A MECHANISTIC APPROACH TO SOIL VARIABILITY AT DIFFERENT SCALE LEVELS

A case study for the Atlantic Zone of Costa Rica

MSc thesis by Susan Klinkert

July 2014

Wageningen UR

A mechanistic approach to soil variability at different scale levels

A case study for the Atlantic Zone of Costa Rica

MEE – Earth & Environment
Soil Geography and Landscape Group

Susan Klinkert
Registration number: 920511444060

Supervisors:
Dr. J.J. Stoorvogel
C. Hendriks MSc

Date:
July 2014, Wageningen

Wageningen UR

TABLE OF CONTENTS

SUMMARY

Knowledge on soil variability is of high importance for agricultural management. Often, soil variability is studied by mapping the spatial distribution of soils. Broadly speaking, two different ways of mapping soils exist: traditional soil mapping and digital soil mapping. Traditional soil mapping uses a judgemental sampling scheme to create polygon based soil maps. Digital soil mapping uses a limited number of soil observations to derive statistical relationships between existing data on environmental co-variables and soil properties. In this way, continuous soil property maps are created. Both methods take into account the mechanisms behind soil variability to some extent; traditional soil mapping by mental models and in digital soil mapping by analysis of driving factors. In order to understand the mechanisms behind soil variability in the landscape, the soil-forming factors which drive the soil-forming processes are important to assess. These factors are generally known as climate, organisms, relief, parent material and time. However, on landscape scale, a real mechanistic, i.e. process-based, description of soil variability is very difficult. By obtaining understanding about the soil-forming factors which drive processes, a mechanistic view on soil variability in the landscape can be obtained. By defining the most important drivers of soil variability at different scale levels, more efficient and accurate soil sampling schemes can be developed. Therefore, this study focuses on determining the most important driving soil-forming factors in the study area at different scale levels by using a mechanistic view onto the landscape. The Atlantic Zone of Costa Rica is used as a case study.

When assessing soil variability, it is important to take into account the different spatial scale levels at which soil-forming factors work. This is done by a conceptual model which shows the relation between spatial scale levels and the soil-forming factors. At all scale levels, all factors will cause soil variability to some extent. However, at each scale level different factors will be most important in driving the processes which cause soil variability.

The study area where soil variability is studied is characterised by tropical conditions, which means tropical temperatures are present all year round and an average rainfall amount of 3058 mm/y is present, although high variation in climatic conditions exist. In general soils are from volcanic origin or developed in young fluvial deposits. To determine which factors are driving soil variability in the study area, two methods are used. First an analysis of legacy and auxiliary data was carried out. This means, the soil map of the Northern Atlantic Zone of Costa Rica was overlain with auxiliary data on soil-forming factors and elevation. In this way, variation in auxiliary data within and between mapping units was assessed. High variation in auxiliary data between mapping units points to high importance in driving soil variability at the regional scale.

Besides this, a multi-scale analysis of soil variability was done by carrying out a fieldwork campaign. Eight transects of five kilometres were sampled. At ever transect, six locations were studied with a spacing of one kilometre in between. At every location, short distance variability was examined. In this way, three different scale levels could be assessed: regional, local and point scale. By linking variability in soil properties to variability in soil-forming factors, a mechanistic view onto the

soil variability could be established. The data was analysed by calculating statistical summaries.

The analysis of legacy and auxiliary data demonstrated the importance of climatic factors on soil variability at the regional scale. The percentage of total variation in temperature explained by the mapping units is 88.3% and in rainfall 69.1%. This means processes like heat, water and solute flow which are induced by respectively temperature and rainfall are most important in causing soil variability on this scale level. The percentage of total variation in elevation explained by the mapping units is 89.0%, however elevations was almost one to one correlated with temperature (r=-0.99) and elevation in itself is not a soil-forming factor.

The multi-scale analysis showed the high variability in the study area. Whereas on regional level climatic conditions are highly important, on local level parent material was very important. On point level, the importance of land use became visible. From regional to local scale, variation in organic matter content and pH of the topsoil was increasing, although only slightly. This points to the importance of climate as an important factor driving the processes which cause variability in the study area, which confirms the result of the analysis of the auxiliary and legacy data.

The question which of the factors is most important in describing soil variability cannot be answered with a general statement based on the conceptual model alone. This has to be determined for every specific case, since every area will have its specific environmental conditions. Both methods examined here work complementary in determining important drivers of soil variability. The analysis of available data is a quick and cheap method, using the advantage of digital soil mapping that drivers of soil variability could be quantified. The multi-scale assessment in the field is much more expensive and time-consuming. However, by using the latter method, soil properties could be taken into account and different scale levels could be studied by usage of expert knowledge, as is done in traditional soil mapping as well.

Concluding it can be stated that a combination of analysis of legacy and auxiliary data together with field observations provides a solid basis for more efficient and well-founded studies on soil variability.

1. INTRODUCTION

1.1 General introduction

Knowledge on soil variability is of high importance for agricultural management. With growing interest in the landscape perspective to address diverse environmental, ecological and agricultural issues, an adequate understanding of soil variability as a function of space and time becomes essential (Lin et al., 2005).

When describing soil variability in a mechanistic way, a genetic approach is used in assessing soil variability (Minasny & McBratney, 2006). This means, the soil-forming processes which shape the current soils are taken into account when soil variability is described. Since soils and their properties are not static in space and time and are influenced by the soil-forming processes, a mechanistic view onto soils is a highly preferred method to understand the occurrence of different soils in the landscape.

A common way to describe soil variability is by mapping soils and their properties. In this thesis two ways of mapping soils are highlighted: the mapping of soils in a traditional way and secondly by means of digital soil mapping.

For a long time, an often used method of assessing soil variability was by traditional soil mapping. This means, first aerial photographs of the study area are examined and expert knowledge is used to get a general view on soil variability in the area. The subsequent soil survey is done based on judgmental sampling. This means, conclusions are drawn based on professional judgement and not on statistical theory. The factors and processes which are influencing soil variability are taken into account in the mental model used by the soil surveyor. In this way, mechanisms driving soil variability are taken into account, although not quantitatively. Based on the field observations on representative sites, soil maps are drawn consisting of discrete polygons. To define borders between polygons, a large amount of field observations is needed.

In recent times, digital soil mapping has become a common way of mapping soil variability. In this technique, the study area is stratified based on legacy and auxiliary data. After stratifying the area, a probability-based sampling design is used. Based on a limited number of soil observations, statistical relationships are derived between soil properties and auxiliary data. In this way, continuous soil property maps are created. Mechanisms driving soil variability can be taken into account via the usage of soil-forming factors as auxiliary data. However, derived relations are determined statistically without taking into account expert knowledge on spatial soil variability in the landscape.

To be able to describe soil variability based on a mechanistic approach, knowledge on the driving soil-forming factors is important. Quantification of soil-forming processes on a landscape scale is very difficult. Most mechanistic models of soil formation developed up to now focus on the horizon or pedon scale level (Minasny & McBratney, 2001). Therefore, to understand the landscape from a mechanistic point of view, the driving soil-forming factors are studied in this research. In this way, the causes of soil variability can be determined and a mechanistic view onto the landscape can be established. Besides this, the scale levels at which these factors and processes work are important to take into consideration. The amount of observed variability in soils is highly dependent on the scale level which is taken into account, both on the temporal as well as the spatial scale level (Burrough, 1983). For land use planning and agricultural management, knowledge on variability at different scale levels is important. At farm scale different management strategies are important than on for example the scale of a complete region based on differences in soil variability.

To study soil variability in this mechanistic, scale-dependent way, an inventory of legacy and auxiliary data of the study area is done and secondly, a multi-scale analysis is carried out to describe soil variability.

This context results in the following main research question:

How to determine which soil-forming factors are most important in describing soil variability at different scales?

In this research, the Atlantic Zone of Costa Rica is used as a case study.

In the remainder of this chapter more background information on soil mapping and soil variability will be given. Chapter 2 describes the conceptual model used in this study. More information is given about the soil-forming factors and processes and the linkage with the different scale levels at which they act. In chapter 3 the materials and methods used in this research are explained and a description of the study area will be given. Furthermore, a description of the two different methods which are used is given. In chapter 4, the results of these two methods will be explained and discussed. Chapter 5 gives a general discussion of the conceptual model and the methods. Chapter 6 gives the final conclusion of the study.

1.2 Modelling of spatial soil variability

1.2.1 Describing soil variability traditionally

At the end of the 19th century, Dokuchaev was the first to recognize the soil as a function of the interaction between climate, organisms, relief and parent material, working over time. He considered soil-forming factors as the cause of soil formation and the soil properties as their effects (Bockheim et al., 2005).

In 1941, Hans Jenny formalized these ideas and he described the soil based on its factors of soil formation. Jenny defined the soil as a system determined by properties that are interrelated. As is proposed by Jenny, this interrelation of soil properties will cause that if a sufficient number of them is fixed, all other properties are fixed too (Jenny, 1941). He described the soil (S) as a function of five independent variables that define the soil system: climate (cl), organisms (o), topography (r), parent material (p) and time (t), which are known as the soil-forming factors.

He summarized this in this first mechanistic model for soil development as (Jenny, 1941):

$$S = f(cl, o, r, p, t, \dots)$$

This model became the basis of describing soil variability. Traditional mapping of soils does take into account the importance of these factors. Polygon-based soil maps are drawn and it is assumed that the polygons are more or less homogeneous and with discrete borders separating the polygons. Although soil classification in the field is based on soil properties, classification systems like the USDA Soil Taxonomy (Soil Survey Staff, 1999) are founded onto the processes which cause the development of certain soil types. Hence, the soil variability is described in a mechanistic way in traditional soil survey to some extent, being it mainly via mental models (Bui, 2004).

When mapping soils in a traditional way, substantial time and money is needed in carrying out the study (Domburg et al., 1997). By a preliminary assessment of the importance of certain soil-forming factors based on legacy data, sampling can be done more efficiently. This can be done by assessment of soil variability at multiple scales in the field as well. This can be a good strategy to get a general idea of variability in the study area, before carrying out more detailed fieldwork.

1.2.2 Digital soil mapping

The model of Jenny forms the basis for most of the soil modelling in recent times. Based on the model of Jenny, McBratney et al. (2003) developed the *scorpan* model for use in digital soil mapping applications (Gray et al., 2009). The *scorpan* model can be written as:

$$S_c = f(s, c, o, r, p, a, n)$$

in which S_c= soil class, s= soil, c= climate, o= organisms, r= topography, p= parent material, a= age and n= space. Soil as a factor was included because McBratney et al. state that soil can be predicted from its properties. It refers to soil information from legacy data, remote sensing or from expert knowledge. Here the factors *s* (a soil attribute predictor) and *n* (a geographic position predictor) are added to the traditional model described by Jenny (Gray et al., 2009).

Recently, models using quantitative techniques derived from geostatistics for spatial prediction of soil properties are developing quickly (McBratney et al., 2000). This is a result of new developments in this field as well as a global need for soil data and information for environmental monitoring and modelling (McBratney et al., 2003). This means, pedometric techniques are increasingly used in mapping the soil and her properties. Pedometric techniques use quantitative, mathematical and statistical methods in assessing pedology (McBratney et al., 2000). Statistical relationships between soil and key environmental factors form the basis for digital soil mapping techniques (Gray et al., 2009). Regression relationships based on auxiliary data from geophysical and satellite image sources are used to represent the state factors described (Gray et al., 2009). The soil-forming factors are taken into account as auxiliary data and thus driving factors behind the mechanisms are assessed. However, relationships with soil properties are predominantly determined statistically without taking into account knowledge onto spatial soil variability in the landscape.

Besides this, it is important to notice that the accuracy of the auxiliary often is low compared to data acquired by laboratory and field measurements due to sensor noise, directional reflectance and topographic and atmospheric distortions (Mulder et al., 2011). Furthermore, spatial resolution of satellites is often too coarse to be able to describe small scale processes. For these reasons, field observations of soil properties keep being important for analysing soils and their characteristics.

1.2.3 Mechanistic description of soil variability

Some research is available on soil modelling based on soil formation. For example, Minasny & McBratney (2001) developed a mechanistic model for soil formation considering soil formation in a landscape at the catena scale. In recent times however, little effort has been put into the quantitative science of soil formation (Minasny & McBratney, 2006). As stated by Minasny and McBratney (2006), a mechanistic model for soil formation is still hardly developed yet. As proposed by Bockheim and Gennadiyev (2000), nowadays most soil analysis is based increasingly onto soil properties instead of soil-forming processes. As stated by them, a consideration of soil-forming processes is important to understand the genetic underpinnings of soil taxonomic systems and also for the development of pedogenetic systems. Pedogenetic systems study the history of the soil and use this knowledge to classify the soil in different types, taking into account the relationships among soil taxa (Minasny & McBratney, 2006). Furthermore, by thinking in a mechanistic way, a more realistic description of soil variability can be obtained. As stated by Hartemink et al. (2008), knowledge-driven approaches based on qualitative models of pedogenesis are expected to provide more robust soil predictions that the usual data-driven approaches.

More efficient soil mapping can be carried out when only the most important soil-forming factors which drive these soil-forming processes and thus soil variability are taken into account. By including existing information on soil variability, efficiency of soil sampling can be increased (Domburg et al., 1997). For example this can be obtained by an inventory of legacy and auxiliary data and the usage of expert knowledge about the study area.

The quantification of soil variability at multiple scales is often desirable for mechanistic modelling and prediction, because this provides a basis for developing a better understanding of scales of influence on variability and besides this, it gives a framework upon which scaling of data may be possible (Lin et al., 2005). In chapter 2, the relation between soil-forming factors and spatial scale levels will be elaborated more extensively.

2. CONCEPTUAL MODEL

2.1 Scale levels

Nowadays the soil formation model of Jenny (1941) is widely used as a basis for studies in soil variability. However, when determining statistical relationships between soil-forming factors and environmental co-variables in digital soil mapping, interactions between soil-forming factors are often not taken into account. Besides this, different importance of soil-forming factors at different scale levels is often not acknowledged. As stated by Burrough in 1983: *"the often quoted statement that soil is the complex product of the interaction of parent materials, climate, hydrology, relief and biological activity acting over time, glosses over the point that each of these soil factors may act over a different spatial scale, and that within each soil-forming factor there can be many spatial scales of interaction"* (Burrough, 1983).

Scale levels are an important aspect of describing variability in general and for soils in specific. At every scale level in soils, pattern structures, i.e. spatial correlations, have been found (Burrough, 1983). The study of Yemefack et al. (2005) is an example of a study of soil variability at different scales. They acknowledge that variability at certain scales may be substantially larger than on other scales and spatial dependence between data points may exist.

In mapping, the scale trilogy is important to take in consideration when assessing soil variability. The scale trilogy comprises out of extent, resolution and support. Changing one of these properties of a map, will lead to a different view of soil variability. McBratney et al. (2000) distinguished three general scale levels at which soil information is presented; each linked to a typical resolution and extent.

2.2 Soil-forming processes

Soil variability is induced by variability in soil-forming processes causing differences in soil formation. If an area would be perfectly uniform in soil-forming factors and processes, than no soil variability would be found. Interacting soil-forming processes at different spatial scales will give rise to the observed soil patterns (Burrough, 1983). Thus, the occurrence of certain soil properties is driven by the soil-forming processes. These are in turn driven by the soil-forming factors as described in section 1.2.1. Soil-forming processes provoke distinction between different soil types by formation of diagnostic horizons, properties and materials (Bockheim & Gennadiyev, 2000). In figure 1 a graphical representation of these relations is given.

Figure 1 Relation between soil-forming factors, processes and soil properties.

2.2.1 Groups of soil-forming processes and the scale level on which they act

Many different soil-forming processes which are occurring at a wide range of spatial and temporal scales are driving soil variability. In general, the soil-forming processes can be divided into four broad groups: transformations, translocations, additions and losses (Simonson, 1959), as is shown in figure 2. Within each of these classes, numerous different soil-forming processes can be observed.

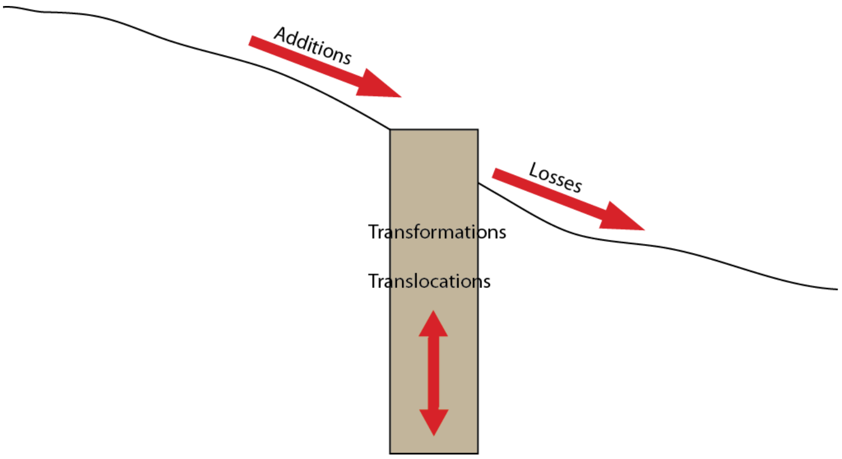

Figure 2 The four main groups of soil-forming processes influencing the soil pedon.

Transformations occur when soil constituents are chemically or physically modified or destroyed and others are synthesized from parent materials (Brady & Weil, 2004). Transformation processes comprise chemical and physical weathering processes, decomposition of organic residues, as well as for example aggregation of mineral particles. Other examples of transformations are gleization and the weathering of parent material. These processes occur at microscopic scale and thus are inducing soil variability at the very small scale within the soil pedon.

Translocations involve the movement of inorganic and organic material laterally within a horizon or vertically from one horizon up or down to another (Brady & Weil, 2004). Examples of such processes are leaching of bases and clay illuviation. These processes occur within the profile and occur thus at a somewhat larger scale level than the transformations.

Additions comprise inputs of materials to the soil profile from external sources (Brady & Weil, 2004). Examples of additions present in most soil systems are the addition of organic matter from plant materials and the addition of fresh materials by sedimentation. These processes occur at larger scales again than the transformations and translocations since these processes work in a lateral direction.

The fourth large group of soil-forming processes are the losses. Losses from the soil profile, just like additions, can be observed both in lateral direction and in downward direction. An important process altering soil development in sloping terrain in lateral direction is erosion of surface materials. Erosion often removes the finer particles, leaving the surface horizon relatively sandier and less rich in organic matter than before (Brady & Weil, 2004). Other important processes causing losses in dissolved salts, silica and organic acids from the soil profile are drainage and leaching of the soil profile. Hence, these processes occur at different scales, i.e. in-profile but also on the hill slope level.

2.2.2 Soil-forming processes and mapping soil variability

In theory, the scales of variation in soil properties are determined by the combined action of environmental processes (Holmes et al., 2005). The measurement of rates of soil-forming processes is in practice however very difficult. Most processes act on large time scales and are affected by a large range of variables.

For this reason, in digital soil mapping often the soil-forming processes are treated as a black box and are estimated by the soil-forming factors which are readily measurable. See figure 3 for a graphical representation of the relation between soil-forming factors, processes and properties and the different mapping strategies.

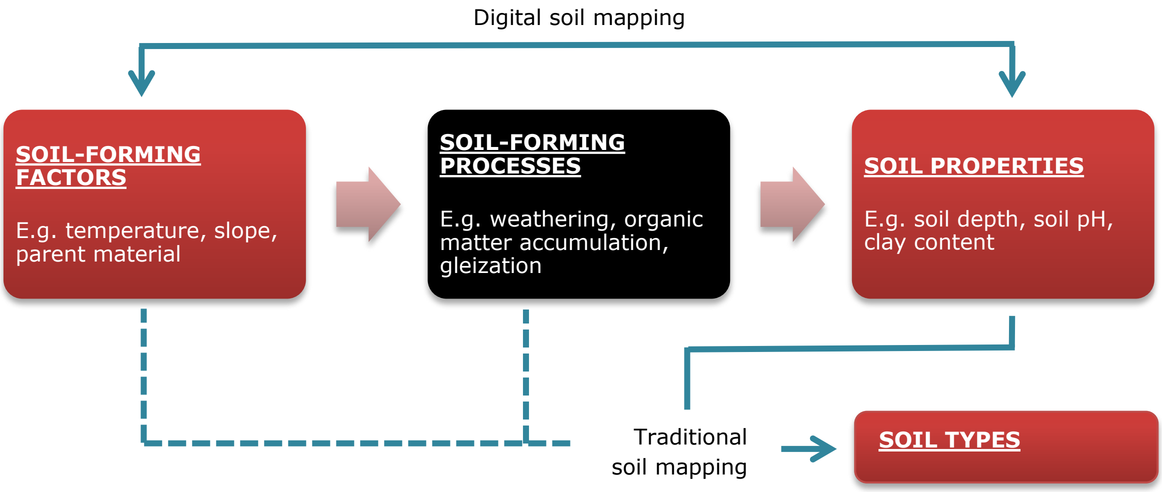

Figure 3 In different mapping strategies, soil-forming processes are treated as a black box. Digital soil mapping relates the soil-forming factors to soil properties. Traditional soil mapping determines soil types based on soil properties using mental models on soil-forming factors and processes.

As said, in digital soil mapping, the soil properties are determined based on information on soil-forming factors. First, auxiliary data on soil-forming factors are correlated to field observations and by usage of geostatistics, predictions for the whole study area are done. In traditional soil mapping in contrast, soil types are determined based on soil properties observed in the field by including expert knowledge on factors and processes in the study area. Both methods do not take in account soil-forming processes directly. Digital soil mapping takes into account the factors driving the processes, but based on statistical relationships instead of pedological knowledge. Traditional soil mapping takes into account the mechanisms via the soil classification system used, but in a more qualitative way. In this study, it is attempted to looked upon soil variability mechanistically, thus from the perspective of the soil-forming processes. However, as already stated in the introduction, the soil-forming factors are used to get insight in the soil-forming processes and thus variability in soil properties. As explained in section 2.2.1, soil-forming processes work at different scales, which are directly related to the different scales at which soil-forming factors are working.

2.3 Model set-up

The concept that different soil-forming factors play a role at different scale levels gives rise to the development of the model as shown in figure 4. Most factors present in the *scorpan* model are taken into account. The factors age and space are not taken into account. Here, only spatial variability is assessed, temporal variability is not considered. Besides this, the factor space is obviously linearly related to the scale level. Several examples of the soil-forming factors working at different spatial scale levels are presented.

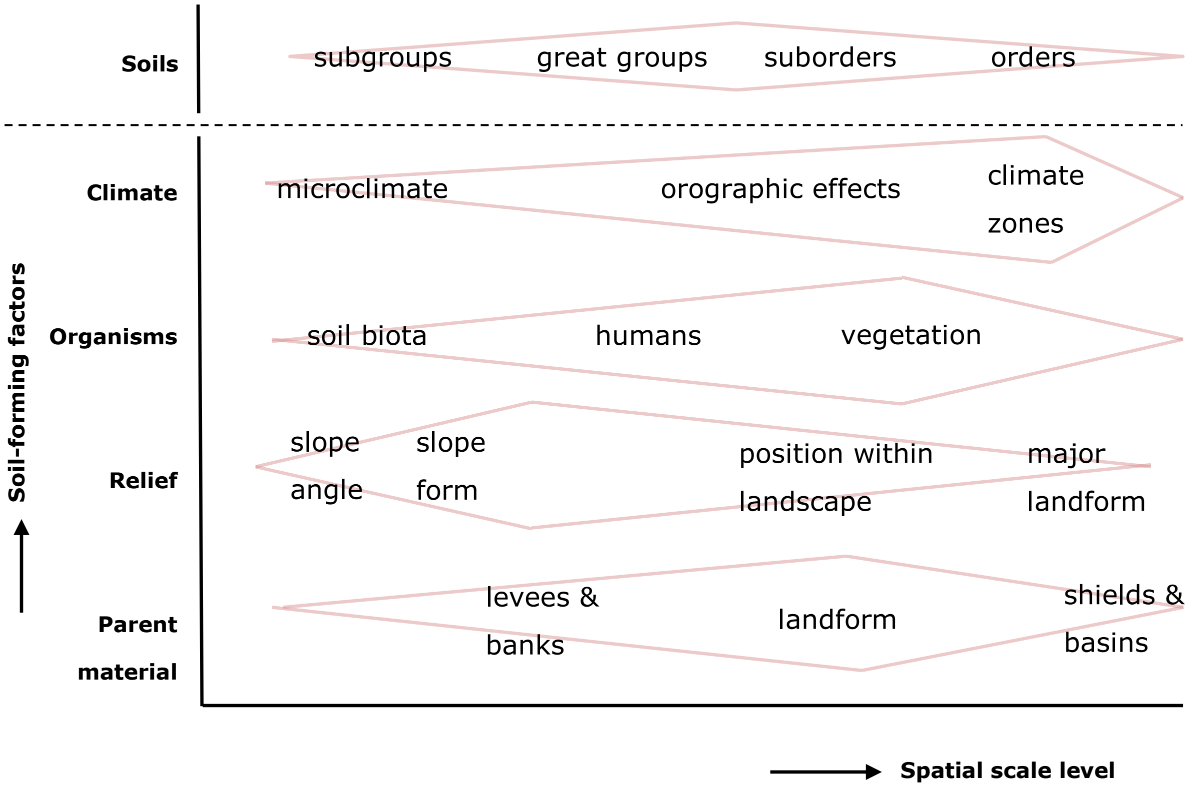

Figure 4 Conceptual model linking spatial scale levels to the soil-forming factors. Examples are shown for every of the four soil-forming factors at different scale levels. Shapes in red show at which spatial scale level the soil-forming factors are generally acknowledged to have major impact. Variation in soil-forming factors leads to soil variability at all scale levels.

2.3.1 Climate

Generally speaking, climate influences soil variability at large scales. Based on differences in climatic conditions, zonal soils can be found. These soils are mainly formed by processes influenced by climatic conditions and occur on large scales related to climatic variability. However, small scale variability in relief will give rise to small variations in climate, such as soil temperature, which will affect soil properties. For example, differences in aspect will cause differences in soil temperature thus variability in soil properties on a small scale. This is an example of the correlation between the different soil-forming factors as well as also was stated by the quote of Burrough (1983) in section 2.1.

2.3.2 Organisms

Vegetation is driving soil variability by soil-plant interactions for example by organic matter accumulation. The main soil-forming organisms are humans, although on small scale other organisms may have a significant impact as well (Hole, 1981; in McBratney et al., 2003). Humans are affecting soil-forming processes by the way soils are managed. Within land use, different scale levels can be determined: from agro-ecological zone to farm level to field level. At each scale level, soil formation will be influenced in a different way. On large scale, differences between types of land use will cause differences in soil properties. At the other hand, within one agricultural field, different management practices will cause soil variability on a small scale.

2.3.3 Relief

On a relatively small scale compared to climate and organisms, relief affects soil variability, mainly by slope processes. However, as done by the FAO, relief can be subdivided into four levels (Jahn et al., 2006). At the highest scale level major landforms are distinguished; which means in mountains different processes will occur than in for example flood plains. Within a landform, different slope positions will induce different rates of soil-forming processes and thus different soil properties. At a certain slope, slope angles varying at the very small spatial scales may induce small variation in soil properties.

2.3.4 Parent material

On large scale, large complexes of parent material can be defined. For example, on continental scale, the African shields can be defined as being one parent material. Zooming in, many different parent materials within these large geological units can be distinguished. For example, on small scale, along rivers different parent materials on levees and banks can be found at short distance.

2.3.5 Effect of soil-forming factors and scale levels on soil variability

Soils are highly variable over different spatial scales and its variability is driven by the factors described above. The USDA Soil Taxonomy classification system (Soil Survey Staff, 1999) classifies soils on different levels, which are: orders, suborders, great groups, subgroups, families and series. These levels include more details about the soil at every level of classification. Broadly speaking, it might be expected that as spatial extent or spatial resolution increases, the magnitude of soil variability will increase, reaching a possible maximum and then starting to stabilize or decrease as space scales continue to increase. However, the magnitude of such changes depend on where the soil is located in the landscape and which soil type or specific soil property is of concern (Lin et al., 2005).

On first sight, every soil-forming factor is expected to drive processes at a certain scale level. For example, climate is thought to drive soil-forming processes at large scales. However, as shown above, every factor drives processes at every scale level. With including knowledge about the study area and being aware of different scale levels, it is expected that a more efficient and accurate description of soil variability can be achieved.

3. MATERIALS & METHODS

3.1 The study area

The study area is located in the eastern part of Costa Rica, between 9.5^0 and 11^0 N and 82.5^0 and 84^0 W. In figure 5 the delineation of the study area can be found. The major part of the study area is located in the Limón province, a small part is part of the Heredia province. At the east, the area is bordered by the Caribbean Sea, at the west by the Central Cordillera, which divides Costa Rica in an Atlantic Zone and a Pacific Zone. The area is characterized by a humid, tropical climate with an udic soil moisture regime. This means, rainfall is evenly distributed throughout the year and in general, the soils never will be dried out. Mean yearly rainfall in the area is 4100 mm, although rainfall amount is highly variable in the area. Near Sixaola in the southeast of the area, average rainfall is 3058 mm/y, while in the north eastern part the average rainfall is more than 5000 mm/y (Arias, et al., 2010a). Mean yearly temperature in the lowlands is 26 ^{0}C, with a fluctuation of 2^{0}C between the warmest and coolest months (Nieuwenhuyse, 1996). Elevation ranges from sea level near the coast up to more than 2500 metres in the Central Cordillera in the western part of the study area (Bouman et al., 2000).

Historically, the area was entirely covered with tropical rain forest. However, the area has been deforested and nowadays the most important land use in the area has become pasture. Still, secondary and primary forest are covering a substantial part of the study area. However, since the construction of a railway at the end of the 19th century, the region has become an important agricultural area. Nowadays, anthropogenic influence on soils is large and the zone is the main banana production region of Costa Rica. From an economic perspective, intensive banana production systems are the most important land use type. On poorer soils, large pineapple plantations can be found. Higher up in the mountains, main land use is forest and pasture, although some coffee plantations with banana as shadow crop and small-scale agriculture are present.

Figure 5 Map of Costa Rica including the location of the study area in red. In blue, the Rio Reventazón is indicated, which traditionally divides the soils in the study area in soils developed in volcanic parent materials at the northwest and soils developed in calcareous parent materials southeast of it. Mountain ridges present in the study area are indicated in white (source: Google Earth).

3.1.1 Geology and geomorphology

Costa Rica is part of the Chorotega oceanic block, which has been the site of a volcanic arc since the Late Cretaceous (Coates & Obando, 1996). Nowadays, this volcanic arc is recognised as the Central Cordillera. There are three main tectonic mechanisms which controlled the development of the Chorotega block and the adjacent Choco block. These three mechanisms are the formation of the volcanic arc, the subduction of the Cocos Ridge and the collision of the arc with South America (Coates & Obando, 1996).

Strong subduction of the Cocos plate under the Caribbean plate in the Cenozoïcum caused the formation of an active volcanic arc along the western part of the Chorotega block with well-developed fore- and back-arc basins (Dengo, 1962; Lonsdagle and Klitgord, 1978; Case and Holcombe, 1980; Escalante, 1990; Astorga et al., 1991; as cited in Coates & Obando, 1996). These geological features are related to subduction zones and are lower lying areas at both sides of a volcanic ridge as a result of the subduction.

The second main tectonic mechanism is the subduction of the Cocos Ridge under the Chorotega block. The Cocos Ridge is an oceanic ridge which is part of the Cocos plate. This had a strong effect on the southern half of the geology of Costa Rica. One of these effects is the formation of the Cordillera de Talamanca, an elevated magmatic arc.

The collision of the eastern end of the Central American arc with the South American Plate is the third main tectonic mechanism which has controlled the development of the Chorotega block.

As a result of this tectonics, high variability in geomorphology can be found in the study area. In the Atlantic Zone three major geographic units can be distinguished: the Caribbean lowlands bordering the Caribbean Sea which consist of alluvial plains, the northern slopes of the active stratovolcanoes which are part of the Central Cordillera in the south, and the northern part of the Talamanca Cordillera (figure 5). In the north eastern part of the area, extended peat swamps near the coast can be found.

Generally speaking the Reventazón river is dividing the lowland area in two different geological units (figure 5). Northwest of this river alluvial soils derived from the Central Cordillera can be found. These are soils with an volcanic origin. In this area, the soils develop in parent material of volcanic origin, consisting of lava flows, lahars, pyroclastic materials and other materials of volcanic origin (Arias et al., 2010a). South of the Reventazón river alluvial soils originating from the Talamanca Cordillera can be found. These are a mixture of volcanic and more calcareous, sedimentary parent materials. The south eastern part of the area has soils which are derived from fine marine deposits, limestone and alluvial materials (Arias et al., 2010a).

3.1.2 Soils

Traditionally, the soils north western of the Reventazón river are defined as being soils of volcanic origin (figure 5). These soils are in general low in pH and base saturation, but have a good physical structure. Soils found in the north western lowlands can be characterized in general by good drainage, containing andic properties and being relatively low in bases. In this area Inceptisols and some Ultisols can be found. The Inceptisols are developing in young parent materials which are deposited by rivers and the volcanoes. On more stable, older positions, Ultisols can be found. Entisols are found near (former) river systems.

In the area southeast of the Reventazón river, more calcareous soil materials are present, which are derived from the Talamanca Cordillera. Here soils are more fertile but often have a poorer physical structure than the soils at the other side of the river. Here, at some places older, reddish clay soils occur. However, many soils found in the lowlands in the southeast are characterized by problems with drainage, relatively high amounts of bases and no andic properties. At these places mainly Inceptisols are found, especially Endoaquepts.

In the Turrialba area, volcanic soils can be found. These soils are formed in materials originating from the Turrialba volcano. The closer to the volcano, the more clearly andic properties are expressed. In this area, Inceptisols, Andisols and some Ultisols can be found. Processes like erosion, organic matter accumulation and weathering of volcanic parent materials determine soil formation in this area.

More to the south, soils are developing in parent materials which are derived from the Talamanca Cordillera. This means no volcanic soils can be found here, but soils with high base content due to the basic parent material are present. Inceptisols and some Ultisols can be found here.

3.2 Analysis of legacy and auxiliary data

Many data on soils and soil-forming factors is already available for the study area. For this reason, an analysis of legacy and auxiliary data was done to study which are the important soil-forming factors driving soil variability in the study area.

3.2.1 Description of the data used

The legacy data which was used was the 1:100,000 soil map of the Northern Atlantic Zone of Costa Rica (Wielemaker & Vogel, 1993). Since no detailed soil map exists for the total Atlantic Zone only the northern part of the study area was assessed. However, conditions in the northern part of the study area are assumed to resemble conditions in the southern part. Besides this, the map covers more than two third of the total study area. The soil map was converted in raster format in order to be able to calculate zonal statistics for every of the 151 mapping units. Some mapping units consist out of one soil type, but most mapping units are complex and consist out of several soil types.

All data was projected in Ocotepeque 1935 Lambert Norte, which is the projection in which the soil map of the Northern Atlantic Zone of Costa Rica was drawn originally. A list of auxiliary data used in the analysis can be found in table 1. All data was processed in ArcMap10.2.1.

Table 1 Overview of the GIS data used in the analysis and the soil-forming factor they represent.

Soil-forming factor	ArcGIS layer	Source	Format
-	Soil map Northern Atlantic Zone of Costa Rica	Wielemaker & Vogel (1993)	Polygon data
-	Elevation	ASTER GDEM NASA LP DAAC, USGS/Earth Resources Observation and Science (EROS) Center	Raster data, resolution of 30 m.
Climate	Yearly average temperature	Worldclim data Hijmans et al. (2005)	Raster data, resolution of 1 km
Climate	Yearly average rainfall	Worldclim data Hijmans et al. (2005)	Raster data, resolution of 1 km
Organisms	NDVI	LANDSAT 7 ETM+ NASA LP DAAC, USGS/Earth Resources Observation and Science (EROS) Center	Raster data, resolution of 15 m.
Relief	Slope	Derivative of ASTER GDEM	Raster data, resolution of 30 m.
Relief	Cosine of aspect	Derivative of ASTER GDEM	Raster data, resolution of 30 m.
Relief	Profile curvature	Derivative of ASTER GDEM	Raster data, resolution of 30 m.
Relief	Plan curvature	Derivative of ASTER GDEM	Raster data, resolution of 30 m.

Climatic soil-forming factors which were used were yearly average temperature and yearly average rainfall, since these are of high importance in soil formation. Air temperature influences soil temperature and has in this way strong influence on rates of soil-forming processes, especially biological processes. Yearly average rainfall influences water movement in the landscape, which highly influences soil-forming processes at the local scale (Lin et al., 2005).

In representing the soil-forming factor organisms, the Normalized Difference Vegetation Index (NDVI) was used. The NDVI is used as a proxy for the amount of biomass. Higher NDVI values indicate higher greenness of the landscape. NDVI was calculated from the different LANDSAT bands based on the following formula:

$$NDVI = \frac{NIR-Red}{NIR+Red}$$ Morawitz et al. (2006)

Thus, NDVI was calculated by taking the difference between the near infrared and visible red band of the LANDSAT data, and dividing this by the sum of both bands.

The soil-forming factor relief is considered by inventory of slope, aspect and curvature. As also mentioned under section 2.3, slope processes are highly important in inducing soil variability along the hill slope. Steeply sloping soils are in general less deep than soils at flat positions. Besides this, the slope affects the movement of water through the soil and in this way, processes like erosion. Furthermore, the aspect of a slope is influencing the soil temperature and wetness of the soil. In this way the aspect of a slope provokes the occurrence of different soils on northern and southern slopes. After calculation of the aspect, all values were converted into radials and the cosine was calculated for all values. For aspect, large values resemble small values, i.e. an aspect of 360^o is equal to an aspect of 0^0. By taking the cosine, the aspect is changed into a continuous variable and differences between north and south facing slopes are exaggerated. North and south facing slopes differ in the amount of sunlight they receive during the year, which affects soil temperature and in this way the rate of soil-forming processes. In this way, soil development will be different between north and south facing slopes. Soils on south facing slopes are often deeper than soils on north facing slopes due to the higher rate of the soil-forming processes.

The curvature of a position determines whether water is accumulating or diverging. This defines the water content of the soil and thus the translocations, transformations and losses from the soil profile. Both plan and profile curvature are assessed, which means the curvature in the direction of the slope as well as perpendicular to the slope are studied.

Although not being a soil-forming factor itself, elevation is taken into account as well. In many digital soil mapping exercises, elevation is considered when mapping soil properties.

Besides this, parent material was not studied in this analysis although it is one of the soil-forming factors. In developing the soil map, parent material is one of the assessed variables and thus a strong relation between the soil map and parent material will exist.

3.2.2 Calculations to determine the importance of the different auxiliary data in driving soil variability

After collecting all legacy and auxiliary data, the percentage of total variation in the auxiliary data explained by the mapping units of the soil map of the Northern Atlantic Zone of Costa Rica was examined. In this way it could be studied which of the soil-forming factors are inducing most soil variability in the northern part of the study area.

The analysis carried out is based on the ANOVA procedure. This means, total variability within the data is divided in two parts: variation between the means of the mapping units (expressed as the Sum of Squares Between) and variation within the mapping units (expressed as Sum of Squares Within). If the variation between mapping units is large compared to the variation within mapping units, the null hypothesis of ANOVA is rejected. The null hypothesis states that all the means of all mapping units are the same (Moore et al., 1994).

These calculations are based on the following equations which are used in an ANOVA:

$$SS_{within} = \frac{\sum(x_{ij} - \bar{x}_i)^2}{n}$$

$$SS_{total} = \frac{\sum(x_{ij} - \bar{x})^2}{n}$$

The Sum of Squares Within gives insight in the amount of variability within the groups; the Total Sum of Squares gives insight in the total variability found. For every of the auxiliary data layers which were studied, the mean value for the total area was determined, which is $\bar{x}$. Besides this, the mean value per mapping unit was calculated, i.e. $\bar{x}_i$. This was done by calculating zonal statistics for every auxiliary data layer based on the mapping units of the soil map. After this, the following calculations were done for every pixel of every layer of auxiliary data: $(x_{ij} - \bar{x}_i)^2$ and$(x_{ij} - \bar{x})^2$, which created two new layers with a unique value for every pixel in these layers. In these equations, x_{ij} is the value of one pixel, $\bar{x}_i$ is the mean value of the specific soil-forming factor for one mapping unit and $\bar{x}$ is the mean value of the specific soil-forming factor for the total area.

The values were not divided by *n*, since in the subsequent calculation these would cancel out. After determination of these two parameters for every pixel, the total Sum of Squares Within and Total Sum of Squares for every raster was determined by determining the mean of SS_{within} and SS_{total} for every raster. To calculate the percentage of variation between the mapping units of the total variation, the following equation was used:

$$SS_{between} = 100\% - \left(\frac{SS_{within}}{SS_{total}} * 100\%\right)$$

This equation is based on the fact that:

$$SS_{total} = SS_{within} + SS_{between}$$

In this way, the percentage of total variation in the auxiliary data explained by the mapping units of the soil map of the Northern Atlantic Zone of Costa Rica was determined. If this percentage is high, it means the auxiliary variable is important in explaining soil variability in the area.

Furthermore, correlations between auxiliary data were studied by calculating Pearson correlation coefficients. If the coefficient is close to one, the variable will be large when the other variable is large as well. If the coefficient is close to minus one, one variable will be small when the other variable is large. A coefficient close to zero indicates no relation between both variables exits. This was done to study whether auxiliary data is changing simultaneously. In this way, one variable can be used as proxy for the highly correlated ones.

3.3 Multi-scale analysis of soil variability

3.3.1 Sampling design

In order to be able to determine which soil-forming factors determine soil variability at different scales, the method of transect sampling was used during the field work. Transect sampling is a type of clustered sampling that allows the study of environmental gradients that occur within the landscape. Provided that transects are representative of a larger physiographic region, the collected data can be used to represent the region (Lin et al., 2005).

Since it is a multi-scale analysis, observations are done at different spatial scales. First, the study area is divided into zones and then these zones are subdivided until the smallest units are involved, which is the point scale level in this study. The advantage of this sampling method is that a wide range of spatial scales can be covered in one field study. This is particularly valuable where variation occurs in spatial scales that differ by several orders of magnitude simultaneously, thus when the variation is nested (Webster and Olive, 1990 in: Lin et al., 2005).

3.3.1.1 Assessment of soil variability at regional level

The fieldwork was carried out in collaboration with another student, who was mapping the agro-ecological variability in the area for a large research program on the soil-borne Panama disease in bananas. Besides this, the fieldwork was carried out in cooperation with Corbana, which is a Costa Rican institution which main objective is to develop the Costa Rican banana industry and to serve the banana producers. For this reason, the fieldwork was structured around the landscapes where bananas are grown.

First, the study area was divided into four main zones. The four zones are perceived as being distinctly different in soils, land use and climate. As being explained in section 3.1, the soils of the lowlands of the Atlantic zone are traditionally divided into the soils east and west of the Reventazón river, related to different management advices. Based on this assumption, the first two zones were determined. Large scale banana plantations can be found in these zones. Furthermore, a zone in the area of Turrialba was determined, which is a mountainous area where banana plants are used as a shadow crop for coffee bushes. The fourth main zone is the southern part of the Atlantic Zone. Here, soils are under influence of the Talamanca Cordillera and are assumed to be different than the soils in the three other regions due to the mountainous landscape and the parent material which is a mixture of volcanic and calcareous materials.

After determining the four main zones, every zone was subdivided in a subzone with low altitude and a subzone with high altitude. For the first two main zones, the 75 meter contour line was used to divide the zone in a low and a high altitude part. The 75 meter contour line is perceived as dividing these zones in a relatively steep, higher zone and a more flat lower zone. Besides this, using this contour line, banana plantations could be found in both the high and low part of the zones. The Turrialba area was divided in two zones based on the 1000 metres contour line to have two zones with reasonable size and different climatic conditions and parent material. The southern Atlantic Zone was divided in an area with low altitude and more rainfall close to the Caribbean Sea and a zone with higher altitude and less rainfall.

3.3.1.2 Assessment of soil variability at local level

In every of these eight zones described under section 3.3.1, one transect of five kilometres has been assessed. In figure 6 the location and names of the eight transects are shown.

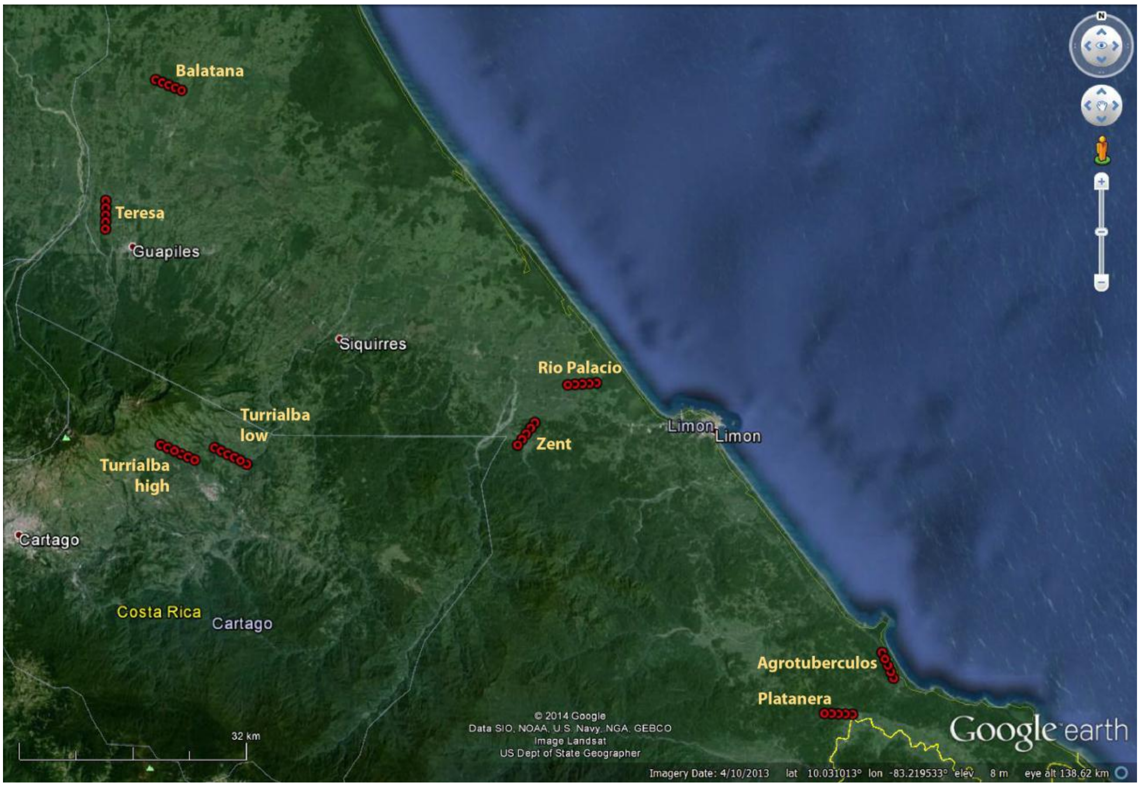

Figure 6 Locations of the eight transects which are studied during the fieldwork. The red dots are the sampling locations; in yellow the names of the transects are shown. Except for the Turrialba transects, five locations are shown per transect, since the exact location of location 1 was determined in the field. For the Turrialba transects, location 1 was determined beforehand, since on these transects no random location in a banana plantation could be determined (source: Google Earth).

If present, a banana plantation was used as starting point of each transect, which is location 1. Only for the two transects in the Turrialba area this was not possible. For the six other transects, randomly a banana plantation for the starting point was chosen with help of ArcGIS and the random number generator in Excel, since every plantation was assigned a number. For the Turrialba area, the starting point was chosen pragmatically. This means, for the low part of the Turrialba area a coffee plantation was used as starting point. For the high part of this zone, the starting point was located in a garden. Except for the transects in the Turrialba area, the name of the transect is equal to the name of the banana plantation in which the first point is located. In table 2 an overview of the eight transects can be found.

Table 2 Names of the eight transects and the zone they are representing.

Area	Name of transect
West of Reventazón river	
low altitude	Balatana
high altitude	Teresa
East of Reventazón river	
low altitude	Rio Palacio
high altitude	Zent
Turrialba area	
low altitude	Turrialba low
high altitude	Turrialba high
Southern Atlantic Zone	
low altitude	Agrotuberculos
high altitude	Platanera

For each of the eight transects, the soil was assessed in detail at six locations at the transect (figure 7). The first location of each transect was positioned randomly within the selected plantation. Locations 2 till 6 were spaced regularly with a distance of one kilometre in between, starting at the border of the plantation. The distance between location 1 and 2 was variable, depending of the location of location 1 in the banana plantation. At least the distance was one kilometre, however dependent onto the randomly chosen site of location 1, this could be more. The decision of the direction of the transects was based on the direction of highest variability in land use, soils and elevation which was determined by examination of Google Earth images and the soil map. Besides this, accessibility of the transect was taken into account.

After deciding the location of a transect, the transect was drawn in Google Earth and the GPS coordinates of the locations were used to find the locations in the field (figure 6). A detailed soil description was made at every of the six locations, which resulted in six soil descriptions per transect, which means in total 48 soil descriptions were made. In table 3 the assessed soil properties are listed. See appendix I for the complete soil description form which was used. First the soil texture was estimated in the field and afterwards, percentages clay were derived from the texture triangle of the FAO.

Table 3 Soil properties studied at every location.

Surface parameters	Surface description	Soil profile description
Altitude	Drainage class	Horizons
Slope	Permeability	Depth of horizons
Profile curvature	Surface stoniness	Soil colour
Plan curvature	Effective soil depth	Texture
Aspect	Tixotrophy	Mottling
	Parent material	Stoniness
	Banana suitability class	Reaction with NaF

After field assessment of these properties, the soil type was determined for every of the 48 locations based on the USDA Soil Taxonomy classification system (Soil Survey Staff, 1999). Besides this, the auxiliary data described under section 3.2.1 were linked to the locations in ArcGIS.

3.3.1.3 Assessment of soil variability at point level

The lowest scale level studied in the field is the scale of twenty metres. As described under section 3.3.1.2, within each transect six locations are studied. At location 1, 3 and 6, short distance variability (i.e. at 20 metres) was assessed as well. At these locations, five samples of the topsoil (i.e. 0-30 cm depth) are taken and one sample of the subsoil (i.e. 30-60 cm depth). Four samples are positioned at a square with a side of 20 metres around the centre point, which together with the centre point gives five topsoil samples at one location (figure 7). In this way, short distance variability in soil properties of the topsoil could be assessed. The subsoil sample and the soil description are taken from the centre point of the square. At location 2,4 and 5, the same procedure has been followed, however here a composite sample was made of the five topsoil samples to reduce the total number of samples.

The samples have been chemically assessed in the soil laboratory of Corbana, which means pH, organic matter content and soil fertility parameters are determined. In table 4 below the number of soil samples in total is shown.

Table 4 Number of soil samples per location, transect and for the total study area

Location	Number of soil samples
1	6
2	2
3	6
4	2
5	2
6	6
Total number of soil samples per transect	24
Total number of soil samples of eight transects	192

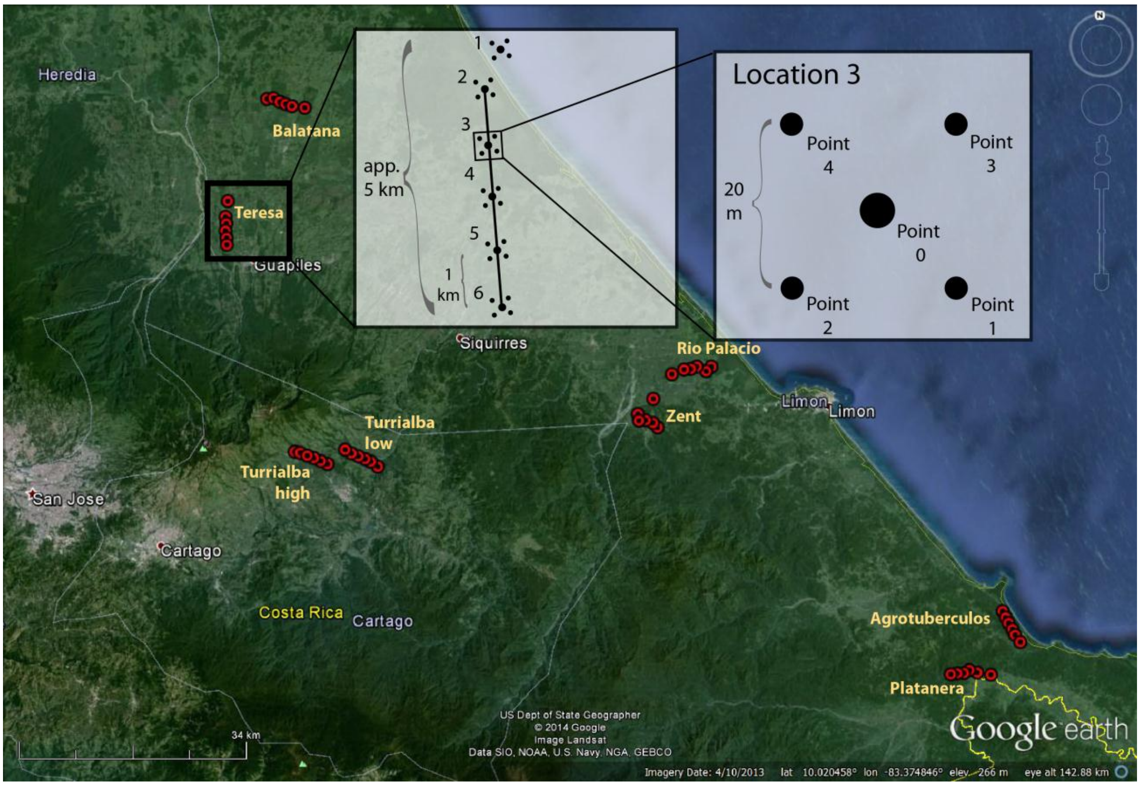

Figure 7 The figure shows the three different scale levels taken into account in the multi-scale analysis. At regional scale, eight transects (in red) are studied. Within the transects, six locations are determined. Location 1 is randomly located within a banana plantation. Locations 2 till 6 are spaced on a regular distance of one kilometre. Within each location, five points are sampled: one centre point and four corner points spaced on a squared with a side of twenty metres.

3.3.2 Calculations on soil variability and variability in auxiliary data based on gathered data

As explained above, a large variety of soil properties was assessed in the field. However, to describe soil variability in relation with the soil-forming factors in the area effectively, a focus is put upon three important soil fertility parameters which are used to describe soil variability: clay content, pH of the topsoil and the organic matter content of the topsoil. The clay content of a soil can be directly related to the parent material. The pH of a soil gives insight in the acidification of a soil, which is influenced by parent material as well as soil management practices. The soil-forming process accumulation of organic matter determines the organic matter content of the soil.

For assessment of spatial variability of soil properties in an area, often the semivariogram is used, which relates the semivariance to the distance between points (Heuvelink & Webster, 2001). However, in this study all data points should be grouped together in order to have enough points to develop the semivariogram. The semivariogram is often used to describe soil variability from a geostatistical point of view (Heuvelink and Webster, 2001). However, different characteristics of the different zones which are assessed are omitted using one semivariogram, since every zone would give its own semivariogram. When taking all points together, this variation between zones would be omitted. This resembles the often used technique in digital soil mapping to use all auxiliary data available without usage of pedological knowledge. In this way, important processes driving soil variability could be overlooked.

For this reason, a different way of analysing the data was used. For the three soil properties described above as well as for the auxiliary data, a statistical summary was calculated. This means, the average value, standard deviation and coefficient of variation for the data was calculated. For the auxiliary data this was done for the regional and local scale level alone, since no data was available at the point scale. For the pH and organic matter content of the topsoil, variability was studied at all three scale levels. The standard deviation and coefficient of variation were used to give insight in variability of both soil-forming factors and soil properties. By calculating the coefficient of variation, variation soil properties between the transects could be studied. Besides this, the standard deviation of the soil properties was determined for the different scale levels and compared between the scale levels. By doing this, the scale level at which most variation occurs could be determined and linked to the driving processes.

As stated in the introduction, determination of statistical connections between co-variables and soil properties is used in describing soil variability in digital soil mapping as well. It was critically stated that in this way knowledge on soil formation and the distribution of soils in the landscape is not taken into account. Although here a statistical summary of soil properties and soil-forming factors will be given as well, this serves a different purpose. The soil-forming factors and their linkage with inducing soil variability will be used to describe the landscape from a mechanistic point of view by means of the field observations. Field observations are the central part of the analysis; the statistical summary of properties and factors serves to support in this description.

4. RESULTS & DISCUSSION

4.1 Analysis of legacy and auxiliary data

When studying which of the auxiliary data is most important in determining soil variability in the study area at regional level, it is clear that elevation is inducing most variability between mapping units, whereas aspect, profile curvature and plan curvature hardly induce any variability (table 5). Climatic factors are highly important factors driving soil variability as well: yearly average temperature explains 88.3% of total variation explained by the mapping units, whereas the mapping units explain 69.1% of the total variation in yearly average amount of rainfall. The mapping units explained some of the total variation in slope percentage and NDVI as well, although substantial less than elevation and the climatic factors: respectively 31.2% and 12.8%.

Clearly, besides elevation, climatic factors are determining soil variability at the regional scale level based on this assessment of the available data. This means, processes like heat, water and solute flow which are induced by respectively temperature and rainfall (Opolot, et al. 2014), are most important in causing soil variability on this scale level. Water movement affects the transformations, translocations, additions, and losses to and from a soil profile. Combined with differences in physical soil properties varying soil properties will occur (Lin et al., 2005). Factors changing on the small scale, which are curvature and aspect, are not responsible for soil variability at regional scale according to this analysis, which is in agreement with the conceptual model (figure 4).

It should be noticed that elevation in itself is not a soil-forming factor, which means this factor is not responsible for inducing soil variability. Statistically it is an important factor but from a mechanistic point of view, it has no value. Elevation summarizes the effect of several soil-forming factors which are often closely linked to it, which are factors like temperature, rainfall and parent material.

Table 5 The percentage of total variation in the auxiliary data explained by the mapping units of the soil map of the Northern Atlantic Zone of Costa Rica.

	Percentage (%)
Elevation	89.0
Yearly average temperature	88.3
Yearly average amount of rainfall	69.1
Slope	31.2
NDVI	12.8
Cosine aspect	2.3
Profile curvature	0.015
Plan curvature	0.008

When studying the correlation between the assessed factors, it becomes clear several soil-forming factors are strongly correlated (table 6). Especially elevation and yearly average temperature are highly correlated. Furthermore, also slope percentage and yearly average temperature are correlated, as well as rainfall and temperature and cosine aspect and temperature.

Based on these relations, it can be stated that only a limited set of soil-forming factors is needed to explain soil variability in this particular area. For assessment of soil variability at this scale level climatic factors like yearly average temperature and yearly average rainfall are most important. Slope percentage and NDVI are also somewhat important. However, NDVI is correlated to yearly average temperature (r=-0.43) and thus temperature can be used as a proxy for NDVI.

Table 6 Pearson correlation coefficients for auxiliary data for the Northern Atlantic Zone of Costa Rica. Slope is expressed in percentage.

	Temperature	Rainfall	NDVI	Elevation	Slope	Profile curvature	Plan curvature	Cosine aspect
Temperature	1.00							
Rainfall	0.47	1.00						
NDVI	-0.43	-0.34	1.00					
Elevation	-0.99	-0.47	0.45	1.00				
Slope	-0.79	-0.36	0.42	0.79	1.00			
Profile curvature	0.43	0.26	-0.31	-0.46	-0.22	1.00		
Plan curvature	0.20	0.17	-0.11	-0.18	-0.16	0.07	1.00	
Cosine aspect	-0.52	-0.26	0.24	0.55	0.33	-0.36	-0.33	1.00

4.2 Multi-scale analysis of soil variability

4.2.1 Assessment of soil variability at regional level

The study area is part of a highly variable environment which is reflected in a high variability in the soil-forming factors. Mountainous areas characterized by volcanic parent material, steep slopes and relatively low temperatures are found in the western part of the area, while in the eastern part a slightly rolling landscape under strong fluvial influence can be found. Large variation in relief and climatic factors are found, especially in rainfall. This variation is reflected in the variation found in the assessed soil-forming factors (table 7).

Table 7 Statistical summary of the assessed soil-forming factors at regional level

Soil-forming factor		Value
Elevation (m)	Mean	295.8
	Std. dev.	474.2
Slope (%)	Mean	10.6
	Std. dev.	7.9
Plan curvature (-)	Mean	0.116
	Std. dev.	0.228
Profile curvature (-)	Mean	-0.027
	Std. dev.	0.204
Cosine aspect (-)	Mean	-0.125
	Std. dev.	0.377
Yearly average temperature (^{0}C)	Mean	24.8
	Std. dev.	2.7
Yearly average rainfall (mm)	Mean	3216.7
	Std. dev.	701.7
NDVI (-)	Mean	0.32
	Std. dev.	0.14

Large variation in elevation is reflected in a standard deviation of 474 metres. The standard deviation of yearly average rainfall is 702 mm, which is high for this relatively small area. Curvature values are most variable, whereas the yearly average temperature is varying least, with a standard deviation of 2.7 ^{0}C.

Variation in rainfall is related to orographic effects and oceanic influences. High rainfall amounts can be found near the volcanic slopes and in the north eastern part of the study area. Due to the mountainous landscape in the southern and western part of the area and the more flat landscape in the eastern part, high variability in elevation and slopes can be found. In the low, flatter eastern part, the influence of curvature is important. NDVI is high everywhere in the area caused by the greenness of the landscape due to the tropical conditions, which means high temperatures are present year-round and plant growth is possible all year.

Table 8 A statistical summary of clay content, pH and organic matter content of the topsoil for the regional scale.

	Mean	Standard deviation
Clay content (%/profile)	30.1	9.64
pH topsoil	5.64	0.38
Organic matter topsoil (%)	5.98	3.78

As a result of the variation in soil-forming factors, large variation of organic matter content in the topsoil was found. Mean organic matter content in the study area is almost 6% with a standard deviation of 3.78% (table 8). This variation is mainly induced by variation in parent material. The Andisols found in the Turrialba area are characterized by high organic matter contents. By formation of complexes between aluminium in allophanes and organic matter, high organic matter contents can be retained in these soils. This process is typical for these volcanic soils. Besides this, these soils are situated higher in the landscape which is related to lower temperatures and thus lower breakdown rates of organic materials. This will result in quicker accumulation of organic matter in the soil as well.

In the study of Powers & Schlesinger (2002), relationships between soil carbon and soil-forming factors were assessed in a study area in the north of Costa Rica. Amongst others, they studied the correlation between soil carbon and soil-forming factors. They found a strong relation between elevation, although not being a real soil-forming factor, and soil carbon concentration. When correlation is determined between soil-forming factors and soil properties for the data points of this study, a correlation coefficient of -0.84 is found between organic matter content of the topsoil and temperature, taking into account all locations. Given that a high correlation exists between temperature and elevation (table 6), a strong relation between elevation and soil carbon concentration exists as well in this study area.

Surprisingly enough, the variation in topsoil pH is small; the mean pH of the topsoil is 5.64 with a standard deviation of 0.38. With a mean pH of 5.64, soils in the study area are in general slightly acidic. Significantly higher pH values were expected for the soils east of the Reventazón river relatively to the soils west of the Reventazón river due to the calcareous parent material originating from the Talamanca Cordillera in that area. However, acidification by the use of nitrogen fertilizers can lower soil pH substantially (Bolan et al., 1991). Besides this, high weathering rates are present in the area due to the tropical climate. High, constant temperatures accelerate processes within the soil, especially in combination with high yearly average rainfall which will induce high leaching within the profile. During the weathering process, acidifying products will be produced. In combination with the acidifying effect of fertilizers, which are used in high amounts on banana plantations, soil pH will decrease.

A correlation coefficient of -0.402 was found between pH and yearly average rainfall in the study area. This could be related to the fact that on the transects Platanera and Agrotuberculos substantial lower amounts of rainfall can be found (i.e. 2349 mm against a mean of 3217 mm for all data points), but high average soil pH, which is related to the calcareous parent material derived from the Talamanca Cordillera.

The clay content in the study area is on average 30.1%. The clay content has a correlation of -0.42 with the cosine of the aspect. This means, a more negative aspect will lead to higher clay contents, i.e. south facing slopes will have higher clay content, probably due to higher weathering rates at the warmer southern slopes. The weathering of parent materials in the Caribbean lowlands is strongly affected by three main processes: the amount of precipitation, causing leaching of soluble elements, the frequency of deposition of fresh sediments by rivers and the increase of soil temperature closer to the Caribbean Sea. These three factors induce a higher weathering rates of the soils in the lowest parts of the study area (Arias et al., 2010b).

4.2.2 Assessment of soil variability at local level

At local level a standard deviation of 0.53 can be found for the mean value of topsoil pH when all locations of all transects are taken into account. For organic matter content in the topsoil, a standard deviation of 4.31 is found. Clay percentage has a standard deviation of 12.4. This means, comparing these values with the values showed under section 4.2.1, the variation between transects in clay, pH and organic matter content of the topsoil is somewhat larger than the variation in the total study area.

Table 9 A statistical summary of clay content, pH and organic matter content of the topsoil at local scale

	Mean	Standard deviation
Clay content (%/profile)	30.1	12.4
pH topsoil	5.64	0.53
Organic matter topsoil (%)	5.98	4.31

Variation in soil-forming factors was, just like the scale of the total study area described under section 4.1, highest for curvature and lowest for temperature and rainfall, see appendix II for a statistical summary of the soil-forming factors per transect.

However, variation in elevation within transects was much smaller than variation between transects. Also slope percentage variability was lower on transect scale than on scale of the whole study area. Related to a decrease in variability of the soil-forming factors within transects with respect to the total study area is the decrease in variability of soil properties within transects. In comparison with the highest scale level, thus that of the total study area, variation within transects is lower. Only for Zent and the low Turrialba transect the variation of the pH of the topsoil was higher than the variation in the total study area. This can be related to the high variability between the different zones in climate, relief and parent materials.

In appendix III a statistical summary of the soil properties specified for every of the eight transects can be found. In the next sections, this table will be shown separately for each transect. Clay content was highly variable between transects. The Teresa transect is characterized by the lowest average clay content, 11.6% and the highest coefficient of variation of all transects, 60%. Rio Palacio has the highest average clay content of 44% and the lowest coefficient of variation: 15.4%.

The mean pH of the Agrotuberculos transect is with a value of 6.16 the highest, while the transect Turrialba low has the lowest average with a value of 4.99. This can be related to the different parent materials, respectively calcareous and volcanic materials. Most variation in pH is found in the Zent transect with a standard deviation of 0.65. Lowest variation is found in the transect of Platanera with a standard deviation of 0.15.

The organic matter content was clearly higher in the high altitude Turrialba transect, with a value of 14.20%. Lowest mean organic matter content is found in the Platanera transect with a value of 3.36%. Most variation in organic matter content was found in the transect of Teresa with a CV of 55.84%. Least variable is the transect of Rio Palacio with a CV of 18.53%.

These figures can be explained based on the following perspective on the landscape in the study area. The western part of the study area consists out of volcanic mountains. Here, volcanism has influenced the soil development: lava flows at different points in time, erosion and sedimentation cause the soils we find nowadays. In that area, mostly Andisols can be found, to which the high organic matter contents found on the Turrialba transects are related. Rivers that drain these mountains are flowing into the lowlands towards the Caribbean Sea. In this lower part of the study area, rivers exert strong influence on the development of soils. They cause the regular deposition of fresh materials which will lead to stratification in parent materials within soil profiles. Due to the tropical climate, i.e. the high amount of yearly rainfall, many rivers are found and are exerting large influence on soil development, especially in the rainy season. The distance to river is very important in soil development in this area. Close to rivers, more sandy soils are found, containing high amounts of stones, due to the high flow velocity of rivers. Further away from the rivers, more clayey soils can be found. This gives rise to the variability in clay and organic matter content found on the Teresa transect.

The smaller the distance to the Caribbean Sea, the more clayey the soils are, which explains the high average clay content of the Rio Palacio transect. Besides this, drainage problems are common in this area. Due to poor drainage, strong gleization occurs in these soils. In the low area, Inceptisols are the most commonly found soils. Close to rivers or former rivers, Entisols can be found.

In the higher parts within the low part of the study area, highly weathered soils can be found. These soils are not under influence of rivers and have not received fresh deposits in recent times. These soils are high in clay content and developed red colours, related to long periods of weathering and leaching of weatherable materials. These soils can be characterized as Ultisols.

4.2.2.1 Teresa

The Teresa transect is located in an area which is highly influenced by rivers (figure 8). With a mean clay content of 12%, soils on this transect are a lot sandier than other soils in the area; the mean clay content in the area is 30% (table 10). A lot of textural differences can be found between the locations as well as within the profile. Besides this, high variability in organic matter content in the topsoil was found, reflected in a standard deviation of 2.64. All soils have andic properties here, although no soils from the soil order Andisols were found. The general land use was pasture, with in between some small scale agriculture.

Table 10 Clay content, pH and organic matter of the topsoil of the Teresa transect.

	Mean	Std. dev.	CV (%)
Clay content (%/profile)	11.6	6.9	60
pH topsoil	5.59	0.18	3.29
Organic matter topsoil (%)	4.74	2.64	55.84

Since all observations were done in proximity of the fast flowing Toro Amarillo river, relatively low average clay content was found. Variability in texture in this transect can be related to the strong influence of rivers onto the landscape, causing deposition of variable parent materials. The high variability in organic matter content in the topsoil can be associated to this variation in parent materials as well since clay soils are in general better in retaining organic matter. Furthermore, the distance to rivers is important for soil development at this transect, which means variability in parent material is important in inducing soil variability. Close to the rivers, high stoniness and sand content and low soil depth is found, changing with distance to rivers. Besides this, the transect is located in alluvial fan materials, which is another explanation for the high sand and stone content of the soils (Nieuwenhuyse, 1996). The alluvial fans are originating from the volcanic Central Cordillera, causing the expression of andic properties. Due to the young age of the soils and regular renewing of parent materials by rivers, no real Andisols have developed here. Due to the sandiness and relative high stoniness of the soils here, the landscape is used extensively.

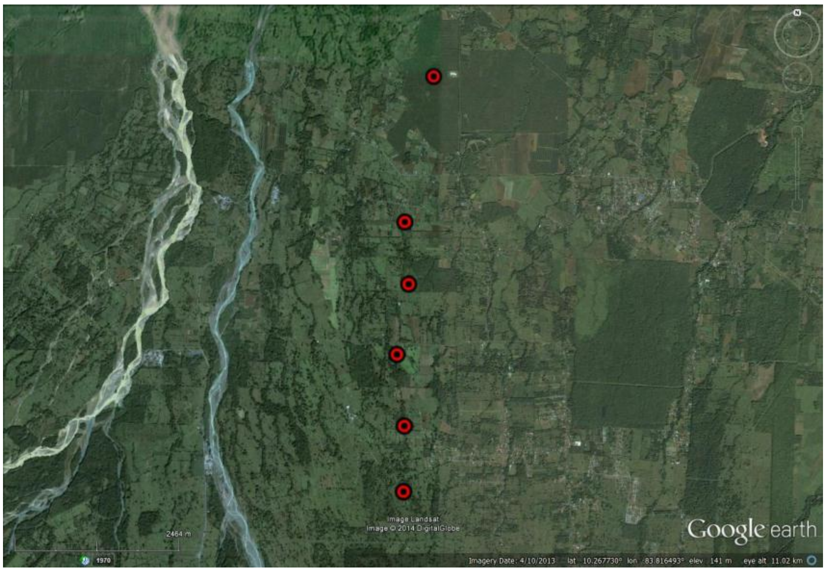

Figure 8 The six locations sampled at the Teresa transect. Location 1 is the uppermost red dot (source: Google Earth).

4.2.2.2 Balatana

Table 11 Clay content, pH and organic matter of the topsoil of the Balatana transect.

	Mean	Std. dev.	CV (%)
Clay content (%/profile)	26.6	13.2	49.5
pH topsoil	5.28	0.47	8.94
Organic matter topsoil (%)	5.54	1.67	30.08

This transect is representing the lower parts of the fluvial dominated zone west of the Reventazón river (figure 9). Although mean elevation is lower than transect Teresa, more height differences were assessed in the landscape; a more rolling landscape is present. On top of the small hills, higher sand contents were found. In the lower lying parts of the area, higher clay content was found (up to 47%) and mottling can be observed. No stones were found in the assessed profiles and clay content was higher than in the soils in the Teresa transect, i.e. on average 27% (table 11). Here, large scale banana plantations are interfered with pineapple plantations, pasture and small scale agriculture.

Based on the observations on this transect, micro relief is highly influencing soil development in the area, inducing small differences in rates of soil-forming processes like leaching and organic matter accumulation and thus soil variability at small scale. These small hills with a flat top are believed to be remnants of an old Pleistocene terrace level (Nieuwenhuyse, 1996). Rivers have even more influence on soil development than in transect Teresa; bigger differences in elevation were observed in the field. On all locations soil auguring up to 120 centimetres was possible in contrast to the Teresa transect. On larger scale, the Balatana transect is located lower in the landscape, which means more at the extremities of the large alluvial fans systems from the mountains. This gives rise to soils with higher clay contents, deeper soils and less stoniness in comparison to the Teresa transect. Related to larger soil depth and higher clay content, land use is more variable than in the Teresa transect and the area is more suitable for agriculture than the zone of the Teresa transect due to a better soil structure.

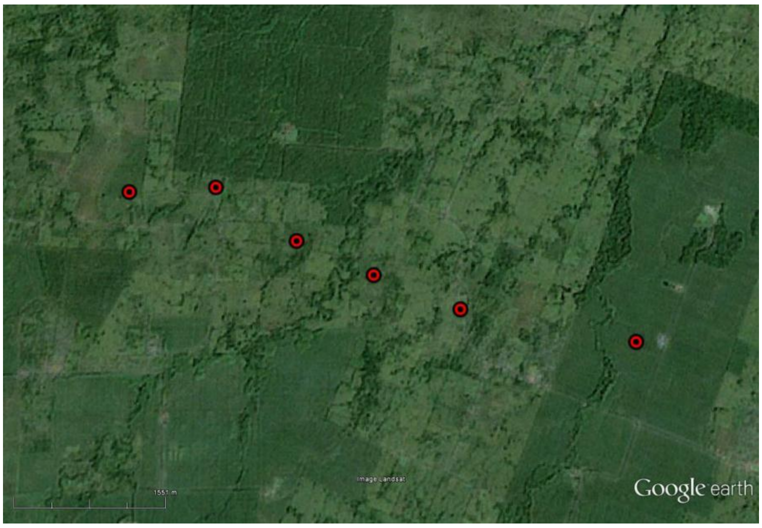

Figure 9 The six locations sampled at the Balatana transect. Location 1 is the most right red dot (source: Google Earth).

4.2.2.3 Rio Palacio

Table 12 Clay content, pH and organic matter of the topsoil of the Rio Palacio transect.

	Mean	Std. dev.	CV (%)
Clay content (%/profile)	44	6.8	15.4
pH topsoil	5.7	0.49	8.61
Organic matter topsoil (%)	3.36	0.62	18.53

In the zone of transect Rio Palacio, the landscape is very flat and, related to this, poorly drained. Figure 10 shows the locations of the transect. In all locations, no significant slope was observed, just as can be seen in appendix II. Less fluvial influence than in the previous two transects was observed. The zone is characterized by many large-scale banana plantations and besides this, large-scale pasture fields. Compared to the previous two transects, less variation in soils occurred. In general, all soils were poorly drained clay soils, i.e. Endoaquepts. This decrease in variation in soil properties is also observed in the relatively low variation in pH, soil organic matter content and clay content compared to the Teresa and Balatana transect (table 12).

This means, topography does not play an important role in inducing variability within this transect. Gleization, related to the poor drainage, is an important in-profile process occurring in this zone based on the observations made. The intensive use of soils can be related to the high nutrient status of these soils, due to non-volcanic, calcium rich origin of the parent material derived from the Talamanca Cordillera. Major limitation for crop growth in the area is the high groundwater table; at the point closest to the coast, groundwater could be found at 20 cm below the surface. This can be related to the low position in the landscape of the transect.

When studying the soil-forming factors in appendix II the variability seems to be relatively high, although this was not observed in the field. This fact points to the importance of field observations with respect to remotely sensed satellite data. Low variation in soil properties was found as well, which confirms the low variation in soil-forming factors.

Figure 10 The six locations sampled at the Rio Palacio transect. The most left red dot is location 1 (source: Google Earth).

4.2.2.4 Zent

Table 13 Clay content, pH and organic matter of the topsoil of the Zent transect.

	Mean	Std. dev.	CV (%)
Clay content (%/profile)	38.2	7.3	19.2
pH topsoil	5.52	0.65	11.81
Organic matter topsoil (%)	3.99	1.8	44.98

In general, this transect is located in a distinctly different landscape than the transects described above. Figure 11 shows the locations of the transect. The landscape can be characterised as mountainous with parent materials derived from the Talamanca Cordillera. In general, pasture is the main land use type here, with some subsistence farming and natural vegetation in between. The area is crossed by a large river and several smaller ones. Further away from the rivers on more stable positions, old, red clay soils can be found. Closer to the rivers more fluvial influenced soils can be found with younger materials and lighter textures.

The high variability of pH and organic matter content of the topsoil in the Zent transect (table 13) can be linked to its mountainous character and thus large variability in topography which will cause variability in soil properties and rates of the soil-forming processes. Besides this, the transect was located just on the border between fluvial parent materials derived from rivers Chirripó and Zent and materials derived from the Talamanca Cordillera. These differences in parent materials and landscape positions induced the development of different soils. This means, close to the rivers young, fresh parent materials could be found, while higher up in the mountains strongly weathered clays soils were found.

It should be noted that the planned transect was not accessible because the road that was intended to drive did not exist. For this reason, a transect along another road was made. Starting point of the transect was at the corner of the Copemas banana plantation instead of the Zent plantation and the points were not completely located on a straight line. If the intended transect would be sampled, locations would be positioned all in the mountainous area and the fluvial derived soils as found now would not be observed.

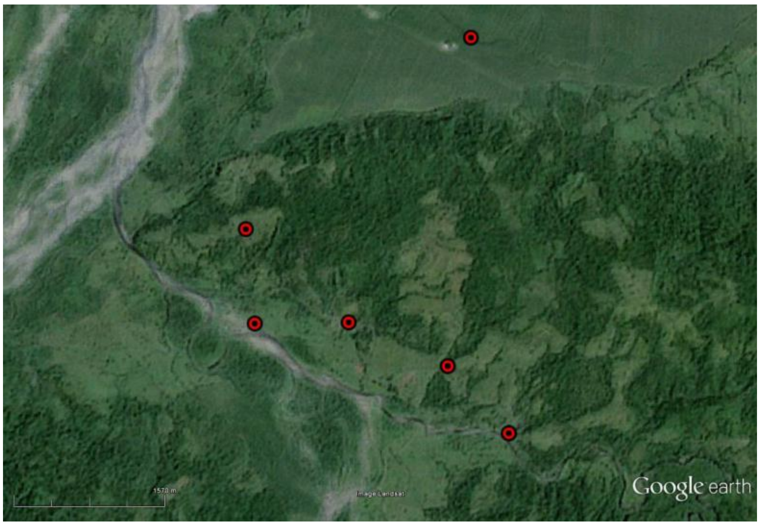

Figure 11 The six locations on the Zent transect. Location 1 is the uppermost red dot (source: Google Earth).

4.2.2.5 Turrialba low

Table 14 Clay content, pH and organic matter of the topsoil of the Turrialba low transect.

	Mean	Std. dev.	CV (%)
Clay content (%/profile)	29.2	11.8	40.4
pH topsoil	4.99	0.5	9.93
Organic matter topsoil (%)	8.83	3.99	45.18

This transect which is located in the lower part of the Turrialba area has the largest altitude difference of the eight transects. Figure 12 shows the locations of the transect. Along the transect a height difference of 500 metres is encountered. Due to strong correlation with elevation, this results in a coefficient of variability for temperature and rain of respectively 3.6% and 6.1%, which is the highest of all transects (appendix II). Mean pH of this transect is lowest and most variable of all transects. Organic matter content of this transect is 8.8% with a coefficient of variation of 45.2% (table 14). Soils are characterised by low soil depth (<70 cm) and in general a loamy texture. Higher in the landscape andic properties are more clearly expressed in the profile. Land use exists in general out of very diverse small scale agriculture. Almost no pasture is present in this area, and a high diversity in planted crops exist: amongst others coffee, papaya and banana.

The soils found on this transect have good physical structure in general due to the volcanic parent materials present here which are related to low bulk densities. However, soil depth, stoniness and low base status are limiting factors for development of intensive agriculture in this area. Relatively high variability in climatic conditions induced by topography result in high variation in rates of soil-forming processes and thus in soil properties in the landscape. Linked to this, higher in the landscape the distance to the Turrialba volcano is smaller and volcanic parent materials cause the explicit expression of andic properties in the soil. Here, different soil-forming processes will play a role than lower in the landscape, thus closer to the rivers Reventazón and Guayabo, where young, fluvial parent materials can be found. The weathering of volcanic parent material containing volcanic glass closer to the volcanoes result in soils rich in organic matter.

Figure 12 The six locations of the Turrialba low transect. Location 1 is the most right red dot. (source: Google Earth).

4.2.2.6 Platanera

Table 15 Clay content, pH and organic matter of the topsoil of the Platanera transect.

	Mean	Std. dev.	CV (%)
Clay content (%/profile)	30.1	11.5	38.2
pH topsoil	5.95	0.15	2.56
Organic matter topsoil (%)	3.36	1.32	39.21

This transect is mainly located in the floodplains of the fast flowing river Sixaola (figure 13). This causes the transect to be characterized by fluvial derived soils. Many small rivers are found in the floodplain of the river Sixaola. A layering of different textures is present within the profile at many places due to river deposits, ranging from loamy sand to clay loam. In old riverbeds, more stony soils are found, with stone percentages up to 40% in one horizon. In the floodplains of the rivers, the landscape is relatively flat, while the surrounding landscape is mountainous. Almost everywhere bananas plants can be found, usually mixed with other crops. Variation in pH is smallest of all transects on this transect (table 15).

In general, the area is highly influenced by river Sixaola. Furthermore, many small rivers are present, exerting high influence on the soils by frequent deposition of fresh materials. Parent material found here is derived from eroded materials from higher parts of the Talamanca Cordillera which are transported by the rivers to the lower parts of the area. This causes the relatively high pH of the topsoil compared to the other transects. In the higher parts of the transect, weathering of the parent materials lead to the transition of the soils from Inceptisols to Ultisols.

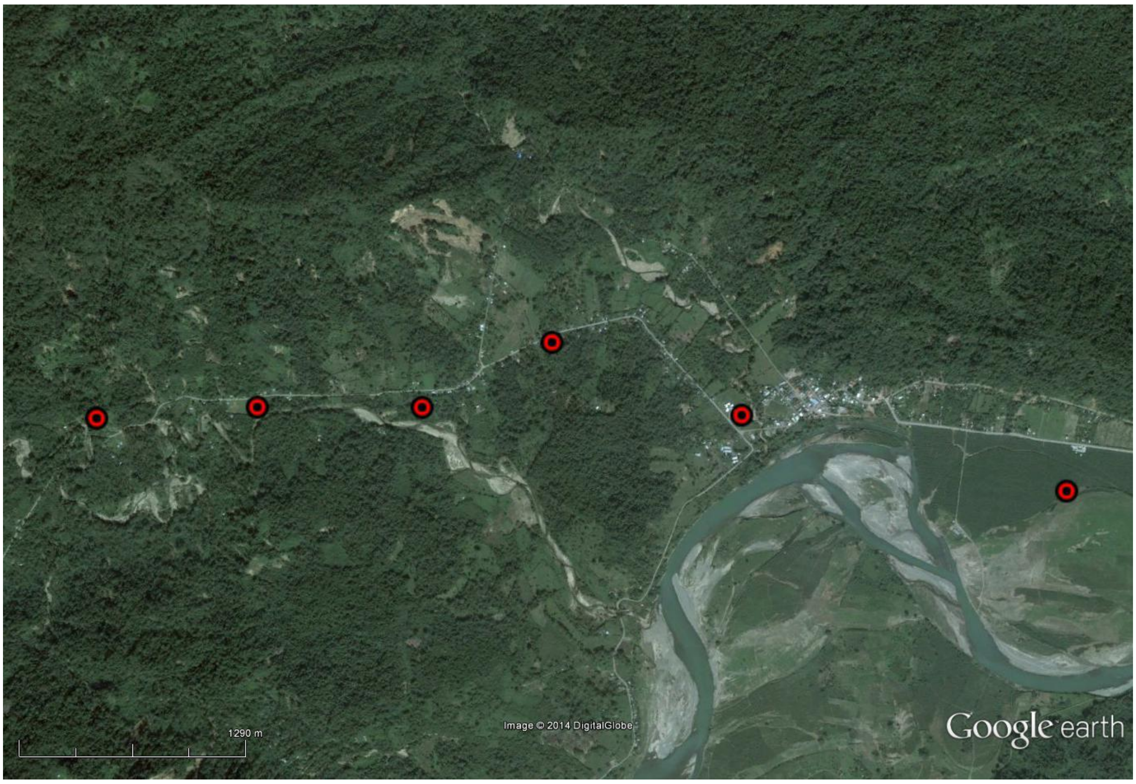

Figure 13 The six locations on the Platanera transect. Location 1 is the most right red dot (source: Google Earth).

4.2.2.7 Agrotuberculos

The Agrotuberculos transect is located parallel to the Caribbean Sea (figure 14). On this transect, the landscape is flat and at the west side the transect is bordered by a small mountain ridge. The transect is in general quite uniform, which is reflected in the relatively low coefficients of variation compared to the other transects (table 16). No significant variations in topography were observed. Most variation could be found in parent materials assessed within the profiles. Stratification of different parent materials were observed due to fluvial deposition.

Table 16 Clay content, pH and organic matter of the topsoil of the Agrotuberculos transect.

	Mean	Std. dev.	CV (%)
Clay content (%/profile)	26.5	7.7	29
pH topsoil	6.16	0.34	5.48
Organic matter topsoil (%)	3.79	1.13	29.85

At three locations, clay on sand was found, while the other locations are characterised by clay loam. However, within the transect little variation in soils was observed. However, the transect is located parallel to the Caribbean Sea, which has an impact on the soil development along this transect. Parallel to the ocean, similar parent materials will be found, although slightly modified by rivers draining into the ocean. Besides this, the locations on this transect have the highest pH of the topsoil of all transects. This can be linked to calcareous deposits from former sea high stands, affecting parent materials on the transect.

Figure 14 The six locations on the Agrotuberculos transect. Location 1 is the lowest red dot (source: Google Earth).

4.2.2.8 Turrialba high

Table 17 Clay content, pH and organic matter of the topsoil of the Turrialba high transect.

	Mean	Std. dev.	CV (%)
Clay content (%/profile)	35	3.7	10.4
pH topsoil	5.93	0.39	6.56
Organic matter topsoil (%)	14.2	4.74	33.37

This transect is located in the Central Cordillera and is positioned close to the Turrialba volcano. Figure 15 shows the locations of the transect. This higher part of the Turrialba area is highly influenced by deposits from the Turrialba volcano. Closer to the volcano, which means higher in the landscape, the andic properties get more clearly expressed. In some lower areas, Ultisols can be found. In the lowest part, Inceptisols can be found due to sedimentation of rivers. Mean organic matter content of the topsoil is distinctively higher than on the other transects with a value of 14.2% and coefficient of variation of 33.4% (table 17). Land use is dominated by pasture and natural vegetation in the area; no banana plantations can be found. The NDVI is highest for this region: 0.612 with a coefficient of variation of 26.5%.

This area can be described as very active and young. Signs of active soil movement could be seen in the landscape in form of erosion (i.e. landslides) and recent deposits from the volcano. The dominance of pasture as common land use is probably related to the good physical structure of the soils, but the low chemical fertility. Besides this, allophanes present in Andisols form complexes with organic matter, which explains the high mean organic matter content of this transect. Volcanic soils here are characterised by phosphorus retention and a low base status. By weathering of the volcanic parent material, old Andisols have weathered towards Ultisols at some stable positions. The high NDVI can be related to the large amount of pastures in the area. As stated by Morawitz et al. (2006), grasslands can have same values for NDVI as forests.

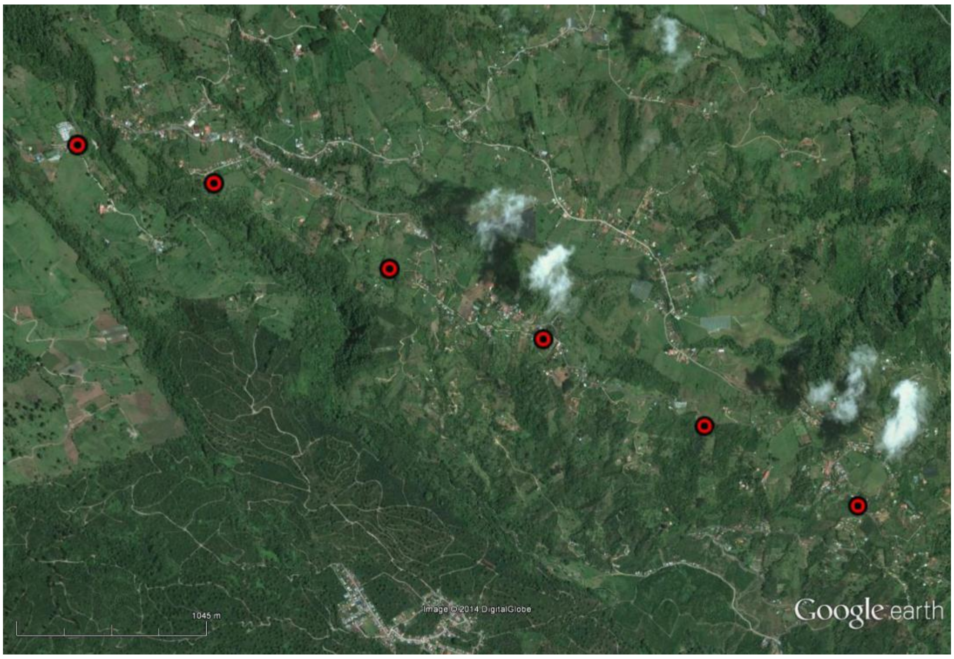

Figure 15 The six locations on the Turrialba high transect. The most right red dot is location 1 (source: Google Earth).

4.2.3 Assessment of soil variability at point level

At the point level, variation in clay content was not examined, since manual estimated differences in soil texture were not expected beforehand at this scale. For this reason, only variation in pH and organic matter content of the topsoil are described in this section. In appendix IV a statistical summary of data of soil variability at point level can be found.

Taking into account all 24 locations where short distance variability of the soil properties was assessed, variation of the pH in the topsoil is somewhat larger on the point scale level than on the scale of a transect. Standard deviation in pH of the top soil at point level is 0.63 (table 18); while the standard deviation in pH on local level is 0.53.

In contrast, the variation of organic matter content in the topsoil is smaller on the field scale level than on the level of the transects. Standard deviation in organic matter content of the top soil at point level is 4.11 (table 18); while the standard deviation in pH on local level is 4.31.

Table 18 A statistical summary of pH and organic matter content of the topsoil at point scale

	Mean	Standard deviation
pH topsoil	5.73	0.63
Organic matter topsoil (%)	5.93	4.11

When studying variability in pH of the topsoil at the scale of a single point, location 3 of the Rio Palacio transect is most variable (appendix IV). This location is situated in the banana plantation Carrandi. This may explain the high variability in pH of the topsoil. Due to fertilizer applications close to the plants, at short distance high variability in pH may occur. In this case, soil variability is mainly driven by soil management, which is part of the soil-forming factor *organisms*. When assessing all locations situated on a banana plantation, which are nine locations of the twenty four locations where short distance variability was assessed, mean standard deviation of the pH of the locations at a banana plantation was 0.48 whereas for the variation within the other locations a mean standard deviation of 0.31 was found. In fact, variability in pH is often high, since it is influenced by many different processes. Soil management will have inevitably strong effect on soil pH. For example, high fertilizer amendments are applied to banana plantations, which may have acidifying effects on the soil. Besides this, organisms will have high impact on soil pH, structure and nutrient levels due to the high activity under tropical conditions.

When assessing variability in organic matter content of the topsoil, a similar picture emerges. Location 1 of the Zent transect, which is located in a banana plantation as well, has most variation in organic matter content, which is 62% (appendix IV). Comparing the locations at banana plantations with locations under different land use gives a view onto the difference between these two groups of points. Mean standard deviation of organic matter content of the topsoil of the locations positioned at a banana plantation is 1.11%, while the other locations have a mean standard deviation of 1.66%. Although standard deviation at the banana plantations is lower, the coefficient of variation is 34%, compared to 26% at the other locations. This is a result of the low mean organic matter content of the locations on a banana plantation compared to the other locations; respectively 3.83%, compared to 7.18%. Due to high extraction of biomass due to high production rates at the banana plantations, organic matter content of the topsoil will be lower than on the other locations.

4.2.4 Integration of scale levels

Relating the variation in soil properties over the different scale levels which are studied in the previous sections gives a view on the important driving factors of soil variability in the study area. In figure 16 the variation in pH and organic matter content of the topsoil from regional scale to point scale can be found. When including lower scale levels, more variation in organic matter content and pH can be found. However, in general, the change of variation over the different scale levels is small. This shows that large scale climatic factors, as shown in the conceptual model in section 2.3 are important drivers in soil variability. The climatic factors will induce soil variability at large scales, since these factors are most variable at large scale levels. This means, processes like heat, water and solute flow are important in driving soil variability in the study area. However, variability in both properties is somewhat increasing over the scale levels, which means a thorough examination of small scale variability should not be underestimated, just as was proposed by Lin et al. (2005). Land use exerts an important influence at variability in organic matter content and pH at the point scale, as was shown under section 4.2.3.

Differences in standard deviations between organic matter content and pH cannot be compared directly since both properties work at a different range, which means variation in pH is in absolute numbers always smaller than variation in organic matter content.

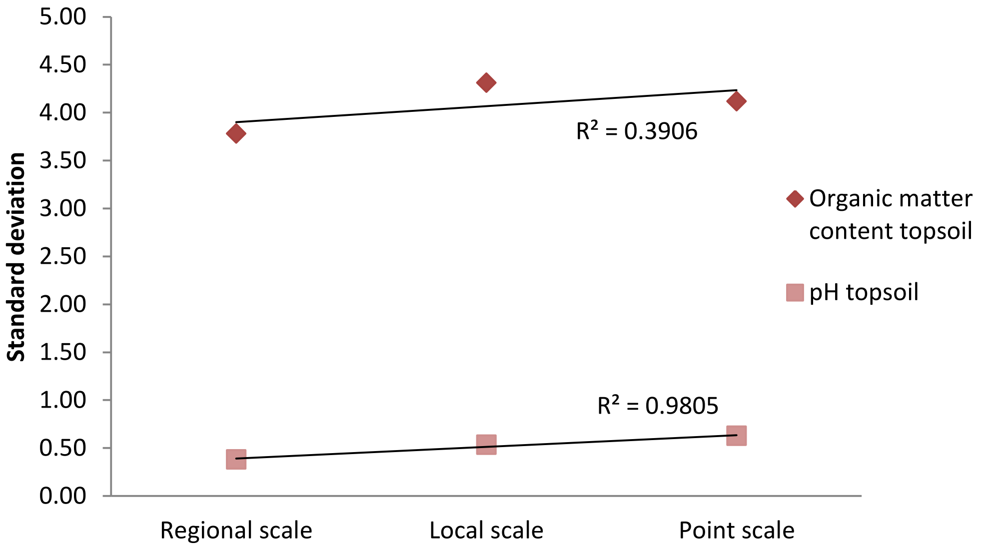

Figure 16 Variation in organic matter content and pH of the topsoil from regional scale to point scale.

5. GENERAL DISCUSSION

5.1 Conceptual model

The conceptual model developed under chapter 2 gives a framework which can be used as a support in figuring out which soil-forming factors are most important in describing soil variability at different spatial scales. As shown by figure 4, to some extent all soil-forming factors play a role at each scale level. However, as indicated by Kirkby et al. (1996) for erosion models, different processes become important when moving from large to small spatial scales.

The question which of the factors is most important in describing soil variability cannot be answered with a general statement based on the model alone. This has to be determined for every specific case, since every area will have its specific environmental conditions. The Atlantic Zone of Costa Rica is characterized by a large variety in parent material and climate conditions. Thus, in this area these factors will be important drivers of soil variability, in contrast to an area with small variation in parent material and climate conditions. However, the development of such a conceptual model helps to understand the importance of considering scale levels when studying soil variability and environmental processes in general.

Based on understanding of the main drivers of soil variability, more efficient soil sampling schemes can be developed. As also mentioned by Domburg et al. (1997), prior knowledge on soil variability will help in developing these schemes. This means, time and money can be saved and better predictions of for example spread of soil-borne diseases based on soil variability can be established.

As stated in the introduction, the soil-forming factors are used in this study to develop a mechanistic understanding about soil variability in the study area instead of the processes itself, because the processes are difficult to quantify on landscape scale. Within traditional soil mapping and digital soil mapping this problem occurs as well. In digital soil mapping, quantitative interpolation techniques often ignore soil formation, while methods based on traditional soil mapping often lack a quantitative framework in including soil-forming processes (Moore et al., 1993). This means, to come to a better mechanistic approach to soil variability on the landscape scale, soil-forming processes should be studied more thoroughly at all scale levels, especially in relation to the landscape.

Furthermore, variation in soils in time was not included in this study. Since rates of soil-forming processes will not be static in time, a study about the relation between the soil-forming factors and time scales would be interesting. When the temporal variability of soils will be taken into account in combination with spatial variability of soils, an even better understanding of soil variability in the landscape will be developed and more fine-tuned agricultural management practices can be established.

5.2 Analysis of legacy and auxiliary data

Based on the analysis of legacy and auxiliary data carried out, the importance of the climatic factors temperature and rainfall on soil variability became clear. On the regional scale, the factors curvature and aspect had only little effect. This is in accordance with the conceptual model presented in chapter 2; these factors exert influence at lower scale levels by influencing small scale soil-forming processes. Furthermore, strong correlation between soil-forming factors was proved in the study area. This means a limited set of soil-forming factors can be used to model soil spatial variability in the study area. An example of prediction of soil properties based on a limited number of soil-forming factors in digital soil mapping is the study carried out by Mora-Vallejo et al. (2008).

Legacy data used here was covering mainly the northern part of the study area. However, it is expected this region is representative for the total study area. The assessed part consists out of two third of the total study area and besides this, environmental conditions are assumed to be the same for this area. Besides this, mapping units were taken as the basis of the inventory of variability. However, mapping units consist often out of several soil types and may contain in this way a large variation within soil properties. This means, these results cannot be compared one to one to results on variability in soil properties. However, this method can give a quick view on which factors influence soil variability in an area and thus on which soil-forming processes are important in the area. As such, it is a cost-efficient way of studying the drivers of soil variability in a study area, since the collection of field data is a relative slow and expensive method of data collection (McBratney et al., 2003). By such a desk study before carrying out fieldwork, most important soil-forming factors can be determined and more efficient and purposeful fieldwork can be carried out.

The resolution of the auxiliary data differed between the climatic data, relief data and NDVI data. The resolution of the climate data was one kilometre, which automatically will result in lower variability than the data on relief which has a 30 meter resolution or the NDVI which is derived from LANDSAT data with a resolution of 15 metres. However, on regional scale the variation in climatic conditions will not be substantial within a scale of one kilometre, especially because climatic variations at this resolution were already substantial. Due to the resolution of the auxiliary data it was not possible to study variability in soil-forming factors on point scale. However, spatial resolution of most satellite images is not yet high enough to be able to distinguish these scale levels, which also gives limitations to digital soil mapping (Mulder et al., 2011).

Besides this, the result of the analysis of the legacy and auxiliary data will be influenced by the positioning of the delineated areas which are part of one mapping unit. These might be scattered over the study area or located close together. This can have a major influence on the results. If delineated areas are close together, climatic factors will be less variable between delineated areas than in the case the delineated areas are scattered around the region. When delineated areas are scattered, variability in climatic conditions within mapping units will be large.

5.3 Multi-scale analysis of soil variability

The multi-scale analysis of soil variability showed that soil variability is increasing when smaller spatial scales are taken into account in the study area. Climatic conditions are influencing variability at the scale of the total study area. Within the transects, parent material can be marked as the most important factor driving soil variability. On point scale, the importance of land use became clear. This means, processes heat and water flow are mainly important in driving soil variability, whereas processes related to soil management will exert influence at short spatial scales. However, the fact that one transect of five kilometres with six locations is used to characterize large areas may give rise to uncertainty about this data. Especially in the mountainous areas, it is difficult to characterize the soils based on six locations. Whether a location is situated ten metres up or down slope can have a large impact on soil properties. For getting a more detailed view onto soil variability in the study area, more and longer transects should be assessed. However, limitations in accessibility, budget and time result in pragmatic decisions in fieldwork. Besides this, the fieldwork carried out serves to give a general overview of the variability in the study area and can be used as a starting point for more detailed soil sampling schemes in soil variability for usage in agriculture.

Some question marks can be placed by the GIS data which is linked to the locations, which may have influenced the results substantially. For example, the mean slope percentage of 11% which is found for the total area is a very high value compared to the mean value of 5.6% which was observed in the field. This shows however that the spatial resolution of 30 metres of the DEM derived soil-forming factors is too small; variability in slope percentages will occur often within the scale of 30 metres.

Finally, it is important to notice the importance of some randomness in soil variability on small scale levels; not all variability can solely be attributed to variation in soil-forming factors. As is stated by Phillips (2001), spatial soil variability may to some extent be unrelated to measurable, observable variations in soil-forming factors. Intrinsic factors like dynamical instability and divergent self-organisation may lead to spatial soil variability which cannot be explained by the soil-forming factors (Phillips, 2001).

5.4 Comparing the two methods used in this study

Both methods used to study important drivers of soil variability in this research have shown their advantages. By analysis of the readily available data, knowledge on processes working in the area could be obtained, even before observations in the field were made. In this way, a more efficient and well-founded fieldwork campaign can be carried out afterwards. Based on the work carried out, it became clear certain soil-forming factors are of less importance in inducing soil variability on a regional scale than on small scale, for example land use. However, to be able to carry out such an analysis, the importance of the existence of legacy data should not been underestimated. Field observations are of paramount importance in understanding the landscape. For this reason, soil science should focus not only on development of statistical models and GIS approaches to soil variability, but also the collection of field data should not be neglected.

By the multi-scale analysis carried out in this study, conclusions are drawn based on these field observations. By an in-field observation of the landscape and soil variability, a real-world view on the variability at different scale levels could be obtained. For future research it will be important to include scale levels in assessment of soil variability.

In this way, both methods work complementary. This means, disadvantages in the first method are taken into account by the other method. Whereas the analysis of available data is a quick and cheap method, giving a quantitative view on drivers of variability, the multi-scale assessment in the field is much more expensive. However, by using this method soil properties could be taken into account, different scale levels could be studied and expert knowledge could be applied. In this way, the analysis of auxiliary and legacy data used the advantage of digital soil mapping that drivers of soil variability could be quantified. The multi-scale analysis included the usage of expert knowledge as is done in traditional soil mapping as well. In this way, both methods can complement each other. After the relatively quick and cheap analysis of already existing data, fieldwork can be carried out based on the results of this analysis to get a better understanding of soil variability at different spatial scales. Subsequently, this inventory of drivers behind soil variability can form the basis of more efficient soil sampling schemes in agriculture.

6. CONCLUSION

Based on the analysis of legacy data and auxiliary data, climatic factors proved to be most important drivers of soil variability in the Atlantic Zone of Costa Rica. The multi-scale analysis confirmed this result by linking the three studied scale levels. Besides this, this method showed the high variability in soil properties and soil-forming factors in the study area. Whereas on regional level climatic conditions are highly important in driving soil variability, within the local level variation in parent material was very important. On point level, the importance of land use became clear.

The analysis of legacy and auxiliary data is a relatively quick and cheap method to study the important soil-forming factors driving soil variability. This method used the advantage of digital soil mapping that drivers of soil variability could be quantified. The multi-scale analysis of soil variability at the other hand gave a real-world view on soil variability and showed the importance of different soil-forming factors at different scale levels. This method included the usage of expert knowledge as is done in traditional soil mapping as well. In this way, both methods work complementary and can be used jointly to study soil variability in a mechanistic way on landscape scale.

A combination of analysis of legacy and auxiliary data together with field observations on different scale levels provides a solid basis for more efficient and well-founded studies on soil variability, which are of high importance for agricultural management.

ACKNOWLEDGEMENTS

The LANDSAT 7 ETM+ and ASTER GDEM data was obtained from the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota.

Furthermore, I would like to express my sincere gratitude to Corbana, Corporación Bananera Nacional, in Guapíles, Costa Rica, in collaboration with whom I carried out my fieldwork. Special thanks to Rafael Segura Mena, Maria Araya and Juan Samuels, who helped a lot during the fieldwork. Also I want to give a special thanks to the people of the physical and chemical laboratories of Corbana, who analysed all soil samples quick and accurate.

REFERENCES

Arias, F., Mata, R., Alvarado, A., Serrano, E., & Laguna, J. (2010a). Caracterización química y clasificación taxonómica de algunos suelos cultivados con banano en las llanuras aluviales del Caribe de Costa Rica. *Agronomía Costarricense, 34*, 177-195.

Arias, F., Mata, R., Alvarado, A., Serrano, E., & Laguna, J. (2010b). Mineralogía de la fracción arcilla de algunos suelos cultivados con banano en las llanuras aluviales del Caribe de Costa Rica. *Agronomía Costarricense, 34*, 197-222.

Bockheim, J. G., & Gennadiyev, A. N. (2000). The role of soil-forming processes in the definition of taxa in Soil Taxonomy and the World Soil Reference Base. *Geoderma, 95*(1–2), 53-72. doi: http://dx.doi.org/10.1016/S0016-7061(99)00083-X

Bockheim, J. G., Gennadiyev, A. N., Hammer, R. D., & Tandarich, J. P. (2005). Historical development of key concepts in pedology. *Geoderma, 124*(1–2), 23-36. doi: http://dx.doi.org/10.1016/j.geoderma.2004.03.004

Bolan, N. S., Hedley, M. J., & White, R. E. (1991). Processes of soil acidification during nitrogen cycling with emphasis on legume based pastures. [Plant and Soil]. *134*(1), 53-63. doi: 10.1007/bf00010717

Bouman, B. M., Jansen, H. P., Schipper, R., Hengsdijk, H., Nieuwenhuyse, A., Nieuwenhuyse, A., . . . Alfaro, R. (2000). The physical and socio-economic setting: the northern Atlantic Zone of Costa Rica *Tools for Land Use Analysis on Different Scales* (Vol. 8, pp. 9-34): Springer Netherlands.

Brady, N. C., & Weil, R. R. (2004). *The nature and properties of soils* (Second edition ed.). New Jersey: Pearson Education.

Bui, E. N. (2004). Soil survey as a knowledge system. *Geoderma, 120*(1–2), 17-26. doi: http://dx.doi.org/10.1016/j.geoderma.2003.07.006

Burrough, P. A. (1983). Multiscale sources of spatial variation in soil. I. The application of fractal concepts to nested levels of soil variation. *Journal of Soil Science, 34*(3), 577-597. doi: 10.1111/j.1365-2389.1983.tb01057.x

Coates, A. G., & Obando, J. A. (1996). The geologic evolution of the Central American Isthmus. *Evolution and environment in tropical America*, 21-56.

Domburg, P., de Gruijter, J. J., & van Beek, P. (1997). Designing efficient soil survey schemes with a knowledge-based system using dynamic programming. *Geoderma, 75*(3–4), 183-201. doi: http://dx.doi.org/10.1016/S0016-7061(96)00090-0

Gray, J. M., Humphreys, G. S., & Deckers, J. A. (2009). Relationships in soil distribution as revealed by a global soil database. *Geoderma, 150*(3–4), 309-323. doi: http://dx.doi.org/10.1016/j.geoderma.2009.02.012

Hartemink, A. E., McBratney, A. B., & de Lourdes Mendonça-Santos, M. (2008). *Digital soil mapping with limited data*: Springer.

Heuvelink, G. B. M., & Webster, R. (2001). Modelling soil variation: past, present, and future. [Developments and Trends in Soil Science]. *Geoderma, 100*(3–4), 269-301. doi: http://dx.doi.org/10.1016/S0016-7061(01)00025-8

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology, 25*(15), 1965-1978. doi: 10.1002/joc.1276

Holmes, K. W., Kyriakidis, P. C., Chadwick, O. A., Soares, J. V., & Roberts, D. A. (2005). Multi-scale variability in tropical soil nutrients following land-cover change. [Biogeochemistry]. *74*(2), 173-203. doi: 10.1007/s10533-004-3544-x

Jahn, R., Blume, H. P., Asio, V. B., Spaargaren, O., & Schad, P. (2006). *Guidelines for soil description, 4th edition*. Rome: FAO.

Jenny, H. (1941). *Factors of soil formation*: McGraw-Hill Book Company New York, NY, USA.

Kirkby, M. J., Imeson, A. C., Bergkamp, G., & Cammeraat, L. H. (1996). Scaling up processes and models from the field plot to the watershed and regional areas. *Journal of Soil and Water Conservation, 51*(5), 391-396.

Lin, H., Wheeler, D., Bell, J., & Wilding, L. (2005). Assessment of soil spatial variability at multiple scales. [Scaling, fractals and diversity in soils and ecohydrology]. *Ecological Modelling, 182*(3–4), 271-290. doi: http://dx.doi.org/10.1016/j.ecolmodel.2004.04.006

McBratney, A., Mendonça Santos, M. d. L., & Minasny, B. (2003). On digital soil mapping. *Geoderma, 117*(1), 3-52.

McBratney, A. B., Mendonça Santos, M. L., & Minasny, B. (2003). On digital soil mapping. *Geoderma, 117*(1–2), 3-52. doi: http://dx.doi.org/10.1016/S0016-7061(03)00223-4

McBratney, A. B., Odeh, I. O. A., Bishop, T. F. A., Dunbar, M. S., & Shatar, T. M. (2000). An overview of pedometric techniques for use in soil survey. *Geoderma, 97*(3–4), 293-327. doi: http://dx.doi.org/10.1016/S0016-7061(00)00043-4

Minasny, B., & McBratney, A. B. (2001). A rudimentary mechanistic model for soil formation and landscape development: II. A two-dimensional model incorporating chemical weathering. [Estimating uncertainty in soil models]. *Geoderma, 103*(1–2), 161-179. doi: http://dx.doi.org/10.1016/S0016-7061(01)00075-1

Minasny, B., & McBratney, A. B. (2006). Mechanistic soil–landscape modelling as an approach to developing pedogenetic classifications. [Advances in landscape-scale soil research]. *Geoderma, 133*(1–2), 138-149. doi: http://dx.doi.org/10.1016/j.geoderma.2006.03.042

Moore, D. S., McCabe, G. P., & TransVorm, V. (1994). *Statistiek in de praktijk*: Academic service.

Moore, I. D., Gessler, P. E., Nielsen, G. A., & Peterson, G. A. (1993). Soil Attribute Prediction Using Terrain Analysis. *Soil Sci. Soc. Am. J., 57*(2), 443-452. doi: 10.2136/sssaj1993.03615995005700020026x

Mora-Vallejo, A., Claessens, L., Stoorvogel, J., & Heuvelink, G. B. M. (2008). Small scale digital soil mapping in Southeastern Kenya. *CATENA, 76*(1), 44-53. doi: http://dx.doi.org/10.1016/j.catena.2008.09.008

Morawitz, D., Blewett, T., Cohen, A., & Alberti, M. (2006). Using NDVI to Assess Vegetative Land Cover Change in Central Puget Sound. [Environmental Monitoring and Assessment]. *114*(1-3), 85-106. doi: 10.1007/s10661-006-1679-z

Mulder, V. L., de Bruin, S., Schaepman, M. E., & Mayr, T. R. (2011). The use of remote sensing in soil and terrain mapping — A review. *Geoderma, 162*(1–2), 1-19. doi: http://dx.doi.org/10.1016/j.geoderma.2010.12.018

Nieuwenhuyse, A. (1996). *Soils, Geology, and Soil-related Sustainability Aspects of the Perhumid Tropical Limon Basin, Costa Rica*: CATIE.

Opolot, E., Yu, Y., & Finke, P. (2014). Modeling soil genesis at pedon and landscape scales: Achievements and problems. *Quaternary International*.

Phillips, J. D. (2001). The Relative Importance of Intrinsic and Extrinsic Factors in Pedodiversity. [Annals of the Association of American Geographers]. *Annals of the Association of American Geographers, 91*(4), 609-621. doi: 10.1111/0004-5608.00261

Powers, J. S., & Schlesinger, W. H. (2002). Relationships among soil carbon distributions and biophysical factors at nested spatial scales in rain forests of northeastern Costa Rica. *Geoderma, 109*(3–4), 165-190. doi: http://dx.doi.org/10.1016/S0016-7061(02)00147-7

Simonson, R. W. (1959). Outline of a generalized theory of soil genesis. *Soil Science Society of America Journal, 23*(2), 152-156.

Soil Survey Staff. (1999). *Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys*: Natural Resources Conservation Service. U.S. Department of Agriculture Handbook 436.

Wielemaker, W. G., & Vogel, A. W. (1993). *Un sistema de información de suelos y tierras para la zona atlántica de Costa Rica*: Universidad Agricola.

Yemefack, M., Rossiter, D. G., & Njomgang, R. (2005). Multi-scale characterization of soil variability within an agricultural landscape mosaic system in southern Cameroon. *Geoderma, 125*(1–2), 117-143. doi: http://dx.doi.org/10.1016/j.geoderma.2004.07.007

APPENDICES

Appendix I – Soil description form of field work

SOIL DESCRIPTION FORM

Site code:	GPS–reading		
	UTM Zone	Easting [m]:	Northing [m]:
Date:			

Soil Site Description

Soil-forming factors	*Surface description*
Altitude [m]:	Drainage class:
Slope [%]:	Permeability:
Profile curvature:	Surface stoniness [%]:
Plan curvature:	Effective soil depth:
Aspect [°]:	Tixotrophy:
Parent material:	Banana suitability class:
Land use:	
MA Rainfall (GD):	
MA Temperature (GD):	

Remarks:

Soil Profile Description

		Colour	Mottles		Texture	Stones in profile		Structure	NaF
Horizon	*Depth (cm)*	*Moist*	*%*	*Type*	*Class*	*%*	*Weath.*	*Type*	*Andic? Yes/no*

ST Classification:

Catena Position:

Remarks:

Appendix II – Statistical summary of the soil-forming factors of all transects

Transect		Elevation	Slope	Plan curvature	Profile curvature	Cosine aspect	Yearly average temperature	Yearly average rainfall	NDVI
Teresa	mean	142	7.4	-0.009	0.183	0.355	256.2	4081.8	0.256
	std.dev.	31	5.0	0.469	0.358	0.678	1.5	23.4	0.082
	CV (%)	22	67.3				0.6	0.6	31.9
Balatana	mean	50	6.7	0.396	-0.266	0.017	260.2	4297.0	0.283
	std.dev.	6	4.9	0.662	0.489	0.797	0.4	9.5	0.079
	CV (%)	13	73.0				0.2	0.2	27.9
Rio Palacio	mean	16	3.2	-0.098	0.146	-0.145	263.8	3156.0	0.307
	std.dev.	4	4.5	0.269	0.281	0.646	0.4	18.8	0.110
	CV (%)	28	141.7				0.2	0.6	35.8
Zent	mean	55	8.6	-0.290	0.215	-0.136	261.3	3242.8	0.185
	std.dev.	27	9.6	0.723	0.481	0.825	1.8	99.3	0.132
	CV (%)	50	111.2				0.7	3.1	71.4
Turrialba low	mean	695	14.7	0.176	-0.155	-0.438	225.8	3117.7	0.346
	std.dev.	186	10.1	0.431	0.344	0.799	8.2	189.4	0.313
	CV (%)	27	68.9				3.6	6.1	90.5
Platanera	mean	60	9.1	0.289	-0.094	-0.164	262.0	2304.2	0.370
	std.dev.	24	12.6	0.441	0.432	0.951	1.1	45.7	0.342
	CV (%)	40	138.5				0.4	2.0	92.4
Agro Tuberculos	mean	21	6.5	0.223	0.049	0.307	263.8	2393.0	0.169
	std.dev.	6	7.7	0.420	0.565	0.682	0.4	27.5	0.156
	CV (%)	28	118.1				0.2	1.1	92.7
Turrialba high	mean	1327	28.4	0.242	-0.298	-0.800	189.3	3141.2	0.612
	std.dev.	111	17.7	0.466	0.311	0.292	11.3	81.2	0.162
	CV (%)	8	62.3				6.0	2.6	26.5

Appendix III – Statistical summary of clay content, pH and organic matter content of all transects

	Clay content			pH topsoil			Organic matter topsoil		
	Mean (%/profile)	Std. dev.	CV (%)	Mean	Std. dev.	CV (%)	Mean (%/profile)	Std. dev.	CV (%)
Teresa	11.6	6.9	60.0	5.59	0.18	3.29	4.74	2.64	55.84
Balatana	26.6	13.2	49.5	5.28	0.47	8.94	5.54	1.67	30.08
Rio Palacio	44.0	6.8	15.4	5.70	0.49	8.61	3.36	0.62	18.53
Zent	38.2	7.3	19.2	5.52	0.65	11.81	3.99	1.80	44.98
Turrialba low	29.2	11.8	40.4	4.99	0.50	9.93	8.83	3.99	45.18
Platanera	30.1	11.5	38.2	5.95	0.15	2.56	3.36	1.32	39.21
Agrotuberculos	26.5	7.7	29.0	6.16	0.34	5.48	3.79	1.13	29.85
Turrialba high	35.0	3.7	10.4	5.93	0.39	6.56	14.20	4.74	33.37

Appendix IV – Statistical summary of clay content, pH and organic matter content of the locations where short distance variability was assessed

		pH topsoil			Organic matter topsoil		
Transect	Location	Mean	Std. dev.	CV (%)	Mean (%)	Std. dev.	CV (%)
Teresa	1	5.7	0.5	8.2	9.1	0.7	8.2
	3	5.7	0.1	1.2	5.7	1.3	22.8
	6	5.5	0.3	4.8	2.8	0.7	23.5
Balatana	1	6.1	0.6	10.2	4.8	1.6	32.7
	3	5.2	0.2	3.5	6.4	1.3	20.0
	6	5.3	0.1	1.8	8.4	0.7	8.3
Rio Palacio	1	4.8	0.2	5.1	2.8	0.4	15.9
	3	5.5	0.9	15.9	2.6	1.3	51.7
	6	5.9	0.4	7.3	3.9	1.8	45.5
Zent	1	5.8	0.3	5.4	1.9	1.2	61.9
	3	6.5	0.5	7.0	2.8	0.9	33.8
	6	5.7	0.1	2.2	4.3	2.4	56.9
Turrialba low	1	4.4	0.2	5.0	5.2	0.8	15.5
	3	5.6	0.4	7.9	6.5	1.9	30.0
	6	5.3	0.3	5.1	16.2	1.5	9.2
Sixaola	1	6.0	0.4	6.2	2.4	0.3	10.8
	3	5.7	0.3	5.6	2.8	1.2	43.2
	6	6.2	0.6	9.6	5.6	1.1	19.7
Agrotuberculos	1	5.7	0.7	12.3	4.4	1.4	32.0
	3	5.9	0.3	4.5	3.6	0.7	18.3
	6	6.5	0.4	6.0	3.7	1.8	50.0
Turrialba high	1	6.0	0.7	11.2	9.7	2.1	21.8
	3	6.1	0.4	6.6	13.4	4.4	32.5
	6	6.3	0.2	2.4	13.3	3.4	25.3