The influence of environmental and gully pot specific factors on solid deposition in gully pots

Applying a Generalized Linear Model (GLM) to assess the relationship of environmental and gully pot specific factors on the solid build-up in gully pots

The influence of environmental and gully pot specific factors on solid deposition in gully pots

Applying a Generalized Linear Model (GLM) to assess the relationship of environmental and gully pot specific factors on solid-build up in gully pots

Type document	Bachelor thesis
Date of publication	07-06-2018
Commissioning institution	Delft University of Technology, Department of Watermanagement, Section Sanitary Engineering
Institution	Hogeschool Van Hall Larenstein
Study programme	Land- and Watermanagement
Specialisation	Applied Hydrology
External supervisor	Matthijs Rietveld (TU Delft) M.W.J.Rietveld@tudelft.com
Internal supervisor	Ad Bot ad.bot@hvhl.nl
Author	André Vallendar andi.vallendar@hvhl.nl
Student number	000005163

Preface

Before you lies my bachelor thesis 'The influence of environmental and gully pot specific factors on solid deposition in gully pots', concluding the Bachelor study programme Land- and Water Management, with a specialization in Applied Hydrology. This thesis has been written to fulfil the graduation requirements from Van Hall Larenstein (VHL) University of Applied Sciences. Research has been conducted at Delft University of Technology from February to June 2018.

Together with my external supervisor, Matthijs Rietveld, a research question was formulated to guide this research. Overall, the conducted study was complex due to my unfamiliarity with statistical modelling and the various factors which had to be taken into account through the modelling process. But time, struggle and relentless pursuit to fully grasp the concepts of this research and the model, led me to answering the research question. With these research results, I hope to contribute to the PhD research of Matthijs Rietveld which ultimately aims to improve sewer asset management by cost reduction and the improvement of serviceability.

In September 2014, I embarked on a BSc journey that sparked my passion for resource water and this journey finds its completion with this BSc thesis which, without the contribution of some important and special people would not have come to fruition.

I would like to thank the members of the section Sanitary Engineering at TU Delft (room 4.64) for being so welcoming and open to any questions I had. I would particularly like to express my gratitude to Matthijs Rietveld for sharing his expertise and for always being available when needed. During my graduate internship, he has given me excellent supervision, guidance and encouragement. He listened to any suggestions I had and I am going to miss our great discussions. I am grateful to have had you as my supervisor. I also would like to thank my internal supervisor, Ad Bot, from Van Hall Larenstein University of Applied Sciences for supporting me in the task of completing this BSc thesis and giving me constructive criticism during our feedback sessions to improve my thesis.

Lastly, I want to thank my family Albrecht, Manuka and Saskia Vallendar. Thank you for encouraging me in my pursuits and inspiring me to follow my dreams. I am grateful to my parents, Albrecht and Manuka who have supported me emotionally and financially. I have always known that you believe in me and want what is best for me. Thank you for teaching me all the fundamental values on how to be a good human being. And Saskia, who I have always looked up to as a younger brother, you have always been a great source of inspiration for me and I could not have asked for a better sister. I am very fortunate to have you in my life. You are the best family someone could have asked for and I love you dearly. I also want to thank my grandfather, Thilak Wijesinghe, who was a Civil engineer himself. He sparked my interest in the resource water by taking me to ancient water reservoirs in Sri Lanka and saying the same words over and over again:

'Now, our fellows do not have the genius to manage water like they used to. The one nation which has truly perfected the way of managing water are the Dutch. One must go to the Netherlands to learn water management.'

You were a 'once-in-a-lifetime' person, which shall always be remembered and remain in our hearts. Thank you for being such a great inspiration Attha. I hope you are having a ball up there.

Delft, June 7, 2018

A. Vallendar

André Vallendar

Summary

The combined impacts of urbanization and climate change pose significant threats for flooding and water quality in urban areas *(Miller and Hutchins, 2017)*. A research carried out by *ten Veldhuis and Clemens, 2011* shows that on average 83% of the flooding events were caused by gully pot blockages, stating the crucial importance of gully pot maintenance. Roadside gully pots are an essential asset of the surface water collection infrastructure *(Butler et al., 1995)* by transporting storm water and retaining road surface solids in order to reduce solid deposition in sewer systems, ensuring water treatment plant efficiency and protecting the water quality of receiving water bodies *(Butler and Karunaratne, 1995; Deletic et al., 2000)*. Presently, it is an area of concern that municipalities in the Netherlands do not distinguish between areas depending on their susceptibility to gully pot blockage. This leads to redundant and delayed cleanings. To move away from this inefficient cleaning pattern, insight needs to be gained about the factors influencing solid build-up in gully pots. Due to the scarcity of research on this topic, this research aims to assess the relationship of environmental and gully pot specific factors on the solid deposition in gully pots.

To evaluate the influence of environmental and gully pot specific factors, solid depths needed to be measured in order to develop a statistical model. Therefore seven monitoring streets were selected in The Hague and Rotterdam and sediment bed heights from 409 gully pots in these streets were collected in three to four week intervals, spanning from November 2017 to May 2018. The monitoring streets were selected based on the presence of shops and trees. Knowledge about influential factors was gained through literature research and ultimately a selection of model parameters was made, depending on data availability and the possibility and complexity of quantifying a parameter. This yielded seventeen environmental and gully pot specific parameters for the statistical model, aiming to describe the behaviour of the solid build-up (L/day).

For the assessment of significant factors, a probabilistic Generalized Linear Model (GLM) was used and it was chosen, based on the normally distributed data of the solid build-up, to use a normal distribution and the identity link in the GLM. The significance of factors was assessed by a significance threshold of 0.05, whereby insignificant parameters (p-value > 0.05) were stepwise eliminated from the model. The final model results yielded that the following parameters influence the solid build-up in gully pots: Tree number 2 Leaf abscission, Tree number 2 Leafless and Tree number 2 Leaf growth, top and combination inlet, effective sand trap volume and rainfall volume. The tree numbers (TN) represent the influence of trees with seasonal variations on the solid deposition.

Results show, that TN2 leaf abscission yielded the highest contribution to solid deposition amongst all other tree factors. This coincides with findings from *Chen et al., 2017* who found a large amount of complaints (>50%) regarding gully pot blockage in fall, overlapping with the leaf abscission phase of trees. Furthermore the positive β value of the effective sand trap volume indicates that increasing volume also increases the solid deposition, indicating that more extensive deposition takes place due to longer retention times in larger sand traps. Also the rainfall was found to be a significant factor, coinciding with findings by *Ellis and Harrop, 1984* who proved that sediment loading is highly correlated with the total rainfall volume ($r = 0.98$). Also gully pots with top and combination seemed to experience more enhanced solid deposition than side inlets. Lastly, the model results showed that the explanatory variables such as shops, rainfall intensity, traffic intensity and road surface were not relevant factors contributing to solid deposition in gully pots. It was yet surprising that the connected surface area, and the rainfall intensity were not considered significant factors based on the model results, although studies from *Ellis and Harrop, 1984, Post, 2016* and *Butler and Karunaratne, 1995* have shown that these factors were found to have a strong to modest correlation with the solid-build in gully pots. The findings from this study can help municipalities to improve maintenance strategies by safeguarding gully pot performance, preventing urban flooding and reducing costs.

Table of contents

1 Introduction

1.1 Context

It cannot be denied: our climate has changed and continues to change drastically. Changes in precipitation patterns and the magnitude of rainfall events are predicted to extensively alter flooding events in urban areas. With lacking mitigation and adaptation measures this can lead to increased flood risks and associated damages in the near future *(Wheater and Evans, 2009)*. Simultaneously the world is undergoing its largest upsurge of urban growth and it is anticipated that 70% of the world's population will be living in urban areas by 2050 *(UN, 2008)*. In the 21st century, urbanization has been one of the major trends *(McDonald et al., 2014)*, leading to uncontrolled urban sprawl, industrial growth and infrastructural changes and development *(UN, 2006)*. The infrastructural changes also comprise an increase in artificial areas, and paired with intensified rainfall, the routing of all storm water runoff by a storm water drainage network can lead to much flashier responses and higher peak flows *(Miller et al., 2014)*, consequently increasing the probability of urban flooding. Flooding and the subsequent spreading of water over adjacent areas can lead to potential health risks *(ten Veldhuis et al., 2011)*, traffic disruption, material damage *(Runhaar et al., 2012)* and economic losses.

The 2011 Administrative Agreement on Water Affairs *(Ministry of Infrastructure and the Environment, 2011)*, which foresaw a total reduction of 450 million euros for the production of drinking water, sewage and wastewater purification, challenged the water sector to improve cost-effectiveness *(Post, 2016)*. Research by *Ashley et al., 2004* has found that blockages are the prevailing cause for sewer service losses. By considering the influence of blockages on flood risk, proactive management approaches to prevent blockages have proven to be cost-effective *(Ashley et al., 2000; ten Veldhuis et al., 2011)*. Nevertheless it remains an area of concern that currently mostly the main sewers are considered for proactive measures, while the remaining system (lateral house connections, lateral connections and gully pots) is still mostly treated by reactive measures *(Post, 2016)*.

1.2 Problem description

In a research carried out by *ten Veldhuis and Clemens, 2011* approximately 83% of urban flooding events were caused by gully pot blockage, emphasizing the crucial importance of gully pot maintenance. Roadside gully pots, also known as a catch basin in North America and Canada *(Osborne et al., 1998)* are the link between surface water runoff and the sewer system *(Bolognesi et al., 2008)*. These street inlets are an essential asset for storm and combined sewers system by being an important and integral part of the surface water collection infrastructure *(Butler et al., 1995)*. Often these street inlets are designed as gully pots, referring to the presence of a sand trap *(Post et al., 2016)* as shown in Figure 1.1. Next to transporting storm water, gully pots are supposed to retain road surface solids in order to reduce solid deposition in sewer systems, ensure water treatment plant efficiency and protect environmental quality of receiving water bodies *(Butler and Karunaratne, 1995; Deletic et al., 2000)*.

Through solid retention in gully pots, silting and deterioration of downstream sewer components is diminished. It can be said that gully pots are the first and last interception point before solids enter the main sewer. This is a crucial task, because pollutants (e.g. heavy metals, hydrocarbons and organic matter) are bounded to retained particles in the gully pot *(Post, 2016)*. A study by *Bryan Ellis and Revitt, 1982* revealed that 70% of the metals are bounded to particles smaller than 100 µm, indicating a link between the highest pollutant load and smaller particle sizes. Depending on the retaining efficiency of the sand trap, the supply of solids induces progressive silting over time *(Post et al., 2016)*. If the sand trap capacity is exceeded, the hydraulic efficiency of the gully pot will progressively decrease, leading to increased urban flood risk during rainfall events *(Silvagni and Volpi, 2002)*.

This can be confirmed by several studies which recognized that gully pot blockage is one of the most prominent reasons leading to events of sewer flooding (*ten Veldhuis et al., 2011*; *Cherqui et al., 2015*).

Presently it is an area of concern that municipalities in the Netherlands do not distinguish between areas depending on their susceptibility to gully pot silting and consequently flooding. This leads to redundant and mostly delayed cleanings, subsequently leading to increased costs and a higher probability of flooding if the gully pot is blocked. To move from this inefficient cleaning pattern it is important to get a better understanding of factors influencing the solid build-up rate in gully pots. Due to the scarcity of research on this topic, this research will therefore aim to assess the relationship of environmental and gully pot specific factors on the solid deposition in gully pots.

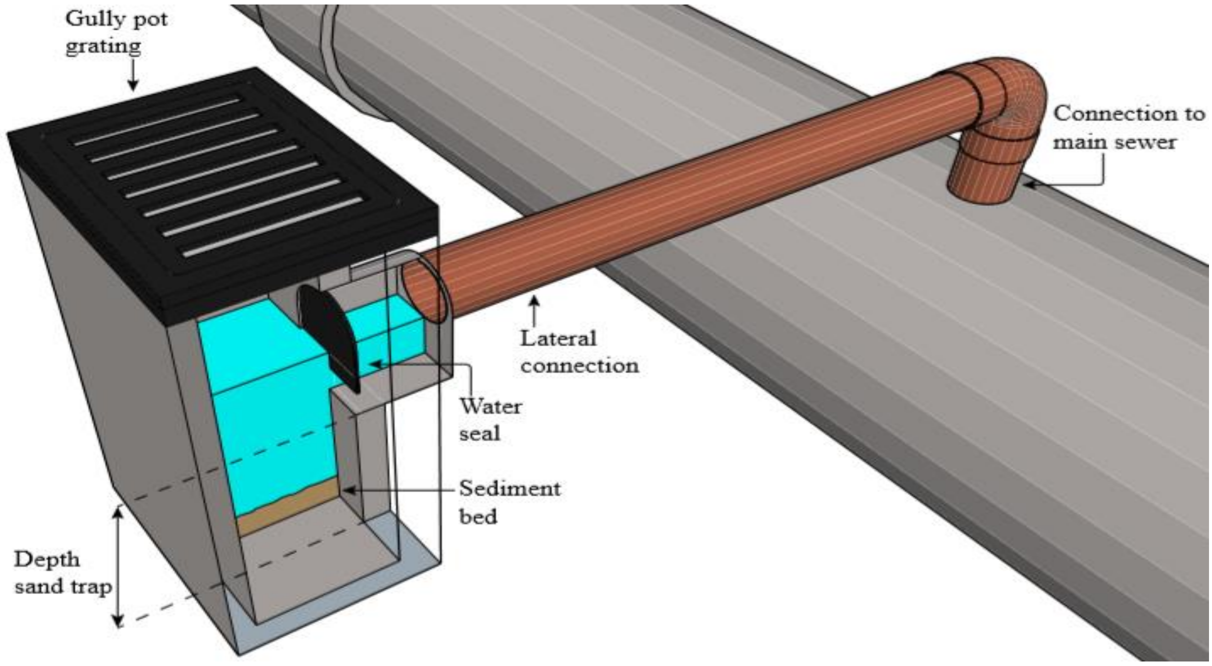


Figure 1.1 Schematisation of a gully pot and lateral connection (Post,2016)

1.3 Research framework

This thesis is part of a bigger project (Figure 1.2) which was set up by Matthijs Rietveld (TU Delft, PhD student, Section Sanitary Engineering). The project aims to improve sewer asset management, both regarding costs and serviceability. Overall it comprises five individual experiments. Three of these experiments (Gully pot inflow, Street sweeping reducing sediment and Gully pot retention) focus merely on the streets and two experiments (Physical modelling test gully pot, CFD calculation gully pot) focus on the gully pot itself. This research will focus on the experiment 'Gully pot retention' by looking into the influence of environmental and gully pot specific factors on solid deposition in gully pots. Ultimately the entire project aims to improve sewer asset management by decreasing costs and improving the functionality of the sewer system by more efficient maintenance. The entire project is part of the Dutch 'Knowledge Program Urban Drainage' (Dutch: 'Kennisprogramma Urban Drainage').

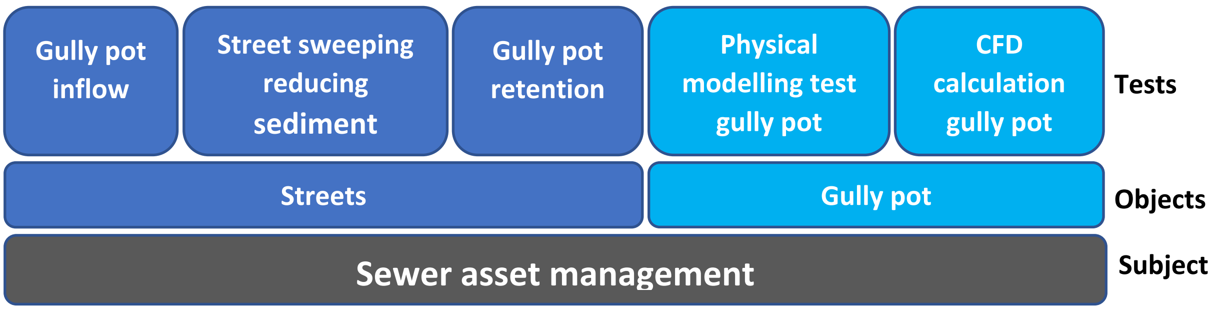


Figure 1.2 Research representation of tests, objects and subject by Matthijs Rietveld

1.4 Research question and sub questions

To gain better insight into the relationship between environmental and gully pot specific factors on the solid deposition in gully pots, this research postulates the following research question:

To what extent do environmental and gully pot specific factors influence solid deposition in roadside gully pots, being the entrance of an urban water system?

Following sub questions have been defined in order to answer the research question:

1) Based on which selection criteria were the monitoring streets chosen?
2) What are relevant internal gully pot processes influencing solid build-up in gully pots?
3) Which environmental and gully pot specific factors were considered as model parameters in order to explain the solid build - up rate in gully pots?
4) What is the relationship between environmental factors and the solid deposition in gully pots based on the model results?
5) What influence do gully pot characteristics have on the solid build-up in gully pots based on the model results?

1.5 Objective

The objective of this research is to gain better insight into the influence of environmental and gully pot specific factors on solid deposition in gully pots. In order to improve maintenance strategies, both by flood prevention and cost reduction, it is important to conduct research on the accumulated solid load in gully pots with respect to different environmental settings. Currently, municipalities do not base cleaning frequencies of gully pots on the physical mechanisms influencing the silting rate. The results of this report can be used to determine the factors that limit or enhance solid deposition and cleaning frequencies can consequently be adapted based on areas in which gully pots experience higher silting rates. Nevertheless maintenance strategies cannot be solely based on this research. It is important to evaluate this research and its results in the context of the entire project of Matthijs Rietveld in order to develop an integrated approach to improve overall maintenance strategies in the future.

1.6 Delimitations

The influence of the environment will most likely be dependent on seasonal variations. Because the measurements have not been monitored throughout the entire year, seasonal variations in solid deposition cannot be fully determined. Measurements were taken over a 7-month period, starting in November 2017 and continued by me from February to May 2018. The entire dataset can give an indication about the solid deposition in a specific timeframe, however, in the broader picture the measurement values of the entire research need to be viewed and assessed integrally in order to draw accurate and sustainable conclusions. No measurements of particles present on the road surface will be taken due to the time-consuming nature of this process which additionally influences and disturbs the present sediment layer.

1.7 Target audience

This report is written for lecturers and students from Van Hall Larenstein University of Applied Sciences, for the members of the Sanitary Engineering section from TU Delft and generally for people who are interested in the field of Watermanagement and Sanitary Engineering. Furthermore this report and the gathered data can be used by members of the Sanitary Engineering section from TU Delft as reference to conduct further research.

2 Materials and methods

This chapter describes the materials and methods that were used in this research. Firstly the project area will be described which will be followed up by three main sections, namely: literature research, data acquisition and data analysis. The selection criteria and the literature research were used to find relevant environmental and gully pot specific parameters. To assess the relationship between these factors and the solid build-up, a statistical model based on field measurements was developed.

2.1 Project area

Figure 2.1 Images from Paul Krugerlaan (left) and Keukenhoflaan (right) (Google Maps)

For this project seven streets were chosen in Rotterdam and The Hague. The location of each street is presented in Appendix 1. Following streets were monitored for this research:

Rotterdam Ludolf de Jonghstraat, De Lugt
The Hague Kanaalweg, Van Stolkweg, Keukenhoflaan, Leuvensestraat, Paul Krugerlaan

Before the start of the project in November 2017 the municipalities from Rotterdam and The Hague offered several streets to conduct monitoring. Originally, the selection of the streets was based on the presence of shops and trees in each street. During the first measurement activities in November 2017, it became evident that a notable amount of heterogeneity exists between the monitored streets. As an example the Paul Krugerlaan and Keukenhoflaan are shown in Figure 2.1. Enlarged images of all the monitoring streets are presented in Appendix 2. The Keukenhoflaan is located in a quiet and residential area with low traffic intensity. It is also noticeable that the road consists of bricks. The Paul Krugerlaan on the other hand is a paved road with a high traffic intensity and the presence of numerous shops. Previous research has been conducted in more homogenous environments, however, this research gave the opportunity to assess the solid build-up in more heterogeneous settings. This makes it easier to generalize and apply the results and conclusions from this research on other areas. For every street, a set of gully pots were selected for monitoring from either the entire street or a segment of the street. Overall 409 gully pots were monitored and Table 1 shows the number of chosen gully pots per street:

Street	Number of gully pots
De Lugt	63
Ludolf de Jonghstraat	58
Kanaalweg	49
Keukenhoflaan	71
Leuvensestraat	62
Paul Krugerlaan	55
Van Stolkweg	51
TOTAL	409

Table 1 Number of gully pots per street

2.2 Literature research

To evaluate which environmental and gully pot specific can be considered to explain the solid build-up rate, literature research was conducted. Based on the following paragraphs, the following table (Table 2) will give a list of possible environmental and gully pot specific factors which can be considered for the statistical model. The factors in this section are merely based on findings from literature research. The selection of the definite model parameters will be described in section 2.4.3.1:

Factor	Source
Environmental (Type 1)	
Traffic intensity	*(Deletic et al., 2000), (Post et al., 2016)*
Construction activities	*(Ashley and Crabtree, 1992)*
Weathering of buildings	*(Jartun et al., 2008)*
Animal waste	*(Brinkmann, 1985)*
Street sweeping	*(Brinkmann, 1985)*
Shops	*(ten Veldhuis and Clemens, 2011)*
Trees + seasonal variation	*(Pratt et al., 1987), (Chen et al., 2017), (Grottker, 1990)*
Antecedent dry period	*(Pratt and Adams, 1984), (Ellis and Harrop, 1984)*
Environmental (Type 2)	
Rainfall volume	*(Ellis and Harrop, 1984)*
Rainfall intensity	*(Ellis and Harrop, 1984)*
Storm duration	*(Ellis and Harrop, 1984)*
Flow volume discharge	*(Ellis and Harrop, 1984)*
Connected surface area	*(Post et al., 2016), (Pratt and Adams, 1984), (Butler and Karunaratne, 1995)*
Slope	*(Post et al., 2016)*
Road type	*(Post et al., 2016), (Garofalo et al., 2014), (Brinkmann, 1985)*
Gully pot specific factors	
Sandtrap depth	*(Memon and Butler, 2002), (Post et al., 2016), (Lager et al., 1977), (Butler and Karunaratne, 1995)*
Inlet type	*(Faram and Harwood, 2003)*
Position outlet pipe	*(Post et al., 2016)*
Water seal	*(Post et al., 2016)*

Table 2 Possible environmental and gully pot specific factors based on field observations and literature

2.2.1 Environmental factors

Within the environmental factors the following distinction can be made between two types of factors:

Type 1 Environmental factor influencing the characteristics, composition and availability of solids

Type 2 Environmental factors influencing the solid mobilisation and transport of solids to gully pots

2.2.1.1 Type 1 - Characteristics, composition and availability of solids

A study carried out by *Xanthopoulos and Augustin, 1992* shows from empirical data that the composition of solids is highly dependent on the characteristics of the area and local circumstances. The rate of solid supply to gully pots is therefore highly variable, depending on spatial and temporal variability in a catchment *(Pratt et al., 1987)*. Examining urban catchment areas, particles in these kind of environments consist predominantly of inorganic material, similar to sand and silt *(Lager et al., 1977; Sartor and Boyd, 1972)*. These can originate from numerous sources, such as: local traffic *(Deletic et al., 2000)*, construction activities *(Ashley and Crabtree, 1992)*, weathering of buildings *(Jartun et al., 2008)*, animal waste, litter and de-icing materials *(Brinkmann, 1985)*. Street cleaning with sweeping machines have found to be an effective measure in removing litter and coarsely sized material. Nevertheless, finer sized fractions of e.g. silt and clay are scarcely removed by street sweeping and the majority of fine particulate material remains in the catchment even if street sweeping is adjusted to more frequent cleaning intervals *(Brinkmann, 1985)*. Therefore it can be assumed that the frequency of street sweeping is a possible environmental factor affecting the solid deposition in gully pots by reducing the availability of solids on the road surface.

In the Netherlands, municipalities preventively clean gully pot chambers approximately once a year by and two to four times a year in vulnerable locations such as markets and shopping streets *(ten Veldhuis and Clemens, 2011)*. This indicates that the presence of markets and shopping streets is considered by municipalities as a possible factor for increased solid deposition. Markets and shopping streets can therefore also be taken into account as a possible environmental factor.

Furthermore *Garofalo et al., 2014* examined the particle size distribution of particulate matter (PM) in two basins were analysed, respectively located near a parking lot and a traffic intersection with high traffic intensity. The experiment revealed that traffic intensity also influences the solid load to be found in an environment of a gully pot. According to *Brinkmann,* 1985, if moderate driving is considered, the average tire wear is approximately 80 mg/vehicle km^{-1}, indicating that the traffic intensity could possibly influence the solid deposition in gully pots.

Looking at the influence of trees and seasonal variation, an analysis by *Pratt et al., 1987* revealed that there seems to be seasonal variation due to a peak In material supply during summer, autumn and after snow- and frost melt. This is supported by an analysis conducted in a study by *Chen et al., 2017*, who found through maintenance records of gully pots that a large amount of complains (>50%) were made in fall, indicating that leaf fall and other vegetation debris can be a potential reason for gully pot blockage. Through the analysis of dry weight, *Grottker, 1990* found that the loss on ignition for dry gully pot samples is about 6 – 10 % greater in autumn than in spring, suggesting an increased quantity of organic material in gully pots during autumn. However, a study carried out by *Scott, 2012* in England showed surprising results with the summer months having a higher organic matter percentage in gully pots than the remaining seasons.

Lastly, a study by *Ellis and Harrop, 1984* shows that the antecedent dry period (Appendix 3, Graph D) is weakly correlated to sediment loadings in gully pots (r = 0.41). This implies that the duration of the preceding dry period and the amount of accumulated sediments on the urban surface prior to a rainfall event, only have a limited relationship to sediment removal rates and deposition. This phenomenon is confirmed by *Pratt and Adams, 1984* who claim that sediments are not necessarily exhausted from the surface, suggesting that factors affecting solid mobilisation (such as rainfall volume, flow volume discharge and rainfall intensity) are the most important, not the period for accumulation of solids before the runoff event.

2.2.1.2 Type 2 - Solid mobilisation and transport

For solid mobilisation and the subsequent transport and supply of solids to gully pots, the following factors play an important role: rainfall volume, rainfall intensity, storm duration and flow volume discharge. This is based on an experimental research where nylon sieves (63 – 2000 µm) were placed in gully pots on the north west margins of London (*Ellis and Harrop, 1984)*. Peak sediment removal rates of 60 g/day^{-1} were measured in summer, in contrast to levels of 1.7 to 35.5 g/day^{-1} recorded in early spring (*Ellis and Harrop, 1984)*. These records suggest seasonal variation in solid deposition, indicating a high correlation between the solid build-up and rainfall characteristics and flow properties. The graphs in Appendix 3 prove that sediment loading is highly correlated with the total rainfall volume (r = 0.98; Graph A), flow volume discharge (r = 0.99; Graph C) and storm duration (r = 0.96; Graph E). The rainfall intensity is also an important measure to determine and explain the seasonal difference in sediment removal rates. Nevertheless, Graph B in Appendix 3 shows that rainfall intensity (r = 0.47) is statistically less significant than rainfall volume and storm duration.

Several studies also found, that the connected surface area should be a relevant factor when it comes to solid deposition in gully pot *(Post et al., 2016, Pratt and Adams, 1984, Butler and Karunaratne, 1995)*. It can be assumed that a larger contributing area increases the total flow volume and quantity of available solids for transport, consequently leading to higher solid deposition rates. However, the contributing area is also connected with the slope which can induce a more rapid flow of water, causing particles present on the road surface to get detached and transported more easily.

Lastly, research carried out by *Post et al., 2016* found a relationship between the road type (main road or local road) and the sediment deposition in gully pots, with a higher build-up rate for main roads than residential roads. In the Netherlands, it is common that the road surface residential areas consists of bricks (therefore also lower traffic intensity) and areas with a higher traffic intensity have commonly a road surface consisting of asphalt. The road type can influence the transport of solids, because washout of sediments between bricks can occur, slowing down the flow of water in the process *(Mobron, 2018)*. Asphalt on the other hand has a smoother surface, indicating that higher flow rates can be achieved on these kind of surfaces. The road surface can therefore be a hampering or enhancing factor regarding solid transport.

2.2.2 Gully pot specific factors

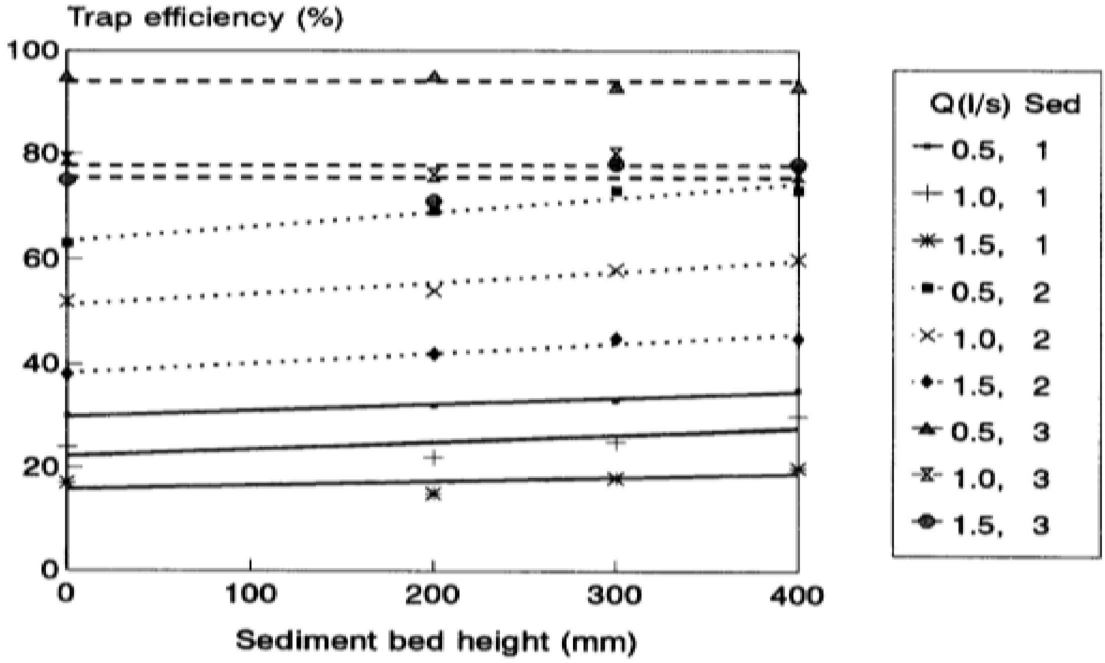


Figure 2.2 Relationship between the trap efficiency, sediment bed height and discharge (Butler and Karunaratne, 1995)

Memon and Butler, 2002 found in previous research that the sand trap depth is a crucial parameter regarding the retaining efficiency of the gully pot. One of the first laboratory studies on this topic was conducted by *Lager et al., 1977* who discovered that increased flow rates resulted in a lower efficiency of solid retention and that finer particles were trapped less efficiently than coarser particles. This can be confirmed by *Butler and Karunaratne, 1995* who reported, based on their laboratory study to examine solid trap efficiency (Appendix 4), that the trap efficiency reduces with increasing flow rate and decreasing particle sizes. Appendix 4 shows that the retaining efficiency drops significantly for sizes below 200 µm and overall efficiency of trapping sediments decreases with increasing flow rates. However, for sizes above 500 µm the data would suggest a relatively high level of solid trapping by effectively screening most large particles from entry into the drainage system. Furthermore, a study by *Lager et al., 1977* also showed that a gully pot can attain a negative retaining efficiency if the sediment depth in a gully pot reaches a threshold of 40 – 50% of the total storage height. This, however, is contradictory to the laboratory studies carried out by *Butler and Karunaratne, 1995* who found that an increase in the sediment bed height can marginally increase the trap efficiency (Figure 2.2). These contradictory outcomes and the fact that the research by *Lager et al., 1977* was conducted in North America with unknown gully pot specifications, makes the sand trap depth an interesting factor to consider in this research.

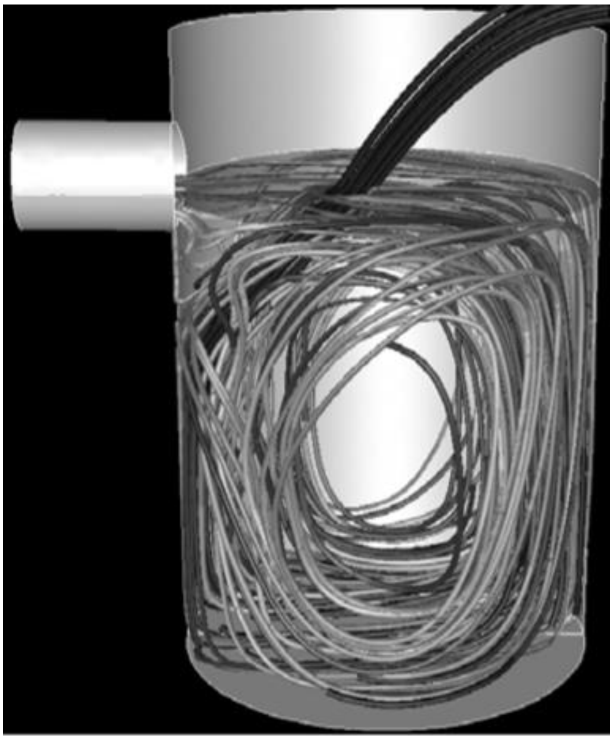


Figure 2.3 Gully pot fluid path lines (Faram and Harwood, 2003)

Furthermore, a study by *Faram and Harwood, 2003* shows that the entry of flow from above causes a circular motion of flows in the gully pot chamber, resulting in high flow velocities in the base region, (Figure 2.3) consequently increasing the probability of eroding the sediment layer. The fluid path lines in Figure 2.3 indicate, that the water is merely entering from the side of the gully pot. A different inlet type could possibly influence and alter the flow of water in a gully pot and, therefore, have an impact on the solid deposition in a gully pot. *Post et al., 2016* also suggests that the position of the outlet pipe can possibly influence the circular motion of flow in a gully pot, as gully pots with an outlet located at the front and the side were found to have higher sediment bed levels than gully pots with an outlet located at the back. Lastly, *Post et al., 2016* also assessed the impact of a water seal on the sediment accumulation which, however, did not deem to be an influential factor.

2.2.3 Internal gully pot processes

According to *Butler et al., 1995* gully pots operate under two distinct regimes: dry weather and wet weather. Biochemical processes dominate when the gully pot is operating under dry weather conditions, and physical processes dominate during wet weather conditions *(Butler et al., 1995; Scott, 2012)*. This research will however only focus on physical internal gully pot processes because these processes can largely influence the quantity of solid build-up, whereas biochemical processes mostly influence the quality of solids (i.e. increase of heavy metal concentrations during dry weather conditions through bacterial mobilisation as suggested by *Mance and Harman, 1978*). Therefore this section will look into the governing physical processes influencing solid build-up during dry and wet weather conditions.

2.2.3.1 Suspension

Butler et al., 1995 suggests that after a storm event, the inflow of solids into a gully pot ceases and conditions rapidly become quiescent. A large amount of solids transported to the gully pot in suspension, but which not have been discharged, will settle and ultimately deposit in the gully pot. The degree of settlement depends primarily on the size and specific gravity (SG) of the particles, temperature, and the period of time until the next storm event *(Butler et al., 1995)*. The concentration of suspended solids at any given time will be associated with the initial conditions following the end of a rainfall event *(Butler et al., 1995)* making suspension a dry weather process.

2.2.3.2 Erosion and re-suspension

Two physical processes govern internal gully pot processes during wet weather conditions: erosion and re-suspension (Figure 2.4). According to several studies, the solid mass contribution due to re-suspension is quantitatively not a significant factor and is limited to a period of 20 – 40 seconds of a rainfall event *(Butler and Karunaratne, 1995; Butler and Memon, 1999; Deletic et al., 2000; Fletcher and Pratt, 1981).* Results from a simulation model by *Fletcher and Pratt, 1981* highlights for a majority of storm events, that the concentration of suspended solids was derived from solids already in suspension rather than by solids in re-suspension *(Butler et al., 1995)*.

Appendix 5 shows a graph from an experimental study by *Butler and Karunaratne, 1995*, representing the concentration of eroded solids in effluent time and confirming that erosion is indeed limited to the first 20 – 40 seconds of a constant inflow rate. The peak of erosion is followed by a rapid inverse exponential reduction which has also been noticed by *Fletcher and Pratt, 1981* who explained that phenomenon as the 'unavailability of suitable material for release'. If the sediment layer is disturbed and ungraded, the sediment is more susceptible to erosion, explaining the short erosion period due to a substantial decrease in particles and the gradual grading of the sediment bed.

Nevertheless, at a low sediment level there is still a continuous erosional output observed for the smallest particle at the highest flow rate and bed depth *(Butler and Karunaratne, 1995)*. By applying a series of flushing tests equivalent to a heavy storm event, *Sartor and Boyd, 1972* found that only 1% of the sediment bed experiences re-suspension. This confirms findings by *Fletcher and Pratt, 1981* who found that the majority of solids discharged from gully pots are caused by a lack of sedimentation rather than re-suspension.

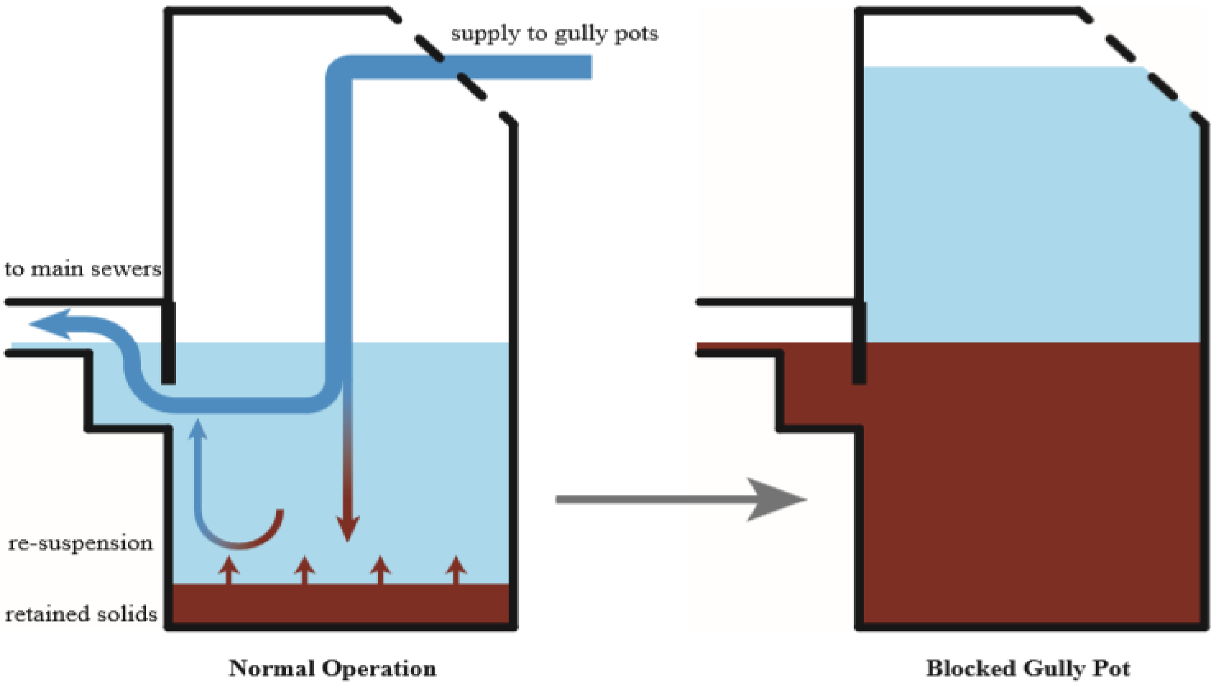


Figure 2.4 Solid transport to gully pots and sediment deposition leading to blockage (Post 2016)

2.3 Data acquisition

Prior to the first measurement period (November 2017) the gully pots in all monitoring streets were cleaned by the municipality. To measure the sediment depth a similar measuring device was used as in the research by *Post et al., 2016*. The principle of the device is illustrated in Figure 2.5.

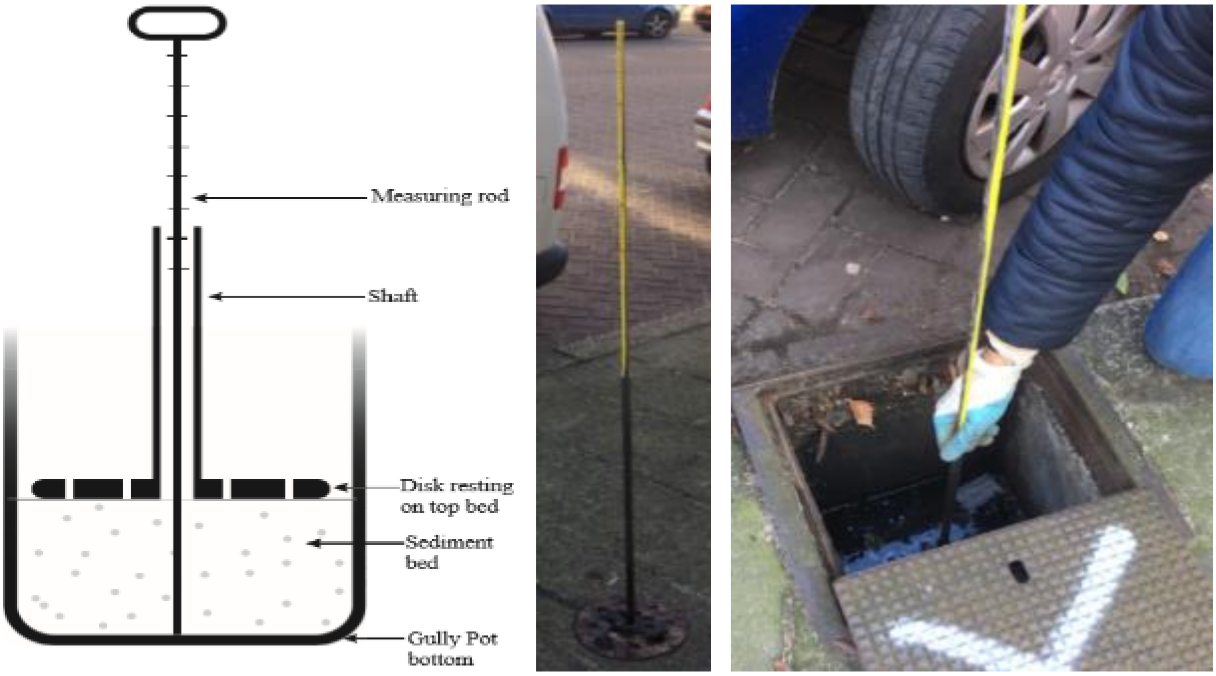


Figure 2.5 Schematisation measuring device (left) (Post et al., 2016), real life measurement device (middle), application of measurement device in gully pot (right)

The measurement device consists of a punctured disk attached to a shaft, with a centrally located retractable measuring rod. During a measurement the disk rests on the sediment bed, while the rod is forced through the sediment layer until the bottom of the gully pot is reached. The measuring rod is equipped with a sequence of marks at intervals of $2{\cdot}10^{-3}$ m, enabling the person conducting the measurement to determine the height of the sediment bed after the gully pot was opened.

Measurements of the sediment bed were taken with a precision of $5{\cdot}10^{-3}$ - $1{\cdot}10^{-2}$ m. Nevertheless the organic matter content of the sediment bed could negatively influence the precision of a measurement. Through the compression of sediment by the punctured disk a large quantity of carbon dioxide and air gets released, significantly decreasing the height of the sediment bed. Due to unintended varying compression per gully pot and the dislocation of solids to the sides of the gully pot during pressure application, the measuring error gets magnified. Furthermore the inner shape of the gully pots also posed a source for flawed measurements. Several gully pots had an ellipse shaped bottom (Figure 2.6), which makes it in a lot of cases impossible for the punctured disk to rest on top of the sediment layer because the ellipse shaped bottom was too narrow. This could lead, in case of low sediment bed levels, to a higher measurement value for the height of the sediment bed because the punctured disk would be located above the actual top of the layer. As last measurement problem the degree of saturation of the sediment bed with water was an area of concern. If the sediment bed of a gully pot was completely dry, all the solids accumulated at the bottom of a gully pot, making measurements easier and more precise. A higher water

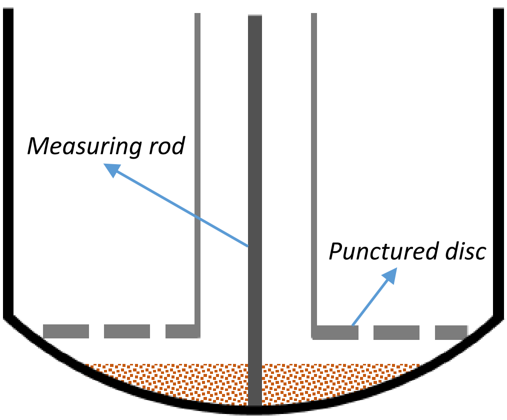


Figure 2.6 Failure mechanism measurement

level in a gully pot in contrast caused some solids (in particular organic material) to float and elude the measurement device, making it impossible to compress all the solids in a gully pot.

The measurements activities approximately took place in three week intervals. During measurements it could happen that gully pots were not measurable due to inaccessibility by parking cars, clogging issues and construction work. It has to be mentioned that gully pots in the Ludolf the Jonghstraat had severe clogging issues which appeared to be the result of a damaged or blocked lateral connection. This assumption can be made based on a minor to non-existent development of a sediment layer, while the gully pots were entirely filled with water. To what extent these missing data points were considered will be explained in the following subchapter. Furthermore measurements could not be executed in the period from 3 April 2018 until 18 May 2018 in parts of the Kanaalweg due to construction work. Therefore twenty gully pots could not be measured in this period.

2.4 Data analysis

This section will give an insight about the process of data analysis. The first two steps included the conversion of measurement values (cm) into a value for the daily solid growth (L/day) and the detection of outliers. The last step comprises all the aspects revolving around the Generalized Linear Model (GLM) which was applied for the statistical analysis in this research.

2.4.1 Conversion of measurement values

Because this research focuses on the daily build-up rate of solids, the obtained data had to be converted into a daily growth between the measurement periods. Following formula was applied to determine these values:

[1] $$Build - up\ rate = \frac{m_i - m_{i-1}}{t}$$ [L/day]

in which m_i is the sediment depth at measurement i, m_{i-1} is the sediment depth from the previous measurement and t the elapsed time between measurements. For a missing measurement due to inaccessibility, construction work or clogging, the following measurement point was used to compute the build-up rate. Consequently, the elapsed time was also adjusted to the period between these two measurements. A distinct difference between this research and the one conducted by *Post et al., 2016* is that this research merely focuses on the build-up rate of sediment between measurements. *Post et al., 2016* focusses on the total solid layer thickness after several months, resulting in a sequence of correlated measurements. This research focuses on the accumulation between two measurements. Therefore, the build-up of one measurement only influence the build-up of the subsequent measurement.

2.4.2 Outlier detection

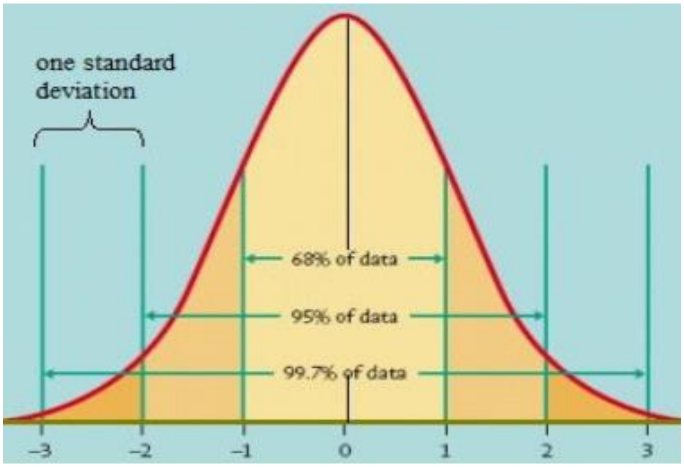


Figure 2.7 Visualization of 'Empirical Rule' (StatisticsHowTo, 2013)

The following step was the detection of outliers in the dataset. Therefore the 'Empirical Rule' (also called 68% – 95% – 99.7% rule) was applied which is a statistical rule stating that 68% of the values fall within the first standard deviation, 95% of the value fall within the first two standard deviations and 99.7% fall within the first three standard deviations (Figure 2.7). For this research the values within three standard deviations were considered as valid measurements.

The outlier detection was applied on all 1750 data points ranging from November 2017 until May 2018. However, it has to be noted that outlier detection and removal was conducted per street, thus the mean (μ) and standard deviation (σ) were computed separately for every monitored street. This was done because the characteristics of each street were different and therefore also the range of measurement values. The subsequent statistical analysis was consequently done with a new dataset excluding the outliers. The results regarding outlier detection and removal will be presented in the following chapter.

2.4.3 Statistical analysis

As the transport and sedimentation processes of solids are complex and not fully understood, a deterministic approach to storm water modelling is connected with numerous difficulties *(Deletic et al., 2000).* Additionally the application of a deterministic approach requires the approximation of site specific parameters, which are prone to uncertainty and inaccuracy *(David and Matos, 2002)*. This can be supported by a study from *Deletic et al., 2000* who found limited model accuracy for larger rainfall events due to deficiencies in the understanding of in-pot erosion processes. Based on these considerations this research will therefore use a probabilistic approach by applying a Generalized Linear Model (GLM).

2.4.3.1 Selection, description and data of explanatory variables

Variable	Factor	Index	Type	Unit	Range
Connected surface area	Environmental	X_1	continuous	(m²)	[6 – 793.5]
Traffic intensity	Environmental	X_2	continuous	(cars/day	[300 – 6000]
Effective sand trap depth	GP specific	X_3	continuous	(cm)	[-18.5 – 28.5]
Effective sand trap volume	GP specific	X_4	continuous	(L)	[-22.66 – 45.57]
Shops	Environmental	X_5	categorical	-	Yes [0] No [1]
TN1 Leaf abscission	Environmental	X_6	continuous	(m)	[0 – 14.16]
TN1 Leafless	Environmental	X_7	continuous	(m)	[0 – 14.16]
TN1 Leaf growth	Environmental	X_8	continuous	(m)	[0 – 12.54]
TN1 Leaf max. capacity	Environmental	X_9	continuous	(m)	[0 – 0.15]
TN2 Leaf abscission	Environmental	X_{10}	continuous	(m²)	[0 – 167.86]
TN2 Leafless	Environmental	X_{11}	continuous	(m²)	[0 – 167.86]
TN2 Leaf growth	Environmental	X_{12}	continuous	(m²)	[0 – 119.28]
TN2 Leaf max. capacity	Environmental	X_{13}	continuous	(m²)	[0 – 0.10]
Road surface	Environmental	X_{14}	categorical		Asphalt [0] Bricks [1]
Inlet	GP specific	X_{15}	categorical		Side [0] Top + Combi [1]
Rainfall volume	Environmental	X_{16}	continuous	(mm/day)	[0.658 – 12.732]
Max. rainfall intensity	Environmental	X_{17}	continuous	(mm/h)	[3.6 – 36.84]

Table 3 Explanatory variables with index, type, unit and range

A Generalized Linear Model (GLM) comprises a set of explanatory variables that aim to describe the behaviour of the response variable. In this case, the explanatory variables (or covariates) are the environmental and gully pot specific factors and the response variable is the solid build-up rate. Section 2.2.1 ('Environmental factors') and 2.2.2 ('Gully pot specific factors') give an indication about possible environmental and gully pot specific factors which could have been used as model parameters in this research. The list was trimmed down to seventeen explanatory variables. The factor, index, type, unit and range of all explanatory variables is listed in Table 3. These variables will be explained more detailed in the following subparagraphs regarding their quantification.

The selection of model parameters was mainly based on data availability and the possibility and complexity of quantifying a parameter. Firstly, factors such as construction activities, weathering of buildings and animal waste could not be taken into account due to the difficulty and complexity of quantifying these parameters. Although construction activities took place during the measurement period, the sample size of affected gully pots by construction activity was so small that the inclusion of this parameter would have been redundant. The same accounts for the weathering of buildings and animal waste which are a factors of uncertainty by being extremely difficult to monitor. Furthermore data regarding street sweeping could not be taken into account because the municipalities of The Hague and Rotterdam were not yet able to provide data for this factor. For the remaining variables, time to accurately assess and quantify the variables was a limiting factor in this research.

2.4.3.1.1 Traceability and explanation of dataset

A dataset with information of the explanatory variables and the response variable was already provided at the start of this research in February 2018. New explanatory variables were added during this research period, including the presence of shops and seasonal variation of the leaf load. Furthermore, values for traffic intensity were adjusted based on newly provided data by the municipalities and the RVMK model (section 2.4.3.1.4). Additionally the distribution of the road surface was adjusted due to an observation error in the preceding measurement period. One dataset includes the information of every explanatory variable and the response variable; managed and provided by Matthijs Rietveld (PhD student, Sanitary Engineering, TU Delft). The dataset includes all measurement values ranging from November 2017 until May 2018. However, the dataset cannot be released publicly and shared in context with this research. For any questions regarding the accessibility of the dataset, the managing person of the dataset, Matthijs Rietveld, should be contacted.

2.4.3.1.2 Tree factor with seasonal variations

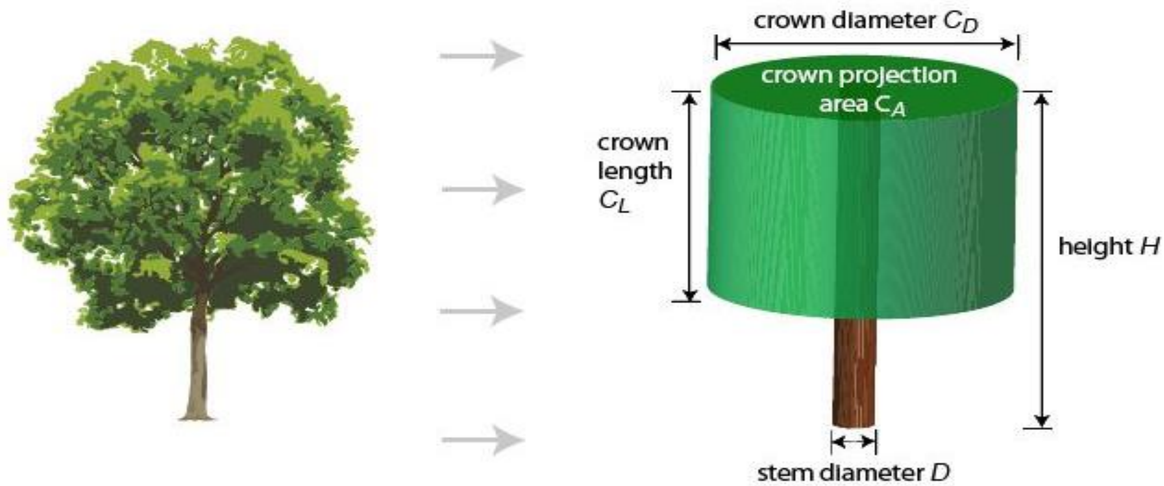


Figure 2.8 Translation of real-world tree to tree model with important parameters (Mobron, 2017)

To determine the effects of trees on the solid deposition in gully pots, two equations were used to respectively compute tree factor 1 (TN1) and tree factor 2 (TN2), giving an approximation of the leaf load per street:

[2] $$TN1 = \frac{1}{L_{street}} \sum_i H(i) * C_D(i)$$ [m]

[3] $$TN2 = \frac{1}{L_{street}} \sum_i H(i) * C_D(i)^2$$ [m^2]

In equations [2] and [3] L_{street} represents the length of the street, H is the height of the tree and C_D is the crown diameter of the tree (Figure 2.8). The main difference between TN1 and TN2 is that the crown diameter in equation of TN2 is squared. This adds, compared to the two dimensional TN1 factor, another dimension to the TN2 factor and could potentially be a more accurate approximation of the leaf load. Furthermore height data for the trees was obtained from the 'Actueel Hoogtebestand Nederland' (AHN3). The measurements of the crown diameters for Rotterdam were provided by the municipality of Rotterdam. For The Hague, only the trunk diameters of the trees were available which were determined with the help of the application 'Haagse Bomen App' which was made available by the municipality of The Hague. For the translation of the trunk diameter to the crown diameter, the relationship between the trunk diameter and crown diameter for semi-detached trees proposed by *Hasenauer, 1997* was used *(Mobron, 2018)*, making it possible to compute the crown diameter for each tree. The length of the monitored streets was determined by the website 'afstandmeten.nl'.

The tree factor was combined with a seasonal factor which is based on different tree phases throughout the year as observed by *Halverson et al., 1986*. The observations distinguish between the following phases: leaf development, full leaf capacity, leaf abscission and leafless. Trees do not act within the boundaries of the classical seasonal distribution of spring, summer, autumn and winter. Field observations yielded that every street behaved differently in terms of tree phases. Therefore, with every measurement activity photos were taken of the trees in each street. Subsequently the photos were analysed and based on the characteristics of the trees, the measurement activity was then categorised into one of the mentioned tree phases. If the measurement activity would fall into one of the four categories, the model would equal the three other phases to a value of zero. In the model, the four tree phases are considered as four different tree factors with seasonal variation.

2.4.3.1.3 Connected surface area

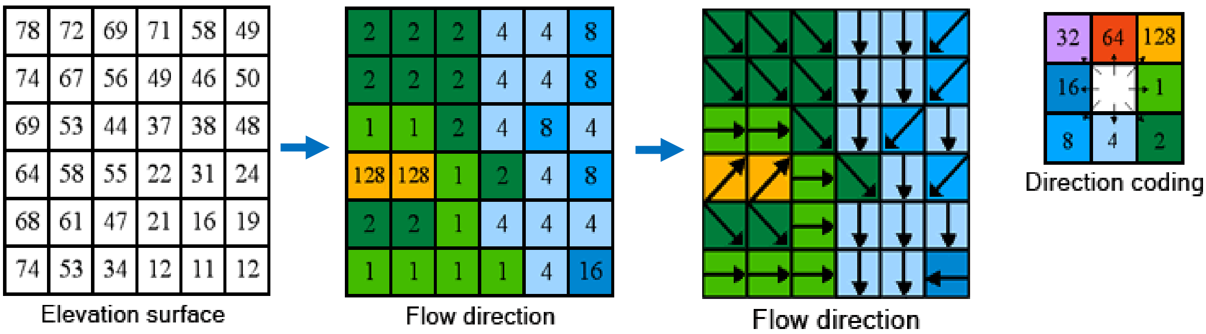


Figure 2.9 Visualization of the determination of the flow direction (Esri, 2012)

The connected surface area of each gully pot was determined by the means of the eight-direction flow approach (D8), which has been one of the earliest and simplest method for specifying flow directions *(Tarboron, 1997)*. This flow model follows an approach as presented by *Jenson and Domingue, 1988* and it assigns the flow from each pixel to one of its eight neighbours that are either adjacent or diagonal (Figure 2.9), depending on the steepest downward slope *(Tarboron, 1997)*. The distance is calculated between the cell centres of each pixel. To derive the flow direction from a digital elevation model (DEM) as presented by *Jenson and Domingue, 1988* the 'Actueel Hoogtebestand Nederland' (AHN3) with a systematic error of maximum 5 cm and a standard deviation of 5 cm was used. The data for the DEM was obtained through airborne laser scanning with a spatial resolution of 0.5m * 0.5m from 2008 *(Zon, 2013)*. Confounding factors such as cars and trees were extracted and filtered from the DEM and were subsequently interpolated from the surrounding data by applying Kriging as used by *Post et al., 2016*.

2.4.3.1.4 Traffic intensity

The traffic intensity (cars/day) was determined for all seven monitored streets in Rotterdam and The Hague, ranging from 50 – 6000 cars/day. The numbers were taken partially from the 'RVMK Holland Rijnland v3.0-verkeersmodel', data from the municipalities and assumptions. The RVMK is a traffic- and environmental model that simulates traffic flows for various future scenarios. This model however primarily computes values for main streets and can be, according to field observations, relatively inaccurate. Therefore data on traffic intensity was also provided by both municipalities. Furthermore the traffic intensity for streets in residential areas (f.e. Keukenhoflaan) needed to be assumed based on field observations and traffic intensities in other streets.

2.4.3.1.5 Road surface

The road surface distinguishes between the usage of asphalt or bricks. The type of road surface and the traffic intensity are correlated, since bricks are mainly used in streets of residential areas and asphalt in streets with increased traffic intensity.

2.4.3.1.6 Shops

As aforementioned, during field measurements it became evident that the presence of shops also needed to be taken into consideration as a possible factor influencing solid deposition in gully pots. This became apparent in the Paul Krugerlaan where a high quantity of shops were located and the composition of solids in the gully pots differed largely from other streets. Therefore every street was assessed based on the presence and location of shops.

2.4.3.1.7 Inlet

Figure 2.10 Top inlet (left), Combination inlet (middle), Side inlet (right)

In this research three types of inlets were present, namely: combination, side and top (Figure 2.10). From measurements it became apparent that top inlets accumulated a much larger quantity of solids which is the reason for adding this parameter. Therefore the model distinguishes between these three types, whereby the top and combination inlet are defined as one value within the inlet type due to a low amount of combination inlets in this research.

2.4.3.1.8 Effective sand trap depth

This factor does not only take the respective sand trap depth of a gully pot into account, but also the present sediment layer. Equation [4] shows the formula for the computation of the effective sand trap depth in which t is the measurement (or observation) number, i is the gully pot number, $ESTD_{t,i}$ is the effective sand trap depth of specific gully pot at a given measurement number, STD_i is the sand trap depth of a specific gully pot and $m_{t,i}$ is the measurement of the sediment layer for a specific gully pot at a given measurement number. The effective sand trap depth could attain a negative value which can be explained by the fact that the present sediment depth exceeded the actual sand trap depth.

$$[4] \qquad ESTD_{t,i} = STD_i - m_{t,i} \qquad [\mathrm{cm}]$$

2.4.3.1.9 Effective sand trap volume

Most studies focus primarily on the sand trap depth rather than the sand trap volume, most likely due to the constant length and width. The experimental set-up from this research however comprises gully pots with varying sizes and therefore the sand trap volume was also added as a parameter. To calculate the effective sand trap volume, the effective sand trap depth is multiplied by the cross section of the gully pot as given in equation [5] in which t is the measurement (or observation) number, i is the gully pot number, $ESTD_{t,i}$ is the is the effective sand trap depth of specific gully pot at a given measurement number, STD_i is the sand trap depth of specific gully pot, $m_{t,i}$ is the measurement of the sediment layer for a specific gully pot at a given measurement number and CS_i is the cross section of a specific gully pot. Negative values for the effective sand trap volume can be explained by the fact that the present sediment depth exceeded the actual sand trap depth. At each measurement the remaining sand trap volume gets newly calculated (the same applies for the effective sand trap depth).

[5] $$ESTV_{t,i} = (STD_i - m_{t,i}) * CS_i$$ [L]

2.4.3.1.10 Rainfall volume and intensity

For the rainfall volume and intensity the climatological radar rainfall dataset of 5-min precipitation accumulation from KNMI ('Koninklijk Nederlands Meteorologisch Instituut') was used. This dataset combines validated and complete rain gauge data from both KNMI rain gauge networks in De Bilt and Den Helder *(KNMI, 2017)*. Data is available for the entire land surface of the Netherlands and the pixels have a spatial resolution of 1km. From January 2018 onwards the radar in De Bilt was replaced by a new radar in Herwijnen. For this research the rainfall data was extracted per street for the following coordinates (Table 4):

Location	Latitude	Longitude
De Lugt	51.940	4.424
Ludolf de Jonghstraat	51.930	4.429
Kanaalweg	52.101	4.283
Keukenhoflaan	52.029	4.294
Leuvensestraat	52.109	4.292
Paul Krugerlaan	52.069	4.289
Van Stolkweg	52.100	4.284

Table 4 Coordinates street locations

It is important to mention that rainfall data was only available from the 26th October 2017 until the 31st March 2018 which is equivalent to a dataset of 1271 data points. For this reason two separate models needed to be run for firstly all explanatory variables excluding rainfall volume and intensity (*'Model 1'*) and secondly only for rainfall volume and intensity (*'Model 2'*).

2.4.3.2 Generalized Linear Model (GLM)

To assess the relationship between the solid build-up rate and environmental and gully pot specific factors, this research will apply a Generalized Linear Model (GLM). The model comprises three components, namely:

- **a linear predictor**: incorporates the information of the explanatory variables
- **a distribution function**: describes the response variable following an exponential distribution
- **a link function**: connects the mean of the distribution function with the linear predictor

A general point of confusion can be the distinction between generalized linear models and the general linear model, two extensive statistical models. The general linear model can be considered as a unique case of the generalized linear model with normally distributed responses and the usage of an identity link. A generalized linear model however is a further generalization and an extension of the general linear model, allowing the response to have a distribution other than the normal distribution.

2.4.3.2.1 Linear predictor

The linear predictor η contains the deterministic part of the model, which is a linear function of k explanatory variables *(Post et al., 2016)*. The linear predictor is the component that incorporates the information of the explanatory variables into the model and it is denoted by the following equation:

[6]
$$\eta_{t,i} = \beta_0 + \beta_1 x_{t,i,1} + \beta_2 x_{t,i,2} + \cdots + \beta_k x_{t,i,k}$$

where t represents the respective observation number and i the gully pot. β refers to the weights assigned to the respective explanatory variable X_k, determining to what extent X_k contributes to the prediction of the response variable Y_i. β_0 is a constant and in geometrical terms an intercept that refers to the predicted value of Y_i if all explanatory variables equal 0. Generally it can be said that the β values give the predicted change in Y_i for a unit increase in X_k.

2.4.3.2.2 Distribution function

The distribution function specifies the conditional distribution of the response variable Y_i. According to *Nelder and Wedderburn, 1972* original formulation, the distribution of the response variable Y_i is part of the exponential family which can include the following distributions: Gaussian, binomial, Poisson, gamma or inverse Gaussian. For this research the random component was defined by a Gaussian normal distribution, given by the following equation:

[7]
$$f\left(\eta_{t,i} | \mu_i, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\eta_{t,i} - \mu)^2}{2\sigma^2}}$$

where μ is the mean of the distribution, σ is the standard deviation and σ^2 the variance. A normal distribution was fitted to all 1750 measurement points, as can be seen in Appendix 6. It has to be noted that the fit is not ideal, however this was the most accurate distribution for the given data. Characteristics of the remaining distributions will be given below, indicating that the remaining distributions from the exponential family are not suitable to describe the response variable:

Binomial: is suitable when the response variable is categorical, meaning that it has two possible outcomes (i.e. binary data) or if proportions need to be modelled *(CU Statistics, 2018; Dunn, 2008)*

Poisson: is suitable when the data is collected in counts (i.e. the number of tails in 12 flips of a coin) and many discrete response variables have counts as possible result *(CU Statistics, 2018)*

Gamma: can be used for positive continuous data *(Dunn, 2008)* and is often used for the analytics of waiting time, time between failure or accident rate, supposing some event occurs n times in unit intervals *(Matsuzaki, 2017)*

Inv. Gaussian: is suitable when the response variable is strictly positive and skewed to the right *(Phillips, 2018)*

Therefore we assume that Y_i follows a normal distribution with a mean μ_i and variance σ^2

[8] $$Y_i \sim N(\mu_i, \sigma^2)$$

Generally the normal distribution is used to model continuous data with a symmetric distribution *(Dobson, 2002)*. Modelling with a normal distribution is according to *Dobson, 2002* widely used for three main reasons:

1) Naturally occurring phenomena are described by the normal distribution f.e. height and blood pressure of people
2) If the data is not normally distributed (i.e. skewed distribution) the average or the total of a random sample of values will be normally distributed by approximation which is proven by the Central Limit Theorem
3) The large availability of statistical theory and applications for the normal distribution

2.4.3.2.3 Link function

The link function $g(\,)$ serves to connect the linear predictor with the distribution function of the model. The link function transforms the expectation of the response variable $E(Y_i) = \mu_i$ to the linear predictor $\eta_{t,i}$ by using a non-linear function. This also allows the non-linearly transformed expectation $g(\mu_i)$ to range from $-\infty$ to ∞ *(Dobson, 2002)*. For a normal distribution the canonical link is the identity function *(Rodríguez, 2017)* and according to *Turner, 2008* the canonical link leads to desired properties of the GLM. The transformation is given by equation [9] and it can be noted that the identity link function simply returns the linear predictor $\eta_{t,i}$ as unaltered argument. Through the transformation $g(\mu_i) = \eta_{t,i}$ the expectation μ_i depends on β_k of the explanatory variables.

[9] $$g(\mu_i) = \eta_{t,i} = \beta_0 + \cdots + \beta_k x_{t,i,k}$$

2.4.3.2.4 Model fitting

To attain β values for every explanatory variable, model fitting was conducted by a method called 'Iteratively Resquared Least Square' (IRLS). This method firstly starts with initial estimates of μ_i and $g(\mu_i)$ and subsequently computes the adjusted response variable (Y) at each iteration, from the values of μ_i and $g(\mu_i)$ from the preceding iteration along with the respective weights (W). The weights are determined by the inverse covariance matrix which is newly computed at each cycle. Every cycle of the iterative scheme yields a new linear predictor value $\eta^{(t+1)} = X\beta^{(t+1)}$ and the adjusted response value $Y^{(t+1)}$ *(Agresti, 2002)*. Fitting a linear model by weighted least squares for a response variable (Y^t) is done by the following equation:

[10] $$\beta^{(t+1)} = (X'W^{(t)}X)^{-1}X'W^{(t)}Y^{(t)}$$

2.4.3.2.5 Hypothesis testing

In order to identify factors influencing the solid build-up in gully pot, hypothesis testing was applied. Following null hypothesis and alternative hypothesis were defined for this research with a chosen significance level of α = 0.05:

H_0 *The factor does not cause a change in the solid deposition in gully pots* **($p > 0.05$)**

H_a *The factor causes a change in the solid deposition in gully pots* **($p < 0.05$)**

The Z-score, which is a measure of standard deviation, is a statistical tool to test statistical significance by helping to decide whether or not to reject the null hypothesis *(ESRI, 2018)*. In order to reject the null hypothesis, a judgement needs to be made regarding the degree of risk of falsely rejecting the null hypothesis. This is the reason for defining a significance level which exemplifies the risk of rejecting the null hypothesis although it is true. The p-value on the other hand is a calculated probability, measuring the strength of evidence for or against the null hypothesis, if the null hypothesis is actually true. Most authors refer to statistically significant as $p<0.05$ and statistically highly significant as $p<0.001$ *(Statsdirect, 2018)*. The Z-score value and p value statistic are related to the standard normal distribution (also exemplified in Figure 2.5) with $N(0,1)$, relating standard deviations with probabilities and allowing significance levels and confidence intervals to be attached to Z-score values and p-values *(ESRI, 2018)*.

In the model, the Z-score value for every explanatory variable is attained by dividing the computed regression coefficients (β_k) by the standard error (SE_k) of the respective variable (X_k). The critical Z-score values, for a 95% confidence interval (based on a significance level of α = 0.05), are -1.96 and 1.96 *(Medcalc, 2018)*. Consequently the associated p-value with a confidence interval of 95% is 0.05. Assuming that a calculated Z-score for one of the parameters is between -1.96 and 1.96, the p-value will be larger than 0.05, indicating that the null hypothesis cannot be rejected and the factor does not cause a change in the solid build-up. However, if the Z-score of a parameter falls outside of the given range, the null hypothesis will be rejected, indicating that the factor causes a change in the solid-build up.

3 Results

This chapter will describes the obtained results for this research. Firstly the results of data exploration will be presented, which will be followed up by the results of the statistical analysis. The entire data analysis was conducted in the programming language Python and the source code for the GLM is presented in Appendix 7.

3.1 Data exploration

The common problem of all statistical methods is 'rubbish in, rubbish out' *(Zuur et al., 2010)*. Influential factors such as outliers and heterogeneity (differences in variation) may cause major problems in linear regression and analysis of variance models *(Fox, 2008)*, leading to biased parameter estimates and unreliable model results. Therefore, detailed data exploration firstly needs to be applied before any statistical analysis can be conducted *(Zuur et al., 2010)*.

3.1.1 Field measurement data

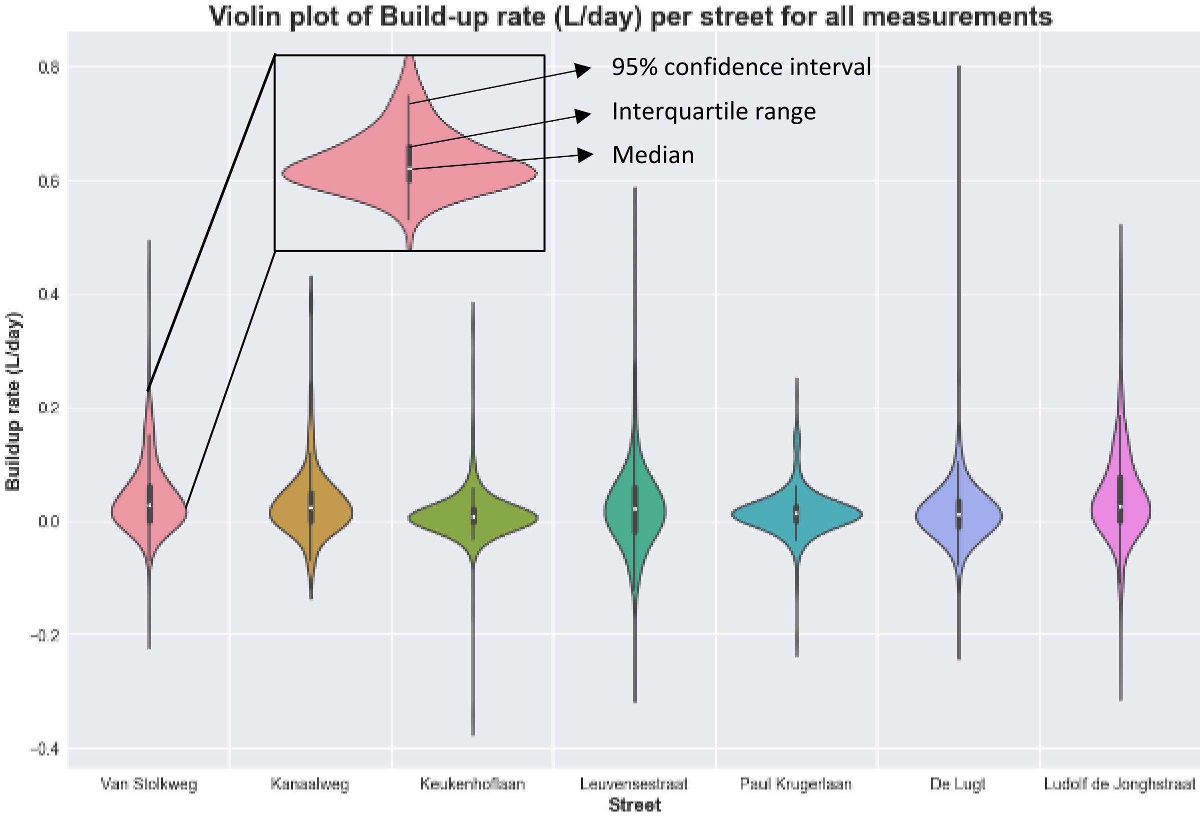


Figure 3.1 Violin plot of the build-up rate (L/day) per street for all measurements (Python)

In the period from 27th October 2017 until 8th May 2018 the measurement activities yielded 1750 data points. To visualize the build-up rate (L/day) per street for all measurements a violin plot was chosen for data visualization (Figure 3.1). Violin plots are comparable to box plots, however they can be interpreted as hybrid between box plots and kernel density plots because they include both these plot types. The outer shape (kernel density estimation) of the violin plot represents all measured results, with the broadest section indicating the highest probability that members of the dataset take on the given value. Furthermore, a boxplot is present within the kernel density estimation, describing the median with a white dot, the interquartile range with a thick grey bar and the 95% confidence interval with a thin grey line. An advantage of the violin plot is, that it is able to give information about the distribution of the chosen variable.

In the case of this research, the data for all streets have a unimodal distribution, which can be explained through the fact that the plots do not have more than one peak (multimodal). Furthermore, it is noticeable that for all the plots the broadest section of the kernel density is located close to the median. All the median values are above zero, indicating a positive net accumulation throughout the entire measurement period. Several plots (Keukenhoflaan, Leuvensestraat, De Lugt and Ludolf de Jonghstraat) have long tails which can be explained by the presence of outliers in the dataset.

The boxplots per street in Appendix 8 highlight, that the largest variety in values was recorded in Van Stolkweg, Leuvensestraat and Ludolf de Jonghstraat. It is also noticeable that that the measurement values in Keukenhoflaan and Paul Krugerlaan are centred close to zero, indicating a minor to non-existent variation in the build-up rate. Furthermore, the boxplots indicate the presence of outliers (especially in De Lugt) which can have a significant impact on the dataset and the subsequent statistical analysis as indicated by *Zuur et al., 2010*.

3.1.2 Outlier detection and removal

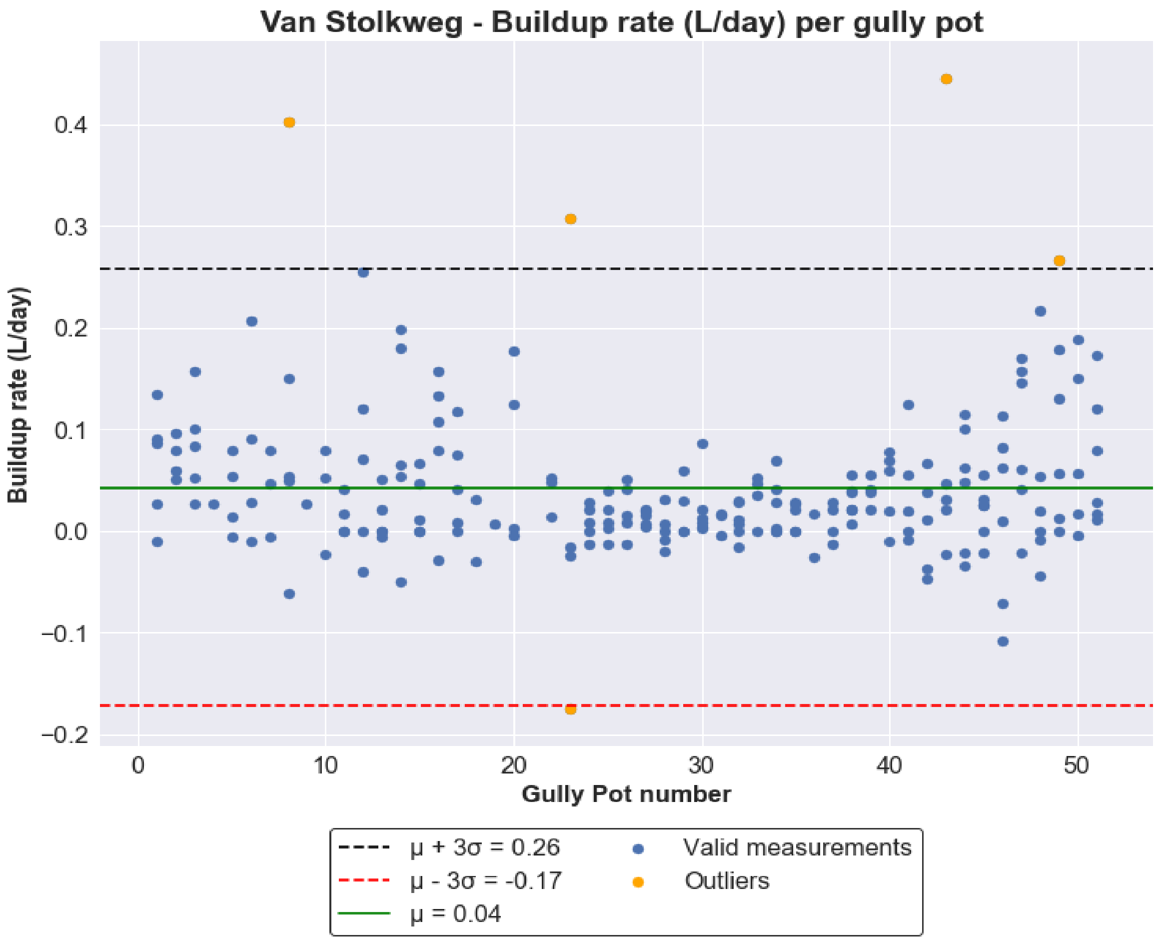


Figure 3.2 Scatterplot Buildup rate (L/day) with all valid measurements and outliers for Van Stolkweg

As described in section 2.4.2, outliers were detected and removed from the dataset by applying the Empirical Rule. The presence of outliers may influence statistical data analysis and cause overdispersion *(Gumedze and Chatora, 2014)*, consequently leading to biased parameter values and standard errors in the model *(Harrison, 2014; Hilbe, 2011)*. From 1750 data points, 40 values were classified as outliers and therefore removed from the dataset. 97.7% of the data points fall thus in the range of $\mu \pm 3\sigma$ and 2.3% fall out of that range. As an example all valid measurement values and outliers are plotted for Van Stolkweg in Figure 3.2. Appendix 9 comprises six scatterplots with valid measurements, the outliers and the given μ and $\mu \pm 3\sigma$ values for every other monitored street.

3.1.3 Correlation matrix and Variation Inflation Factor (VIF)

Another factor of uncertainty is the correlation between explanatory variables. Strong collinearity amongst covariates might lead to unreliable and falsified parameter estimates, as the parameters might react erratically to minor changes in the data *(Zuur et al., 2010)*. Therefore, firstly a correlation matrix (Appendix 10) with the Pearson standard correlation coefficient got computed in order to detect variables that cannot be identified separately. The results show, that especially the respective explanatory variables for tree number 1 (X_6 - X_{10}) and tree number 2 (X_{11} - X_{15}) have a very strong to perfect correlation (r = 0.98 – 1.00). Furthermore, the effective sand trap depth (X_3) and the effective sand trap volume (X_4) are also strongly correlated (r = 0.86).

To verify these outcomes, the variance inflation factor (VIF) was calculated for every covariate. This is a method, using a covariate X_k as a response variable and all remaining covariates as explanatory variables. A high R^2 value in these models indicate that the primary variation in covariate X_k, is explained by the remaining covariates, establishing the state of collinearity (or multicollinearity). If multicollinearity is given, one predictor variable can be linearly predicted with a substantial degree of accuracy by other predictor variables; therefore this variable should be dropped from the covariates *(Zuur et al., 2010)*. According to *Zuur et al., 2010*, the choice of elimination can be merely based on the variation inflation factor. For this research following VIF values were computed (Table 5):

Variable	**Index**	**VIF (1st run)**	**VIF (2nd run)**
Connected surface area	X_1	3.55	3.54
Traffic intensity	X_2	3.27	2.96
Effective sand trap volume	X_3	43.60	4.49
Effective sand trap depth	X_4	53.47	-
Shops	X_5	2.77	2.14
TN1 Leaf abscission	X_6	39.73	-
TN1 Leafless	X_7	77.35	-
TN1 Leaf growth	X_8	65.21	-
TN1 Leaf max. capacity	X_9	*inf*	-
TN2 Leaf abscission	X_{10}	37.59	1.31
TN2 Leafless	X_{11}	70.51	2.03
TN2 Leaf growth	X_{12}	63.36	1.22
TN2 Leaf max. capacity	X_{13}	*inf*	-
Road surface	X_{14}	5.28	4.42
Inlet	X_{15}	2.32	1.85

Table 5 VIF values for all covariates (Python)

The results show, based on the proposed cut-off range of 5 – 10 by *Montgomery and Peck, 1992*, that the values for following covariates are very high: effective sand trap depth (VIF = 43.60), effective sand trap volume (VIF = 53.47), TN1 Leaf abscission (VIF = 39.73), TN1 Leafless (VIF = 77.35), TN1 Leaf growth (VIF = 65.21), TN2 Leaf abscission (VIF = 37.59), TN2 Leafless (VIF = 70.51), TN2 Leaf growth (VIF = 63.36). Due to the fact that the effective sand trap depth (X_4) has a higher VIF value than the effective sand trap volume (X_3), the effective sand trap depth was dropped as covariate from the model.

Furthermore all TN1 factors have a higher VIF value than their respective TN2 counterparts. Therefore it was chosen to drop all the covariates related to TN1. It needs to be noted that the TN1 Leaf max. capacity (X_9) and TN2 Leaf max. capacity (X_{13}) have perfect collinearity, suggesting that these are completely redundant covariates. This can be explained by the fact that in none of the conducted measurement activities, the trees possessed maximum leaf capacity. Therefore all the values for X_9 and X_{13} were equal to zero. It was chosen to retain the road surface (X_{14}) because the VIF value was equal to the lower boundary of the cut-off range. All in all the following covariates were dropped from the model: X_3, X_6, X_7, X_8, X_9 and X_{13}.

3.2 Statistical analysis

After conducting data exploration in order to diminish Influential factors such as outliers and multicollinearity, statistical analysis can be used on the improved dataset. The following subchapters will present the results of the Generalized Linear Model (GLM).

3.2.1 Model quality – Akaike Information Criterion (AIC)

The Akaike Information Criterion (AIC) is an estimator, which assesses the relative quality of a statistical model. Given the proposed probability distribution of a normal distribution, there are three possible link functions according to *Dunn, 2008*, namely: identity, log and inverse. Each model comprised the same set of explanatory variables and response variable. The AIC values given in Table 6 reveal that the normal distribution and the identity link had best relative performance compared to the other two model structures, indicated by the lowest AIC value. Therefore the normal distribution and identity link were used as model structure for this research.

Model	AIC
Normal distribution + Identity link	-5286.56
Normal distribution + Log link	-5379.84
Normal distribution + Inverse	-5316.40

Table 6 AIC values for competing model structues (Python)

3.2.2 Model 1 results

This section assesses the relationship of environmental and gully pot specific factors on the build-up rate in gully pots (excl. rainfall volume and intensity). Through stepwise elimination of explanatory variables, the covariates that cause a change in the solid build-up were identified. After every model run, the covariate with the highest p-value was consequently removed from the model. Model running ended when every remaining covariate had a p-value < 0.05. All in all the model was run five times. The explanatory variables with their respective p-value per run are presented in the Table 7:

Variable	Index	1st run	2nd run	3rd run	4th run	5th run
		p-value	*p-value*	*p-value*	*p-value*	*p-value*
Connected surface area	X_1	2.05E-01	2.00E-01	1.99E-01	1.49E-01	-
Traffic intensity	X_2	2.80E-01	2.43E-01	3.29E-01	-	-
Effective sand trap vol.	X_3	8.74E-08	3.76E-08	1.29E-08	1.81E-08	7.48E-09
Shops	X_5	5.36E-01	5.21E-01	-	-	-
TN2 Leaf abscission	X_{10}	3.32E-14	1.46E-14	1.65E-14	2.21E-14	1.72E-15
TN2 Leafless	X_{11}	1.96E-05	1.50E-05	1.79E-05	2.54E-05	1.75E-06
TN2 Leaf growth	X_{12}	5.01E-02	4.96E-02	5.82E-02	8.10E-02	4.54E-02
Road surface	X_{14}	9.46E-01	-	-	-	-
Inlet	X_{15}	1.45E-07	1.29E-07	9.66E-05	1.41E-07	2.22E-07

Table 7 P-values for every covariate per model run (Python)

After the fifth model run, the connected surface area (X_1), traffic intensity (X_2), shops (X_5) and road surface (X_{14}) were excluded from the model. According to the model the following environmental and gully pot specific have an influence on the build-up rate in the monitored gully pots: Effective sand trap volume (p = 7.48E-09), TN2 Leaf abscission (p = 1.72E-15), TN2 Leafless (p = 1.75E-06), TN2 Leaf growth (p = 4.54E-02) and Inlet (p = 2.22E-07). It is noticeable, that the p-value of TN2 Leaf growth is close to the threshold of $\alpha = 0.05$, indicating a near elimination from the model. Following mean weight values (β) were computed in the final model run (Table 8):

Variable	Index	β value
Constant	β_0	-1.03E-02
Effective sand trap volume	β_3	9.21E-04
TN2 Leaf abscission	β_{10}	2.82E-04
TN2 Leafless	β_{11}	1.01E-04
TN2 Leaf growth	β_{12}	9.69E-05
Inlet	β_{15}	1.36E-02

Table 8 β values for significant factors and constant (Python)

All the covariates positively contribute to solid build-up in gully pots. The positive β value for the inlet indicates that the category in the covariate that has the higher number accounts for the solid build-up. Top or combination inlets consequently account for solid-build up in gully pots, rather than side inlets. Furthermore the results yield that TN2 Leaf abscission has the highest β value compared to all TN2 factors, indicating that this factor accounts for the highest build-up rate amongst all TN 2 factors.

Based on the results, the deterministic component of the estimated GLM with standardised and dimensionless covariates and mean weight values can therefore be written by the following equation:

$$[11] \quad \eta_{t,i} = -1.03\mathrm{E}^{-02} + 9.21\mathrm{E}^{-04} * Effective\ sand\ trap\ volume_{t,i} + 2.82\mathrm{E}^{-04} * TN2\ Leaf\ abscission_{t,i} + 1.01\mathrm{E}^{-04} * TN2\ Leafless_{t,i} + 9.69\mathrm{E}^{-05} * TN2\ Leaf\ growth_{t,i} + 1.36\mathrm{E}^{-02} * Inlet\ (Top/Combi)_{t,i}$$

3.2.3 Model 2 results

As described in section 2.4.3.1.10 'Rainfall volume and intensity', the rainfall data had to be run in a separate GLM due to a time lag in data provision by KNMI. The same routine was applied as for the model excluding rainfall volume and intensity. Following results were attained for the model including rainfall volume and intensity (Table 9):

Variable	Index	1st run	2nd run
		p-value	*p-value*
Rainfall volume	X_{16}	1.59E-01	1.15E-04
Rainfall intensity	X_{17}	8.05E-01	-

Table 9 P-values for rainfall volume and intensity (Python)

After the second model run, the rainfall volume (X_{16}) deemed to be a significant factor ($p = 1.15E-04$), whereas the rainfall intensity (X_{17}) was largely insignificant ($p = 8.05E-01$) after the first model run and therefore had to be eliminated from the model. The positive β value for the rainfall volume ($\beta = 3.25E-03$) indicates, that an increase in the rainfall volume causes an increase in the solid build-up and consequently a decrease in the rainfall volume causes a decrease in the solid build-up.

3.2.4 Standardized residuals vs. fitted values

Residuals are defined as the difference between any data point and the fitted regression line and can be also referred to as stochastic error. By subtracting the observed value from the predicted (or fitted) value, the residual can be calculated for every data point. In practice standardized residuals are expected to follow a normal distribution *(Sutton et al., 2000)*. This can be explained by the Central Limit Theorem which states, that the distribution of the sum of a large number of random (or independent) variables or observations tend to be normally distributed. If non-normality is present, then the underlying assumptions of regression have been violated. Figure 3.3 depicts the standardized residuals with a normal distribution fit for this research. As it can be seen, the normal distribution fits the residuals reasonably well. However, the normal distribution does not perfectly fit the data, indicating that possible factors (such as a missing variable, a missing interaction between terms in the model, correlation of residuals etc.) have not been taken into account in the model *(Minitab, 2012)*.

To support this assumption, a normal probability (Q-Q) plot has been plotted for the residuals (Appendix 11). A Q-Q plot is a plot, where the axes are transformed in order to plot a normal distribution in a straight line and in this is case it has been used to assess the normality of the residuals. It takes the quantiles of the residuals versus the quantiles of the normal distribution and it can be seen, that some non-normality is present in the residuals (indicated by blue boxes in Appendix 11). Furthermore, the standardized residuals vs. fitted values (Appendix 12) show, that the residuals follow a random pattern and as indicated by the third polynomial fit, the residuals and fitted values balance each other out and approximate to zero (in the range of 0 – 0.06 of the fitted values). This, however, does not account for values in the range from -0.02 to 0 and 0.06 – 0.08 of the fitted values. The positive tilt of the fit (between -0.02 to 0) indicates a larger quantity of positive residuals and therefore lower prediction than observation values. Consequently, the negative tilt (between 0.06 – 0.08) indicates a larger quantity of negative residuals and therefore higher prediction than observation values.

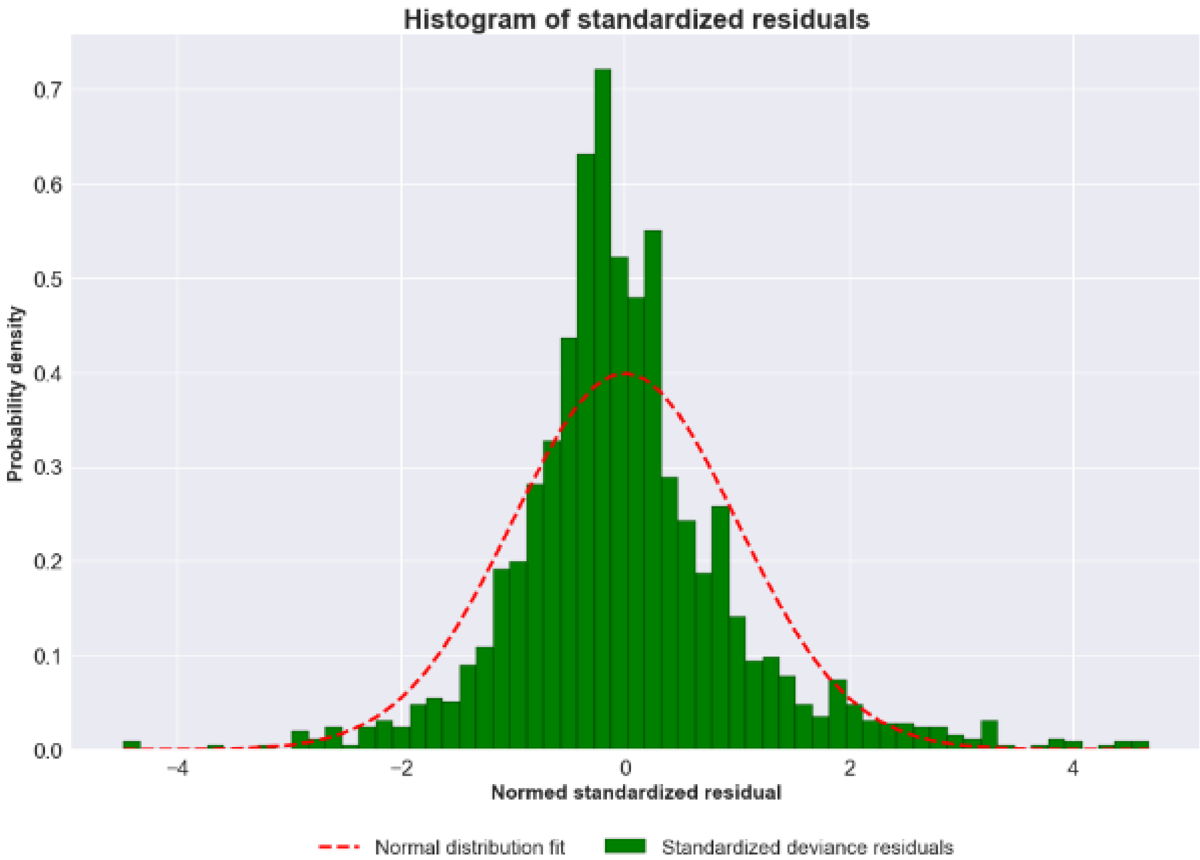


Figure 3.3 Histogram of standardized residuals with normal distribution fit (Python)

4 Conclusion

The main objective of this research was to examine the influence of environmental and gully pot specific factors on the solid build-up rate gully pots. Therefore, a Generalized Linear Model (GLM) was developed in Python in order to detect the significant factors influencing solid deposition. Based on the predefined research question, this research comes to the conclusion that the effective sand trap volume, rainfall volume, top/combination inlet and trees with seasonal variation are positively correlated with solid build-up, and thus influencing the solid deposition in gully pots. In order to answer the main research question, five sub questions functioned as guidelines to draw key conclusions from these research findings.

Before the start of the project in November 2017, the municipalities of Rotterdam and The Hague offered several streets to conduct monitoring and the street selection was originally based on two environmental factors, namely the presence of shops *(ten Veldhuis and Clemens, 2011)* and trees *(Pratt et al., 1987; Chen et al., 2017; Grottker, 1990)*. However, during the first measurement activities in 2017 it became evident that a distinct heterogeneity existed between the streets in terms of for example the traffic intensity and road surface. A long list of environmental and gully pot specific factors was created based on literature research, in order to get a better understanding of influential factors. This ultimately yielded 17 model parameters which were selected mainly based on the availability of data and the complexity of quantifying a parameter. Following model parameters were thus selected as explanatory variables (X_k) to describe the behaviour of the solid build-up as response variable (Y_i): connected surface area, traffic intensity, effective sand trap depth and volume, shops, tree number 1 (for leaf abscission, leafless, leaf growth and leaf max. capacity phase), tree number 2 (for leaf abscission, leafless, leaf growth and leaf max. capacity phase), road surface, inlet type, rainfall volume and rainfall intensity.

Further important factors which need to be considered relating to the quantity of solid deposition, are three physical internal gully pot processes: suspension, erosion and re-suspension. An experimental study by *Butler and Karunaratne, 1995* suggests that erosion of the solid layer limited to 20 – 40 seconds at the beginning of a constant inflow rate, which is also the time frame in which re-suspension occurs. From a research by *Fletcher and Pratt, 1981* it can be assumed that suspension is the most dominant physical internal gully pot process controlling solid-build, because the contribution of suspended solids in gully pots could predominantly be derived from suspension rather than re-suspension.

After defining model parameters and getting a better understanding of internal gully pot processes, the Generalized Linear Model (GLM) could be applied. The results indicate that following environmental factors were significant: TN2 Leaf abscission, TN2 Leaf growth, TN2 Leafless and rainfall volume. The positive β values indicated, that TN2 leaf abscission (β = 2.82E-04) accounts for the highest quantity of solid deposition amongst all the tree factors. These findings support a research carried out by *Chen et al., 2017* who found through maintenance records of gully pots, that a large amount of complains (>50%) were made in fall, which coincides with the leaf abscission phase of trees.

The TN2 Leafless factor has the second highest β value (β = 1.01E-04) which can be explained by the fact that this period is the subsequent phase of the leaf abscission phase. Leaf- and litter from the preceding abscission phase might still be present on the streets and transported to gully pots during rainfall events, explaining the higher β value than in the leaf growth phase (β = 9.69E-05). However the p-value of TN2 Leaf growth (p = 4.54E-02) indicates, that this factor was almost eliminated from the model based on a threshold of α = 0.05, indicating that this parameter could be dropped in the future if more records from this period are available. Furthermore the rainfall volume, which together with the rainfall intensity had to be run in a separate due to a time lag in data provision from KNMI, was also found to be a significant factor, coinciding with findings by *Ellis and Harrop, 1984* who proved that sediment loading is highly correlated with the total rainfall volume (r = 0.98).

Next to the environmental factors, following gully pot specific factors were found to be significant: top/combination inlet and the effective sand trap volume. From field observations it became evident that top and combination inlets accumulate a higher quantity of solids than side inlets and for this reason this parameter was also considered in this research. The observations coincide with the model results, indicating that top or combination inlets (β = 1.36E-02) also have an influence on the solid build-up in gully pots. The main difference between a top and a side inlet is that the water with entrained solids can enter a gully pot with a top inlet from every direction whereas the water entrance from a gully pot with a side inlet is limited to one side. This can potentially result in higher solid deposition rates for top inlets during rainfall events. Furthermore the side inlet can cause more rapidly flowing water to enter the gully pot from a diagonal angle, causing increased circular motion of flows in the gully pot chamber. This results in high flow velocities as suggested by *Faram and Harwood, 2003* and therefore increased eroding probability of the present sediment layer.

Lastly, the effective sand trap volume also deemed to be an influential factor when it comes to solid deposition in gully pots. The positive weight β_3 corresponds to a positive correlation between the variable and the build-up rate, presumably indicating that increased sand trap volume increases the retention of inflowing solids. It can be assumed that an increased sand trap volume leads to prolonged retention time for particles, suggesting that suspended solids deposit more extensively. Next to that, larger sand traps could experience less turbulence than smaller sand traps during rainfall events, which would lead to less erosion. However, decreasing the effective sand trap volume leads to decreased solid deposition, indicating the reaching of an equilibrium state as suggested by *Post et al., 2016* who found stabilising sediment levels after several months for 95% of the gully pots.

In conclusion, the findings from this research may aid to support and improve maintenance strategies of gully pots and its surroundings. Knowledge of environmental and gully pot specific properties that contribute to progressive accumulation of solids in gully pots might justify investments during the design- and maintenance phase. By knowing the contributing factors and the solid build-up rate in heterogeneous environments, this study can provide more cost-efficient sewer asset management and help to facilitate a move from reactive to more proactive measures.

4.1 Discussion

With regards to models, the British statistician George Box gave the following aphorism in his 1987 book 'Empirical Model-Building and Response Surfaces':

'Essentially, all models are wrong, but some are useful' *- George Box*

Models are basically a simplification of the reality and therefore they never give an exact representation of the reality. Nevertheless a model can help us to approximate the reality by helping us to understand processes and events that we would not have been able to comprehend without a model. As stated by *Welsing, 2015*, models are supposed to result in clinically useful prediction or improved understanding of the interaction and relation between factors. This surely also applies to the GLM used in this research. The following paragraphs will therefore address a number of possible error sources, which could have negatively influenced the modelling process and results. Furthermore possible explanations will be given relating to the insignificance of model parameters.

Firstly, the standardized residuals have shown that they are not perfectly normally distributed. As aforementioned, this can be caused by various factors, such as: a missing variable, a missing interaction between model terms, correlation of residuals with another variable and the correlation of residuals amongst each other *(Minitab, 2012)*. This research assumes that there is no correlation between successive observations from the same gully pot. In practice however, successive observations can be correlated, leading to the conclusion that the used model structure does not account for the violation of independence. In the study by *Post et al., 2016*, the deterministic component got extended with a correlation structure to model intergully pot variation and the correlation produced by this variation, in order to resolve the violation of independence. The inclusion of for example an auto-regressive process, where the correlation between measurements declines exponentially with time as used by *Post et al., 2016*, might reduce the error of the random component and lead to overall better model performance.

Moreover it was remarkable that the connected surface area (X_1) and rainfall intensity (X_{17}) were not significant factors in this research. It is intuitive to think, that a larger connected surface area leads to a higher discharge and therefore increasing solid deposition. However from a physical standpoint, a higher discharge can lead to higher erosion rates and therefore re-suspension, ultimately decreasing the sediment depth. Furthermore the provided height data was measured for The Hague in 2014 and for Rotterdam in 2016. Since then land subsidence might have taken place, altering the contributing area of each gully pot and influencing the accuracy of the connected surface area computation. Nevertheless, the connected surface area was the last factor to be eliminated from the model (p = 1.49E-01), suggesting that this factor is close to being significant. Studies from *Post, 2016*, *Pratt and Adams, 1984* and *Butler and Karunaratne, 1995* found the connected surface area to be a relevant factor.

On the other hand, the insignificance of rainfall intensity can possibly be explained through the exhaustion of solids on the road surface throughout the rainfall event. This assumption can be supported by *Ellis and Harrop, 1984* who only found moderate correlation (r = 0.47) between the rainfall intensity and sediment loading, indicating that the availability of solids on the road surface is an important factor to consider in relation to the rainfall intensity.

Furthermore the presence of shops (p = 5.21E-01) did not deem to be a significant factor. A possible explanation could be an increased street sweeping frequency in these areas. It was suggested by *ten Veldhuis and Clemens, 2011* that the gully pot chambers in vulnerable locations such as markets and shops are preventively cleaned two to four times a year. Therefore it can be assumed that the street sweeping frequency is more frequent in the Paul Krugerlaan than for example the Keukenhooflaan. It is therefore a logical conclusion that the factors street sweeping and shops cancel each other out.

The traffic intensity (p = 3.29E-01) and the road surface (p = 9.46E-01) did also not deem to be significant factors. In a research carried out by out by *Post et al., 2016* the traffic intensity is not a significant factor, however a relationship was found between the road type (asphalt or brick) and the sediment deposition in gully pots, with a higher build-up rate for main roads than residential roads. These are contradicting results, because the road surface was the first model parameter in this research to be eliminated from the model. Other studies, such as *Garofalo et al., 2014* and *Brinkmann,* 1985 show, that the traffic intensity should have an influence on the solid deposition in gully pots predominantly through tire wear as indicated by *Brinkmann,* 1985.

Lastly the flawed internal communication of the municipalities and the hired companies for cleaning activities was an area of concern. Throughout the research it was clearly communicated when to clean the gully pots, however it happened numerous times that the hired companies did not follow the schedule by either cleaning roads they were not supposed to clean, or not cleaning roads they were supposed to. Without being noticed by the municipality, it only became evident through personal observations during measurement activities that gully pots were either cleaned or not cleaned. Subsequently the municipalities had to be contacted in order to confirm whether the gully pots in a specific street were cleaned or not cleaned. It is crucial that the communication between all involved parties improves in the near future, since good communication improves the overall process of the research and diminishes errors, which could possibly influence the results of future research.

4.2 Recommendations for future research

Based on the results, conclusion and discussion of this research, recommendations can be made regarding future research. Following recommendations can be made to potentially improve future research regarding the solid deposition in gully pots:

[1] It is recommended to continue this research at least until November 2018, in order to draw more accurate conclusions regarding the seasonal behaviour of solid deposition. A research by *Scott, 2012* showed surprising results, with gully pots experiencing higher organic matter percentage during the summer months than in the remaining seasons, potentially indicating that increased solid deposition takes place during this period. This could not be assessed in this research due to missing measurements during the maximum capacity phase of trees. The relationship between the maximum capacity of trees and solid deposition should be assessed in further research. Furthermore, a larger quantity of data points should also improve the credibility and performance of the statistical analysis.

[2] For the following research period, the inclusion of for example an autoregressive component should extensively be studied. This can improve the model performance by possibly resolving the violation of independence and minimizing the error terms in the model.

[3] It is recommended to look into the environmental and gully pot specific factors which were insignificant according to the model results and additionally add possible influential factors to the model which are not yet included. There is proof by several studies that the factors which were gathered through literature research should have an impact on the solid deposition in gully pots. More detailed data can possibly be acquired from the municipalities regarding the traffic intensity and new data for the street sweeping frequency. The improvement of the model is a continuous process.

[4] The results from this research should be assessed jointly with the results from the other four experiments within the research framework. The results can be generalized, potentially leading to the construction of a model that can accurately predict solid build-up rate with given environmental settings and gully pot specific factors. This model can be very helpful for municipalities for assessing areas based on its susceptibility to gully pot blockage and moving a redundant and reactive cleaning pattern.

References

Agresti, A., 2002. Categorical Data Analysis, Wiley series in probability and statistics. doi:10.1198/tech.2003.s28

Ashley, R., Hvitved-Jacobsen, T., Bertrand-Krajewski, J.-L., Verbanck, M., 2004. Solids in sewers - Characteristics, effects and control of sewer solids and associated pollutants. IWA Publishing.

Ashley, R.M., Crabtree, R.W., 1992. Sediment origins, deposition and build-up in combined sewer systems. Water Sci. Technol. 25, 1–12.

Ashley, R.M., Fraser, A., Burrows, R., Blanksby, J., 2000. The management of sediment in combined sewers. Urban Water 2, 263–275. doi:10.1016/S1462-0758(01)00010-3

Bolognesi, A., Casadio, A., Ciccarello, A., Maglionico, M., Artina, S., 2008. Experimental study of roadside gully pots efficiency in trapping solids washed off during rainfall events. Int. Conf. Urban Drain. 1–10.

Brinkmann, W.L.F., 1985. Urban stormwater pollutants: Sources and loadings. GeoJournal 11, 277–283. doi:10.1007/BF00186341

Bryan Ellis, J., Revitt, D.M., 1982. Incidence of heavy metals in street surface sediments: Solubility and grain size studies. Water. Air. Soil Pollut. 17, 87–100. doi:10.1007/BF00164094

Butler, D., Karunaratne, S.H.P.G., 1995. The suspended solids trap efficiency of the roadside gully pot. Water Res. 29, 719–729. doi:10.1016/0043-1354(94)00149-2

Butler, D., Memon, F., 1999. Dynamic modelling of roadside gully pots during wet weather, Water Research. doi:10.1016/S0043-1354(99)00050-0

Butler, D., Xiao, Y., Karunaratne, S.H.P.G., Thedchanamoorthy, S., 1995. The gully pot as a physical, chemical and biological reactor. Water Sci. Technol. 31, 219–228. doi:10.1016/0273-1223(95)00339-O

Chen, Y., Cowling, P., Polack, F., Remde, S., Mourdjis, P., 2017. Dynamic optimisation of preventative and corrective maintenance schedules for a large scale urban drainage system. Eur. J. Oper. Res. 257, 494–510. doi:10.1016/J.EJOR.2016.07.027

Cherqui, F., Belmeziti, A., Granger, D., Sourdril, A., Le Gauffre, P., 2015. Assessing urban potential flooding risk and identifying effective risk-reduction measures. Sci. Total Environ. 514, 418–425. doi:10.1016/J.SCITOTENV.2015.02.027

CU Statistics, 2018. Generalized Linear Models [WWW Document]. URL http://www.stat.columbia.edu/~martin/W2024/R11.pdf

David, L., Matos, R., 2002. Wet weather water quality modelling of a Portuguese urban catchment: difficulties and benefits. Water Sci. Technol. 131–140.

Deletic, A., Ashley, R., Rest, D., 2000. Modelling input of fine granular sediment into drainage systems via gully-pots. Water Res. 34, 3836–3844. doi:10.1016/S0043-1354(00)00133-0

Dobson, A.J., 2002. An Introduction to Generalized Linear Models, CRC Press Company. doi:10.1198/tech.2002.s91

Dunn, P., 2008. Generalized Linear Models in R [WWW Document]. URL http://maths-people.anu.edu.au/~johnm/courses/r/ASC2008/pdf/glm-ohp.pdf

Ellis, J.B., Harrop, D.O., 1984. Variations in solids loadings to roadside gully pots. Sci. Total Environ. 33, 203–2011. doi:https://doi.org/10.1016/0048-9697(84)90394-2

ESRI, 2018. What is a Z score and what is a p-value? [WWW Document]. URL http://resources.esri.com/help/9.3/arcgisdesktop/com/gp_toolref/spatial_statistics_toolbox/what_is_a_z_score_what_is_a_p_value.htm

Faram, M.G., Harwood, R., 2003. A method for the numerical assessment of sediment interceptors. Water Sci. Technol. 47, 167–174.

Fletcher, I.J., Pratt, C.J., 1981. Mathematical simulation of pollutant contributions to urban runoff roadside gully ponds, in: In Procdeeings of the 2nd International Conference on Urban Storm Drainage. Urbana, USA.

Fox, J., 2008. Applied Regression Analysis and Generalized Linear Models. SAGE Publication, Hamilton.

Garofalo, G., Carbone, M., Piro, P., 2014. Sampling, testing and modeling particle size distribution in urban catch basins. Water Sci. Technol. 70, 1873–9. doi:10.2166/wst.2014.409

Grottker, M., 1990. Pollutant removal by gully pots in different catchment areas. Sci. Total Environ. 93, 515–522. doi:10.1016/0048-9697(90)90142-H

Gumedze, F.N., Chatora, T.D., 2014. Detection of outliers in longitudinal count data via overdispersion. Comput. Stat. Data Anal. 79, 192–202. doi:10.1016/j.csda.2014.05.018

Halverson, H.G., Gleason, S.B., Heisler, G.M., 1986. Leaf duration and the sequeunce of leaf development and abcission in northeastern urban hardwood trees. Urban Ecol. 9, 323–335.

Harrison, X.A., 2014. Using observation-level random effects to model overdispersion in count data in ecology and evolution. PeerJ 2. doi:10.7717/peerj.616

Hasenauer, H., 1997. Dimensional relationships of open-grown trees in Austria. For. Ecol. Manage. 96, 197–206. doi:10.1016/S0378-1127(97)00057-1

Hilbe, J., 2011. Negative binomial regression, 2nd editio. ed. Cambridge University Press, Cambridge.

Jartun, M., Ottesen, R.T., Steinnes, E., Volden, T., 2008. Runoff of particle bound pollutants from urban impervious surfaces studied by analysis of sediments from stormwater traps. Sci. Total Environ. 396, 147–163. doi:10.1016/j.scitotenv.2008.02.002

Jenson, S.K., Domingue, J.O., 1988. Extracting Topographic Structure from Digital Elevation Data for Geographic Information System Analysis 54, 1593–1600.

KNMI, 2017. Precipitation - 5-min precipitation accumulations from climatological gauge-adjusted radar dataset for The Netherlands (1 km, extended mask) [WWW Document].

Lager, J., Smith, W., Tchobanoglous, G., 1977. Catchbasin technology overview and assessment. U.S. Environmental Protection Agency, Cincinnati, Ohio.

Mance, G., Harman, M.M.., 1978. The quality of urban stormwater runoff. Urban Storm Drainage, Helliwell (Ed.), Pentech Press 603–618.

Matsuzaki, T., 2017. Understanding the basis of GLM Regression (Logistic, Gaussian, Gamma, etc) [WWW Document]. URL https://tsmatz.wordpress.com/2017/08/30/glm-regression-logistic-poisson-gaussian-gamma-tutorial-with-r/

McDonald, R.I., Weber, K., Padowski, J., Flörke, M., Schneider, C., Green, P.A., Gleeson, T., Eckman, S., Lehner, B., Balk, D., Boucher, T., Grill, G., Montgomery, M., 2014. Water on an urban planet: Urbanization and the reach of urban water infrastructure. Glob. Environ. Chang. 27, 96–105. doi:10.1016/j.gloenvcha.2014.04.022

Medcalc, 2018. Values of the t-distribution (two-tailed) [WWW Document]. URL https://www.medcalc.org/manual/t-distribution.php

Memon, F., Butler, D., 2002. Assessment of gully pot management strategies for runoff quality control using a dynamic model. Sci. Total Environ. 295, 115–129.

Miller, J.D., Hutchins, M., 2017. The impacts of urbanisation and climate change on urban flooding and urban water quality: A review of the evidence concerning the United Kingdom. J. Hydrol. Reg. Stud. 12, 345–362. doi:10.1016/j.ejrh.2017.06.006

Miller, J.D., Kim, H., Kjeldsen, T.R., Packman, J., Grebby, S., Dearden, R., 2014. Assessing the impact of urbanization on storm runoff in a peri-urban catchment using historical change in impervious cover. J. Hydrol. 515, 59–70. doi:10.1016/j.jhydrol.2014.04.011

Ministry of Infrastructure and the Environment, 2011. 2011 Administrative Agreement on Water Affairs.

Minitab, 2012. Why You Need to Check Your Residual Plots for Regression Analysis: Or, To Err is Human, To Err Randomly is Statistically Divine [WWW Document]. URL http://blog.minitab.com/blog/adventures-in-statistics-2/why-you-need-to-check-your-residual-plots-for-regression-analysis

Mobron, N., 2018. Factors influencing sediment load to gully pots. TU Delft.

Montgomery, D.C., Peck, E.A., 1992. Introduction to Linear Regression Analysis, Wiley. Wiley, New York. doi:10.1111/insr.12020_10

Nelder, J.A., Wedderburn, R.W.M., 1972. Generalized Linear Models. J. R. Stat. Soc. 13517213, 370–384. doi:10.2307/2344614

Osborne, M., Butler, D., Clarke, P., Memon, F., 1998. Management of gully pots for improved runoff quality. CIRIA, London.

Phillips, N.D., 2018. Regression on non-Normal data with GLM() [WWW Document]. URL https://bookdown.org/ndphillips/YaRrr/regression-on-non-normal-data-with-glm.html

Post, J.A.., 2016. A statistical approach to guide the management of the anterior part of the sewer system. TU Delft. doi:10.4233/uuid:8cb16d1a-44e0-4798-b638-88b5b74d6bd6

Post, J.A.B., Pothof, I.W.M., Dirksen, J., Baars, E.J., Langeveld, J.G., Clemens, F.H.L.R., 2016. Monitoring and statistical modelling of sedimentation in gully pots. Water Res. 88, 245–256. doi:10.1016/j.watres.2015.10.021

Pratt, C.J., Adams, J.R.W., 1984. Sediment supply and transmission via roadside gully pots. Sci. Total Environ. 33, 213–224. doi:10.1016/0048-9697(84)90395-4

Pratt, C.J., Elliott, G.E.P., Fulcher, G.A., 1987. Suspended solids discharge from highway gully pots in a residential catchment. Sci. Total Environ. 59, 355–364. doi:10.1016/0048-9697(87)90459-1

Rodríguez, G., 2017. Generalized Linear Model [WWW Document]. URL http://data.princeton.edu/wws509/notes/a2s1.html

Runhaar, H., Mees, H., Wardekker, A., van der Sluijs, J., Driessen, P.P.J., 2012. Adaptation to climate change-related risks in Dutch urban areas: Stimuli and barriers. Reg. Environ. Chang. 12, 777–790. doi:10.1007/s10113-012-0292-7

Sartor, J.D., Boyd, G.., 1972. Water Pollution Aspects of Street Surface Contaminants. U.S. Environ. Prot. Agency, USA.

Scott, K.M., 2012. Investigating Sustainable Solutions for Roadside Gully Pot Management 246.

Silvagni, G., Volpi, F., 2002. A model able to estimate the phenomenon of silting of roadside gully pots, in: 2nd International Conference “New Trends in Water and Environmental Engineering for Safety and Life: Eco-Compatible Solutions for Aquatic Environments.” Capri.

Statsdirect, 2018. P value [WWW Document]. URL https://www.statsdirect.com/help/basics/p_values.htm

Sutton, A.J., Abrams, K.R., Jones, D.R., Sheldon, T.A., Fujian, S., 2000. Methods for Meta-Analysis in Medical Research. Wiley.

Tarboron, G., 1997. A new method for the determination of flow directions and upslope areas in grid digital elevation models 33, 309–319.

ten Veldhuis, J.A.E., Clemens, F.H.L.R., 2011. The efficiency of asset management strategies to reduce urban flood risk. Water Sci. Technol. 64, 1317–1324. doi:10.2166/wst.2011.715

ten Veldhuis, J.A.E., Clemens, F.H.L.R., van Gelder, P.H.A.J.M., 2011. Quantitative fault tree analysis for urban water infrastructure flooding. Struct. Infrastruct. Eng. 7, 809–821. doi:10.1080/15732470902985876

Turner, H., 2008. Introduction to Generalized Linear Models [WWW Document]. URL http://statmath.wu.ac.at/courses/heather_turner/glmCourse_001.pdf

UN, 2008. United Nations expert group meeting on Population Distribution, Urbanization, Internal Migration and Development. United Nations Popul. Div.

UN, 2006. World Urbanization Prospects: The 2005 Revision. doi:10.2307/2808041

Welsing, P.M.J., 2015. Statistical modelling: essentially, all models are wrong, but some are useful. Rheumatology 54, 1133–1134. doi:10.1093/rheumatology/kev116

Wheater, H., Evans, E., 2009. Land Use Policy Land use , water management and future flood risk 251–264. doi:10.1016/j.landusepol.2009.08.019

Xanthopoulos, C., Augustin, A., 1992. Input and Characterization of Sediments in Urban Sewer Systems. Water Sci. Technol. 25, 21–28.

Zon, N. Van Der, 2013. Kwaliteitsdocument AHN2.

Zuur, A.F., Ieno, E.N., Elphick, C.S., 2010. A protocol for data exploration to avoid common statistical problems. Methods Ecol. Evol. 1, 3–14. doi:10.1111/j.2041-210X.2009.00001.x

Appendix

Appendix 0 - Self assessment

Choice of graduate internship and activities

Based on the great experiences I made during my second internship at the University of Aveiro, I also wanted to carry out a study at a research institution for my graduate internship. Therefore, I chose TU Delft in the Netherlands as the place for my final thesis. Urban water management has been an area of fundamental interest to me, which lead to choosing TU Delft to conduct my research. Under supervision of Matthijs Rietveld, I started conducting my research on 'The influence of environmental and gully pot specific factors on the solid deposition in gully pots' from the 12th February 2018 until the 7th June 2018. The first few weeks included primarily desk research and measurement activities. During the course of the graduate internship the focus shifted more towards the programming of the Generalized Linear Model (GLM) in Python, also including additions aspects such as the correlation matrix, the variation inflation factor and the visualization of data. Measurement activities were conducted mostly in three week intervals and the gathered data was used to feed the statistical model. Overall, the entire study comprised of the following activities: desk research, data acquisition and data analysis.

Research approach and difficulties

I have to admit that I was intimidated by the name of TU Delft because it is such highly ranked and prestigious university and one of the leading universities in the field of water worldwide. I doubted myself and I was thinking whether I am capable of successfully completing my BSc thesis at TU Delft. Especially in the beginning I was overwhelmed by the immense quantity of available literature on those topics relevant to my research. The most difficult part was to understand the statistics in this research, due to my unfamiliarity with statistical modelling and the vast variety of available models. I spent a majority of the time and late nights trying to understand statistical modelling. It took me quite a while to grasp the concepts and program the Generalized Linear Model (GLM) in Python. I highly benefited from the minor I followed at Wageningen University as I was able to extensively use the programming language Python (Module: 'Programming in Python') and in combination with linear algebra (Module: 'Mathematics 3') it came in very handy for this research. This research made me realize that programming is a great tool for data analysis and visualization.

Nevertheless, programming tended to take a lot of time because occasionally a minor mistake in the script occurred. Detecting these minor errors would often take several hours and it was sometimes followed by another error in the script. Overall the programming part consisted of a lot of trial and error, making it a lengthy and sometimes also very frustrating process.

Furthermore, the writing of the thesis was a tedious and lengthy process due to the scientific nature of the topic. As aforementioned, I had to conduct a lot of literature research to familiarize myself with the topic. The writing process itself was difficult as many sources were used to prove statements and assumptions. I also struggled with the citing sources in a scientific manner, as this was not covered in depth during my studies at Van Hall Larenstein..

Communication with supervisors

Communication during this research with my external supervisor Matthijs and internal supervisor Ad was very helpful. With Matthijs I had frequent discussions regarding the process of the research and aspects he wanted me to look into more thoroughly. He was always open to questions and he reacted quickly when I sent him mails. With Ad communication was more difficult, which can be explained by the fact that communication was mainly via mail. For Ad I had fewer questions as the process of the research was clear to me. For questions regarding the content of the research Matthijs was my contact person. Nevertheless, I found the feedback Ad gave me very useful and it helped me to improve my report hugely.

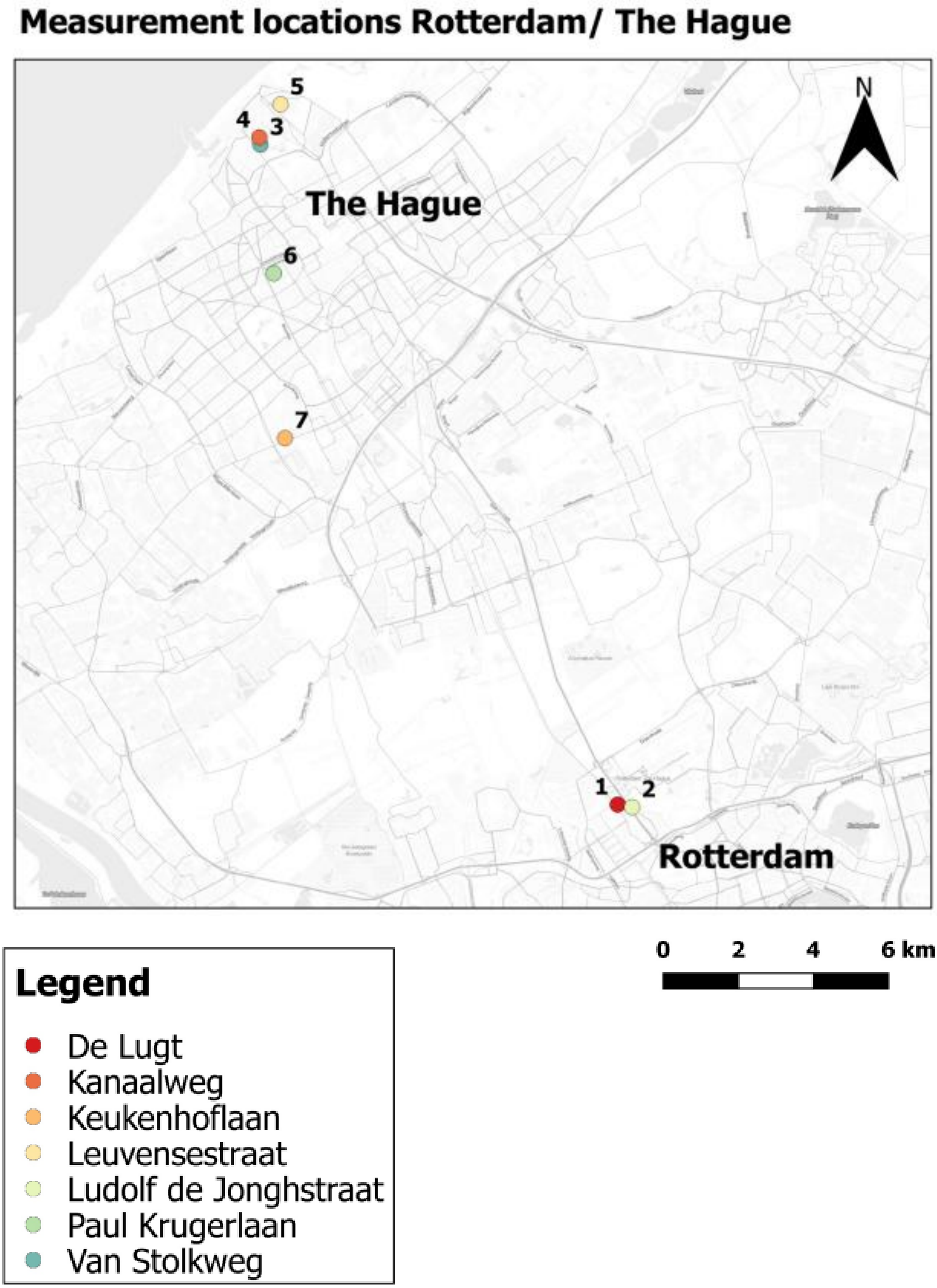
Personal experience

I really cherish the Sanitary Engineering group (people from room 4.64) and especially my supervisor Matthijs for their acceptance and help. Everyone let me feel like I am part of the group, although I was only a Bachelor student amongst PhD students. I had a lot of great talks and discussions with Matthijs who really gave me the freedom to think creatively but also pushed me to scrutinize different behaviours and processes regarding solid deposition in gully pots. He gave me the freedom to bring in and implement my ideas and jointly we were able to include new factors in the model. Especially the newly implemented seasonal variation for the tree factor was one of the major contributions to this research. I personally felt that Matthijs treated me as one of his colleagues, rather than a graduate intern.

I am very pleased I conducted my research at TU Delft as I found it to be the optimal place and environment to conduct my final research. Nevertheless I also realized, in terms of knowledge, that there is still so much I need to learn and can learn by looking at all the other PhD students and professors I encountered during my graduate internship. I feel I have only scratched the surface of what I can accomplish and I am eager to learn much more in the context of Watermanagement. This graduate internship fully convinced me to broaden my horizon and deepen my knowledge and therefore I have decided to follow a Master programme at TU Delft.

Furthermore, this research also highlighted my time management skills. I did not encounter any problems in regards to completing this report on time. Considering the nature of this research, where fixed office hours are not necessarily required, I realized I am yet able to work independently and in a structured manner. Looking at the future, this will certainly be an ability I will benefit from.

Appendix 1 - Measurement locations Rotterdam and The Hague



Appendix 2 - Impression monitoring streets

Paul Krugerlaan

Keukenhoflaan

Kanaalweg

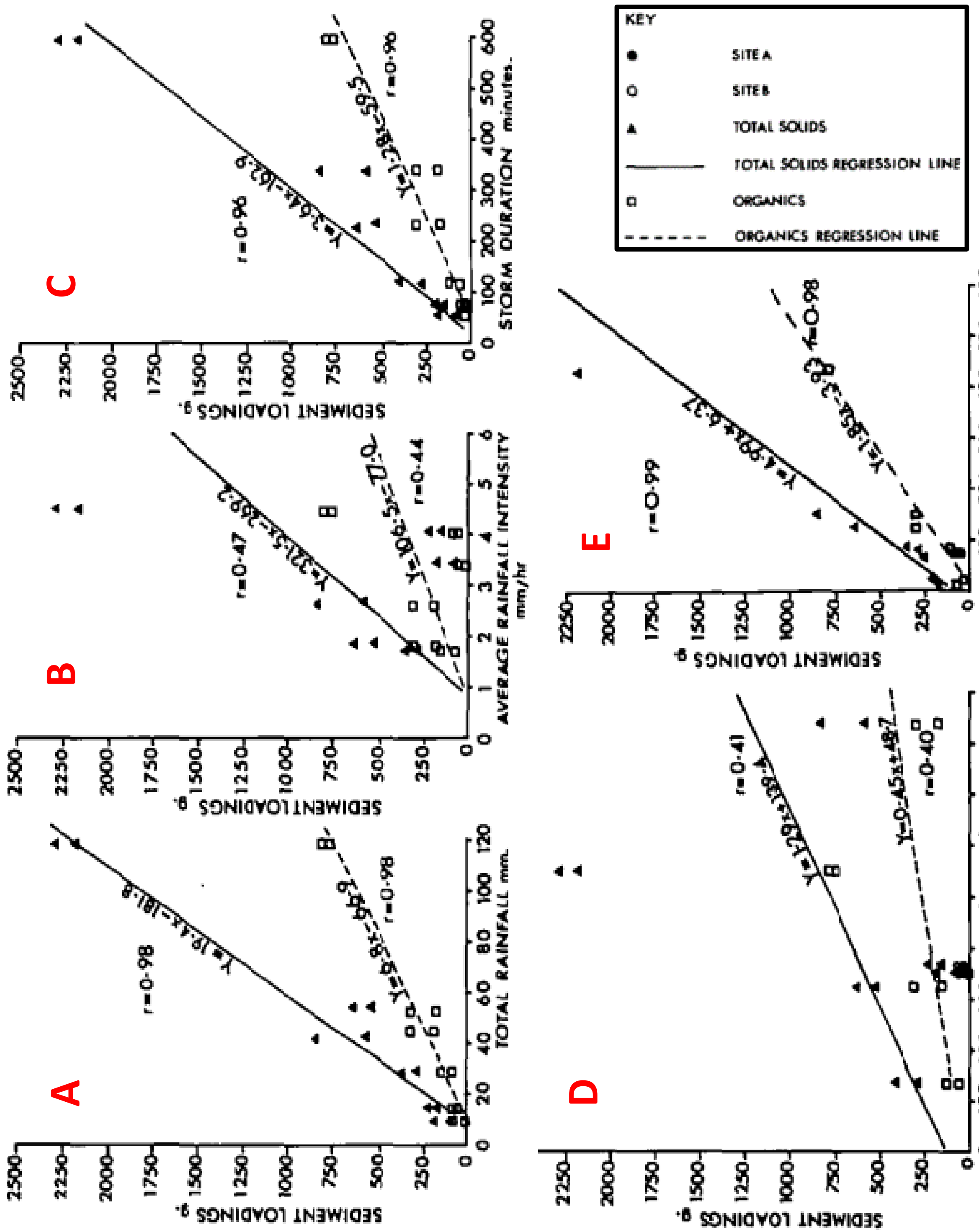
Van Stolkweg

Leuvensestraat

Ludolf de Jonghstraat

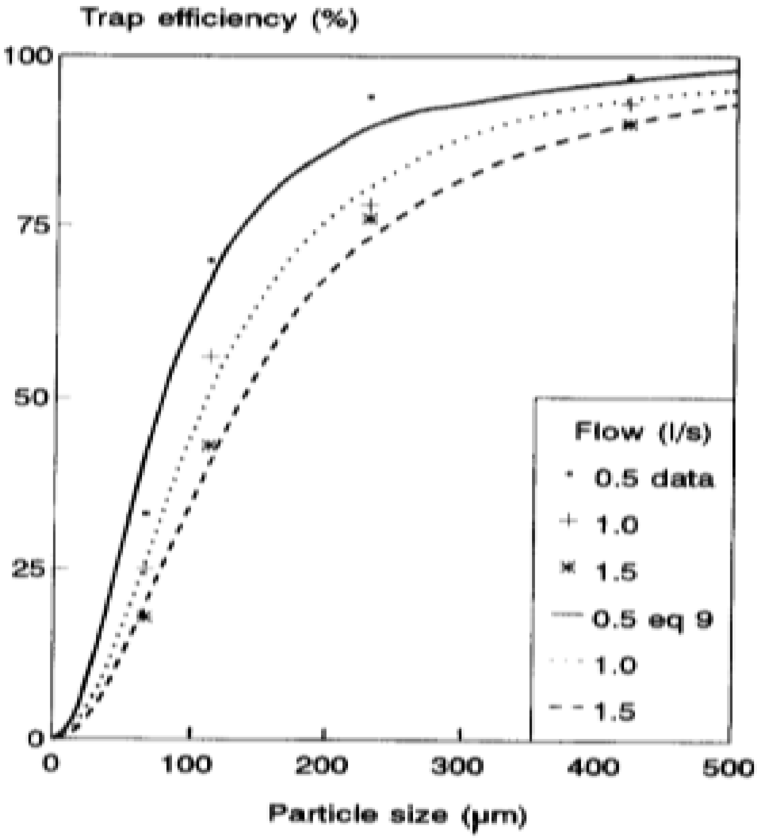
De Lugt

Appendix 3 - Hydrological characteristics and sediment loadings



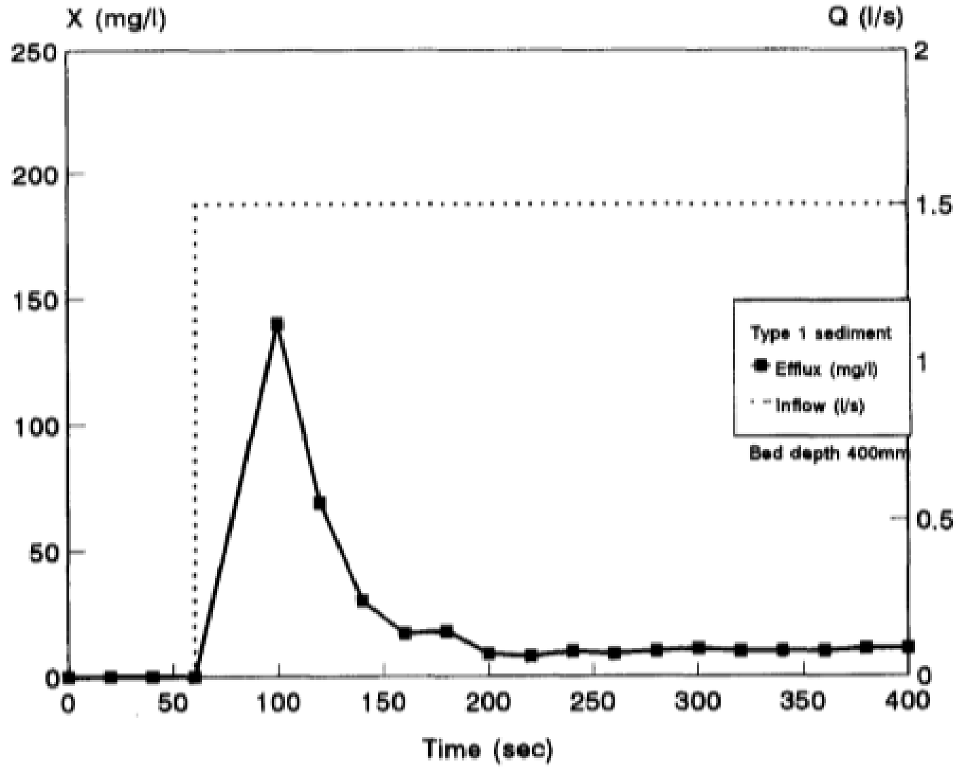
(Ellis and Harrop, 1984)

Appendix 4 - Relationship between trap efficiency, particle size and discharge



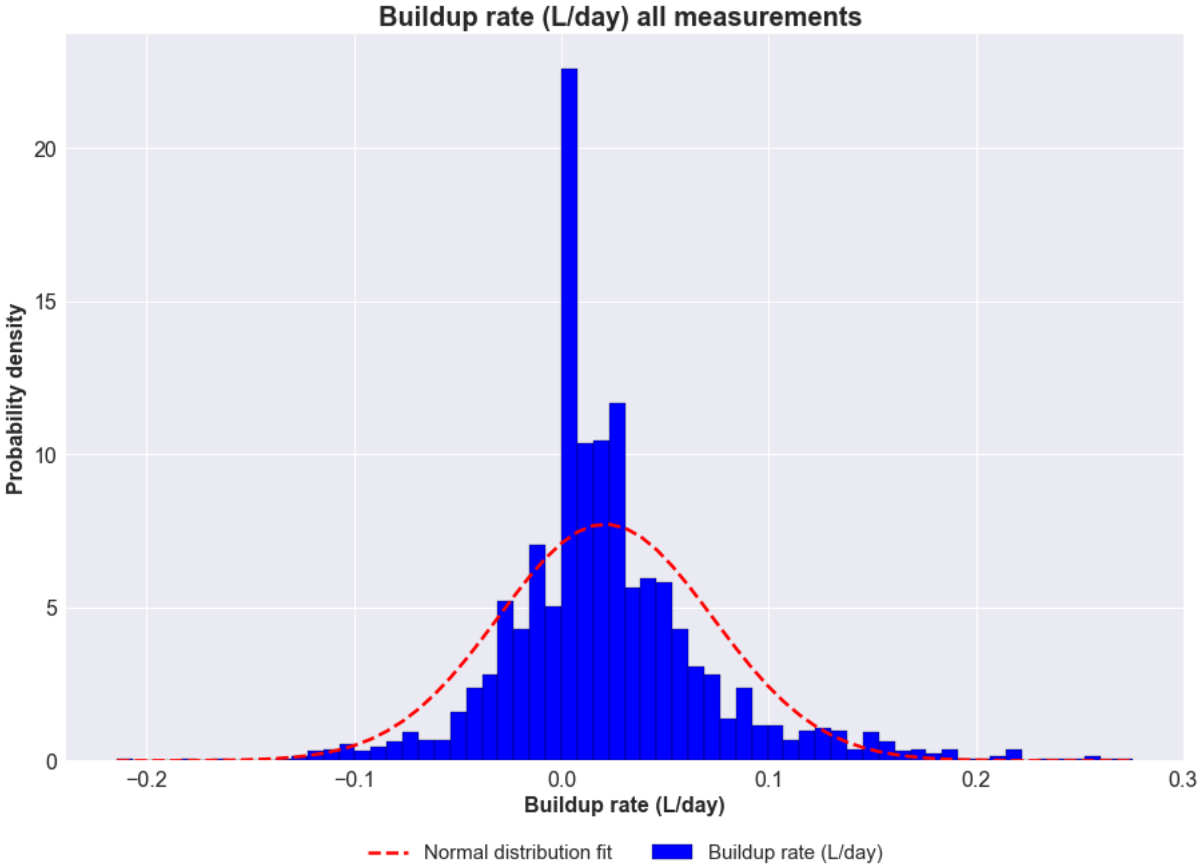
(Butler and Karunaratne, 1995)

Appendix 5 - Concentration of eroded solids in effluent with time



(Butler and Karunaratne, 1995)

Appendix 6 - Normal distribution fit measurement data



Appendix 7 - Source code (Python) for Generalized Linear Model (GLM)

```
import numpy as np
import pandas as pd
import statsmodels.api as sm
from statsmodels import graphics
from statsmodels.graphics.api import abline_plot
from scipy.stats import norm
import matplotlib.mlab as mlab
from scipy import stats
from matplotlib import pyplot as plt
from matplotlib import cm as cm
import seaborn as sns
sns.set()
```

```
# Reading data from CSV file

data = pd.read_excel(r"C:\Users\André\Desktop\Jupyter Notebook\alldata.xlsx", 'Sheet1', index_col = None)
data = data.to_csv(r"C:\Users\André\Desktop\Jupyter Notebook\alldata.csv", encoding = 'utf-8', index = False)
data = pd.read_csv(r"C:\Users\André\Desktop\Jupyter Notebook\alldata.csv", sep = ',', engine = "python")
data = data[pd.notnull(data['Inlet'])]
```

```
# Splitting data into all streets

df1 = data[data['Street'] == 'Van Stolkweg']
df2 = data[data['Street'] == 'Paul Krugerlaan']
df3 = data[data['Street'] == 'Leuvensestraat']
df4 = data[data['Street'] == 'Keukenhoflaan']
df5 = data[data['Street'] == 'Kanaalweg']
df6 = data[data['Street'] == 'Ludolf de Jonghstraat']
df7 = data[data['Street'] == 'De Lugt']
```

```
# VAN STOLKWEG Outliers

mean1 = np.nanmean(df1['Buildup rate (L/day)'])
std1 = np.nanstd(df1['Buildup rate (L/day)'])

o1 = df1.loc[(df1['Buildup rate (L/day)'] > mean1 + 3 * std1)]
o2 = df1.loc[(df1['Buildup rate (L/day)'] < mean1 - 3 * std1)]
out1 = pd.concat([o1, o2])

for value in df1['Buildup rate (L/day)']:
    if (value < (mean1 - 3 * std1)):
        df1 = df1[df1['Buildup rate (L/day)'] != value]
    elif (value > (mean1 + 3 * std1)):
        df1 = df1[df1['Buildup rate (L/day)'] != value]
```

```
# PAUL KRUGERLAAN Outliers

mean2 = np.nanmean(df2['Buildup rate (L/day)'])
std2 = np.nanstd(df2['Buildup rate (L/day)'])

o3 = df2.loc[(df2['Buildup rate (L/day)'] > mean2 + 3 * std2)]
o4 = df2.loc[(df2['Buildup rate (L/day)'] < mean2 - 3 * std2)]
out2 = pd.concat([o3, o4])

for value in df2['Buildup rate (L/day)']:
    if (value < (mean2 - 3 * std2)):
        df2 = df2[df2['Buildup rate (L/day)'] != value]
    elif (value > (mean2 + 3 * std2)):
        df2 = df2[df2['Buildup rate (L/day)'] != value]
```

```
# LEUVENSESTRAAT Outliers

mean3 = np.nanmean(df3['Buildup rate (L/day)'])
std3 = np.nanstd(df3['Buildup rate (L/day)'])

o5 = df3.loc[(df3['Buildup rate (L/day)'] > mean3 + 3 * std3)]
o6 = df3.loc[(df3['Buildup rate (L/day)'] < mean3 - 3 * std3)]
out3 = pd.concat([o5, o6])

for value in df3['Buildup rate (L/day)']:
    if (value < (mean3 - 3 * std3)):
        df3 = df3[df3['Buildup rate (L/day)'] != value]
    elif (value > (mean3 + 3 * std3)):
        df3 = df3[df3['Buildup rate (L/day)'] != value]
```

```
# KEUKENHOFLAAN Outliers

mean4 = np.nanmean(df4['Buildup rate (L/day)'])
std4 = np.nanstd(df4['Buildup rate (L/day)'])

o7 = df4.loc[(df4['Buildup rate (L/day)'] > mean4 + 3 * std4)]
o8 = df4.loc[(df4['Buildup rate (L/day)'] < mean4 - 3 * std4)]
out4 = pd.concat([o7, o8])

for value in df4['Buildup rate (L/day)']:
    if (value < (mean4 - 3 * std4)):
        df4 = df4[df4['Buildup rate (L/day)'] != value]
    elif (value > (mean4 + 3 * std4)):
        df4 = df4[df4['Buildup rate (L/day)'] != value]
```

```python
# KANAALWEG Outliers

mean5 = np.nanmean(df5['Buildup rate (L/day)'])
std5 = np.nanstd(df5['Buildup rate (L/day)'])

o9 = df5.loc[(df5['Buildup rate (L/day)'] > mean5 + 3 * std5)]
o10 = df5.loc[(df5['Buildup rate (L/day)'] < mean5 - 3 * std5)]
out5 = pd.concat([o9, o10])

for value in df5['Buildup rate (L/day)']:
    if (value < (mean5 - 3 * std5)):
        df5 = df5[df5['Buildup rate (L/day)'] != value]
    elif (value > (mean5 + 3 * std5)):
        df5 = df5[df5['Buildup rate (L/day)'] != value]
```

```python
# LUDOLF DE JONGHSTRAAT Outliers

mean6 = np.nanmean(df6['Buildup rate (L/day)'])
std6 = np.nanstd(df6['Buildup rate (L/day)'])

o11 = df6.loc[(df6['Buildup rate (L/day)'] > mean6 + 3 * std6)]
o12 = df6.loc[(df6['Buildup rate (L/day)'] < mean6 - 3 * std6)]
out6 = pd.concat([o11, o12])

for value in df6['Buildup rate (L/day)']:
    if (value < (mean6 - 3 * std6)):
        df6 = df6[df6['Buildup rate (L/day)'] != value]
    elif (value > (mean6 + 3 * std6)):
        df6 = df6[df6['Buildup rate (L/day)'] != value]
```

```python
# DE LUGT Outliers

mean7 = np.nanmean(df7['Buildup rate (L/day)'])
std7 = np.nanstd(df7['Buildup rate (L/day)'])

o13 = df7.loc[(df7['Buildup rate (L/day)'] > mean7 + 3 * std7)]
o14 = df7.loc[(df7['Buildup rate (L/day)'] < mean7 - 3 * std7)]
out7 = pd.concat([o13, o14])

for value in df7['Buildup rate (L/day)']:
    if (value < (mean7 - 3 * std7)):
        df7 = df7[df7['Buildup rate (L/day)'] != value]
    elif (value > (mean7 + 3 * std7)):
        df7 = df7[df7['Buildup rate (L/day)'] != value]
```

```python
# Concatenating all filtered datafames

data_filtered = pd.concat([df1,df5,df4,df3,df2,df7,df6])
outliers = pd.concat([out1, out5, out4, out2, out3, out4, out5, out7, out6])
```

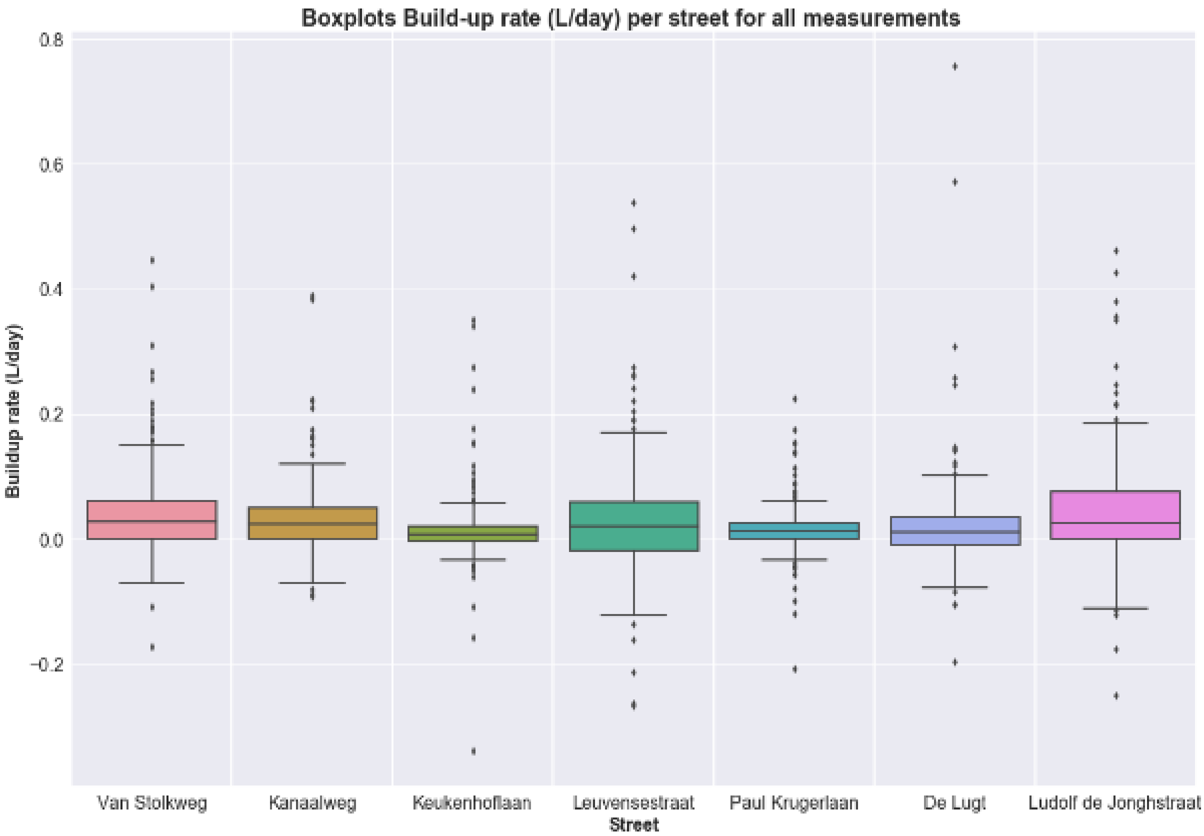
```python
# Splitting data into response and explanatory variables

resp_var = data_filtered[[x for x in data.columns if 'Buildup rate (L/day)' in x]].values.flatten()
x1 = data_filtered[[x for x in data.columns if 'Connected area (m^2)' in x]].values.flatten()
x2 = data_filtered[[x for x in data.columns if 'Traffic intensity (mv/day)' in x]].values.flatten()
x3 = data_filtered[[x for x in data.columns if 'Effective sand trap volume' in x]].values.flatten()
x5 = data_filtered[[x for x in data.columns if 'Shops' in x]].values.flatten()
x10 = data_filtered[[x for x in data.columns if 'Tree number 2 leaf fall' in x]].values.flatten()
x11 = data_filtered[[x for x in data.columns if 'Tree number 2 no leaf' in x]].values.flatten()
x12 = data_filtered[[x for x in data.columns if 'Tree number 2 leaf grow' in x]].values.flatten()
x14_temp = data_filtered[[x for x in data.columns if 'Road surface' in x]].values.flatten()
x15_temp = data_filtered[[x for x in data.columns if 'Inlet' in x]].values.flatten()
```

```python
# Converting strings from data into numerical values

x14 = list()
x15 = list()

for elem in x14_temp:
    if elem == 'Asphalt':
        x14 += [0]
    elif elem == 'Bricks':
        x14 += [1]

x14 = np.asarray(x14)

for elem in x15_temp:
    if elem == 'zij':
        x15 += [0]
    elif elem == 'boven' or elem == 'combi':
        x15 += [1]

x15 = np.asarray(x15)
```

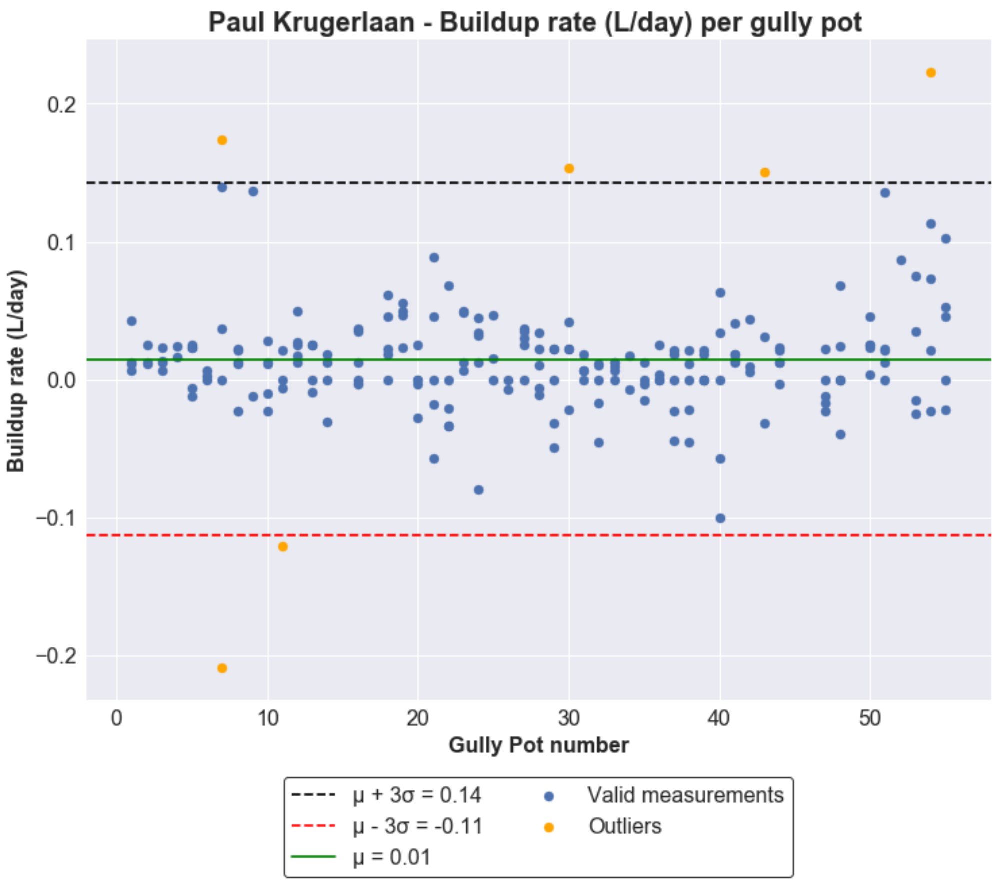
```python
# Stacking all explanatory variables and converting it into required matrix form

x_all = np.vstack((x3, x10, x11, x12, x15))
x_all = np.transpose(x_all)
x_all = sm.add_constant(x_all, prepend = True)
```

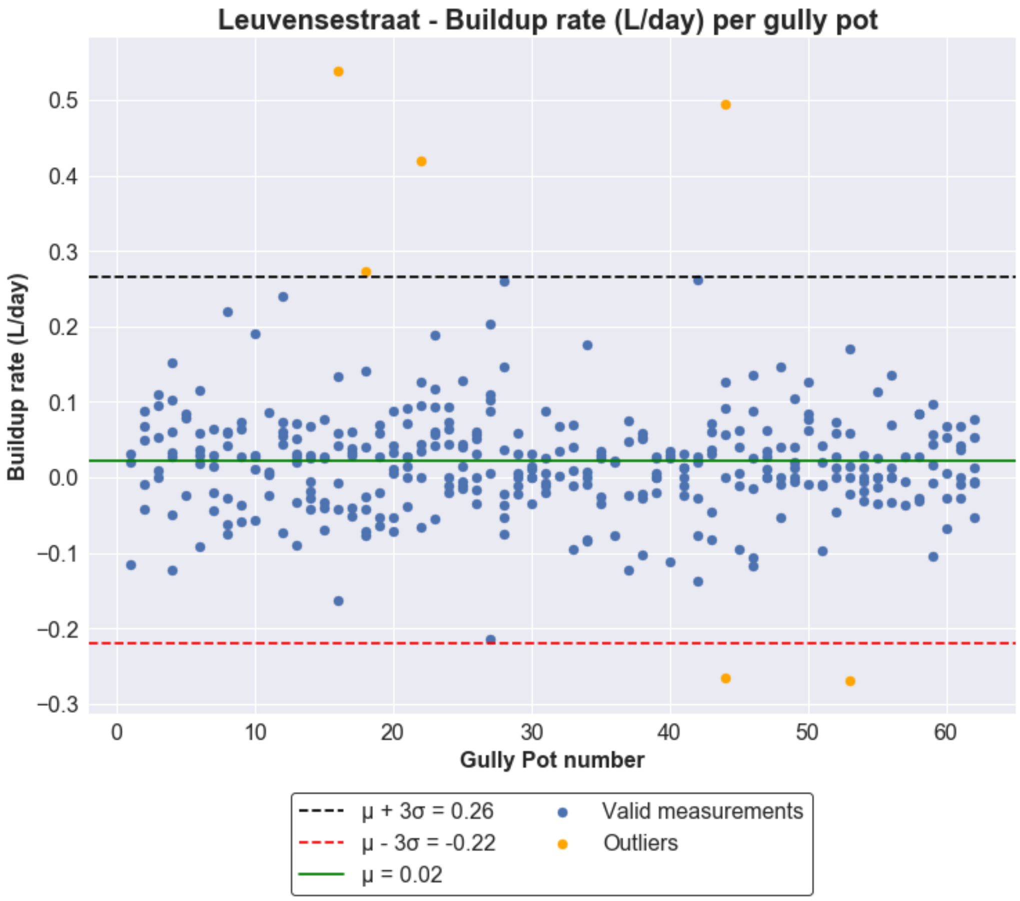
```python
# Applying General Linear Model

glm_normal = sm.GLM(resp_var, x_all, family = sm.families.Gaussian(sm.families.links.identity))
res = glm_normal.fit()
res1 = glm_normal._fit_gradient()
const, b3, b10, b11, b12, b15 = res.params
print(res.summary())
```

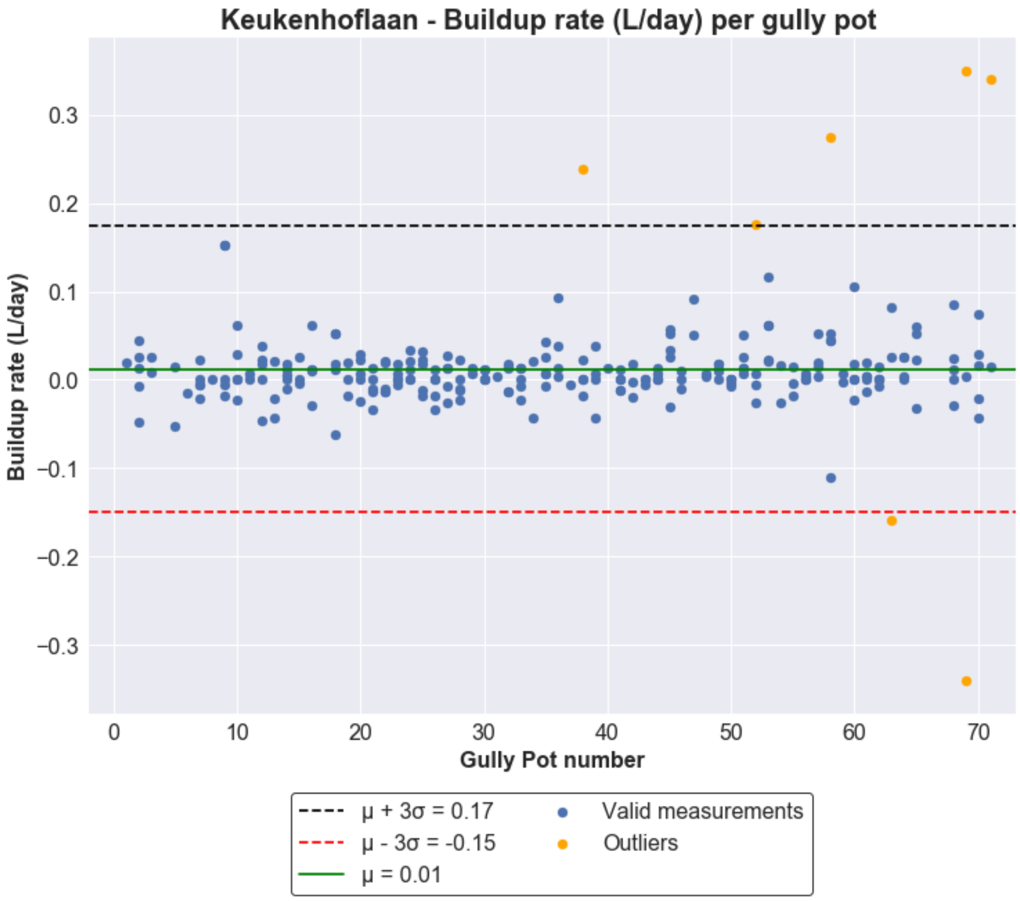
Appendix 8 - Boxplots Build-up rate (L/day) per street for all measurements

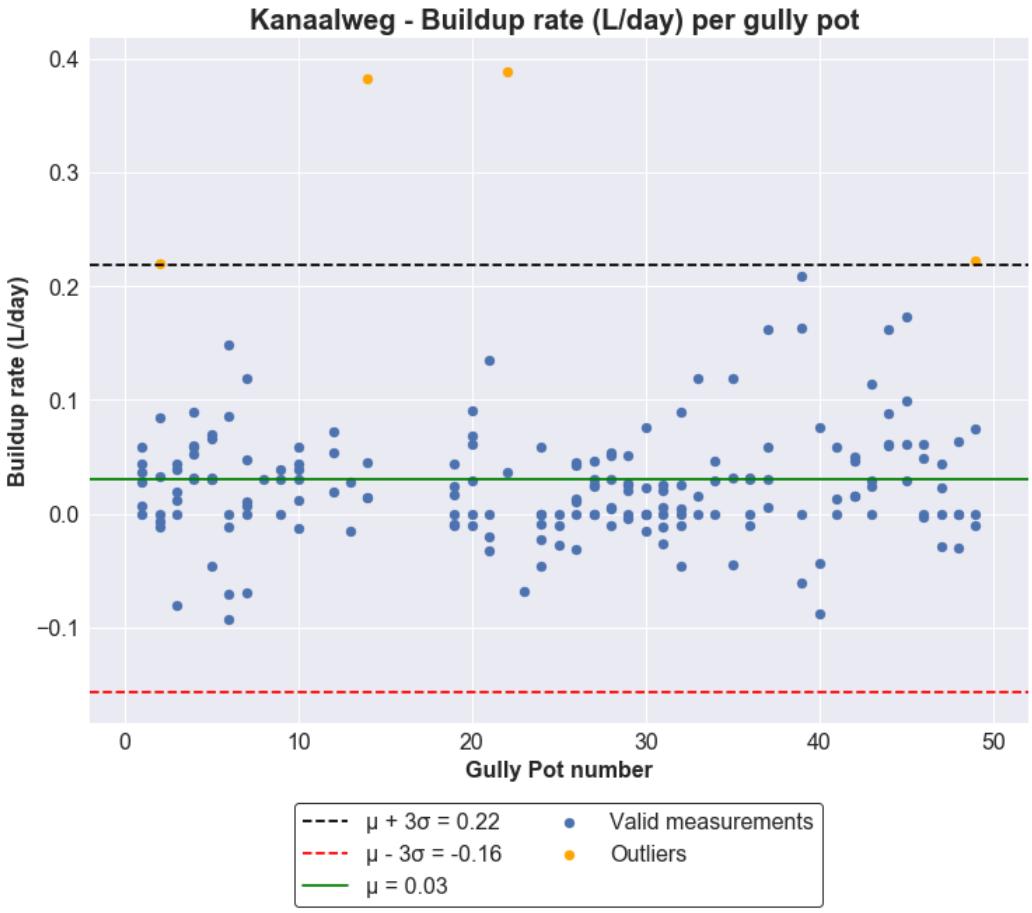


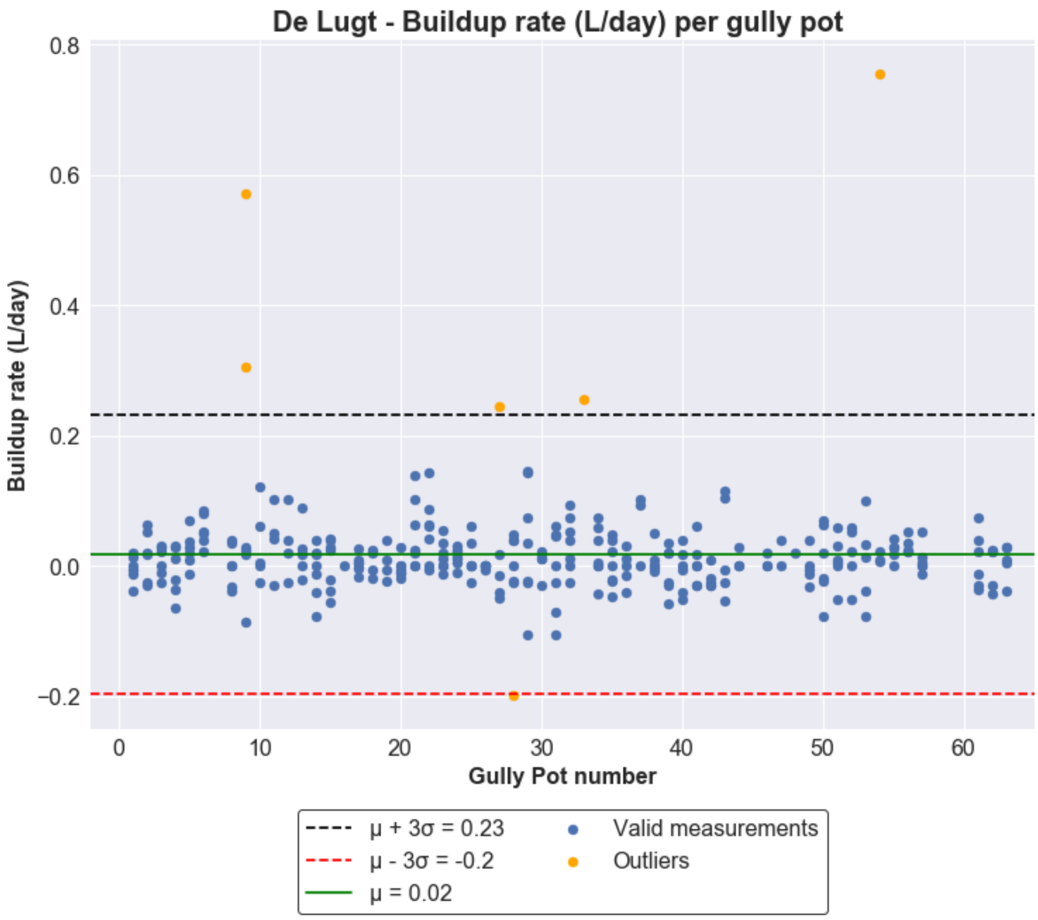
Appendix 9 - Outlier detection and removal per street

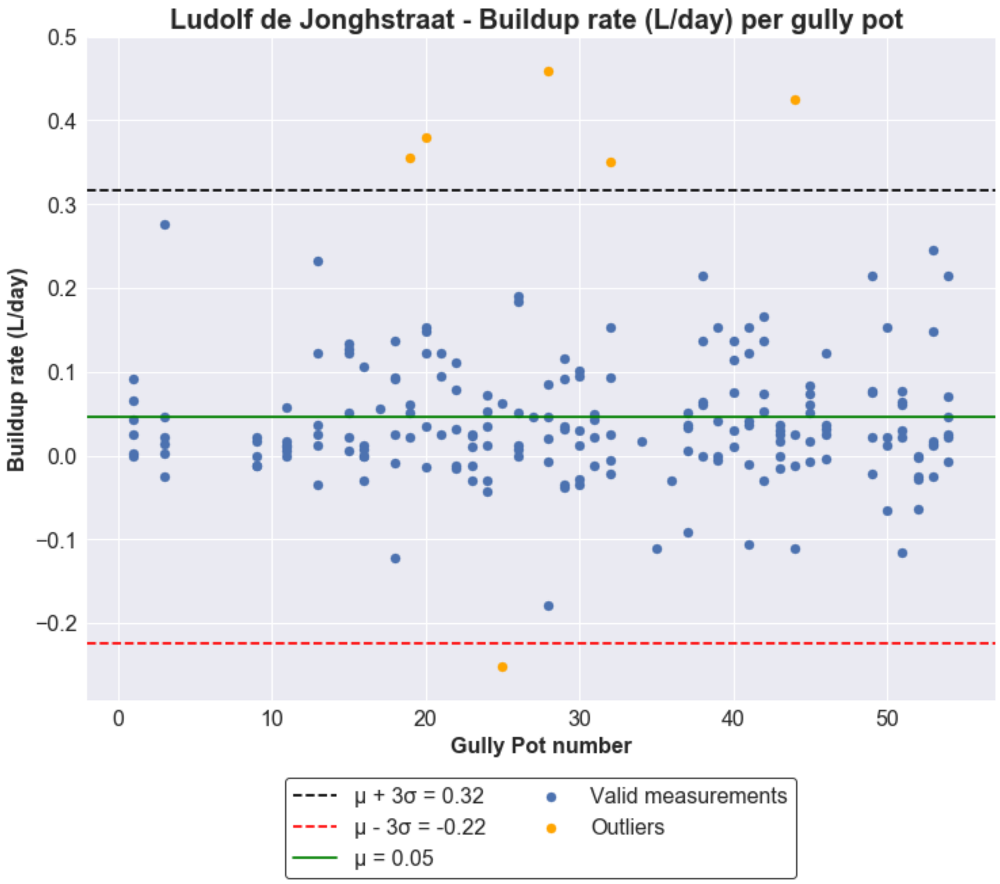



Leuvensestraat - Buildup rate (L/day) per gully pot
Buildup rate (L/day)
Gully Pot number
μ + 3σ = 0.26
μ - 3σ = -0.22
μ = 0.02
Valid measurements
Outliers

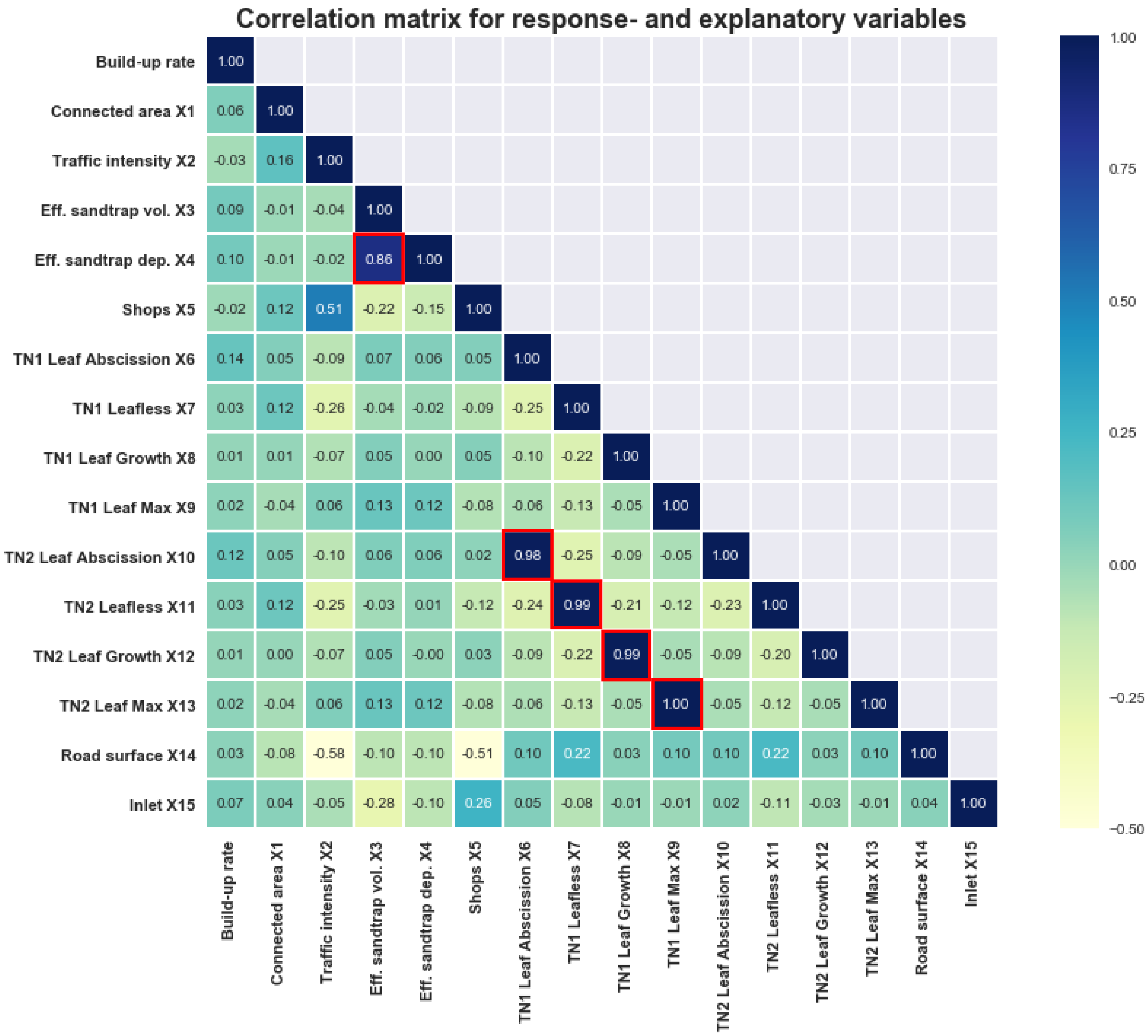

Keukenhoflaan - Buildup rate (L/day) per gully pot
Buildup rate (L/day)
Gully Pot number
μ + 3σ = 0.17
μ - 3σ = -0.15
μ = 0.01
Valid measurements
Outliers


Kanaalweg - Buildup rate (L/day) per gully pot
Buildup rate (L/day)
0.4
0.3
0.2
0.1
0.0
−0.1
0
10
20
30
40
50
Gully Pot number
μ + 3σ = 0.22
μ - 3σ = -0.16
μ = 0.03
Valid measurements
Outliers

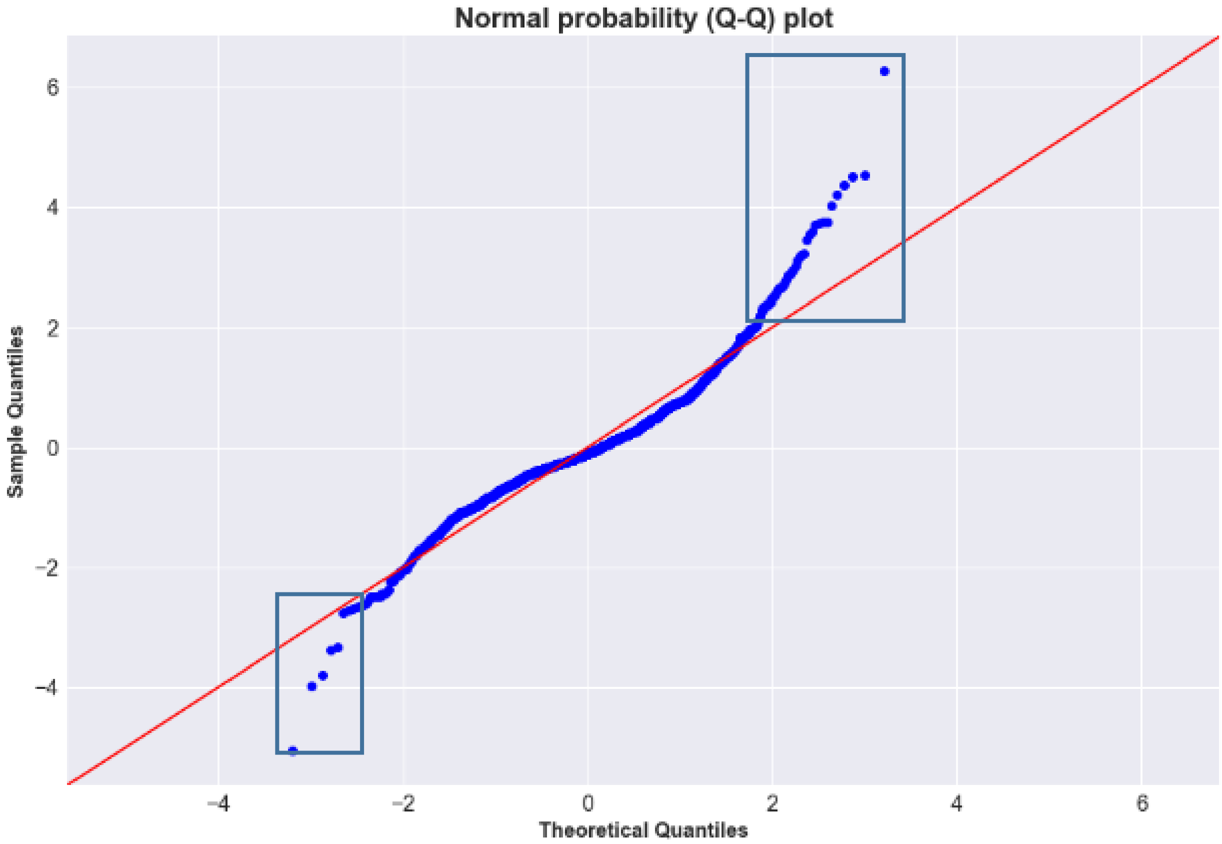

De Lugt - Buildup rate (L/day) per gully pot
Buildup rate (L/day)
0.8
0.6
0.4
0.2
0.0
−0.2
0
10
20
30
40
50
60
Gully Pot number
μ + 3σ = 0.23
μ - 3σ = -0.2
μ = 0.02
Valid measurements
Outliers


Ludolf de Jonghstraat - Buildup rate (L/day) per gully pot
Buildup rate (L/day)
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
0
10
20
30
40
50
Gully Pot number
μ + 3σ = 0.32
μ - 3σ = -0.22
μ = 0.05
Valid measurements
Outliers

Appendix 10 - Correlation matrix response- and explanatory variables



Appendix 11 - Normal probability (Q-Q) plot



Appendix 12 - Standardized residuals vs. fitted values