

Soil Characteristics Changes

Using spectral and photographic data to assess the promise of PLS regression and object-oriented image segmentation to track the impacts of erosion



M.Sc. Minor Thesis by William Cook

August 2012

Land Degradation and Development Group



WAGENINGEN UNIVERSITY
WAGENINGEN UR

Soil Characteristics Changes

Using spectral and photographic data to assess the promise of PLS regression and object-oriented image segmentation to track the impacts of erosion

Master Minor Thesis Land Degradation and Development Group submitted in partial fulfilment of the degree of Master of Science in International Land and Water Management at Wageningen University, the Netherlands

Study Program:

MSc. International Land and Water Management (MIL)

Student Registration Number:

880605161020

LDD 80318

Supervisors:

Dr. Saskia Keesstra

Dr. Harm Bartholomeus

Examiner:

Prof. Dr. Coen Ritsema

Date: August 2012

Wageningen University, Land Degradation and Development Group

Acknowledgements

I would like to take this opportunity to first thank Saskia Keesstra and Harm Bartholomeus for being so patient, and for guiding me through the topics I had no familiarity with. I would also like to thank Martine Hagen and the Amsterdam Free University Sediment Analysis Laboratory for the guidance and use of their sediment distribution analysis equipment. Finally, I would like to thank Tamás Lassu and Amaia Marruedo Arricibita for their help and insight in the experimental process.

Contents

List of Figures	ii
List of Annexes	ii
Introduction and Purpose	1
Research Questions	2
Research Principles	3
Erosion Processes.....	3
Hyperspectral Analysis	5
Partial Least Squares (PLS) Regression	7
Rill Formation Identification	7
Key Take-Away	8
Methodology	11
Lab Setup	11
Hyperspectral Data Collection.....	13
Photograph Collection	15
Soil Sample Collection	15
Soil Texture Analysis.....	16
Texture Analysis	16
Rainfall Intensity Relationship	16
ENVI/SAMS and Continuum Removal.....	17
ParLeS and PLSR	17
Photographic Analysis	18
Results.....	19
Soil Texture and Rainfall Intensity	19
Full Frequency Table Graphs.....	19
Size Category < 8 μm	21
Size Category 8-63 μm	21
Size category 63-2000 μm	21
PLS Regression	25
Rill Identification	26
Conclusion	30
Research Answers	30
Discussion	30
Wider Setting – Rill Identification.....	30
Wider Setting – Soil Texture Analysis	31
This Thesis – Rill Identification	32
This Thesis – Soil Texture Analysis	32
Bibliography	34

List of Figures

Figure 1. The erosion processes along a hill slope (Alberta Department of Transportation n.d.).	4
Figure 2. Example spectral signatures with substance-specific absorption features.	5
Figure 3. Model predictive capacity based on RPD and R^2 (Chang, et al. 2001).	7
Figure 4. Relationship of the research principles.	10
Figure 5. The rainfall simulator, with soil boxes inside.	11
Figure 6. Flowchart of methodological items.	12
Figure 7. Soil box and approximate position of spectral and soil sample collection points.	13
Figure 8. Comparing the laboratory setup of Rong 2009 (upper left), Calabrò, Leone, and Amenta 2010 (upper right), and this thesis (lower).	14
Figure 9. Example photograph. Note the hotspot.	15
Figure 10. An example of the untreated spectral data with region of interest highlighted (top), and continuum removed data region (bottom).	17
Figure 11. Full frequency table graphs for Box 1.	20
Figure 12. Size category $< 8 \mu\text{m}$ by box.	22
Figure 13. Size category $8\text{-}63 \mu\text{m}$ by box.	23
Figure 14. Size category $63\text{-}2000 \mu\text{m}$ by box.	24
Figure 15. Photo and quadtree of dry box. Note the identified objects.	27
Figure 16. Photo and quadtree of wet box, 50% zoom. Note the similar linear feature.	28
Figure 17. Photo and quadtree of wet box, 100% zoom. The features are more easily seen, but image is noisier.	29

List of Annexes

Annex 1: Soil texture by box and texture class.	37
Annex 2: Soil texture by box and date.	41
Annex 3. ParLeS PLS Regression Steps.	43

Introduction and Purpose

It is an accepted fact that rainfall events cause erosion on a slope, depending on the characteristics of both the slope and the rainfall event. Furthermore, the erosion type and the erosion potential (carrying capacity) are different along the slope. Regarding the former, typically, erosion follows the sequence of splash erosion being dominant at the top of the slope, laminar erosion dominating the middle distance of the slope, and rill erosion taking hold towards the toe of the slope as the carrying capacity of the eroding water increases. Each type of erosion represents relatively more soil material being eroded. The erosion potential varies along with this carrying capacity, resulting in preferential erosion along the slope, starting with clay particles eroding first, followed by larger particles. This forms the basis of this thesis, which aims to explore using technology to track the relationship between rainfall events and erosion processes.

This project falls under the purview of the SMART INSPECTORS programme, which aims to develop innovative methods for non-satellite based remote sensing in order to provide “Decision Support” to decision makers on the topics of agriculture and environmental issues. One such issue is “erosion control”, part of which is the process of monitoring and characterizing soil properties such as chemical content, texture, and surface features (Waal Hochschule Rhein 2011). An important aspect of this overarching programme is the inclusion of remote sensing technologies as a main focus. This means that while the monitoring changes in soil structure along a slope directly would be the simplest and most reliable method; this research must be carried out utilising remote sensing as an integral analytical tool. On this topic, we are again dealing with well-established information. Remote sensing technologies, specifically hyperspectral data, has been used successfully in the past to identify and track changes in clay *mineral* content (Ben-Dor and Benin 1995). Using similar data, which is easily obtainable with standard and widely available spectral equipment, we will attempt to track changes in clay *particle* concentrations under the assumption that clay size particles are proportional to clay mineral concentrations in the soils available. Furthermore, in alignment with the SMART INSPECTOR goal of increasing ability to monitor surface soil features, we are interested in monitoring the last type of erosion, rill erosion.

This type of erosion impacts the soil structure just as the preferential erosion of different sized soil particles, but results in far more visible changes. Furthermore it is the starting point for gully erosion, which is far more mechanically challenging on the field level. For this reason, early detection of rill development is in line with the goals of the SMART INSPECTOR programme. The challenge lays in being able to incorporate remote sensing methods in the process of rill formation identification. The existence of object-oriented image analysis software simplifies this greatly. Using a series of computing logarithms, this type of software is capable of identifying coherent objects. Examples in the past include buildings, trees and bushes, and gullies at the field level (Chávez and Clevers 2012, Shruthi, Kerle and

Jetten 2011). Having proven successful for these purposes, we will attempt to turn this software to more micro-scale analysis of soil surface changes.

With these topics as the guiding themes of this research, the overarching reasons for this study can be discussed in an intelligible manner. On a practical level, the purpose of this research is to explore the possibility of using commonly available technologies to monitor and identify changes in soil surface characteristics. This is primarily focused on utilising already established software packages (discussed later) in order to obtain this goal. Such a method, which does not require purpose-specific techniques, nor does it require re-inventing the wheel in terms of analytical algorithms, is in line with the umbrella programme discussed previously. The purpose is to be accessible, practical, and readily translated into a standard procedure. On a more academic level, this research serves to replicate, substantiate, and further develop already established or proposed methods and uses of these types of analyses and analytical tools. As such, the methods and content of this research is focused on avoiding the pitfall of re-inventing the wheel.

Research Questions

1. Can changes in clay mineral be detected by a mobile “field” spectrometer under laboratory conditions in natural soils?
 - a. Can the spectral data be used to predict clay concentrations using modelling software?
 - b. Is there a detectable correlation between rainfall intensity and changes in soil characteristics?
2. Can the formation of erosion features (rills) be detected in natural soils using photographic analysis software under non-ideal conditions (directional lighting)?
 - a. Is the formation of rills (number, confirmation, etc.) correlated with rainfall intensity?
 - b. Is the formation of rills (number, confirmation, etc.) correlated with soil slope?

In order to better detail the content of the questions and to more completely describe the underlying topics that must be addressed in the process of research, the principles of the topic will be examined. Following this delineation of the research concepts, the specific methodology used to answer these questions will be described in detail.

Research Principles

Three main topics need to be addressed in order for the research to be detailed in a comprehensive manner. In this section, we discuss the topics of erosion processes, hyperspectral data, partial least squares regression, and rill formation identification. This provides the background for understanding the research topic as based on scientific principles. Following the theoretical discussion, each topic is briefly discussed more in the context of the research purpose, as presented in the previous section. In so doing, we may link the scientific ideas with the research questions at hand. Furthermore, the discussion provides an introduction to the specific methodological process of the research, which will be discussed in the next section.

Erosion Processes

As discussed previously, there is one primary topic which serves as the basis for the entire body of research. This fundamental principle is that rainfall characteristics and the slope of the soil surface will have an impact on the erosional processes at work. These erosional processes, as mentioned in the previous section, are called splash erosion, laminar (or sheet) erosion, and rill erosion. Each erosion type adheres to different mechanics, and is dominant at different locations along the slope. The dominance is primarily governed by the carrying capacity of the eroding water. Splash erosion is the detachment or destabilisation of soil particles and aggregates due to the kinetic energy of the raindrops as they initially hit the surface of the soil (Lado, Ben-Hur and Shainberg 2004). Given the relatively small size of raindrops, this amount of energy is comparatively small, and as such will only affect the smallest or most unstable particles. Laminar erosion represents the movement of a sheet of water over the soil surface. The sheer stress of this water sheet overcomes the cohesive forces of the soil matrix, causing soil particles to break off and be transported away (Alberta Department of Transportation n.d.). While this type of erosion can be severe it is less noticeable than rill erosion, and may be considered more diffuse. Rill erosion occurs when sufficient water is flowing over the soil surface to cause an incision into the surface. This results in a vertical groove in the soil, which is more easily identified visually than laminar erosion. The types of erosion discussed above do not occur exclusively, however, the dominance and intensity of each is influenced by the nature of the rainfall event and the slope of the soil surface (Lal 2001).

The most easily generalised impact soil surface slope has on erosion processes is the relative dominance of each type according to the relative position along the slope. Figure 1 demonstrates the generalised form of this principle. At the top of the slope, splash erosion dominates the erosion processes. This is primarily because the carrying capacity of the water is small compared to the kinetic energy of the raindrop itself. In the middle of the slope, there has been more area accumulating water and it has also had time to gain in velocity. This increases the carrying capacity of the water to the point where it is greater than the kinetic energy of the raindrops. Thus, laminar flow dominates in the middle slope by forming a thin sheet of moving water, picking up soil particles as it flows. Further down the slope, after the expected increase in amount and velocity of moving water, the soil surface integrity is compromised, resulting in an incision which serves as a channel for future water, increasing

the presence of rills. Rill erosion becomes the most noticeable erosion process at this point. According to this series of events, as the carrying capacity of the water increases down the slope, the dominant erosion process progresses from splash erosion to laminar erosion to rill erosion, as depicted in Figure 1. Not included in the image, however, is the logical extension of this increasing carrying capacity. Not only will the type of erosion change, but so will the amount and size of material that is eroded. As the energy available to the water increases down slope, more soil particles can be carried by the eroding water, as well as larger soil particles (Merritt 1984, Lal 2001). As such, at the top of the slope, only the smallest particles (clay size) will be eroded, and will become increasingly more eroded further down the slope. This provides a general overview of the influence of slope on erosion processes, but does not address the influence of rainfall characteristics on erosion processes.

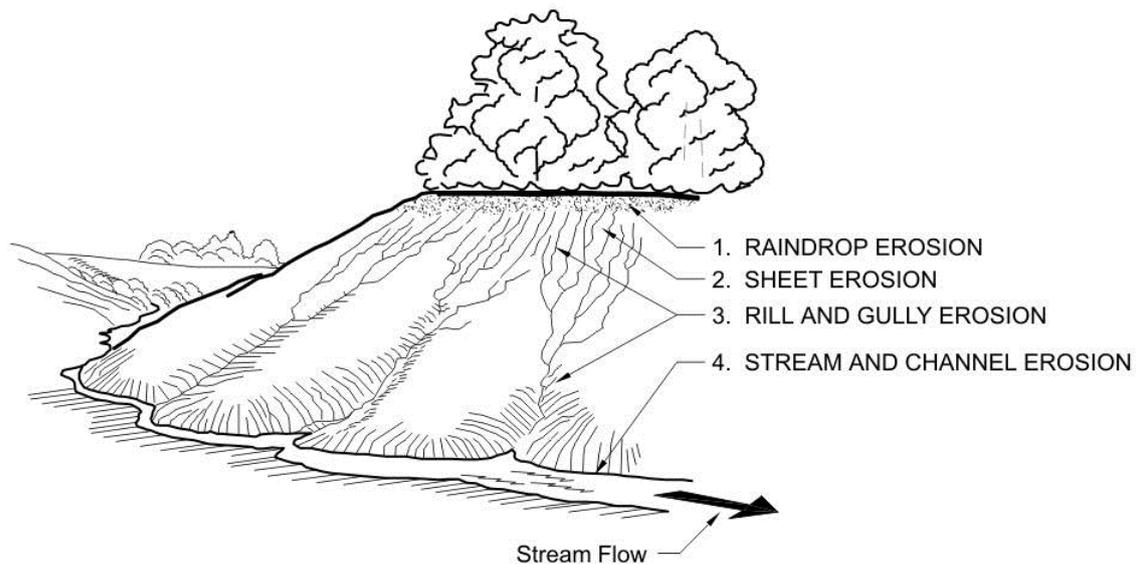


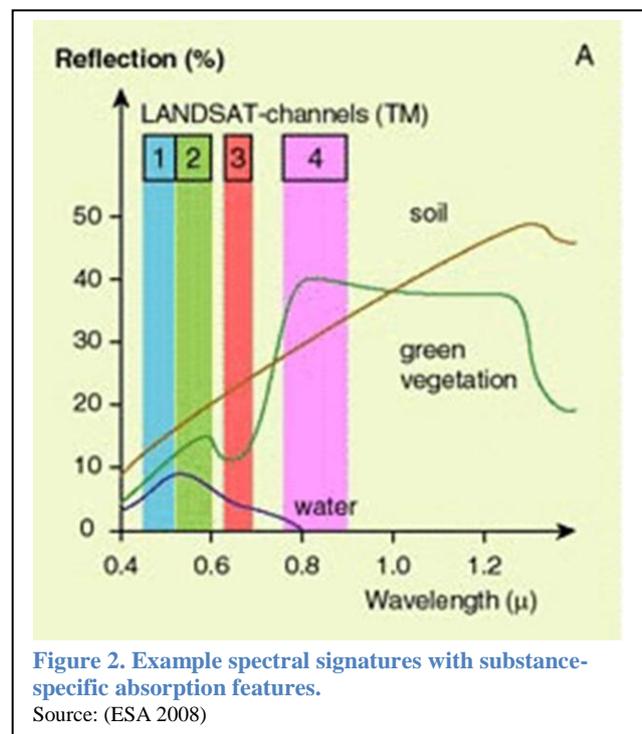
Figure 1. The erosion processes along a hill slope (Alberta Department of Transportation n.d.).

For this research, three rainfall characteristics are the most relevant. These rainfall characteristics are rainfall intensity (depth/time), raindrop size, and duration of rainfall event. The lesser of the two, raindrop size has an impact on the amount of splash erosion that occurs per raindrop. This can influence the erosion over the whole slope simply due to the additional kinetic energy being able to dislodge more and bigger soil particles. However, the logical extension of this is that it also influences the overall erosion process along the slope by raising the threshold at which the different types of erosion become dominant. The intensity of the rainfall event represents what is typically thought of as “how hard it’s raining”. Defined more precisely, rainfall intensity is represented as a cubic amount of water per area of soil per time. This is then simplified to depth per time (mm/hr). The final characteristic is the duration of the rainfall event, which need not be explained beyond its name. Soil loss equations typically combine these rainfall characteristics into one term, representing the erosive potential of a rainfall event without considering the soil conditions. As any of these characteristics increase, so too does the potential for erosion.

In an experimental setting, controlling the rainfall characteristics and soil slope can be relatively easy. Slope can be controlled by the use of movable soil boxes, which can be adjusted to meet whatever slope requirements are needed. Rainfall intensity, duration, and drop size can be calibrated in a rainfall simulator. With these factors fixed or adjustable, the only remaining factor, then, is the amount and type of runoff. This allows for the initial link between variations in rainfall intensity and changes in soil texture distribution along the slope (erosion). However, linking back to the previous discussion of including remote sensing in the analysis, it becomes clear that being able to perform soil texture analysis with samples taken directly from the soil surface is not enough. We also need to be able to link remote sensing data, in this case hyperspectral data, to the soil texture changes.

Hyperspectral Analysis

The most fundamental principle in multi- or hyper-spectral remote sensing is that different substances reflect electromagnetic radiation differently and uniquely. From this principle, different substances may be consistently and accurately identified using multi- or hyper-spectral data. More specifically, it is the regions of the spectrum that are either strongly reflected or not strongly reflected that result in distinct spectral signatures. Figure 2 demonstrates this principle by showing the spectral signatures of three commonly sensed materials. In analysing spectral data, focusing on specific wavelength ranges (or channels if the sensor utilises that method of organisation) allows for individual pixels to be categorised as one substance or another. Examining Figure 2



Error! Not a valid bookmark self-reference. shows that if a pixel has a deep absorption feature near 0.7 μm , and a strong reflectance feature near 0.8 μm , the pixel is most likely covered with vegetation. While the overall shape of a spectral signature is telling for a guided analysis, automated analysis will necessarily have to rely on specific absorption and reflectance features at specific electromagnetic wavelengths. Fortunately, most substances have been studied, meaning that many spectral indices exist in order to detail the absorption and reflectance features of said substances.

This simple method of identifying broad groups of substances (“soil”, “green vegetation”, “water”) works well if the analysis is directed towards coarse categorization and measurement. However, it is possible to further differentiate between constituents of a heterogeneous substance. Past experience with multispectral data reveals that determining the difference between green vegetation species (bulrushes versus reeds) is possible with

sufficient fine-scale spectral data. An extensive body of literature indicates that similar distinctions in soil components are also possible under similar data treatment methods (Stenberg, et al. 2010; Ben-Dor 2002; Ben-Dor and Benin 1995; Ben-Dor, et al. 2009). However, the fine level of spectral resolution required to differentiate between the small spectral features is complicated due to the heterogeneity soil. Therefore, while the general principle of spectral signatures is useful in monitoring possible changes in soil characteristics, it is necessary to process the overlapping features of the various components of soil in order to ensure that detected changes are true changes, not just some other substances spectral signature. Various methods are automatically available in standard spectral analysis software.

One such promising method for processing spectral signals is the “near infrared analysis” method (NIRA). NIRA operates under the assumption that “a concentration of a given constituent is proportional to the linear combination of several absorption features” (Ben-Dor 2002, 177). The NIRA method has the benefits of “rapidity, convenience, simplicity, accuracy, and ability to analyse many constituents at the same time” (Ben-Dor and Benin 1995, 364), while requiring a small number of wavelength “channels” (Ben-Dor 2002). Therefore, given the technical limitations of the equipment available (only a mobile field spectrometer and general software is available), the NIRA method, or a similarly simple method, seems to be the most suitable for this research. Furthermore, simple procedures meet the overall goal of “decision support” by increasing the accessibility of the process itself.

NIRA is a two-step process whereby the user first develops a predictive equation for spectral responses based on empirical data in order to calibrate their model. The second step is to test the now calibrated predictive equation against “unknown” samples (Ben-Dor and Benin 1995). Software packages exist that automatically develop the necessary model, performing the necessary calibration and testing with minimal room for human intervention (and related inconsistencies). The ParLeS chemometric analysis software is one such software package, and seems to be representative of the capabilities and processes expected for NIRA. This software is capable of taking in two different types of data (spectral and soil texture data), and by performing a regression on it, determine the level of correlation between them. This allows for the analysis to go forward using already established methods and “standard” settings in order to evaluate the results. It is important to remember that in this form of NIRA, we are not meant to be looking for causal relations, but instead are focused entirely on the *correlation* between hyperspectral data and soil texture data.

While it is easy to become caught up in the process of developing a predictive equation to determine soil component concentrations based on spectral data, it is important to keep the overall logic of the process in mind. The research goal is to explore the possibility of determining correlation between soil components, rainfall intensity, and/or slope. Rainfall intensity and slope are already known, while soil component concentration is not. Untreated spectral data cannot directly reveal the concentration of soil components, however. As such, it is necessary to follow the NIRA method to establish the concentrations through spectral data.

Partial Least Squares (PLS) Regression

In order to determine if a predictive correlation can be established between the hyperspectral data and soil texture measurements, the data must be subjected to a statistical test for determining such relationships. Partial Least Squares Regression (PLSR) represents such a statistical test. It is a “recently” developed method to establish a predictive model between one response variable and many possible predictor variables. PLSR has the benefit over other forms of regression in avoiding artificially creating a relationship, called overfitting, (Tobias 1997) and “increasing predictive power” (Gibbons 2009, 13). Furthermore, PLSR is focused entirely on determining predictive correlation between the variables, while no attention is given to causality. This matches the requirement of the previous research principle.

PLRS has been used in many different settings, and a large body of “chemometric” (relating chemical and spectral data) studies exist (Calabrò, Leone and Amenta 2010, Rong 2009, Stenberg, et al. 2010, Tallada, Palacios-Rojas and Armstrong 2009, Vincentie 2012). Calabró et al. (2010) even determined that marrying hyperspectral data with PLSR can successfully determine texture changes. This is due to PLRS’ ability to deal with a large number of possible predictors such as the hundreds of wavelengths in spectral data. Additionally, PLRS is a widespread model development method because it is readily available in robust, well-established software packages that allow for easy use, in line with one of the guiding purposes of this research.

While developing the PLSR model represents the main body of the PLRS model, without a form of verification, it can only be determined how internally consistent the model is. In order to assess the predictive capacity of the model, a set of data not included in the model construction needs to be tested. After testing the model in this way, the predictive validity of the model can be assessed using another set of statistical values. Chang, et al. (2001) recommend a three-category system in assessing the predictive power of such models based on the ratio of the standard deviation to the root mean standard error of cross validation (RPD) and the R^2 . This scheme is shown in Figure 3.

Category	RPD	R^2
A (highly predictive)	>2.0	>0.8
B (can be improved)	1.4-2.0	0.5-0.8
C (not predictive)	<1.4	<0.5

Figure 3. Model predictive capacity based on RPD and R^2 (Chang, et al. 2001).

Rill Formation Identification

Taking a step away from the inclusion of hyperspectral remote sensing data and returning to the initial principle of erosion processes, we can address the final aspect of the research at hand: rill formation. As discussed previously, rill formation can be conceived of as the final, most concentrated form of erosion. This concentration of relatively large volumes of water in small spaces increases the velocity of the water, increasing the carrying capacity of the eroding water. The result is the formation of small incisions in the soil surface. Definitions of what constitutes a rill vary from source to source, including the practical definition of “small enough to not interfere with field machinery operations,” (Wall, Baldwin and Shelton 1987) and the more precise “75 mm or less in depth” (Alberta Department of Transportation n.d.).

This leaves a great deal of interpretation as to what a rill exactly is. This represents the first point of discussion for examining the formation of rills in an experiment. If a rill is any water caused incision in the soil surface less than 75 mm deep, or shallow enough not to cause problems with field equipment, what one person says is a rill might be another person's early gully.

In order to avoid this ambiguity, and to make the process more accessible as per the purpose of the research, it is useful to look for more rules-based identification methods. One possibility is the use of object-oriented image analysis software. Object-oriented software contains a set of rules or algorithms that allow for the contextualisation of individual pixels in order to "cluster pixels into salient image regions" (University of Toronto n.d.) rather than relying on a range of depths of the feature. This method allows for the relationship between pixels to be considered during the classification process, including shading, geometry, hue, and colour (Chávez 2012). Since rills are identifiable by human pattern-seeking methods, it should likewise be possible to identify them with such algorithms. The difference being, as mentioned previously, that the software has not preconceived understanding of what constitutes a rill, and as such must follow the exact same identification rules regardless of the surroundings. In order to assess the validity of such a method, however, a human mind is still required. This strength of being completely without context- or knowledge-bias also means that the software cannot tell the difference between a twig, rock, or rill. It only knows coherent objects. As such, a human training set can serve as a baseline for assessing such an assessment (Shruthi, Kerle and Jetten 2011).

All of this requires obtaining rill erosion, however. As discussed previously, the degree of erosion is the only factor that is not easily controllable. Therefore, it is entirely dependent on the rainfall and soil characteristics.

Key Take-Away

Figure 4 represents the interaction between the key research principles under the framework of the research purposes and questions. The three key concepts are erosion processes, hyperspectral analysis, and rill erosion identification. These concepts can be linked directly and indirectly in order to attempt to answer the questions put forward previously. Each one has a few important related issues that must be kept in mind throughout the research process, however.

Erosion processes are the most fundamental principle. Three types of erosion, operating under different mechanisms, dominate the erosion process under different conditions including position along the slope, soil slope, and rainfall intensity. In general, we should expect a preferential sequence of erosion, with the smaller soil particles being eroded first. This holds true regardless of which erosion type is dominant. Furthermore, erosion increases downslope and as slope increases. This is the result of an increase in the amount of water available for erosion and the velocity of this water increasing, both of which increase the carrying capacity of the eroding water. Rainfall characteristics represent the core controllable factors, and increasing rainfall intensity in particular is generally deemed to result in

proportionally increasing erosion potential which we expect to increase the changes in soil texture.

Direct soil texture measurements are not in line with the overarching programme purposes of the research, however, and as such it is crucial to include a remote sensing aspect in identifying changes in soil texture. Using hyperspectral data, focusing on the near infrared region of the spectrum, has been successfully used in the past to identify soil chemical properties (such as clay mineral content). Identifying texture content, however, has received less widespread attention. In order to link hyperspectral data to the texture content, it is necessary to perform a regression in order to establish the degree of correlation between the two. Software expressly for this type of analysis is already in use, and has been used for a wide array of chemical-spectral relationships. Even though its use for correlating with soil texture is less well explored, there does not seem to be any reason it should not work, since the correlation methods do not need a particular *type* of data.

Rill formation identification is relatively simple for humans, but the definition of what is and is not a rill varies, resulting in a possibly inconsistent method of identification. Object-oriented image analysis software is capable of eliminating that ambiguity of identification by following specific rules in identifying coherent objects in an image. However, since the software is not “smart”, a human training set is still necessary in order to see how well it does in comparison to the contextualising human mind. This kind of software has been successfully used in the past at the macro- and meso-level scale of gully development, but the literature is somewhat lacking on the micro scale of rill formation. All of this, however, is dependent on achieving rill erosion, the final stage of the erosion processes discussed previously.

Having now discussed the research principles, we can begin examining how the principles were put into action in the research. The following section will detail the methodology used to carry out the research of this thesis.

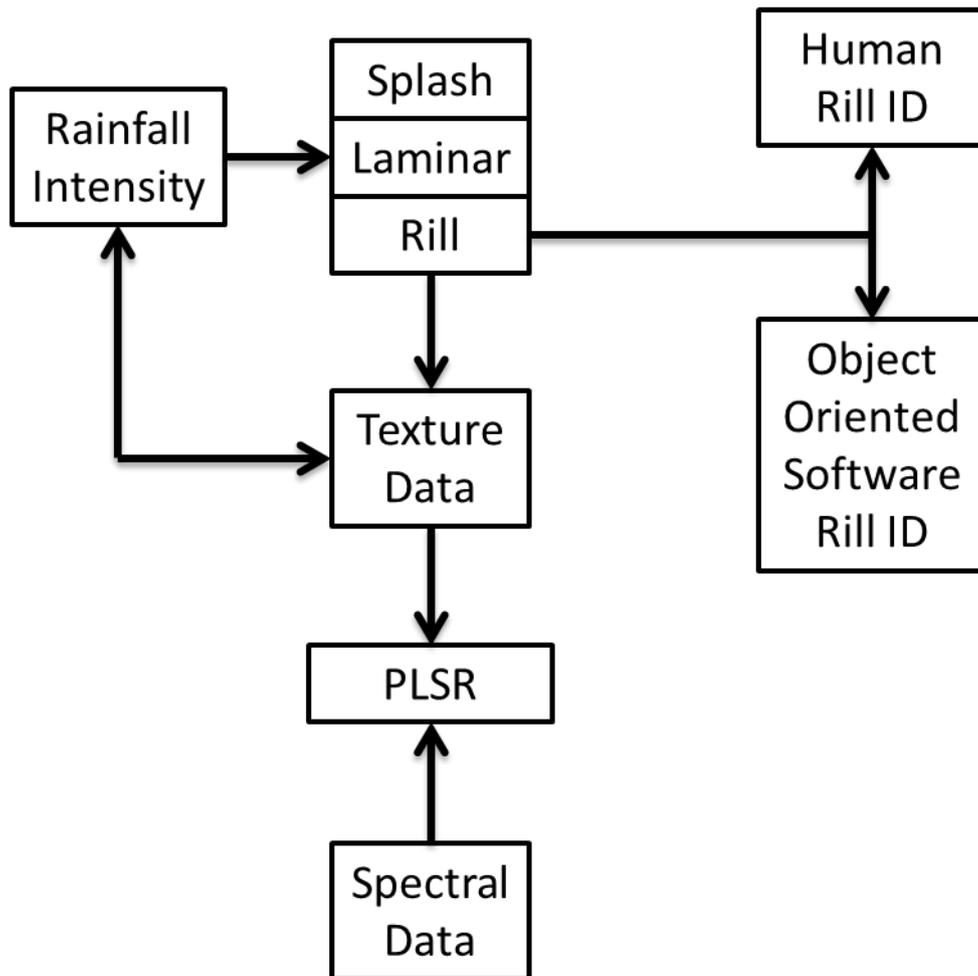


Figure 4. Relationship of the research principles.

Methodology

The methods used in enacting this research have been split into three distinct but related parts. The lab setup includes the basic layout of the laboratory portion of the research. This includes the collection of the raw data (photographs, spectral measurements, and soil samples). The texture analysis topic includes all the analytical steps that have to do with the changes in soil texture. Included in this topic are the methods for measuring the soil texture, relating it to the rainfall intensities, and the development of the regression model using ParLeS chemometric software. Finally, the topic of rill identification based on photographic evidence using eCognition, an object-oriented image analysis software. Figure 6 demonstrates the methodological flow of the research. Each of the items present in the flowchart will be discussed.

Lab Setup

Four soil boxes have been placed inside of a large rainfall simulator that has been calibrated previously. Each of the soil boxes have been placed in order to provide a range of rainfall intensities, although only a *range* of intensities are known per box, instead of a more fine scale intensity map. The boxes were filled at the beginning of each day with soil taken from a nearby plot. This soil represents a “natural” soil, despite the fact that it has been extensively disturbed and turned. First, baseline photograph, hyperspectral measurements, and soil measurements were taken. The rainfall simulator was then turned on and allowed to run until runoff occurred, at which point a timer was started allowing for 15 minutes’ rainfall. At that point, the rainfall simulator was stopped, and another set of measurements were taken. This process was repeated four times, with measurements being taken each time. This procedure was designed to accommodate not only myself, but another experiment examining soil surface roughness, and so soil samples could not be taken until the very end in order to avoid disturbing the soil surface.



Figure 5. The rainfall simulator, with soil boxes inside.

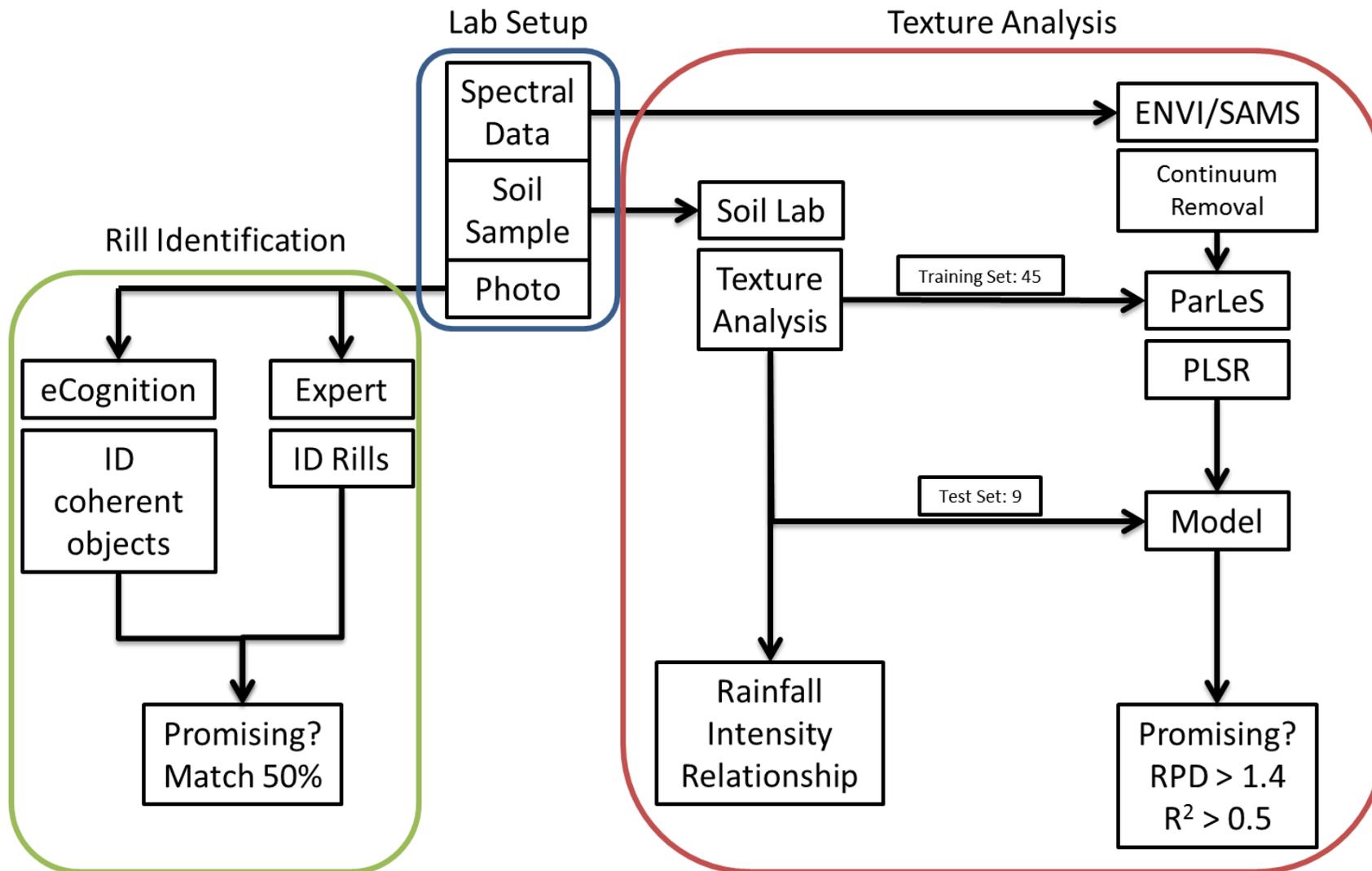


Figure 6. Flowchart of methodological items.

Hyperspectral Data Collection

The main body of the experimental design focuses on how to reliably collect spectral data over the same area of each soil box while maintaining approximately the same distance from the soil surface, and not being influenced by changes in the ambient illumination. In order to address these issues, a basic rig has been designed to hold the spectrometer sensor in place. The spectrometer's fibre optic sensor is mounted first on a "fore-optic" which limits the view of the fibre to 8 degrees. This serves to reduce the size of the "picture" the spectrometer takes (field of view). Functionally, reducing the field of view also reduces the heterogeneity of the sensed area, which reduces the possibility of overlapping spectral signatures due to multiple types of substances being in the same pixel. The fore-optic is then mounted on the rig, which rests on the frame of soil box. This allows the sensor to be perpendicular with the soil surface, and to be approximately 8.5 cm away. Beside the fore-optic is a lamp which is also mounted on the rig. This lamp is meant to reduce the effect of



Figure 7. Soil box and approximate position of spectral and soil sample collection points.

variations in the ambient lighting by simply overpowering it. The rig with lamp and sensor can then be slid along the frame of the soil box, taking measurements every 11.6 cm for three of the four soil boxes. This particular method of spectral data collection is not novel, and has been successfully used in other studies (Calabrò, Leone and Amenta 2010, Rong 2009). Figure 8 demonstrates the similarities between the setups, despite the different subject matter.

Measurements are taken before the first rainfall simulation, and each time the rainfall simulator stops. Therefore, each day's dataset consists of 6 sets of 18 measurements per set. The procedure for taking each measurement set is as follows:

1. Optimize spectrometer with 100 W lamp shining on spectralon reference panel
2. Calibrate spectrometer on white reference panel
3. Turn off lamp
4. Mount rig on box
5. Turn on lamp
6. Wait three spectrometer measurement cycles
7. Take measurement
8. Repeat from step 3

The process for one set of measurements took approximately 10 minutes, with the timing of each measurement varying up to 10 seconds.



Figure 8. Comparing the laboratory setup of Rong 2009 (upper left), Calabrò, Leone, and Amenta 2010 (upper right), and this thesis (lower).

Photograph Collection

Following each set of spectral measurements, a set of photographs were taken using a digital single-lens reflex (DSLR) camera. The conditions of these photographs varied significantly, as each photograph was taken free-hand with the automatic flash and automatic focus enabled. However, each photograph was taken from roughly the same position at the low end of the box. These photographs took less than 5 minutes to take. The type of digital camera is not important, as the purpose of this research is to explore common technologies, which implies that no specialised hardware should be necessary. Typically images for object-oriented software analysis are high resolution aerial or satellite image (Shruthi, Kerle and Jetten 2011), and a similar laboratory setup has been used to minutely observe rill formation (Merritt 1984).



Figure 9. Example photograph. Note the hotspot.

The conscious decision to allow the photographs to be non-ideal has distinct implications for the analysis. Specifically, the presence of directional lighting from the flash varies from the typical non-directional lighting expected from the sun or ideal laboratory lighting. Even a cursory examination of such photographs reveals a very distinct “hotspot” (see Figure 9) where the light of the flash is reflected very intensely. This may prove a confounding factor in the analysis because the analytical software includes brightness and hue in its algorithms, and countering such artefact brought about by directional lighting may be beyond the scope of this thesis.

Soil Sample Collection

In order to validate the relationship between spectral data and soil clay content, it is necessary to have a measurement of the actual clay in the soil. This is done by manually collecting a daily reference sample before any rainfall, and the crust of the soil at the end of each day, consisting of the top five centimetres of the soil profile in an approximately 8 cm diameter in addition to a single reference sample at the beginning to represent all of the boxes (assuming homogeneity between boxes). The actual thickness of the sample varied significantly due to the use of a soil knife and attempting to cut through sometimes sandy or cobbly soil. Each soil sample was placed in its own uniquely labelled specimen jar, and put aside for the soil texture analysis at a later date.

The soil samples were taken in approximately the same location as the hyperspectral measurements at the end of each day’s set of rainfall events. Figure 7 shows these approximate locations. It should be noted that the soil sample area is considerably bigger than the area that is spectrally sampled. The spectrometer analysed anywhere between 1.1 and 5.3

cm² for each “picture” taken due to uneven spots in the soil surface caused by the box filling process. The soil sample, on the other hand included the soil surface of approximately 50.2 cm² or less, depending on the presence of large cobbles or other coarse materials that would need to be separated later. This represents a significant difference between the two types of data. The soil samples were much more likely to contain heterogeneous material, but also avoided being inadvertently placed on a boundary between two very different substances (a large cobble and the soil surface, for example). This type of sample collection is unremarkable, and might be expected in any introductory soil science course.

Soil Texture Analysis

The analysis of the soil texture data proceeded in two parallel tracks. After the collection of the soil samples, they were taken to the Amsterdam Free University soil analysis laboratory and subjected to a standard texture analysis procedure. Meanwhile, the spectral data was subjected to continuum removal in order to make the data more manageable and meaningful. These data, once reconciled, were then fed into the ParLeS chemometric software programme. The resulting regression was then assessed for worthiness. Each step is detailed below.

Texture Analysis

To establish the texture distribution of each soil sample, they were taken to the Amsterdam Free University soil analysis laboratory. Due to limitations on laboratory availability only 57 samples (three days-worth of samples and reference samples) were taken to be analysed. These samples were subjected to the laboratory’s standard treatment procedure which included chemically removing organics and breaking up aggregates, as well as mechanically screening for large soil particles or stones. In the available time, all the samples were pre-treated according to the laboratory’s procedure, and approximately 2/3 of the samples were placed in the laboratory’s laser distrometer. The remaining samples were completed by the laboratory manager on my behalf, and the results were sent to me electronically. Nine of these measurements were randomly chosen (three from each days-worth of samples) and set aside as a test set for the ParLeS software later. Based on the fact that the procedure in the lab was presented in a laminated folder that was readily accessible to anyone in the lab, conversation with the laboratory manager, and on personal experience, this method of determining texture is extremely common to the point of being introductory soil science material. Nevertheless, the warning given about the actual *sampling* of the soil still holds true for the analysis step.

Rainfall Intensity Relationship

After receiving the soil texture data, the task of relating the changes in soil texture to rainfall intensity was undertaken using Microsoft Excel. The data was arranged according to box (intensity), date, and size category to allow for more meaningful comparison between samples. The raw analysis values were first included, and then the differences between the reference samples and the raw analysis values were treated similarly. A linear trendline was superimposed on the daily data series along with its equation and R² value for ease of comparison.

ENVI/SAMS and Continuum Removal

The hyperspectral data collected in the laboratory were run through two spectral analysis software packages. The first was the commercial software ENVI. In this software, the spectral data were subject to the process of continuum removal. This process, simply described, makes the actual magnitude of absorption and reflection features inconsequential by laying a smooth curve spanning the highest points in the spectral region selected. This smooth curve counts as the baseline, and the data is transformed from an actual magnitude of reflectance to a proportion of the baseline reflection. This is a standard spectral data pre-processing step if there is the likelihood of some interference (such as water) in some of the spectral samples that depresses the magnitude of reflectance over the whole spectrum. This was determined to be likely, and so continuum removal was considered a safe step. The removal was limited to a specific region of the NIR range, focused on between 2160 and 2260 nm wavelength. The second spectral software was the freeware package SAMS, developed by CSTARs of the University of California, Davis. This software was utilised to verify the continuum removal had taken place, and to compile the samples into a spectral library, and then into a useable form (ASCII). Nine samples that corresponded to the nine soil samples were set aside for testing the PLS regression.

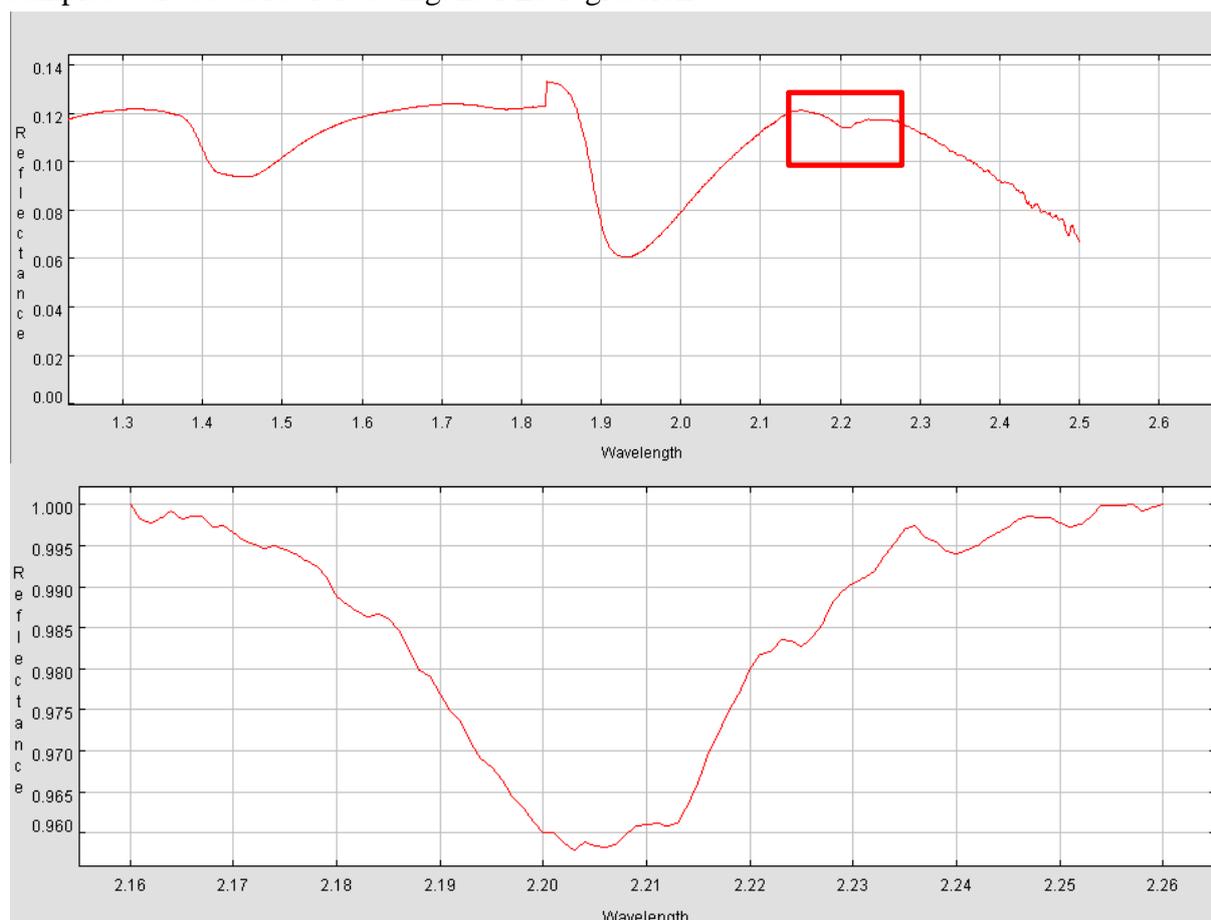


Figure 10. An example of the untreated spectral data with region of interest highlighted (top), and continuum removed data region (bottom).

ParLeS and PLSR

I chose to use the ParLeS chemometric software programme to develop a PLSR model based on its availability, ease of use, and widespread acceptability. See Rossel (2008) for the exact

nature of the software algorithms. The following procedure was followed when developing and testing the PLSR model:

1. Collate the soil and continuum removes spectra into one tab-delineated ASCII file.
2. Import the calibration dataset (45 observations) into the software.
3. Perform no data manipulations.
4. Perform PCA using 10 principle components
5. Perform PLSR cross-validation using a leave-1-out method to avoid overfitting
 - a. Choose the number of factors minimising first AIC and then RMSE
6. Perform PLSR model creation
 - a. Use the number of factors determined from step 5
7. Import the model testing dataset (9 observations) into the software.
8. Perform the PLSR prediction
9. Assess predictive validity
 - a. $RPD > 1.4$, $R^2 > 0.5$ is promising.

Photographic Analysis

The photographic analysis proceeded in two steps. The first was the printing of the digital photos on photo-quality paper and giving them to an expert in the field of erosion. In this case, the photos were given to Dr. Saskia Keesstra. Along with the photos, instructions were given to identify any rills in the photograph. The digital counterparts of these photographs were then entered into the eCognition software package. This object-oriented image analysis software follows the principles discussed previously, and is a widely used software programme for identifying erosion features on the meso- and macro-scales (Shruthi, Kerle and Jetten 2011). Following the general procedure recommended by Chávez and Clevers (2012) for identifying distinct features, a quadtree segmentation was performed on the clearest photographs (where the autofocus had functioned as expected). Quadtree segmentation is an algorithm that divides the image into quarters, and assesses the homogeneity of those quarters. If they are not homogeneous, they are again divided. This process is repeated until the threshold is met, and a single object is identified (Definiens AG 2007, Smith and Chang 1994). This represents the most basic type of image segmentation for identifying objects, and takes the least amount of pre-processing, but also does not conform well to irregular shapes. It was determined based on the work of Knight, et al. (2007) that a 50% match between the human identified rills and computer identified rills would be the cut-off between a promising and not promising methodology.

Results

In this section we will discuss the outcomes of the various analyses. There are three main groups of findings that will be presented. The first group is the relationship between rainfall intensities (which varied by box) and soil texture changes (focused on the smaller soil particles). The second group of findings relates to the correlation between the hyperspectral and soil texture data using the ParLeS chemometric software. Finally, the results of the rill identification will be presented.

Soil Texture and Rainfall Intensity

As discussed previously, each box represents a different range of rainfall intensities. Box 1 included the range of 40-50 mm/hr. Box 2 included the range of 60-70, and Box 4 included the range of 30-40. In increasing order of rainfall intensity, the boxes are arranged as Box 4, Box 1, and Box 3. Based on the principles discussed previously, since the soils are assumed to be the same in each box, we would expect the amount of erosion to increase with the rainfall intensity. Here we will only examine a few representative examples, but the complete compilation of graphs is included in Annex 1 and Annex 2.

Full Frequency Table Graphs

Figure 11 shows the soil texture distribution graphs for Box 1, which is representative of the other boxes. Each data series represents the texture distribution for a measurement point (see Figure 7). We can see a very strong s-shaped curve on these graphs, which is consistent on each day, indicating that the soils, regardless of whatever processes occurred, remained similar in texture distribution. After the internal consistency is established, the similarities begin to vanish. The graph of the 26th appears promising in that the reference sample has the highest clay percentage and among the lowest larger particle percentage. We could infer from this that as the rainfall events occurred, the clay was eroded preferentially, resulting in a relatively larger share of silts and sands. However, we can see that the concentrations for each position do not follow in any logical order. The clay fraction, for example, is not the highest in position 1 (the top of the box). Instead it is preceded by positions 4 and 2. If deposition were occurring, then we would expect the lower sample positions to have proportionally more clay, not one near the top and one near the bottom.

Furthermore, the sample positions' rankings in concentrations vary from day to day. On the 26th, position 1 ranks 3rd in $< 8 \mu\text{m}$ and 4th in $8-16 \mu\text{m}$. These rankings are not matched in either of the other days. It is possible that we may have seen a more definite trend in relation to the reference samples if we had taken one reference per box rather than one reference per day, but the amount of confusion in the full data range defies this explanation. It is possible that the soil was more heterogeneous than assumed, which would result in each position in each box having a different initial ranking, and might propagate that ranking through the experiment.

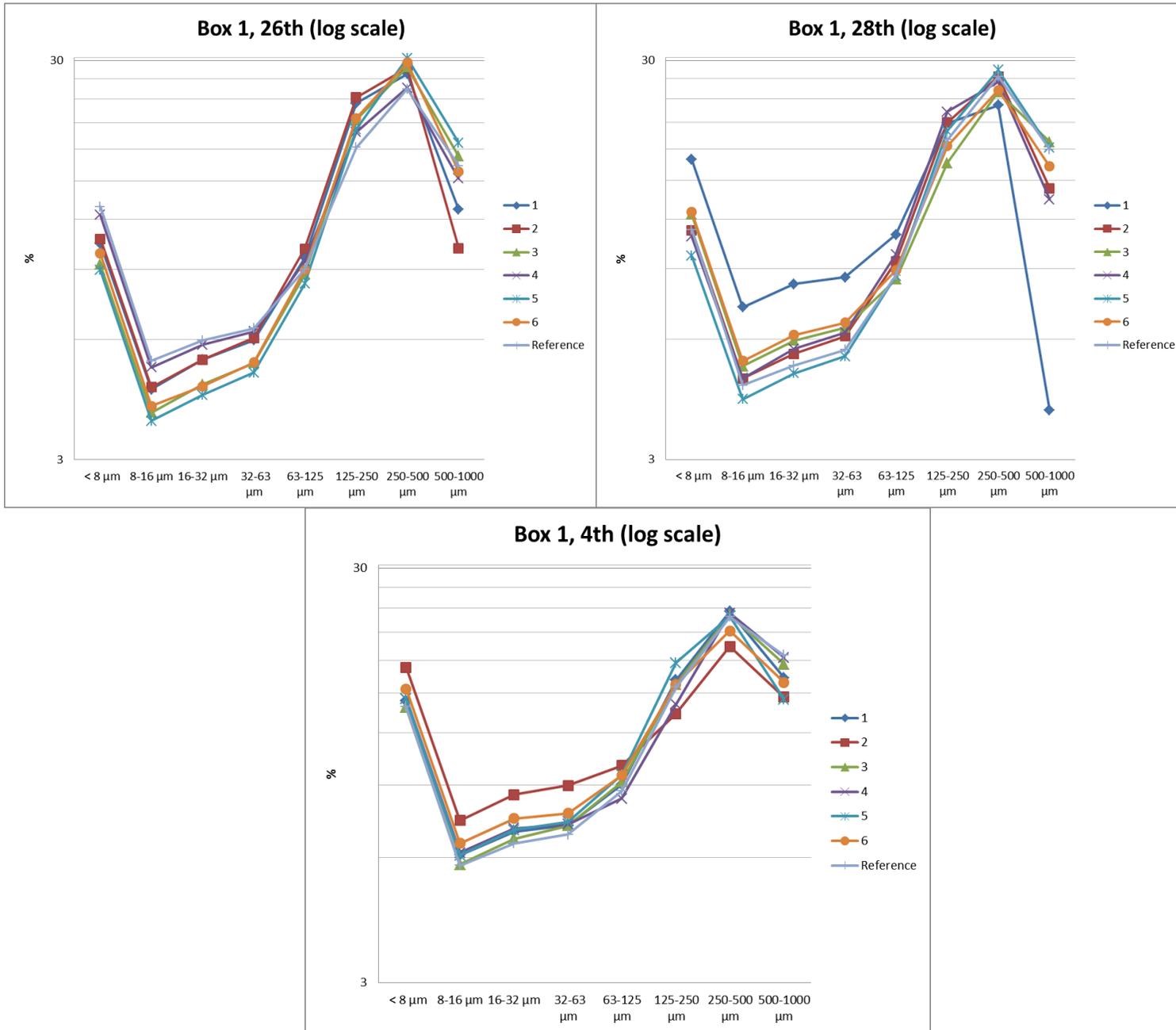


Figure 11. Full frequency table graphs for Box 1.

Size Category < 8 μm

Boxes 4 and 1 respond as expected in any rainfall event sufficiently intense enough to cause transport of eroded material. Figure 12 shows that the per cent clay particles per sample decreases further down the slope. This is in line with the erosion processes principles discussed earlier. However, what we do not see is an increase in erosion, which would be indicated by consistently lower values and a steeper slope, as the rainfall intensity increases. On this topic, the data becomes inconsistent. While the concentrations of clay particles in the first sample position are lower for the 28th and 4th, the opposite is true for the 26th. Likewise, the general trend of the 28th becomes steeper, indicating an increased amount of erosion, this is not mirrored by the results of the 4th and 26th, which actually show a markedly less steep line, indicating a decrease in erosion. Box 2 does not fit well with the other two boxes, as it shows an increasing percentage of clay particles further down the slope, in contradiction to the research principles and other boxes. Furthermore, it also is internally inconsistent with one date showing the expected trend, and two days showing the opposite. We will see as the results are discussed that this is not an isolated event, and that Box 2 consistently follows this pattern. That leaves us with two explanations: 1) A systematic error somewhere in the analysis process, or 2) The actual concentrations increase down the slope. I find the first option to be more appealing from a common-sense perspective. However, given the validity of the other two boxes, I feel forced to leave the Box 2 data as is for fear of correcting something that is actually reflecting reality.

Regardless of possible errors, the majority of daily samples indicate a general negative trend in clay texture content down the slope. However, these trends are not particularly strong, judging by their R^2 values. The largest R^2 belongs to the 26th series of Box 2, set at 0.36. This is in comparison to the lowest value of 0.01 in Box 1 on the 4th series.

Size Category 8-63 μm

This size category presents the same trends and issues as the previous section. Figure 13 details the distribution of mid-sized soil particles. Discussion of this size category is essentially the same as the previous size category, and so we will not examine it in any depth.

Size category 63-2000 μm

Figure 14 shows the soil texture data covering the coarser end of the texture spectrum. We can see that the trends of these graphs are the opposite of the previous graphs. This allows us to say that the measurement technique, at least, is consistent. As the remaining percentage is indeed the difference between 100% and the sum of the smaller size categories, we know that there was no error in the actual measurements. Thus internal consistency is obtained. Once again we see in general that the relative amount of large particles increases down the slope, in agreement with expectations. However, Box 2 presents the same confused situation as the previous size category graphs. Likewise, the rate of increasing percentage does not increase in any noticeably consistent manner.

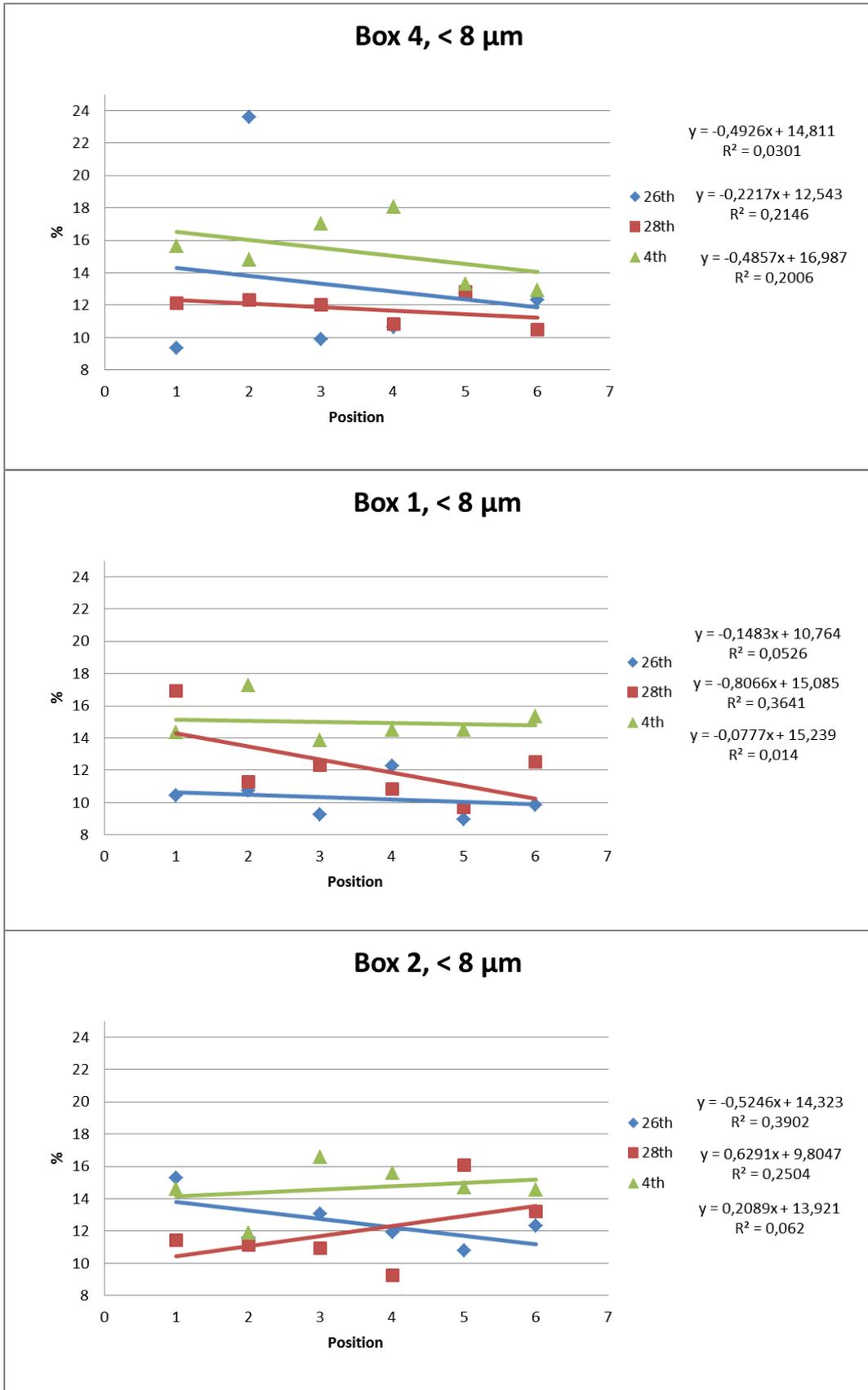


Figure 12. Size category < 8 μm by box.

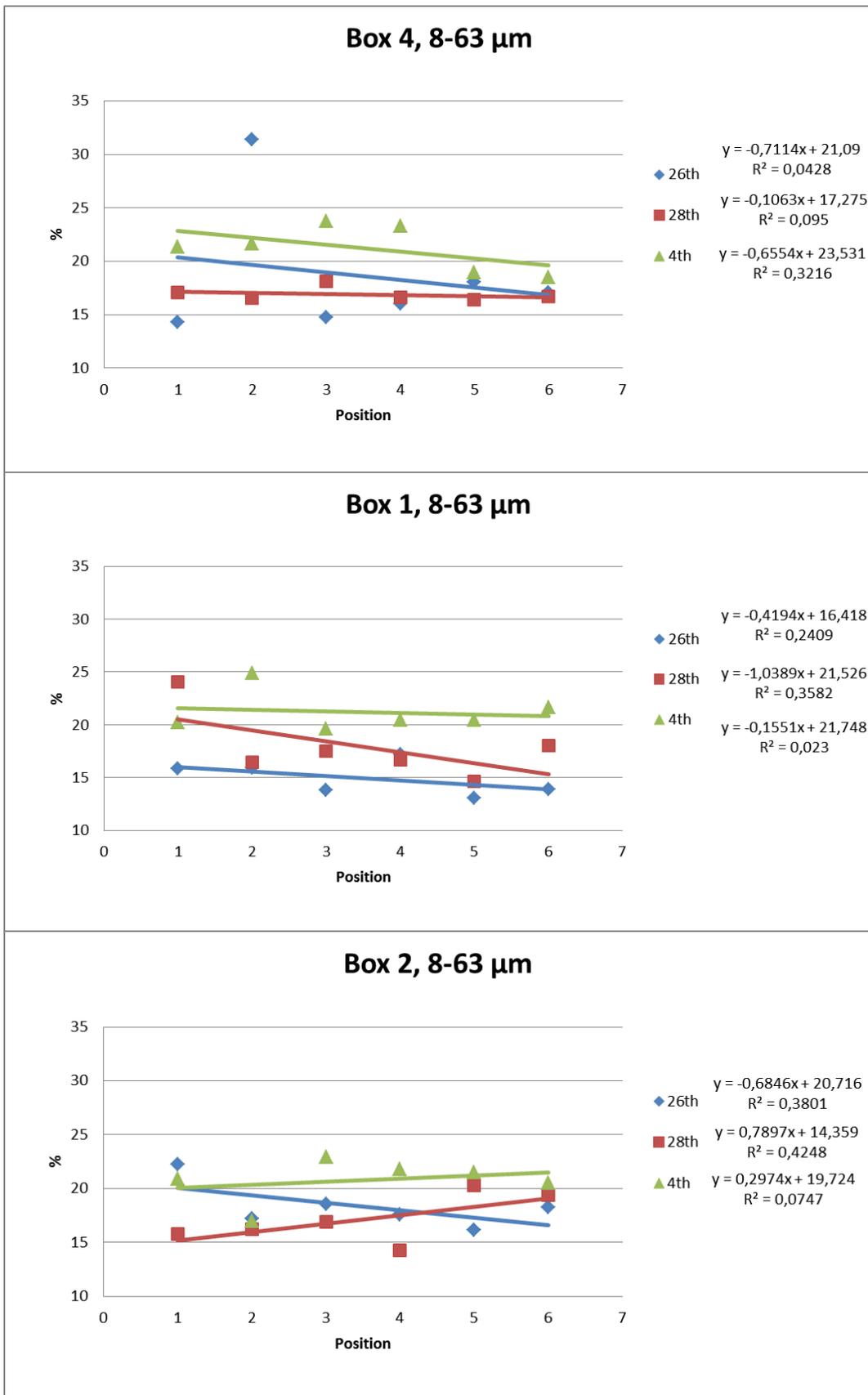


Figure 13. Size category 8-63 μm by box.

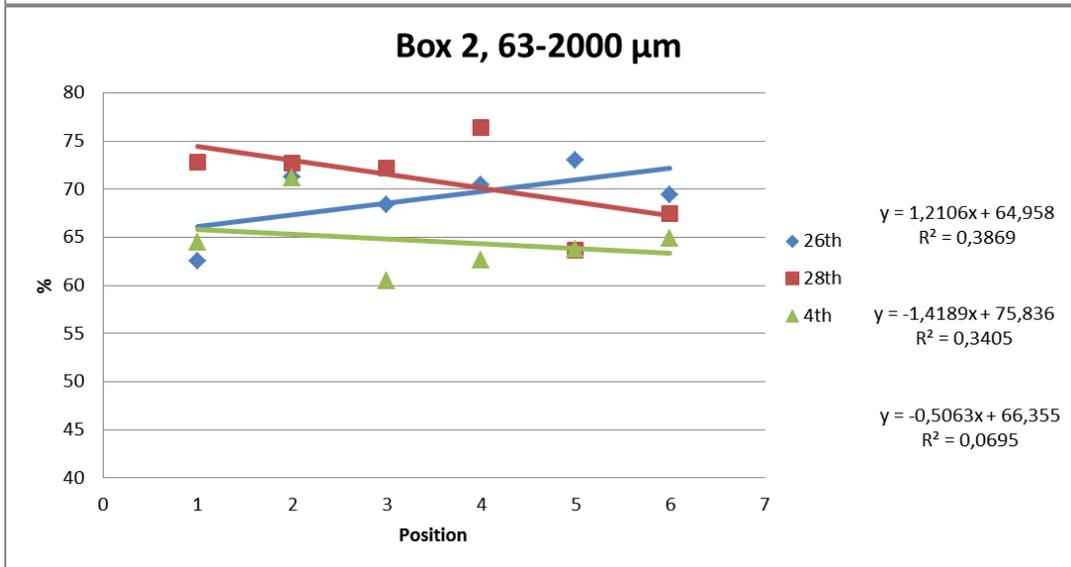
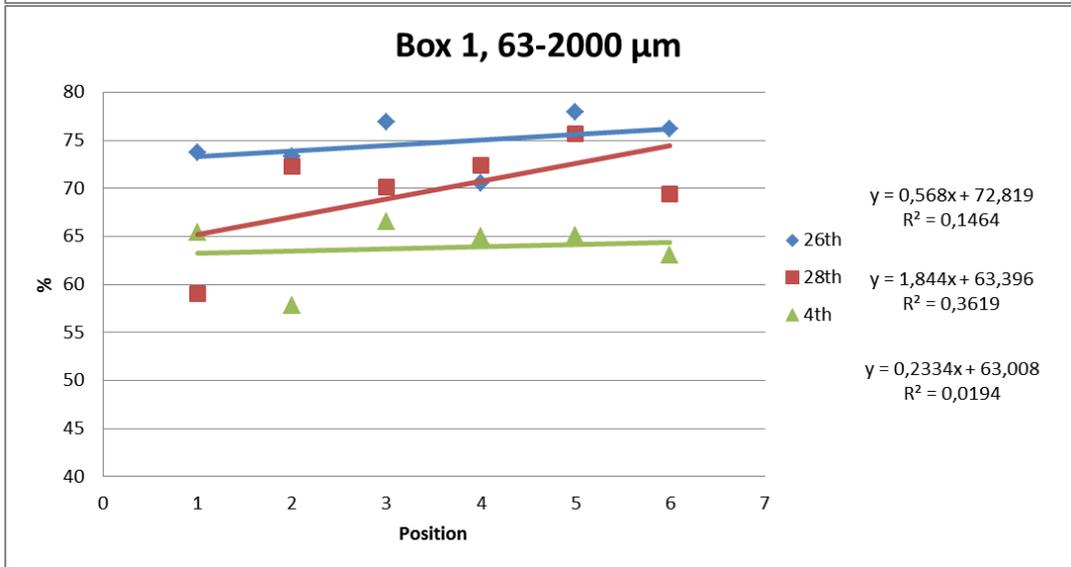
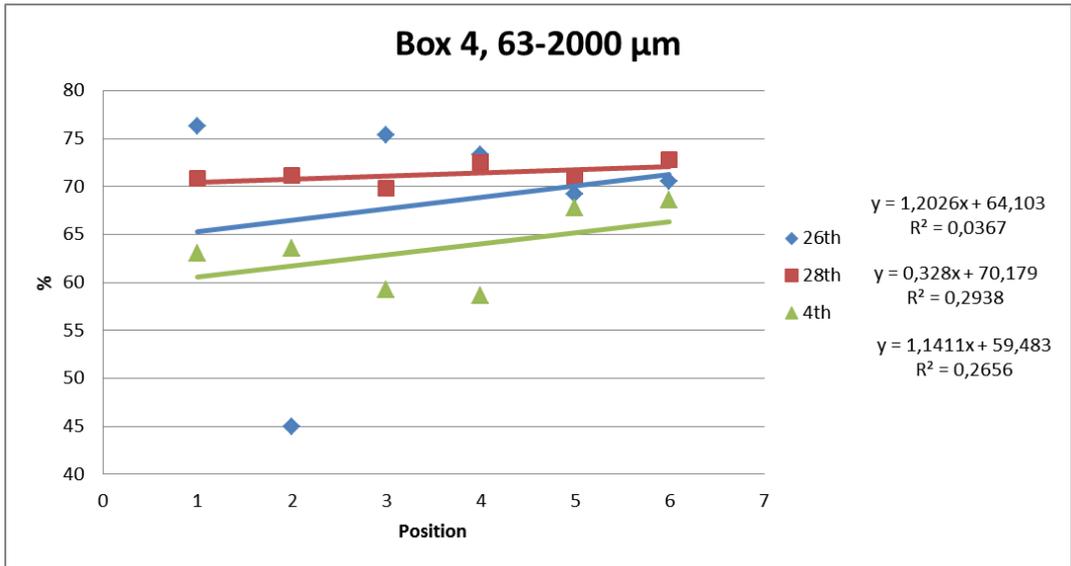


Figure 14. Size category 63-2000 µm by box.

PLS Regression

While examining the soil texture by itself serves the purpose of determining if the results make sense in light of the research principles (which they do in very general terms), the main goal is to explore the possibility of being able to indirectly determine changes in soil texture. Here, we focus on the processing of the soil texture measurements ($< 8 \mu\text{m}$) with its corresponding spectral measurements between 2160 and 2260 nm wavelength in the ParLeS chemometric software programme. Annex 3 documents each step of the modelling and testing process.

What we find is that the software determines that a single factor best explains the variation in the data. This seems to imply that the software has determined the clay content and spectral measurements to be directly correlated. This is supported by the internal verification the software performs, where the model that was just developed is tested against the data used to calibrate it. The result has an R^2 value of approximately 0.84, an RMSE of 2.95, and an RPD of 0.99. Visually, this trend appears quite strong and reasonable. However, it only includes the 45 samples dedicated to developing the model, and it is only a measure of *internal* validity of the model.

To actually assess the validity of the model, the nine samples that were selected previously for testing the model were entered into the software. While we can see that the RMSE has dropped to approximately 1.80, the R^2 has also been halved, resulting in a value of 0.41, and the RPD value is 1.02 (see Figure 15 for details). We can see that the model over predicts higher values (five of the nine samples are predicted as higher than they actually are). Not until we get to the highest values do we see an under prediction. Based on the RPD value that was established as the cut-off in the methodology section, it would seem that this method of inquiry is not a fruitful line. However, it should be noted that this was an intentionally “quick-and-dirty” approach to modelling, and so with some refinement of the methods, may yet yield a more reasonable result.

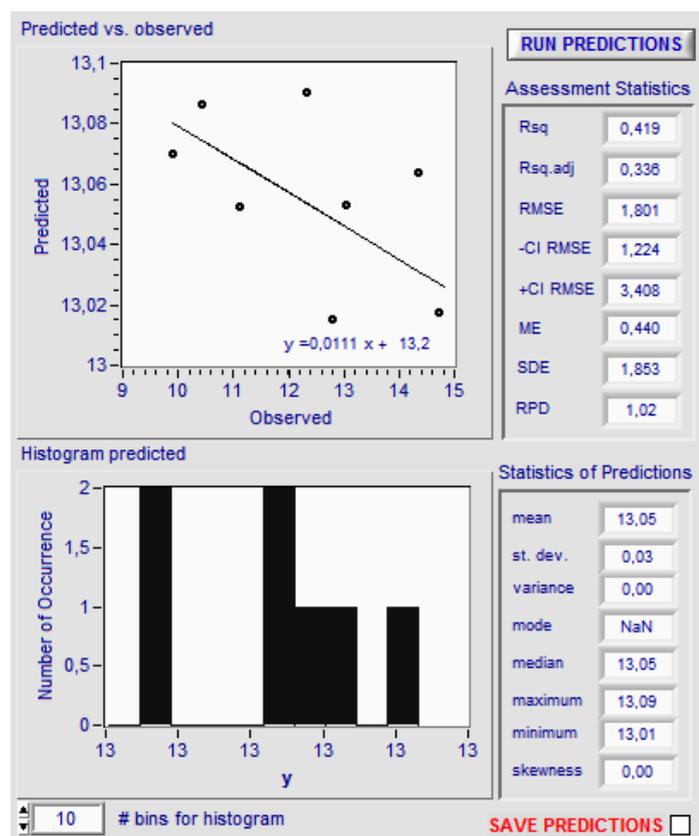


Figure 15. PLSR model test information.

Rill Identification

The rill identification portion of the research proved to be more fundamentally flawed than either of the other two portions. This is due primarily to the failure to obtain rill erosion. Upon giving the photographs to the expert for rill identification, I was informed that there were simply no rills present in any of the photographs. This meant that any attempt to assess the ability of eCognition to identify rills would be flawed since 1) there would be no baseline to compare with, and 2) according to the expert, there was nothing for the software to identify. This would mean any objects identified would be artefacts of the research methods, or something else entirely. Given the body of literature existing proving the utility of eCognition in capable hands for identifying anything from building to trees and gullies, I decided to perform only the most basic quadtree segmentation process using the standard eCognition threshold settings in order to provide a qualitative assessment of the capabilities of eCognition when dealing with micro-level soil surface changes.

Figure 16 shows the box before any rainfall events. In the dry photographs, we can clearly see the programme is capable of picking out the lighter coloured soil aggregates at a relatively low zoom level of 50%. However, we also see a large area in the lower right corner of the box is identified as coherent despite the fact that it appears as homogeneous as the rest of the box. This is probably due to the unevenness in the lighting of the soil surface. Since that area all appears to be in shadow, the software identifies it primarily due to colour or hue, resulting in a false identification.

Figure 17 shows the same box after a day's worth of rainfall events when we would most expect rill erosion to have occurred. As the expert indicated, not rills formed. At 50% zoom, we can begin to see the hints of some linear features at the lower end of the box. While there are not rills, they do provide a point of reference. In general, however, we can see that the quadtree segmentation proceeded further in the wet photographs, indicated by more black spaces. This means that the rainfall event increased the heterogeneity of the soil surface, forcing the software to divide the image to smaller portions than the dry photographs. We still see the area in the lower right corner is falsely identified as more homogeneous than the rest of the box.

By zooming in further, the linear feature of the previous image is more clearly shown as consisting of relatively large homogeneous areas. However, it also increases the noise of the image, making the linear features harder to distinguish from the background noise. These quadtree segmentations represent only the most basic steps in the capabilities of eCognition. Several other initial segmentation methods were attempted, but none provided any clearer images than these, and so were not included in this thesis. If the segmentation process were to continue, in capable hands, it seems that such linear objects might be made more distinct.

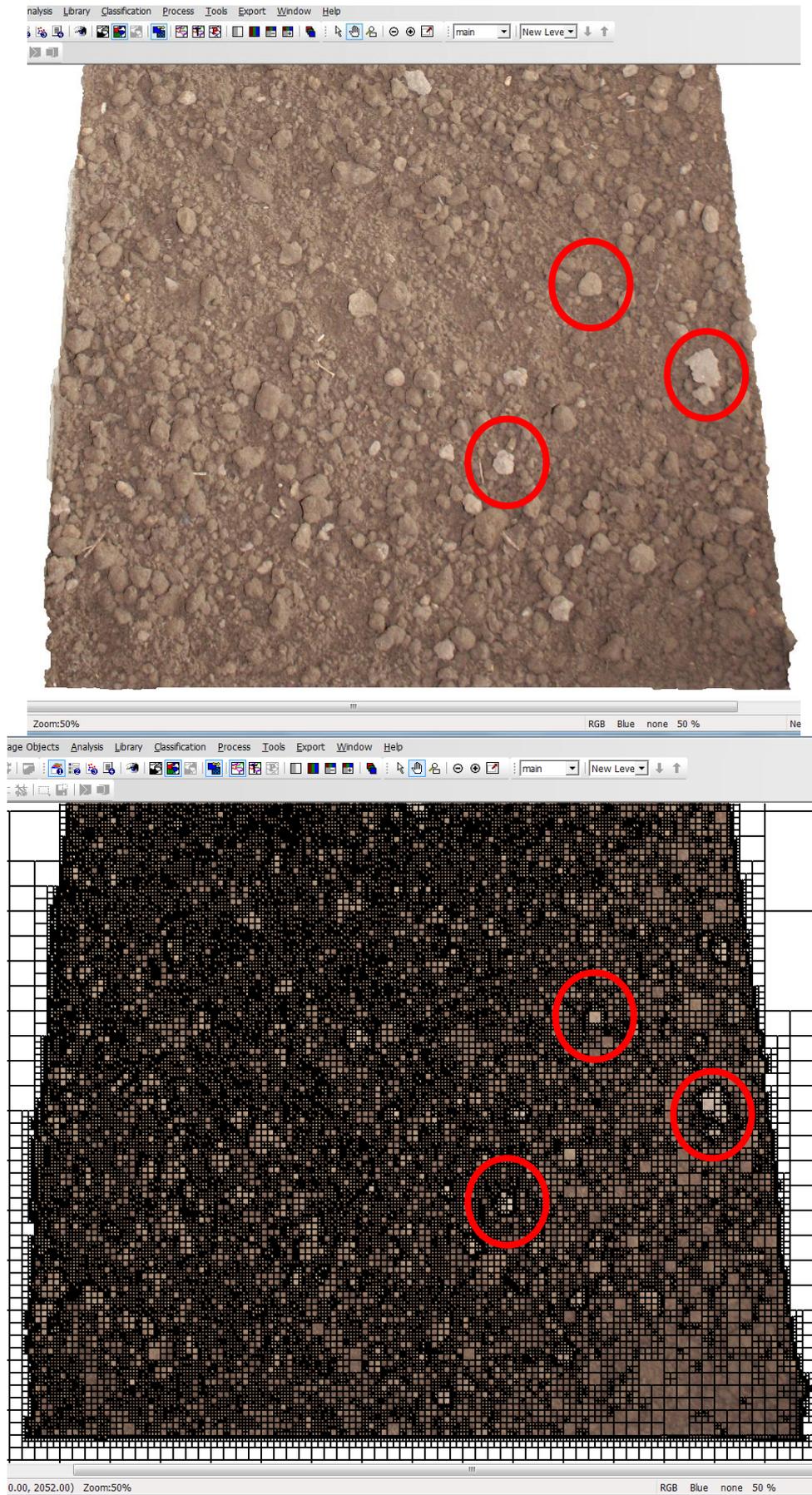


Figure 16. Photo and quadtree of dry box. Note the identified objects.

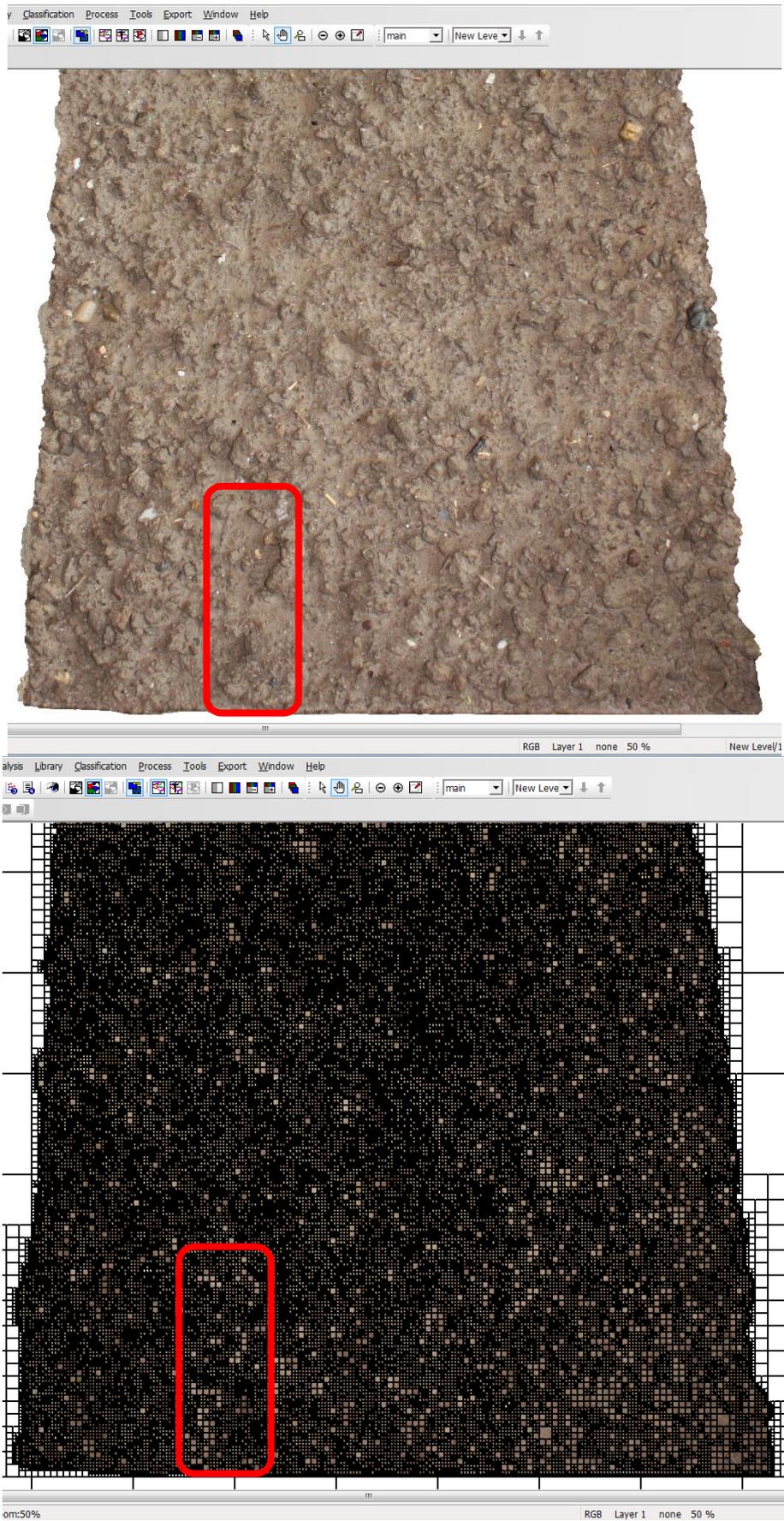


Figure 17. Photo and quadtree of wet box, 50% zoom. Note the similar linear feature.



Figure 18. Photo and quadtree of wet box, 100% zoom. The features are more easily seen, but image is noisier

Conclusion

A wide variety of topics have already been discussed in the previous section as supporting explanations for the results, but in this final section I wish to summarize these points and include some more general comments relating the results with the methods used. Furthermore, this section serves to answer the research questions as concisely as possible.

Research Answers

Q1: Can changes in clay texture be detected by a mobile “field” spectrometer under laboratory conditions in natural soils?

A1: The spectral measurements taken indicate that such changes are detectable.

Q1a.: Can the spectral data be used to predict clay concentrations using the ParLeS modelling software?

A1a.: Under the conditions of this experiment, the PLS regression model of the ParLeS software is not a valid tool to predict clay content using NIR spectral data.

Q1b.: Is there a detectable correlation between rainfall intensity and changes in soil characteristics?

A1b.: The data obtained does not indicate a consistent correlation between rainfall intensity and changes in soil texture.

Q2: Can the formation of erosion features (rills) be detected in natural soils using photographic analysis software under non-ideal conditions (directional lighting)?

A1: Rill formation was not achieved in this experiment, so no conclusion may be reached on this topic.

Discussion

There already exists a wide body of literature on the topics of soil erosion in simulated rainfall settings, spectral-soil characteristic relationships, PLS regressions, and object-oriented image analysis. It is useful to consider how this thesis contributes to this body in order to assess in what ways the outcomes compliment and contrast the extant literature. We will first briefly examine the rill identification part of the thesis, followed by a discussion of the soil texture analysis portion. This then leads us into a concluding discussion focused only on the results of this thesis.

Wider Setting – Rill Identification

In the most general terms, this experiment differs from the general body of literature in the focus placed upon the use of “normal” (red-green-blue) digital photography. Blaschke (2010) demonstrates how the very definition of “remote sensing” in the context of object-oriented image analysis limits its use almost entirely to satellite or high-altitude aerial photography (such as Quickbird). Because of this, the most common discussion of object-oriented analysis is a discussion of meso- to macro-scale phenomena. For soil related studies this primarily

means the formation of gully systems and tracking extant and potential landslide locations (Martha, et al. 2010, Shruthi, Kerle and Jetten 2011). Regardless, the sources acknowledge object-oriented analysis as a method that is still in development, and as such is subject to the birth-pains of any method that has not be thoroughly tested. This is evidenced through the very study-specific methods used, wherein each study requires a different set of identification parameters. In this thesis, we attempted to demonstrate the ability of object-oriented analysis to identify micro-scale features using “standard” out-of-the-box settings, which runs contrary to the typical application of the method. Regardless of these issues, the results of this study, even with the failure to obtain rill erosion, seems supportive of the literature’s assessment of object-oriented image analysis as a way to overcome the inherent problem of encountering heterogeneous surfaces (Blaschke 2010).

Wider Setting – Soil Texture Analysis

The discussion of the soil texture analysis in reference to the wider discussion and body of literature falls into two distinct categories. First is the topic of the methodology, and the second is a discussion of the actual results. The methodology of this experiment is far from novel, and the general setup has been utilised in many other experiments. Rainfall simulators are widely available, and widely used in conjunction with small (1.5 m²) soil boxes. Spectrometers have been used to determine soil characteristics, and PLSR has been used to assess the relationship between soil (and other substances) and their characteristics. However, there are a number of key differences between these past uses and this study. First, this soil was meant to represent a “natural” soil, and therefore was not crushed, packed, nor sieved, representing a departure of standard practice (Merritt 1984, Shainberg, Warrington and Laflen 1992, Parsons and Stone 2006). Furthermore, rainfall intensities and pre-rain soil conditions are typically better understood and more important in rainfall simulation studies (Reicher, Norton and Huang 1994, Chaplot and Le Bissonnais 2000, Frauenfeld and Truman 2004). Not having thoroughly integrated these topics into the research may be a methodological weakness brought about expediency.

The results of this thesis represent an agreement with the implicit (and sometimes explicit) results found elsewhere. Typically, the results of tracking erosion in simulated rainfall events is not expressed as changes in points along the slope, but rather as a total amount of erosion through time. This is a gap in the literature, but perhaps this thesis demonstrates the infeasibility of performing such a study in a simulated rainfall situation. We can speculate along these lines primarily because previous studies found that variations in soil characteristics and variable rainfall intensities, both of which were excluded in the methodology, were found to have important impacts on the total amount of erosion, concentration of erosion through time, and peak amount of erosion (Chaplot and Le Bissonnais 2000, Merritt 1984, Parsons and Stone 2006). Furthermore, while erosion is proportional to intensity (Lal 2001), the slope length found in small soil boxes “does not allow flow velocity to increase sufficiently to reveal a slope effect” (Chaplot and Le Bissonnais 2000, 151), which may lead to the dominance of the less differentiating splash erosion along the entirety of the slope (Agassi and Levy 1991). Thus, we might not expect the texture differentiation in erosion that we would see on a normal hillslope.

The relationship between soil *chemical* characteristics and spectral data has been well explored in the past (Ben-Dor and Benin 1995), while the relationship between soil *texture* characteristics and spectral data is less well explored. Therefore, we cannot definitively place the results of this thesis in the wider discussion of the topic. However, we can say that the use of PLSR to explore such relationships is very well established as a practical testing tool (Abdi 2003, Rong 2009, Rossel 2008). Furthermore, we may go on to say that PLSR worked exactly as expected, resulting in a definitive assessment of the statistical relationship between spectral data and soil characteristics.

This Thesis – Rill Identification

The simplest failure in the results to explain is the rill identification. The failure to form distinct rills meant that any analysis would necessarily be qualitative. In order to obtain rill formation, the methodology would have to be changed at the lab setup level. It is clear that the carrying capacity of the water never reached the threshold at which rill erosion would dominate. This could be due to insufficient rainfall intensity, insufficient rainfall event duration, insufficient slope or slope length, or a combination of these factors. If the experiment were designed from the ground up with obtaining rill formation in mind, it would be possible to obtain it. However, the laboratory set up was designed with multiple purposes in mind, and so to limit the impact of one experiment on the other, rill erosion could not be guaranteed.

Nevertheless, the basic image analysis performed on the images of the boxes does display the software's capability to identify coherent objects. This is particularly strong in the dry image, where colours have very high contrast and the surface is covered with many distinct aggregates. The wet image, by contrast, is noticeably more heterogeneous according to the quadtree algorithm. Not continuing the segmentation process in an attempt to unify those few distinct objects discussed in the previous section severely hinders any possibility of providing a meaningful qualitative assessment of the process. In the future, focusing on refining this technique could bring such objects into sharper relief.

This Thesis – Soil Texture Analysis

There were two parts to the texture analysis which depended on each other. First, we wished to see if rainfall intensity impacted soil texture in a measurable way. This was to determine if the methods and setup resulted in consistent and understandable results. Second, we wished to see if the soil texture could be predicted from spectral data using a PLS regression. The theoretical use of this would be, in the future, to take spectral measurements of soil and use it to infer the soil texture distribution (in this case, clay content).

As noted in the results section, the general trend of the first part matched with the expectations. There was less clay at the bottom of the slope than at the top, in accordance with the erosion processes principles. However, the changes in intensity did not appear to have any consistent impact on the total percentage of clay present, nor the rate at which the clay content decreased down slope. Furthermore, the results of *part* of one of the boxes runs

counter to the trends of the other samples. It is possible that a very simple human error could have caused this confusion, since two people were separately responsible for the running of the soil samples through the laser distrometer. However, it is also possible that the soil itself simply had a drastic increase in clay concentrations in that region of the box. This is not unreasonable, considering how different the soils were from day to day. If the rainfall intensities were indeed the same, and the soils were taken from the same source, we would expect the texture distribution to be similar. Instead we saw a variation of 5% in the finer particles, and 20% in the coarser particles. While a 5% variation might not be considered much difference under normal circumstances, when the changes in clay content are in the neighbourhood of 2% or less, that margin of error can completely contain all visible variation.

Possibilities for error abound in every aspect of this line of analysis. For example, when taking the spectral measurements, illumination and distance from the soil surface varied. Furthermore, as the soil surface subsided and changed shape, the spectral samples experience variable fields of view. This could influence both the magnitude of reflectance and the spectral features of the sample, as rocks and organic matter was exposed throughout the experiment. While this lends credibility to the statement that these soils represent natural soils, it confounds the hope of providing consistent measurements.

In taking the soil samples themselves, there was a great degree of variation in the depth of sampling, which even at its thinnest included a great deal of material that the spectrometer had no way of sensing. Additionally, due to the coarseness of the sampling method, it was necessary to take soil from an area that greatly exceeded the spectral sample field of view. This inclusion of unsensed material could have any effect at all on the final analyses, and it is impossible at this point to know how it influenced the texture analysis, texture-intensity relationship, and PLS model.

On the topic of the ParLeS modelling process, it is quite unlikely, given the robustness of the software and PLS regression process, that the resulting failure to provide a meaningful predictive model was an error on the part of the software. Instead, it was probably a result of the faulty sample taking already mentioned. Nevertheless, the resulting PLSR model was close to the cut-off criteria, and so by refining the methodology of sample-taking it seems likely that a more useful model may be developed.

Despite the nearly universal failure to achieve expected or desired outcomes in this experiment, modifying the methodology to correct for the errors discussed and tailoring the process with one outcome in mind does seem like a promising course of action.

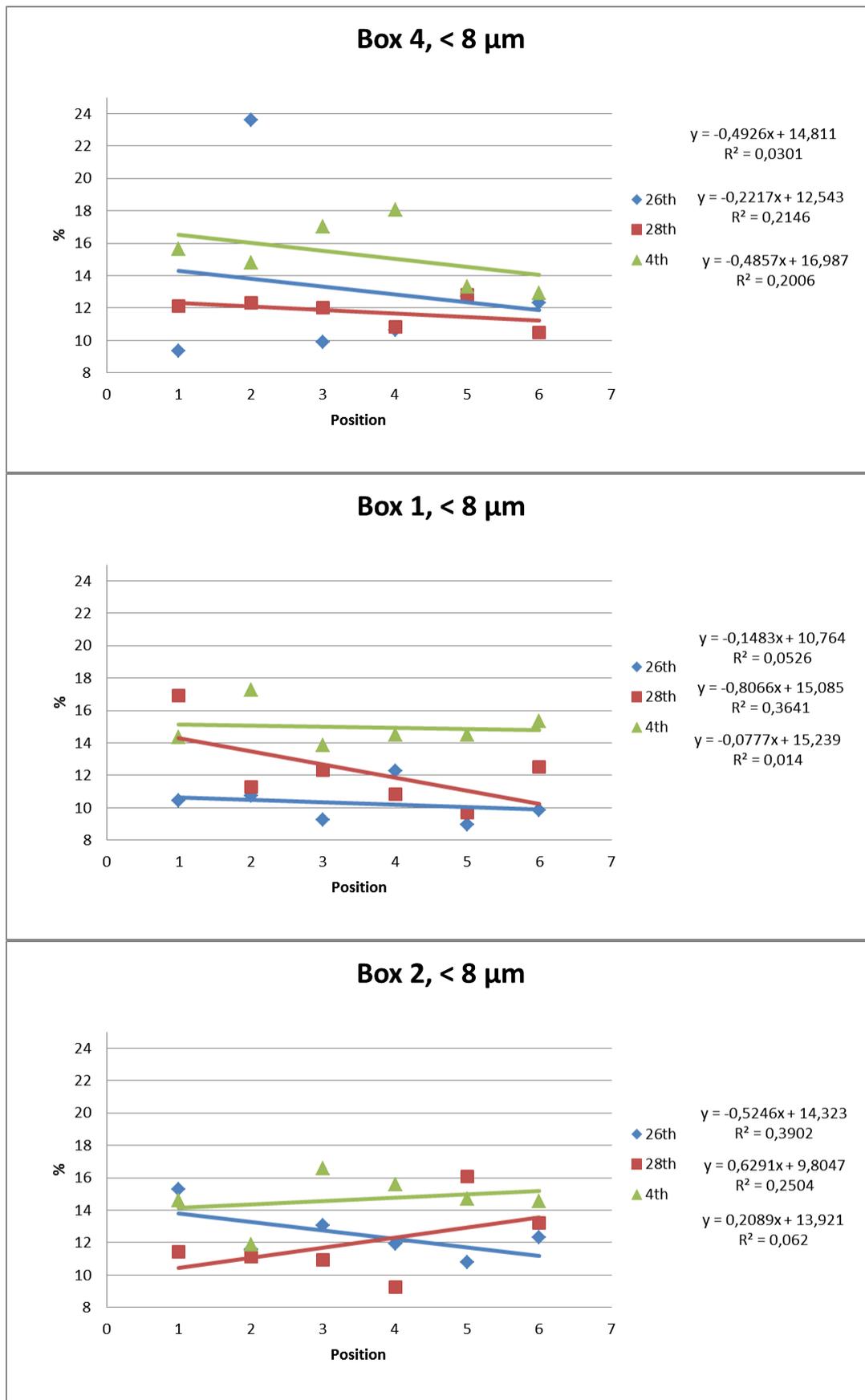
Bibliography

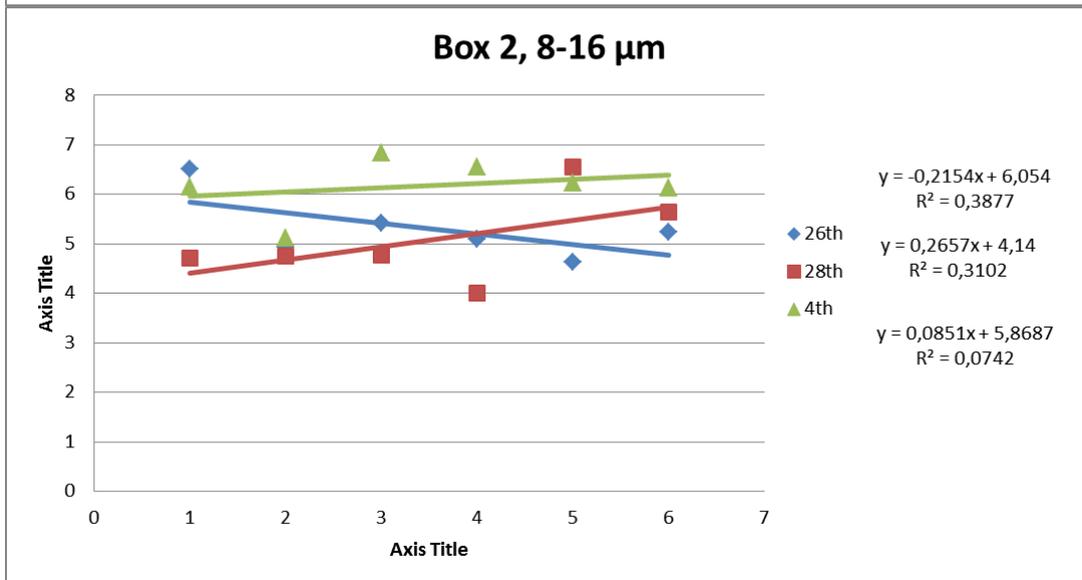
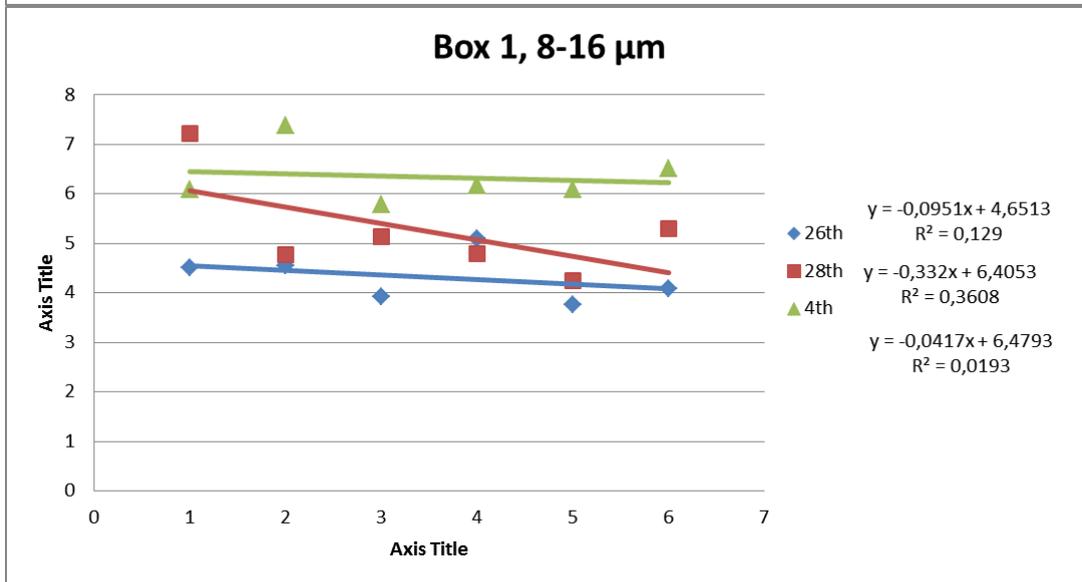
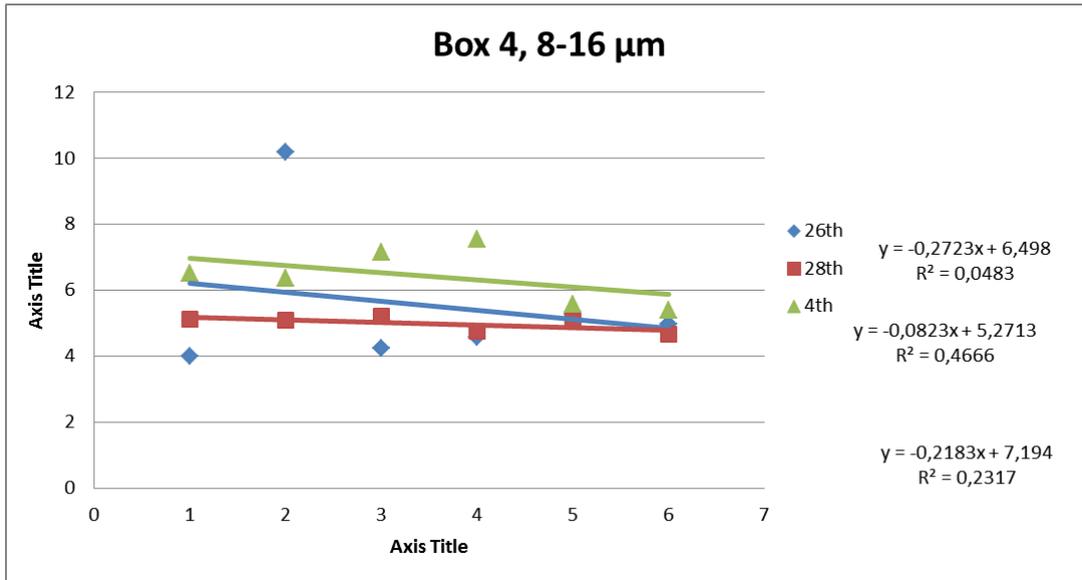
- Alberta Department of Transportation. n.d.
<http://www.transportation.alberta.ca/Content/docType372/Production/4ErosSediPrs.pdf> (accessed August 8, 2012).
- Abdi, H. "Partial Least Squares (PLS) Regression." In *Encyclopedia of Social Sciences Research Methods*, edited by M. Lewis-Beck, A. Bryman and T. Futing. Thousand Oaks (CA): Sage, 2003.
- Agassi, M., and G.J. Levy. "Stone cover and rain intensity - Effects on infiltration, erosion and water splash." *Australian Journal of Soil Research* 29, no. 4 (1991): 565-575.
- Assouline, S., and M. Ben-Hur. "Effects of rainfall intensity and slope gradient on the dynamics of interrill erosion during soil surface sealing." *CATENA* 66, no. 3 (2006): 211-220.
- Ben-Dor, E. "Quantitative Remote Sensing of Soil Properties." *Advances in Agronomy* 75 (2002): 173-243.
- Ben-Dor, E., and A. Benin. "Near-Infrared Analysis as a Rapid Method to Simultaneously Evaluate Several Soil Properties." *Soil Science Society of America Journal* 59 (1995): 364-372.
- Ben-Dor, E., et al. "Using Imaging Spectroscopy to study soil properties." *Remote Sensing of Environment* 113 (2009): 38-55.
- Benz, U.C., P. Hofmann, G. Willhauck, I. Lingenfelder, and M. Heynen. "Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information." *Journal of Photogrammetry & Remote Sensing* 58 (2004): 239-258.
- Blaschke, T. "Object based image analysis for remote sensing." *ISPRS Journal of Photogrammetry and Remote Sensing* 65, no. 1 (2010): 2-16.
- Calabrò, G., A.P. Leone, and P. Amenta. "Prediction of soil properties from reflectance spectroscopy using multivariate statistical analysis and artificial intelligence methods: a comparative study." *11th European Symposium on Statistical Methods for the Food Industry*. Benevento, 2010.
- Chang, C.W., D.A. Laird, M.J. Mausback, and R. Hurburgh. "Near-Infrared Reflectance Spectroscopy: Principle Components Regression Analysis of Soil Properties." *Soil Science Society of America Journal* 65, no. 2 (2001): 480-490.
- Chaplot, V., and Y. Le Bissonnais. "Field measurements of interrill erosion under different slopes and plot sizes." *Earth Surface Processes and Landforms* 25, no. 2 (2000): 145-153.
- Chávez, R. *Instruction* (28 March 2012).
- Chávez, R., and J.G.P.W. Clevers. "Practical Manual: Object oriented image analysis using eCognition." Wageningen University, 2012.
- Definiens AG. *Developer 7 Reference Book*. 2007.
- FAO. "Water Harvesting." *FAO Corporate Document Repository*. 1991.
<http://www.fao.org/docrep/U3160E/u3160e05.htm> (accessed August 8, 2012).
- Frauenfeld, B., and C. Truman. "Variable Rainfall Intensity Effects on Runoff and Interrill Erosion From Two Coastal Plain Ultisols in Georgia." *Soil Science* 169 (2004): 143-154.

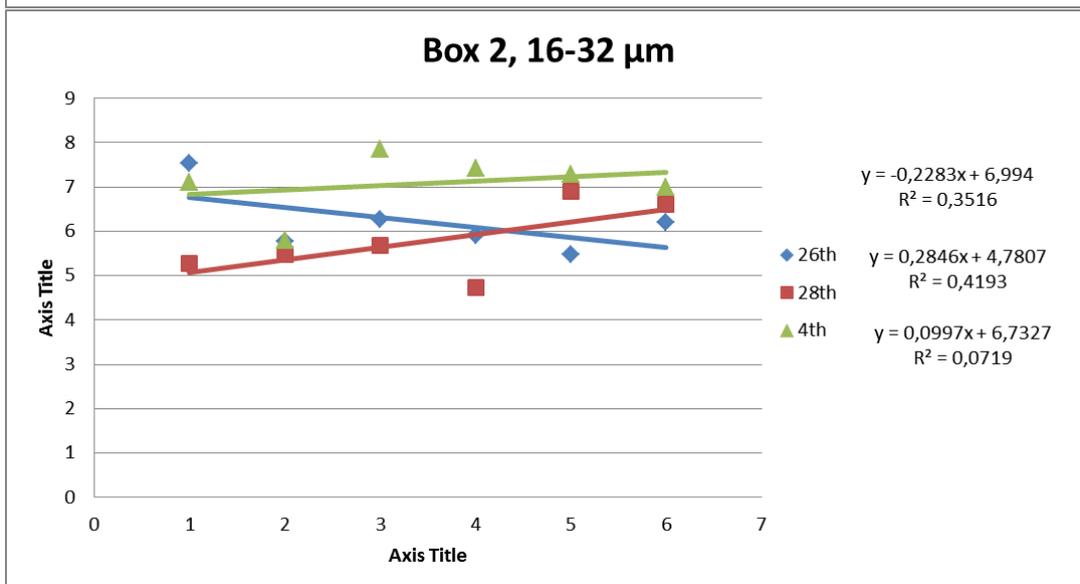
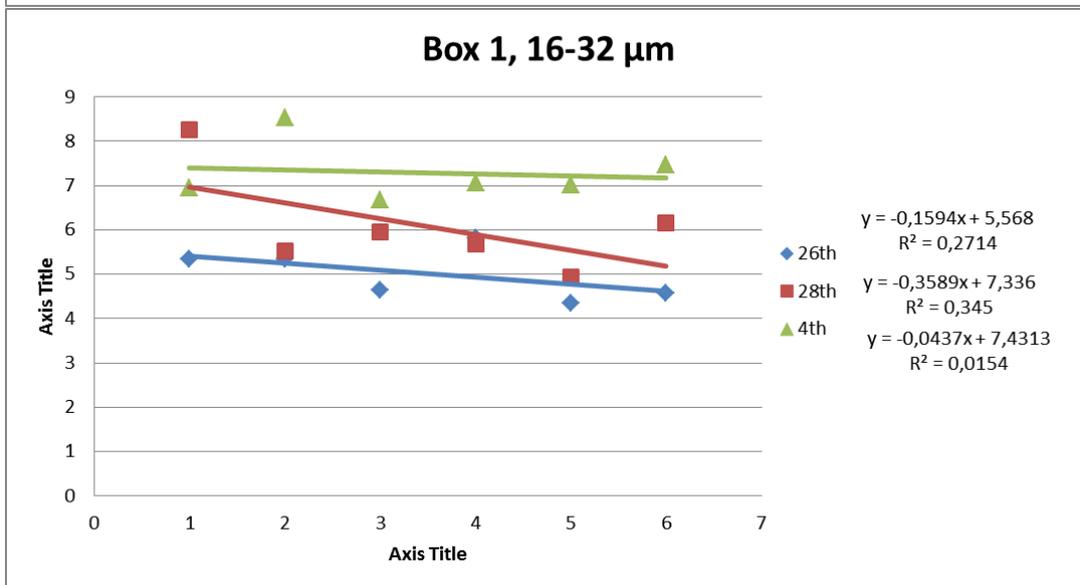
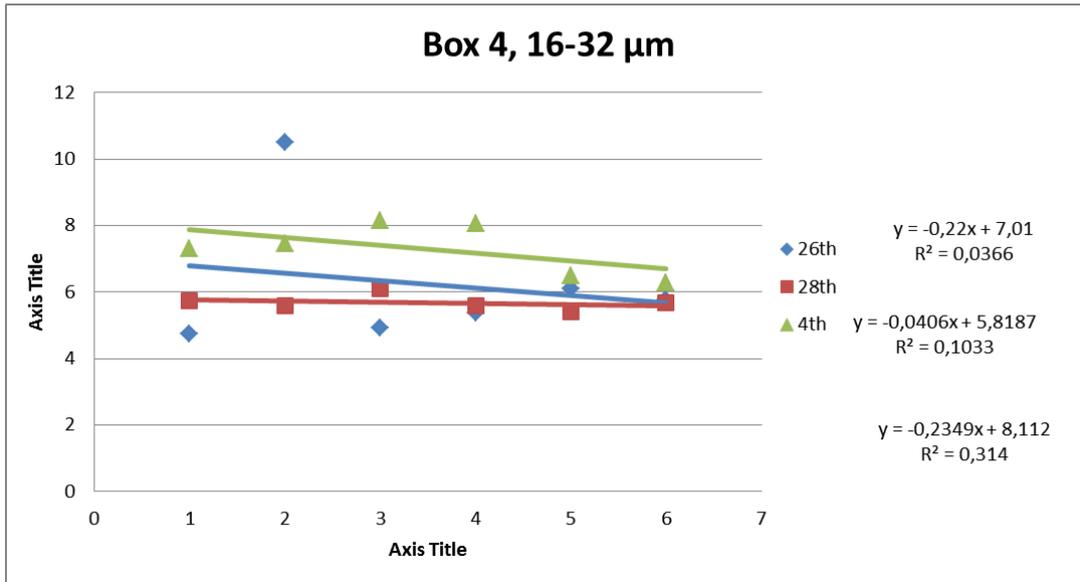
- Gabriels, D., and W.C. Moldenhauer. "Size Distribution of Eroded Material from Simulated Rainfall: Effect over a range of texture." *Purdue Journal Paper*, 1978: 954-958.
- Gibbons, C. "Understanding Multivariate Regression." 10 October 2009.
- Knight, J., J. Spencer, A. Brooks, and S. Phinn. "Large-area, high-resolution remote sensing based mapping of alluvial gully erosion in Australia's tropical rivers." *Proceedings of the 5th Australian Stream Management Conference - Australian Rivers: Making a Difference*. Thurgoona: Charles Sturt University, 2007. 199-204.
- Lado, M., M. Ben-Hur, and L. Shainberg. "Soil Wetting and Texture Effects on Aggregate Stability, Seal Formation, and Erosion." *Soil Science Society of America Journal* 68 (2004): 1992-1999.
- Lal, R. "Soil Degradation by Erosion." *Land Degradation & Development* 12 (2001): 519-539.
- Martha, T.R., N. Kerle, V. Jetten, C.J. van Westen, and K.V. Kumar. "Characterising spectral, spatial and morphometric properties of landslides for semi-automatic detection using object-oriented methods." *Geomorphology* 116, no. 1-2 (2010): 24-36.
- Merritt, E. "The Identification of Four Stages During Micro-Rill Development." *Earth Surface Processes and Landforms* 9 (1984): 493-496.
- Meyer, L.D. *How rain intensity affects interrill erosion*. ASAE Paper, Oxford, MS, USA: USDA Sedimentation Lab, 1980.
- Parsons, A.J., and P.M. Stone. "Effects of intra-storm variations in rainfall intensity on interrill runoff and erosion." *CATENA* 67, no. 1 (2006): 68-78.
- Reicher, J.M., L.D. Norton, and C.H. Huang. "Sealing, Amendment, and Rain Intensity Effects on Erosion of High-Clay Soils." *Soil Science Society of America Journal* 58, no. 4 (1994): 1199-1205.
- Rong, Z. *Estimation of foliar chemicals as indicators of tea (Camellia sinensis L) quality using reflectance spectroscopy*. Master Thesis, the Netherlands: International Institute for Geo-information Science and Earth Observation, 2009.
- Rossel, R.A.V. "ParLeS: Software for chemometric analysis of spectroscopic data." *Chemometrics and Intelligent Laboratory Systems* 90 (2008): 72-83.
- Shainberg, I., D. Warrington, and J.M. Laflen. "Soil Dispersibility, Rain Properties, and Slope Interaction in Rill Formation and Erosion." *Soil Science Society of America Journal* 56 (1992): 278-283.
- Shruthi, R.B.V., N. Kerle, and V. Jetten. "Object-based gully feature extraction using high spatial resolution imagery." *Geomorphology* 134 (2011): 260-268.
- Smith, J.R., and S.F. Chang. "Quad-Tree Segmentation for Texture-Based Image Query." *AMC 2nd International Conference on Multimedia*. San Francisco, 1994.
- Stenberg, B., R.A.V. Rossel, A.M. Mouazen, and J. Wetterlind. "Visible and Near Infrared Spectroscopy in Soil Science." *Advances in Agronomy* 107 (2010): 163-215.
- Stone, R.P., and D. Hilborn. "Universal Soil Loss Equation." *Ontario Ministry of Agriculture, Food and Rural Affairs*. May 2000.
<http://www.omafra.gov.on.ca/english/engineer/facts/00-001.htm> (accessed August 8, 2012).

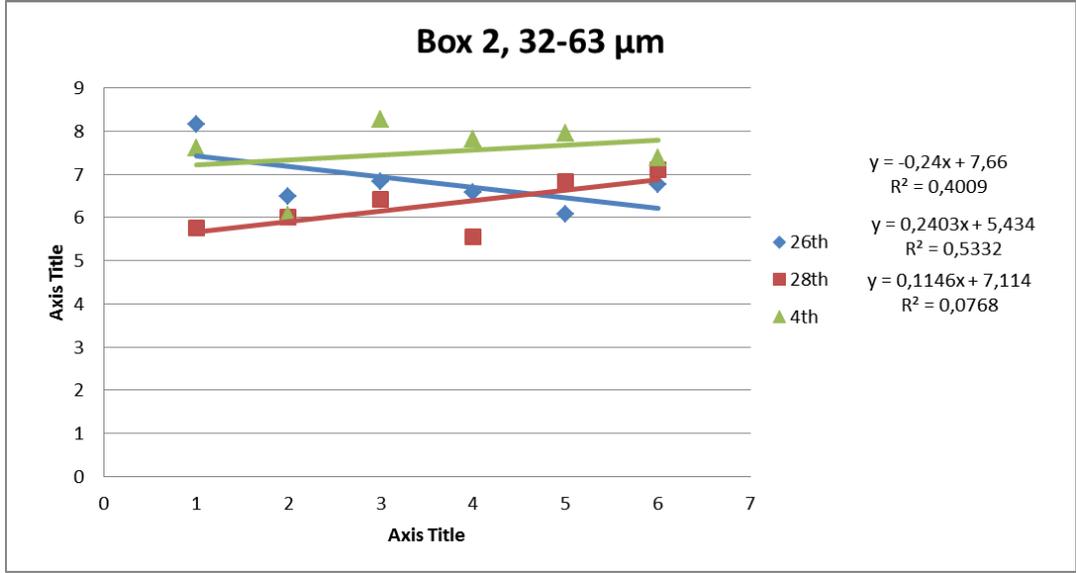
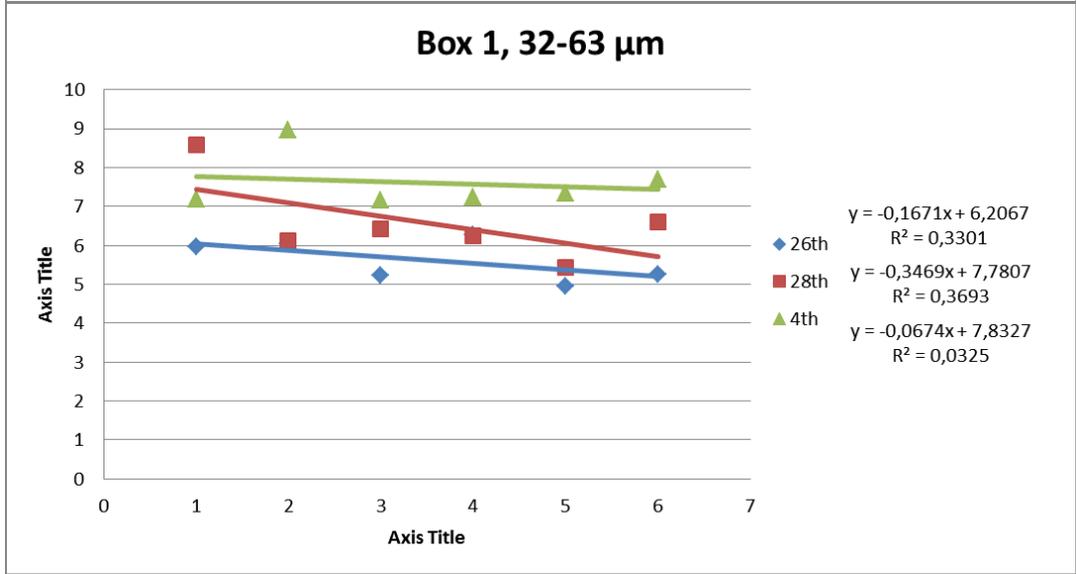
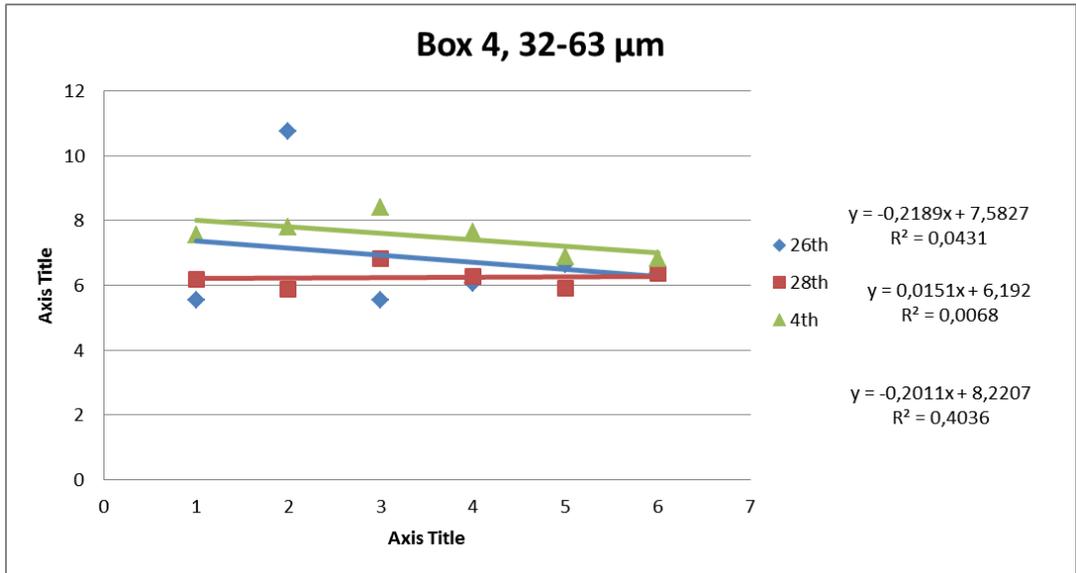
- Tallada, J.G., N. Palacios-Rojas, and P.R. Armstrong. "Prediction of maize seed attributes using a rapid single kernel near infrared instruments." *Journal of Cereal Science* 50 (2009): 381-387.
- Tobias, R.D. *An Introduction to Partial Least Squares Regression*. Cary (NC): SAS Institute, 1997.
- UBC Faculty of Land and Food Systems. *Soil Erosion*. n.d.
http://www.landfood.ubc.ca/soil200/soil_mgmt/soil_erosion.htm (accessed August 8, 2012).
- University of Toronto. *Image Segmentation*. n.d.
<http://www.cs.toronto.edu/~jepson/csc2503/segmentation.pdf> (accessed March 30, 2012).
- Victoria Department of Primary Industries. *Soil Erosion: Mechanics*. 14 November 2011.
http://vro.dpi.vic.gov.au/dpi/vro/vrosite.nsf/pages/soilhealth_mechanisms (accessed August 8, 2012).
- . *Water Erosion*. 14 November 2011.
http://vro.dpi.vic.gov.au/dpi/vro/vrosite.nsf/pages/soilhealth_water (accessed August 8, 2012).
- Vincentie, H. *Ash Unraveled: A descriptive study about physical, chemical and spectral characteristics of ash and soil in a Mediterranean forest after a large wildfire*. MSc. Thesis, Wageningen University, 2012.
- Waal Hochschule Rhein. "SMART INSPECTORS." *Projectconcept*. 30 March 2011.
- Wageningen University. "Instructions on how to run ParLeS." n.d.
- Wall, G., C.S. Baldwin, and I.J. Shelton. "Soil Erosion - Causes and Effects." *Ontario Ministry of Agriculture, Food and Rural Affairs*. 1987.
<http://www.omafra.gov.on.ca/english/engineer/facts/87-040.htm> (accessed August 8, 2012).
- Wischmeier, W.H., and D.C. Smith. *Predicting Rainfall Erosion Losses: A Guide to Conservation Planning*. US Department of Agriculture, 1978.

Annex 1: Soil texture by box and texture class

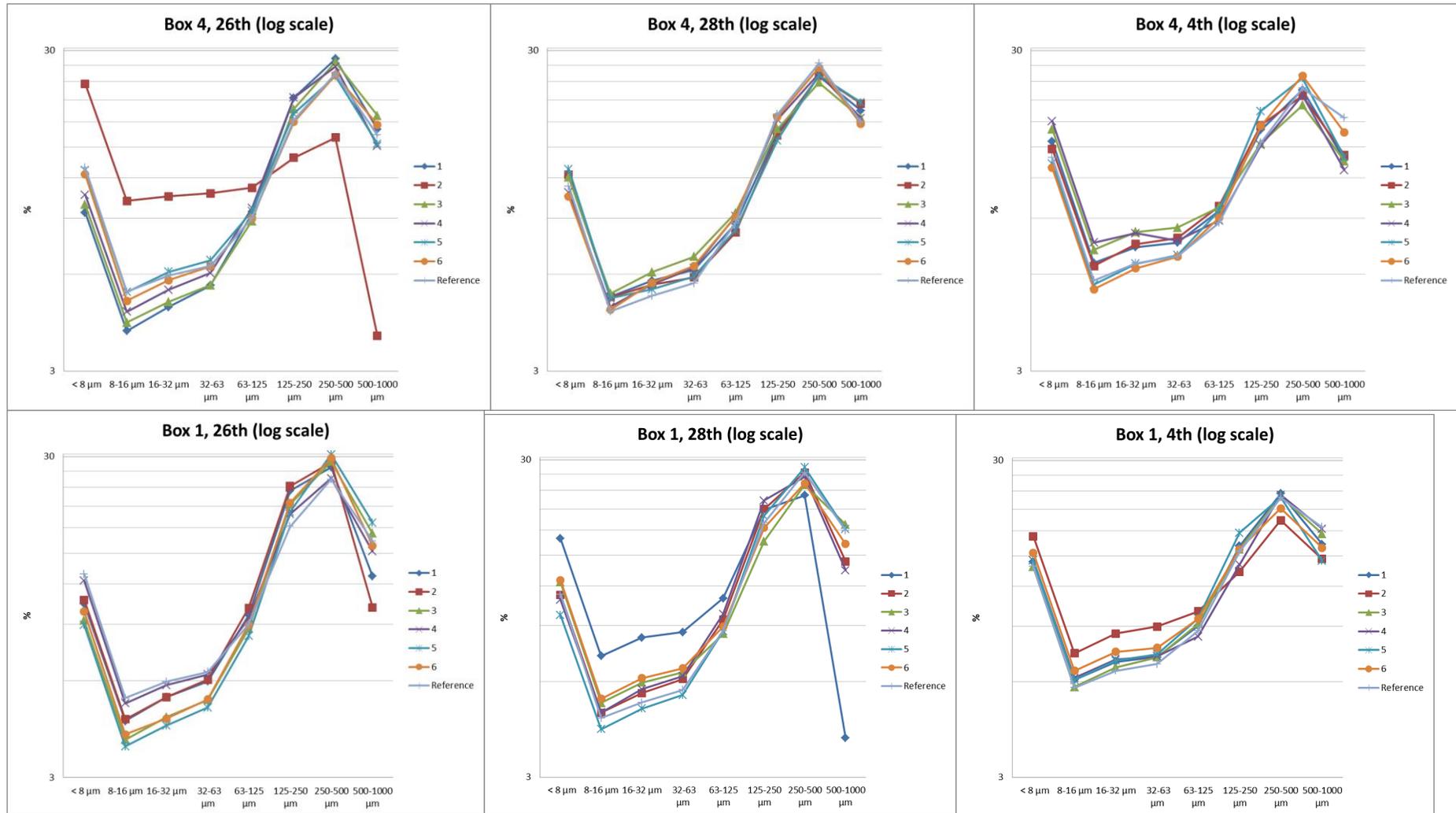


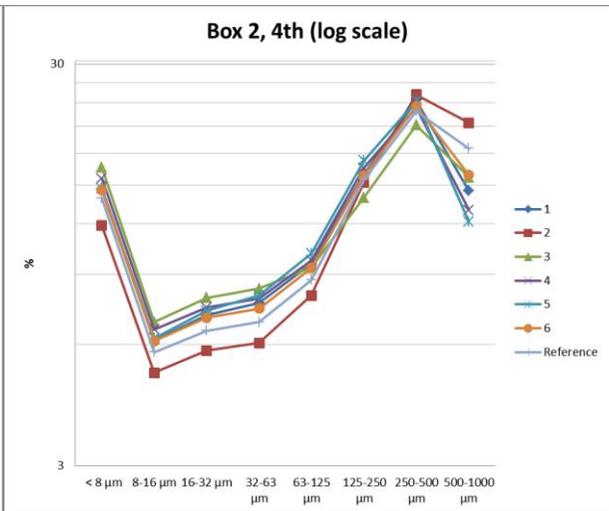
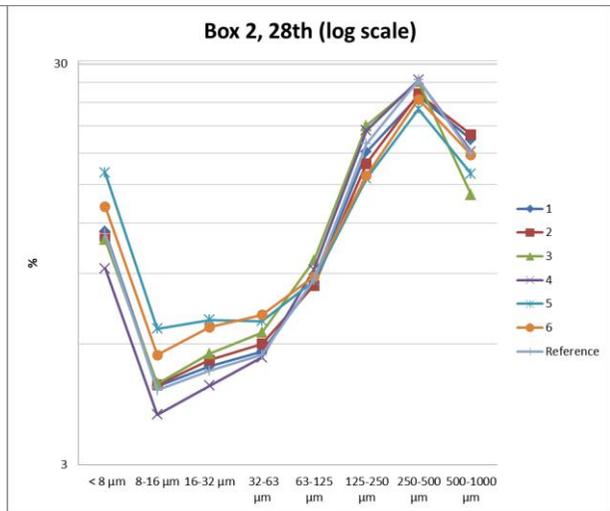
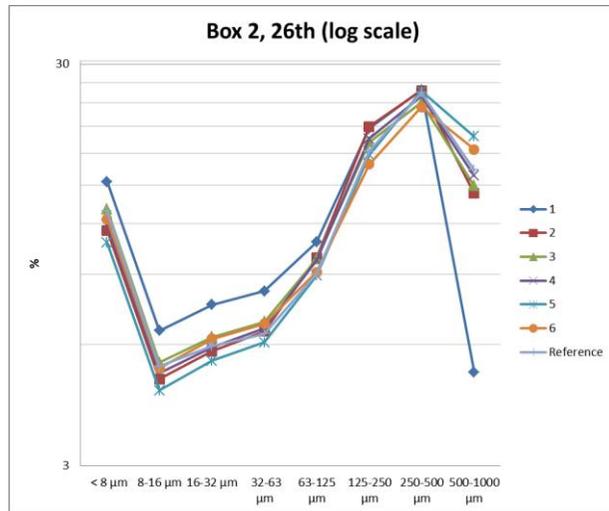






Annex 2: Soil texture by box and date





Annex 3. ParLeS PLS Regression Steps

ParLeS version 3.1

ParLeS Import Data Modelling Data Manipulations PCA PLSR Cross Validation PLSR Model Import Data Prediction PLSR Predict Bagging-PLSR

DATA FOR MODELLING IMPORT DATA FOR MODELLING

Get file for modelling

Header information - data for modelling
 Sample # Clay % 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189

Total Number of y variables

y variables

10,72	0	0	0	0	0	0	0	0
9,24	0	0	0	0	0	0	0	0
12,29	0	0	0	0	0	0	0	0
8,94	0	0	0	0	0	0	0	0
9,85	0	0	0	0	0	0	0	0
15,29	0	0	0	0	0	0	0	0

Select y variable for modelling

Labels	Selected y	X variables	X-label
a26-1-2	10,72	1	0,998962 0,998438 0,998071
a26-1-3	9,24	1	0,999521 0,999377 0,999385
a26-1-4	12,29	1 1	0,999682 0,999015
a26-1-5	8,94	1 1	0,99998 0,998977
a26-1-6	9,85	1	0,999427 0,998291 0,998801
a26-2-1	15,29	1	0,999303 0,99905 0,999286

size y size X

DATA TO MERGE

Check to merge files from a single directory

Directory with files to merge

File extension (e.g. *.txt) Size merged data SAVE MERGED FILE

Histogram y

No. Occurrence vs y

y Statistics

mean	13,05
st dev	2,91
med.	12,31
max.	23,60
min.	8,94
skew	27,48

Sample spectra

Units vs X-label

