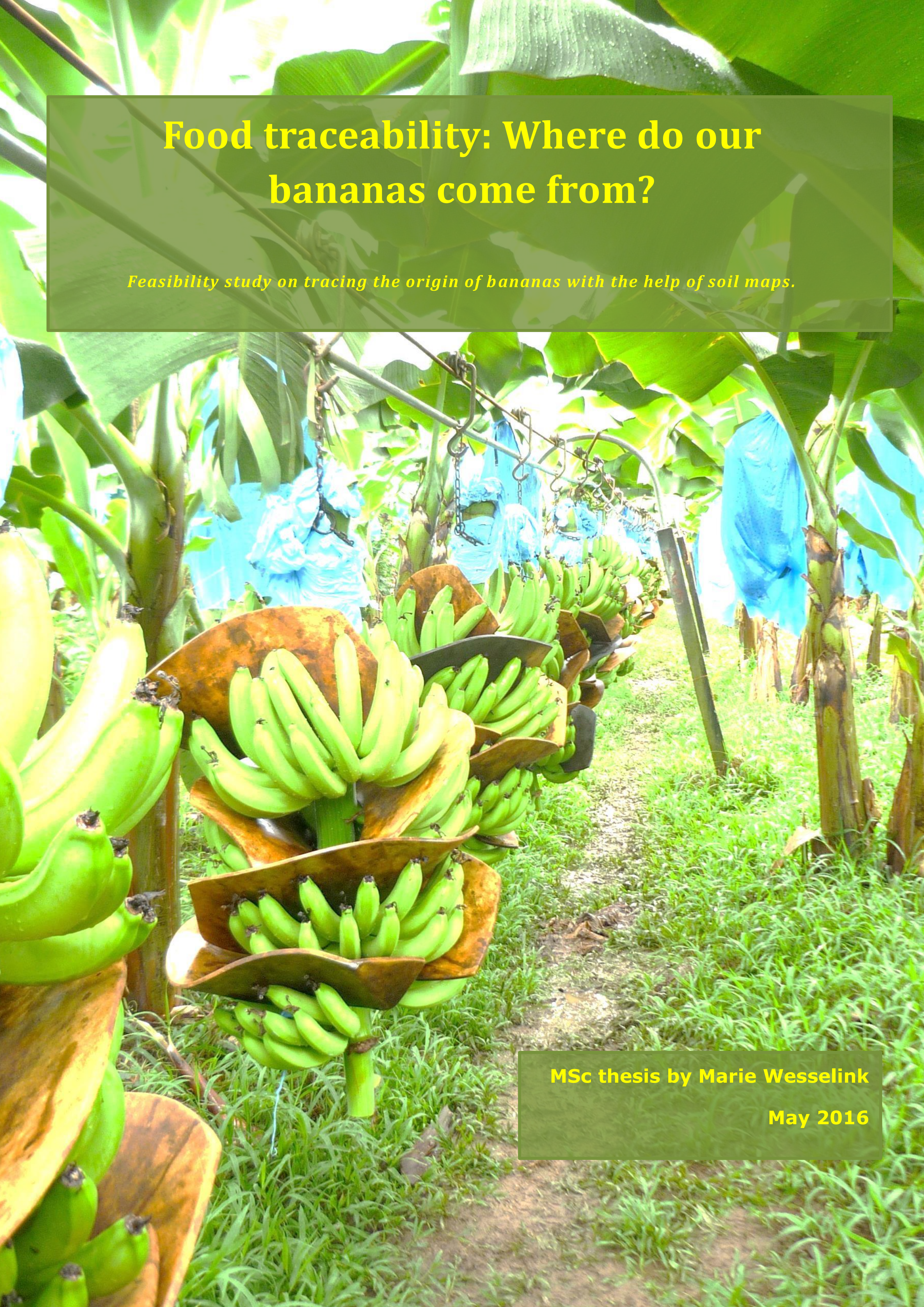

Food traceability: Where do our bananas come from?

Feasibility study on tracing the origin of bananas with the help of soil maps.

MSc thesis by Marie Wesselink

May 2016

Food traceability: Where do our bananas come from?

A feasibility study on tracing the origin of bananas with the help of soil maps.

Master thesis project for Master Earth and Environment (MEE) – Specialisation Soil Geography and Earth Surface Dynamics.

Marie Wesselink
910123942130
marie.wesselink@wur.nl

Supervisors:

Dr. Jetse Stoorvogel – Soil Geography and Landscape, Wageningen UR
Rafael Segura Mena MSc – CORBANA, Costa Rica

Abstract

Various track and trace systems exist throughout the world. The information on the origin of products is provided by the producer or seller of the product. For food products it is of importance that retracing the product is also possible, in the light of food safety issues. Consumers want certainty on where their food comes from and whether products labelled as organic are actually organic. This study explores whether it is possible to retrace banana plant material considering known soil map properties. Soil and banana plant sample data were gathered in Costa Rica. Soil samples were compared to various soil property strata which were derived from the soil map. Banana plant samples were compared to those same strata. Significant differences for soil and plant material with the strata were found. Predictability percentages for a correct prediction of plant material belonging to a certain stratum were calculated. Results showed that it is possible to predict where banana plant material originated. Although the step from banana plant material to the banana fruits still has to be made the results from this study show great potential for a retracing system for bananas based on soil map properties.

Acknowledgements

There are many people who helped me and contributed in the process of this research. First of all I want to thank my supervisors Jetse and Rafael. Jetse, thanks for your patience with me throughout the whole process, especially during report writing. Rafael, thanks for making me feel at home in Costa Rica and thinking along in what direction my research should go. Thanks also for showing me around, taking me on your work trips and inviting me over to try traditional Costa Rican dishes at your home.

There were several people in Costa Rica that contributed to (parts of) this research and all my experiences there. I want to special thanks Juan Samuels for taking me on numerous field trips, always patiently explaining everything going on and willing to answer all my questions. Thanks to all the other people at CORBANA that I was somehow involved for data gathering or lab work. Thanks to Wendy, Alvaro, Sheyris, Ana Maria and Sofie for all the breakfasts, lunches and dinners together!

At last, thanks to everyone in the SGL thesis room, for the moral support and the frustration moments we had together. Thanks to Christian, for being there the whole period of my thesis, which was quite long, and for the daily distraction moments!

Table of Contents

List of figures

List of tables

1. Introduction

Track and trace systems are used throughout the whole world for a large variety of goods. Track and trace systems are set up by the producer or supplier of the goods to inform the customer where their purchased goods are during transport, or to check where they exactly come from. In the Netherlands a track and trace system is frequently used within the mail industry. When a package is to be delivered, you will receive a code to check online where your package is at the moment and when it will be delivered. Another example of such a track and trace system is used by Dole, the large international banana company. Dole bananas in the supermarkets have a sticker with a farm number on it, see Figure 1-1 a). On their website this number can be filled in and it will be revealed at what farm the banana was grown. Next to this additional information about the farm, climate and banana certificates are provided.

The responsibility of such a system is completely in the hands of the producer or seller of the product. As a customer there is no choice or to believe the information on the product you receive from the producer. To be able to check or retrace these products back to their origin a new approach is needed. To trace back the origin of your food product without relying on the information of the producing company a systematic control system needs to be set up. When only the actual product, like for example the banana without the Dole code, is available, somehow a linkage between the properties of the product and the conditions of the growing location needs to be made.

Amongst others, the food industry demands for such a control system in the light of food traceability and authenticity. Nowadays consumers demand information on where their products come from and under what circumstances they are produced. The health, safety and quality of a product are concepts consumers associate with traceability (van Rijswijk et al., 2008). Food safety is a very important issue, since when consumers get ill or experience complaints after consuming particular foods, it should be possible to trace the responsible parties. And what to think of the situation when deadly animals like spiders are found in a bunch of bananas? (Petre, 2014)

Producers and banana companies come up with more and more quality labels for their bananas, see also Figure 1-1. All kinds of certifications are introduced to convince the consumer of the quality or sustainability of the banana production. The Rainforest Alliance is one of those certifications. This certification stands for a certain degree of sustainability, promoting conservation of the environment and the biodiversity (van der Waal & Moss, 2013). In general a higher price is asked for such certified products, and consumers are willing to pay a slightly higher price compared to non-labelled products (Padilla et al., 2007; Velcovska, 2012).

Another way to justify higher prices is by claiming that your product has a certain added value over other similar products. Think for example of organic production (Figure 1-1 b)) or the claim that your product originated in a certain region. This second example is especially relevant for products where the origin location determines part of the price, for example wine (Versari et al., 2014). For bananas the claim of being organically produced will play a larger role.

Every producer will always want that his product stands out in comparison with his competitors. It might therefore be tempting to add a certain label or qualification to your product that is actually not true. Over the past decades a few large food safety / fraud incidents have occurred in the meat industry (Jonge et al., 2004) and the powder milk industry (Linhai et al., 2011). These fraud issues makes that consumers are more suspicious towards the quality of their food. This highlights the importance of a retracing system for food.

Figure 1-1. Banana labelling. From top-left to bottom-right: a) A Dole banana with a traceability code. b) A fair-trade Max Havelaar banana. c) A certified Rainforest Alliance banana. d) A bio-organic labelled banana.

This retracing system needs to be based on general available environmental properties of the growing locations. To be able to establish a link between the product and its location worldwide it is desirable that this linkage is based on characteristics that are available over a widespread range of locations all over the world. Examples of such characteristics are climate conditions and soil properties. Already in 1982 Turner and Barkus showed that climatic conditions have an impact on a plant's nutrient status (Turner & Barkus, 1982). The nutrient status of a soil, whether it is natural or with additional fertilisation, is also partly reflected in plants grown on top of it (Bugaud et al., 2009; Johns & Vimpany, 1999; Moreira et al., 2011).

To test the feasibility of a so-called retracing system banana plants and their soil from plantations in Costa Rica are used. Bananas are an ideal crop for this case study because unlike other crops they show almost no internal genetic variation. All commercially grown bananas are the same clone, which is called Cavendish (Robinson & Galán Saúco, 2010). Therefore it can be assumed that variation in banana composition can be ascribed to external influences like the conditions of the soil they were grown on and internal variation is assumed to be negligible.

To summarise: there is a demand for a control or so called re-traceability system for food. Such a check should be possible when local conditions as for example the soil composition is known.

Relationships between those conditions and the full-grown (export) product are still under investigated.

To be able investigate and work on a solution for the above posed problem statement the following research questions were raised:

- Can banana plant characteristics reveal information on their origin?

To step by step come to an answer for this question two guiding sub questions are proposed:

- Is the available soil map of Costa Rica representative for the local soil conditions?
- Is banana plant material grown on different soils significantly different from each other?

2. Methodology

2.1 General methodology

In the complete process of banana growth until the moment it ends up in the supermarket there are various factors that can influence its composition. It all starts with the environment. Important environmental factors which might affect the banana composition are the soil it grows on (1) and the climate. A banana plant will grow (2) if it is situated in the appropriate surroundings and managed well. The plant will produce the fruits (3) which will be picked and exported once they are full-grown. As a last step after transport they will be ripened (4) and brought to the supermarkets. The composition of a banana can be explained as being a function of these four steps. In addition to the just mentioned steps also the variety and the management could play a substantial role as well.

In this study we will do a first investigation focussed on the linkage between the soil and the banana plant. This means that of the previously explained steps number 1 and 2 are examined and 3 and 4 are left for what they are. Next to that, the variety of the banana plants and the management practices are not taken into account as a factor of potential variation. This is justified by the fact that all data used comes from large export banana plantations. All export bananas currently are the same variety, regardless of to which part of the world they are exported. Management is of course dependent on local (soil) conditions but assumed to be quite uniform over large commercial banana plantations. In addition to that management details are not openly available for most locations.

To in the end be able to retrace the supermarket bananas step 3 and 4 are important as well. This is beyond the limits of this research. For now it is assumed that a good start can be made investigating the first two steps. This is relatively easy and when positive results are found this will be promising for the further steps.

As an end result it is desired to link the banana (-plant) to general available information. Establishing a link between a soil sample and a banana plant sample is therefore not directly the most useful solution. Soil sample data might not be openly available for locations all over the world. In the end the desired result is to be able to link the banana (in this case the foliar data) to available soil or climate information. For locations all over the world maps are available. Soil maps might not be at the same level of detail everywhere, but this is of minor importance. Most important for this study is that relations can be found at a general level of detail, or that we can exclude certain areas to be the originated area of the banana plant.

The same goes for climate data. All over the world rainfall levels are recorded. This data is available through weather institutes, in many cases even for free. A general aim of this research is to find a relation between the banana plant and such simple information which is freely available everywhere. These relations could then later be extrapolated to the bananas itself.

To see whether the available soil map is at all representative for the area a check was done with the soil sample data. There does not necessarily have to be a connection between the two. Soil maps are built-up with knowledge on the origin or age of the material, occurrence of certain processes or general characteristics for example. Soil samples analysed in a laboratory mostly contain information on concentrations of particular elements within the sediment. It is possible or in some cases even likely that areas that get a different classification according to the soil map have about the same Aluminium levels when samples are analysed in a laboratory for example. When there is no linkage found between soil samples and the map it is highly unlikely that there will be a significant relation between the plant samples and the map since variation in the plant samples will most likely be due to variations in the soil.

On the other hand, no relation between the map and soil samples does not yet have to be conclusive for the plant data. Therefore the next step will be to check the foliar sample data with the soil map. It could be argued to add a check of the soil sample data directly with the foliar sample data. This was not done in this investigation. Our goal is to find relationships between the plant material and readily available soil information like a soil map. When retracing a banana bought in the supermarket there is also no soil sample data available, but soil maps are available widespread all over the world.

2.2 Study area

The study area is the Atlantic Zone in Costa Rica. The Atlantic Zone covers an area of about 5400 km^2 in the north eastern part of the country. Costa Rica is an important country in the banana production, the third largest exporting country in the world according to (FAOSTAT, 2010), obtained via (Robinson & Galán Saúco, 2010). Bananas are mainly produced in large plantations and exported by multinationals like Chiquita, Del Monte and Dole (Arias et al., 2004). Most banana plantations are situated in the northeast side of the country, the so called Atlantic Zone, see also Figure 2-1. The large export plantations all maintain more or less the same production technique, which is a large scale monoculture production with the use of high rates of fertilisers and pesticides (Bellamy, 2013). There are only a very few farms organically producing bananas, those plantations are small and oriented at the local market.

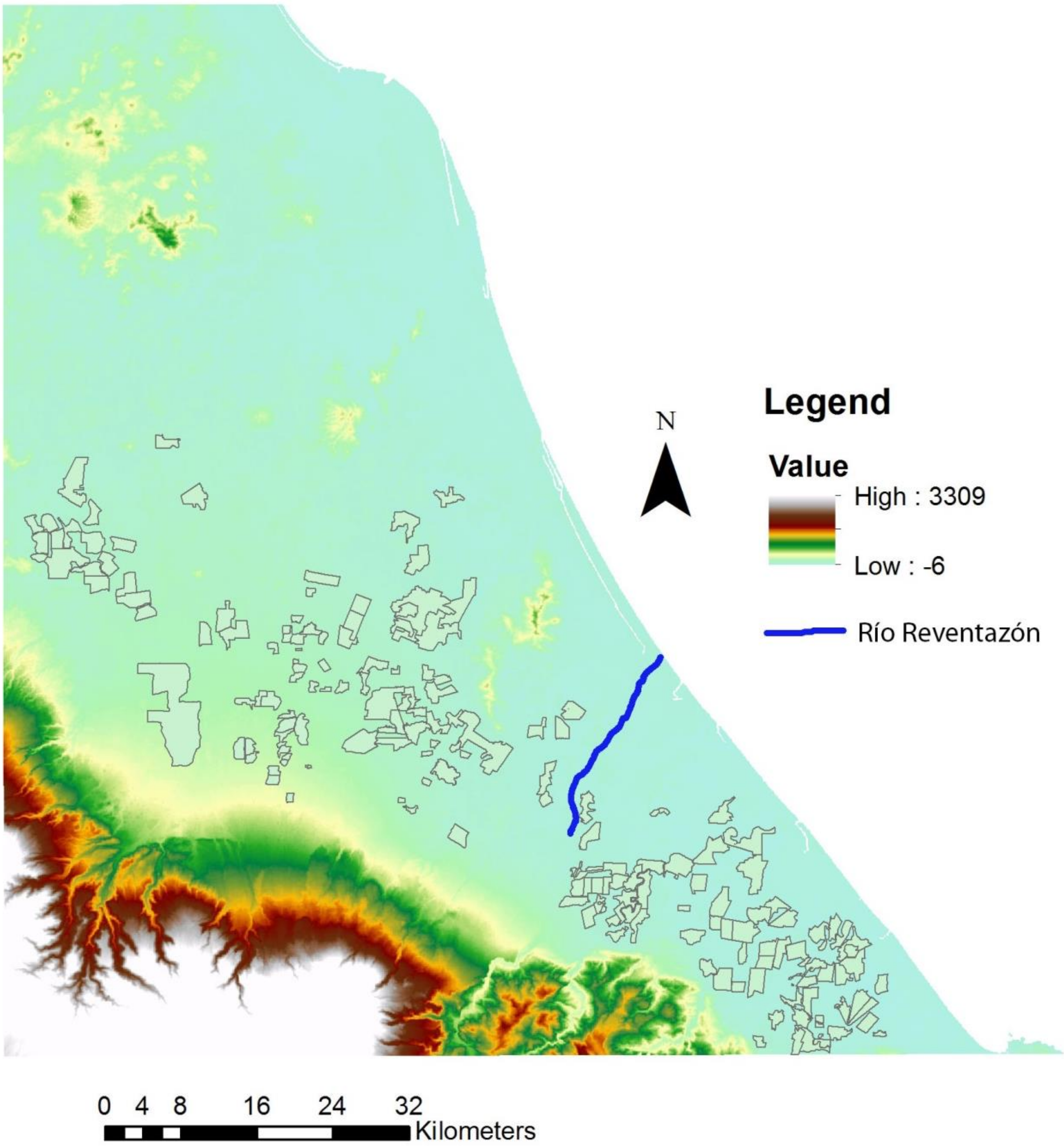

Figure 2-1. Elevation map of the Atlantic Zone of Costa Rica with banana plantations drawn in.

Within the study area there is variation in soil characteristics. In terms of soils the area can roughly be divided in two parts, the division line being the Rio Reventazón. This river is drawn in in the map in blue. Soils north of this river have parent material with a volcanic origin in contradiction to soils south of this river which have mainly sedimentary rock as origin (Nieuwenhuyse, 1996). The northern volcanic soils get fresh volcanic material from volcanoes that were recently active. These soils are therefore relatively young. South of the Rio Reventazón soils can be much older, and even some very old tropical red soils are found.

Temperature shows no clear seasonality in the Atlantic Zone. Differences between the minimum and maximum temperature are about 2°C (IMN, 2014). Rainfall rates do vary widely per month, but also per region. The regional distribution is illustrated in the map in Figure 2-2. Within the Atlantic Zone the annual rainfall levels vary from 3000-3500 mm up to 6000-8000 mm.

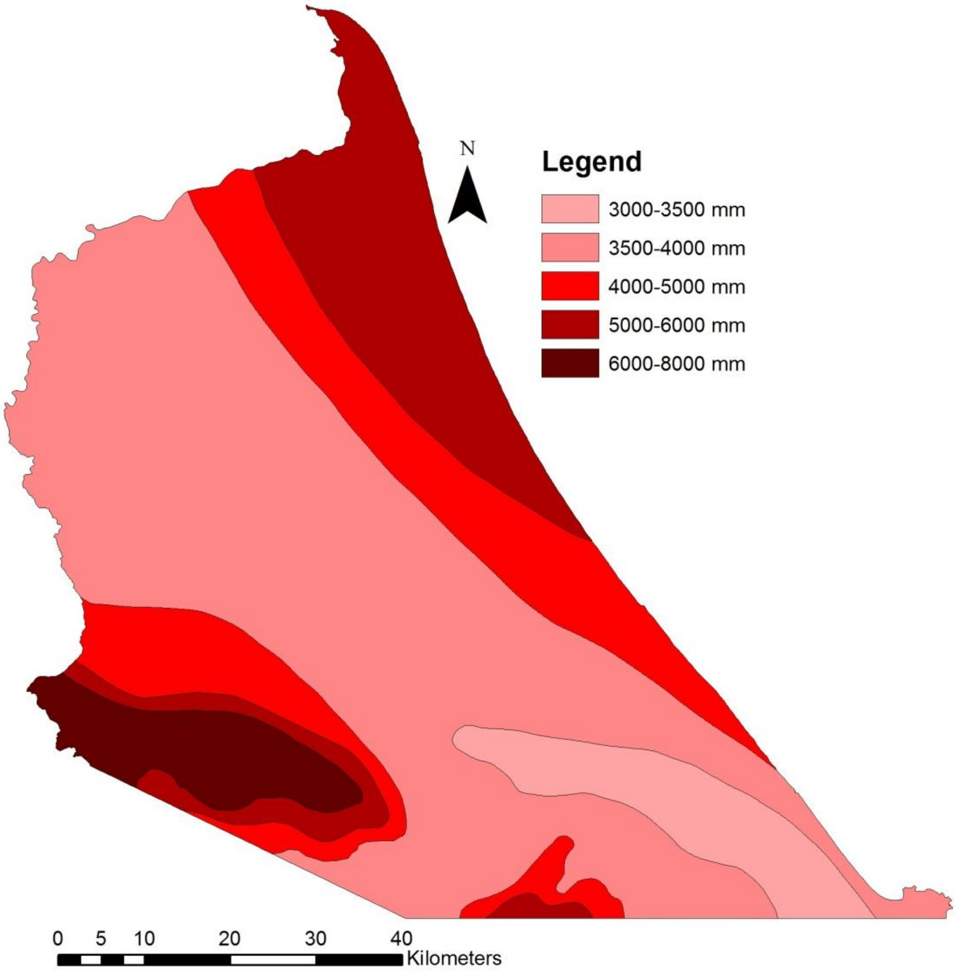

Figure 2-2. Yearly rainfall pattern for the Atlantic Zone.

2.3 Methodology step 1: Soil data and stratification

As mentioned in section 2.1 we first investigated the relation between soil sample data and the soil map. The available soil map will therefore be stratified three times. The first stratification will be based on soil types, the second on texture and the third time stratification will take place on andic properties. A fourth stratification was done on annual rainfall levels. Within all four stratifications the strata were analysed on available soil sample data. In the subchapters below all those elements are highlighted and explained in more detail.

2.3.1 Soil map

A digital soil map of the Atlantic Zone was available for use. The map is built-up from associations. This means that one delineated unit most probably consists of more than one subunit. The following example illustrates this concept: imagine the texture map; one unit might be pictured as being clay, while it actually is only 70% clay and 30% another texture. This unit is then an association of a clay and a sand unit. From the map it is not readable where exactly the clay or sand patches are located. The data behind the map just states in numbers the relative area percentages per unit.

Figure 2-3 shows a scheme of the built-up of the used soil map. The largest unit level is the mapping unit (MU). Every MU has one or more accompanying terrain unit (TU). Every TU has one soil unit (SU). These SUs all have a value for 14 different soil properties amongst which are: texture class, depth, pH and andic properties.

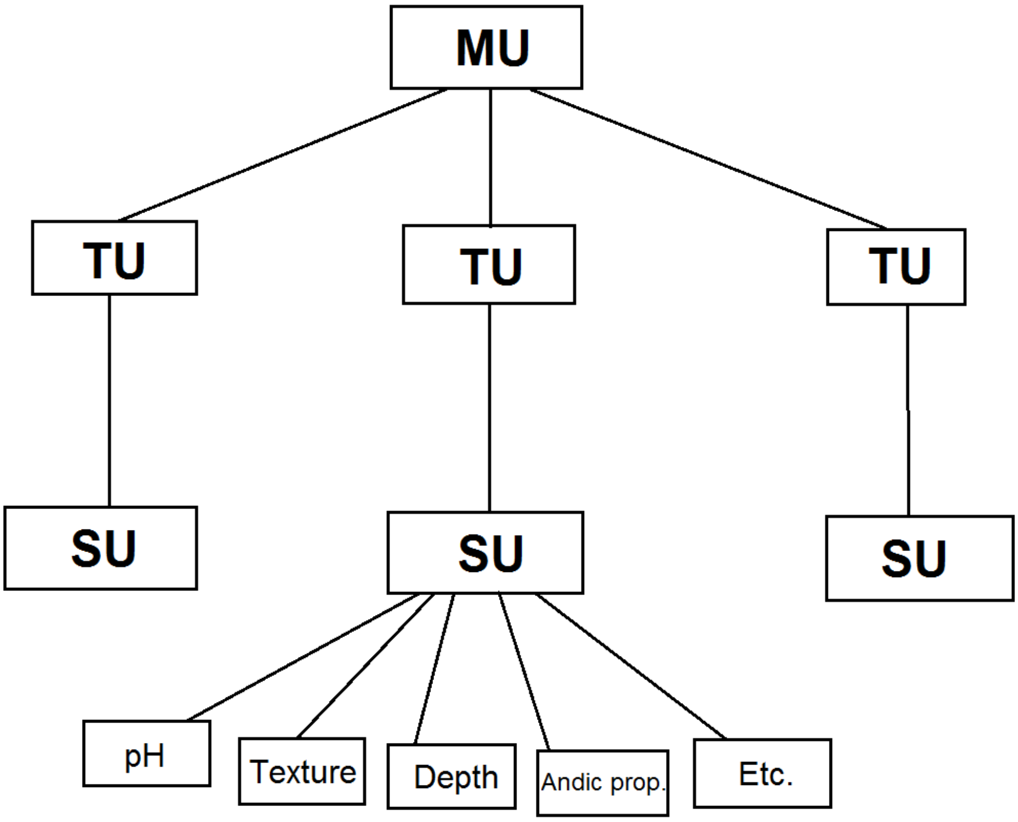

Figure 2-3. Schematic overview of how the units in the soil map are built up.

The 'actual' soil map is shown in Figure 2-4. Actual is put between apostrophes here, since the units displayed are not necessarily 100% that unit but an association of more units. The displayed unit in the map is the major unit.

2.3.2 Stratification of the study area

Properties of a soil map can be spread over a wide range of values. Take for example the soil texture which can vary from pure sand to heavy clay, and multiple mixtures of those in between. To be able to do a more general analysis that makes sense the soil properties were stratified over a maximum of 5 strata. Stratification of the soil map was based on general soil type, texture class and andic properties. The rainfall map was also used for stratification.

For every farm's acreage the relative area falling in a certain stratum was calculated. These relative area percentages are not per definition similar to the relative SU area of a farm. They can be the same, or higher. This is caused by the fact that different texture classes were combined, and SUs can differ in their andic classification but at the same time be similar in their general soil type. The rainfall map is not built-up of associations, therefore it works more straightforward; the relative area falling in a mapping unit represents the actual area for that stratum directly.

It could be possible that a farm has its area spread equally over more strata, up to an area of 20 percent in all five texture strata for example. For the analysis to make sense it was chosen to only use farms that have the majority of their area in one stratum. To quantify this, the lower boundary was set at 75 percent. So, a farm is only considered to represent a stratum if at least 75% of its area actually falls within it. This has as a direct consequence that possibly some farms will be omitted for some stratifications because their area is equally spread over two or more strata. In that case no major stratum can be appointed. All the farms that do have 75 or more percent of their area in a similar stratum are taken together for further analysis.

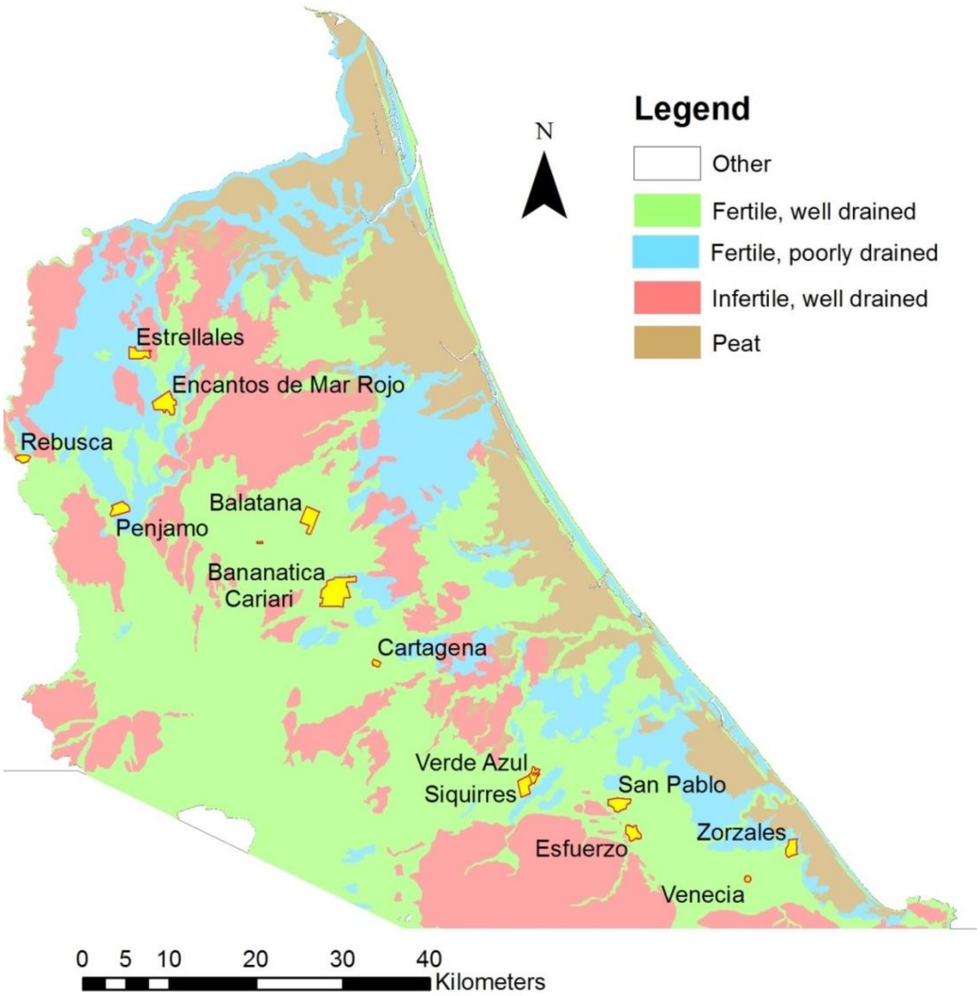

Figure 2-4. Generalised soil type map of the Atlantic Zone of Costa Rica with the banana farm locations drawn in.

2.3.3 Soil sample data

Soil sample data were provided by CORBANA. CORBANA is the Corporacíon Bananera Nacional, or in other words the national banana cooperation of Costa Rica. This cooperation has done research on various banana plantations in Costa Rica for many years. Therefore they have an extensive collection of soil data of banana plantations all throughout the Atlantic Zone. In total the data of 14 farms was used. An overview of the amount of samples per farm can be found in Appendix I. The locations of all the 14 farms are shown in the map in Figure 2-4.

Every sample contains information on 14 soil properties. The soil properties that were taken into account for analysis are OM, Zn, Mg, Cu, Fe, Mn and Al. A detailed overview of the measured soil properties is given in Appendix II.

B and S measurements were not taken into account since far less than 10% of all samples had measurements on these elements. N, P and K were not analysed since those three nutrients are important if not essential for growth of banana plants and are therefore frequently anthropogenically added to the soil as fertiliser (Robinson & Galán Saúco, 2010). This makes that the levels of those nutrients will be more or less similar on all plantations, regardless of their location. Ca is regularly added to banana plantation soils to counteract the acidifying effects of fertilisation. Ca is therefore not analysed. With the Ca treatment a certain optimal pH level is maintained and pH is for that reason also not analysed.

2.3.4 Soil data statistics

The geographical locations of all the chosen banana farms were digitally available. The farm locations are depicted in Figure 2-4. The soil sample data were linked to these areas so they could be compared to the soil map. By intersecting the farm locations with the soil map data a table was retrieved that reveals information on which part of what farm is situated in which MU. With this table a relative area (in percent) was calculated for each part of a banana farm that belongs to a different MU.

For all MU areas the link was made to the associated TUs and accompanying SUs with the soil properties information. The area of one SU within a farm was calculated by multiplying the relative area (MU) and the fraction of the particular SU in that MU. If there were more areas of the same SU within one farm, they were combined, regardless of the MU. This could be done since only the properties related to the SU are of importance for the analysis.

Once all the strata were defined and the area of each farm in those strata was calculated a statistical analysis was applied. This was done to see whether the stratification actually makes sense considering the available soil sample data. In the statistical analysis the values of a measured soil property were compared per stratification. So this was done four times; for the general soil types, for the texture, for the andic properties and for the annual amount of rainfall.

To test if certain correlations or differences exist, for all the above described strata the mean and standard deviation was calculated for a set of data variables. Also the significance of the difference between those values was calculated. This was done with a Welch Two Sample t-test. The outcome of this test is a certain p value. In principle this t-test works with only two variables (in this case that would be the strata). When more than two strata needed to be compared, all were one on one compared (e.g. 1-2, 1-3, 2-3). The highest p value (lowest significance) was then chosen. When only two strata are to be compared they will automatically have the same p-value.

2.4 Methodology step 2: Plant data and strata

2.4.1 Leaf sample data

At all the banana farms where soil samples were taken foliar samples were taken as well. Soil and foliar samples were not always taken simultaneously, sometimes only the soil was sampled while another time only the banana leaves were sampled. This makes it possible that the amount of foliar samples on a particular farm is not equal to the amount of soil samples. This is shown in the table in Appendix I.

For the foliar data the exact same set of components was chosen as for the soil data, except for the OM since this is not measured in plants. The whole set of available foliar components is given in Appendix III. A main reason to use the same components is to be able to compare the results between the soil and the plants. Since the analysis is all about relative differences and comparisons of the data it is of no importance that the unit in which a component was measured differs between plant and soil data.

2.4.2 Leaf data statistics

The data of the foliar analyses was overlaid on and compared with the soil map, just as was done with the soil sample data (explained in section 2.3.4).

Next to that a predictability analysis was done with the use of the calculated means and standard deviations. If the values of a foliar component differ between the set strata it is interesting to see whether it is possible to predict the other way around. In other words, how big is the chance that the origin of banana leaves is correctly predicted when comparing foliar analysis with soil properties? Since for the banana data in this research the origin is known the correctness of the prediction can be checked directly.

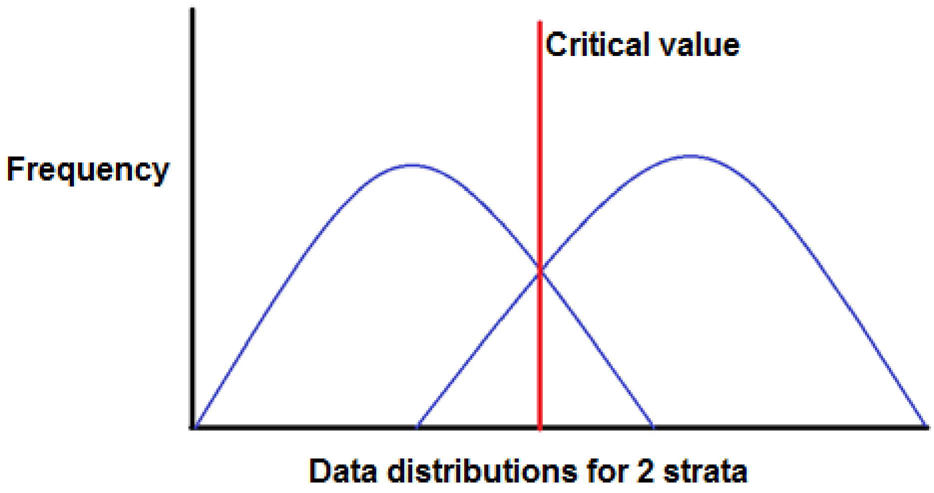

Figure 2-5. Example with the data distributions of two strata (blue lines) and the arbitrary critical value.

Rules were set up for values to fall in a specific stratum. When there were only two possible strata to distinguish between a critical value was defined between those strata distributions, based on the mean and standard deviation. In a graph it could look like Figure 2-5. With three or more strata divisions were made between every individual stratum. A possible result concerning stratification into three strata could be: all values lower than X belong to stratum 1, values between X and Y belong to stratum 2 and all values higher than Y belong to stratum 3.

The data of the banana farms that were assigned to be a certain stratum are then checked to see how many of the sample points actually fall below / above the set critical value and are thus correctly predicted.

3. Results and discussion

In this chapter all the different map stratifications are displayed along with the area distribution of the different farms over the strata. This is followed by statistics for all the strata, and the plant predictability analysis. Results are directly discussed; a general discussion follows at the end of the chapter.

3.1 Strata

3.1.1 Results

General, easy distinguishable strata were defined. Stratification was done for soil type, texture class and andic properties. Table 3-1 shows the strata. In total 4 times 3/4/5 strata were defined. To come to these strata some original strata were reclassified. An overview of the original classification of the soil types, texture classes, andic properties and annual rainfall is given in Appendix IV.

Table 3-1. Stratification of the Atlantic Zone.

Soil type	Fertile well-drained	Fertile poorly-drained	Infertile well-drained	Peat	-
Texture class	Sandy	Sandy loam	Silty	Clayey loam	Clayey
Andic properties	Andic	Partly andic	Non-andic	-	-
Rainfall (mm/year)	3000-3500	3500-4000	4000-5000	5000-6000	6000-8000

3.1.2 Discussion

During stratification banana preferences were not taken into account. Strata were defined by taking comparable textures (for example) together. It might occur that no banana farms are situated on pure clay soils because those are too dense to root in for the plants for example. In that case those strata are not further analysed in this research. It should be kept in mind that it could be coincidence that no farms occur in certain strata. It does not necessarily indicate that there are no banana farms in that stratum.

3.2 Results step 1: Soil data and strata

3.2.1 Soil type strata

3.2.1.1 Results

Figure 2-4 shows the four soil type strata. By means of an intersect with the map and the farm locations the relative area percentages per stratum within each farm were calculated, see Table 3-2.

Table 3-2. Overview of farms with area distribution (in %) per soil type stratum.

Farm	% Fert. well-drained	% Fert. poorly-drained	% Infert. well-drained	% Peat
Balatana	74.4	25.3	0.3	
Bananatica	58.5	41.5		
Cariari	98.1		1.9	
Cartagena	100			
El Esfuerzo	90.8	9.2		
Encantos de Mar Rojo	48	52		
Estrellales	48.2	41.3	9	1.5
Penjamo	36.6	63.4		
Rebusca	32.8	30	37.2	
San Pablo	11.5	88.5		
Siquirres	94	6		
Venecia	70	30		
Verde Azul	100			
Zorzales	63.4	36.6		

From this table the farms that have an area of 75% or higher in one stratum were used. No farm has a majority of its area in the infertile or peat stratum, so those were not analysed any further. Farms from which the soil data was used are:

Fertile well drained: Cariari, Cartagena, Esfuerzo, Siquirres and Verde Azul.
Fertile poorly drained: San Pablo.

All the calculated means and standard deviations with accompanying significance levels are presented in section 3.4 in Table 3-9. Column 2 and 3 of this table show the soil type strata, the rows give the statistical values for the soil sample data. The soil strata differ most in Mg levels. Fertile well-drained soils have a mean Mg level of 3.9 cmol+/L and fertile poorly drained soils 8.0 cmol+/L. The standard deviations are 2.0 and 1.1 for fertile well-drained and fertile poorly drained soils respectively.

3.2.1.2 Discussion

The infertile well drained and peaty soils were really underrepresented in the banana farm areas. When a farmer is given a choice he will grow his bananas on fertile and/or better drained soils. Only one farm (Rebusca) has a relatively large part in the infertile well-drained stratum, 37.2 percent. It might be that the farmer deals with this lack of fertility by adding extra or different fertilisers, or the infertile parts of the plantation were not used.

During the analyses in this research the outer borders of banana farms were used as outline of the farm area. It was assumed that the complete area within these borders was used for banana production. Therefore this complete area was used for the investigations in this study. However, from field observations it is known that not every part of every farm is always used. It was observed that some farms have part of their area lie fallow, for various reasons. It could be temporarily, because the area needs to be replanted with new young banana plants, but it might also be the case that a certain area within the farm its borders is unsuitable for the growth of bananas due to local soil conditions, see also the pictures in Figure 3-1a & b.

Figure 3-1a & b. Part of the San Pablo banana farm. This soil is currently intensively re-worked and fertilised to be able to be taken into production for the growth of bananas.

The soil map initially classified parts of the Atlantic Zone of Costa Rica as being poorly drained. There is a footnote to be made to this classification. It might be that a location or particular soil area originally was poorly drained, but nowadays there are various techniques to counteract those drainage conditions. All banana farms visited had some sort of drainage system with a network of drainage ditches to be able to react to intense rainfall events. Anthropogenic influences have enforced changes to the original natural soil conditions. This may induce that the separation between poorly drained and well drained soils is not actually present in the banana farms. This could explain the little variation found in soil data between the two strata.

3.2.2 Texture strata

3.2.2.1 Results

When the banana farms were intersected with the texture strata as posed in Table 3-1, the relative areas were calculated (Table 3-3). Notice that the area percentages of the farm Estrellales do not add up to one hundred percent. This can be explained by the fact that 1.5 percent of its area was classified as peat. Peat has no texture and was left out in the texture stratification.

Table 3-3. Overview of farms with area distribution (%) per texture stratum

Farms	Sand	Sandy loam	Silt	Clay loam	Clay
Balatana	6.2	60.5		33	0.3
Bananatica	8.4	73.4		18.2	
Cariari		75.2		22.9	1.9
Cartagena		70		30	
Esfuerzo		90.8		9.2	
Encantos de Mar Rojo	12	24		40	24
Estrellales	2.8	42.6		38.5	14.6
Penjamo	4	45.1		50.9	
Rebusca				62.8	37.2
San Pablo		35.8		64.2	
Siquirres		94.5		5.5	
Venecia				100	
Verde Azul		100			
Zorzales		11.6		88.4	

Not one of the banana farms has any percent of their area in the silt stratum. The farm area occurrences in the first and last stratum are minor, with a maximum up to 37 percent. Major farm parts are located in either the sandy loam stratum or the clay loam stratum:

Sandy loam: Cariari, Esfuerzo, Siquirres, Verde Azul
Clayey loam: Venecia, Zorzales

The two texture strata differ most in Mg content. For the other soil properties the means are rather close and/or they have substantial standard deviations.

3.2.2.2 Discussion

The third stratum contains the silty soils. None of the farms is situated on such a soil. When the soil map is checked it was found that the silty soils are mainly found in the far east of the Atlantic zone close to the ocean, or in the far south-east. There are no soils with this kind of texture in the regions where the banana farms are situated. Bananas can grow on silty soils, research even showed that a higher silt content makes banana plants less vulnerable for diseases (Freitas et al., 2016). Most likely there are other reasons why these areas are hardly used for banana production.

3.2.3 Andic strata

3.2.3.1 Results

Also for the andic strata the map was overlaid and intersected with the farm locations. This gave the area distribution amongst the strata as found in Table 3-4. Every stratum is represented with at least one farm that has a relative area larger than 75 percent.

Table 3-4. Overview of farms with area distribution (%) per andic stratum.

Farms	Andic	Part andic	Non-andic
Balatana	29.2	56.6	14.2
Bananatica		91.7	8.3
Cariari	75.3	1.9	22.8
Cartagena	70		30
El Esfuerzo			100
Encantos de Mar Rojo	8	40	52
Estrellales	14.2	43	42.8
Penjamo	10.9	63.4	25.7
Rebusca		37.2	62.8
San Pablo		84	16
Siquirres		76.5	23.5
Venecia			100
Verde Azul		100	
Zorzales			100

The farms that were used per stratum are:

Andic: Cariari
Partly andic: Bananatica, San Pablo, Verde Azul
Non-andic: Esfuerzo, Venecia, Zorzales

The 3 andic strata differ significantly from each other for the properties Mg, Fe and Mn. Per property the means for the 3 strata are far apart and standard deviations are rather small (Table 3-9). Based on Mg or Fe or Mn it is easy to differentiate between the andic soil strata.

3.2.3.2 Discussion

Differentiate between areas based on their andic properties can be risky. Volcanic areas are very prone to changes over time. Next to that it should be kept in mind that andic properties do not occur worldwide, since there are not volcanoes everywhere. When banana plantations in other regions of the world are analysed where no volcanic activity occurs the results of this particular analysis cannot be used. Volcanic activity is however largely present in Costa Rica, so it will not raise any problems in this research.

3.2.4 Rainfall strata

3.2.4.1 Results

When the farm locations are overlaid on the rainfall map it looks like they all fall for 100 percent within one rainfall stratum, see also Figure 3-2. This is actually the case, visually every farm occurs only in one stratum and since the rainfall map of the Atlantic Zone of Costa Rica is not built-up out of associations this is true. With this map it is therefore not necessary to intersect and calculate relative areas per stratum for every farm. It was simply red from the map. The banana farms occur only in two rainfall strata.

3000-35000 mm/year: San Pablo, Siquirres, Verde Azul, Venecia and Zorzales.
3500-4000 mm/year: Balatana, Bananatica, Cariari, Cartagena, Esfuerzo, Encantos de mar Rojo, Estrellalas, Penjamo and Rebusca.

The spatial distribution of those farms in the strata is visualised in Figure 3-2 with different colours. The farms which receive less rainfall are depicted in light blue while the farms with higher annual rainfall are depicted in darker blue.

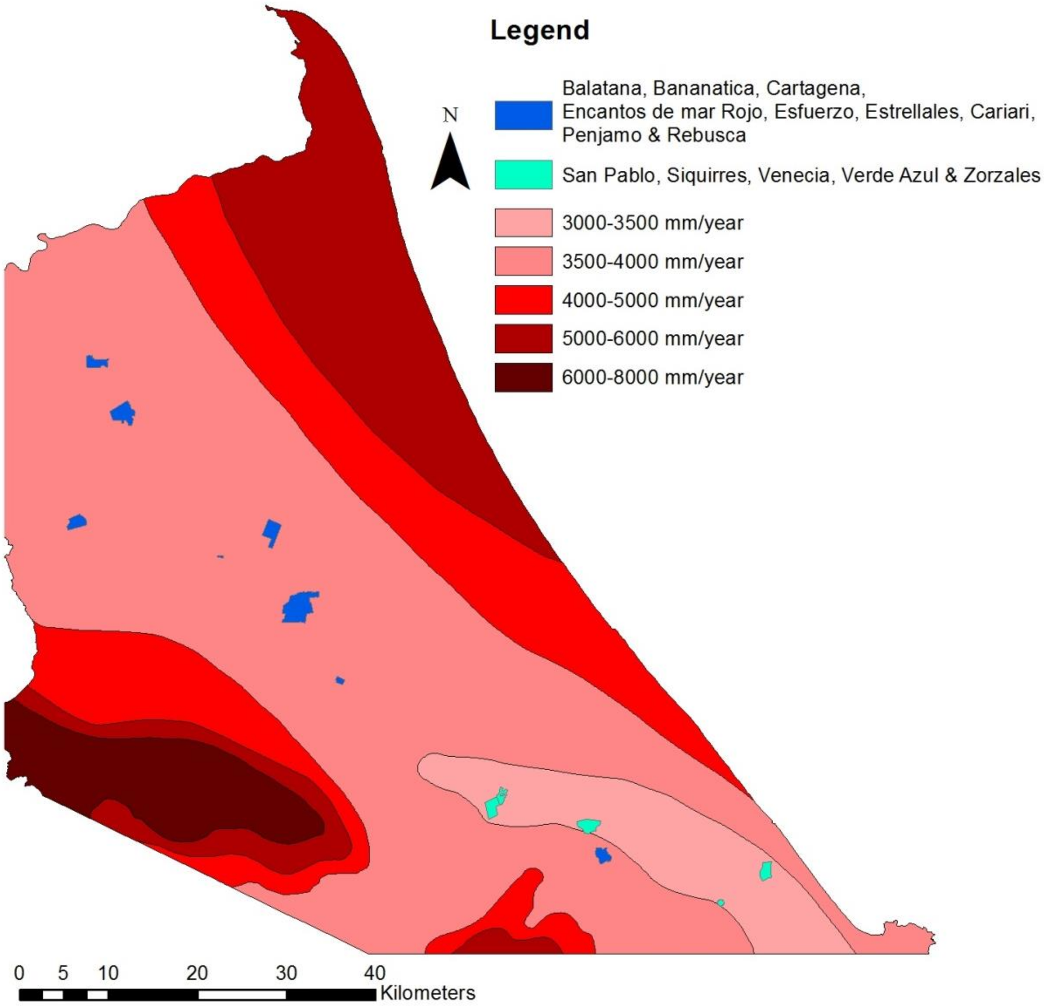

Figure 3-2. Map of different rainfall areas for the Atlantic zone of Costa Rica.

The soil property means are not very different for the two rainfall strata (Table 3-9) and if there is a difference the high standard deviation cancels out this difference. An exception here is the Fe concentration. For the lower rainfall stratum the mean Fe concentration is 278.3 ± 42.5, for the higher rainfall stratum it is 194.3 ± 39.3. These results are promising for the plant data for these strata.

3.2.4.2 Discussion

All the farms were divided over only two broad rainfall strata. Within a stratum the difference between the minimum and maximum annual rainfall is at least 500 millimetres. Ideally these strata would be subdivided in narrower rainfall areas. Possibly then a distinction could be made between the farm locations that now all fall into one stratum. This is a limitation of the available rainfall map. The strata as defined here cover the maximum possible level of detail given the available data.

3.3 Results step 2: Plant data and strata

In this section the results of the comparisons between the stratifications and the banana leaves data are presented. For all the strata the intersect was made with the farm areas, as already described in section 3.2. Also the foliar properties were analysed on their mean, standard deviation and a significance level was calculated. All these results are to be found in Table 3-10. In the subchapters coming up the results of the predictability analysis will be presented per property. The overall results of this are shown in Table 3-11. This table shows the percentage of the sample data that was correctly predicted with the set critical values. A colour scheme was used to make higher values stand out. All light green coloured cells represent correct predictability percentages between 60 and 70 percent, middle green colours represent percentages between 70 and 80 percent and dark green highlights percentages of 80 percent or more. There were never more than two strata that needed to be directly compared. A score of 50% would mean that it is equal where a sample would end up, so in that case there is no added value of analysing that particular plant component.

When the result of one analysis is for example 75% it means that 75% of the banana leaves from that specific stratum are correctly predicted to be originated there. The other 25% will be wrongly predicted (false negative prediction). Every banana that can be correctly predicted does not need screening anymore in case of a quality conflict. Only the wrongly predicted 25% needs an extra control check.

3.3.1 Soil type strata

3.3.1.1 Results

Only two of the four soil type strata contained relative farm areas of more than 75% per farm and were taken into account in analysis, see also section 3.2.1.1. The only property that could be used to distinguish between the strata based on the plant data is Mg. However, these values are also not extremely far apart. The means of the two strata differ 0.03 and the standard deviations are 0.03 and 0.04 for fertile well-drained soils and fertile poorly drained soils respectively.

To predict in which stratum a sample would end up a division between the two strata is made for every plant property taken into account. Table 3-5 shows the values chosen as critical value between the two strata.

Table 3-5. Set critical values between soil type strata for prediction analysis.

Plant elements	Value fert. well-drained	Value fert. poorly-drained
Zn	< 20	≥ 20
Mg	< 0.285	≥ 0.285
Cu	< 9	≥ 9
Fe	-	-
Mn	<190	≥190
Al	≥ 6.4	< 6.4

The highest and best results were obtained for the elements Mg and Cu (Table 3-11). A correct prediction could be done for 69 and 70 percent of the samples, for Mg and Cu respectively, for fertile well-drained soils. So to say, banana plant material with less Cu than 9 mg/kg will be predicted to have come from an area with the properties of a fertile well-drained soil. In 70% of these cases this will be correct.

3.3.1.2 Discussion

The statistics of the plant data match well with the soil data. For both the Mg levels are most suitable to distinguish between the two soil type strata. The difference between the values is however for the plants much smaller than for the soil data.

When looking at the values in Table 3-5, most of the values make sense compared with their means and standard deviations as in Table 3-10 except for the values for Fe and Mn. The Fe distributions for the two different strata were so alike that it was impossible to differentiate between the two and therefore no critical value is given.

Comparing the critical value of 190 for Mn with the means of 209 and 201, for fertile well drained and fertile poorly drained soils respectively, seems to make no sense. Further investigation showed that the means for Mn for the two strata were not so representative, because the data is not normally distributed. In Figure 3-3 the distributions are plotted. Both sandy loam soils (green) and clay loam soils (blue) have some outliers to the far right that are responsible for 'too high' means. The graph shows that in general the clay loam stratum data are a bit higher and that makes why the critical value was set as it is.

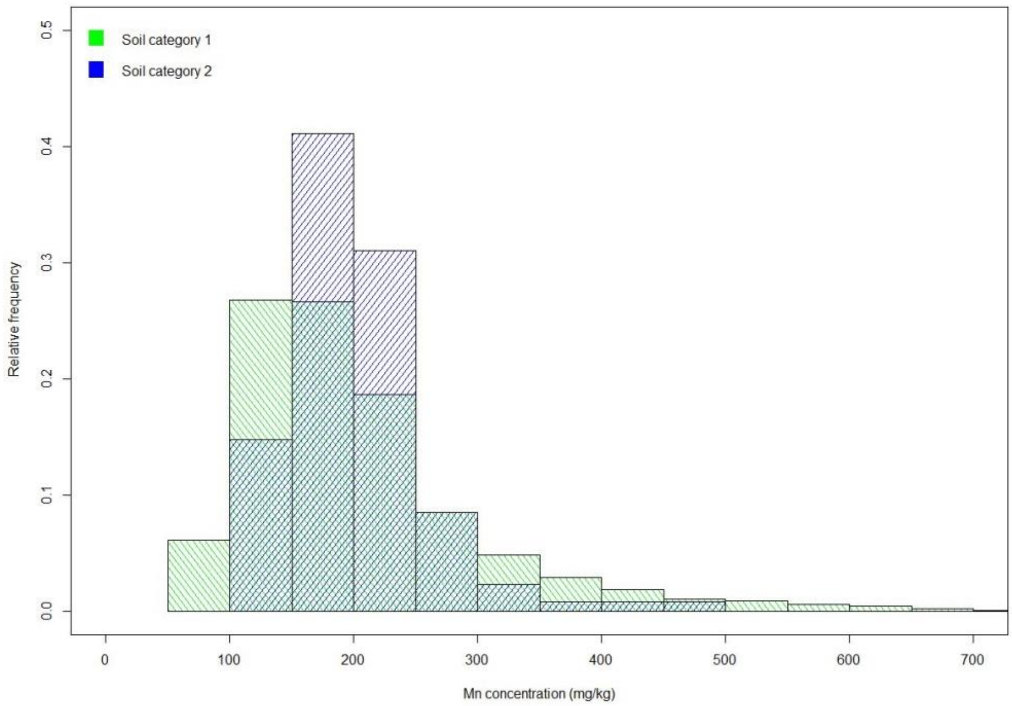

Figure 3-3. Histogram showing the distribution of Mn in plants for the two soil strata.

3.3.2 Texture strata

3.3.2.1 Results

The means and standard deviations for the two texture strata are not very far apart from each other, except for the Mg levels. This trend was present for the soil data (section 3.2.2.1) as well.

The critical values that were set for the two texture strata are given in Table 3-6. It was not possible to define critical values for the elements Mn and Al for the texture strata.

Table 3-6. Critical values between texture strata for prediction analysis.

Plant elements	Values sandy loam	Values clay loam
Zn	≥ 19.8	< 19.8
Mg	< 0.29	≥ 0.29
Cu	≥ 8	< 8
Fe	≥ 56	< 56
Mn	-	-
Al	-	-

For the four elements where critical values were set the prediction analysis results from Table 3-11 look promising. For all of them there is at least one stratum that has a correct prediction percentage of 60 or higher. For Mg the percentages of both strata are even 76%, so more than three-quarters of the samples were predicted correctly considering the Mg levels.

3.3.2.2 Discussion

Table 3-6 shows no critical values for the properties Mn and Al. For both of these elements the distributions for both strata were impossible to take apart. The standard deviations were for both cases much higher than the difference between the means of the two strata.

3.3.3 Andic strata

3.3.3.1 Results

Three andic strata were defined. All three contained relative farm areas of 75 percent or more. When analysing the plant data it was found that for all the elements two out of the three strata were indistinguishable. There was always 1 stratum standing out from the other two. It was however not always the same stratum that was different from the other two, see also the statistics in Table 3-10. For the prediction analysis the strata that were alike were then taken together for that specific element. In Table 3-7 the strata that were taken together are depicted in bold. For Zn and Cu the first and third stratum were taken together, for Mg, Fe and Al the second and third stratum and for Mn the first and second stratum were taken together.

Table 3-7. Set critical values between andic strata for prediction analysis. Values printed in bold are strata which were taken together, see text for more information.

Plant elements	Values andic	Values partly andic	Values non-andic
Zn	**< 19.8**	≥ 19.8	**< 19.8**
Mg	< 0.28	**≥ 0.28**	**≥ 0.28**
Cu	**< 8.65**	≥ 8.65	**< 8.65**
Fe	< 58	**≥ 58**	**≥ 58**
Mn	**< 205**	**< 205**	≥ 205
Al	≥ 8	**< 8**	**< 8**

Figure 3-4 gives an example of the just above explained phenomenon. It shows the plant Mg concentrations for the three andic strata. The partly andic stratum (blue) and non-andic stratum (red) overlap almost completely, while the andic stratum (green) could still be seen separately.

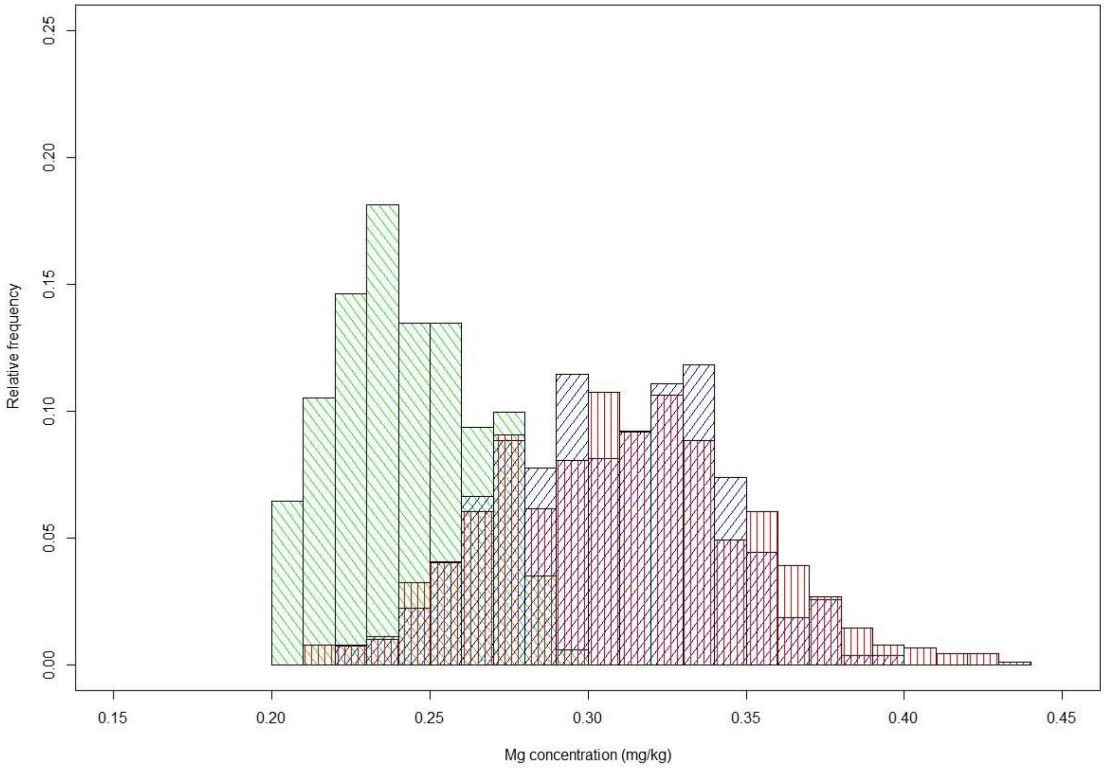

Figure 3-4. Histograms showing the distribution of Mg in andic strata.

Except for Mn there is for every (combined) andic stratum a predictability percentage over 60. There is even an 86 percent chance for a correct prediction of a banana plant to come from an andic soil when concerning Mg concentrations.

3.3.3.2 Discussion

When the plant element concentrations per stratum are compared with the soil data results it strikes that they do not at all align. Take for example the values for Fe for soil and plant data for all the strata. The means of the soil data for the three strata are far apart (143, 202 and 301 for stratum 1, 2 and 3 respectively, see also Table 3-9) with relatively low standard deviations. These values are not at all represented in the plant analysis. The second and third strata are almost similar in Fe concentrations and the first stratum has the highest Fe mean, while in the soil it had the lowest value.

3.3.4 Rainfall strata

3.3.4.1 Results

Also between the two rainfall strata critical values were tried to be defined to distinguish plant material between the two. As can be seen from Table 3-8 this only succeeded for half of the plant elements. Zn, Mg and Mn were indistinguishable from each other. This is supported by the mean and standard deviation values in Table 3-10. All the elements show only minor differences between the two strata.

Table 3-8. Set critical values between rainfall strata for prediction analysis.

Plant elements	Values 3000-3500 mm/y	Values 3500-4000 mm/y
Zn	-	-
Mg	-	-
Cu	< 8	≥ 8
Fe	< 58	≥ 58
Mn	-	-
Al	< 7	≥ 7

When comparing the values in Table 3-8 with Table 3-10 the critical value for Cu might be striking, because both means are higher than 8 but still the value is set at 8. The graph in Figure 3-5 gives some clarification. Both distributions show more outliers to the higher (right) side and are therefore not normally distributed. The means give a too high distorted image of both strata.

The prediction percentages as found in the last columns of Table 3-11 are not very notable. The highest percentage found was 68%, for Fe levels in the lower rainfall stratum. Four out of the six plant elements showed only very minor differences between the two strata. This follows logically from the statistics that were calculated (Table 3-10).

3.3.4.2 Discussion

A better result was expected to be found for the plant statistics for Fe. The soil Fe levels were far apart for the two strata and this same result was not found for the plant data. In the plant strata the means are 56.4 and 59.7 for the low and high rainfall stratum respectively. This is also striking because for the soil data the lower rainfall stratum had a much higher value. For the plant data this is turned around, albeit to a lesser extent.

The rainfall strata were proven to be not very useful to distinguish between banana plants. This was to be expected after the first soil data analyses. Literature does not reveal much about the influence of rain on soil composition. It is however known that rainfall intensities do influence crusting of the soil (Nciizah & Wakindiki, 2014). In their research they describe that this affects the time it takes for a seed to emerge above ground. Further effects on the plants were not found.

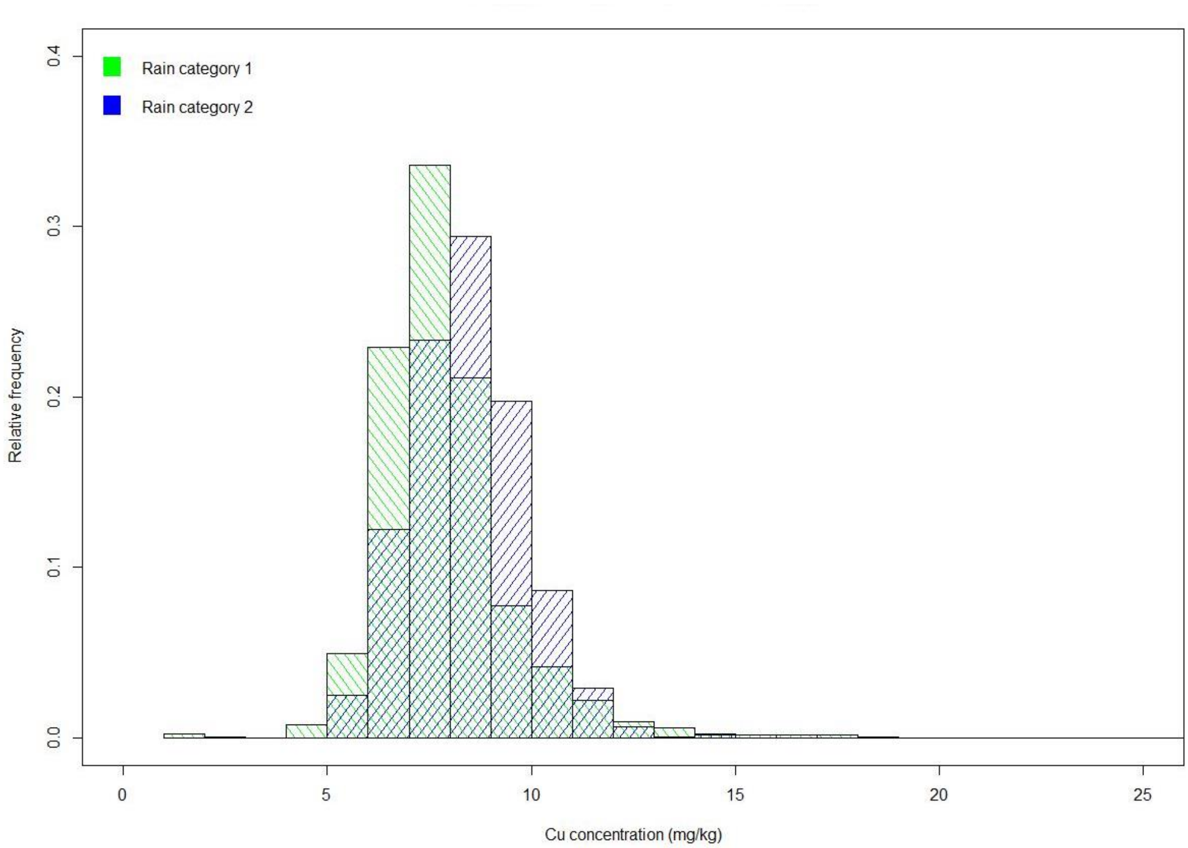

Figure 3-5. Histogram showing Cu distribution in plants for the rainfall strata.

3.4 Results: tables

The three tables below show the results as described and discussed in the sections above. The first two tables present the statistics. The significance levels (p-values) are represented with symbols. '***' stands for a p-value between 0 and 0.001. '**' stands for a p-value between 0.001 and 0.01. '*' stands for a p-value between 0.01 and 0.05. '.' stands for a p-value between 0.05 and 0.1. ' ' stands for a p-value between 0.1 and 1.

Table 3-9. Means, standard deviations and significance levels of soil elements per stratum.

SOILS	Fert. well-drained	Fert. poorly-drained	Sandy loam	Clay loam	Andic	Partly andic	Non-andic	3000-3500 mm/y	3500-4000 mm/y
OM	3.2 ± 1.4***	1.9 ± 0.57***	3.0 ± 1.4.	2.8 ± 0.8.	5.2 ± 1.1*	2.4 ± 1.1***	4.8 ± 2.4*	2.7 ± 0.7***	3.5 ± 1.3***
Zn	8.9 ±4.8*	8.2 ± 4.1*	9.0 ± 4.4***	7.0 ± 3.4***	8.0 ± 3.5**	6.0 ± 3.2***	7.2 ± 3.4***	8.4 ± 4.4***	6.8 ± 4.3***
Mg	3.9 ± 2.0***	8.0 ± 1.1***	4.4 ± 1.9***	7.0 ± 0.8***	2.4 ± 0.8***	3.5 ± 1.0***	7.2 ± 1.2***	5.7 ± 1.9***	3.5 ± 1.9***
Cu	4.8 ± 1.9***	7.8 ± 1.7***	5.7 ± 2.1*	5.5 ± 0.92*	3.4 ± 1.4***	5.2 ± 2.8**	6.0 ± 1.3**	6.1 ± 1.7	6.0 ± 3.8
Fe	207 ± 78***	257 ± 32.1***	242 ± 62.9***	308 ± 40.6***	143 ± 45.2***	202 ± 7.9***	301 ± 41.9***	278 ± 42.5***	194 ± 39.3***
Mn	38.6 ± 27.4***	52.5 ± 19.3***	40.0 ± 27.0***	58.5 ± 20.1***	19.8 ± 6.6***	30.6 ± 13.3***	65.3 ± 24.7***	45.6 ± 20.7***	33.2 ± 26.6***
Al	0.27 ± 0.36**	0.40 ± 0.69**	0.23 ± 0.34***	0.93 ± 0.86***	0.47 ± 0.46***	0.19 ± 0.32***	0.85 ± 0.84***	0.49 ± 0.73***	0.32 ± 0.45***

Table 3-10. Means, standard deviations and significance levels of plant elements per stratum.

PLANT	Fert. well-drained	Fert. poorly-drained	Sandy loam	Clay loam	Andic	Partly andic	Non-andic	3000-3500 mm/y	3500-4000 mm/
Zn	19.9 ± 3.6**	21.6 ± 4.9**	20.5 ± 4.6***	18.9 ± 3.7***	19.5 ± 3.5*	20.4 ± 4.1*	19.0 ± 3.6*	20.0 ± 4.6	20.3 ± 3.8
Mg	0.27 ± 0.04***	0.30 ± 0.03***	0.26 ± 0.03***	0.32 ± 0.04***	0.25 ± 0.02***	0.31 ± 0.03	0.32 ± 0.04	0.29 ± 0.05***	0.28 ± 0.04***
Cu	8.3 ± 1.3***	10.0 ± 3.4***	8.2 ± 1.4***	7.8 ± 1.2***	7.7 ± 1.2.	9.5 ± 2.5***	7.9 ± 1.2.	8.2 ± 1.8***	8.9 ± 1.8***
Fe	58.5 ± 9.8**	56.5 ± 7.6**	58.2 ± 14.3***	54.8 ± 8.5***	60.5 ± 10.2***	55.7 ± 7.6	55.0 ± 8.6	56.4 ± 12.4***	59.7 ± 9.8***
Mn	209 ± 115*	201 ± 54.9*	209 ± 88.6***	242 ± 154***	205 ± 63.3	201 ± 53.3	255 ± 148***	213 ± 121	213 ± 113
Al	6.9 ± 4.3*	6.3 ± 2.5*	7.1 ± 4.2**	8.0 ± 5.0**	10.2 ± 4.5***	6.2 ± 2.6***	7.6 ± 4.7***	6.9 ± 4.3***	7.9 ± 4.2***

Table 3-11. Correct predictability percentage (actual=predicted) for plants in the various strata. Cells highlighted light green are all percentages between 60 and 70%. The middle green colour represents percentages from 70 – 80%. Dark green cells are percentages of 80% or higher. X[1] represents values of the strata partly andic and non-andic taken together. X[2] represents values of the strata andic and non-andic taken together. X[3] represents values of the strata andic and partly andic taken together.

PLANTS	Fert. well-drained	Fert. poorly-drained	Sandy loam	Clay loam	Andic	Partly andic	Non-andic	3000-3500 mm/y	3500-4000 mm/y
Zn	55	54	62	57	64[2]	51		50	50
Mg	69	61	76	76	86	67[1]		50	50
Cu	70	60	63	52	75[2]	50		61	62
Fe	50	50	50	61	56	66[1]		68	56
Mn	53	57	50	50	57[3]		50	50	50
Al	49	57	50	50	77	61[1]		56	56

3.5 Overall discussion

The variety in the amount of samples per banana farm is extremely large; soil sample numbers are ranging from 79 samples (Estrellales) to 368 samples (Siquirres). In the foliar analyses the range is even more extreme; from 11 samples (Balatana) up to 519 samples (Zorzales). When sample quantities in strata become larger the relative difference declines since multiple farms are taken together in most of the strata. The Balatana farm is the only one that really stands out in a negative sense, but in none of the analyses the results are based only on this farm. Overall, the least number of samples used in one analysis is in the plant analysis for the fertile poorly drained soils, this stratum is represented solely by the farm San Pablo, with a total of 129 foliar samples. A number of 129 samples is still a reasonable amount to base conclusions on.

So the amount of samples does not fall short in any of my analyses. On the other hand, too much samples can cause a distorted image of the significance levels. An immense number of samples can cause even the smallest difference to be significant. This can also be seen in my data. Look for example at Table 3-10 at the plant Mg values for the rainfall strata. They are 0.29±0.05*** and 0.28±0.04***. The stars (*) indicate a p-value smaller than 0.001. By reasoning and by looking at the histogram plot of the two distributions (Figure 3-6) it can be concluded that the distributions are very similar. Sample numbers were 1414 (3000-3500 mm/y stratum) and 1383 (3500-4000 mm/y stratum). This large quantity of samples is able to create a false positive significance level. On the other hand, the significance levels calculated in this research were not of essential interest for the outcomes of the research. They were only used as a first insight in comparing strata.

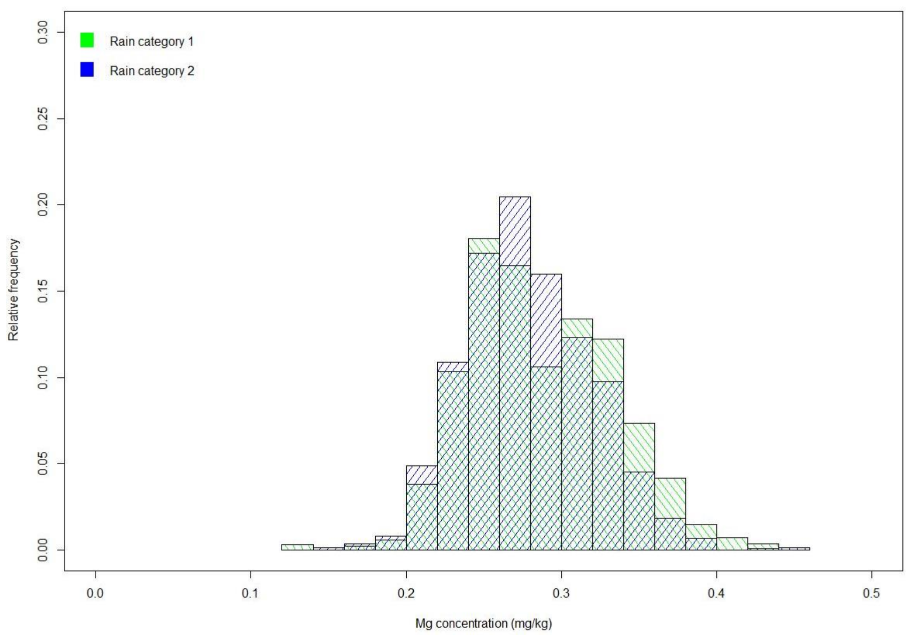

Figure 3-6. Histogram plot of Mg distribution in plants for the rainfall strata.

For all the soil and plant sample data a normal distribution was assumed. Some results already revealed that this was not always correct, for example look back at Figure 3-3. When taking a closer look at the distributions of the sample data to estimate critical values between distributions non-normally distributed values were found. Means and standard deviations are not applicable to non-normal distributions. These measures were however used in this research. It was chosen to continue the analysis with those means and standard deviations, since it was applicable to most of the data distributions. Next to that it was convenient to use one and the same simple technique to analyse all the data distributions.

As explained in the Methodology chapter already (section 2.3.1) the soil map is built-up from associations. To cope with this it was chosen to take a farm into account only if it had a relative area of 75% or higher in one stratum. Best results would probably have been obtained if only farms with 100% of their area in a stratum were used. This is practically impossible with an association map. It would make it hard to come up with a general algorithm that is applicable for all the bananas in the area. Banana farm areas that have a decent area percentage in several strata would not have been taken into account then.

When the rainfall strata are not considered (all farms fall completely in one stratum there) there are 5 farms that do not have a 75% or higher area percentage in either one of the strata. These 5 farms are: Balatana, Encantos de mar Rojo, Estrellales, Penjamo and Rebusca. The approach of this research is not fulfilling for farms that are diverse in their soil properties. The banana leaves from those plantations are probably incorrectly predicted to belong to a certain area. To prevent this from happening a solution could be to stratify the area in more parts, by stratifying on a more detailed level. This more detailed stratification could be based on the original soil types for example, instead of on the generalised soil types as was done in this case. The used soil map contains 74 different soil units (SU). With such a large amount of units it is likely that more farms can be distinguished from each other.

To get completely rid of the downsides of the associated soil map it could be decided to not use the map at all. If a significant correlation could be established between the soil samples and the leaf samples directly the soil map would be of no need anymore. Soil sample data are available for almost all large commercial banana farms (at least in Costa Rica). Further investigation into the soil-plant relation is needed in that case, since comparing the soil data with the plant data does not always show a one-to-one relationship for the data used here. This is amongst others shown by the Fe concentrations for the soil and plant data of the rainfall strata. Another clear example of this is the Mg levels in the andic strata. All three strata have very different soil concentrations but the plant concentrations in the partly andic and non-andic stratum are practically similar. It is assumable that plants will only take up the amount of a nutrient that it needs. Most nutrients have an optimum concentration, too high levels can become toxic. This can explain that higher soil nutrient concentrations not necessarily lead to higher concentrations within the plant. This also shows from the data for Fe levels in the soil and plant data. For the soil data there are large differences for Fe between the different strata while all strata have approximately similar Fe levels in the plants.

On the other hand there are also interactions between the nutrients in the soil. A higher concentration of Mn has a negative influence on the uptake of Mg and Ca for example (Turner & Barkus, 1983). The effects of an element can also have positive effects. Zn for example enhances uptake of other nutrients and higher Zn levels significantly increase growth of the banana plants and fruits (Moreira & Fageria, 2009).

Since there are relationships between different soil elements and their uptake rate by plants, unique combinations of elements might occur in one of the strata. In this research only one nutrient at the time was examined for the prediction analysis. It is very plausible that left over unexplained variation could be explained by (a combination of) other plant elements.

Just like combining different elements for analysis it would also be possible to combine strata of the different stratifications, for example soils that are andic and receive the higher amount of rainfall. If it can be assumed that the combined properties where stratification is based on are not correlated it is possible that they will explain an extra portion of the variation. Therefore this could make it easier to say with more certainty where the banana plant material originated.

As already explained in the first section of the Methodology (2.1) 4 steps can be identified that influence the composition of a banana. The first step is the composition of the soil, the second step comprises the nutrient transport from the soil to the banana plant. Those are the two steps as investigated in this research. The following steps are the production of fruits on the plant (3) and the fourth and last step is the transport and ripening of the bananas. Results after step 2 look promising, but what is the importance of step 3 and 4 still?

The transition from step 2 to step 3 involves the distribution of nutrients from the banana plant into the actual fruits. Logically there is a relation between those two, after all if a nutrient is not at all present in the plant it will also not be found in the fruits. The relation between plant nutrient concentrations and fruit concentrations is not necessarily a one-to-one relationship, since different nutrients are translocated differently within the plant (Figure 3-7). Nutrients like Fe and Mn are mostly retained in the leaves, while Mg has a high translocation rate towards the fruits (Moreira & Fageria, 2009). Other sources say that nutrient concentrations in the fruits reach the same levels always, irrespective of soil or plant concentrations (assuming soil and plant concentrations are sufficiently high) ((Milthorpe & Moorby, 1979) via (Turner & Barkus, 1983)).

Table 3-12. Average N, P and K percentages in banana leaves and fruits.

	N leaves	N fruit	P leaves	P fruit	K leaves	K fruit
Bananatica	2.58	0.67	0.18	0.09	3.55	2.11
San Pablo	2.71	0.92	0.18	0.10	3.58	2.04

From two of the farms used in this research data was available about the N, P and K percentages in leaves as well as in the fruits (Table 3-12). It is by far not enough data to conclude on, but it seems there is a relation between the N content in the leaves and the fruits. The N percentage in the leaves of banana plants from Bananatica is lower than for San Pablo, and the same goes for N concentrations in the banana fruits.

Foliar concentration, increment and fraction of retranslocation of nutrients in banana plants cultivated in Central Amazon, Amazonas State, Brazil[1]

	NF_1	NF_2		
	g kg^{-1}		Δ %	FNR
N	26.1	21.7	−16.9	−0.73
P	1.7	1.4	−17.6	−0.72
K	27.5	22.5	−18.2	−0.74
Ca	4.7	6.6	40.4	0.00
Mg	2.9	2.2	−24.1	−0.83
S	1.7	1.9	11.8	−0.24
	NF_3	NF_4		
	mg kg^{-1}			
B	19.3	17.5	−9.3	0.00
Cu	5.5	3.4	−38.2	−0.56
Fe	52.4	92.6	76.7	0.45
Mn	280.0	540.1	92.9	0.50
Zn	13.0	14.0	7.7	0.05

[1]NF_1 and NF_3—foliar nutrients concentration at the beginning of inflorescence; NF_2 and NF_4—foliar nutrients concentration at the bunches yield; Δ—increment of foliar level-macronutrients = $(NF_2\text{-}NF_1)/100$ or micronutrients = $(NF_4\text{-}NF_3)/100$; FNR—fraction of nutrient retranslocated.

Figure 3-7. Re-translocation rates of nutrients in banana plants. Adapted from (Moreira & Fageria, 2009).

After bananas are full grown and harvested they are transported and artificially ripened. During this stage the chemical content of the banana changes. The pH of the fruits might decrease, while the sugar content largely increases (Kulkarni et al., 2011). No more changes occur to the elements investigated in this research (Al, Mg, Fe, Zn, Mn and Cu).

To come from the steps made in this research to a complete method that covers all the steps it is most important to investigate the exact relation between step 2 and 3. So far progression is already made by having more insight in relationships between the soil map and banana plant data.

Overall, large steps are made to come up with a decent retracing system for bananas. The prediction analysis results are promising. The percentage of samples that could be correctly predicted to its origin does not need further screening in case of a quality or safety issue. Only the leftover percentage, the false negative samples, stills needs a control check in this case.

4. Conclusion

The research questions posed at the start of this research were the following:

- Is the available soil map of Costa Rica representative for the local soil conditions?
- Is banana plant material grown on different soils significantly different from each other?
- Can banana plant characteristics reveal information on their origin?

All the questions posed can be positively answered. The soil map of the Atlantic Zone is representative for the local soil conditions. This was shown by comparing soil sample data over different soil map stratifications. The sampled data were significantly different for the strata. Largest variation in the soil sample data was found when the area was stratified on soil type and on andic properties. From all the soil elements sampled, Mg and Fe proved to be most useful to distinguish the defined strata since the largest variation amongst strata was found for these elements.

Banana leaves grown on different soils do also differ significantly from each other. Soil differences do lead to differences in leaf composition. Similar as for the soil data, stratification based on soil type and on andic properties showed the best results for the banana leaf data as well. The leaf Mg concentration most clearly differed over the different strata. Unlike the soil sample results, the leaf Fe concentration did not show much variation over the strata.

Banana plant characteristics most certainly do reveal information on their origin. With the stratification used it turned out it is possible to correctly predict the origin of banana leaves. The Mg and Cu concentrations in the leaves are most effective for this analysis. When considering stratification based on andic properties and using Mg as predictor variable, the origin of up to 86% of the leaf samples is correctly predicted.

Results of this study are very promising, however there are always some challenges to overcome. A downside of the soil map used is the way it is built-up, with associations. Not knowing the exact soil properties for 100% at a certain location is a limitation factor in this research. Ideally this is known, so all sample data could be used in the analyses. Now farms that fall in various strata are not analysed and nothing can be concluded from those leaf samples.

This study shows the potential of retracing bananas to their original growing location. Stratifying the area on very general properties like the soil type or the andic properties already makes it possible to distinguish between banana leaves.

5. References

Arias, P., Dankers, C., Liu, P., & Pilkauskas, P. (2004). La Economía Mundial del Banano 1985–2002. Estudios Food and Agricultural Organization (FAO) *Productos Básicos 1*.

Bellamy, A. S. (2013). Banana production systems: Identification of alternative systems for more sustainable production. *Ambio, 42*(3), 334-343. doi: 10.1007/s13280-012-0341-y

Bugaud, C., Daribo, M. O., Beauté, M. P., Telle, N., & Dubois, C. (2009). Relative importance of location and period of banana bunch growth in carbohydrate content and mineral composition of fruit. *Fruits, 64*(2), 63-74. doi: 10.1051/fruits/2009001

FAOSTAT. (2010). *FAO Database*.

Freitas, A. S., Pozza, E. A., Alves, M. C., Coelho, G., Rocha, H. S., & Pozza, A. A. A. (2016). Spatial distribution of Yellow Sigatoka Leaf Spot correlated with soil fertility and plant nutrition. *Precision Agriculture, 17*(1), 93-107. doi: 10.1007/s11119-015-9409-6

IMN. (2014). Monthly weather averages for San José over the period 1996-2014. Retrieved 14-4-2016, from Instituto Meteorológico Nacional de Costa Rica https://www.imn.ac.cr/en/inicio

Johns, G. G., & Vimpany, I. A. (1999). Interaction of pH amendment and potassium fertiliser on soil chemistry and banana plant growth. *Australian Journal of Agricultural Research, 50*(2), 199-210.

Jonge, J. d., Frewer, L., Trijp, H. v., Renes, R. J., Wit, W. d., & Timmers, J. (2004). Monitoring consumer confidence in food safety: an exploratory study. *British Food Journal, 106*(10/11), 837-849. doi: 10.1108/00070700410561423

Kulkarni, S. G., Kudachikar, V. B., & Keshava Prakash, M. N. (2011). Studies on physico-chemical changes during artificial ripening of banana (Musa sp) variety 'Robusta'. *Journal of Food Science and Technology, 48*(6), 730-734. doi: 10.1007/s13197-010-0133-y

Linhai, W., Lingling, X., & Jian, G. (2011). The acceptability of certified traceable food among Chinese consumers. *British Food Journal, 113*(4), 519-534. doi: 10.1108/00070701111123998

Milthorpe, F. L., & Moorby, J. (1979). *An Introduction to Crop Physiology*: Cambridge University Press.

Moreira, A., Castro, C., & Fageria, N. K. (2011). Effects of Boron application on yield, foliar Boron concentration, and efficiency of soil Boron extracting solutions in a xanthic ferralsol cultivated with banana in central Amazon. *Communications in Soil Science and Plant Analysis, 42*(18), 2169-2178. doi: 10.1080/00103624.2011.602447

Moreira, A., & Fageria, N. K. (2009). Yield, Uptake, and Retranslocation of Nutrients in Banana Plants Cultivated in Upland Soil of Central Amazonian. *Journal of Plant Nutrition, 32*(3), 443-457. doi: 10.1080/01904160802660750

Nciizah, A. D., & Wakindiki, I. I. C. (2014). Rainfall intensity effects on crusting and mode of seedling emergence in some quartz-dominated South African soils. *Water SA, 40*(4), 587-594. doi: 10.4314/wsa.v40i4.2

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

Padilla, C., Villalobos, P., Spiller, A., & Henry, G. (2007). Consumer preference and willingness to pay for an officially certified quality label: Implications for traditional food producers. *Agricultura Tecnica, 67*(3), 300-308.

Petre, J. (2014, 18 October 2014). Family's terror as they find world's deadliest spider trapped in Waitrose home delivery. *Mail on Sunday*.

Robinson, J. C., & Galán Saúco, V. (2010). *Bananas and plantains. 2nd edition.* Wallingford: CABI.

Turner, D. W., & Barkus, B. (1982). Yield, chemical composition, growth and maturity of 'Williams' banana fruit in relation to supply of potassium, magnesium and manganese. *Scientia Horticulturae, 16*(3), 239-252. doi: 10.1016/0304-4238(82)90072-3

Turner, D. W., & Barkus, B. (1983). The uptake and distribution of mineral nutrients in the banana in response to supply of K, Mg and Mn. *Fertilizer research, 4*(1), 89-99. doi: 10.1007/bf01049669

van der Waal, J. W. H., & Moss, J. R. J. (2013). *JUST GREEN BANANAS: TOWARDS FULL SUSTAINABILITY OF THE EXPORT BANANA TRADE.* Paper presented at the Acta Horticulturae.

van Rijswijk, W., Frewer, L. J., Menozzi, D., & Faioli, G. (2008). Consumer perceptions of traceability: A cross-national comparison of the associated benefits. *Food Quality and Preference, 19*(5), 452-464. doi: 10.1016/j.foodqual.2008.02.001

Velcovska, S. (2012). Food quality labels and their perception by consumers in the Czech Republic. *World Academy of Science, Engineering and Technology, 66*, 154-160.

Versari, A., Laurie, V. F., Ricci, A., Laghi, L., & Parpinello, G. P. (2014). Progress in authentication, typification and traceability of grapes and wines by chemometric approaches. *Food Research International, 60*, 2-18.

6. Appendices

I. Overview of banana farm samples used for analysis.

Farm	# of soil analyses	Total amount of soil records	# of foliar analyses	Total amount of foliar records
Balatana	10	96	1	11
Bananatica	10	158	5	142
Cariari	10	190	7	171
Cartagena	11	342	13	406
El Esfuerzo	14	203	14	182
Encantos de Mar Rojo	12	250	9	228
Estrellales	6	79	4	89
Penjamo	6	82	5	65
Rebusca	6	103	5	89
San Pablo	11	218	7	129
Siquirres	11	368	9	340
Venecia	8	252	5	193
Verde Azul	9	173	11	233
Zorzales	12	229	29	519

The second and fourth column show the amount of sample moments over the past 5 years for the soil and plant respectively. The third and fifth column show the total amount of individual observations. An example: at the farm Balatana the soil has been sampled 10 times over the past five years, and this resulted in a total amount of 96 records, which means that on average almost 10 (96/10) different samples were taken every time the farm was visited for soil samples.

II. Soil samples

Every soil sample was analysed on the following characteristics:

- **Percentage of organic matter (OM)**
- pH
- **Aluminium in centimol plus per litre (Al)**
- Acidity in centimol plus per litre (Acid)
- Calcium in centimol plus per litre (Ca)
- **Magnesium in centimol plus per litre (Mg)**
- Potassium in centimol plus per litre (K)
- Phosphorus in centimol plus per litre (P)
- **Iron in centimol plus per litre (Fe)**
- **Copper in centimol plus per litre (Cu)**
- **Zinc in centimol plus per litre (Zn)**
- **Manganese in centimol plus per litre (Mn)**
- Boron in milligram per litre (B)
- Sulphur in milligram per litre (S)

Elements that are printed in bold are the ones used in the analyses.

III. Foliar samples

The banana leaves were analysed on the following components:

- Nitrogen as percentage of the total dry weight (N)
- Phosphorus as percentage of the total dry weight (P)
- Potassium as percentage of the total dry weight (K)
- Calcium as percentage of the total dry weight (Ca)

- **Magnesium as percentage of the total dry weight (Mg)**
- Sulphur as percentage of the total dry weight (S)
- **Iron in milligram per kilogram (Fe)**
- **Copper in milligram per kilogram (Cu)**
- **Zinc in milligram per kilogram (Zn)**
- **Manganese in milligram per kilogram (Mn)**
- Boron in milligram per kilogram (B)
- **Aluminium in e-milligram per kilogram (Al)**

Elements that are printed in bold are the ones used in the analyses.

IV. Strata

The chosen soil properties all had some initial classification, since most properties are nominal data. Below an overview of the present classes per property is given. The number between brackets behind each represents the reclassified stratum.

General soil type:

1. Fertile well drained soils	(1)
2. Fertile poorly drained soils	(2)
3. Infertile well drained soils	(3)
4. Peaty soils	(4)

Texture:

1. Sand	(1)
2. Loamy sand	(1)
3. Sandy loam	(2)
4. Fine sandy loam	(2)
5. Very fine sandy loam	(2)
6. Loam	(2)
7. Silty loam	(3)
8. Silt	(3)
9. Clay loam	(4)
10. Sandy clay loam	(4)
11. Silty clay loam	(4)
12. Sandy clay	(5)
13. Silty clay	(5)
14. Clay	(5)

Andic properties:

1. Andic	(1)
2. Containing (pseudo) vitric properties	(2)
3. Andic subgroup	(2)
4. Andic subgroup	(2)
5. Non-andic	(3)

Rainfall:

1. 3000-3500 mm	(1)
2. 3500-4000 mm	(2)
3. 4000-5000 mm	(3)
4. 5000-6000 mm	(4)
5. 6000-8000 mm	(5)