Universität für Bodenkultur Wien
University of Natural Resources and Life Sciences, Vienna

Department of Water, Atmosphere and Environment
Institute of Hydrobiology and Aquatic Ecosystem Management

DESIGN OF VEGETATED BUFFER STRIPS FOR USE IN PHOSPHORUS RETENTION FROM NON-POINT AGRICULTURAL SOURCES.

IDENTIFICATION OF KEY PARAMETERS AND ANALYSIS OF PRACTICAL IMPLICATIONS

Master thesis
in partial fulfilment of the requirements for the degree of
Diplomingenieur
in
Water Management and Environmental Engineering

executed at the
Institute for Water Quality, Resource and Waste Management
Vienna University of Technology

under the supervision of
Ao.Univ.Prof. Matthias Zessner

submitted by:
Oriol MOLLÓ MANONELLES
0941666

Vienna,
9. Oktober 2017

Signature

_____

Acknowledgements

Table of contents

List of tables

List of figures

Abstract

The use of vegetated buffer strips to control phosphorus transfers from agricultural fields to surface waters is an established and accepted practice to improve water quality of freshwater bodies. However, its long-term performance under field conditions is not yet well understood.

In this study, current literature was reviewed to identify key parameters affecting buffer performance and data from 20 previous experiments was analysed to find out if long-term buffer performance under field conditions is any different from results observed in single-event experimental plots.

The results showed an average TP retention of 69% in experimental plots compared to just 58% in cultivated fields, pointing to a 11% difference in performance. We conclude this difference is related to spatial phenomena like flow concentration and buffer bypass, which have been observed to limit buffer function under field conditions but are seldom promoted in experimental plots. Differences due to the time frame of observations were also detected: on average, a performance of 70% was achieved when considering single runoff-generating events, whereas for long-term multi-annual observations performance dropped to 51%, suggesting that buffer strips lose retention capacity with time.

Further, buffer width, slope and soil texture together with width ratio were identified as the best predictors of buffer performance. A regression model built with the variables buffer slope and width was able to explain up to 60% of the sample variance for field conditions. However, due to the low number of records used to adjust the model and lack of validation, we recommend it to be used for obtaining a first performance approximation rather than as a robust planning tool.

We conclude that although vegetated buffer strips successfully contribute to reducing phosphorus transfers, their long-term performance under field conditions is less than what is expected from experiments in plot studies.

Kurzzusammenfassung

Der Einsatz von Gewässerrandstreifen zur Kontrolle des Phosphoreintrags aus landwirtschaftlich genutzten Flächen in Oberflächengewässer gilt als etablierte Maßnahme zur Verbesserung der Wasserqualität. Ihre langfristige Wirksamkeit unter realen Bedingungen ist jedoch bisher unzureichend erforscht.

In dieser Arbeit wurden durch umfangreiche Literaturrecherche Schlüsselparameter für die Wirksamkeit von Gewässerrandstreifen identifiziert. Die Daten von 20 ausgewählten Experimenten wurden auf Unterschiede in der langfristigen Wirksamkeit unter Feldbedingungen mit jener in Versuchsflächen mit Einzelereignissen untersucht.

Die Ergebnisse zeigten eine durchschnittliche Phosphor-Retention von 69% in Versuchsflächen, jedoch nur 58% in Feldbedingungen. Der Unterschied von rund 11% scheint wesentlich durch räumliche Faktoren wie Abflusswege, die an Gewässerrandstreifen vorbeiführen erklärt. Diese verringern die Wirksamkeit von Filterstreifen, werden aber in Versuchsflächenstudien kaum berücksichtigt. Weiters wurde eine tendenzielle Abnahme der Retentionskapazität im Zeitverlauf beobachtet: bei Einzelereignissen betrug die durchschnittliche Wirksamkeit 70%, bei langfristiger mehrjähriger Beobachtung nur 51%.

Als relevanteste Parameter für die Effektivität von Filterstreifen wurden Filterbreite, Hangneigung, Bodentextur sowie Breiteverhältnis identifiziert. Anhand eines multiplen Regressionsmodelles konnten bis zu 60% der Stichproben-Varianz unter Feldbedingungen erklärt werden. Aufgrund der geringen Anzahl von Datensätze sowie mangels Validierung dient das Modell jedoch eher einer initialen groben Schätzung der potentiellen Wirksamkeit des Filterstreifens als einer exakten Planung.

Abschließend kann gefolgert werden, dass Gewässerrandstreifen eine erfolgreiche Maßnahme zur Reduktion des Phosphoreintrages von landwirtschaftlich genutzten Flächen darstellen, ihre langfristige Wirksamkeit jedoch unter den von Versuchsflächenstudien abgeleiteten Erwartungswerten liegt.

Abbreviations

AIC   Akaike Information Criterion

ANOVA Analysis of variance

DF   Degrees of freedom

DP   Dissolved Phosphorus

NCRS   Natural Resource and Soil Conservation Services

MLR   Multiple Linear Regression

OLS   Ordinary Least Squares

ÖPUL   Österreichische Programm für umweltgerechte Landwirtschaft

PP   Particle Phosphorus

TP   Total Phosphorus

TSS   Total Suspended Solids

USDA   United Stated Department of Agriculture

VBS   Vegetated Buffer Strip

VFS   Vegetated Filter Strip

VIC   Variance Inflation Factor

WFD   Water Framework Directive

1 Introduction

A vegetated buffer strip is an area of vegetation maintained down slope of pollutant sources to remove sediments, nutrients and other materials from incoming runoff before they reach freshwater bodies (I. Chaubey et al., 1995; T.A. Dillaha et al., 1989).

Vegetated buffer strips are not a new measure. Already in the 1970s they were in use in civil engineering to control erosion from open pit mining, road works or forest soil restoration (Haycock, 1997). Studies back then focused mainly on the suitability of grassed buffers in controlling soil erosion and trapping sediments.

With increasing environmental pressure of agriculture on freshwater bodies, the measure was contemplated in the early 1980s as a way to control pollution from nutrient-enriched agricultural runoff to surface water. Since then, vegetated buffer strips and related types of buffer zones, like field margins and riparian forests, have been the subject of interest of many studies on the mechanisms and efficiency of these structures in reducing nutrient transfer from runoff-generating events. Many experimental trials confirm their effectiveness in trapping nutrients, and in particular phosphorus, during runoff-generating events (Abu-Zreig et al., 2004; Blanco-Canqui et al., 2006; Daniels and Gilliam, 1996; Patty et al., 1997; Yang et al., 2015).

Nowadays, the use of vegetated buffer strips is an established management tool for controlling sediments and nutrients yield from agricultural fields and is viewed as a practical, low-cost management tool for improving the quality of runoff from source areas (Lim et al., 1998). Environmental agencies and policy makers of many countries have incorporated the measure in their programs as an effort to improve surface water quality. The US Department of Agriculture already introduced the measure for field erosion control 30 years ago through the Natural Resources Conservation Services (NCRS) and keeps adapting the requirements for the establishment of buffer strips as needed every few years (Natural Resources Conservation Service, 2008). In Europe, since the introduction of the European Water Framework Directive in year 2000 many Member States implemented subsidies on vegetated buffers as part of landscape management programmes (Moreddu, 2011). Austria has been funding the installation of vegetated filter strips alongside permanent streams through the landscape management program (ÖPUL) since 2007 (Hösl et al., 2012).

Yet, there is a lack of experiments on the long-term performance of buffers under field conditions. The few existing studies suggest that the efficiency is less when compared to single-event plot-scale studies (Daniels and Gilliam, 1996; Uusi-Kämppä, 2008). The explanation to these differences is still not quite clear. In fact, Sheppard et al. (2006) go even further to say that VBS use is perhaps advocated without sufficient evidence of value.

So, what can we really expect from vegetated buffer strips as a long-term measure to control phosphorus emissions to surface waters under long-term field conditions? This central question motivated a literature review, which laid the basis for the present master thesis.

2 Objectives

The present study investigates the effect of identified key parameters on phosphorus retention efficiency of vegetated buffer strips based on experimental data from literature and compares the performance between experimental plots and field conditions, proposing a multiple regression model to assess buffer performance under long-term field conditions. It aims at improving the comprehension of VBS systems and drawing practical implications to help environmental managers and decision-makers to achieve a better implementation of the measure.

3 Fundamentals

3.1 Overview of the vegetated buffer strip system

Buffer strips function by modifying phosphorus transfers from source fields to surface waters through deposition of particulate phosphorus and infiltration of dissolved phosphorus (Roberts et al., 2012).
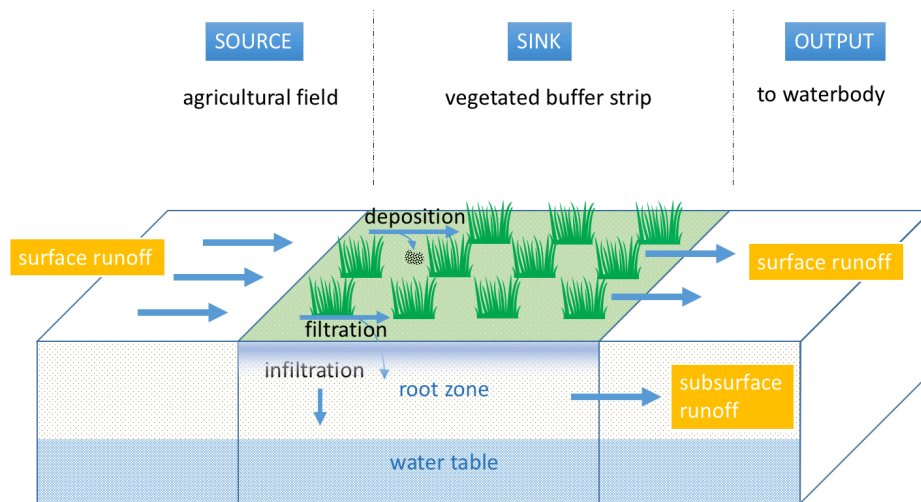
Figure 3.1-1 The vegetated buffer strip system.

At source, phosphorus from fertilizers applied to agricultural fields is detached from the parent soil by runoff during intense precipitation events via two mechanisms. The first of them is physical detachment and involves phosphorus-containing particles being pulled out from parent soil by runoff shear stress and transported adhered to suspended particles in runoff flow. The second is solubilisation of phosphorus from particles of the soil matrix and transport in the runoff solution (Roberts et al., 2012).

Phosphorus-loaded runoff travels from the source fields to the vegetated buffers, where it encounters a rougher and more porous surface causing it to slow down and lose transport capacity. In turn, this increases the contact time between surface flow and the soil-plant system, enhancing runoff infiltration and settling of transported particles at the soil surface. Stems and aerial parts of plants are responsible for the increase in hydraulic roughness, whereas roots form channel-systems where infiltration is enhanced (Doroiz et al., 2006). Under common field conditions not all the runoff is infiltrated at buffer and is released downslope eventually reaching the drainage network.

The deposition of particulate phosphorus at buffer does not follow a random distribution, in fact, there is a granulometric sorting. Coarse particles are deposited at the front and as transport capacity of water is decreased, smaller particles settle, causing a gradient in phosphorus load in the

downslope direction. The consolidation of the deposited particles occurs between periods of runoff, so if runoff episodes follow too frequently one another, sediment deposited but not consolidated can be remobilised by subsequent erosion (Dorioz et al., 2006).

After the initial deposition and infiltration, phosphorus is actively adsorbed to soil particles or locked in pores with interstitial water. Following these initial physical fast processes, phosphorus enters biological or geochemical pools by sorption and assimilation and its cycling is then governed by much slower processes occurring in the time frame of days to month (Roberts et al., 2012).

Finally, phosphorus is stored in long-term biological and geochemical pools, the most relevant of them is vegetation. Plants absorb phosphorus through the roots by releasing exudates and accumulate it as part of their biomass to build new tissues. Humus, decaying organic matter and associated microbes are other important phosphorus pools because of its role in phosphorus cycling. Seasonal changes in soil characteristics such as temperature, pH and aeration affect microbial fauna and can cause sudden release of dissolved phosphorus by leaching phosphates to freshwater bodies (Dosskey et al., 2010).

3.2 Current state of the knowledge

The potential effectiveness of VFS in trapping phosphorus from agricultural runoff is well established in plot studies (Abu-Zreig et al., 2004; I. Chaubey et al., 1995; Schmitt et al., 1999; T.A. Dillaha et al., 1989; Uusi-Kämppä, 2008), showing performances on total phosphorus retention between 30% to 99% in strips 3 m to 50 m wide.

Most of these studies were conducted on regular plots physically separated from its environment under simulated rainfall or simulated runoff to allow for experiment replication. They build the basis for today's understanding on the mechanisms governing buffer performance and served to identify several key influencing factors.

The most important of these factors is filter width. An increase of width is associated to an increase on buffer performance (Blanco-Canqui et al., 2006; Dosskey et al., 2008; T.A. Dillaha et al., 1989; Yang et al., 2015). Another important factor is buffer slope. Most of the experimental studies on buffer slope observe a decrease of buffer performance with increasing slope (Syversen, 2005; T.A. Dillaha et al., 1989; Yang et al., 2015).

Considerable attention has been paid in studying the role of vegetation and plant morphology to buffer performance. Most studies have been conducted on grassed buffers rather than on forested buffers but there is no conclusive answer as if the former or the latter perform better (Daniels and Gilliam, 1996; Mayer et al., 2007; Yuan et al., 2009).

Soil characteristics has been the object of intensive research as well. Schmitt et al. (1999) came up to the conclusion that the best retention performance is achieved under average soil texture conditions. Whereas White and Arnold (2009) associated clay-rich soils retained almost no runoff

in buffer strips at moderate runoff loadings. Concerning the runoff inflow, higher sediment trapping efficiencies are associated with decreased flow rates (Abu-Zreig et al., 2004). Finally, Yang et al. (2015) argue that a higher suspended solid concentration on inflow runoff yields higher TP removal efficiencies. A more in-depth review of the key factors affecting buffer performance at plot level is given in Chapter 3.4.

Experimental studies on long-term field conditions are less abundant than single-event plot studies and they point out to smaller performances as well. One of the longest-term studies to date was performed by Uusi-Kamppa (2008) who analysed data on buffer performance from 1992 to 2002. She reported a 40% decrease of total P loss through surface runoff compared to plots without buffer zones for the period. Borin et al. (2005) even observed smaller performance on a 4-year period reporting just a 22% of total P reduction.

Studies on the impact of buffer strips at watershed level are rare. The difficulty of designing experiments linking phosphorus reduction at watershed outlet and performance of single buffer strips might probably be the reason behind it. Gilliam (1987) measured the phosphorus redistribution in sediment deposited within riparian areas and reported about 50% of the total P leaving agricultural fields appeared to be removed from the runoff water in the riparian areas. That said, the riparian areas did not merely include buffer strips but also ephemeral streams and flood plains. Meals and Hopkins (2002) evaluated the effectiveness of riparian restoration in two agricultural watersheds for a period of 2 years. They monitored water quality and streamflow at watershed outlet, documenting reductions of 20% in mean total P concentration and between 20% to 50% in mean total P load in streamflow at watershed outlet. The riparian restoration did not just include the creation of buffer strips, but also riparian fencing, alternative water supplies for cattle and protected stream crossing. The results are not directly comparable to that of the experiments in buffer strips because streamflow loads and concentration instead of runoff concentrations at buffer entrance and exit was being used as a parameter.

Current research efforts on vegetated buffer strips are concentrated around two main topics. The first of them is how flow convergence and field connectivity affect the buffer performance. It has been demonstrated that flow convergence impairs buffer capacity to retain sediments and phosphorus (Blanco-Canqui et al., 2006; Muñoz-Carpena et al., 1999; White and Arnold, 2009) and that man-made linear structures concentrate flow bypassing VBS severely compromising the performance of buffer strips (Hösl et al., 2012). The second aspect currently being studied is the role of soil phosphorus cycling on long-term buffer performance. There is sound evidence, that the increased biological and geochemical activity in buffer strips enhances P turnover, releasing previously retained phosphorus in soluble forms (remobilization) and leaching to surface and ground waters (Roberts et al., 2013; Stutter et al., 2009; Uusi-Kämppä, 2008).

3.3   Current state of the knowledge in Austria

The interest in vegetated buffer strips in Austria is relatively new and is tightly related to the introduction of the measure in environmental conservation programmes.

Fundamentals

The first documented reference to VBS, if anecdotal, dates back to the 1993 Landscape Conservation Programme form the former Ministry for the Environment, Youth and Family. The measure was being subsidized as part of a package to mitigate soil erosion and improve natural habitat connectivity. It consisted in keeping areas adjacent to rivers free of cultivation practices and the prohibition to use fertilizers on them. Inclusion criteria and subsidize scheme was based solely on soil climate and area of the included parcel. No guidelines or technical specifications accompanied the documentation. The next reference comes in the environmental conservation programme LIFE. The project "Wasserwelt March-Thaya-Aven" contemplated VBS along other measures for the conservation of riparian habitat (BMLFUW, 2001). A few references appear since the middle of the 90s in documents from the Torrent and Avalanche Control and communal flood protection plans, but no apparent connection to water quality exists so it is of little relevance in this context.

The real introduction of the measure came with the first ÖPUL Programme, Austrian Programme for the promotion of environmentally friendly, extensive agriculture to protect natural habitats. In an effort to test the measure, a pilot-project consisting on the "development and conservation of surfaces for the protection of freshwater bodies" in selected watersheds was implemented and coupled to a 2-year study. The study, named GERAST (wpa Beratende Ingenieure, 2009), examined the mechanisms of action of VBS, validated a model to estimate its performance and reviewed the implementation of the measure.

The authors drew important conclusions from their research; the first of them was that the maintenance of a 50-m width strip along rivers provides enough protection against surface runoff from agricultural sources for water quality. Secondly, they also affirmed, that even under convergent flow conditions, the measure proves efficient in the majority of catchments adding that a higher performance can be achieved if linear structures are mapped and considered in the design phase. Despite the proven benefits of the measure, it could not be considered effective in the context of the ÖPUL Programme because of the low number of adhering participants. In fact, just 0,08% for 2007 and 0,21% for 2008 of all the potential fields were included.

Based on the experiences of the GERAST project, Hösl (2009) published a master thesis on "the effect of convergent surface runoff in the landscape and its impact on the efficiency of vegetated filter strips." With the help of GIS and field surveys, she found out that linear flow paths can concentrate runoff flow before it enters the aquatic ecosystem, bypassing vegetated buffer strips. These areas, described by the author as Unprotected Areas (UA), were observed in seven out of the ten surveyed catchments, with an extent varying from 10% to 38% of the total catchment area and had a negatively impact on the overall performance of VFS at catchment level.

Her work was later extended with the article "Man-made linear flow paths at catchment scale: Identification, factors and consequences for the efficiency of vegetated filter strips" (Hösl et al., 2012). In this essay, the authors identified man-made linear flow paths at catchment scale using DEM of varying resolution. They identified the length of the road network and annual precipitation as factors influencing the extent of the Unprotected Areas, arguing that without integrating the mapped linear structures, UA could not be detected in the broad-gridded, i.e. 10 m, DEMs.

However, after integration of mapped linear structures, DEM resolution did not influence the calculated extent of UA. Adding, that for their particular environmental settings, GIS-based design of placement of retention structures leads to considerable errors and should be verified by fieldwork.

During the planning phase of ÖPUL 2014 different implementation modalities of VBS were discussed. An important point of discussion was how the choice of the channel network influenced the situation and extension of potential VBS sites under the ÖPUL criteria, as well what impact this had on the efficiency of VBS at reducing phosphorus at watershed level. Zessner and Hepp (2014) studied this by modelling particulate phosphorus loads with the model PhosFate (Kovács, 2013), a high-resolution model for the identification of erosive areas contributing to phosphorus emission, in two depictive watersheds considering a coarse channel network (including all watersheds equal or bigger than 10 km^2 and lakes equal or bigger to 0,5 km^2; as required in WFD management plans) and a fine channel network (including all watersheds equal or bigger than 1 km^2) based on 25-m wide buffer strips adjacent to the channel network. Their results indicated that the areal efficiency on phosphorus retention using the finer channel network was almost as twice as that of using the coarser channel network, pointing out to the important contribution of small headwater channels, i.e. gullies and brooks; to erosion and phosphorus transfer to rivers.

Huber (2015) proposed on his master thesis an approach to estimate the dimension and costs to establish or extend riparian buffers for water quality using GIS applications. He developed a variable-width model based on Mander's equation (Mander et al., 1997) and identified a total of 80.400 m of rivers and stream banks directly connected to agricultural land use, which could potentially host riparian buffers.

Finally, the most recent work on the topic is a technical report on the development of a prediction tool to evaluate the performance of measures for reducing nutrient load in surface waters of Upper Austria (Zessner et al., 2016). The authors modified and validated a version of PhosFate (Kovács, 2013), , which allowed the use of 10-m cell raster maps. The modified model identified correctly the surveyed runoff paths in around 50% of the fields assessed. Further, the authors argue that with realistic model improvements it would be possible to correctly predict phosphorus emissions on up to 75% of the single fields.

3.4   Factors controlling the effectiveness of vegetated buffer strips

3.4.1   Classification of factors

Dorioz et al. (2006) proposed a classification of the factors controlling buffer strip effectiveness according to their relative position to the buffer system:

- Internal factors: endogenous to the buffer system. They influence the contact time between runoff and vegetation, rate of infiltration and the phosphorus transformation processes. These include width, slope, soil and vegetation of the buffer and flow concentration within the buffer.

- External: exogenous to the buffer system: They determine the characteristics of the incoming flow, flow concentration at source and transport pathways. These factors depend on climate, meteorology, topography and agricultural practices among others. They include: source area, source slope, soil characteristics, phosphorus load and concentration of runoff, drainage channel morphology.

The bulk of the literature consulted focuses on assessing the internal factors of buffer strips. A possible explanation to that might be the interest of researchers in studying the factors which can be influenced through management practices. In the next pages, we review the factors identified in the literature as the most significant to buffer performance.

3.4.2 Vegetation

Two characteristic features of vegetation aid to the physical retention of particles in buffer strips: aerial body parts and a root system. Aerial body parts increase hydraulic roughness, decrease runoff velocity and flow energy available to transport particles (Yang et al., 2015). The root system of plants increases the permeability and porosity of the soil, enhancing runoff infiltration (Roberts et al., 2012).

Vegetation influences in-soil processes occurring within the VBS (Roberts et al., 2012). Plants absorb phosphorus from the soil through the roots accumulating it in their biomass to form new tissues. They do that by releasing exudates who transform the phosphorus particles in absorbable ionic forms.

Whether grassed or wooded filter strips are more suitable for P retention depends on the physical form considered and the prevailing soil transformation processes in the buffer. Grassed filter strips are best suited to intercept sediment-bound particles and are favoured to control phosphorus form agricultural runoff (Schmitt et al., 1999). The high density and total cover of grass stems oppose a higher resistance to the flow than bushes and trees, capturing sediment-bound particles more efficiently. Further, the dense root system of grasses, with increased number of fine roots, increases soil porosity and permeability (Roberts et al., 2012). However, the lodging of grass tillers, the process in which they bend due to the weight of rain, runoff or adverse weather conditions, can reduce the ability of plants to retain particulate matter and create preferential flow paths (Dorioz et al., 2006).

Forested buffer strips on the other side, are best suited to capture dissolved nutrients. Their deeper root system can trap and uptake nutrients from deeper soil layers compared to grasses and incorporate them as part of their biomass (Yang et al., 2015).

That said, there is no clear consensus in the literature as to whether grassed or wooded filter strips are more effective. While some studies did not show any clear differences in performance (Daniels and Gilliam, 1996; Mayer et al., 2007; Yuan et al., 2009), others show an advantage to grassed vegetation (Schmitt et al., 1999). Wenger (1999) sees in the combination of grassed and wooded

buffers a reasonable compromise with the advantages of both vegetation types. Finally, Dorioz et al. (2006) claims that mixtures of grass, trees and bushes are less effective in retaining nutrients from agricultural sources.

The height of vegetation has been also the subject of study. Cole et al. (1997) claims that in order to be effective, grass has to be higher than the expected flow height because submerged vegetation loses the ability to interact with the flow. However, an increase in height far beyond this threshold does not noticeably improve the buffer performance. Pearce et al. (1997) even affirm the contrary: close-cut vegetation is preferable because it does not fall under the influence of rain or runoff.

Mander et al. (1997) reviewed the role of forest age in VBS and concluded that younger stands and bushes perform better than older stands in capturing soil nutrients because of the higher nutrient uptake and high microbial activity in the soil during the growing phase. A similar appreciation is made by Dosskey et a. (2010): in early stages of succession, the incorporation of phosphorus into plant biomass increases rapidly, it reaches its maximum and then stagnates until a state of equilibrium is approached, reducing steadily with increasing age. Noting that even in mature communities with a net uptake of near zero, plants can provide a phosphorus sink by storing it in tissues.

Vegetation harvesting has proved a successful technique to maintain a high phosphorus uptake and to permanently remove phosphorus from the system (Dosskey et al., 2010; Mander et al., 1997; Uusi-Kämppä, 2005).

3.4.3   Slope inclination

With increasing slope inclination, the flow velocity and kinetic energy of runoff increases. This results in a higher erosive potential and sediment transport capacity of flow. Depending on the geomorphology of the watershed, slope at source and at buffer might not be the same: higher slopes at source fields mean more soil particles can be eroded and transported downwards whereas higher slopes at buffer decrease contact time between runoff and vegetation, resulting in less sedimentation and infiltration. The interaction of slope at source and buffer is far more complex than that depending on the landform and scale considered.

The link between higher VBS slope and lower total phosphorus trapping efficiency has been demonstrated in the literature (Daniels and Gilliam, 1996; Dillaha et al., 1988; Patty et al., 1997; Syversen, 2005). Interestingly, few experiments have considered buffer slope as an independent parameter and none examined the role between different slope at source fields and buffers. Dillaha et al. (1988) observed that as buffer slope increased from 11% to 16%, total phosphorus retention declined from 63 to 52% in 4,6 m wide strips and from 80% to 57% in 9,1 m buffers, revealing an inverse linear relationship between buffer slope and retention of total phosphorus. Syversen (2005) found out higher retention efficiencies on the sites with lesser slope but he attributed the differences to vegetation conditions rather than decreased slope gradient. In a more recent study Dosskey (2008) simulated the behaviour of vegetated buffer strips at different slopes and concluded that trapping efficiency is lower at greater slope. One author argues that removal

efficiency is maximized on slopes around 9% with lower and higher slopes yielding less performance (Zhang et al., 2010).

A study conducted in Canada by Sheppard et al. (2006) on the efficiency of vegetated field margins on flat landscapes (slope gradient inferior than 2%) observed that the ponding of runoff water in topographical lows increased the prospect of loss of dissolved phosphorus to drainage network.

The relationship between soluble phosphorus retention efficiency and slope is not clear and results are not conclusive. Patty et al. (1997) conducted experiments on the use of grassed buffer strips to remove soluble phosphorus from runoff water in a range of soil and cropping conditions for filter lengths of 6, 12 and 18 m and 7, 10 and 15% slope gradients. The results showed a reduction between 22 to 89% in soluble phosphorus with increasing slope gradient but no relationship between slope and reduction efficiency was established.

In summary, greater buffer slopes are related to lower performance. Higher slopes are associated to higher runoff potential erosion at source, higher transport capacity of runoff flow and reduced contact time between runoff and vegetation at buffer.

3.4.4 Soil characteristics

Soil characteristics are an important factor affecting the release of phosphorus from source fields and the trapping efficiency of buffers.

Soil texture determines the quantity and size of phosphorus particles available for mobilization at source (Schmitt et al., 1999). Fine, clay soil particles have a greater capacity to hold phosphorus than sand particles because they are more cohesive and allow soil particles to stick together making them less vulnerable to erosion coarser particles. Coarse-textured soils, on the other side, allow for greater infiltration rates and require higher runoff energy to be mobilized. Middle-textured soils fall in the middle: particles are not held together by cohesive forces and have a size small enough to be easily mobilized. It is actually the medium-texture silt clay loam soils that mobilize the highest rate of phosphorus to runoff.

Soil texture in the buffer affects efficiency in a similar fashion. The best retention performance is achieved under average soil texture conditions, i.e. silt or sand (Schmitt et al., 1999). Soils largely made up of sand may drain water so rapidly into the groundwater that roots are not able to effectively trap phosphorus, leaching to groundwater (Hawes and Smith, 2005). According to White and Arnold (2009) buffer strips with a high fraction of clay retain almost no runoff at moderate runoff loadings. It has been observed that colloid-forming clay particles deposition involves turbulent flow between runoff and vegetation (Roberts et al., 2012).

Another soil property affecting buffer performance is the soil redox state. It has the potential to remobilize big quantities of phosphorus in soil reservoirs (Roberts et al., 2012). Under anoxic soil conditions, i.e. hydric, water-saturated and poor ventilated soils, phosphorus precipitates with Fe and Al (acid medium) and with Ca (alkali medium) to form phosphates dissolved in the soil

solution (Dorioz et al., 2006). When phosphorus input surpasses the reaction capacity, precipitation sites may saturate with phosphorus and leach phosphates as return flow to surface waters or percolate to groundwater eventually reaching water bodies by lateral flow.

The literature reviewed reports many cases were buffer strips become a source of dissolved phosphorus due to the microbiological and geochemical soil processes causing phosphorus remobilization and leaching. Yang et al. (2015) detected release of dissolved phosphorus in almost all plots of their studies. In one 10,0 m wide buffer plot, the release of dissolved phosphorus was of 106%. They attributed this leach to two causes: the first, phosphorus desorption from soil particles and the second, to the small quantity of humus in surface soil leading to strong variations of dissolved phosphorus. Similarly, Dillaha (1988) reported an instance were effluent from the filter contained two times more phosphates than the influent, attributing it to the leaching of previously trapped phosphorus in soluble forms.

The most relevant long-term pool of phosphorus in soil is vegetation, other noteworthy biological pools include humus, decaying organic matter and associated microbes. Their function as a pool plays an important role in phosphorus cycling because seasonal changes in soil characteristics such as temperature, pH and aeration can cause sudden release of dissolved phosphorus and occasion leaching to freshwater bodies (Dosskey et al., 2010).

Soil organic matters plays a major role in phosphorus cycling and regulates the release of phosphorus in other soluble forms. Greater biological activity in soils improves its ability to effectively deal with dissolved nutrients and pollutants. That kind of activity is more prevalent in soils rich in plants, soil micro and macrofauna and bacteria than in a soil without these organisms (Grismer et al., 2006).

Finally, humus and plant litter from leaf senescence are an important pool of phosphorus in the soil, they accumulate phosphorus in organic forms which is not easily absorbable by plants and can be mobilized and leached out by runoff (Dosskey et al., 2010).

The extent of phosphorus that the microbial biomass in the soil can retain varies seasonally with humidity and depends on the organic matter content of the soil (Roberts et al., 2013). During drier months, the remobilization of phosphorus from microbial pool increases, whereas during wetter months the retention capacity is higher. Shorter-term fluctuations in microbial biomass phosphorus due to drying and wetting cycles also   can cause phosphorus remobilization.

3.4.5   Flow convergence

The complex geomorphology of hillslopes curves the landscape in various ways, influencing runoff processes. When sections of the terrain spread up, runoff is dispersed across the slope surface. Similarly, when sections of the terrain converge, i.e. come close together, runoff concentrates in depressions of the terrain leaving other areas with little or no runoff. Whereas the first of the processes does not hinder the ability of VBS to trap sediments or phosphorus, the second, flow convergence, can be a major constrain of buffer performance.

Flow convergence may take place at source fields and runoff enter concentrated to the filter strip or it may develop within the buffer. In either case, the results of experimental studies and modelling assessment are conclusive: flow convergence reduces the performance of buffer strips in retaining sediment and phosphorus (Blanco-Canqui et al., 2006; Dillaha et al., 1988; Helmers and others, 2008; Verstraeten et al., 2006; White and Arnold, 2009).

The phenomenon can even be observed in relatively gentle slopes. Flow convergence has even been detected on plot scale experiments where it was not even intended, Dillaha et al. (1988) observed that the cross slopes of fields caused runoff to flow to one side of the plot and advance as deeper channel flow, constraining the flow to a 0,5 to 1 m wide strip. Helmers et al. (2005) developed a method to detect and quantify flow convergence and identified converging areas of overland flow in relatively planar topography. Using 3 cm resolution maps, they succeeded in predicting the actual flow paths. Further, they noted that despite observing areas with uniformly distributed overland flow, filter strips trapped approximately 80% of the incoming sediment.

Dosseky et al. (2002) used the vegetative filter strip model VFSMOD (Muñoz-Carpena et al., 1999) to evaluate the performance of riparian buffers and found that flow concentration reduced buffer effective area between 12% to 81% of the gross buffer area. White and Arnold (2009) adapted the VFSMOD model to consider the effects of flow concentration at field scale and found that on average 10% of the buffer strip received 50% of the source field runoff. Blanco-Canqui et al. (2006) reached similar conclusions and found removal efficiencies dropped up to 25% under inter-rill and concentrated flow conditions compared to with sheet flow.

Flow concentration is influenced by spatial scale as well. Verstraeten et al. (2006) demonstrated in a modelling study that flow convergence is of increasingly importance at larger scales. Their model predicted efficiency dropped from 70% at plot level to 20% at catchment level, attributing it to flow convergence and sediment bypasses.

The phenomenon is not exclusive to natural terrain depressions. Field practices like ploughing along the line of slope concentrate flow in very small portions of the fields bypassing vegetated buffers or entering them as concentrated flow. Man-made linear structures like drainage, ditches, channels and roads discharge directly to the stream network bypassing vegetated buffers. Hösl (2012) studied the impact of man-made linear flow paths in ten Austrian catchments and found out that in seven out of ten catchments between 10% and 38% of the total catchment area was directly discharging to the drainage network. It is possible to identify the extent of areas flowing directly to a stream by integrating maps of the linear structures to DEM models.

4 Material and methods

4.1 Literature research

The background information to determine the current state of knowledge as well as the data from experimental studies was compiled by means of online research in peer-reviewed journals from publishing platforms like Elsevier, Wiley and Science Direct and in Google Scholars between October 2015 and July 2016.

Other sources of information include documents from national environmental agencies, research institutes as well as personal communication from field-related scientist at the Institute for Water Quality, Resource and Waste Management (IWAG) of the TU Vienna.

4.2 Software

The open-source program **Zotero** from the Roy Rosenzweig Center for History and New Media (CHNM) was used to manage bibliographic data and related research material. **Microsoft Office® Excel 2013** for Windows and **Excel 2016** for Mac were used to collect and process the data subject to statistical analysis as well as to create graphs and illustrations. The programming language and software environment **R Studio** was used to perform the statistical analyses of the data. Finally, **Microsoft Office® Word 2016** for Mac was used to write and format the master thesis.

4.3 Study design

The present study is an analysis of experimental data published up to date on the performance of vegetated filter strips for phosphorous retention from non-point agricultural sources.

4.3.1 Data collection and database

Data for the statistical analysis has been gathered from figures reported in experimental studies and published mostly in peer-reviewed journals, with few exceptions; where papers found in universities websites coming from broadly recognized authors have been included.

Data was summarized in variables including information on the author, the year of publication, the soil texture, the duration of the experiments, the extension of the experiment, the source of water, the type of flow, the presence of flow concentration, the field length, the type of vegetation, the width of the VBS, the total phosphorus retention (TP), the particulate phosphorus retention (PP) and dissolved phosphorus retention (DP), the measures and units upon which this reduction is based and the slope inclination of the VBS. Auxiliary variables to process the data were created as necessary

A table including all included data records can be found in Appendix 1: Database. This table is a simplification of the database structure and does not include auxiliary variables, necessary for the organization and sorting of the information but not subject of the analysis.

4.3.2   Eligibility criteria

Eligibility criteria were defined to qualify for inclusion based on the type of study, the phosphorus source, the vegetation, the measured outcome and the magnitude of results. They were chosen to ensure data included is within the scope of interest of the work and to ensure comparable and conclusive results.

Table 4.3-1 Overview of inclusion and exclusion criteria

	inclusion	exclusion
Type of study	experimental evidence	modelling studies
Phosphorus source	diffuse agricultural runoff	point agricultural runoff and non-agricultural runoff sources (i.e. feedlot runoff, street runoff, sewage)
vegetation	grass, woody or mixed	none
structure	Vegetated Buffer Strips	field margins, grassed waterways and other agricultural BMP
measured outcome	TP, PP or DP	other P fractions
magnitude of results	percent (%) retention	absolute values of retained P

4.3.3   Data pre-processing

Raw data was pre-processed to obtain a clean, consistent and meaningful dataset. Data was checked for quality, irrelevant variables were excluded, duplicates were deleted and records not fulfilling the inclusion criteria were cleansed. Where needed, data transformations like reclassification and summarization of variables were applied.

4.3.3.1   Data quality

Data quality was based on the criteria of completeness of records, consistency and plausibility of values. A quality standard, defined by the following rules, was implemented:

- Records not reporting a value for the outcome variable TP in a discrete percentage value, or where a direct calculation by addition of particulate (PP) and dissolved (DP) phosphorus fractions was not possible, were deleted.

- Records missing values for three or more explanatory variables were deleted. This check was implemented as a function in R: records with three or more "NA" were cleansed.

- Records with physically implausible values or with values in conflict with values from other variables were deleted.

- ▪ Records were manually doubled-checked for consistency. If a typing or transcription error was found, it was corrected.

4.3.3.2   Transformation of variables

4.3.3.2.1 Reclassification

Numerical explanatory variables given as intervals instead of discrete values were reclassified into discrete values by averaging the range to its central value.

4.3.3.2.2 Recoding

Recoding consists in combining several categories into fewer categories with higher frequency counts and less detailed information, with the goal to reduce the total number of possible outcomes (Templ, 2017).

Categorical variables with low relative frequency outcomes, i.e. less than 5% of the total observations falling in a category or factor, were recoded into fewer factors in order to increase the relative frequencies.

4.3.3.2.3 Categorizing

Categorizing consists in the transformation of continuous data into groups of bins or categories by dividing the range of a probability distribution with cut-off values (Altman, 2014). This technique allows for comparison between categories and allows the use of determinate inference techniques but has the trade-off of loss of information, statistical power and efficiency.

The explanatory variable "VBS slope" was categorized into the variable "VBS slope class" to allow for a comparison of different slope classes. This step allowed to extract information in the cases were the values for the variable "VBS slope" were given as an interval rather than in a discrete form, but a reclassification of these values into discrete values was not desired.

4.3.3.3   Data selection

After removing duplicate entries, the database consisted of 119 records. From them, 7 observations did not fulfil the inclusion criteria: phosphorus sources other than diffuse agricultural runoff, observations on bare-soil buffers as well as flow types other than surface runoff and were excluded from the analysis.

From the remaining 112 records, a total of 13 were deleted for failing to meet the quality requirements, mostly due to missing values on three or more explanatory variables. Finally, one single record was considered to be an outlier and was as well removed.

In total 98 records were included for the analysis, 69 were classified as experimental plots and 29 as cultivated land.
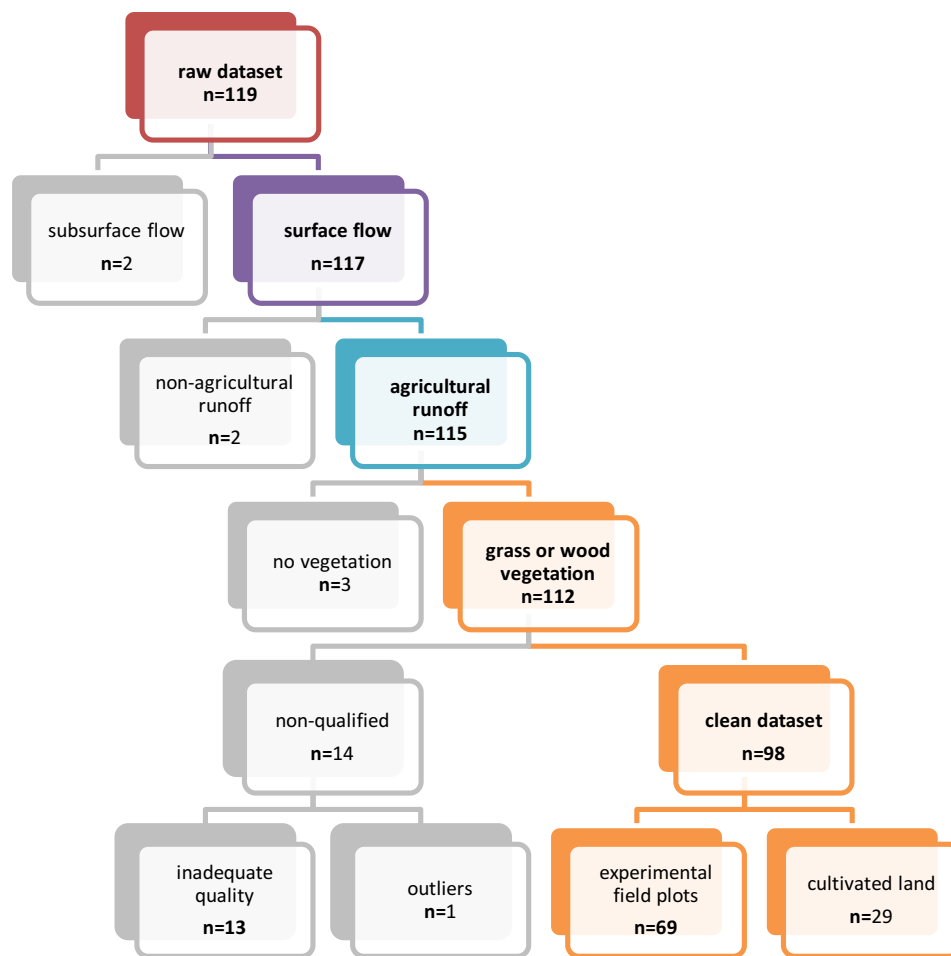
Figure 4.3-1 Data pre-processing chart (n represent the number of observations)

4.3.4 Variable overview

Table 4.3-2 Variable overview depicts the variables subject of analysis, including its name, its unit when applicable, the type of data and the analysis performed on them. A total of nine variables were considered, including the outcome variable TP retention and eight explanatory variables.

Basing on the objectives of this work and depending on the type of data, variables were subject to different descriptive statistics, an inference analysis was performed on them and depending on their relevance they were selected for regression modelling. Note that other auxiliary variables, i.e. author and year of publication, quality checks or intermediate values, are present in the original database but they are not listed here as they are not relevant for the analysis.

Table 4.3-2 Variable overview

Variable	Unit	Analyses	Data Type
soil group (soil.group)	*(category)*	descriptive statistics regression analysis inference	categorical
time frame (time.frame)	*(category)*	descriptive statistics regression analysis inference	categorical
experiment type (extension)	*(category)*	descriptive statistics regression analysis inference	categorical
vegetation cover (vegetation)	*(category)*	descriptive statistics inference	categorical
VBS width (VBS.width)	m	descriptive statistics regression analysis inference	continuous
width ratio (width.ratio)	*category*	descriptive statistics regression analysis inference	categorical
TP retention (TP.retention)	%	descriptive statistics regression analysis inference	continuous
VBS slope (VBS.slope)	%	descriptive statistics regression analysis inference	continuous
VBS slope class slope.class	*(category)*	descriptive statistics regression analysis inference	categorical

In parenthesis, the name of the variable in the database CSV file.

4.3.5   Variable definition

4.3.5.1   Soil textural class (soil.texture)

Soil textural class of the surface soil layer according to the United States Department of Agriculture (USDA) textural soil classification. A total of twelve major soil textures are defined on the name of the primary constituent fractions clay, silt, sand or a combination of them plus loam, a mix of similar proportions (Soil Science Division Staff, 2017). The USDA soil texture triangle names all the classes and the proportion of the primary constituents.
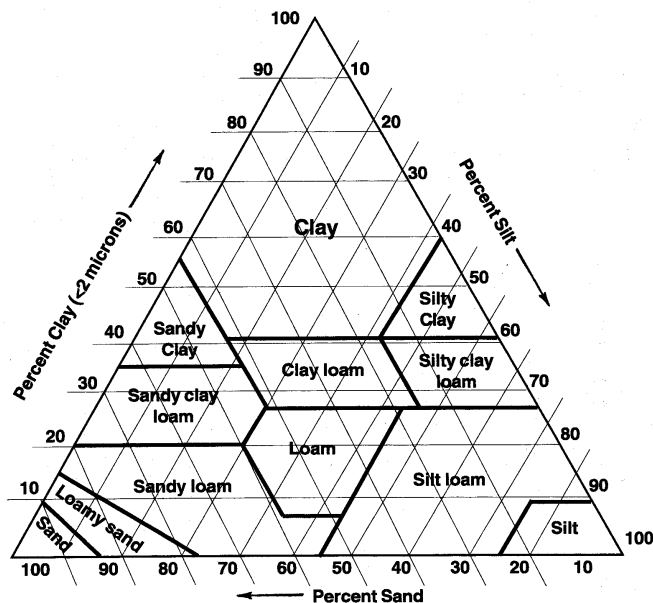
Figure 4.3-2 USDA soil texture triangle (C. Ditzler, K. Scheffe, and H.C. Monger, 2017)

4.3.5.2   Soil textural group (soil.group)

Categorical variable summarizing the USDA soil textural class into broader, soil texture groups or classes (Soil Science Division Staff, 2017).

By reducing the number of possible textural classes into broader groups, this variable aims at increasing the relative frequencies of each category.

Table 4.3-3 Soil textural groups

soil textural group	USDA soil textural class
coarse	sandy
	sandy loam
	loamy sand
medium	loam
	silt
	silt loam
fine	sandy clay
	sandy clay loam
	clay loam
	silty sandy clay loam
	silty clay
	clay

4.3.5.3   Time frame (time.frame)

Variable describing the time frame in which the observations were made. It aims at identifying differences in outcomes due to time-related accumulation or decomposition processes and the effects of climatic patterns.

- **Single events:** observations during a precipitation event or a series of precipitation events, separated from the next event period of precipitation longer than the recessing hydrograph limb caused by event. Time-related patterns are limited to hours following the event: receding runoff, decreasing soil-moisture and day-night variations.

- **Seasonal**: observations during a series of events accumulated over a period of time greater than one month and less than one year. Seasonal soil and climatic patterns like phosphorus accumulation, organic-matter decay or daily temperature variations might be observed.

- **Annual**: observations during a series of events accumulated over a period of time of at least one year. Annual climatic patterns and annual-related accumulation or decomposition processes might be reflected.

4.3.5.4 Experiment type (extension)

Variable describing the environmental setting under which the experiment has been conducted. The main aim of this variable is to differentiate between controlled experiment field settings and field conditions.

- Experimental field plot: intentionally planned plot where the observer has control of the experiment events and all the explanatory variables. The experiment might be replicated by other researchers.

  This implies control of external factors like precipitation, runoff and nutrient fluxes at plot level, so that a balance between inputs and outputs is possible. It is achieved by separating the plot from adjacent fields through installation of lateral barriers and use of runoff collections. Rainfall and inflow runoff are controlled by using rainfall or runoff simulators. In the few cases where observations are held under natural rainfall conditions, rainfall intensity and volume is monitored in such a manner that the calculation of a flow balance is possible.

- Cultivated land: plot, where the observer has just partial or incomplete control of the experiment events and the explanatory variables. Distinct events occur and are monitored and quantified by the observer. Replication of the experiment under the exact same conditions is not possible, but it is possible to compare some characteristics of the experiment. In this work, this category is alternatively referred as field conditions.

It includes observations under field conditions, held in VBS adjacent to agricultural land under natural rainfall. A water and nutrient balance is still possible.

4.3.5.5 Vegetation cover (vegetation)

Variable describing the type of vegetation cover established or used in the VBS according to its stem structure.

- Grass: herbaceous plants with soft stems. Its small and flexible stem bends and might be overflown in the presence of runoff.

- Wood: vascular planta with hard stems, including trees and bushes. Its lignified stem allows for great plant heights and shear resistance. Their stem does not bend and is not overflown in the presence of runoff.

4.3.5.6 VBS width (VBS.width)

Distance between upper and lower border of the VBS, equivalent to the minimum flow length of runoff in the VBS.

4.3.5.7 Field-VBS width ratio (width.ratio)

Quotient between the VBS width and the width of the field contributing to runoff.

$$width.ratio \ (m) = \frac{VBS.width}{field \ width}$$

Equation 4.3-1

4.3.5.8 Total Phosphorus retention (TP.retention)

Percentage fraction (%) of total phosphorus retained in the VBS calculated as the quotient between the TP concentration in runoff from the source field (inflow) at the upper boundary of the VBS and the TP concentration in runoff (outflow) abandoning the VBS at the lower limit.

$$TP \ (\%) = \left(\frac{TP \ inflow - TP \ outflow}{TP \ inflow}\right) \times 100$$

Equation 4.3-2

This definition assumes that the VBS is a closed system with no other paths of phosphorus entering or exiting the system.

4.3.5.9 VBS slope (VBS.slope)

Mean slope-gradient of the VBS surface to the horizontal plane expressed as percentage (%). This definition assumes the VBS profile does not present undulations or irregularities and the slope-gradient is constant.

4.3.5.10 VBS slope class (slope.class)

As none of the slope-gradient classification systems found in the literature i.e. Soil Survey Manual (Soil Science Division Staff, 2017), Guidelines for Soil description (FAO, 2006) fitted the needs of this work, a process-based classification according to the potential dominant erosive processes was adopted:

- Gentle: slope-gradient up to 8%.

Mostly dominated by sheet and inter-rill erosion.

- Moderate: slope-gradient greater than 8% and up to 20%.

Mostly dominated by rill and gully erosion. The latter dominating with increasing slope.

4.3.6 Influential points: leverage points and outliers

Data was checked for potential leverage points and outliers using the Cook's distance method. Cook (1977) proposed a measure of influence based on the extent to which parameter estimates for a regression would change if a particular observation was omitted from the model.

Cook's Distance (CD) is defined as the standardized distance between the parameter estimates by omitting a particular observation and the parameter estimates by including all the observations (Cook, 1977), revealing which cases are most influential in affecting the regression equation by causing large leverage (Stevens, 1984). The higher the distance, the greater the leverage a point exerts on the regression line and the more influential it is. If the distance is above a critical value, observations are identified as causing leverage and further consideration is required. The critical value has been set at 4/n, where n indicates the number of observations. Values above this threshold are considered leverage points.

A chart with the CD on the y-axis and the number of the observation on the x-axis was plotted. A red-dashed line representing the critical distance was include in the graph. Points above the critical distance were subject to further analysis to consider if they shall be included in the analysis.

This method was chosen against other suitable methods for multivariate data like robust regression i.e. Least Median Squares, Mahalanobis Distance for its sensibility to identify leverage points and residuals of the observation (Reimann et al., 2008).

4.3.7 Data summary

Figure 4.3-3 summarizes the number of observations for each factor of the variables included in the analysis. Values of continuous variables have been classified as either available or no available. No available values were not considered in the statistical analysis.
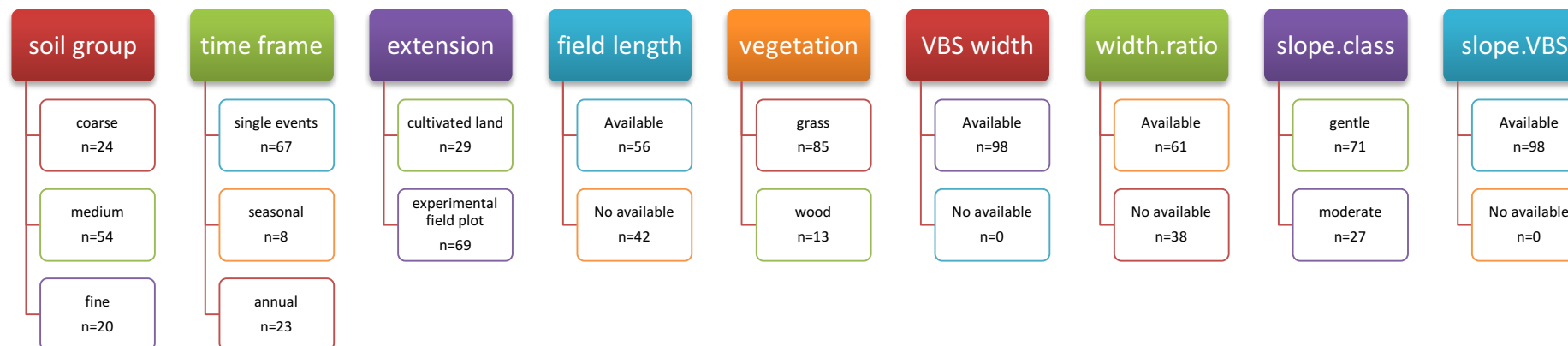
Figure 4.3-3 Data summary (n represents the number of observations)

4.4 Statistical analyses

4.4.1 Descriptive statistics

Main outcome variables were analysed as outlined in Table 4.3-2, data was presented graphically in histograms, box-plots or scatterplots to facilitate interpretation. Special focus was laid on showing data distributions, detection of spread and revelation of bivariate relationships between outcome and explanatory variables. These descriptive statistics constitute an initial exploration of data before performing inferential tests.

Box-plots were constructed using quartiles, the space between first and third quartiles, the Interquartile Range (IQR), was filled in colour. The second quartile is built with the median and is depicted as a labelled point. The lower whisker extends to a maximum of 1,5*IQR below the first quartile and the upper whisker to a maximum of 1,5*IQR above the third quartile respectively. Outcomes beyond the range of the whiskers were plotted as single dots.

Scatterplots include all data points. The outcome variable TP retention is always shown on the Y axis extending from 0% to 100% on a linear scale. The explanatory variables are represented on the X axis with linear or log-transformed scale as needed. Single data observations are represented as coloured points with the colour representing the value for a second explanatory variable. In a few cases a tendency line was added to the scatterplot to help with visual interpretation but no parameters or strength of regression was given.

The Spearman's rank-order correlation coefficient (r_s) was calculated between the outcome and each numerical explanatory variable. This statistic was chosen in favour of Pearson's correlation coefficient because it was believed to better capture the non-normal nature between the outcome and explanatory variables.

The Spearman's rank-order correlation coefficient (r_s) is a non-parametric (distribution-free) rank measure of the strength of the associations between two variables. It assesses how well an arbitrary monotonic function can describe the relationship between two variables, without making any assumptions about the frequency distribution of the variables (Hauke and Kossowski, 2011). Its value is analogous to that of Pearson's correlation coefficient: the closer it is to one or minus one, the stronger its association of rank.

4.4.2 Inferential statistics

Inferential tests were performed between the outcome and each of the explanatory variables to identify significance and degree of association. For each test result, the probability of rejecting the null hypothesis (p-value), the significance and the degrees of freedom were reported.

The significance level α was set to 0,05 for all tests. Probability values (p-value) smaller than 0,05 were considered statistically significant, rejecting the null hypothesis and assuming the alternative

hypothesis to be true. Probability values bigger than 0,05 failed to reject the null hypothesis and the alternative hypothesis was not accepted. There is a 5% risk of committing a type I error (false positive), rejecting the null hypothesis when it is actually true.

4.4.2.1 Continuous variables

Significance of association between numerical explanatory variables and outcome variable was tested with an independent t-test on the calculated Spearman's correlation coefficient r_s.

The t-test statistics uses a Student's distribution with n-2 degrees of freedom to calculate probability values. The value for the t-statistic was calculated using the degrees of freedom (n) and Spearman's r_s:

$$t = r_s \sqrt{\frac{n-2}{1-r_s^2}}$$

Equation 4.4-1

A. Assumptions:

The t-test is based in the following assumptions (Yeager, 2017):

Table 4.4-1 Assumptions of a t-test

Assumptions
A. Independence of observations
B. Random sample
C. Normal distribution of residuals
D. Homogeneity of variance (homoscedasticity)

The t-test is robust against small deviations from the mathematical assumption of normal distributions of residuals for medium or large (n > 30) sample sizes, yielding fairly accurate p-values for non-normal distributions (Hampel, 2002). In practice, this means that even if the assumption of normal distribution of residuals is violated the test can still be considered valid as long as the other assumptions are met.

In parametric tests like the t-test, homogeneity of variance (assumption D) is considered acceptable if the largest variance is not more than four times the smallest (De Winter and Cahusac, 2014). Higher inequalities of variance are considered heterogeneous. Where data distribution suggested the variance was not homogeneous a Leven test for equality of variance was performed to exclude heteroscedastic distribution form the analysis.

B. Hypothesis testing:

The aim of the hypothesis was to test if the population Spearman's correlation coefficient r_s was zero or alternatively, different from zero:

Null hypothesis	H_0: $r_s = 0$	→ no association between the two variables
Alternative hypothesis	H_1: $r_s \neq 0$	→ association between the two variables

If we fail to reject the null hypothesis H_0 and assume it is true, we accept that the correlation between the explanatory and outcome variable is zero and no significant association between the two exists. If we reject the null hypothesis (p-value $< 0{,}05$), we assume the alternative hypothesis H_1 is valid, we accept the correlation between the explanatory and outcome variable is different from zero and a significant association between the two variables exists.

4.4.2.2 Categorical variables

Significance and strength of association between categorical explanatory variables and the outcome variable was tested with a one-way analysis of variance (ANOVA).

The statistic computed in ANOVA to generate probability values is a F-test, which follows a Fischer distribution. This test compares the means of two or more independent groups or factors, assuming a normal distribution of variance within the groups, to determine whether their associated population means are statistically the same or on the contrary, different (Kao and Green, 2008).

A. Assumptions:

ANOVA is based in the following assumptions (Yeager, 2017):

Table 4.4-2 Assumption of ANOVA

Assumption
A. Independence of observations
B. Random sample
C. Normal distribution of residuals
D. Homogeneity of variance (homoscedasticity)

ANOVA is robust against small deviations from the mathematical assumption of normal distributions of residuals for medium or large (n $>$ 30) sample sizes, yielding fairly accurate p-values for non-normal distributions (Hampel, 2002). In practice, this means that even if the assumption of normal distribution of residuals is violated the test can still be considered valid as long as the other assumptions are met.

In parametric tests like ANOVA, homogeneity of variance (assumption D) is considered acceptable if the largest variance is not more than four times the smallest (De Winter and Cahusac, 2014). Higher inequalities of variance are considered heterogeneous. Where data distribution suggested the variance was not homogeneous a Leven test for equality of variance was performed to exclude heteroscedastic distribution form the analysis.

B. Hypothesis testing:

The aim of the hypothesis was to compare the means of two or more independent groups in order to determine whether there is statistical evidence that the associated populations are the same or, on the contrary, there is significant probability to believe they are different.

| Null hypothesis | $H_0: \mu_k = 0$ | $\rightarrow$ all k population means are equal |
| Alternative hypothesis | $H_1: \mu_k \neq 0$ | $\rightarrow$ not all k population means are equal |

If we fail to reject the null hypothesis H_0 and assume it is true, we accept that all population means are equal and no differences between the groups exists. If we reject the null hypothesis (p-value < 0,05), we assume the alternative hypothesis H_1 is valid, we accept that not all the k population means are equal and significant differences between the groups exist with at least one of the means differing from the others.

Note that ANOVA does not indicate which mean is different from the others. To determine it requires multiple pair comparisons of means or post-hoc tests (Yeager, 2017). The performance of post-hoc tests is beyond the scope of this work.

4.4.3 Regression models

Explanatory variable found to show association with the outcome variable were used to build multiple linear regression models (MLR) using the ordinary least squares (OLS) technique to adjust the equation parameters.

The models' parameters are reported to two significant figures or the minimum necessary which produces a deviation of ±1,5 percentage points compared to the full precision model. Intermediate figures are reported and calculate using the fullest precision possible.

A total of three models were built:

4.4.3.1 Model 1: General model

This model aims at identifying the set of explanatory variables which yield the best goodness-of-fit, understood as the highest adjusted coefficient of multiple determination (adj. R^2), at predicting the TP retention capacity of VBS. It is built considering all data records in the database and is valid for the conditions of the modelled data: phosphorus from agricultural sources in VBS

between 2 and 29 m wide for slopes bigger than 2% and smaller or equal to 20%, independently of vegetation type.

4.4.3.2   Model 2: Width-ratio model

This model aims at assessing the power of width-ratio at predicting TP retention compared to that of VBS width by examining goodness of fit. It is built considering those records in the database where the width-ratio variable is available and is valid for the conditions of the modelled data.

4.4.3.3   Model 3: Field performance

This model aims at assessing power of chosen explanatory variables at predicting TP retention under field conditions for VBS slopes smaller than 8%. It is built considering records in the database of the experiment type "cultivated land" with a VBS slope smaller than 8% and is valid for data fulfilling these conditions.

4.4.3.4   The multiple linear regression model (MLR)

The multiple linear regression model assumes there is a linear relationship between a dependent outcome variable Y and a set of independent explanatory variables X_i (Schmidheiny, 2013). The outcome Y is the result of a combination of deterministic parameters β_i, a constant α_0 and random error ε_i. The relationship between the outcome variable Y and the explanatory variables X_i is described by the following equation:

$$Y_i = \alpha_0 + \beta_i \cdot X_i + \varepsilon_i$$

Equation 4.4-2

This equation is called the line of best fit. The constant term α_0 represents the intercept of the linear relationship and catches the outcome of the regression when the independent variables are at reference level. The slope terms β_i correspond to the multiple regression coefficients, which are the independent contribution of each independent variable to the predicted outcome, whereas X_i are the given values for the independent variable i. The term ε_i corresponds to the random error, the unexplained part of the variance or the part of the outcome which cannot be explained by the explanatory variables.

4.4.3.5   The ordinary least squares (OLS)

The ordinary least squares method was used to estimate the regression parameters. This technique minimizes the squared distances between the observed and the predicted dependent variable y, called residuals (Schmidheiny, 2013):

$$\sum_{i=1}^{n}(Y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(\hat{\varepsilon}_i)^2 = min$$

Equation 4.4-3

The residuals are defined as the deviation (distance) between the predicted and the observed values (see Figure 4.4-1) and provide a measure of how well the predicted model fits each single data point (Moutinho and Hutcheson, 2011). Positive values for the residual mean the prediction was too low whereas negative values mean the prediction was too high.

The sum of all squared residuals, known as the residual sum of squares (RSS), provides a measure of the model fitness (Moutinho and Hutcheson, 2011). A poorly fitting model will deviate markedly from the observed data and have a consequently large RSS, whereas a good-fitting model harmonizes with observed data and has a small RSS.
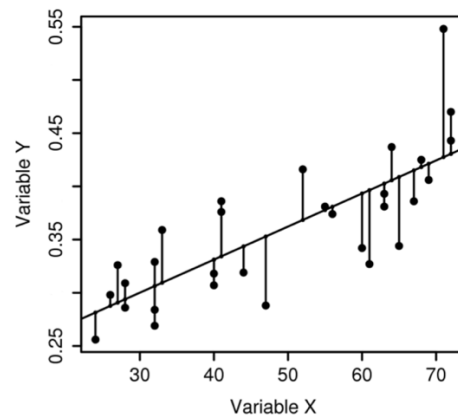
Figure 4.4-1 Residuals in Ordinary Least-Squares. After Hutcheson, G.D. (2011)

The adjusted coefficient of multiple determination adj. R^2 was used as a measure to indicate the percentage of variance in the response variable that is explained by the model. It can take values between 0 and 1 denoting the strength of the linear association, with a value of 0 indicating no association and a value of 1 indicating that the regression line passes exactly through every point on the scatter plot. It allows to determine how certain one can be in making predictions from a determinate model (Draper and Smith, 1998).

This statistic is a modification of the coefficient of determination R^2 in simple linear regression, which takes into consideration the number of independent variables used in the model and is a better measure of the explained variation than R^2 in multiple regression (Moutinho and Hutcheson, 2011). It can be calculated as follows:

$$adj.R^2 = R^2 - \frac{k(1 - R^2)}{n - k - 1}$$

Equation 4.4-4

where n is the number of observations used to construct the model and k is the number of independent terms. A further discussion of measures of fitness for regression models can be found in Moutinho and Hutcheson (2011).

4.4.3.6   Regression assumptions

The use of linear regression is based in particular assumptions. To be fit for prediction or statistical inference, a model needs to verify the assumptions (Osborne and Waters, 2002):

1.   Independence of observations
2.   Linearity between dependent and independent variables
3.   Normal distributions of residuals
4.   Homogeneity of variance

Whereas some of these assumptions are robust to violation e.g. normality of error distribution, others are more sensitive to it and in case of violation, model predictions might be biased or even totally invalid.

Independence of observations refers to the input data and is achieved through proper design of the study. All of the other assumptions refer to the model itself and can just be tested *a posteriori* once the model is built. The tools with which the assumptions will be assessed are discussed in chapter 4.4.3.10.

4.4.3.7   Multicollinearity

Multicollinearity exists when two or more of the predictors in a regression model are moderately or highly correlated (Simon and Young, 2017), causing redundant information. What a predictor explains about the response is overlapped by what another predictor or a set of predictors already explain with the result that one of the predictors can be linearly predicted from the others (Yoo et al., 2014).

When collinearity occurs, the correlation coefficients β_i may change erratically in response to small changes of the data. It does not reduce the predictive power of the model as a whole, but it affects calculations regarding individual predictors. This may result in coefficients to be found statistically significant, when they actually are not.

A measure to assess collinearity is the Variance Inflation Factor (VIF). It expresses the degree to which collinearity among predictors degrades the precision of an estimate (Burns and Burns, 2008). It is calculated as the reciprocal of the inverse of the coefficient of determination R^2 of the regression of a single explanatory variable against another explanatory variable:

$$VIF_i = \frac{1}{1 - R^2}$$

Equation 4.4-5

From the Equation 4.4-5 it can be deduced that if the two explanatory variables are not correlated, their coefficient of determination R^2 is zero and the VIF will be one. Similarly, the stronger they correlate, the higher will be the coefficient of determination and the stronger their collinearity.

Material and methods

Because the standard VIF cannot be used for predictors with more than one degree of freedom such as categorical variables with more than two levels, multicollinearity was assessed using the Generalized Variance Inflation factor (GVIF). This notion, proposed by Fox and Monette (1992), corrects the VIF to account for the number of degrees of freedom of the predictor.

There is no universal agreement as what the cut-off based on values of VIF should be used to detect multicollinearity (P. Vatcheva and Lee, 2016). Rogerson (2001) suggests a VIF higher than 5 indicates multicollinearity problems, whereas other authors suggest a VIF value greater than 10 is a concern (Burns and Burns, 2008). For the purpose of this work variables with a VIF equal to or higher than 4 were considered collinear and were removed from the model.

4.4.3.8   Model testing

A test was performed to show if the regression model showed a statistically significant relationship between the response and explanatory variables. The test used to compute probability values was an F-test (Fischer distribution) under the following hypothesis:

Null hypothesis   H$_0$: $\beta_i = 0$ $\rightarrow$ all β_i regression parameters are equal
Alternative hypothesis   H$_1$: $\beta_i \neq 0$ $\rightarrow$ not all β_i regression parameters are equal

If we fail to reject the null hypothesis, we conclude that all β_i regression parameters are equal, indicating that the explanatory variables have no effect on the outcome variable and no linear relationship exists between the two. If we reject the null hypothesis, the alternative hypothesis H$_1$ is assumed to be true indicating that a linear relationship between the explanatory variables and the outcome variable exists.

4.4.3.9   Model selection

Models were selected backwards starting with all explanatory variables and progressively removing one variable at each step until the best adjusted coefficient of multiple determination (adj. R^2) was reached. For each step, regression parameters α_0, β, adj. R^2 and test results were reviewed.

4.4.3.10 Regression diagnostics

The following diagnostics plots were used to check if the models fulfil the regression assumptions or if they deviate considerable form it (see Chapter 4.4.3.6 for a listing of regression assumptions).

4.4.3.10.1   Residuals vs. fitted values plot

A scatter plot of the residuals on the y-axis against the fitted values on the x-axis was used to check the homogeneity of variance.

If the variance is homogeneous, the residuals should be approximately randomly distributed around zero. A model showing evenly distributed residuals meets the regression assumption of homogeneous distribution of residuals. If the residuals follow a distinct pattern, the variance is unequal, indicating a certain characteristic of the data is not properly caught up by the explanatory variables (Osborne and Waters, 2002).

A slight deviation on variance homogeneity has little effect on significance tests. But when heterogeneity is marked it can lead to serious distortions of findings, weakening the prediction power of the regression (Osborne and Waters, 2002) and yield inaccurate model predictions.

4.4.3.10.2   *Normal Q-Q plot of the standardized residuals*

The Quantile-Quantile (Q-Q) plot was used to check the assumption of normal distribution of residuals. The quantiles of the standardized residuals were plotted on the y-axis and compared to the quantiles of a theoretical normal distribution on the x-axis.

The standardized residuals are defined as the residuals divided by the standard deviation:

$$standardized\ residuals = \frac{residuals}{\sqrt{s^2}}$$

Equation 4.4-6

A straight line indicates residuals follow a normal distribution, whereas a curvature indicates a deviation from it. A downward concavity in a Q-Q plot indicates negative skewness (long tail to the left) of the residuals whereas an upward convexity indicates positive skewness of the residuals. Further, an S-shape is an indicator of an excess of extreme residuals relative to a normal distribution (heavy tail).

4.4.3.10.3   *Scale Location: standard deviation of residuals vs. fitted values*

A scatter plot of the standard deviation of the residuals versus the fitted values was used to assess homogeneity of variance. Similar to the residuals vs. fitted plot but using absolute values instead, this graph is more sensible to detect skewness of distribution.

4.4.3.10.4   *Cook's distance plot: influential points*

The Cook's distance was used to identify influential points in the model (see Chapter 4.3.6 for an explanation of Cook's distance method).

A chart with the CD on the y-axis and the number of the observation on the x-axis was plotted. A red-dashed line representing the critical value was included in the graph. Points above the critical distance of 4/n were subject to further analysis to consider if they shall be included in the model.

5 Results

5.1 Descriptive statistics

5.1.1 TP retention

For a total of 98 observations from all the studies included in the assessment, the smallest TP retention observed was 10% and the highest 97%. The mean retention was 66%, whereas the median 70 %. A total of 4 observations show a retention inferior to 25%.

The histogram depicted in figure 5.1-1 shows a highly-skewed modal distribution of observations following a chi-squared-like geometry with the main peak concentrating at retentions around 85%. Further, there is a plateau in the range of 50% to 75% TP retention.
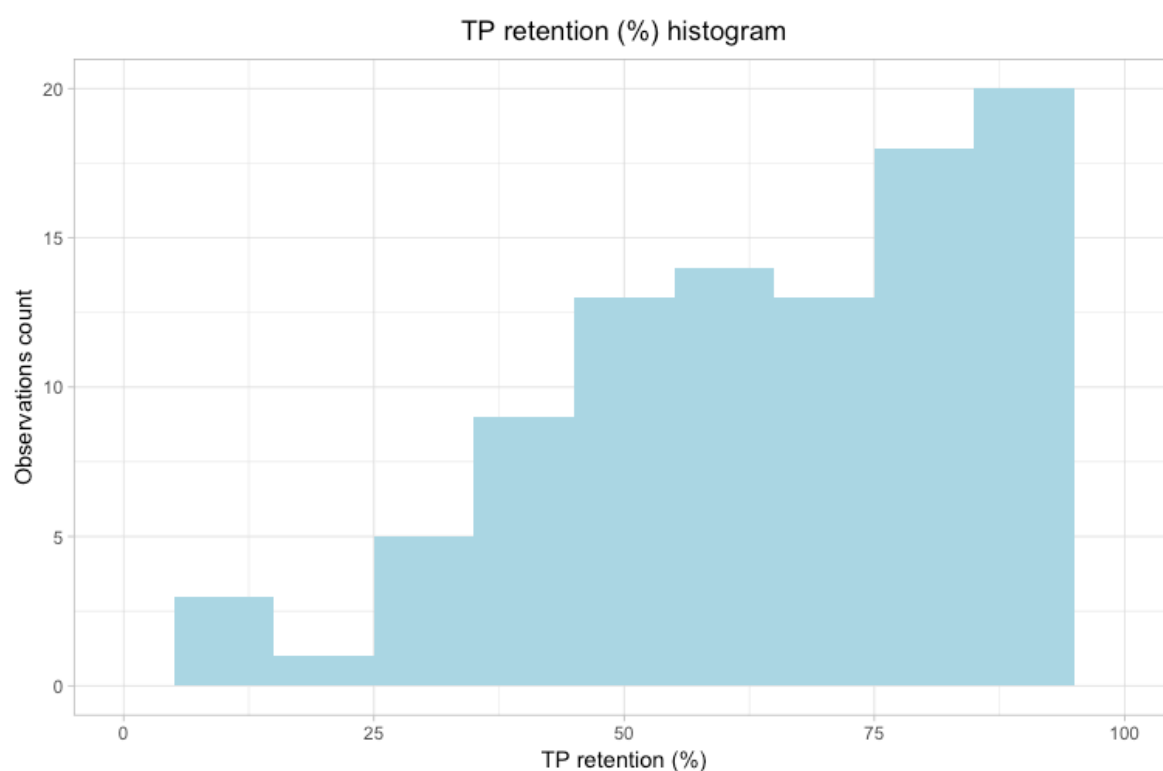
Figure 5.1-1 Histogram of TP retention.

5.1.2 TP retention boxplot by type of experiment

Figure 5.1-2 shows the box-plots of TP retention according to the type of experiment. Overall, the mean and median are higher for "experimental field plot" than to "cultivated land".

Observations of the category "experimental field plot" show a slightly right-skewed box-plot, with a median of 70% TP retention and a mean of 69%. Observations of the category "cultivated land" show a slightly left-skewed box-plot with a median of 57% and a mean of 58%. Interquartile range (IQR) is bigger in "cultivated land" observations with a value of 33% compared to the 28% of "experimental field plot". The whiskers show a similar longitude in both groups, stretching slightly more downwards in "cultivated land".

Further, a single observation in the group "experimental field plot" was marked as an outlier (represented by a black point) by the box-plot tool. A prior outlier analysis on a multivariate basis (see Chapter 4.3.6) discarded the point as an outlier and it is included in the observations.
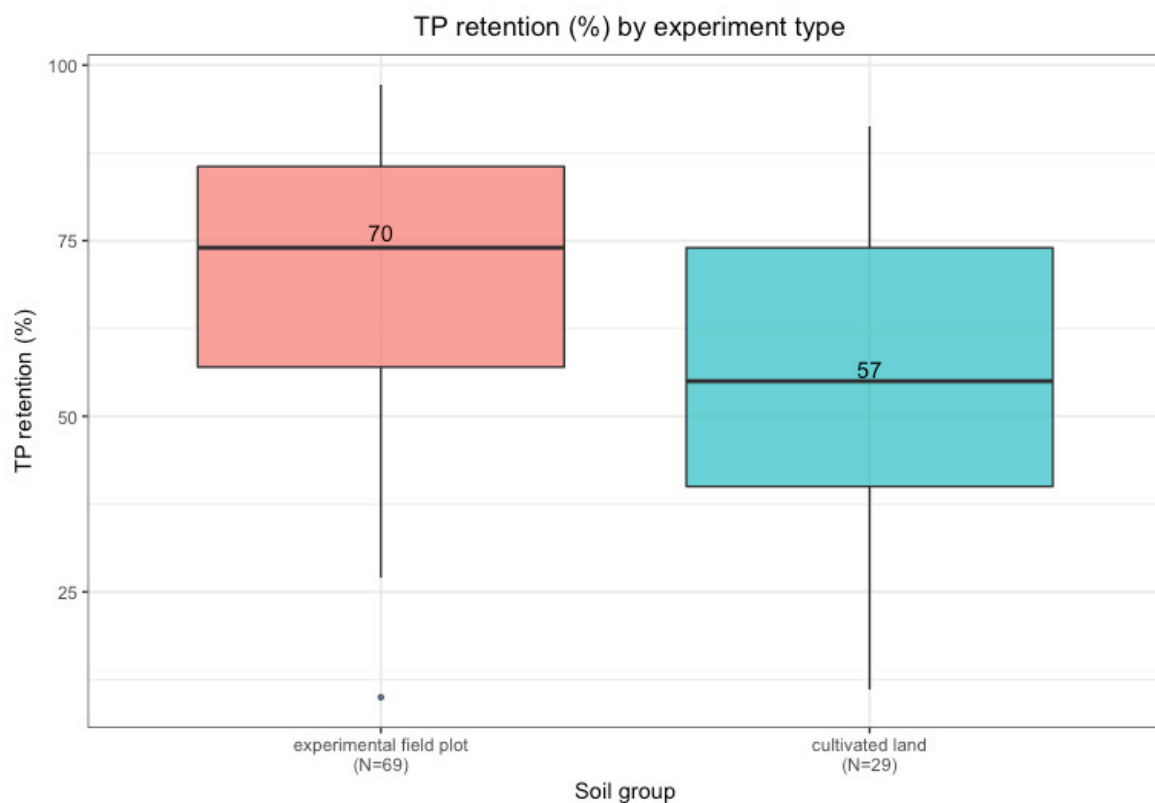
Figure 5.1-2 Boxplot TP retention by experiment type

The results show differences between the two groups, with "experimental field plots" achieving 13% higher TP retention than "cultivated land".

5.1.3 TP retention histogram by time frame

Figure 5.1-3 presents the box-plots of TP retention according to the time frame considered.

Observations of the class "single events" show with a median of 69% and a mean of 69% the highest of TP retention. The spread, with an IQR of 29% is the smallest among all three categories. Observations of the category "seasonal" display a slightly skewed boxplot with a median of 65% and a mean of 58%, almost symmetrically whiskers and an IQR of 35%. The low number of observations (*N=8*) may not properly catch the distribution of the "seasonal" values and caution should be exercised when interpreting this result. Finally, the category "annual" exhibits an appreciable right-skewed distribution with the smallest values for the median, 57% and the mean, 51%.
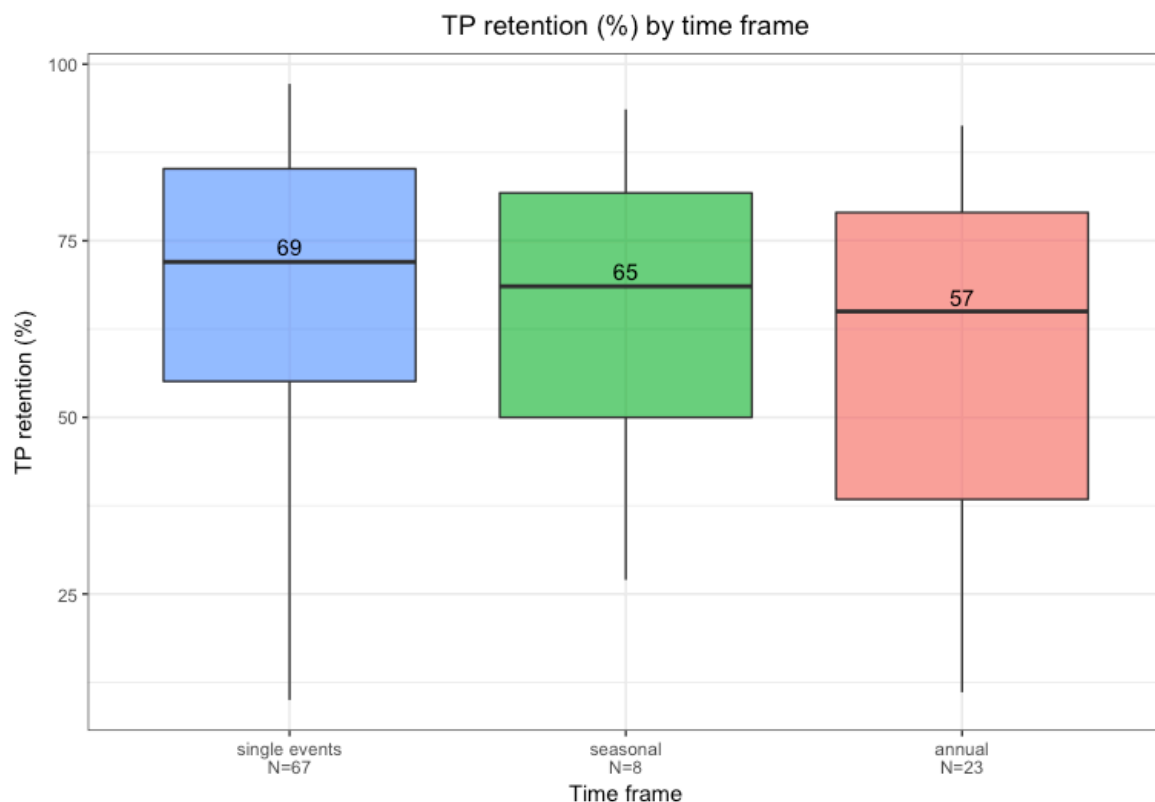
Figure 5.1-3 Boxplot TP retention by time frame

Overall, the results show a negative gradient in TP retention associated to increasing time frame of consideration. The highest TP retention performances are observed in "single events", the lowest to "annual" observations with "seasonal" outcomes falling between the two.

5.1.4 TP retention by soil textural group

Figure 5.1-4 presents the box-plots of TP retention according to soil textural group considered.

The "coarse" textural group shows a marked right-skew with a median of 85%, virtually at the same height as the Q3 value of 86%, rather away from the mean of 73%. The upper whisker is practically non-existent and the lower whisker extents downwards indicating spread of values below the first quantile. The "medium" group presents with a median of 68% and a mean of 67% a rather symmetrical distribution of values, whereas the "fine" shows a left-skewed distribution with a median of 55% and a mean of 61%. Further, the interquartile range in the latter group is with a value of 35% much bigger than in the two other groups.
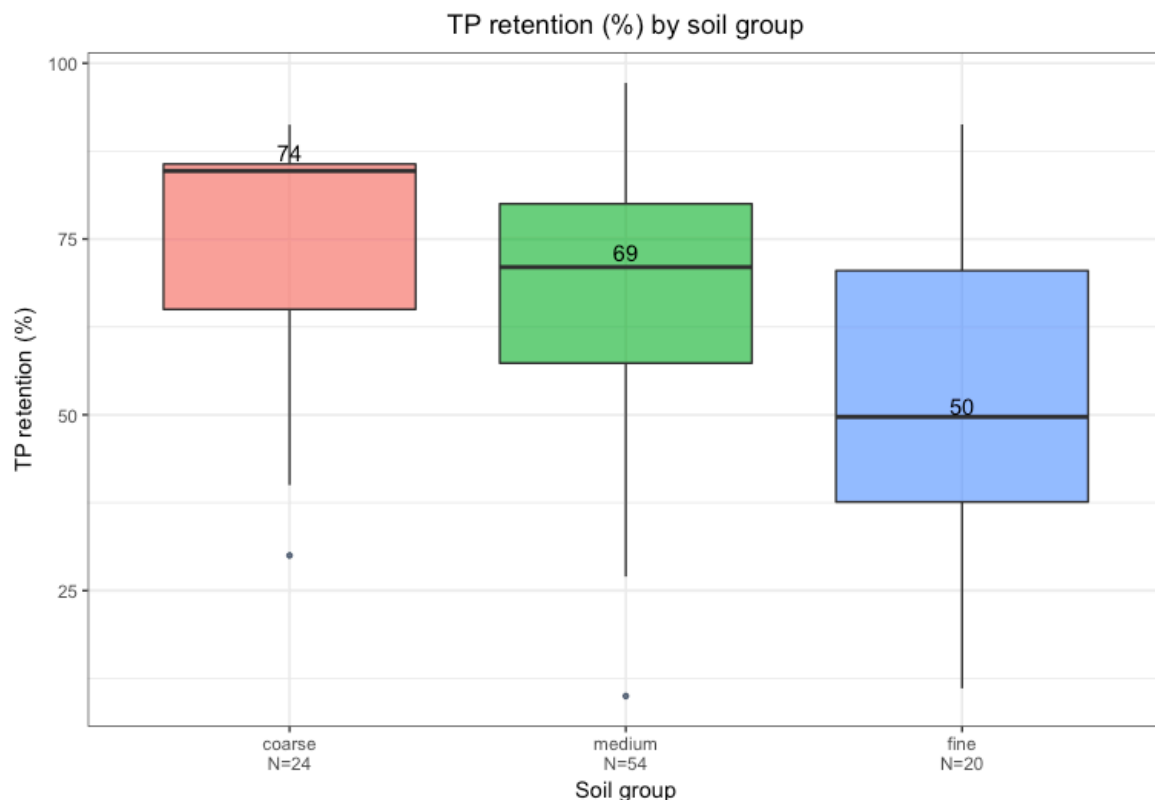
Figure 5.1-4 Boxplot TP retention by soil group

The results show identifiable differences among the three groups. The highest TP retention is associated with the "coarse" soil group, followed by "medium" and with "fine" presenting the least retention.

5.1.5   TP retention by vegetation type

Figure 5.1-5 depicts box-plots of TP retention by vegetation type.

Observations of the type "wood" form a rather symmetrical plot with a median of 72% and a mean of 72% but with just a total of 13 observations (N=13), the results may lack on power to completely represent the variability within the group so caution should be exercised when interpreting this result. Observations of the type "grass" form also a practically symmetrical distribution with a median of 67% and a mean of 65%.
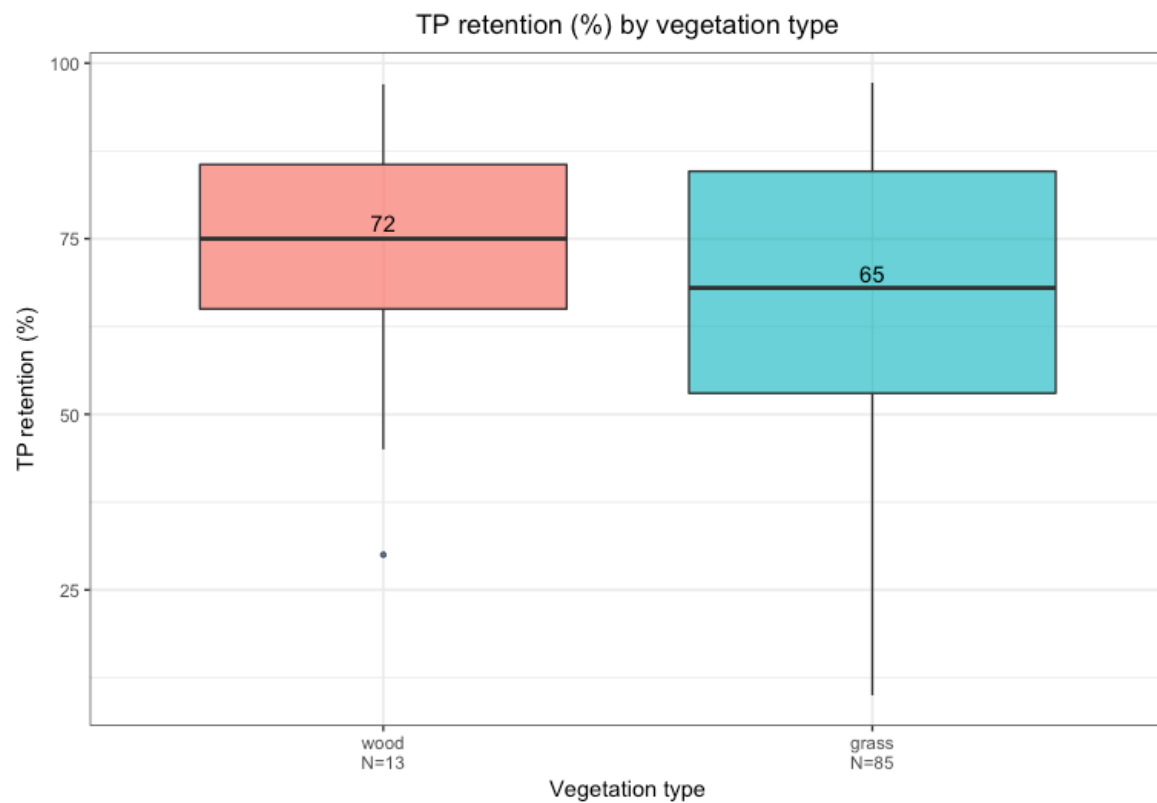
Figure 5.1-5 Boxplot TP retention by vegetation type

The results indicate slightly differences in outcomes according to vegetation group. A higher TP retention is associated with the "wood" vegetation type.

5.1.6   TP retention – VBS width scatter plot and correlation

Figure 5.1-6 illustrates a scatter plot between TP retention and VBS width, with dot colour representing type of experiment. Similarly, Figure 5.1-7 depicts the VBS width variable on a logarithmic scale.

The Spearman's correlation test gives a correlation coefficient $r_s = 0,43$ with *p-value* $= 7.55 \times 10^6$, pointing to a statistically significant positive correlation between TP retention and VBS width. Additionally, the charts reveal a concentration of observations around VBS width between 5 to 15 m, indicating that most of the experiments have been conducted at this length.

Given the strong evidence of association, the variable "VBS width" will be include in the regression analysis.
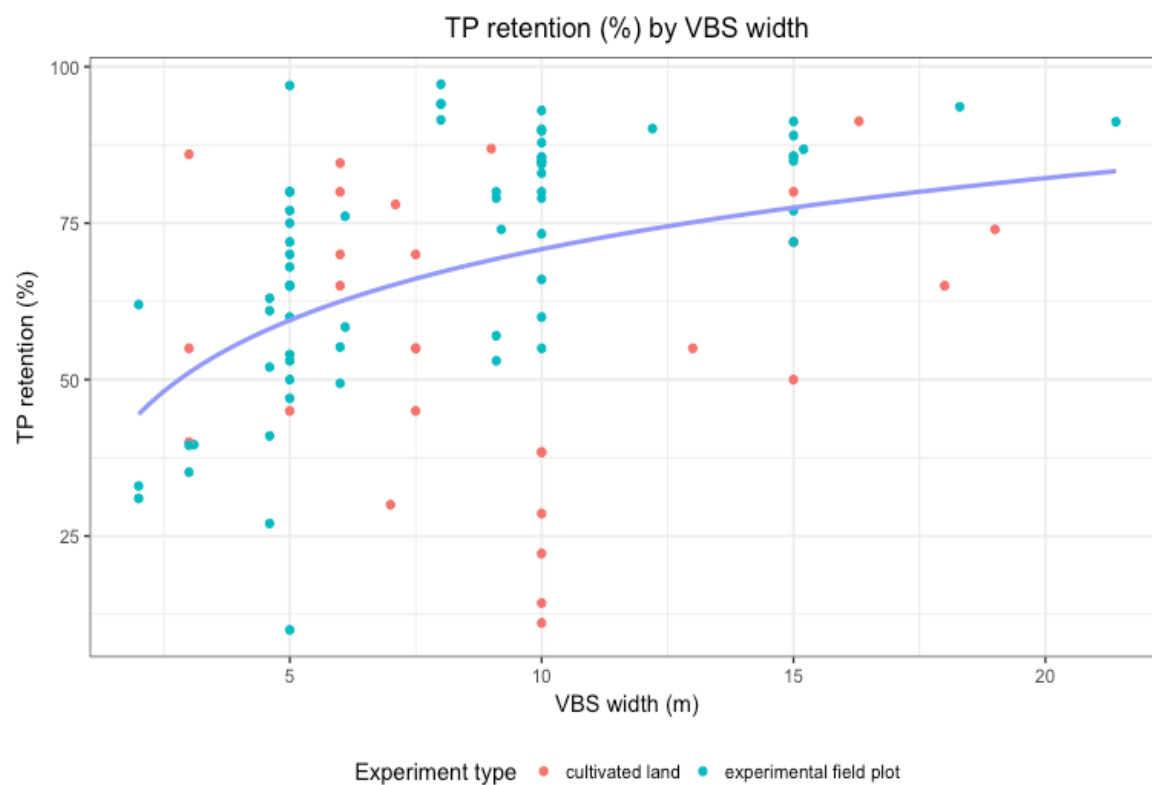
Figure 5.1-6 Scatter plot TP retention - VBS width
Dot colour represents type of experiment.
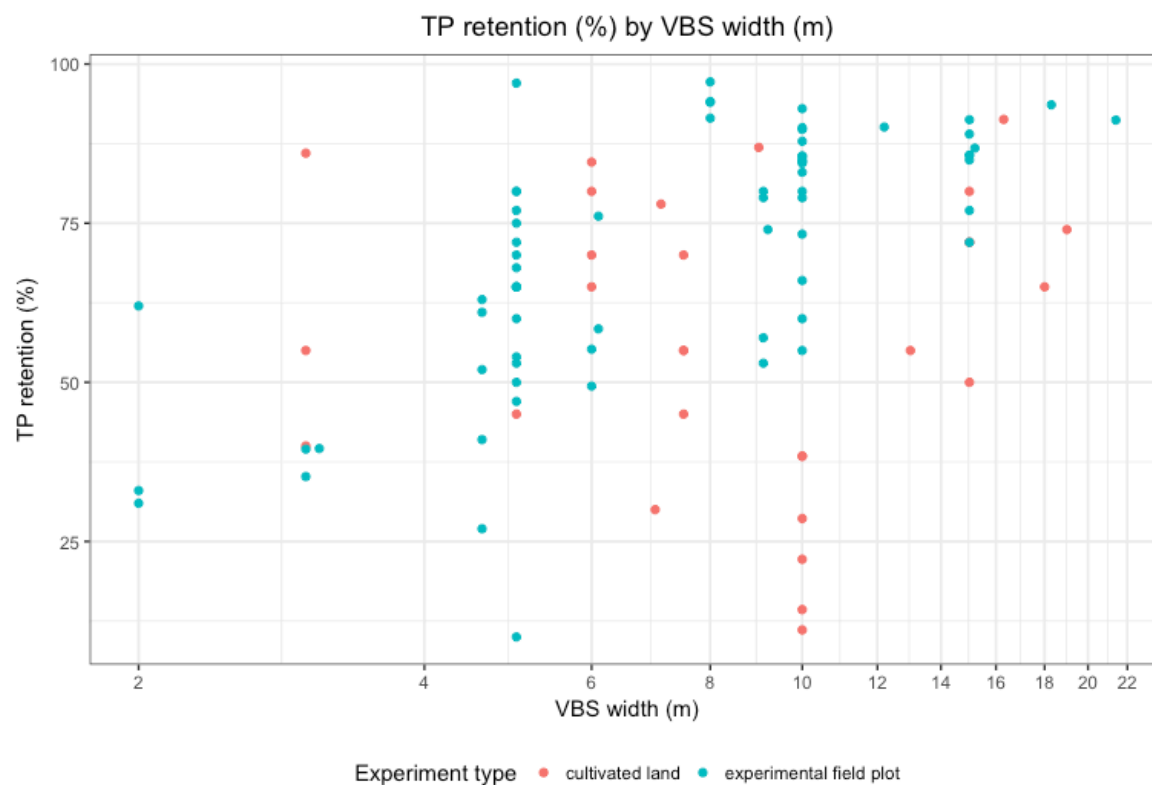
Figure 5.1-7 Scatter plot TP retention – log VBS width
Dot colour represents type of experiment.

5.1.7 TP retention – width ratio scatter plot and correlation

5.1.7.1 Scatter plot

Depicted in Figure 5.1-8 is a scatter plot of TP retention and width ratio. Overall, there is a negative association between the two variables, TP retention decreases with increasing width ratio. TP retention decreases with increasing width ratio. A Spearman's correlation test gives a correlation coefficient r_s = -0,38 with *p-value* = 2,2 x 10^{-3}, pointing to a statistically significant negative correlation. Hence, the variable will be included as part of the regression analysis.

Although the test evidences association between the two variables, a further look to the chart seems to indicate two different tendencies instead of just one. Data was analysed according to a second explanatory variable to try to find out possible underlying patterns.
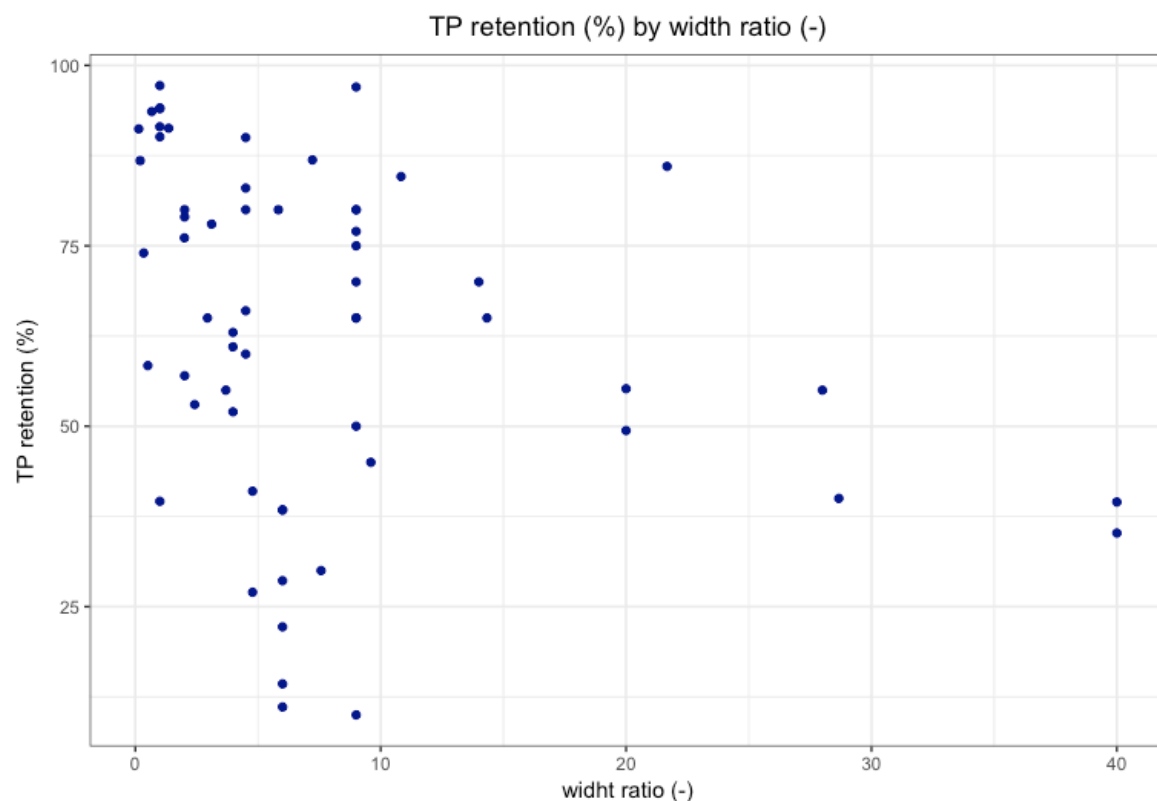
Figure 5.1-8 Scatter plot TP retention - width ratio

5.1.7.2 TP retention – width ratio scatter plot by time frame

As shown in figure 5.1-8 there is a clustering of outcomes according to the time frame of observations. Underlying the general tendency to decrease TP retention with increasing width ratio, we can see that "seasonal" observations tend to decrease much faster in outcome than values of "single events" category and that "annual" values fall whether in one or the other tendency with no clear tendency within the group. Hence, data seems to be clustered in two groups with the same direction but different slope.

A Spearman's correlation for paired "time frame" data groups was performed to find out if a certain group is associated with one of the two tendencies observed in the graphic. The results, reported in Table 5.1-1 reveal how the r_s of the group "seasonal" with a value of -0,94 is significantly different to that of the group "single events" r_s = -0,36 or the aggregated "single events + annual" r_s = -0,48 group. Further, the variability of the group "seasonal" might not be well represented due to the low number of observations. Given the low degree of representativeness and the different degree of association, data with "seasonal" values follow, it will be excluded from the regression analysis.
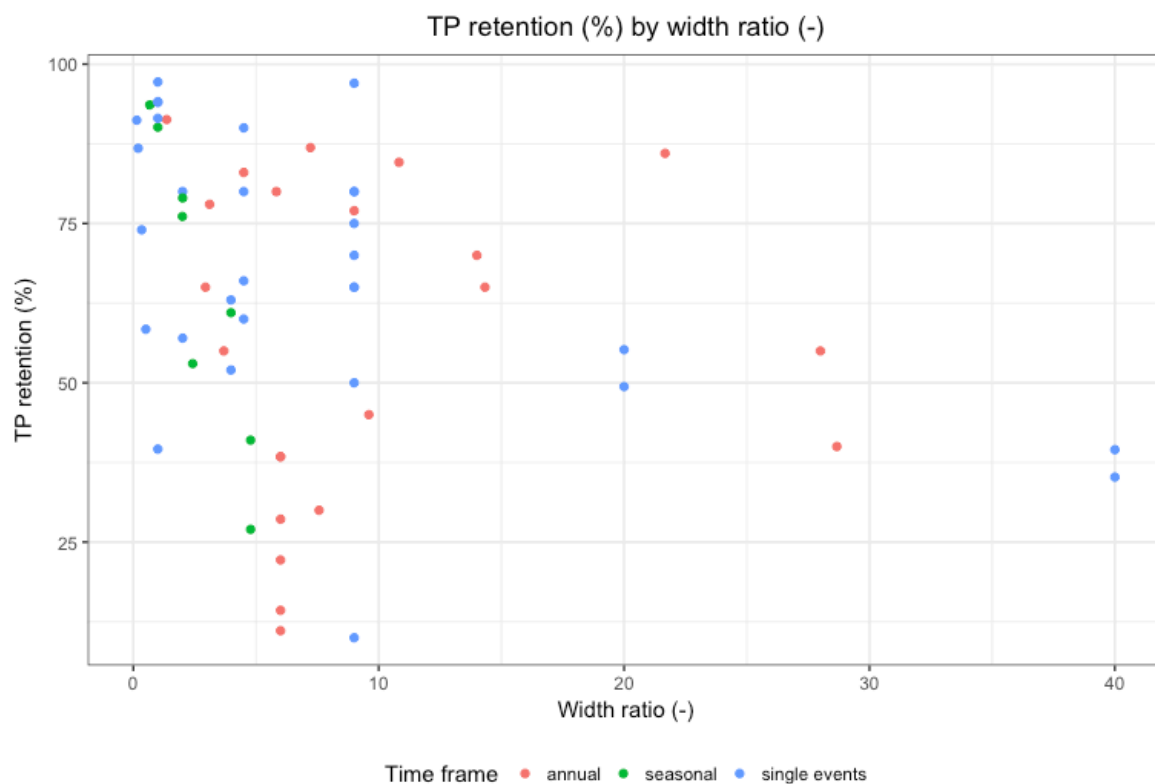
Figure 5.1-9 Scatter plot TP retention - width ratio by time frame
Dot colour represents time frame of observations.

Table 5.1-1 Comparison of data subsets in function of "time frame" factor

	subsets	Spearman's r_s	p-value
comparison 1	single events + annual	-0,36	$8,8 \times 10^{-3}$
	seasonal	-0,94	$3,8 \times 10^{-4}$
comparison 2			
	single events	-0,48	$7,23 \times 10^{-3}$
	seasonal + annual	-0,29	0,11

5.1.7.3 TP retention – width ratio scatter plot by VBS slope

Figure 5.1-10 highlights another possible explanation to the differences in outcomes might be due to the slope class. Observations falling into the category "moderate" slope (VBS slope equal or higher to 8%), follow a different tendency than observations falling in the category "gentle" slope (VBS slope less than 8%). A tendency line has been added to the figure to aid with the interpretation.

The Spearman's correlation gives a correlation coefficient of r_s = -0,57 with *p-value* = 5,5 x 10^{-5} for gentle slopes and of r_s = -0,02 with *p-value* = 0,92 for the moderate slope class, exposing a markedly different degree and quality of association. The moderate slope class shows almost no degree of association and values of this group will not be included in the regression model.
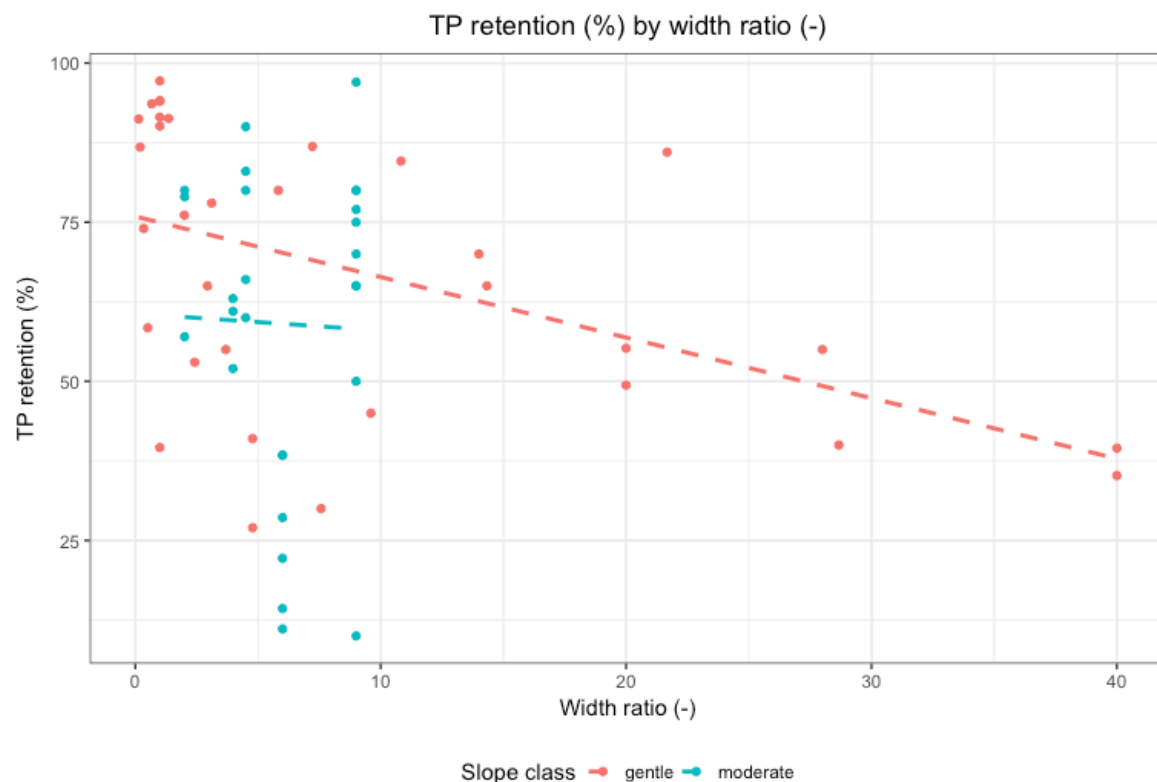
Figure 5.1-10 Scatter plot TP retention - width ratio by slope
Dot colour represents time frame of observations. Tendency line is added to aid with interpretation.

5.1.8   TP retention – VBS slope scatter plot and correlation

No degree of association can be graphically observed between TP retention and VBS slope for the plot in Figure 5.1-11. This contrasts with the results from the Spearman's correlation between the two variables, with a r_s = -0,2473 and a p-value = 0,0136, indicating a weak but statistically significant correlation between the two variables. The tendency line added in Figure 5.1-11 points out to a weak negative tendency in data for each of the groups and slightly higher outcomes for the group "gentle" in comparison to group "moderate".

A further look into Figure 5.1-11 exposes that the variable VBS slope is not equally represented across the slope range and points concentrate in two intervals stretching along slopes below 7% and slopes between 11% and 20%, creating a gap between the two intervals. A revision of the data reveals there is a lack experimental studies for the slope range between 7% to 11% with no apparent reason. Further, no other studies in the literature could be found for this slope range.

In addition, the plot reveals that spread varies with slope. For a slope of 3% outcomes range from 35% to 86% TP retention, whereas for a slope of 6,5% the range varies from 45% to 80% and for a slope of 14% it varies between 50% to 90%.
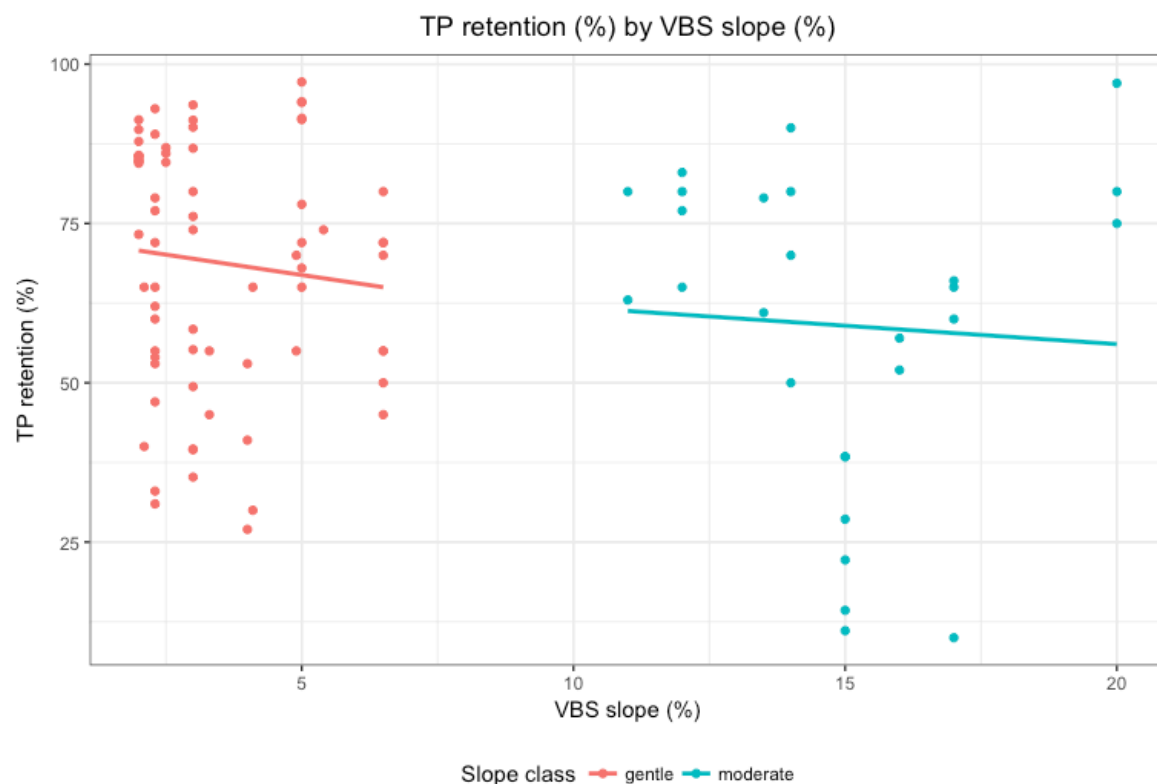
Figure 5.1-11 Scatter plot TP retention - VBS slope
Dot colour represents time frame of observations. Tendency line is added to aid with interpretation.

5.1.8.1   Boxplot for VBS slope class

Shown in Figure 5.1-12 is the box-plots of TP retention by type of VBS slope class. Observations falling into category "gentle", with VBS slopes equal or less than 8%, form a practically symmetrical plot with a median of 69% and a mean of 68% for a total of 71 observations. Observations of the category "moderate", with VBS slopes higher than 8%, presents minimal right-skewed distribution a median of 59% and a mean of 57%.

Although, similarly distributed, the figure shows a difference in outcome between the two groups with higher outcome values and less spread associated with gentle VBS slopes.
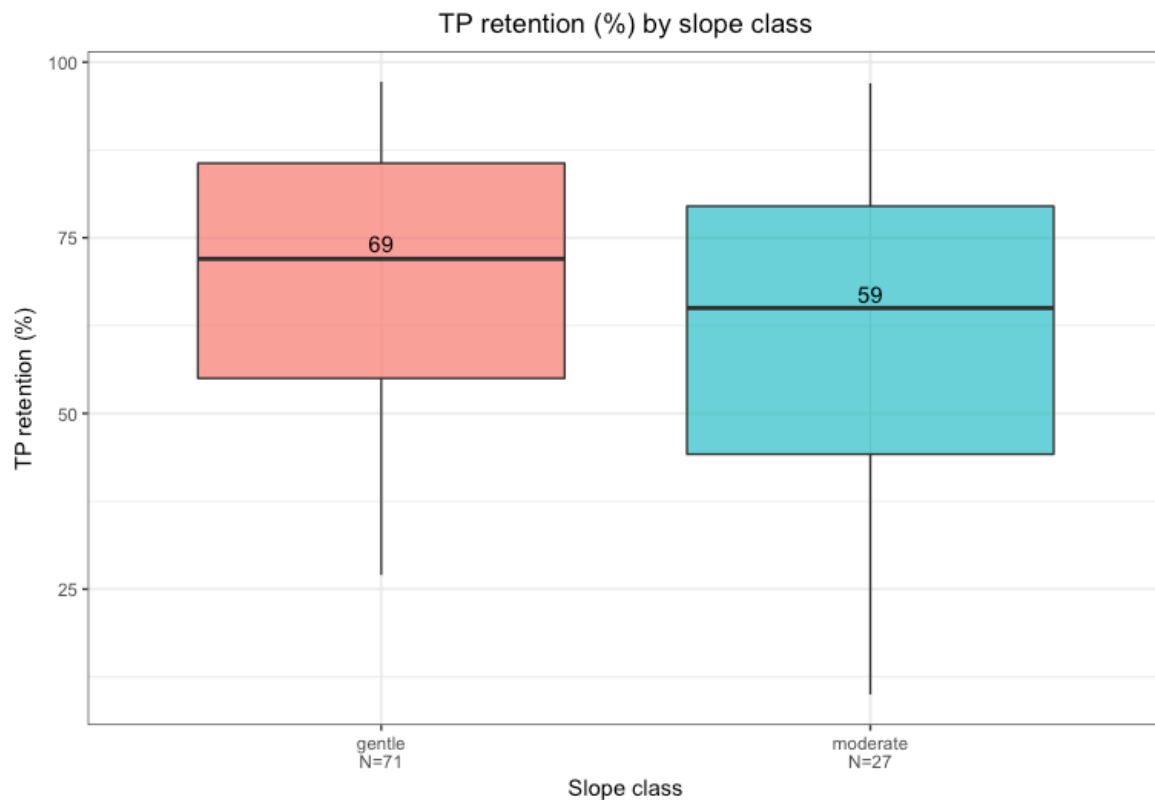
Figure 5.1-12 Boxplot TP retention by slope class

5.2 Deductive Statistics

5.2.1 Correlation between TP retention and numerical explanatory variables

5.2.1.1 Correlation between TP retention and VBS width

A check of the t-test assumptions revealed the data was fit for the inference analysis, even if the assumption of normal distribution of residuals was not met. The residuals present a small skew from the normal distribution but considering the size of the sample is bigger than 30, the results of the t-test can be accepted as valid.

With a probability value *p-value* = 7.55 x 10^{-6} under the significance level of 0,05 for 96 degrees of freedom, there is enough evidence to reject the null hypothesis and accept the alternative hypothesis that states there is an association between TP retention and the VBS width.

From the results of the test we can deduce there is a statistically significant positive correlation between TP retention and the log-transformed VBS width: the higher the VBS width, the higher the TP retention we can expect.

5.2.1.2   Correlation between TP retention and field length

A check of the t-test assumptions revealed the data was fit for the inference analysis, even if the assumption of normal distribution of residuals was not met. The residuals present a small skew from the normal distribution but considering the size of the sample is bigger than 30, the results of the t-test can be accepted as valid.
With a probability value *p-value* = 0,0036 under the significance level of 0,05 for 54 degrees of freedom, there is enough evidence to reject the null hypothesis and accept the alternative hypothesis that states there is an association between TP retention and the field length.

From the results of the test we can deduce there is a statistically significant negative correlation between TP retention and the field length: the higher the VBS length, the lower the TP retention we can expect.

5.2.1.3   Correlation between TP retention and width ratio

A check of the t-test assumptions revealed the data was fit for the inference analysis, even if the assumption of normal distribution of residuals was not met. The residuals present a small skew from the normal distribution but considering the size of the sample is bigger than 30, the results of the t-test can be accepted as valid.

With a probability value *p-value* = 0,0017 under the significance level of 0,05 for 58 degrees of freedom, there is enough evidence to reject the null hypothesis and accept the alternative hypothesis that states there is an association between TP retention and the width ratio.

From the results of the test we can deduce there is a statistically significant negative correlation between TP retention and the width ratio: the higher the width ratio, the lower the TP retention we can expect.

5.2.1.4   Correlation between TP retention and VBS slope

A check of the t-test assumptions revealed the data was fit for the inference analysis, even if the assumption of normal distribution of residuals was not met. The residuals present a small skew from the normal distribution but considering the size of the sample is bigger than 30, the results of the t-test can be accepted as valid.

With a probability value *p-value* = 0,0125 under the significance level of 0,05 for 96 degrees of freedom, there is enough evidence to reject the null hypothesis and accept the alternative hypothesis that states there is an association between TP retention and the VBS slope.

From the results of the test we can deduce there is a statistically significant negative correlation between TP retention and the width ratio: the higher the VBS slope, the lower the TP retention we can expect.

5.2.1.5   Summary of the test results

The Spearman's correlation is statistically significant for all considered variables. The strongest of the relationships with TP retention is shown by the log-transformed VBS width ($r_s = 0{,}44$) and the lowest with VBS slope ($r_s = -0{,}25$).

Table 5.2-1 Correlation of continuous variables with TP retention

	Spearman's rho	p-value	significant	df
log VBS.width	0,4398	$5{,}86 \times 10^{-6}$	yes	96
field.length	-0,3835	0,0036	yes	54
width.ratio	-0,3967	0,0017	yes	58
VBS.slope	-0,2512	0,0125	yes	96

5.2.2   Association between TP retention and categorical variables

5.2.2.1   Association between TP retention and soil group

An assessment of the ANOVA assumptions showed the data was fit for the analysis. The assumption of homogeneity of variance was not met, but the results of the Leven test revealed inequality of variance was still within the acceptable limits for use in ANOVA.

The ANOVA results showed a statistically significant association between TP retention and the variable soil group. With a probability value of *p-value* = $2{,}35 \times 10^{-4}$ under the significance level of 0,05 for 2 degrees of freedom, there is enough evidence to reject the null hypothesis of equality of means and accept the alternative hypothesis of difference in group means, with at least one of the soil groups scoring different from the others, proving an association between TP retention and the variable soil group.

5.2.2.2   Association between TP retention and time frame

An assessment of the ANOVA assumptions showed the data was fit for the analysis. The assumption of homogeneity of variance was not met, but the results of the Leven test revealed inequality of variance was still within the acceptable limits for use in ANOVA.

The ANOVA results showed a statistically significant association between TP retention and the variable time frame. With a probability value of *p-value* = 0,006 under the significance level of 0,05 for 2 degrees of freedom, there is enough evidence to reject the null hypothesis of equality of all population means and accept the alternative hypothesis of difference in means, proving there is an association between TP retention and the time frame considered.

5.2.2.3   Association between TP retention and experiment type

An assessment of the ANOVA assumptions showed the data was fit for the analysis. The assumption of homogeneity of variance was not met, but the results of the Leven test revealed inequality of variance was still within the acceptable limits for use in ANOVA.

Results

The ANOVA results showed a statistically significant association between TP retention and the variable experiment type. With a probability value of *p-value* = 0,006 under the significance level of 0,05 for 1 degrees of freedom, there is enough evidence to reject the null hypothesis of equality of all population means and accept the alternative hypothesis of difference in means, proving there is an association between TP retention and the experiment type considered.

5.2.2.4 Association between TP retention and vegetation

An assessment of the ANOVA assumptions showed the data was fit for the analysis. The assumption of homogeneity of variance (assumption G) was not met, but the results of the Leven test revealed inequality of variance was still within the acceptable limits for use in ANOVA.

The ANOVA results did not show a statistically significant association between TP retention and the variable vegetation. The probability value *p-value* = 0,273 is over the significance level of 0,05, there is not enough evidence to reject the null hypothesis of equality of population means, revealing that the outcome of TP retention is independent of the vegetation type.

5.2.2.5 Association between TP retention and slope class

An assessment of the ANOVA assumptions showed the data was fit for the analysis. The assumption of homogeneity of variance was not met, but the results of the Leven test revealed inequality of variance was still within the acceptable limits for use is ANOVA.

The ANOVA results did not show a statistically significant association between TP retention and the variable slope class. The probability value *p-value* = 0,0087 is over the significance level of 0,05. There is not enough evidence to reject the null hypothesis of equality of population means, revealing that the outcome of TP retention is independent of the slope class.

5.2.2.6 Summary of the test results

The results of the ANOVA showed TP retention was associated with the variables soil group, time frame and experiment type but it was independent of the variables vegetation and slope class.

Table 5.2-2 Association of categorical variables with TP retention

	p-value	significance	df
soil.group	$2,35 \times 10^{-4}$	yes	2
time frame	0,006	yes	2
experiment type	0,006	yes	1
vegetation	0,273	no	1
slope.class	0,087	no	1

Further, the summary of the test reveals the probability values obtained for the variables time frame and the variable experiment type are very similar, but no other statistical conclusions can be derived from this observation.

5.3   General linear model

5.3.1   Variables under consideration

The following explanatory variables were found out statistically significant in the bivariate analysis. Thus, they were selected for the multivariate regression:

soil group   VBS width   time frame   experiment type   field length   width ratio   VBS slope

5.3.2   Model 1: General model

Model selection was done through backward elimination of variables, excluding at each step the variable with the highest p-value until the highest adj. R^2 was reached.

5.3.2.1   Model selection

Table 5.3-1 summarizes the variable elimination process. The initial model (model 1a) was fitted with all variables and one variable was eliminated at each modelling step as follows:

Table 5.3-1 Model 1 backward elimination of variables

Model	V1	V2	V3	V4	V5	V6	V7
1a	soil.group	log(VBS.width)	VBS.slope	time.frame	extension	field.length	width.ratio
1b	soil.group	log(VBS.width)	VBS.slope	time.frame	extension	field.length	
1c	soil.group	log(VBS.width)	VBS.slope	time.frame	extension		
1d	soil.group	log(VBS.width)	VBS.slope	time.frame			
1e	soil.group	log(VBS.width)	VBS.slope				
1f	soil.group	log(VBS.width)					

Table 5.3-2 summarizes the results of each model run. The adjusted coefficient of multiple determination (adj. R^2), the p-value, the Akaike Information Criterion (AIC) and the number of observations (N) were reported. The summary values for the selected model are highlighted in bold letters.

Table 5.3-2 Model 1 summary of results

Model	1a	1b	1c	1d	1e	1f
R-squared	0,418	0,383	0,423	**0,423**	0,388	0,369
adj. R-squared	0,304	0,278	0,378	**0,385**	0,362	0,349
p	0,002	0,002	<0,000	**<0,000**	<0,000	<0,000
AIC	507	506	846	**844**	845	845
N	56	56	98	**98**	98	98

Model 1d produced the highest adjusted coefficient of determination (Adj. R^2 = 0,385) and it was chosen as the model with the best goodness-of-fit. This model was built with the variables soil group, log (VBS width), VBS slope and time frame.

An increase in Adj. R^2 can be appreciated with decreasing number of variables and increasing number of fitting records (N), because the statistic takes into account both the number of variables used for fitting the equation and the records available to do it. A noticeable change in coefficient of determination can be observed when the variable field.length is discarded from the model, increasing the number of observations available for fitting from N=56 to N=98.

5.3.2.2 Model 1

The equation describing the model consists of a constant term α_0 and two parameters, one for each of the continuous variables included in the model:

$$TP\ (\%) =\ \alpha_0 + 18{,}51 * \log\left[VBS.width\ (m)\right] - 0{,}46 * VBS.slope\ (\%)$$

Equation 5.3-1

Figure 5.3-1 depicts the model regression line for different classes of soil texture group under a slope value of 5%.
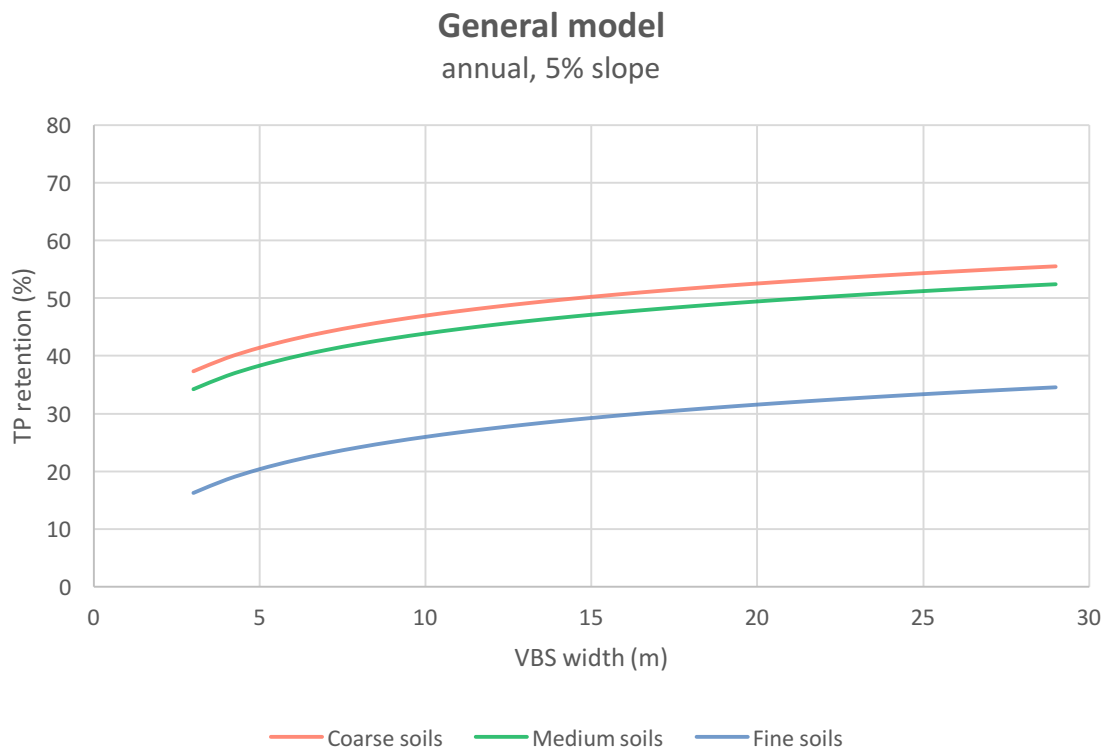
Figure 5.3-1 General model plot for annual observations and 5% VBS slope.

The constant term of the Equation 5.3-1 catches the influence of the categorical variables "soil group" and "time frame" and was calculated at reference level for the factors "coarse" and "annual". For all the other factors its value can be adjusted by adding or extracting as indicated in the Table 5.3-3.

Results

Table 5.3-3 Model 1 constant term

variable	factor	value	p-value
soil group	coarse	*30,74 (reference level)*	<0,000
	medium	-4,48	0,860
	fine	-22,36	<0,000
time frame	annual	*30,74(reference level)*	<0,000
	seasonal	+0,58	0,942
	single events	+9,06	0,051

The following can be deduced about the explanatory variables from the model equation:

▪ Soil group: there is a performance gradient with coarser soils achieving the highest TP retention. Medium-textured soils reach similar results as coarse soils, the former showing 4 percentage points lower retention rate than the latter. A real drop in performance can be observed when compared to fine-textured soils, where TP retention is 22 percentage points lower than in coarse soils.

▪ Time frame: there is a performance gradient in the time frame considered with single events achieving the highest TP retention. Annually the TP retention performance of VBS falls 9% when compared to single events.

Seasonal observations fall between the two with outcomes much closer to annual performance than single-event performance but the results for seasonal performance should be taken with due caution as the number of "seasonal" observations (N=8) used to build the model might not be representative enough.

▪ VBS width: the performance of VBS increased with buffer strip logarithmically for the range of 3 m to 29 m. A perceivable stagnation of performance can be observed from 25 m on.

▪ VBS slope: performance decreased with increasing VBS slope at an approximate pace of 0,46% per slope unit.

5.3.2.3   Collinearity

Collinearity was assessed by means of a VIF analysis which gave the following results:

Table 5.3-4 Model 1 analysis of collinearity. VIF results

Variable	GVIF	DF	GVIF^(1/(2*Df))
soil.group	1,87	2	1,17
time.frame	1,49	2	1,10
VBS.width	1,09	1	1,04
VBS.slope	1,30	1	1,14

As it can be seen in Table 5.3-4 all the results are below the VIF critical level of 4 points. Hence, the model is not affected by collinearity of its variables and no further considerations need to be taken. The highest VIF value 1,87 is shown by variable soil group and the part of the variance already explained by the other variables is significantly small. The smallest VIF value 1,09 is obtained by the variable VBS width.

5.3.2.4   Model diagnostics

The charts show a rather homogeneous variance with residuals very close to zero in the central range (60% to 85% TP retention) and a slight deviation in the extremes. The Cook's distance plot (down-right) shows the records 9, 44 and 63 cause leverage to the model but these points were not eliminated from the model. Overall, the modelling assumptions are met.

A further look at the residual vs. fitted plot (Figure 5.3-2a) shows us the residuals in the range 60% to 85% TP retention are practically zero. A slight deviation of residuals is observed in values below 50% and above 90% TP retention, increasing gradually as we tend to the extremes. The Normal Q-Q (Figure 5.3-2b) and the Scale-location plot (Figure 5.3-2c) seems to corroborate the deviation of residuals. This phenomenon reflects the structure of the data: with few records below 50% and above 90% performance, the model is better adjusted for outcomes in the range 60% to 85% TP retention.

Figure 5.3-2 Model 1 diagnostic plots, a) Residuals vs. Fitted plot b) Normal Q-Q plot) c) Scale-Location plot d) Cook's distance plot.

5.3.2.5   Summary of results

Model 1 was built to analyse the role of considered variables at predicting TP retention in VBS. Variables VBS width and soil group were found out to be the most useful at predicting TP outcome results. The model built with the variables VBS width, VBS slope, soil group and time frame yielded the best coefficient of determination and served to explain 35% of the sample variance.

Finally, model diagnostics showed no particular violation of modelling assumptions.

5.3.3   Model 2: width-ratio and VBS width

5.3.3.1   Model summary

The model wratio was fitted with the variable width.ratio, the model VBSwidth was fitted with the log-transformed variable VBS.width.

The Table 5.3-5 summarizes the results of the models, the adjusted coefficient of multiple determination (adj. R^2), the p-value, the Akaike Information Criterion (AIC) and number of observations (N) were reported.

Table 5.3-5 Model 2 summary of results

	wratio	log_VBSwidth
(Intercept)	78,81	29,18
width.ratio	-1,02	
log(VBS.width)		20,40
R-squared	0,3373	0,2999
adj. R-squared	0,3108	0,2719
p	0,0015	0,0031
AIC	235	237
N	27	27

Model wratio produced with a value of 0,3108 a higher adjusted coefficient of multiple determination than model VBSwidth for the same number of observations (N=27) and degrees of freedom (DF=25), suggesting that width ratio is a better predictor than VBS width for TP retention.

5.3.3.2   Model wratio

The model is described by a linear equation with a constant term (intercept) and one parameter (slope) as follows:

$$TP\,(\%) = 78,81 - width.ratio$$

<div align="right">**Equation 5.3-2**</div>

Results

Figure 5.3-3 depicts the model equation plotted as a blue line together with the single values used to adjust it as black dots. The shaded grey area around the model equation represents the confidence bands generated by the 95% confidence intervals of the regression line. The model parameters Adj. R^2, intercept and slope of model equation as well as the p-value are reported in the header.

The chart reveals an opposing relationship between the two variables, with decreasing TP retention as width ratio increases. Further, an uneven distribution of spread can be observed with a higher variability associated with lower width-ratio (width-ratio < 10).
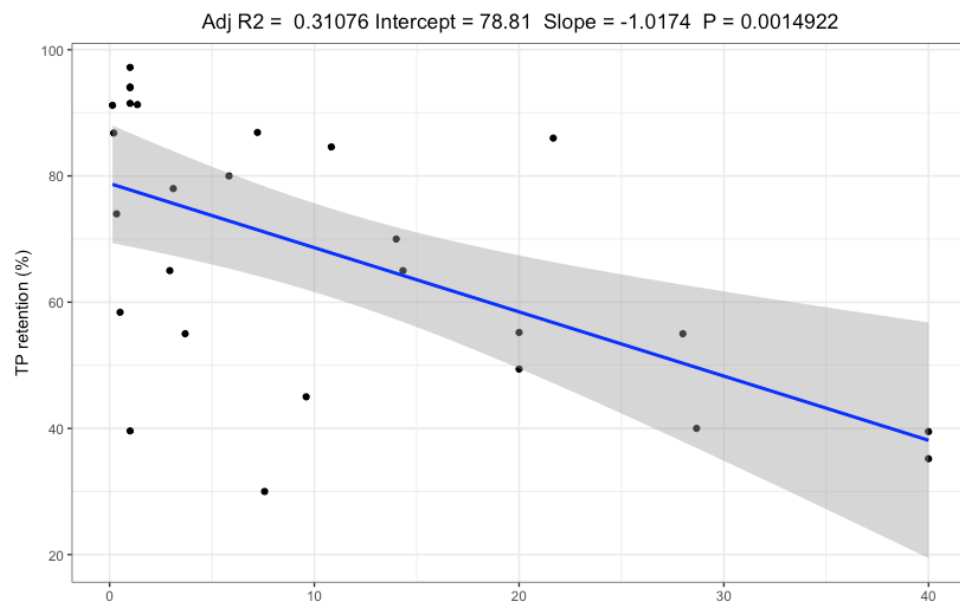
Figure 5.3-3 Model wratio plot

5.3.3.3   Model diagnostics

The residuals diagnostic charts (Figure 5.3-4a+c) show variance heterogeneity increases with increasing outcome of results and density of observations. The normal Q-Q plot (Figure 5.3-4b) shows a downward concavity, indicating a negative skewness on the distribution of residuals. Lastly, the Cook's distance plot (Figure 5.3-4d) shows points 12, 23 and 40 cause significant leverage to the model, the removal of these points would increase the model predictive accuracy but with the reduced number of observations this could come at the cost of overfitting the model so these points were not deleted from the model.

A closer look to the residual vs. fitted plot (Figure 5.3-4a) sheds to light, that variability increases above 70% TP retention. This pattern can also be seen at the Scale-location chart (Figure 5.3-4c) with a knick-point at 70%. This phenomenon is associated to the data structure: the density of observations increases with increasing TP retention and with more observations the residuals increase as well.

Overall, variance heterogeneity is acceptable enough to consider the model is fit for regression. However, the accuracy to predict outcomes might be compromised. The limiting factor is the

number of observations available for the variable width-ratio. If more records would be available for this variable, a more accurate model could be obtained.
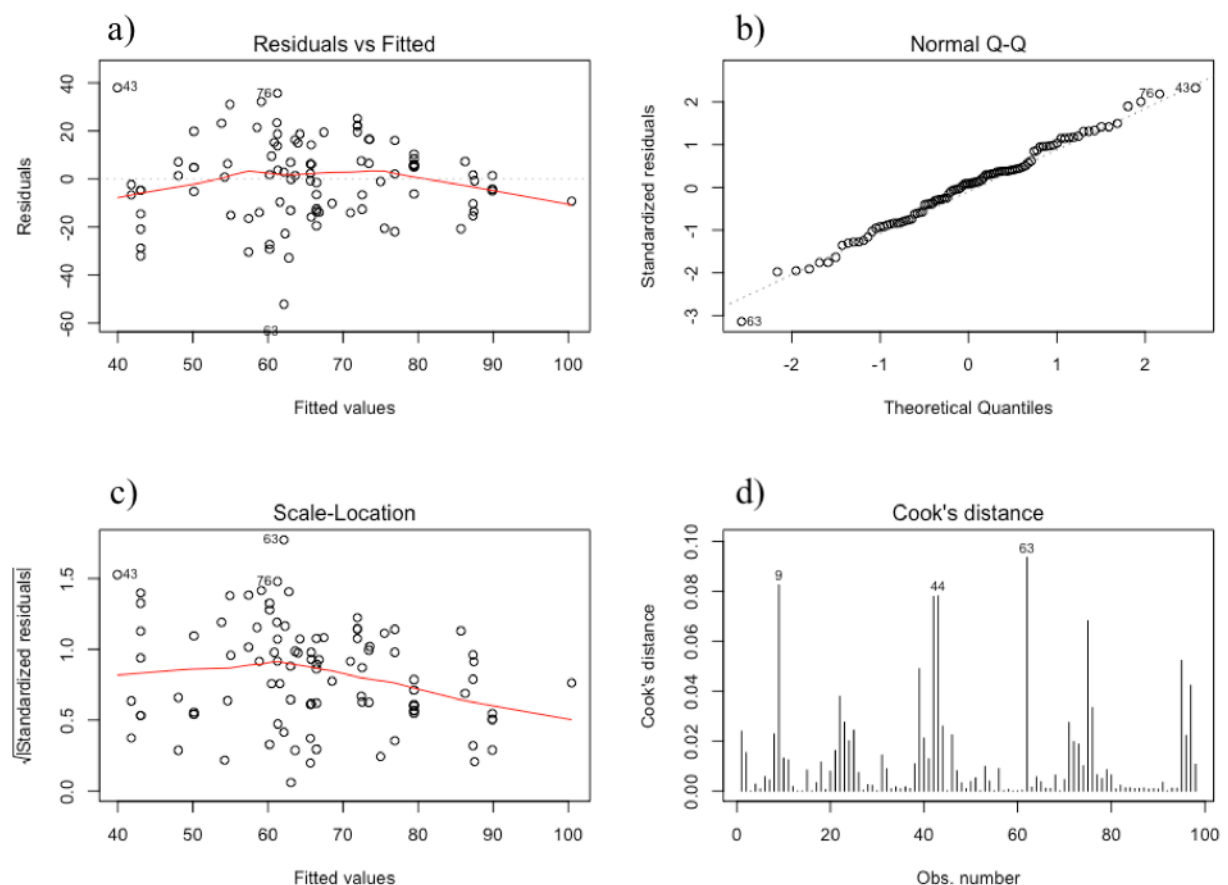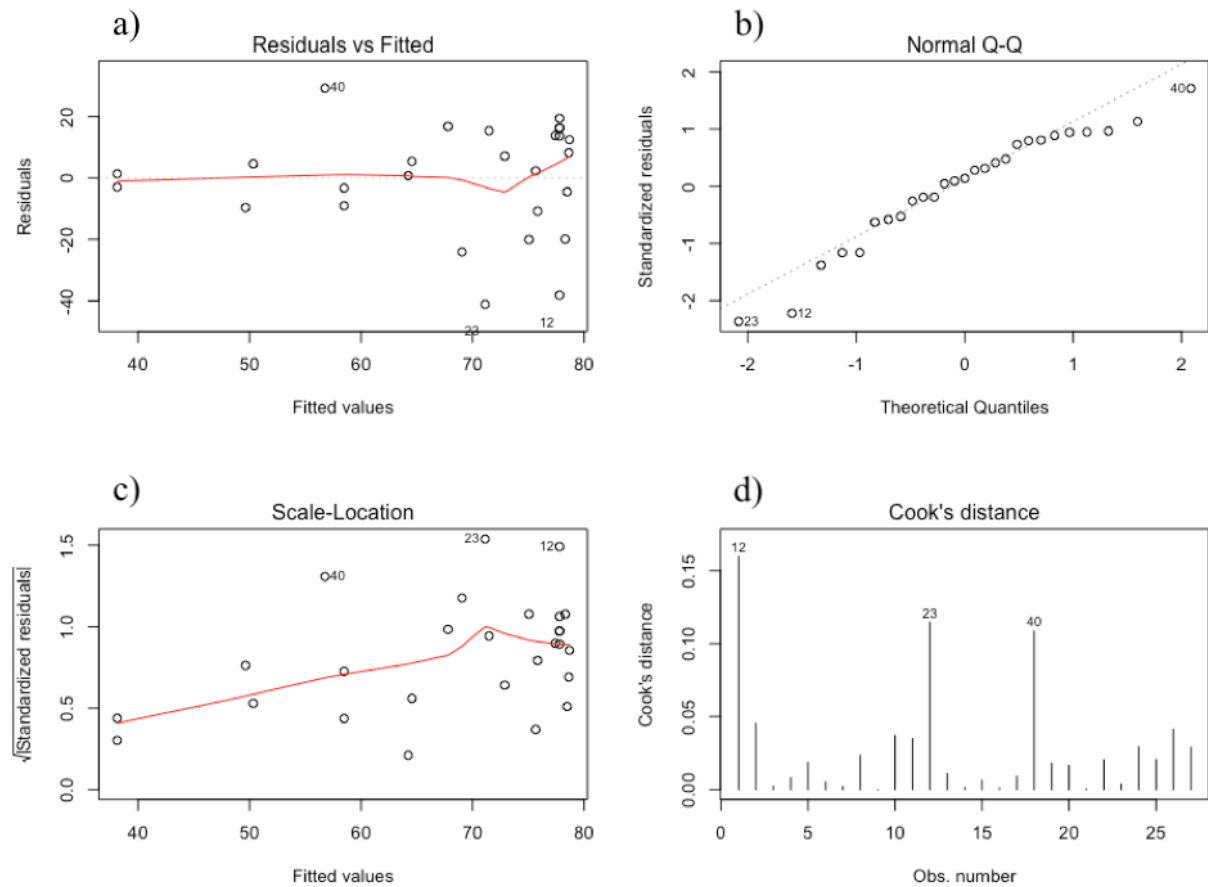

Figure 5.3-4 Model wratio diagnostic plots, a) Residuals vs. Fitted plot b) Normal Q-Q plot) c) Scale-Location plot d) Cook's distance plot.

5.3.3.4 Model VBS width

The model is described by a linear equation with a constant term α_0 (intercept) and one parameter (slope) as follows:

$$TP\,(\%) =\ 29{,}18 + 20{,}40 * \log\,[VBS.width(m)]$$

Equation 5.3-3

Figure 5.3-5 depicts the model quation plotted as a blue line with the single values used to adjust it as black dots. The shaded grey area around the model equationrepresents the confidence bands generated by the 95% confidence intervals of the regression line. The model parameters Adj. R^2, intercept and slope of model equation as well as the p-value are reported in the header.

The chart reveals a positive relationship between the two variables, with increasing expected TP retention as the log-transformed VBS width value increases. Further, it can be seen a concentration of observations in the range between 15 to 20 m.
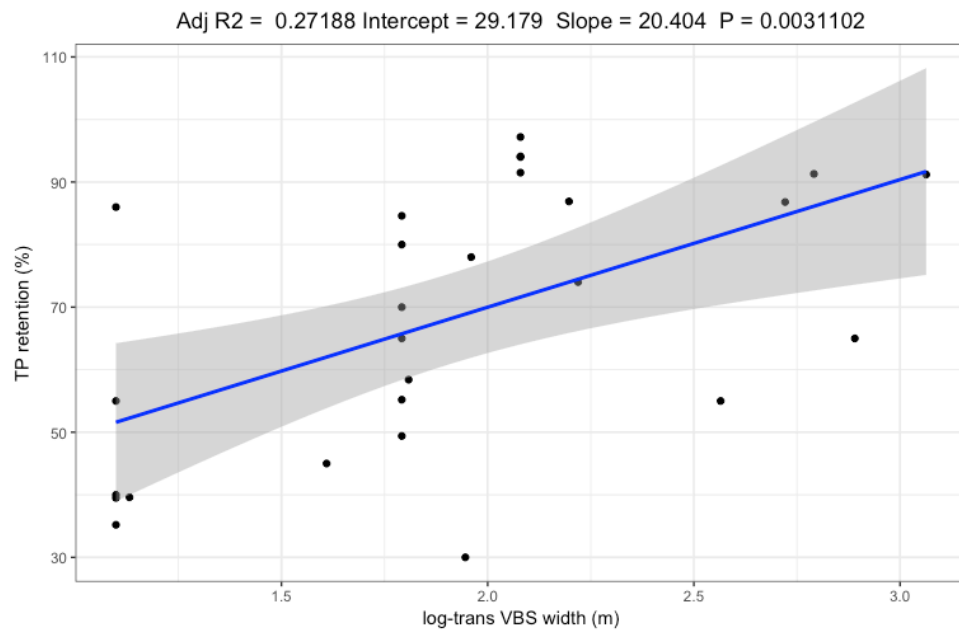
Figure 5.3-5 Model VBS width plot

5.3.3.5 Model diagnostics

The residuals diagnostic plots (Figure 5.3-6a and c) suggest a violation of the condition of variance homoscedasticity. The Normal Q-Q plot (Figure 5.3-6b) displays a S-shaped form indicating heavy tails at both extremes of the residuals distribution. The Cook's distance plot (Figure 5.3-6d) indicate points 23, 24 and 40 cause leverage to the model. Overall, diagnostics seems to indicate the data does not fit a linear regression, other regression approaches might be more adequate to describe the relationship between VBS width and TP retention.

The residuals vs. fitted plot (Figure 5.3-6a) show an irregular distribution of residuals with a triangle-shaped form: the model starts underpredicting outcome results for the lower range of outcome values, rapidly reversing this trend and overpredicting to again, underpredict values for higher outcomes. The scale-location chart (Figure 5.3-6c) shows two knick-points and an error trend, with residuals decreasing with increasing outcome. This problem might be associated to the big density of records with VBS width between 5 to 15 m but the low number of results outside this range. The limited number of observations (N=27) used to adjust the model together with the high residuals may explain the problem.

Summing up, although the model is statistically significant, the extent of the relationship between VBS width and TP retention is not correctly caught up by the model. This contrasts with the goodness of the variable VBS width for predicting TP retention when considering all the 98 available data points (see Chapter 5.1.6).
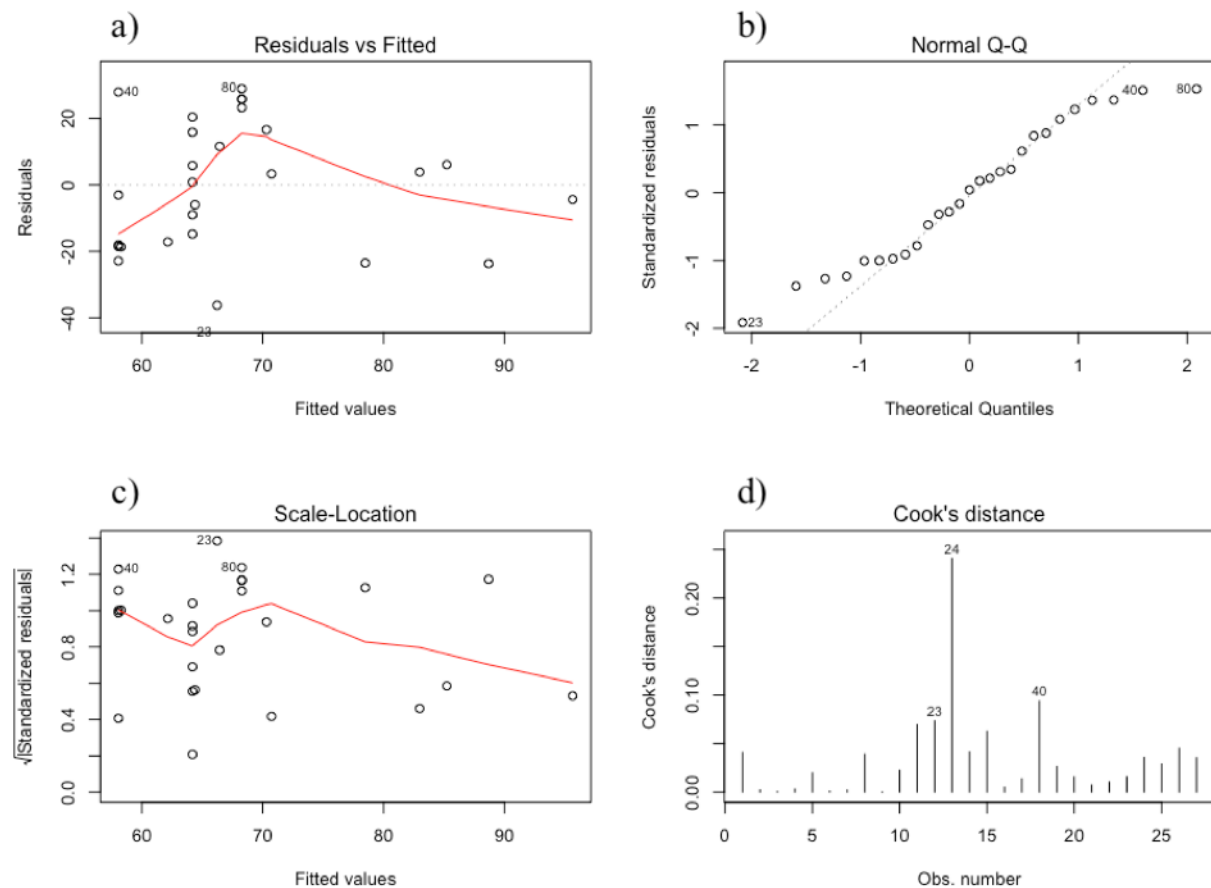
Figure 5.3-6 Model VBS width diagnostic plots, a) Residuals vs. Fitted plot b) Normal Q-Q plot) c) Scale-Location plot d) Cook's distance plot.

5.3.3.6   Summary of results

Model wratio and model VBS width were built to compare the performance of the variables width ratio and VBS width at predicting TP retention in VBS. Considering the same set of observations, the model wratio was found out to be more powerful and accurate than model VBS width, suggesting that the variable wratio is a better predictor of TP retention efficiency than VBS width.

Further, both models suffer from variance heterogeneity. In the case of model wratio variance heterogeneity is small enough not to consider it an issue on model quality, for model VBS width seems to be affected by variance heterogeneity to an extent great enough to compromise model accuracy and reliability of results.

5.3.4   Model 3: field performance

This model aims at identifying the explanatory variables which best predicts TP retention capacity of VBS under field conditions. It is built using a subset of data included in the general model (model 1) with all the entries of experiment type "cultivated land" and VBS slope smaller than 8% considering the follow variables:

soil group	VBS width	time frame	cultivated land	VBS slope

By just considering VBS slopes under 8% we ensure that the model does not suffer from the distribution issues observed on the TP retention – VBS slope scatter plot (see Chapter 5.1.8).

The variables field length and width ratio were excluded from the analysis because the limited availability of records would have considerably compromised the number of observations available to fit the model.

5.3.4.1 Model selection

Table 5.3-6 summarizes the variable elimination process. The initial model (model 3a) was fitted with the variables VBS slope, VBS width, soil group and time frame. One variable was eliminated at each step as follows:

Table 5.3-6 Model 3 backward elimination of variables

	V1	V2	V3	V4
3a		VBS.slope	VBS.width	soil.group	time.frame
3b		VBS.slope	VBS.width	soil.group	
3c		VBS.slope	VBS.width		
3d		VBS.slope			

Table 5.3-7 summarizes the results of each model run. The adjusted coefficient of multiple determination (adj. R^2), the p-value, the Akaike Information Criterion (AIC) and the number of observations (N) were reported. The summary values for the selected model are highlighted in bold letters.

Table 5.3-7 Model 3 summary of results

	model_3a	model_3b	model_3c	model_3d
R-squared	0,640	0,617	**0,569**	0,497
adj. R-squared	0,562	0,553	**0,536**	0,478
p	0,000	0,000	**>0,000**	>0,000
AIC	246	246	**246**	248
N	29	29	**29**	29

Model 3a produced the highest adjusted coefficient of multiple determination (Adj. R^2 = 0,562) and it can be considered the best fitting model. Overall, the fitness of the model decreases with decreasing number of variables, suggesting that all the variables contribute to explain a part of the variance.

If the Akakike Information Criterion (AIC) is used as model selection criterion instead of the Adj. R^2, we observe that the model 3c produces the same amount of information as model 3a with two variables less. Because the elimination of the variables time frame and soil group does not cause

a loss of information and model fitness is not severely affected, the model 3c is selected for further analysis.

5.3.4.2 Model 3

The equation describing model 3 consists of a constant term and two parameters, one for each of the continuous variables included:

$$TP\ (\%) = \ 57{,}32 + 12{,}55 * \log\ [VBS.width(m)\ ] - 3{,}98 * VBS.slope\ (\%)$$

Equation 5.3-4

Figure 5.3-7 depicts the model regression line for different slopes and widths.
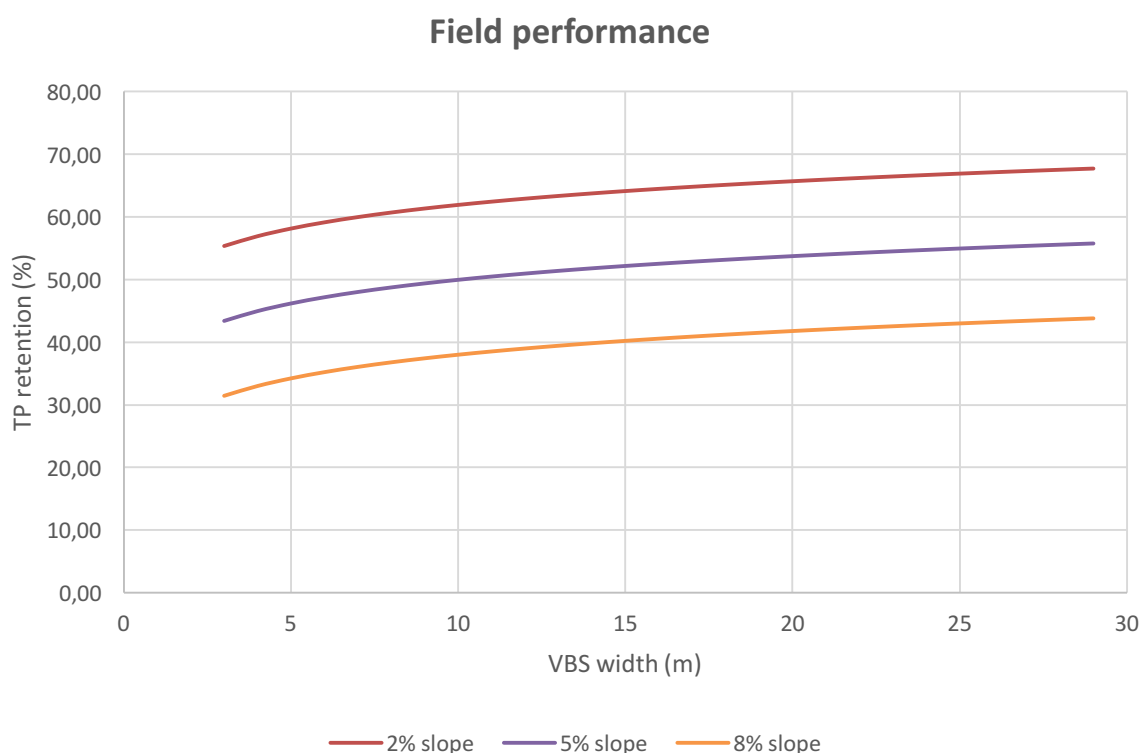
Figure 5.3-7 Field performance model

Table 5.3-8 Model 3 equation terms

variable	value	p-value
(intercept)	57,32	<0,000
VBS.width	12,55	0,0454
VBS.slope	-3,98	<0,000

The following can be deduced about the explanatory variables from the model equation:

- VBS width: the performance increased with buffer width logarithmically for the range of 3 m to 29 m. A perceivable stagnation of performance can be observed from 25 m on.

- VBS slope: performance decreased with increasing slope at an approximate pace of -3,8% per slope percent unit.

- VBS slope has a higher influence on predicted outcome than VBS width.

5.3.4.3   Collinearity

Collinearity was assessed by means of a VIF analysis and gave the following results:

Table 5.3-9 Model 3 analysis of collinearity. VIF results

Variable	GVIF
VBS.slope	1,04
VBS.width	1,04

As it can be appreciated in Table 4.3-1 the results lay below the critical VIF value of 4 points. Thus, it can be considered that there are no collinearity issues between the variables VBS slope and the variable VBS width.

5.3.4.4   Model diagnostics

Model diagnostics show a deviation from the assumption of variance homogeneity. The residuals are, in fact, closely distributed around zero and variance starts to deviate at the very top-end of the fitted values (Figure 5.3-8a). This trend seems even stronger when observing the Normal Q-Q plot (Figure 5.3-8b). The two knick-points in the scale-location plot (Figure 5.3-8c) show variance homogeneity is not satisfied. A possible explanation for the knick-points might be the polarization of data records: while most points concentrate at the lower and upper end of the outcome range, there is a gap in outcomes between 30% to 55% TP retention.

The violation of homogeneity of variance is an issue which should be considered when using the model but is no reason to reject it. In view of the number of records used to build the model, it can still yield fairly accurate results. However, a bigger, random sample would be needed to stabilize variance and provide the accuracy needed for proper forecasting.

That said, the model is consistent with results available in contemporary literature and should be taken seriously as a first approximation of what can we expect of VBS for TP retention.
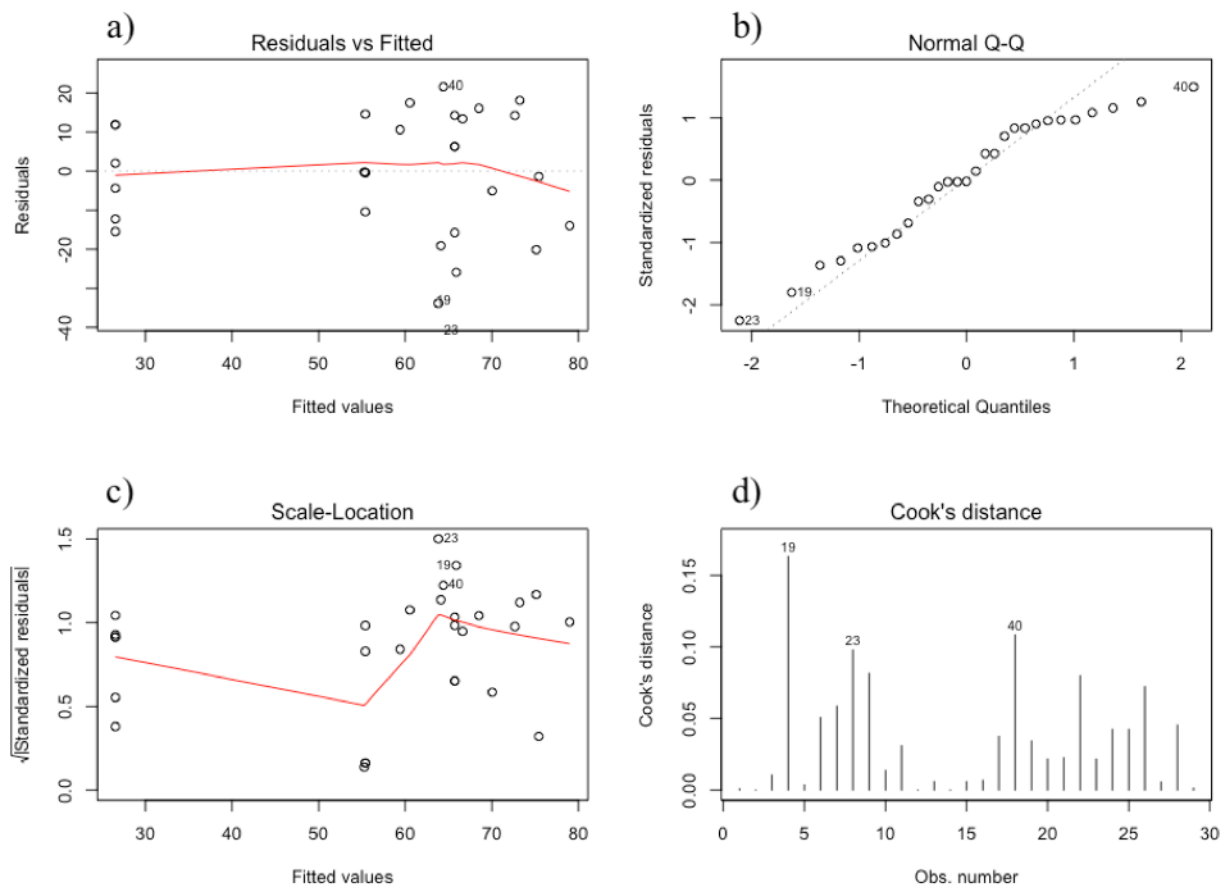
Figure 5.3-8 Model 3 diagnostics, a) Residuals vs. Fitted plot b) Normal Q-Q plot) c) Scale-Location plot d) Cook's distance plot.

5.3.4.5 Summary of results

Model 3 was built to assess the predictive power of the considered variables at predicting TP retention under field conditions in VBS with slopes equal or smaller than 8%. The model built with variables VBS width and VBS slope were chosen as the best model and served to explain 54% of the sample variance.

The model 3 suffers from variance heterogeneity but not beyond a point where it should be rejected.

6   Discussion of results

In this study, we identified key parameters influencing phosphorus retention through vegetated buffer strips and analysed their effect on filter efficiency based on experimental data including 98 records from 22 experiments.

Our findings reveal statistically significant differences (p = 0,006) in outcome between data from experimental plots and cultivated fields, the first group achieving average performances of 69% compared to 58% in the second, hinting at a 11% advantage for experimental plots. Hence, we conclude there is a difference performance under field conditions and experimental plots. This conclusion could not be backed by previous results due to a lack of similar investigations. We understand that the difference in performance must be related to spatial differences between the two groups, in particular by the phenomena of flow convergence (Blanco-Canqui et al., 2006; Muñoz-Carpena et al., 1999; White and Arnold, 2009) and buffer bypass (Hösl et al., 2012), which has been observed to limit buffer function under field conditions but is seldom promoted in experimental plots.

We also found statistically significant (p = 0,006) differences in buffer performance according to the time frame of observations, suggesting that buffer strips lose retention capacity with time. Single-event experiments were able to retain a 69% of total incoming phosphorus, whereas in long-term multi-annual experiments performance dropped to an average of 51%. Our results conform to previous studies pointing out higher performances on single-events (Yang et al., 2015) compared to long-term performance (Borin et al., 2005; Daniels and Gilliam, 1996; Uusi-Kämppä, 2008). We understand that these differences are due to time-related buffer processes like buffer saturation, which only becomes apparent during the course of years (White and Arnold, 2009) but also to the slow velocities of phosphorus cycling in soil (Roberts et al., 2012) and the periodization of geochemical and microbiological phosphorus-related processes (Roberts et al., 2013). These processes take place during days or weeks and are strongly influenced by environmental site conditions i.e. soil temperature and humidity.

The variable time-frame was used to build the regression model 1. The results of the model corroborate what we found out with ANOVA, exposing an almost 10% advantage in performance for single events over annual observations. This result was expected as ANOVA and linear regression base on the same statistical method. Interestingly, model 1 showed a greater p-value for the single events observations. We understand that the equation used to build the model is better adjusted for annual observations, which might partly be due to the greater variability of the group single events. We speculate that the influence of single factors affecting buffer performance might level out with time. However, this last claim should be taken with due care as further in-depth research beyond the scope of this study would be needed to confirm this assumption.

Soil texture of the vegetated buffer also serves as an explanation for buffer performance (p = 0,35 x 10^{-4}), with coarser soil textures associated with higher performance rates. We obtained an average TP retention of 73% in coarse-textured soils, compared to 67% in medium and just 51%

Discussion of results

in fine-textured soils. These results are in consonance with previous findings, claiming an advantage to sandy soils (Dorioz et al., 2006; Dosskey et al., 2008; Magette et al., 1989; White and Arnold, 2009). This relative advantage seems to be explained by the higher permeability of sandy soils, which reduces transport capacity of runoff more effectively than finer soils (Dorioz et al., 2006). Further, we understand that the higher shear forces needed to mobilize coarser soil particles contribute synergistically with the reduction in runoff transport capacity.

In addition to that, the results of the model 1 for the variable soil group agree overall with those of ANOVA, revealing a slight advantage of coarse over medium-textured soils and showing a 22% improvement in comparison to fine-textured soils. The equation seems to better predict results for coarse and fine soils than for medium-textured soils. We presume, the broader range of soil textural classes included under the group "medium" might partly explain this.

In agreement with the vast majority of studies, our analysis further corroborates the positive relationship between buffer performance and buffer width ($r_s = 0,44$, $p = 5,86 \times 10^{-6}$). We found out that for the range of 3 m to 30 m width, the relationship between VBS width and TP retention is more similar to a logarithmic function rather than to a linear function. This result hints that an increase in buffer width beyond a certain point does not necessarily increase buffer performance. Hence, we deduct there is a maximum theoretical width above which buffer performance is not increased.

With regards to width ratio, our findings reveal a moderate oppositional association to TP retention ($r_s = -0,38$, $p = 0,002$) suggesting that small width ratios, i.e. wider VBS in proportion to source field width, give better retention performances. Assuming that crop fields have a geometry close enough to be considered regular, it can be accepted that its width is closely related to its surface area. Because the quantity of phosphorus available for mobilization at source and the retention capacity at buffer depends on the surface area, we conclude the variable width ratio captures the relationship between the source area and the buffer area. White and Arnold (2009) argued that at field scale, area ratio and VFS width may not be interchangeable because fields may not be rectangular. While we understand this is true, we believe the geometry of most fields is regular enough to consider width ratio is closely related to area ratio.

A further look into the width ratio results reveal two underlying patterns in data. After adjusting for a number of suspect cofounder variables, these patterns might be partly explained by slope class. If we consider the slope class "gentle" the degree of association between width ratio and TP retention increases to $r_s = -0,57$ ($p = 5,5 \times 10^{-5}$), whereas for the slope class "moderate" the relationship drops to $r_s = -0,02$ ($p = 0,92$). We speculate the erosion processes associated with each slope type might be behind this difference, however, further research is needed to understand the nature of these patterns.

We performed a linear regression analysis in order to compare the performance of the variables width ratio and VBS width in predicting TP retention (model 2) under the same set of observations (N=27). We found out that the variable width ratio explained a greater amount of TP retention variance ($r_s = 0,31$, $p = 0,015$) than the log-trans VBS width ($r_s = 0,27$, $p = 0,0031$). Hence, we argue

Discussion of results

that width ratio might be a better predictor of buffer performance than VBS width. Even though these findings appear to be coherent, they should be interpreted with precaution due to the small number of records used to adjust the model equations.

We found a weak inverse association between VBS slope and TP retention (r_s = -0,25, p=0,0125). Since both mobilization and transport capacity of runoff are influenced by slope, we expected a stronger link between the two variables. This view is backed up by previous literature results. Dosskey et al. (2008) observed that a 20-m wide strip having a 2% slope would trap 85% of incoming sediment, but only 20% of incoming sediment if the slope was 10%.

A revision of the analysed data reveals a gap for the slope range between 7% to 11%, dividing records in two groups. The first and by far largest group is constituted by the observations with a slope of less or equal than 7% and sums up 71 out of the 98 observations (circa 72% of the total). The second group sums up the remaining 27 observations (close to 28% of the total) and is constituted by slopes higher or equal than 11% up to 20%. We hypothesize that research interests were focused on buffers under sheet and inter-rill conditions which take place at comparably less inclined slopes rather than focusing on flow-concentration processes like rill and gully erosion associated to steeper slopes. However, the real reason for this data gap is not known and may well be explained by other factors. A more accurate and powerful relationship between VBS slope and TP retention can be obtained if the regression analysis is performed for each of the slope groups separately.

With regards to vegetation type, we did not find statistically significant differences ($p \geq 0,05$) in outcome. Our findings seem to agree the broader opinion that no clear advantage can be attributed to grassed or wooded buffers (Daniels and Gilliam, 1996; Mayer et al., 2007; Yuan et al., 2009) and contradicts studies claiming an advantage in favour of grassed buffers (Wenger, 1999). Even though these findings appear to be coherent, they should be interpreted with caution due to the low number of observations on wooded buffers (N=13).

Model 1 was built to analyse the role of chosen variables at predicting TP retention capacity of VBS. The multiple regression analysis identified the variables VBS width, soil group and VBS slope as the most powerful at predicting TP outcome. Further, we identified that the time frame of consideration i.e. single event vs. annual observations plays a role in performance and should be considered when building a model.

Model 3 was built to assess the predictive power of chosen explanatory variables at predicting VBS under field performance and to propose a predictive model. The chosen model was fitted with the variables VBS width and VBS slope and explained a 59% of the sample variance. The results suggest that model 3 better predicts TP retention under field conditions than model 1. However, due to the low number of observations used to fit model 3 it may suffer from overfitting issues and results should be taken with caution. Further experimental studies under field conditions are needed in order to establish and validate a more precise model.

7   Conclusion and outlook

This present study investigates the effect of key parameters on phosphorus retention efficiency of vegetated buffer strips based on experimental data and compares the performance between experimental plots and field conditions.

The outcome presented in this work highlights that field performance of vegetated buffers is less than what we expect on experimental plots. Based on the current knowledge of the mechanisms governing buffer efficiency, we attribute this difference to two main factors:

Spatial heterogeneities under field conditions, meaning flow concentration processes like rills and gullies impairing buffer function (White and Arnold, 2009) and man-made linear structures acting as terrain barriers rerouting flow (Hösl et al., 2012), both of which have been demonstrated to affect buffer efficiency under field conditions but are seldom promoted in experimental plots.

Secondly, the temporal scale of field plot studies, as most of the experiments on field plots are conducted as single-events. Thus, they fairly describe the initial physical processes but omit the effect of buffer saturation over time and microbiological and geochemical soil dynamics which have an important role on phosphorus soil cycling but are merely observed during the course of days to years. Further, most of these cited processes are not steady over time but follow seasonal patterns (periodization) depending on the environmental conditions. An example of this is the release of phosphorus in dissolved forms after decomposition of the litter (Dorioz et al., 2006) occurring in autumn.

The prediction of buffer performance based solely on buffer width is an extended practice. However, we found that buffer performance is better described by a logarithmic law. This implies the existence of a theoretical maximum optimal trapping efficiency above which buffer performance stagnates. Increasing efficiency beyond that point would require disproportionately wide buffers. Based on our observations we speculate that this maximum optimal efficiency is reached at around buffer widths between 25 m to 30 m, but may be higher depending on other influencing factors.

Further, we understand that in order to optimally design buffers, a good prediction tool must include a relation between potential particle mobilization at source and trapping capacity of buffer. Basing on the flow path notion, we introduced the concept of width ratio, a quotient between buffer width and source field width, which we found out to be a better predictor of buffer performance than buffer width. We understand that width ratio expresses a relation between potential mobilization at source, through the longest flow path and potential deposition sites at buffer.

So, coming back our initial question: What can we really expect from vegetated buffer strips as a long-term measure to control phosphorus emissions to surface waters under field conditions?

Conclusion and outlook

We believe the expectations on buffer strips are set too high. Its use is backed mostly on short-term studies on experimental plots which have been demonstrated to not fully correspond to long-term field conditions. We are confident that vegetated buffer strips can be an effective tool to control phosphorus from agricultural sources if considered as a part of an integral watershed-wide conservation program including other best management practice, as it has already been demonstrated by Meals and Hopkins (2002), but it should not be considered as a stand-alone tool.

If a buffer strip is well maintained, it is sensible to expect long-term efficiencies of 50% to 60% TP retention on 25 m wide buffers for up to 5% slope gradient under average soil conditions for periods of at least one to ten years, but in order to preserve retention capacity over time buffers should be maintained. Measures like harvesting vegetation at the end of the season to remove phosphorus from the system and allow for more phosphorus uptake and inspection of buffer condition after intense precipitation events to control the appearance of concentrated forms of erosion should be part of a buffer maintenance plan.

There is no single universal solution. It is important that buffers are designed with local sites conditions in mind, taking into account climate, soil and field practices but also requirements of environmental management plans and possible restrictions. At the end of the day, it is about achieving a feasible compromise between societal needs, in the form of agricultural land and crop growth and environmental conservation.

8 References

Abu-Zreig, M., Rudra, R.P., Lalonde, M.N., Whiteley, H.R., Kaushik, N.K., 2004. Experimental investigation of runoff reduction and sediment removal by vegetated filter strips. Hydrol. Process. 18, 2029–2037. doi:10.1002/hyp.1400

Altman, D.G., 2014. Categorizing Continuous Variables, in: Wiley StatsRef: Statistics Reference Online. John Wiley & Sons, Ltd. doi:10.1002/9781118445112.stat04857

Blanco-Canqui, H., Gantzer, C.J., Anderson, S.H., 2006. Performance of Grass Barriers and Filter Strips under Interrill and Concentrated Flow. J. Environ. Qual. 35, 1969. doi:10.2134/jeq2006.0073

BMLFUW, 2001. Wasserwelt March-Thaya-Auen [WWW Document]. URL https://www.bmlfuw.gv.at/umwelt/natur-artenschutz/life-natur/life-projekte_abgeschl/marchthaya2.html (accessed 10.8.17).

Borin, M., Vianello, M., Morari, F., Zanin, G., 2005. Effectiveness of buffer strips in removing pollutants in runoff from a cultivated field in North-East Italy. Agric. Ecosyst. Environ. 105, 101–114. doi:10.1016/j.agee.2004.05.011

Burns, R.B., Burns, R.A., 2008. Business research methods and statistics using SPSS. SAGE, Los Angeles ; London.

C. Ditzler, K. Scheffe, and H.C. Monger, 2017. Soil survey manual.

Cole, J.T., Baird, J.H., Basta, N.T., Huhnke, R.L., Storm, D.E., Johnson, G.V., Payton, M.E., Smolen, M.D., Martin, D.L., Cole, J.C., 1997. Influence of Buffers on Pesticide and Nutrient Runoff from Bermudagrass Turf. J. Environ. Qual. 26, 1589. doi:10.2134/jeq1997.00472425002600060019x

Cook, R.D., 1977. Detection of Influential Observation in Linear Regression. Technometrics 19, 15. doi:10.2307/1268249

Cooper, J.R., Gilliam, J.W., 1987. Phosphorus Redistribution from Cultivated Fields into Riparian Areas1. Soil Sci. Soc. Am. J. 51, 1600. doi:10.2136/sssaj1987.03615995005100060035x

Daniels, R.B., Gilliam, J.W., 1996. Sediment and Chemical Load Reduction by Grass and Riparian Filters. Soil Sci. Soc. Am. J. 60, 246. doi:10.2136/sssaj1996.03615995006000010037x

De Winter, P., Cahusac, P., 2014. Starting out in statistics: an introduction for students of human health, disease and psychology. Wiley Blackwell, Chichester, West Sussex, UK ; Hoboken, NJ, USA.

Dillaha, T.A., Sherrard, J.H., Lee, D., Mostaghimi, S., Shanholtz, V.O., 1988. Evaluation of vegetative filter strips as a best management practice for feed lots. J. Water Pollut. Control Fed. 1231–1238.

Dorioz, J.M., Wang, D., Poulenard, J., Trévisan, D., 2006. The effect of grass buffer strips on phosphorus dynamics—A critical review and synthesis as a basis for application in agricultural landscapes in France. Agric. Ecosyst. Environ. 117, 4–21. doi:10.1016/j.agee.2006.03.029

Dosskey, M.G., Helmers, M., Eisenhauer, D., Franti, T., Hoagland, K., 2002. Assessment of Concentrated Flow Through Riparian Buffers. J. Soil Water Conserv. 57.

Dosskey, M.G., Helmers, M.J., Eisenhauer, D.E., 2008. A design aid for determining width of filter strips. J. Soil Water Conserv. 63, 232–241. doi:10.2489/jswc.63.4.232

Dosskey, M.G., Vidon, P., Gurwick, N.P., Allan, C.J., Duval, T.P., Lowrance, R., 2010. The Role

References

of Riparian Vegetation in Protecting and Improving Chemical Water Quality in Streams. JAWRA J. Am. Water Resour. Assoc. 46, 261–277. doi:10.1111/j.1752-1688.2010.00419.x

Draper, N.R., Smith, H., 1998. Applied regression analysis, 3rd ed. ed, Wiley series in probability and statistics. Wiley, New York.

FAO, 2006. Guidelines for soil description, 4th ed. ed. Food and Agriculture Organization of the United Nations, Rome.

Fox, J., Monette, G., 1992. Generalized Collinearity Diagnostics. J. Am. Stat. Assoc. 87, 178–183. doi:10.1080/01621459.1992.10475190

Grismer, M., TOBY, A., LEWIS, D., 2006. Vegetative Filter Strips for Nonpoint Source Pollution Control in Agriculture. ANR Publ 8195.

Hampel, F., 2002. Robust inference. Encycl. Environmetrics.

Hauke, J., Kossowski, T., 2011. Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data. Quaest. Geogr. 30. doi:10.2478/v10117-011-0021-1

Hawes, E., Smith, M., 2005. Riparian buffer zones: Functions and recommended widths. Eightmile River Wild Scen. Study Comm. 15, 2005.

Haycock, N. (Ed.), 1997. Buffer zones: their processes and potential in water protection ; the proceedings of the International Conference on Buffer Zones, September 1996. Quest Environmental, Harpenden.

Helmers, M.J., Eisenhauer, D.E., Dosskey, M.G., Franti, T.G., Brothers, J.M., McCullough, M.C., 2005. Flow pathways and sediment trapping in a field-scale vegetative filter.

Helmers, M.J., others, 2008. 4. Buffers and Vegetative Filter Strips.

Hösl, R., 2009. Analyse von linearen Abflusswegen und ihr Einfluss auf die Effektivität von Gewässerrandstreifen. Universität Wien.

Hösl, R., Strauss, P., Glade, T., 2012. Man-made linear flow paths at catchment scale: Identification, factors and consequences for the efficiency of vegetated filter strips. Landsc. Urban Plan. 104, 245–252. doi:10.1016/j.landurbplan.2011.10.017

Huber, M., 2015. Erfassung, Dimensionierung und Gestaltung von Gewässerrandstreifen im Mondseeeinzugsgebiet (Master Thesis). PARIS LODRON UNIVERSITÄT SALZBURG, Salzburg.

I. Chaubey, D. R. Edwards, T. C. Daniel, P. A. Moore Jr., D. J. Nichols, 1995. Effectiveness of Vegetative Filter Strips in Controlling Losses of Surface-applied Poultry Litter Constituents. Trans. ASAE 38, 1687–1692. doi:10.13031/2013.27995

Kao, L.S., Green, C.E., 2008. Analysis of Variance: Is There a Difference in Means and What Does It Mean? J. Surg. Res. 144, 158–170. doi:10.1016/j.jss.2007.02.053

Kovács, Á., 2013. Quantification of diffuse phosphorous inputs into surface water systems. (Doctoral Thesis). Vienna University of Technology.

Lim, T.T., Edwards, D.R., Workman, S.R., Larson, B.T., Dunn, L., 1998. Vegetated filter strip removal of cattle manure constituents in runoff. Trans. ASAE 41, 1375.

Magette, W.L., Brinsfield, R.B., Palmer, R.E., Wood, J.D., 1989. Nutrient and sediment removal by vegetated filter strips. Trans. ASAE 32, 663–0667.

Mander, Ü., Kuusemets, V., Lõhmus, K., Mauring, T., 1997. Efficiency and dimensioning of riparian buffer zones in agricultural catchments. Ecol. Eng. 8, 299–324.

References

Mayer, P.M., Reynolds, S.K., McCutchen, M.D., Canfield, T.J., 2007. Meta-Analysis of Nitrogen Removal in Riparian Buffers. J. Environ. Qual. 36, 1172. doi:10.2134/jeq2006.0462

Meals, D.W., Hopkins, R.B., 2002. Phosphorus reductions following riparian restoration in two agricultural watersheds in Vermont, USA. Water Sci. Technol. 45, 51–60.

Moreddu, C., 2011. Evaluation of agricultural policy reforms in the European Union. OECD, Paris.

Moutinho, L., Hutcheson, G.D. (Eds.), 2011. The Sage dictionary of quantitative management research. SAGE, Los Angeles, Calif. ; London.

Muñoz-Carpena, R., Parsons, J.E., Gilliam, J.W., 1999. Modeling hydrology and sediment transport in vegetative filter strips. J. Hydrol. 214, 111–129. doi:10.1016/S0022-1694(98)00272-8

Natural Resources Conservation Service, 2008. Conservation Practice Standards. Filter strips. [WWW Document]. Conserv. Pract. Stand. URL https://www.nrcs.usda.gov/wps/portal/nrcs/main/national/technical/cp/ncps/ (accessed 10.6.17).

Osborne, J.W., Waters, E., 2002. Multiple Regression Assumptions. ERIC Digest.

P. Vatcheva, K., Lee, M., 2016. Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies. Epidemiol. Open Access 06. doi:10.4172/2161-1165.1000227

Patty, L., Réal, B., Joël Gril, J., 1997. The Use of Grassed Buffer Strips to Remove Pesticides, Nitrate and Soluble Phosphorus Compounds from Runoff Water. Pestic. Sci. 49, 243–251. doi:10.1002/(SICI)1096-9063(199703)49:3<243::AID-PS510>3.0.CO;2-8

Pearce, R.A., Trlica, M.J., Leininger, W.C., Smith, J.L., Frasier, G.W., 1997. Efficiency of Grass Buffer Strips and Vegetation Height on Sediment Filtration in Laboratory Rainfall Simulations. J. Environ. Qual. 26, 139. doi:10.2134/jeq1997.00472425002600010021x

Reimann, C., et al. (Eds.), 2008. Statistical data analysis explained: applied environmental statistics with R. John Wiley & Sons, Chichester, England ; Hoboken, NJ.

Roberts, W.M., Matthews, R.A., Blackwell, M.S.A., Peukert, S., Collins, A.L., Stutter, M.I., Haygarth, P.M., 2013. Microbial biomass phosphorus contributions to phosphorus solubility in riparian vegetated buffer strip soils. Biol. Fertil. Soils 49, 1237–1241. doi:10.1007/s00374-013-0802-x

Roberts, W.M., Stutter, M.I., Haygarth, P.M., 2012. Phosphorus Retention and Remobilization in Vegetated Buffer Strips: A Review. J. Environ. Qual. 41, 389. doi:10.2134/jeq2010.0543

Rogerson, P., 2001. Statistical methods for geography. SAGE Publications, London ; Thousand Oaks, Calif.

Schmidheiny, K., 2013. The Multiple Linear Regression Model. Citeseer.

Schmitt, T.J., Dosskey, M.G., Hoagland, K.D., 1999. Filter Strip Performance and Processes for Different Vegetation, Widths, and Contaminants. J. Environ. Qual. 28, 1479. doi:10.2134/jeq1999.00472425002800050013x

Sheppard, S.C., Sheppard, M.I., Long, J., Sanipelli, B., Tait, J., 2006. Runoff phosphorus retention in vegetated field margins on flat landscapes. Can. J. Soil Sci. 86, 871–884.

Simon, L., Young, D., 2017. Lesson 12: Multicollinearity & Other Regression Pitfalls. STAT 501. [WWW Document]. URL https://onlinecourses.science.psu.edu/stat501/node/343 (accessed 7.23.17).

References

Soil Science Division Staff, 2017. Soil_Survey_Manual.pdf.

Stevens, J.P., 1984. Outliers and influential data points in regression analysis. Psychol. Bull. 95, 334–344. doi:10.1037/0033-2909.95.2.334

Stutter, M.I., Langan, S.J., Lumsdon, D.G., 2009. Vegetated Buffer Strips Can Lead to Increased Release of Phosphorus to Waters: A Biogeochemical Assessment of the Mechanisms. Environ. Sci. Technol. 43, 1858–1863. doi:10.1021/es8030193

Syversen, N., 2005. Effect and design of buffer zones in the Nordic climate: The influence of width, amount of surface runoff, seasonal variation and vegetation type on retention efficiency for nutrient and particle runoff. Ecol. Eng. 24, 483–490. doi:10.1016/j.ecoleng.2005.01.016

T.A. Dillaha, R.B. Reneau, S. Mostaghimi, D. Lee, 1989. Vegetative Filter Strips for Agricultural Nonpoint Source Pollution Control. Trans. ASAE 32, 0513–0519. doi:10.13031/2013.31033

Templ, M., 2017. Statistical Disclosure Control for Microdata: Methods and Applications in R. Springer.

Uusi-Kämppä, J., 2008. Evaluating vegetated buffer zones for phosphorus retention in cereal and grass production. NJF Rep. 4, 68–73.

Uusi-Kämppä, J., 2005. Phosphorus purification in buffer zones in cold climates. Ecol. Eng. 24, 491–502. doi:10.1016/j.ecoleng.2005.01.013

Verstraeten, G., Poesen, J., Gillijns, K., Govers, G., 2006. The use of riparian vegetated filter strips to reduce river sediment loads: an overestimated control measure? Hydrol. Process. 20, 4259–4267. doi:10.1002/hyp.6155

Wenger, S., 1999. A review of the scientific literature on riparian buffer width, extent and vegetation.

White, M.J., Arnold, J.G., 2009. Development of a simplistic vegetative filter strip model for sediment and nutrient retention at the field scale. Hydrol. Process. 23, 1602–1616. doi:10.1002/hyp.7291

wpa Beratende Ingenieure, 2009. Effektivität von Gewässerrandstreifen zum Schutz von Oberflächengewässern.

Yang, F., Yang, Y., Li, H., Cao, M., 2015. Removal efficiencies of vegetation-specific filter strips on nonpoint source pollutants. Ecol. Eng. 82, 145–158. doi:10.1016/j.ecoleng.2015.04.018

Yeager, K., 2017. LibGuides: SPSS Tutorials: Independent Samples t Test [WWW Document]. URL http://libguides.library.kent.edu/SPSS/IndependentTTest

Yoo, W., Mayberry, R., Bae, S., Singh, K., (Peter) He, Q., Lillard, J.W., 2014. A Study of Effects of MultiCollinearity in the Multivariable Analysis. Int. J. Appl. Sci. Technol. 4, 9–19.

Yuan, Y., Bingner, R.L., Locke, M.A., 2009. A Review of effectiveness of vegetative buffers on sediment trapping in agricultural areas. Ecohydrology 2, 321–336. doi:10.1002/eco.82

Zessner, M., Hepp, G., 2014. Überlegungen zur Wirksamkeit der ÖPUL 2014 Massnahme „Vorbeugender Oberflächengewässerschutz auf Ackerflächen "zur Verringerung von Phosphoremissionen im Kontext von Einzugsgebieten. Inst. Für Wassergüte Ressourcenmanagement Abfallwirtsch. Tech. Univ. Wien Wien.

Zessner, M., Hepp, G., Zoboli, O., Molló Manonelles, O., Langer, M., Lechner, A., 2016. Erstellung und Evaluierung eines Prognosetools zur Quantifizierung von Maßnahmenwirksamkeiten im Bereich der Nährstoffeinträge in oberösterreichische Oberflächengewässer. Inst. Für Wassergüte Ressourcenmanagement Abfallwirtsch. Tech. Univ.

References

Wien Wien.

Zhang, X., Liu, X., Zhang, M., Dahlgren, R.A., Eitzel, M., 2010. A Review of Vegetated Buffers and a Meta-analysis of Their Mitigation Efficacy in Reducing Nonpoint Source Pollution. J. Environ. Qual. 39, 76. doi:10.2134/jeq2008.0496

9 Appendix

Appendix 1: Database

author	year	QA	soil.texture	soil.group	time.frame	extension	vegetation	field.length	VBS.width	width.ratio	TP.retention	VBS.slope	slope.class
Abu-Zreig et al.	2003	Y	silt loam	medium	single events	experimental field plot	grass	NA	2	NA	31	2,3	gentle
Abu-Zreig et al.	2003	Y	silt loam	medium	single events	experimental field plot	grass	NA	2	NA	62	2,3	gentle
Abu-Zreig et al.	2003	Y	silt loam	medium	single events	experimental field plot	grass	NA	2	NA	33	2,3	gentle
Abu-Zreig et al.	2003	Y	silt loam	medium	single events	experimental field plot	grass	NA	5	NA	47	2,3	gentle
Abu-Zreig et al.	2003	Y	silt loam	medium	single events	experimental field plot	grass	NA	5	NA	54	2,3	gentle
Abu-Zreig et al.	2003	Y	silt loam	medium	single events	experimental field plot	grass	NA	5	NA	60	2,3	gentle
Abu-Zreig et al.	2003	Y	silt loam	medium	single events	experimental field plot	grass	NA	5	NA	53	2,3	gentle
Abu-Zreig et al.	2003	Y	silt loam	medium	single events	experimental field plot	grass	NA	10	NA	93	2,3	gentle
Abu-Zreig et al.	2003	Y	silt loam	medium	single events	experimental field plot	grass	NA	10	NA	79	2,3	gentle
Abu-Zreig et al.	2003	Y	silt loam	medium	single events	experimental field plot	grass	NA	10	NA	55	2,3	gentle
Abu-Zreig et al.	2003	Y	silt loam	medium	single events	experimental field plot	grass	NA	15	NA	77	2,3	gentle
Abu-Zreig et al.	2003	Y	silt loam	medium	single events	experimental field plot	grass	NA	15	NA	89	2,3	gentle
Abu-Zreig et al.	2003	Y	silt loam	medium	single events	experimental field plot	grass	NA	15	NA	72	2,3	gentle
Abu-Zreig et al.	2003	Y	silt loam	medium	single events	experimental field plot	grass	NA	5	NA	68	5	gentle
Abu-Zreig et al.	2003	Y	silt loam	medium	single events	experimental field plot	grass	NA	5	NA	72	5	gentle
Abu-Zreig et al.	2003	Y	silt loam	medium	single events	experimental field plot	grass	NA	5	NA	65	5	gentle
Abu-Zreig et al.	2003	Y	silt loam	medium	single events	experimental field plot	grass	NA	5	NA	65	2,3	gentle
Blanco-Canqui et al.	2006	Y	silt loam	medium	single events	experimental field plot	grass	8	8	1	94,1	5	gentle
Blanco-Canqui et al.	2006	Y	silt loam	medium	single events	experimental field plot	grass	8	8	1	91,5	5	gentle
Blanco-Canqui et al.	2006	Y	silt loam	medium	single events	experimental field plot	grass	8	8	1	97,2	5	gentle
Blanco-Canqui et al.	2006	Y	silt loam	medium	single events	experimental field plot	grass	8	8	1	94	5	gentle
Borin et al.	2005	Y	loam	medium	annual	cultivated land	wood	35	6	5,83	80	3	gentle
Chaubey et al.	1994	Y	silt loam	medium	single events	experimental field plot	grass	3,1	3,1	1	39,6	3	gentle
Chaubey et al.	1994	Y	silt loam	medium	single events	experimental field plot	grass	3,1	6,1	0,51	58,4	3	gentle
Chaubey et al.	1994	Y	silt loam	medium	single events	experimental field plot	grass	3,1	9,2	0,34	74	3	gentle
Chaubey et al.	1994	Y	silt loam	medium	single events	experimental field plot	grass	3,1	15,2	0,2	86,8	3	gentle

Appendix

Chaubey et al.	1994	Y	silt loam	medium	single events	experimental field plot	grass	3,1	21,4	0,14	91,2	3	gentle
Daniels and Gilliam	1996	Y	sandy loam	coarse	annual	cultivated land	grass	84	3	28	55	4,9	gentle
Daniels and Gilliam	1996	Y	sandy loam	coarse	annual	cultivated land	grass	84	6	14	70	4,9	gentle
Daniels and Gilliam	1996	Y	sandy loam	coarse	annual	cultivated land	grass	86	3	28,67	40	2,1	gentle
Daniels and Gilliam	1996	Y	sandy loam	coarse	annual	cultivated land	grass	86	6	14,33	65	2,1	gentle
Daniels and Gilliam	1996	Y	sandy loam	coarse	annual	cultivated land	grass	48	5	9,6	45	3,3	gentle
Daniels and Gilliam	1996	Y	sandy loam	coarse	annual	cultivated land	grass	48	13	3,69	55	3,3	gentle
Daniels and Gilliam	1996	Y	sandy loam	coarse	annual	cultivated land	wood	53	7	7,57	30	4,1	gentle
Daniels and Gilliam	1996	Y	sandy loam	coarse	annual	cultivated land	wood	53	18	2,94	65	4,1	gentle
Dillaha et al.	1988	Y	silt loam	medium	single events	experimental field plot	grass	18,3	4,6	3,98	63	11	moderate
Dillaha et al.	1988	Y	silt loam	medium	single events	experimental field plot	grass	18,3	4,6	3,98	52	16	moderate
Dillaha et al.	1988	Y	silt loam	medium	single events	experimental field plot	grass	18,3	9,1	2,01	80	11	moderate
Dillaha et al.	1988	Y	silt loam	medium	single events	experimental field plot	grass	18,3	9,1	2,01	57	16	moderate
Dillaha et al.	1989	Y	silt loam	medium	seasonal	experimental field plot	grass	18,3	4,6	3,98	61	13,5	moderate
Dillaha et al.	1989	Y	silt loam	medium	seasonal	experimental field plot	grass	18,3	9,1	2,01	79	13,5	moderate
Duchemin	2002	Y	sandy loam	coarse	annual	cultivated land	grass	65	3	21,67	86	2,5	gentle
Duchemin	2002	Y	sandy loam	coarse	annual	cultivated land	grass	65	6	10,83	84,6	2,5	gentle
Duchemin	2002	Y	sandy loam	coarse	annual	cultivated land	grass	65	9	7,22	86,9	2,5	gentle
Lee et al.	1999	Y	silty clay loam	fine	single events	experimental field plot	grass	NA	3	40	39,5	3	gentle
Lee et al.	1999	Y	silty clay loam	fine	single events	experimental field plot	grass	NA	3	40	35,2	3	gentle
Lee et al.	1999	Y	silty clay loam	fine	single events	experimental field plot	grass	NA	6	20	55,2	3	gentle
Lee et al.	1999	Y	silty clay loam	fine	single events	experimental field plot	grass	NA	6	20	49,4	3	gentle
Lee et al.	2003	Y	silty clay loam	fine	annual	cultivated land	grass	22,1	7,1	3,11	78	5	gentle
Lee et al.	2003	Y	silty clay loam	fine	annual	cultivated land	wood	22,1	16,3	1,36	91,3	5	gentle
Lim et al.	1998	Y	silt loam	medium	seasonal	experimental field plot	grass	12,2	6,1	2	76,1	3	gentle
Lim et al.	1998	Y	silt loam	medium	seasonal	experimental field plot	grass	12,2	12,2	1	90,1	3	gentle
Lim et al.	1998	Y	silt loam	medium	seasonal	experimental field plot	grass	12,2	18,3	0,67	93,6	3	gentle
Magette et al.	1987	Y	silt loam	medium	seasonal	experimental field plot	grass	22	4,6	4,78	41	4	gentle
Magette et al.	1987	Y	silt loam	medium	seasonal	experimental field plot	grass	22	9,1	2,42	53	4	gentle
Magette et al.	1989	Y	silt loam	medium	seasonal	experimental field plot	grass	22	4,6	4,78	27	4	gentle
Peterjohn and Correll	1984	Y	sandy loam	coarse	annual	cultivated land	wood	NA	19	NA	74	5,4	gentle

Appendix

Schmitt et al.	1999	Y	clay loam	fine	single events	cultivated land	grass	*NA*	7,5	*NA*		70	6,5	gentle
Schmitt et al.	1999	Y	clay loam	fine	single events	cultivated land	grass	*NA*	15	*NA*		80	6,5	gentle
Schmitt et al.	1999	Y	clay loam	fine	single events	cultivated land	grass	*NA*	7,5	*NA*		55	6,5	gentle
Schmitt et al.	1999	Y	clay loam	fine	single events	cultivated land	grass	*NA*	15	*NA*		72	6,5	gentle
Schmitt et al.	1999	Y	clay loam	fine	single events	cultivated land	grass	*NA*	7,5	*NA*		55	6,5	gentle
Schmitt et al.	1999	Y	clay loam	fine	single events	cultivated land	grass	*NA*	15	*NA*		72	6,5	gentle
Schmitt et al.	1999	Y	clay loam	fine	single events	cultivated land	wood	*NA*	7,5	*NA*		45	6,5	gentle
Schmitt et al.	1999	Y	clay loam	fine	single events	cultivated land	wood	*NA*	15	*NA*		50	6,5	gentle
Syversen	2005	Y	silt loam	medium	single events	experimental field plot	grass	45	5	9	65	17	moderate
Syversen	2005	Y	silt loam	medium	single events	experimental field plot	grass	45	5	9	10	17	moderate
Syversen	2005	Y	silt loam	medium	single events	experimental field plot	grass	45	10	4,5	66	17	moderate
Syversen	2005	Y	silt loam	medium	single events	experimental field plot	grass	45	10	4,5	60	17	moderate
Syversen	2005	Y	silt loam	medium	single events	experimental field plot	grass	45	5	9	50	14	moderate
Syversen	2005	Y	silt loam	medium	single events	experimental field plot	grass	45	5	9	70	14	moderate
Syversen	2005	Y	silt loam	medium	single events	experimental field plot	grass	45	10	4,5	80	14	moderate
Syversen	2005	Y	silt loam	medium	single events	experimental field plot	grass	45	10	4,5	90	14	moderate
Syversen	2005	Y	silt loam	medium	single events	experimental field plot	grass	45	5	9	65	12	moderate
Syversen	2005	Y	silt loam	medium	single events	experimental field plot	grass	45	5	9	80	12	moderate
Syversen	2005	Y	silt loam	medium	annual	experimental field plot	grass	45	5	9	77	12	moderate
Syversen	2005	Y	silt loam	medium	annual	experimental field plot	grass	45	10	4,5	83	12	moderate
Syversen	2005	Y	silt loam	medium	single events	experimental field plot	grass	45	5	9	80	20	moderate
Syversen	2005	Y	silt loam	medium	single events	experimental field plot	wood	45	5	9	75	20	moderate
Syversen	2005	Y	silt loam	medium	single events	experimental field plot	wood	45	5	9	97	20	moderate
Uusi-Kamppa	2008	Y	clay	fine	annual	cultivated land	grass	60	10	6	38,4	15	moderate
Uusi-Kamppa	2008	Y	clay	fine	annual	cultivated land	grass	60	10	6	38,4	15	moderate
Uusi-Kamppa	2008	Y	clay	fine	annual	cultivated land	grass	60	10	6	11,1	15	moderate
Uusi-Kamppa	2008	Y	clay	fine	annual	cultivated land	grass	60	10	6	22,2	15	moderate
Uusi-Kamppa	2008	Y	clay	fine	annual	cultivated land	grass	60	10	6	14,3	15	moderate
Uusi-Kamppa	2008	Y	clay	fine	annual	cultivated land	grass	60	10	6	28,6	15	moderate
Yang et al.	2015	Y	sandy loam	coarse	single events	experimental field plot	wood	*NA*	10	*NA*		84,44	2	gentle
Yang et al.	2015	Y	sandy loam	coarse	single events	experimental field plot	wood	*NA*	10	*NA*		87,87	2	gentle

Appendix

Yang et al.	2015	Y	sandy loam	coarse	single events	experimental field plot	wood	*NA*	10	*NA*	85,57	2	gentle
Yang et al.	2015	Y	sandy loam	coarse	single events	experimental field plot	wood	*NA*	10	*NA*	73,28	2	gentle
Yang et al.	2015	Y	sandy loam	coarse	single events	experimental field plot	grass	*NA*	10	*NA*	84,84	2	gentle
Yang et al.	2015	Y	sandy loam	coarse	single events	experimental field plot	grass	*NA*	15	*NA*	84,93	2	gentle
Yang et al.	2015	Y	sandy loam	coarse	single events	experimental field plot	grass	*NA*	10	*NA*	85,46	2	gentle
Yang et al.	2015	Y	sandy loam	coarse	single events	experimental field plot	grass	*NA*	15	*NA*	85,65	2	gentle
Yang et al.	2015	Y	sandy loam	coarse	single events	experimental field plot	grass	*NA*	10	*NA*	84,75	2	gentle
Yang et al.	2015	Y	sandy loam	coarse	single events	experimental field plot	grass	*NA*	15	*NA*	85,68	2	gentle
Yang et al.	2015	Y	sandy loam	coarse	single events	experimental field plot	grass	*NA*	10	*NA*	89,73	2	gentle
Yang et al.	2015	Y	sandy loam	coarse	single events	experimental field plot	grass	*NA*	15	*NA*	91,26	2	gentle

10 Affirmation

I certify, that the master thesis was written by me, not using sources and tools other than quoted and without use of any other illegitimate support. Furthermore, I confirm that I have not submitted this master thesis either nationally or internationally in any form.

Oriol Molló Manonelles   Vienna, 9 October 2017

signature