1], [4]	34	0.24	2	Reported scour value above 3 m
W,CW, C	'Jensen et al (2006)'	Original	45	0.13	1	Reported $d_{50} = 17mm$, breaking waves and tidal currents
C	'Jones (unpublished)'	[4]	17	1.17	3	

Table B.1: *Sourcelist*

catg	Author	Source	Nr.	ϵ	W	Remarks
C	'Knight (1975)'	[1]	35	1.14	2	
C	'Kothyari (1989)'	[1]	21	0.17	2	Reported scour value above 3 m
C	'Kothyari et al (1992)'	[2]	73	1.01	2	D and h_w are mixed up, reports also temporal variation
C	'Kwan (1984)'	[1]	9	1.94	2	
C	'Lanca et al. (2013)'	Original, [3]	38	0.36	1	All around mob = 1, wall influence, large piles
C	'Lee, S.L. and Sturm, W. (2009)'	[3]	4	0.75	2	
C	'Maza Alvarez (1966)'	[1]	25	0.99	3	
C	'Melville (1997)'	[2], [4]	17	0.33	1	
C	'Melville and Chiew (1999)'	Original, [3], [4]	27	0.33	1	
C	'Mignosa, P. (1979/1980)'	[3]	13	0.50	2	Low sediment density
C	Molinas (2003)'	Original	184	0.17	1	3 piles behind each other , gravel experiments
W, CW, C	'Mostafa(2011)'	Original	8	0.16	1	
W	'Nielsen (2011)'	Original	25	0.11	2	Breaking waves, slope
C	'Oliveto and Hager (1999)'	[3]	45	0.31	1	Low sediment density
C	'Oliveto and Hager (2000)'	[3]	34	0.20	1	
C	'Oliveto and Hager (2001)'	[3]	6	0.14	1	
C	'Oliveto and Hager (2002)'	[3]	3	0.25	1	
C	'Oliveto (2002)'	Original, [2]	46	0.34	1	
CW, C	Qi (2013)'		28	0.29	1	Wall influence, really small scour depth, only 100min duration
C	'Raikar, V. R., Dey, S. (2005)'	[3]	20	0.17	1	Extremely large sediment
W	'Samaneh (2014)'		2	0.17	1	
C	'Shen and others (1969)'	[1], [4]	23	0.14	2	Different values hoffmans and usgs
C	'Sheppard and Miller (2006)'	Original, [3], [4]	24	0.14	1	
C	'Sheppard and others (2004)'	Original, [3], [4]	14	0.21	1	wall influence, large piles
C	'Simmaro, G. (from Lanca (2013))'	Lanca (2013)	5	0.74	2	
C	'Subdey(?)'	[1]	18	1.07	2	
W, C	'Sumer (1992)'		53	1.25	1	
W,CW, C	'Sumer (2001a)'	Inhouse	27	0.19	2	

Table B.2: *Sourcelist (continued)*

catg	Author	Source	Nr.	ϵ	W	Remarks
W	'Sumer (2001b)'		16	0.04	1	wall influence, large piles
W	'Sumer (2007)'	-	48	0.15	2	Dense silt
W, CW, C	'Sumer (2013)'		83	0.14	1	irregular waves
CW, C	'Thompson (2006)'		4	0.53	1	Backfilling, initial scour and global scour also present (and accounted for), not sure if equilibrium is reached
C	'unknown'	[2]	10	0.24	3	
C	'unknown (Kandasamy?)'	[2]	1	0.34	3	
C	'unknown (Kwan?)'	[2]	1	1.49	3	
C	'unknown (Melville?)'	[2]	2	1.18	3	
CW, C	'Whitehouse et al (2006)'	Inhouse	5	0.44	2	
CW	'WL; Ballast Nedam'	Inhouse	4	0.09	1	
W, CW, C	'WL; R&D 2004'	Inhouse	80	0.18	1	
CW	'WL; R&D 2006'	Inhouse	4	0.07	1	
W, CW, C	'WL; R&D-2007'	Inhouse	25	0.34	1	
C	'Yanmaz and Altinbilek (1991)'	[2], [1], [3], [4]	33	0.17	1	
C	'Zhao (2010)'	-	28	0.64	1	3 piles next to each other, submerged piles
CW	'Zhao, Teng, Li (2004)'	Haddorp (2005)	13	0.04	1	large piles

Table B.3: *Sourcelist (continued)*

Appendix C

Database data

This appendix contains an additional overview of the data available in the database. This data is depicted in dimensional and dimensionless plots.

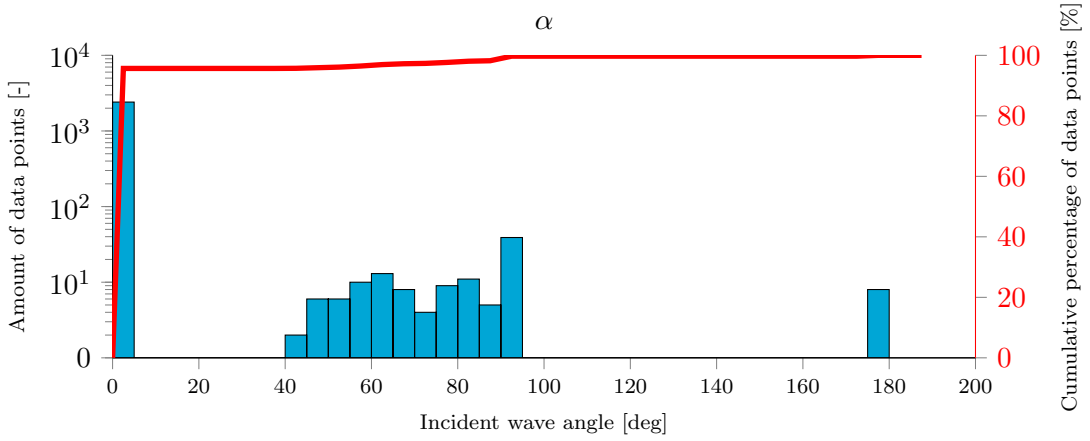


Figure C.1: *Data analysis: range of incident wave angles (α) in the database on logarithmic scale*

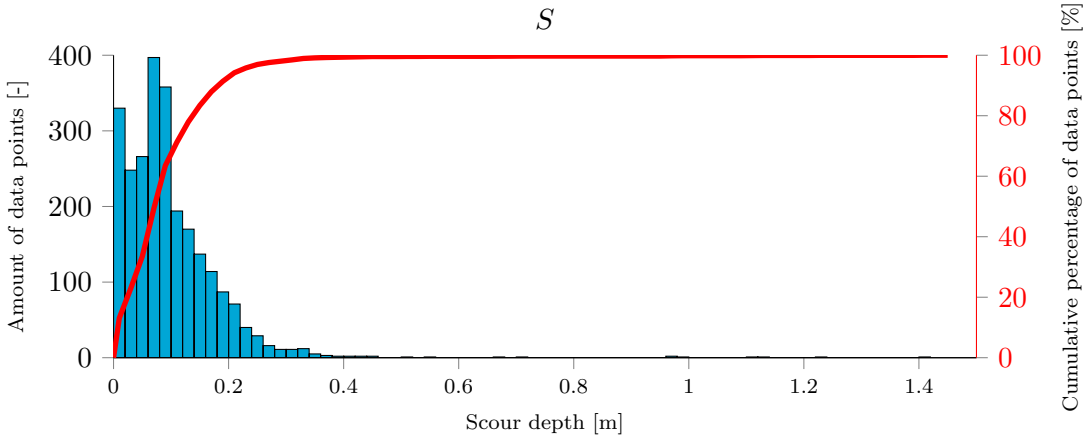


Figure C.2: *Data analysis: range of equilibrium scour depth (S) in the database on linear scale*

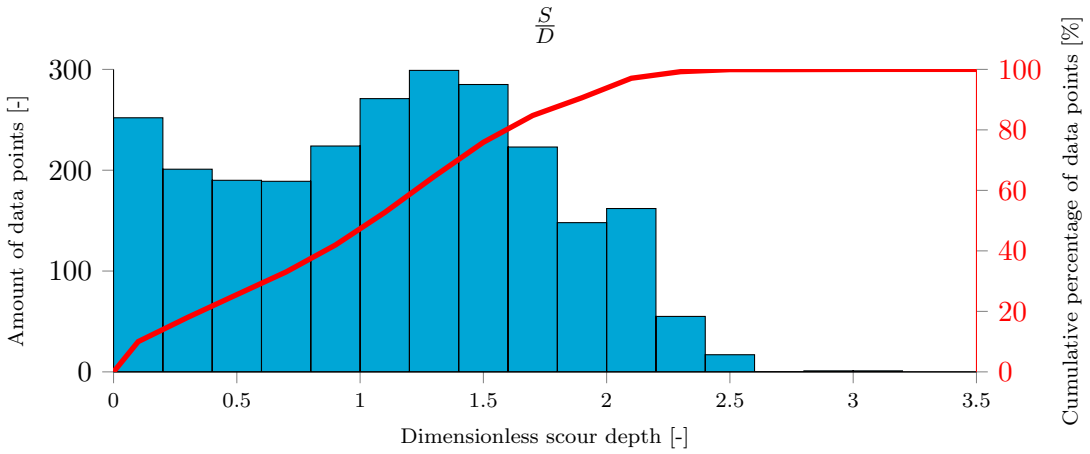


Figure C.3: *Data analysis: range of dimensionless equilibrium scour depth ($\frac{S}{D}$) in the database on linear scale*

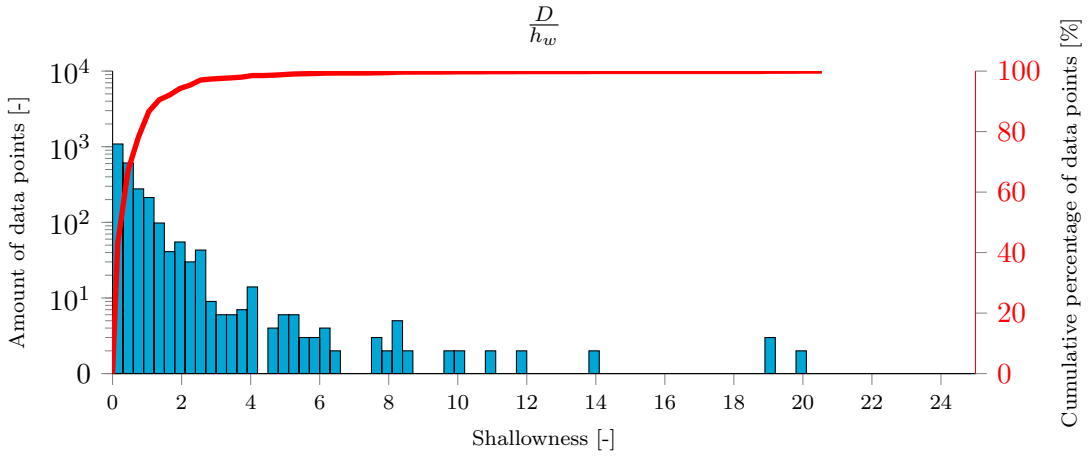


Figure C.4: *Data analysis: range of shallowness ($\frac{D}{h_w}$) in the database on logarithmic scale*

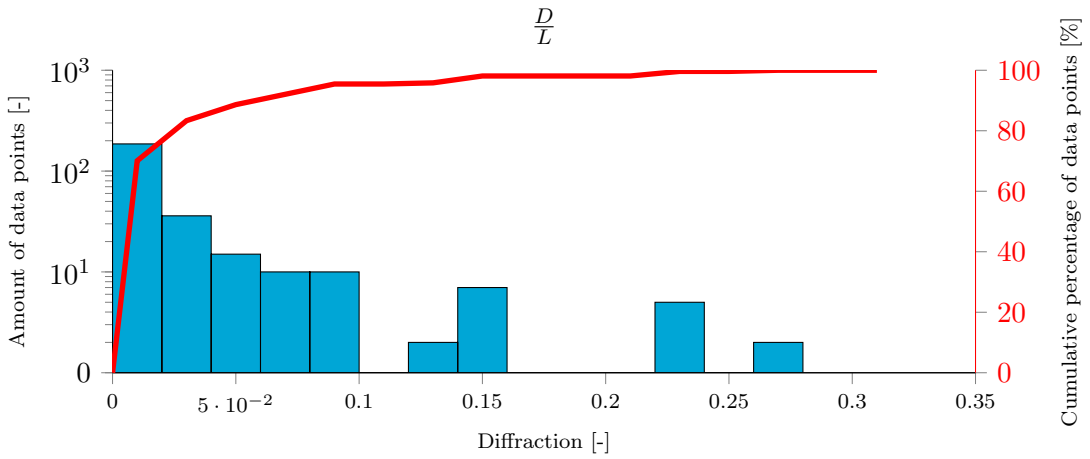


Figure C.5: *Data analysis: range of diffraction ratios ($\frac{D}{L}$) in the database on logarithmic scale*

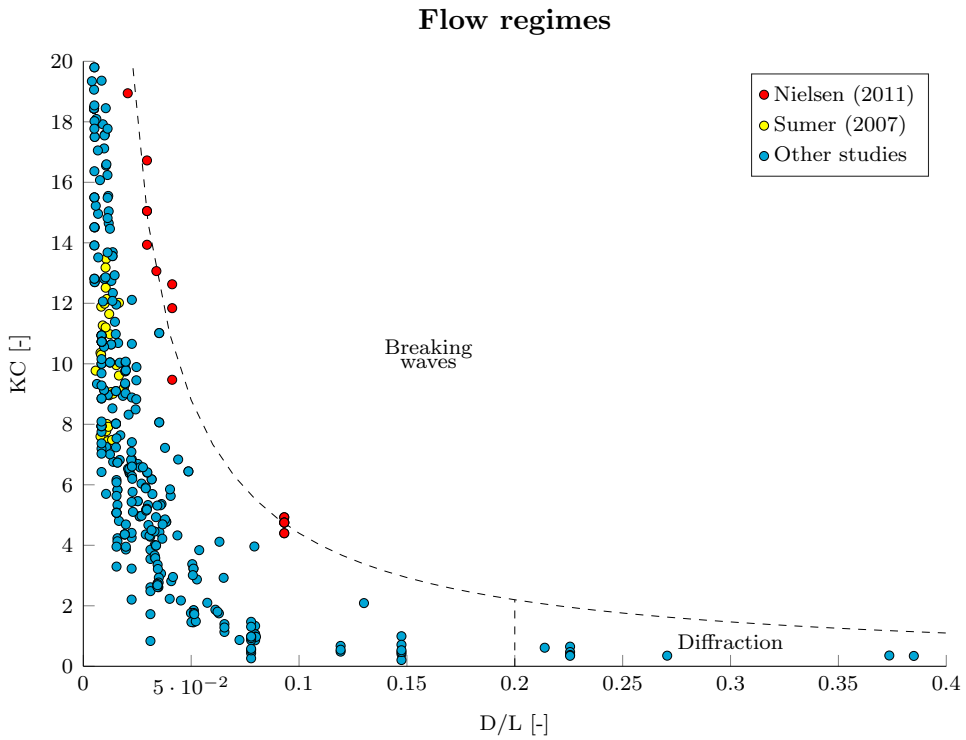


Figure C.6: *Data analysis: range of flow regimes in the database on linear scale, zoomed in*

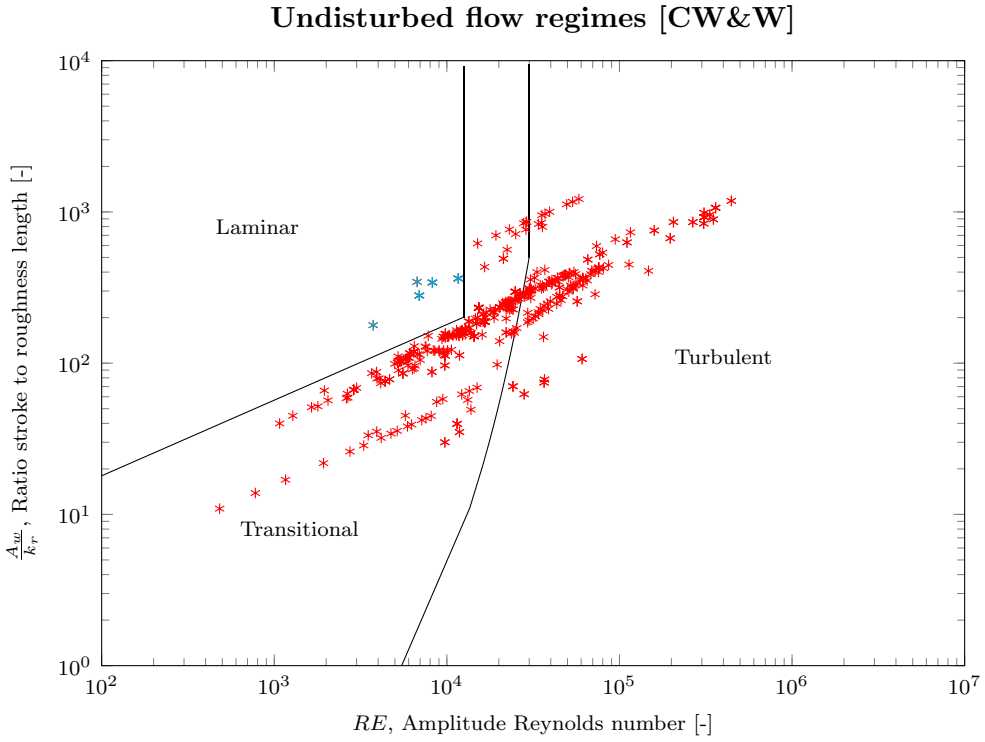


Figure C.7: *Flow regimes based on Jonsson*[52]

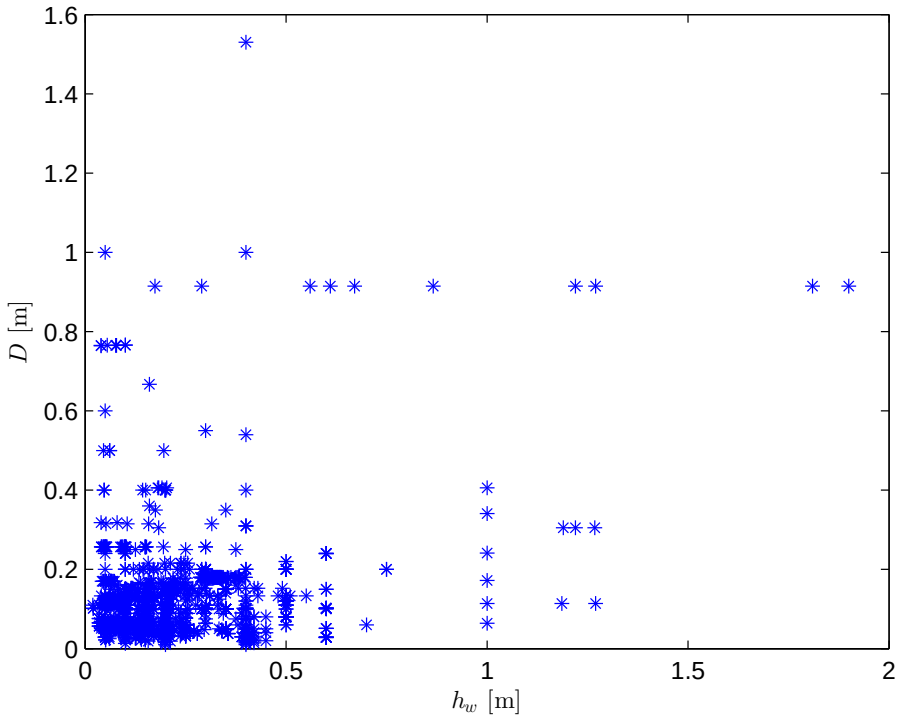


Figure C.8: *Water depth versus pile diameter*

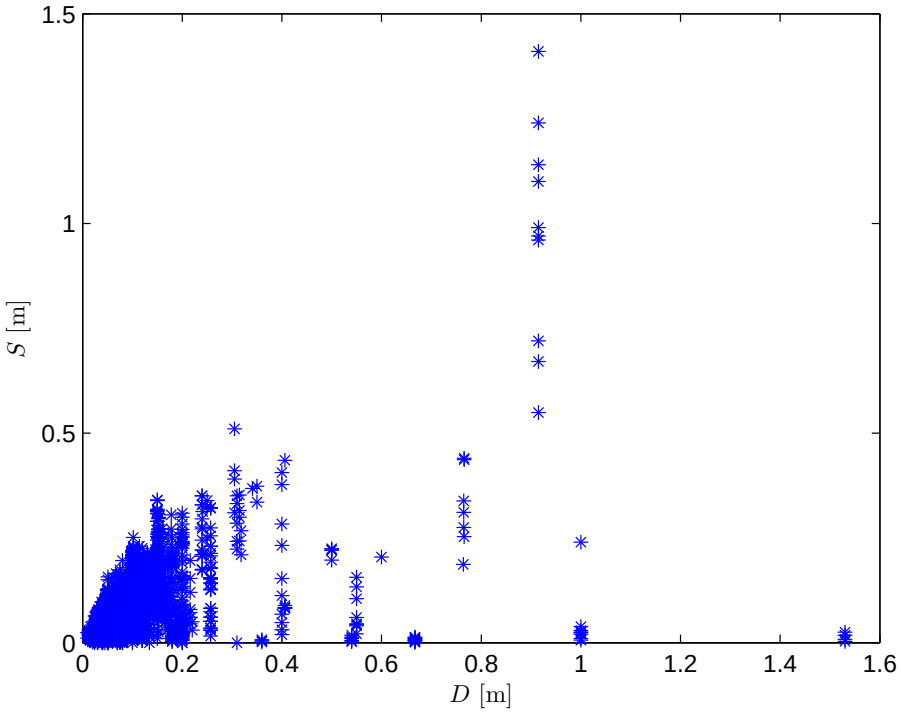


Figure C.9: *Pile diameter versus equilibrium scour depth*

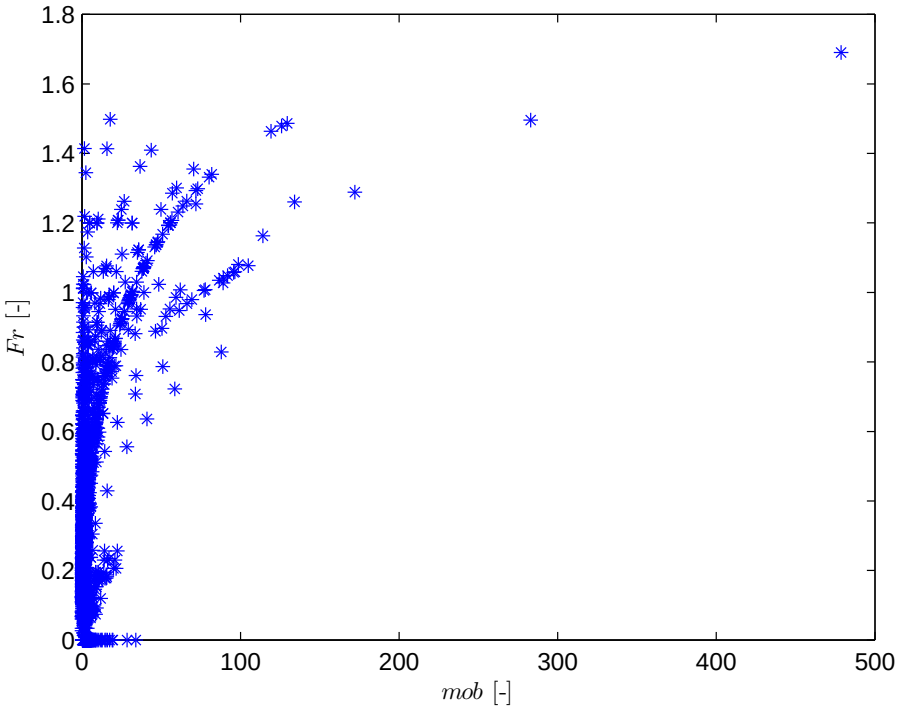


Figure C.10: *Relative mobility versus Froude number*

Appendix D

Data coverage method

The accuracy of a prediction with the proposed equilibrium scour depth formula, depends on how well the conditions of the to-be-predicted datapoint is covered in the database. If similar tests are found, the accuracy of the prediction will be more reliable. In this appendix a simple method is proposed to evaluate if a new set of data is covered by the database.

An example is first given for the [W] formula, which only consists of 2 parameters. Therefore the coverage of these datapoints can be seen as a map, as shown in figure D. The large dot represents the new data point, which will be called datapoint x, with value KC_x and $(D/h_w)_x$.

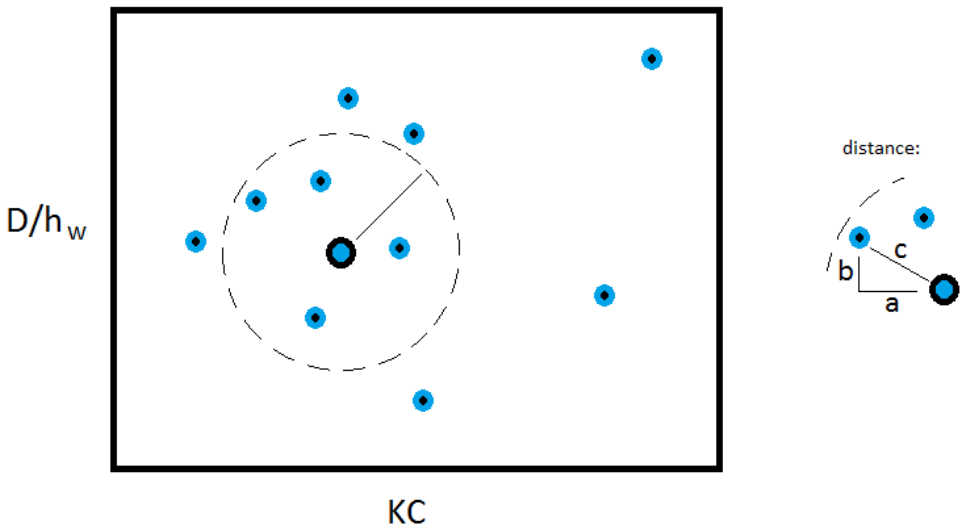


Figure D.1: *Example of data coverage: KC and D/hw*

All data within the circle are considered to be within the range of the datapoint x. The more data in the circle, the better the coverage. To quantify this amount the distance

from each point in the circle to datapoint x will be calculated. This vector can be easily calculated with Pythagoras. However, since D/h_w and KC do not vary on the same scale, this distance needs to be normalized first.

At this point it has to be determined for what distance in terms of KC and D/h_w the values are in reach. For this example it is said that all KC values within 1 distance from KC_x are within range and all data with D/h_w values with 0.2 from $(D/h_w)_x$. So in mathematical terms:

$$v = \sqrt{\left(\frac{KC_x - KC}{1}\right)^2 + \left(\frac{(D/h_w)_x - D/h_w}{0.2}\right)^2} \tag{D.1}$$

Where v is the normalized vector length from datapoint x to a value within the circle. All values larger than 1 can now be discarded, since they are outside the unity circle. The remainder of terms is subtracted from 1 and summed up to a value Q, which represents the quantity of data coverage. This ensures two things: that the value of Q will be high when there is a large amount of data in the circle and that the value of Q will be high when the distance between datapoint x and the database value is zero.

$$Q = \sum (1 - v) \qquad \text{for v} \leq 1 \tag{D.2}$$

For high values of Q the data was covered well, for low values there was no datapoints in the database with similar conditions.

This method is tested for each datapoint in the database, to see how many similar tests there are. The result can be seen in figure D.2, which shows that the method works well. The coverage is ranked in 5 weights, in which 5 is good coverage and 1 is bad coverage. It is seen that the accumulated data at the bottom has a high ranking coverage value, while the more sparse values at the top have low ranking value.

The same principle can be repeated for the formulas with more than two parameters. For instance for the [C] formula, which contains 5 parameters, equation (D.1) can be changed to a 5 dimensional vector length, in the following matter:

$$v = \sqrt{\left(\frac{mob_x - mob}{3}\right)^2 + \left(\frac{Fr_x - Fr}{0.1}\right)^2 + \left(\frac{d^*_x - d^*}{10}\right)^2 + \left(\frac{\sigma_x - \sigma}{0.2}\right)^2 + \left(\frac{(D/h_w)_x - D/h_w}{0.2}\right)^2} \tag{D.3}$$

The result can be seen in figure D.3. This shows that the data that was best covered in the set, is also best predicted.

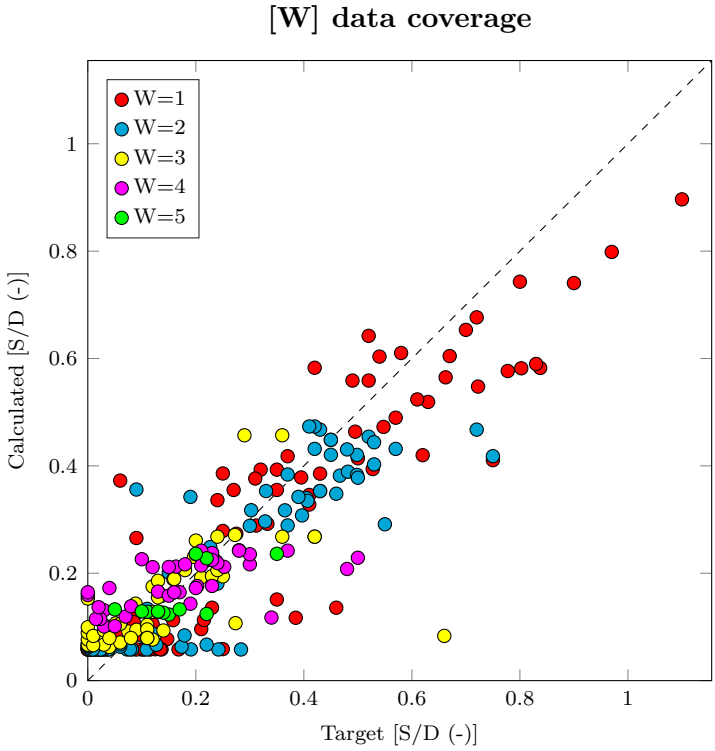


Figure D.2: *Data coverage [W] tests*

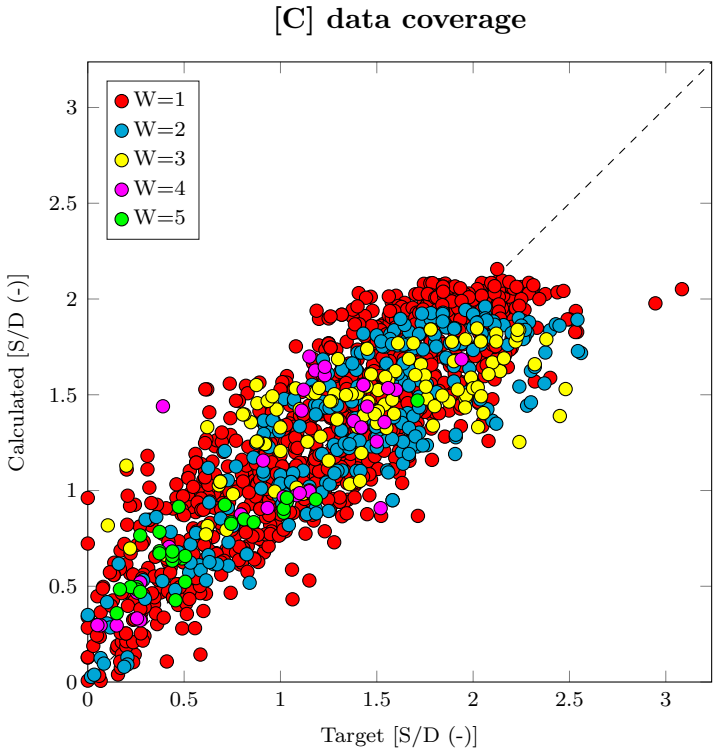


Figure D.3: *Data coverage [C] tests*