


Role of the environment on the incidence of Panama disease in bananas

Marianne Bosman

MSc thesis – Soil Geography and Landscape
Wageningen, August 2016

Role of the environment on the incidence of Panama disease in bananas

Marianne Bosman
930509109010
Marianne.bosman@wur.nl

MSc thesis - Soil Geography and Landscape
Master Earth and Environment (MEE)

Supervisors
Jetse Stoorvogel
Soil Geography and Landscape group, Wageningen University, Wageningen, The Netherlands

Rafael Segura Mena
Research division, CORBANA SA., Guápiles, Costa Rica

Wageningen University, August 2016

Abstract

Panama disease, or also known as Fusarium wilt of banana, is a major thread to global banana production, putting economic development and food security in danger for many countries worldwide, while no solutions are present. Soil management provides a promising alternative to a solution, but it is not clear which environmental variables are important for decreased disease incidence. Previous work addressing this issue lacked confirmation of the results in real operative farms and the knowledge on the influence of landscape patterns is lacking. A field experiment is executed where landscape, climate, land management, soil chemistry, microbial populations and disease incidence are assessed in order to find relations between environmental variables and disease incidence of Panama disease. With these variables the variance in the study area is illustrated via a GIS analysis, the complexity of the problem is shown and t-tests and regression analysis are performed to find those variables that influence Panama disease. The results show that factors such as curvature, temperature, plant density, pH, acidity and Magnesium all have a relation with disease incidence. With the results, soil management practices can be created and adjusted accordingly, to minimize the harm Panama disease causes ta present and can cause in the future.

Table of Contents

Introduction

General introduction

Banana production is in large parts of the tropics and sub-tropics important for both food security and economic security. After rice, wheat and maize banana has the highest global gross value production and is thus a large food commodity worldwide (Zhang et al., 2013). Still, only 15% of all produced bananas are exported on the international market, while the other 85% is for local and national consumption (Ploetz et al., 2015). The total amount of bananas produced is yearly increasing with an average rate of 4.25% (FAOSTAT, 2013).

Costa Rica is after Ecuador and The Philippines, the most exporting country in the world, and Costa Rica is the top banana producing country in Central America and the Caribbean (Robinson & Saúco, 2010). The most large scale banana plantations in Costa Rica can be found in the lowlands, but throughout the whole country farms are located that produce bananas for local and regional consumption. These farms are smaller than the commercial plantations and different types of bananas in all sorts and sizes are produced and sold within Costa Rica.

Unfortunately both the small and large scale banana production is being plagued and threatened by multiple diseases, such as black and yellow sigatoka, bacterial wilt, root rot and Fusarium wilt. There are hundreds of different banana cultivars produced on a small scale, but for large scale cultivation only a few cultivars are economically beneficial. Therefore, between large plantations genetic variation is small, inducing a high vulnerability of pests and diseases (Ghag et al., 2015; Ordonez et al., 2015; Ploetz, 2015b). The economic importance of banana production for Costa Rica and the threat of different banana diseases causing large yield gaps is an important reason for Costa Rica to minimize the damage by finding effective measures to deal with these diseases.

One of the most destructive plant disease for banana is called Fusarium Oxysporum f.sp. cubens (Foc), or better known as Panama Disease or Fusarium wilt of banana. Fusarium wilt is a soil-borne fungus causing problems in the vascular units of a plant (Beckman & Ploetz, 1990). Once Fusarium wilt infiltrates the banana plant it restricts water going up in the vascular unit, causing water shortage and eventually plant wilting (Cook & Papendick, 1972). Fusarium wilt can be recognized by the yellowing of the leaves and/or the collapse of the petiole and the subsequent falling of the leaves. In later phases of the disease also the pseudo stem of the plant can split and eventually the plant will die (Figure 1) (Pérez-Vicente, 2004).

Figure 1
The visual effects of Fusarium wilt of bananas on the plant.
A: yellowing and dying of the leaves. B: collapse of the petiole.
C: stem splitting.

Fusarium wilt caused great destruction in the large scale banana industry in Latin America in the first half of the 20th century. Back then, Gros Michel was the main cultivar, however susceptible to Race 1 of Fusarium wilt. After 50 years of research in pesticide usage, relocation and breeding, Gros Michel was being replaced by a disease resistant cultivar called Cavendish (Marquardt, 2001). By now, Race 1 has spread throughout the world and Cavendish has become the main cultivar for export bananas. Gros Michel is still being produced, but rather on a small scale for local consumption than for the export.

In the 80's a new strain of Fusarium wilt, Race 4, was discover which affects the Cavendish cultivar and many other banana cultivars (Pegg & Langdon, 1986). By now Race 4 is spreading in the Asian tropics, causing important losses in Malaysia and Indonesia among others (Pérez-Vicente, 2004). Because Race 4 effects Cavendish and many other local cultivars (Ploetz, 2006), not only the export but also local and regional food provision is threatened (Ordonez et al., 2015). Race 4 is already disseminated through large parts of Asia. In Africa and the Middle East one and two locations respectively are infested with Fusarium wilt (Figure 2). It is only a matter of time before Race 4 will be introduced in Latin America and the current monoculture production induces a rapid

Figure 2
Map of the world with all areas currently infected by Panama disease (Race 4) (red), areas at risk (green) and areas affected during the previous (Race 1) outbreak of Panama disease (blue)(source: panamadisease.org/en/map).

dissemination once Race 4 has arrived. Even though Race 4 is not yet present in Latin America, many countries, including Costa Rica, are preparing themselves for its arrival and the impact Race 4 will have (Ploetz, 2015b).

A significant part of this preparation consists of finding ways to minimize the damage Fusarium wilt can cause. Local Gros Michel production shows that there are ways to keep producing in infested areas (Race 1). Research on these local farms can provide answers for both local and large scale banana production. However, the incidence of the disease is very variable and it is not well known which factors are important for a low disease incidence.

The use of resistant cultivars was the only known effective method to overcome Panama disease (Ghag et al., 2015). However for Race 4 no suitable cultivars for large export production are known yet. Breeding and GMO can result in a successful resistant cultivar, but has many complications and restrictions due to the perennial nature of the banana plant. Also measures that are useful for other crops with an annual cycle and a short lived host are unsuccessful against Fusarium wilt of banana.

There are various management methods that might bring a solution. Soil management is a sustainable method for disease control and shows prospects for reducing the effect of Fusarium wilt. There are a substantial number of papers showing positive results between disease incidence and biological, chemical, environmental or cultural variables, however there is a lack of practical field results (Pérez-Vicente, 2004; Ploetz, 2015a). Overall, feedbacks, interactions and relations between variables and disease incidence are poorly understood, while spatial variability in both disease incidence and explanatory variables is high. Additional research on soil management and abiotic variables is needed to understand which variables are important and how it is possible to manipulate variables for disease control (Pattison et al., 2014; Ghag et al., 2015; Ordonez et al., 2015).

Problem analysis

Many environmental factors play a role in the interactions that happen in and between the plant and soil. These factors can be of biological, chemical, physical, cultural, or climate nature. All factors have their own contribution to soil quality and plant health and many interaction between factors are present. Also spatial variability can cause that one factor may be helpful for disease control in one situation, but might have a totally different effect in another location with other landscape characteristics and plant surroundings.

There are two main routes to reduce the effects of Fusarium wilt. One way is by minimizing the damage Fusarium wilt can bring to a plant, by making sure the plant is healthy and has sufficient resources. The other way is by supressing Fusarium wilt in the soil, thus minimizing the fungi concentration in the soil. Consequently this minimizes the invasion of Fusarium wilt in the plant. Soils that have the ability to supress Fusarium wilt are called supressive soils. Soils that cannot supress the disease are also known as conducive soils. It has been proven before (Shen et al., 2015b) that some soils are indeed suppressive to Fusarium wilt, however a full conversion form a conducive to a supressive soil has not been successful yet (Palti, 1981; Ploetz, 2015a).

To optimize both the plant health and supress the soil, soil management is introduced as a way to minimize the effect of Fusarium wilt of bananas (Ghorbani et al., 2008). Soil management is a wide concept that includes the application of fertilizers, tillage, managing nutrient efficiency, crop rotation and green cover. Soil management can apply to both soil suppression and plant health. It appears that factors such as pH, water content and temperature are of importance for reducing susceptibility of a banana plant to Fusarium wilt (Peng et al., 1999). Also Amir & Alabouvette (1993) showed that abiotic environmental factors can influence the reaction of Fusarium wilt in supressive soils. Pattison et al. (2014) showed the positive effect of green cover between banana plants on the disease incidence.

Many factors on different scales influence the incidence of Fusarium wilt in the plant and are thus important for proper soil management. On a soil scale, chemical soil properties such as pH, Acidity, Organic Matter, Nitrogen, Magnesium and Manganese are suspected to be of importance (Alabouvette, 1999; Navajothy et al., 2011; Segura et al., 2015a) and also the size of microbial populations seems to have a big influence on the disease incidence (peng et al., 1999; Shen et al., 2015b). Microbial populations are important for the suppression in the soil, while many chemical and physical environmental factors affect the microbial population in their turn.

On a farm scale, land management appears to have a dominant role on the disease incidence. Intercropping, crop rotation and green cover are three examples of the importance of land management (Stover, 1962; Zhang et al., 2013; Pattison et al., 2014).

Landscape and climate vary on an even larger scale than land management and is also expected to be related to disease incidence. Temperature has shown to be important (Pegg & Langdon, 1986), however limited research is done before on the relation between disease incidence and landscape factors. With new information about the role of the landscape and climate on disease incidence the planning of new locations for farms can be improved and additional information about soil suppression can be obtained.

Even though soil management looks like an attractive measure to control the disease, one important knowledge gap is present; field data on farm level proving the positive effects of soil management is lacking. Research has been conducted before on the influence of land management and soil conditions (soil chemistry and microbial populations) on disease incidence in banana (Pattison et al., 2014; Shen et al., 2015a; Segura et al., 2015a). However, this research has been conducted in pot- and field experiments. To confirm that land management and soil conditions are important for disease control, new data needs to be collected on operative farms where bananas are grown under pressure of Fusarium wilt.

This research will focus on confirmation of previously found results in a field survey. This will be done by first illustrating the large spatial variability in the study area and second the large variability in disease incidence. A large variability in disease incidence would mean that there are indeed factors that can explain this variability. The last part of this research will attempt to find those environmental variables that can explain part of the variability in disease incidence. Variables included are of landscape, climate, land management, soil chemical and microbial nature. This study will execute a field survey in the surroundings of Turrialba, Costa Rica.

In this area, small scale farmers are able to produce Gros Michel bananas for the local market, even though Fusarium wilt is present in all of these soils. This area has large spatial landscape variability and each farmer has its own way of managing his fields. This is characteristic specifically for this area, while all the large banana plantations have a much lower spatial variability when it comes to landscape and climate. By including landscape and climate variables, this research is able to put the importance of these variables relative to the importance of land management, soil chemistry and microbial variables. Therefore, the results of these latter variables are also applicable on infested farms and plantations outside of the study area.

The aim of this research is to prove the importance of the environment on the incidence of Panama disease. It provides an overview of the influence of environmental variables on the susceptibility of banana plants to Panama disease. With these results following research methods, planning locations of new farms and the practices of soil management can be improved, which will give rise to more opportunities to control the coexistence between Panama disease and banana production.

For this research three research questions are formulated as following;

- How variable is the landscape in the study area?

- How variable is disease incidence between Gros Michel farms within the study area?

- To what extent can environment explain the variability in disease incidence?

These questions will be assessed by executing a field survey in the surroundings of Turrialba, Costa Rica. Many variables will be taken into account. First this research will show the spatial variability in landscape and climate via a variability assessment done solely on sources available online. A quick scan in the field will be executed to collect data about the variability in disease incidence. It is expected that part of the variability in disease incidence can be explained by the variability in landscape and climate. During the quick scan also data for land management, soil chemistry and microbial populations will be collected. Finally a statistical analysis will be performed. The goal of the statistical analysis is to find out which variables are important for explaining disease incidence and which variables are not. These results are useful for further and more detailed research on using soil management as a method for disease control.

With this research it is possible to confirm previously found results in pot- and field experiments. During the field survey many small growing farms will be visited and depending on the current disease incidence in banana and size of the farm multiple observations will be made within one farm. The results provide important information on the relation between variables and the disease incidence. Relations with landscape and climate factors can be used for future location planning of new farms within the study area. Results on land management, soil chemistry and microbial populations can be used for improving soil management practices both within and outside the study area.

Materials & Methods

This research is executed in three main steps: (i) A variability assessment showing the landscape variability in the area; (ii) the collection of data via a quick scan; and (iii) the statistical analysis to determine the role of the environment.

For this research in total 40 variables are selected to answer the research questions. The selected variables can be divided in five different categories; landscape, climate, land management, soil chemistry and microbial populations. For the variability assessment the landscape and climate variables will be used. All landscape and climate variables are collected via online sources and analysed via ArcGIS. During the quick scan data about disease incidence, land management, soil chemistry and microbial populations is collected. During the statistical analysis all variables will be used.

This chapter will describe the study area and the methods and materials used for the variability assessment, quick scan and statistical analysis respectively.

Study area

The study area is located in the middle of Costa Rica surrounding the city of Turrialba, within the Cordillera Talamanca and where volcanic activity is high. Due to the active plate tectonics and recent volcanic activity, the area is mountainous with variations in altitude ranging from 157 to 2900 meters above sea level. Also the spatial variations in climatic variables such as temperature and precipitations are high, namely 10.8 - 25.5 °C (annual mean) and 2122 - 4586 mm/yr respectively. The climate in the study area is humid tropical, with minimal annual temperature variability and a rainy season from May to November.

The dominant soils in the area are Andosols and Nitisols (Dijkshoorn et al., 2005). Andosols are volcanic soils, mainly located surrounding the Turrialba and Irazú volcanos. Nitisols are deep, well drained tropical red soils with a high clay percentage. These soils are located more south-east of the Turrialba volcano, where volcanic influence on soil formation is lower.

Both soils are suitable for the cultivation of a large variety of crops. Therefore, apart from the natural reserves and extreme steep and rocky slopes, the area is mainly used for the cultivation of coffee and sugar cane, but also crops such as bananas, cacao and peach palm can still be found in the area (Kass et al., 1995). The bananas grown in this area and their accompanying farms are not to be compared with the large plantations that are located on the flat plains in the western part of Costa Rica. In the study area the banana yield is not meant for export, but is sold on the local and regional markets in Costa Rica, providing a wide variety of banana types, including the Gros Michel bananas that are researched in this report.

The farms visited for this research are all small farms, with no additional employees except for the farmer itself and the family. Many farms do not only cultivate Gros Michel, but often a mixture of different crops is present with coffee usually as the main or secondary crop (Figure 3). The size of the farms is very variable, but none are comparable to a normal size banana plantation in Costa Rica.

The study area shown in Figure 4 is delimited as following; North border=10°N, East border=83.4°W, South border = 9.8°N and the East border = 83.8°W.

The borders of the area are based on the locations of the farms and geological structures, so that multiple landforms are included in the stratification.

Figure 3
Banana plantation near Turrialba with coffee plants in between as an intercropping system.

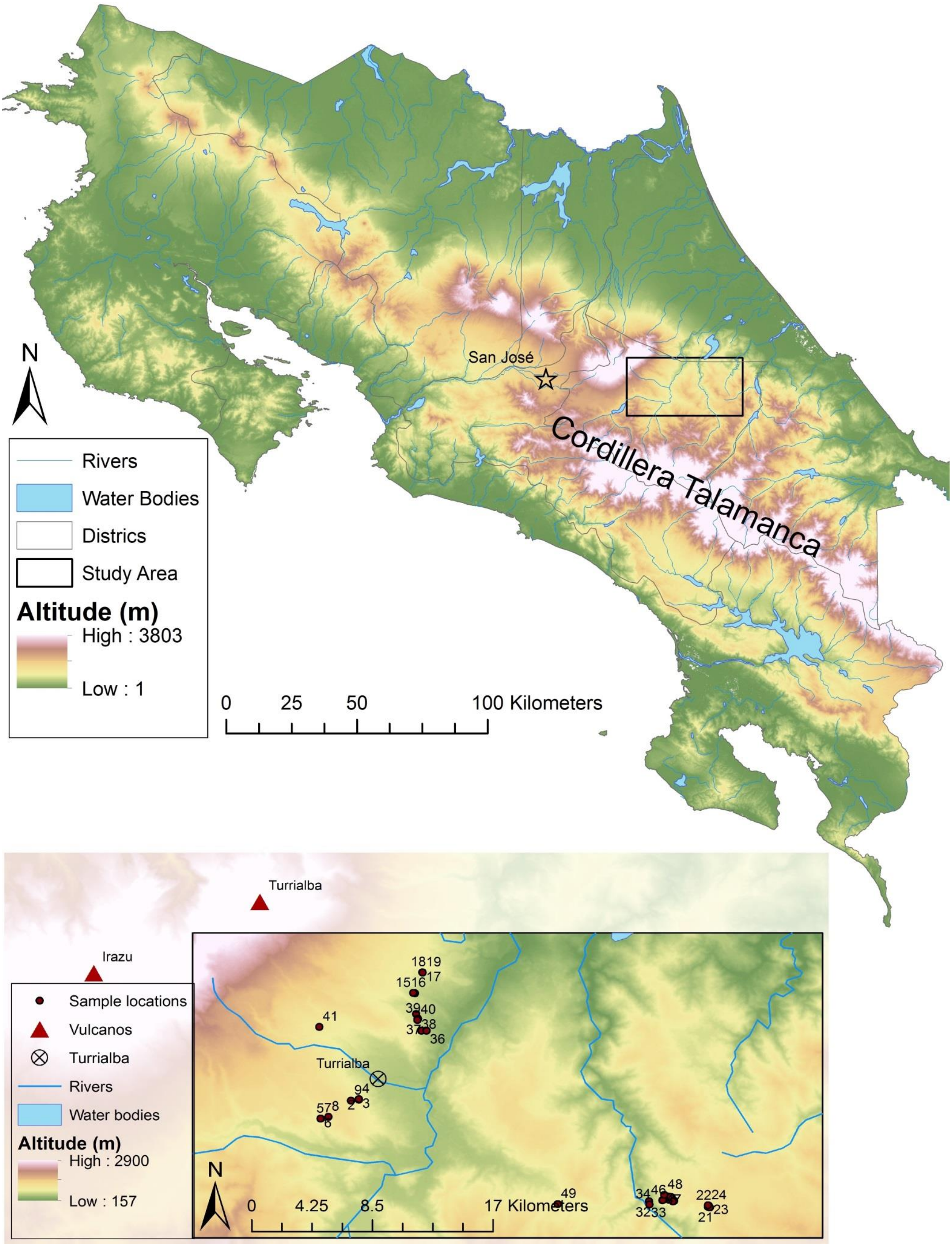


Figure 4
Top: Elevation map of Costa Rica with the study area in the black box. Bottom: Elevation map of the study area with all 50 sample locations and the two important volcanos in the area (Turrialba and Irazú).

Variability assessment

The first step in this research is to illustrate the high variability of landscape in the study area. This is done by creating a stratification map. The goal for creating this map is to illustrate the great variability in the study area and also show that this research deals with a complex problem that does not have a straight solution. The high spatial variability in landscape and climate in the study area is important to include. High spatial variability as in the study area is not present where the large scale banana plantations are located. To find results applicable for both the study area and outside of it, including large scale plantations, the high variability should be taken into account.

The method used to do the variability assessment is comparable to the climatic stratifications of Metzger et al. (2005) and Fairbanks (1999). First a Principal Component Analysis (PCA) is applied to determine the relative importance of all variables. Later a Cluster analysis is performed to create a final stratification map. Only landscape and climate variables are used for the variability assessment since these variables are available online, so that the variability assessment can be performed separate from the field work.

This section will start with a description of the selected variables and an argumentation on why these variables are included. Then the method used to find the relative importance of variables (PCA) and to make the stratification map (Cluster analysis) will be described. Both methods are executed in ArcGIS.

Selection of variables - Landscape

Since limited research before has focused on disease incidence in combination with landscape elements only basic and well known landscape variables are selected. All are expected to be important for disease incidence, however the rate of importance is unknown. All variables are listed in Table 1.

The altitude can influence disease incidence via a combination of different variables. In higher elevated areas, temperature is lower, precipitation patterns are different and slopes are often steeper than in lower elevated areas. These three aspects can make for either positive or negative influence on the disease. The slope is also important. First, because steep slopes have more erosion, thus more nutrient losses, than gentle slopes, where sediments often accumulate. Erosion also contributes to the distribution of Fusarium wilt that is transported with the sediments. Secondly, steepness influences the water availability in the soil and the amount of water going down, washing Fusarium wilt to the root system (Rishbeth, 1957). Sunshine is very important for banana plants and aspect is an important variable influencing the amount of sun received by the plants. Plan and profile curvature are also included. Curvature gives information about the shape of the slope. The shape of the slope, together with the steepness, is determinative for the locations where erosion exceeds sedimentation and the reverse. In concave (hollow) slopes sediment movement is the highest due to the accumulation of sediments during dry periods and the removal of these sediments when surface flow occurs. Convex (bulging) slopes are overall more prone to soil erosion and are dryer than convex slopes. Due to the high erosion rate on convex slopes, important nutrients and organic material erode with it, resulting in lower soil fertility. The erosion patterns of convex and concave slopes are especially important for agricultural lands, including banana cultivation, which are already more prone to soil erosion due to lower vegetation cover (Daniels & Hammer, 1992). Hillshade shows which areas are overshadowed by neighbouring mountains under a certain azimuth and altitude of the sun. It indicates which areas receive less or no sunlight at all due to shadows of landscape structures. Even though an area might be gentle sloping with a south aspect, if it is surrounded by high steep mountains it will be overshadowed and thus have a disadvantage for cultivation and plant health. Consequently it can also be more vulnerable to disease incidence. Distance to river is also selected as a variable. First of all, because its role in the dissemination of disease which can result in higher concentrations of the fungus in the soils closer to the river than farther away (Stover, 1962). Secondly, regular floods of the rivers and streams can influence the biotic and abiotic components of the soil. Thirdly, floods can cause great damage to banana plants and decrease plant health (Rishbeth, 1957). At last the Normalized Difference Vegetation Index (NDVI) is selected, because it is an indicator for the soil conditions for both the banana plant and Fusarium wilt (Bouwmeester et al., 2016).

The DEM used is a 30x30 m resolution tile from the ASTER Global Digital Elevation Model (ASTER GDEM). ASTER GDEM is a product of NASA and METI (NASA JPL, 2009). The altitude, slope, profile curvature, plan curvature and hillshade are calculated with the default settings in ArcGIS. Aspect is converted into radian units, so that north and south slopes are more distinctive compared to east and west

Table 1
All variables used for this research divided in five classes; Landscape, Climate, Land management, Soil chemistry and Microbial populations. For each variable a description, the unit and the source of information is also given

Variable	Description	Unit	Source
Landscape			
Longitude	*Longitude position in CRTM05 projection*	*Meters*	*GPS*
latitude	*Latitude position in CRTM05 projection*	*Meters*	*GPS*
Altitude (GPS)	*Elevation above AMSL according to the used GPS device with CRTM05 Projection*	*Meters*	*GPS*
Altitude (DEM)	Elevation above AMSL according to the 30X30 meter DEM	Meters	ASTERGDEM
Slope	The slope according to the 30x30 meter DEM	Degrees	ASTERGDEM
Aspect (Cosines)	Aspect, converted into radians with Cosine	Radians	ASTERGDEM
Curvature	Measure of the convexity of the slope		
Profile curvature	Convexity of the slope in the vertical direction	1/100 Meters	ASTERGDEM
Plan curvature	Convexity of the slope in the horizontal direction	1/100 Meters	ASTERGDEM
Hillshade	shaded relief with an azimuth of 180 and altitude of 75	-	ASTERGDEM
Distance to River	Distance to river, calculated with an existing river map	Meters	ASTERGDEM
NDVI	Normalized Difference Vegetation index derived from satellite imagery (13 April 2016)	-	LANDSAT 8
Climate			
Precipitation	Total annual precipitation	(mm)	Worldclim.org
Minimum precipitation	Total precipitation in the driest month (March)	(mm)	Worldclim.org
Temperature	Mean annual temperature	°C x 10	Worldclim.org
Minimum temperature	Mean temperature in the coldest month (December)	°C x 10	Worldclim.org
Land Management			
Plant Density	Number of mats divided by the area	Mats/m2	Field observations
Distance between mats	The average distance between two mats in the area	Meters	Field observations
Number of leaves	Average number of leaves per plant	-	Field observations
Number of shoots	Average number of shoots per mat	-	Field observations
Number of other crops	Total number of other crops, beside bananas, present in the area	-	Field observations
Proportion of other crops	The proportion of other crops, beside bananas, present in the area (crop cover/intercropping)	%	Field observations
Tree cover	Total proportion of trees covering the area	%	Field observations
Green cover	Total proportion of vegetation covering the soil in the area	%	Field observations
Brown cover	Total proportion of brown organic material on top of the soil in the area	%	Field observations
Bare soil	Total proportion of non-covered soil in the area	%	Field observations
soil chemistry			
Organic Material	Proportion of Organic Material in the top 30 cm of the soil	%	CORBANA
PH	Soil pH in the top 30 cm of the soil	-	CORBANA
Aluminium	Aluminium concentration in the top 30 cm of the soil	cmol(+)/L	CORBANA
Acidity	Concentration of acid molecules in the top 30 cm of the soil	cmol(+)/L	CORBANA
Calcium	Calcium concentration in the top 30 cm of the soil	cmol(+)/L	CORBANA
Magnesium	Magnesium concentration in the top 30 cm of the soil	cmol(+)/L	CORBANA
Potassium	Potassium concentration in the top 30 cm of the soil	cmol(+)/L	CORBANA
Phosphorus	Phosphorous concentration in the top 30 cm of the soil	mg/L	CORBANA
Iron	Iron concentration in the top 30 cm of the soil	mg/L	CORBANA
Copper	Copper concentration in the top 30 cm of the soil	mg/L	CORBANA
Zinc	Zinc concentration in the top 30 cm of the soil	mg/L	CORBANA
Manganese	Manganese concentration in the top 30 cm of the soil	mg/L	CORBANA
Boron	Boron concentration in the top 30 cm of the soil	mg/L	CORBANA
Carbon	Proportion of Carbon in the top 30 cm of the soil	%	CORBANA
Nitrogen	Proportion of Nitrogen in the top 30 cm of the soil	%	CORBANA
Microbial populations			
Fungi	All threadlike fungi (filiformes) in the top 30 cm of the soil	cfu/g	CORBANA
Total Bacteria	Total number of bacteria in the top 30 cm of the soil	cfu /g	CORBANA
Total Actinomycetes	Total number of Actinomycetes in the top 30 cm of the soil	cfu /g	CORBANA

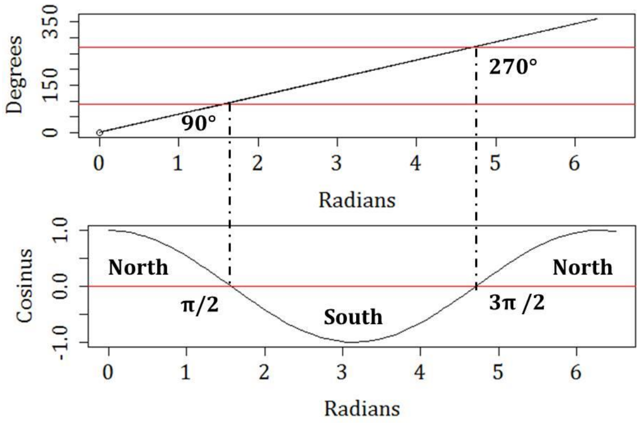


Figure 5
Illustration of the conversion from aspect in degrees to aspect in radians.

slopes (Figure 5). This choice is made, because for plant health south slopes are expected to be in favour against north slopes, while east and west slopes both have the same amount of daily sun and are thus less important. Distance to river is determined by calculating the Euclidean distance from a river network of the area provided by CORBANA. The NDVI is calculated in ArcGIS from a LANDSAT 8 satellite image taken on 13 April 2016, where the whole study area was clear of clouds. The imagery is obtained from http://earthexplorer.usgs.gov/ and info about LANDSAT 8 can be found in the paper of knight & Kvaran (2014). At last in the field also the altitude is measured with a Garmin GPS-device. This is done as a way to validate the position of the longitude and latitude with the DEM and its derivatives.

Selection of variables - Climate

Climate is known to be of importance for the incidence of Fusarium wilt of banana, however this knowledge is mainly based on differences in climate between different climatic areas (Stover, 1962). In this research, variability in precipitation and temperature is assessed on a large scale within the study area, but on a small scale compared to global climatic variability. Total annual precipitation and annual mean temperature are selected to cover the average climate conditions, and total precipitation in the driest month (March) and mean temperature in the coldest month (December) are selected to cover the yearly climatic extremes in the area. Since banana plants prefer lots of water and high temperature the extremes are thus the opposite; low precipitation and cold temperatures. It is decided to include temperature as a separate variable, regardless of its correlation to the altitude. This correlation is expected to be high, but local differences can be present. Since the study area is relatively small for climatic conditions, these local differences are important relative to a large scale.

The data is obtained from www.worldclim.org/current and info about the data can be found in Hijmans et al. (2005). The original resolution is 930x930m, they are resamples in ArcGIS to the same size and resolution as the DEM.

Relative importance of variables

On the selected landscape and climate variables a Principal Component Analysis (PCA) is performed (Wold et al., 1987). A Principal Component Analysis is a statistical method to reduce the amount of data, while still keeping the variability between different variables. PCA is a statistical method that is able to identify patterns and show the similarities and differences between variables.

In this research many variables are considered, however the importance of the variables and the interactions between variables are mainly unknown. The choice is made to use a PCA, because a PCA finds those variables that are dominant in variability by searching the correlated variables and converting them into simplified layers that are uncorrelated from each other. Thus, capturing the variance explained only by uncorrelated variables. This way, the eventual stratification will be composited of variables that are uncorrelated and explanatory for the variability and consequently important for the division between strata (Smith, 2002; Li et al., 2006). The PCA is executed on continuous raster's in ArcGIS.

Before executing the PCA, the used variables needed to be identical to each other in mean, standard deviation, extend and resolution. Therefore all raster's are re-sampled to the same extend and same cell size as the DEM. The second step is to standardize all data. During the PCA, the size of the variance is linear to the size of the range (Wold et al., 1987).Therefore all maps are standardized to a mean of zero and a standard deviation of one. It is important in this case to keep in mind that, when standardizing, a variable with relatively low variance gets the same range as a variable with high variability. The assumption is made that for all variables used the spatial variance is high and thus standardization does not bias the results.

Stratification

With the results of the PCA, the Iterative Self-Organizing Data Analysis Technique (ISO-DATA) (Tou et al., 1974) is implemented to divide the area in 25 different clusters. These 25 clusters are converted into a map with fourteen strata. In ArcGIS an 'ISO cluster unsupervised classification' is performed to cluster the PCA variance. Unsupervised means that no pre-defined data is used and only the variance of the input is used for creating clusters (Mansfield et al., 1999). Since no pre-defined data is present, this is the right approach for the cluster analysis. For the analysis 25 cluster are created based on the variance for each cell calculated with the PCA. The number of clusters, 25, is chosen arbitrary. Each cluster represents the most important variable(s) for the variance in that area. These twenty five clusters are however very divers and spread throughout the area. To decrease this spread and optimize the interpretation of the stratification map, the 25 clusters are manually converted into fourteen strata. The number of strata is also arbitrary, depending on the variability and sizes of the before created clusters. The fourteen strata illustrate the high spatial landscape variability in the study area, based on the importance and correlations between variables.

Quick scan

The idea of a quick scan in this research is to collect as much information as possible within a limited number of days. During this quick scan 21 farms are visited, 49 observations are done and 50 soil samples are collected. The quick scan is performed in two different weeks. Observation numbers 1-34 are collected from 16 to 18 May 2016. The second fieldwork trip is executed on Wednesday 6 and Thursday 7 of July, 2016. During the quick scan both field observations and soil samples are collected. The field observations provide information about land management and disease control. The soil samples were analysed in the lab for soil chemical variables and microbial populations. The quick scan is performed in corporation with both CORBANA and Bioversity International.

This section will start with a description of the collected variables during the quick scan and an argumentation on why these variables are included. This will be followed by a brief description on how the data is collected.

Selection of variables – Land management

For land management the variables observed are selected based on expert knowledge and resource availability. Since the observations in the field should not take longer than 15 minutes ('quick' scan), all variables are immediately observable. Therefore variables such as different soil horizons, soil type and soil texture are not taken into account, even though large spatial differences in soil types and characteristics can be expected in this area.

Plant density is selected as a variable because it is expected to be of importance for disease incidence. A high plant density can cause competition between the plants and consequently result in lower plant health and higher incidence. Plant density (Mat density) is calculated by estimating the number of mats and the area (m^2) of the field. A mat is the complete banana plant including multiple pseudostems and shoots (Figure 6). The distance between mats is the estimated average distance between two mats in the area. The distance between mats is related to plant density, however within an area, distance between mats can still vary. Number of shoots and leaves are an important proxy for the health of a plant and are estimated by taking the average from a random selection of plants.

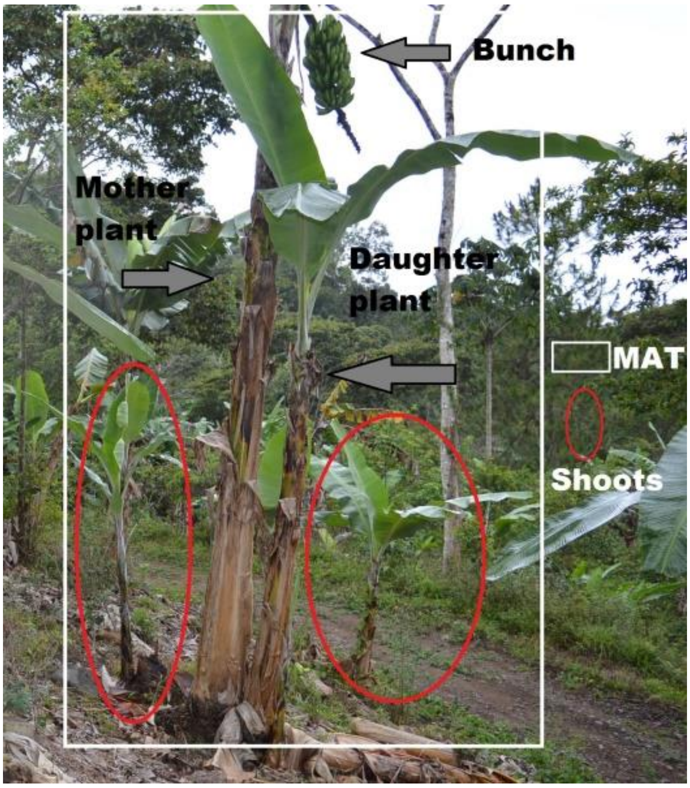


Figure 6
Simple anatomy of a banana mat.

Based on previous research in field experiments (Zhang et al., 2013; Pattison et al., 2014), things such as intercropping and green cover also appear to influence disease incidence. Therefore the plant surroundings are observed and crop cover, tree cover, green cover, brown cover and bare soil on each location are estimated. A description of these variables can be found in Table 1. All these 5 variables influence the soil conditions by increasing soil aeration, changing the water holding capacity, influencing soil temperature and decreasing soil erosion. They are important for both soil suppression and plant health (Sudarma & Suprapta, 2011).

Selection of variables – soil chemistry

The soil chemical variables that are measured are all variables that can be controlled in the soil via fertilization. Aluminium (Al), Acidity (Acid), Calcium (Ca), Phosphorous (P), Magnesium (Mg) and Potassium (K) are macro nutrients and Iron (Fe), Copper (Cu), Zinc (Zn), Manganese (Mn) and Boron (B) are all micro nutrients. Also pH, Organic Matter (OM), Carbon (C) and Nitrogen (N) concentrations are measured, simply because they are very important soil components, easy to measure and part of the standard procedure.

It is known that chemical composition in the soil is important for disease suppression (Segura et al., 2015a; Shen et al., 2015a), however the interactions between nutrients and other chemical components are very complex. Certain nutrients have shown to be significant for explaining disease incidence in pot and field experiments (pH, Acidity, Ca, Mg, Mn, N). However, the chemical processes within a soil can be different for each soil type and each location. Therefore all nutrients are accounted for to make sure other or new processes and relations are not overlooked.

Selection of variables – Microbial populations

The three variables that are used for microbial populations are all three indicators of the microbial activity in the soil. It is said that microbial activity is very important for soil suppression of Fusarium wilt (Scher & Baker, 1980; Shen et al., 2015b). Besides this reason, microbial populations are also measured to see which variables might be enhancing or reducing microbial activity in the soil. This can provide information about the complex processes and interactions present in the soil.

Collection of the data

The data are mainly collected on farms that are already under surveillance by Bioversity International. This results in a convenience sampling design. Advantages are that the locations are known and accessible, Gros Michel is (recently) being cultivated and permission of the farmers to take the samples is no problem. Disadvantages of this sampling design are that the spatial distribution of the different locations is very clustered. This can result in the locations to be biased and thus not completely representative for the whole population (Bouwmeester et al., 2016). Despite these disadvantages, convenience sampling is chosen because the area is relatively large with only a small number of Gros Michel farms, and convenience sampling makes the collection of observations quick.

Even though most of the farms visited are under surveillance of Bioversity International, 4 locations are spotted while driving from one farm to another and samples are taken here as well.

In most of the locations variability in disease incidence within a farm is visible. If this is the case multiple observations are made within one farm. In total 21 farms are visited, 49 observations are made and 50 soil samples are collected. Two observations are performed along the road and do not have accompanying soil samples, but on some farms extra soil samples are taken to account for the high spatial variability within the soil.

At every farm the same procedure is executed for the field observations and soil samples. First the farm overall is observed and the farm owners are asked about the current condition of the banana plants concerning disease, age and production. With this information and a personal judgement, the farm is divided in different sections that show clear differences in disease incidence. The borders of these sections are usually decided by differences in landscape features such as aspect and upslope/downslope or by geographical elements such as roads or tree patches. For every section field observations are made and soil samples were taken. Every field is classified as either high or low incidence.

The disease incidence is also determined using a standard grading system (Appendix 1) resulting in a continuous range. This grade system exists of five classes of severity and for each class the percentage present in the area is estimated. For each grade the percentage is multiplied by the number of the grade (1-5) and all values are added. This results in a continuous variable describing the disease incidence on a scale from 100-500. Thus disease severity is not the proportion of plants affected, but rather the expression of the disease in the plants (peng et al, 1999). The field form can be found in Appendix 2.

The soil samples are collected with a 30 cm deep auger. In every section about 5-10 full augers are collected randomly dispersed throughout the area and mixed together. The distance between two augers is dependent on the size of the area. If it is not possible to auger the 30 cm depth due to rocks in the soil a new spot close to the old one is used. The samples are conserved in an icebox.

Soil samples are chemically analysed using a Mehlich lll extraction (Mehlich, 1984). The microbiological analysis is performed by counting the fungi, bacteria and actinomycetes on petri plates (CFU/g of soil) (Segura et al., 2015b). All samples are analysed in the soil and microbiology laboratories at CORBANA, Guápiles.

Statistical analysis

The statistical analysis is used to find environmental variables that influence the incidence of Fusarium wilt in banana plants. Variables that do show statistical significance with disease incidence can be of importance for future disease control of Fusarium wilt.

All 40 variables selected in this research have their own contribution to disease incidence, whether it might be in plant health or soil suppression. The interactions and relationships are very complex, and the aim of this research is not to completely understand all the interactions. In this research all these variables will be analysed and evaluated, with the goal to determine which variables are important for disease incidence. Those important variables are expected to confirm the results from previous pot- and field experiments. Only a brief analysis will be performed for the interactions between variables.

For the data analysis, four statistical approaches are used to capture all relations and variability between the variables and disease incidence. The four approaches are t-tests, linear regression, quadratic regression and stepwise regression. For all variables a normal distribution is assumed.

The independent t-tests are performed to find the variables that clearly show a dependency with disease incidence. For the t-tests the variables are split in two groups; high disease incidence and low disease incidence. If the t-tests have a p-value < 0.05 there is a significant difference between the averages for high and low disease incidence. The null hypothesis (p-value ≥0.05) is that there is no difference in mean between high and low incidence. The alternative hypothesis (p-value < 0.05) is that there is a statistically significant difference between the mean values of the two groups. If the null hypothesis is rejected, it implies that those variables are important for explaining variability in disease incidence.

Besides the t-tests also a regression analysis is performed. This regression analysis provides additional information on the relationship between disease incidence and the different variables. The t-tests show important information about the variables dependence on disease incidence, however not every relation can be captured with a single t-test. If the distribution of data points is non-linear it can be possible that the t-test did not fully capture the relation with disease incidence. In this research the regression analysis (i) provides extra statistical information on important variables and the relation with disease incidence and (ii) it can capture relations that are present, but which are not visible via a t-test.

It is not known what type of relation there is between environmental variables and disease incidence (i.e. linear, exponential, quadratic). With this reason, both linear and quadratic regression will be executed.

With a regression analysis the relationship between two sets of data is estimated with a formula, which enables a researcher to use this relationship for prediction purposes. The formula used for a linear regression is:

$$\hat{Y} = bX + a \tag{1}$$

In the case of many environmental variables, often an optimum at a certain value is present (Figure 7). Therefore also a quadratic regression analysis is performed. A quadratic regression is similar to a linear regression, except for the formula which is as following:

$$\hat{Y} = cX^2 + bX + a \tag{2}$$

The regression analysis also provides a measure of the size of the correlation (R-squared (R^2)) and the p-value, indicating the statistical significance of the relationship found. A relationship is found to be dominant when the $R^2 > 0.40$ and the p-value < 0.005. An R^2>0.40 meant that 40% of the variation in disease incidence can be explained by that one variable. In this type of research (field surveys for plant diseases) an R^2 of 0.40 is very high. A p-value < 0.005 is determined to tell when a correlation is statistically significant.

To account for interrelationships and multiple variables influencing the disease incidence at the same time, a forward stepwise regression analysis is performed. A stepwise regression tries to find the best possible combination of variables to predict disease incidence the best. It is different from a normal regression analysis, because it combines multiple variables to create the best fitting model for disease incidence prediction. Also it is a good alternative for multiple regression that is not executed in this research due to the high number of variables and low number of observations. In a forwards stepwise regression the model starts with one variable and with each step adds a new variable that improves the model. This happens until adding another variable makes for a worse model than before. As mentioned earlier, in this research a large set of variables with an unknown contribution to disease incidence are used. A forward stepwise regression extracts a few of these variable and creates the best fitting model with less risk for overestimation.

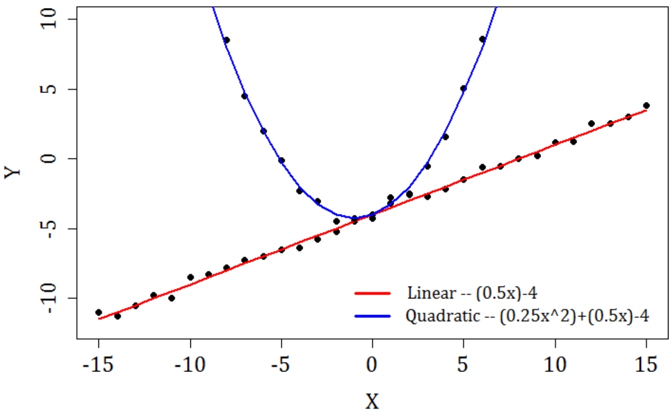


Figure 7
Example of two regression graphs. The red line shows a typical linear regression and the blue line illustrates the shape of a quadratic line with an 'optimum' around x = -1.

In this research, to prevent overestimation, the stepwise regression is executed separately for each sector of variables. In the end all important variables are combined into one last stepwise regression. Thus, accounting for interactions between variables from different sectors as well as to find the most important explanatory variables.

Results & Discussion

In this chapter all the results collected during this research are presented and discussed. The results show the high variability in landscape, the high variability in disease incidence and a number of variables that explain part of the variability in disease incidence.

This chapter starts with the variability assessment and the accompanying stratification map. The great variability in the area is illustrated, showing how many landscape variables are important for the high variability. The high variability in this areas shows that in this research the landscape and climate variables cannot be neglected for disease incidence. After the variability assessment the results from the quick scan data will be presented and extensively discussed. First some general remarks on the quick scan will be made and the variability in disease incidence as observed will be described. Later, the results for all data samples will be described per sector; landscape and climate, Land management and soil conditions (soil chemistry and Microbial populations). The data will be described and discussed with the results from the statistical analysis (t-tests and regression analysis).

These result show that all sectors have variables that can explain part of the variability in disease incidence. The results for these variables will be presented and discussed. The goal of this part is to confirm the dependency of disease incidence for variables that were found before in pot- and field experiments. The main variables this research attempts to confirm in a field survey are intercropping (crop cover), green cover, pH, Acidity, Calcium, Magnesium, Manganese, Nitrogen and microbial populations. The variables that show dependency with disease incidence in these results are profile curvature, temperature, plant density, crop cover, bare soil, pH, Acidity, Magnesium, Potassium, Phosphorous and Fungi concentrations.

At the end of each section a small conclusion and recommendations will be provided. At the end of this chapter the overall discussion is presented and possible implementations for the results are given.

Landscape and climate variability

The goal for this variability assessment is to find how high the spatial variability in landscape and climate is within the study area. Most large scale plantations have much less landscape variability than this study area. Also in the pot- and field experiments, of which this research tries to confirm the results, variability in landscape in minimal. In the study area, landscape variability is high compared to other banana growing areas and therefore landscape and climate can also explain variability in the disease incidence. To account for this, the stratification map is created to illustrate the high variability and, during the data analysis for the quick scan data, all landscape and climate variables are also analysed.

For the creation of the stratification map the following variables are used: altitude, slope, aspect, profile curvature, plan curvature, hillshade, distance to river, NDVI, total annual precipitation, mean annual temperature, total precipitation in the driest month and mean temperature in the coldest month. The mean and standard deviations for all used variables are shown in Table 2 and extensive summary statistics can be found in Appendix 3. From Table 2 and Appendix 3 it is clear that all included variables have a high individual variability within the study area. The standard deviations are high relative to the variable means.

Relative importance of variables

Table 3 shows the correlation matrix of the twelve variables and Table 4 provides the eigenvectors and eigenvalues that represent the results of the PCA. The correlation matrix shows the linear correlations between variables and gives an indication about which variables will be combined in the same PCA layers, in order to minimize the dependency between layers. The eigenvectors and eigenvalues provide information about which variables are dominant in each PCA layer and how much of the variance can be explained by each individual layer.

In the correlation matrix it is clearly visible that altitude is highly correlated to all the climate variables. This is not surprising since both temperature and precipitation are partially explained by the altitude of an area. A clear correlation that is also not surprising is the correlation between hillshade and slope and aspect. Hillshade is a representation of the shadows created on the surface under a certain position of the son. The steepness of the slopes is important for the size of a shadow under a certain altitude of the sun. Also the azimuth of the sun with respect to the slopes is directly correlated to aspect. Therefore the correlation between aspect and slope is expected. At last in the correlation matrix it is also visible that profile and plan curvature are correlated and that all climate variables are highly correlated to each other. Neither of the correlations are unexpected.

Table 2
The mean and Standard deviation values for all twelve variables included in the variability assessment for the study area located surrounding Turrialba, Costa Rica

	Mean	Standard Deviation		Mean	Standard Deviation
Altitude (m)	961.00	397.00	Distance to River (m)	317.16	266.24
Slope (%)	15.94	9.75	NDVI (-)	0.44	0.08
Aspect (Radians)	0.03	0.72	Annual Precipitation (mm)	3031.00	322.00
Profile Curvature (1/100 m)	0.01	0.55	Annual Temperature (°C)	21.23	2.13
Plan curvature (1/100 m)	0.02	0.46	Precipitation driest month (mm)	115.16	30.78
Hillshade (-)	232.49	20.46	Temperature coldest month (°C)	20.25	2.09

Table 3
Pearson correlation coefficients for all twelve raster variables used in the variability assessment. A description of all variables is present in Table 1. Prof_curv and Plan_curv are profile and plan curvature respectively, DisToRiv is the distance to river, Annual_prec is the annual precipitation, Dryest_month is the total precipitation in the dryest month, Annual_temp is the annual mean temperature and Coldest_month is the mean temperature form the coldest month

	Alti-tude	Slope	Aspect	Prof_curv	Plan_curv	Hill shade	DisTo Riv	NDVI	Annual_prec	Dryest_month	Annual_temp	Coldest_month
Altitude	1.00											
Slope	0.06	1.00										
Aspect	-0.16	-0.03	1.00									
Prof_curv	-0.06	-0.02	0.00	1.00								
Plan_curv	0.01	0.00	0.00	**-0.51**	1.00							
Hillshade	0.06	**-0.66**	**-0.61**	0.01	0.00	1.00						
DisToRiv	0.06	0.12	0.00	-0.08	0.03	-0.08	1.00					
NDVI	0.07	0.24	0.04	-0.01	-0.02	-0.17	0.07	1.00				
Annual_prec	**-0.36**	0.02	0.05	0.00	0.00	-0.05	0.06	0.09	1.00			
Dryest_month	**-0.54**	0.01	0.05	0.00	0.00	-0.05	0.08	0.08	**0.85**	1.00		
Annual_temp	**-0.98**	-0.06	0.19	0.02	0.01	-0.09	0.00	-0.05	**0.36**	**0.55**	1.00	
Coldest_month	**-0.98**	-0.06	0.19	0.02	0.01	-0.08	-0.01	-0.05	**0.37**	**0.55**	**1.00**	1.00

Table 4
Principal Component Results for the variability assessment. PCA eigenvectors (above) and the PCA eigenvalues, eigenvalues in percentage and the sum of all percentages (below)

PCA Layer /Variables	1	2	3	4	5	6	7	8	9	10	11	12
Altitude	**-0.48**	0.08	0.04	-0.19	0.19	0.11	0.04	0.00	-0.08	0.01	**0.81**	-0.01
Slope	-0.01	**0.55**	-0.01	-0.16	**-0.35**	-0.31	**0.42**	-0.01	0.05	**-0.53**	0.02	0.00
Aspect	0.13	**0.37**	-0.10	**0.41**	**0.48**	0.37	-0.26	0.01	0.02	**-0.48**	-0.01	0.00
Prof_curv	0.02	-0.06	**-0.69**	-0.12	-0.03	0.04	0.03	**0.71**	0.00	0.00	0.02	0.00
Plan_curv	0.00	0.03	**0.68**	0.16	0.06	-0.11	0.03	**0.70**	0.00	0.00	-0.01	0.00
Hillshade	-0.08	**-0.66**	0.08	-0.17	-0.08	-0.03	-0.17	0.00	0.07	**-0.69**	0.02	0.00
DisToRiv	0.01	0.15	0.16	-0.26	**-0.53**	**0.75**	-0.17	0.07	0.03	0.01	-0.04	0.00
NDVI	0.00	0.29	0.01	**-0.35**	-0.02	-0.39	**-0.80**	0.03	0.01	0.03	-0.01	0.00
Annual_prec	**0.32**	0.02	0.09	**-0.51**	**0.40**	0.10	0.18	0.00	**0.65**	0.05	0.02	-0.01
Dryest_month	**0.40**	-0.01	0.08	**-0.43**	0.26	0.07	0.13	0.00	**-0.75**	-0.07	-0.02	0.00
Annual_temp	**0.49**	-0.06	-0.01	0.18	-0.21	-0.07	-0.07	-0.01	0.05	0.01	**0.40**	**-0.71**
Coldest_month	**0.49**	-0.06	-0.01	0.18	-0.21	-0.07	-0.07	-0.01	0.06	0.02	**0.41**	**0.70**
EigenValues	3.66	1.93	1.50	1.33	0.93	0.92	0.80	0.47	0.13	0.08	0.02	0.00
EigenValues (%)	31.05	16.43	12.73	11.30	7.93	7.79	6.81	4.02	1.08	0.70	0.16	0.00
Sum (%)	31.05	47.47	60.20	71.50	79.43	87.23	94.04	98.06	99.14	99.83	100	100

Table 4 provides the results of the PCA and shows the contribution of the variables to the variance in the area. During the PCA the twelve variables are converted into twelve layers, where each following layer captures less variance and is thus less important. The first 4 layers explain more than 70% of the variance from the in total 12 variables, while the last four layers explain not even 2% of the total variance. The first layer accounts for 31% of the total variance and via the eigenvectors it is visible that it represents altitude (-0.48) and all climate variables (0.32; 0.40; 0.49; 0.49). Since these variables are heavily correlated to each other, during the stratification this layer dominance represents only altitude. The second layer mainly represents hillshade (-0.66), slope (0.55) and aspect (0.37). The third principal component clearly represents the curvature variables (-0.69; 0.68), while the fourth component does not has a clear variable which it represents dominantly, but is more a collective of variables that are already accounted for except for NDVI (-0.35). The eigenvectors of all remainder layers are shown in Table 4 and the variables dominating each layer are highlighted.

Stratification

All PCA layers were clustered with the Iso Cluster Unsupervised Classification, creating 25 clusters within 150 iterations. After 150 iterations no visible changes were present anymore between two following maps. The created clusters where based on the PCA layers and its importance, thus the variables representing the first few layers had a higher relevance for inclusion in the cluster creation than those representing the less important layers. The 25 final clusters were the result of only statistical calculations, but these 25 clusters were manually converted into 14 strata representing the dominant landscape variance in the area.

The conversion from 25 clusters to 14 strata was done in four steps. This conversion was done to make the stratification map as much as possible connected to the possibilities for banana production. Thus the stratification map shows the variability in the area that can influence banana production, plant health and thus disease incidence.

First, the natural areas and low NDVI areas were defined in the cluster map. The natural areas are protected areas and these areas are not accessible for this research, neither are they available for banana production. All areas with a low NDVI are also not suitable or available for agriculture since they represent cities and waterbodies (NDVI < 0.39) (Leal Filho et al., 2016). The second step was to merge those clusters that contained less than 1% of all total cells with its surroundings. The third step was to combine certain clusters that do not have a difference significant enough to account for. An example is the combination of two classes in the lower lying areas that were distinct from each other in curvature characteristics. In the lower lying areas landscape variability is already decreasing and differences in curvature are less dominant than in the middle and high elevated areas. Differences in curvature in the lower areas are, of course, present, however for the purpose of disease incidence, curvature is less important in the lower lying areas. In the end, fourteen strata were defined and to remove the scatterings in the map and make the strata more cohesive the function 'sieve' is used to finish the stratification map.

Description of the stratification map

The stratification map is illustrated in Figure 8. The following paragraphs describe the distinction and importance of the different strata, the bigger landscape units visible in the map, the high variability in the area and its implication on the research and disease incidence.

As visible in Figure 8 the map is divided in four main classes; A: the two items in the map that are present in the whole area and are not necessarily explaining landscape variability, but representing those areas unsuitable for agriculture; B: the higher elevated areas, where temperature is usually lower and precipitation higher; C: middle elevated areas where the main landscape variability is visible and also representing the area's mostly used for Gros Michel production; D: the relatively lower elevated areas, where topography becomes less variable and a clear river valley starts to develop. In Table 5 all strata, the division criteria and hierarchy is shown.

The first two strata (low NDVI and natural areas) represent those dominant strata in the area that are not suitable for the cultivation of crops. The natural areas are protected nature and the areas with a low NDVI illustrate the location of the cities, villages and waterbodies in the study area. The criterion for NDVI means that an NDVI < 0.39 is not suitable for agriculture due to the presence of buildings, infrastructure or waterbodies. This value is relatively high compared to literature (Senay & Elliot, 2000; Bandhari et al., 2012), however in this map, as it is checked with satellite imagery, a value of 0.39 had the best fit, which is a value comparable to the NDVI values found by Leal Filho et al., 2016 for Dutch city cover.

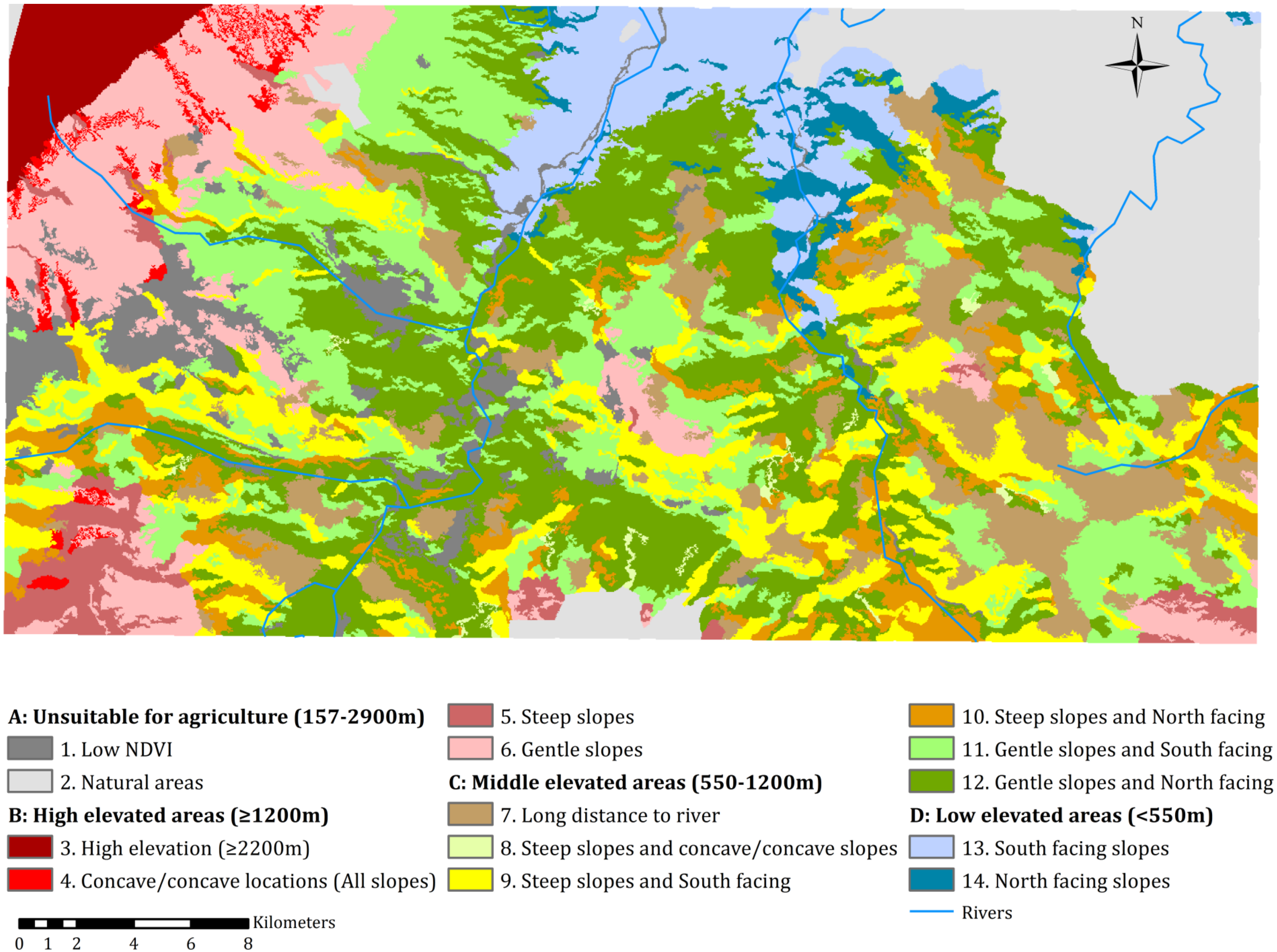


Figure 8
Stratification map of the study area.

Table 5
All criteria used to divide the study area in fourteen strata. The first column shows the hierarchy of the variables per elevation class, while the hierarchy between elevation classes is determined by A (highest) to D (lowest). The hierarchy tells which stratum was selected before the others were included

Hierarchy	Strata	Division Criteria
1.	(A.2)	Natural Parks
2.	(A.1)	NDVI < 0.39
1.	(B.3)	Elevation ≥2200
2.	(B.4)	All slopes; Profile Curvature > 0; Plan Curvature < 0
3.	(B.5)	Slope ≥ 20%
4.	(B.6)	Slope < 20%
1.	(C.7)	Distance to River ≥ 500m
2.	(C.8)	Slope ≥ 20%; Profile Curvature > 0; Plan Curvature < 0
3.	(C.9)	Slope ≥ 20%; Aspect <0
4.	(C.10)	Slope ≥ 20%; Aspect ≥0
5.	(C.11)	Slope < 20%; Aspect <0
6.	(C.12)	Slope < 20%; Aspect ≥0
1.	(D.13)	Aspect < 0
2.	(D.14)	Aspect ≥ 0

In the higher elevated areas there is one stratum representing the area that is equal or higher than 2200m AMSL. Above this height, the temperatures are too low for good banana production. Also, slopes are much steeper on higher elevated areas and erosion is larger. 27 °C Is the optimal mean temperature for banana growth and development (Robinson & Saúco, 2010), while in the study area, above 2200 meters the mean annual temperature is everywhere below 15°C, which is too low for banana cultivation (Calberto et al., 2015). The next stratum represents all concave/concave slopes (head slopes) within the higher elevated areas. This means that there is a positive profile curvature and a negative plan curvature. If the curvature value is zero, the slope is straight and if the value is negative for profile curvature and positive for plan curvature the slope is convex. Concave/concave slopes are important due to the fact that water and sediments accumulates in the down middle of the slope. These factors result in a typical ecology and soil development that is different from other slope conditions (Daniels & Hammer, 1992; Busacca et al., 1993). This, in its turn can influence the nutrient availability and the soil microbial activity, which is expected to be related to disease incidence.

The other two strata in the high elevated areas represent steep or gentle slopes. Steep slopes have more surface runoff and more erosion compared to gentle slopes, however gentle slopes are in a disadvantage when it comes to optimal sunlight conditions. The slope criterion for steepness is 20%, which is chosen arbitrary with personal knowledge about slope processes and the study area.

The middle elevated areas cover the biggest part of the map (60%), have the highest spatial variability and all quick scan observations are located in the middle elevated areas. The first stratum includes all areas that are located far from the rivers and streams in the area. In Figure 8 only the main rivers are illustrated, but rivers and streams can be found throughout the whole area. This stratum illustrates all locations that are more than 0.5km removed from the nearest river, where influence of the river on soil conditions and banana cultivation is negligible. The second stratum illustrates the few concave/concave locations within the steep slopes, mainly present in the south-middle of the map, bordering the natural areas. The importance of the concave/concave locations is the same here as for the higher elevated areas. However, here only the concave/concave locations on the steep slopes are highlighted, while the concave/concave locations on the gentle slopes are not included in a separate stratum. This is because the importance of concave/concave locations is better expressed on steep slopes, and because the number of concave/concave locations on the gentle slopes was too small to include. The next four strata illustrate either south or north facing steep slopes and south or north facing gentle slopes. The difference here is in both slope steepness and aspect. Aspect is divided in north and south, since number of sun hours is important for banana cultivation, and south facing slopes receive more sun than north facing slopes. As illustrated in the Materials & Methods, an aspect ≥ 0 is north facing and has an aspect in degrees between 0 and 90 degrees or 270 and 360 degrees. If aspect is < 0 the slopes are south facing and in degrees this means that the slope is between 90 and 270 degrees.

In the lower elevated areas, there is only a distinction between north and south slopes. This area is the beginning of the large outstretched plains in the west of Costa Rica, where the large Cavendish plantations can be found. These areas have much lower spatial variability concerning landscape features. In these areas soil variability starts to dominate over landscape variability. Nonetheless, aspect is important because there is still a lot of relief in these parts.

Interpretation

In this stratification map a number of larger landscape patterns are clearly visible. In the north-west corner there is a rapid increase in elevation, which represents the foot of the Turrialba volcano. Also the river pattern in the area is visible through the distinction between steep and gentle slopes in the middle elevated areas. The gentle slopes are closer to the main rivers, while the steep slopes are often farther from the rivers. At last, the incision of the rivers is clearly visible in the middle of the map, where there still is a high elevated area surrounded by two rivers in a V-shape closing in to each other.

The map clearly shows that the spatial variability in this area is very high. In this map altitude, slope, aspect, profile curvature, plan curvature, distance to river and NDVI are all included and have their own role to capture the variability in the study area. There are many small patches of strata present in the whole of the map and each stratum has its own shape and size. It is important to keep in mind that a stratification map is already a simplification of reality, and that this map is based on a 30x30 grid, while in reality there can even exist spatial variability within each 30x30m grid cell.

The main thing both the map and the description above illustrate is the high spatial variability in the area. It shows that the cultivation of bananas and the accompanying disease incidence is not a straightforward problem, but rather a complex problem. In this problem, many variables have their own role and influence on the disease incidence, making research less predictable and harder to execute. On the other hand, this high variability is very important to take into account, since it can be a significant predictor for disease incidence. These variables are usually not assessed, because the main banana cultivation takes place in more homogeneous landscape areas. In the study area, however, the landscape is very heterogeneous and might thus explain part of the variability in disease incidence. Therefore the landscape variability should be included when endeavouring to explain variability in disease incidence.

Quick scan

In this section the high variability between farms in disease incidence is briefly discussed. First variability in disease incidence will be described. Second, some restrictions in observing disease incidence are mentioned. Finally, also some brief remarks on variability in land management will be made.

Disease incidence

During the field work it became clear that variety in disease incidence is high. All farmers in the area wanted to grow Gros Michel banana, since it used to be a good and productive crop with a good price on the market. Even now that the disease infiltrated the whole area, farmers still want to grow Gros Michel. Remarkably some farmers succeed very well in this, while other farmers almost give up on Gros Michel production. Disease incidence is very variable in the area, ranging from farms where no plant survives longer than six months to areas where Gros Michel plants are eight to ten years old and are still yearly producing bananas. Sometimes differences in disease incidence were even present within a farm. An example was a farm in a valley, where the valley fields all had a very high incidence, and the two fields on the slopes of the valley had low disease incidence. Another farm had only very young Gros Michel plants, because they die before they produce a bunch. However, one Gros Michel plant was already 3 years old and still showed no signs of disease incidence. This 3 year old plant was fully healthy, but did not produce a single bunch in those three years.

The high variability in disease incidence in the study area is a problem for the local farmers. The areas that show no to low incidence are not secured to have the same low incidence in a few years. Another farm was visited, which had a very high incidence and almost no single plant survived. However, ten years ago at that exact location the plants were healthy and productive.

Disease incidence in the study area is highly variable. Mainly spatial variability is observed, but also in time disease incidence can change. This makes for an uncertain future of Gros Michel production in the study area, however the presence of variability provides opportunities for research. Due to the high variability, it can be concluded that there are variables that can explain this variability in disease incidence. With the right explanatory variables the consequences of Fusarium wilt can be minimized.

Figure 9
Dead plant with a new pseudostem erupting out of the dead mother plant.

Restrictions

One restriction in the field observations is the way the disease incidence is observed. For the disease incidence itself a standard scheme was used (Appendix 1), however there were two important restrictions; (i) The scheme assumed that the plants being observed were all adult plants and of the same age and (ii) the symptoms for Fusarium wilt are not always the same and sometimes look like other diseases or certain nutrient deficiencies (Stover, 1962). The first restriction was sometimes difficult to deal with, since in these farm sometimes plants of all different sizes and ages where standing mixed in the field. The problem with this is that younger plants show less symptoms than older plants under the same infection rate. Especially for very young plants (<6 months) that seldom show any symptoms (Zhang et al., 2013). When this was the case the best estimation possible was made, with a side note mentioning the different ages present in the field.

The second problem was less problematic than the first problem, but nonetheless just as important. Not all Fusarium wilt symptoms are the same for every plant and also other diseases or deficiencies can cause similar symptoms, for example leaf yellowing. Usually it was possible to differentiate between the Fusarium wilt, other diseases and nutrient deficiency due to the knowledge of the Costa Rican experts who joined the fieldwork. If the distinction was still not completely clear, the farmer was asked, who often knew the answer, because of previous experiences.

Another remarkable thing that was observed during the field work, was the process of a dead plant trying to survive. In some cases the plant was dead due to Fusarium wilt, but a new plant erupted out of the dying pseudostem (Figure 9). This is the plant trying to survive, however the chances of this plant producing any bananas is as good as zero. Therefore during the observations, plants that showed this behaviour where assumed dead, even though no direct or lesser disease symptoms where visible on the leaves.

Variability in land management

During the fieldwork a lot was observed, both via data and personal experiences. Interesting variability in land management was noticed while executing the fieldwork. The land management between farms is very different. There was a big difference in the producing capacity between farms and the absolute sizes. Some farms had clustered bananas, or no structure at all. Other farms where very productive, very structured and implemented the system and banana plant pattern that is also used for the big scale Cavendish plantations. Differences between farms were also visible in the proportion of intercropping and the crops used for intercropping. At last also big differences in soil cover were present. The previous sections showed the high variability in landscape, land management and in disease incidence. The next step is to find those variables that can explain this incidence in disease.

Role of the environment on disease incidence

In the next part the results of the statistical analysis for all variables will be discussed. While reading the results and interpretations it is important to keep in mind that the landscape in the area is very irregular (See Variability assessment) as are the farms in land management and size. Also, there are only 50 data points used for the regression analysis resulting in overall low R-squared values.

For the t-tests half of the locations were low incidence and the other half of the locations had a high incidence. The averages for each variable for high and low incidence are compared to each other in a t-test. For the regression analysis the continuous disease incidence data is used, as described in the Materials & Methods section.

During the discussion of the results, all variables with p-value < 0.05 for the t-tests will be discussed. Also all variables with R-squared higher than 0.20 and p-value < 0.005 will be discussed. Even though an R-squared of 0.20 is not statistically high, it still means that 20% of the variability in disease incidence can be explained by this variable. Within this type of research (crop diseases and field surveys) and with the low number of data points available (50), 20% explained variability is high.

Landscape and Climate variables

For landscape and climate variables only profile curvature showed a significant difference (p-value = 0.034) in mean values for high and low disease incidence. The regression analysis also showed a correlation (R^2 = 0.206; p-value = 0.005) between disease incidence and mean annual temperature. These two results are described and discussed below.

Profile curvature seems to have a significant (p-value = 0.034) difference between high and low incidence. A high incidence is accompanied with a negative mean profile curvature (-0.168) that represent convex slopes, while a low incidence has a mean value of 0.079 and is thus more present in the straight to concave locations. The reason why convex slopes have a higher incidence can be explained by the fact that convex slopes loose more nutrients in the soil due to soil erosion and nutrient leaching with water infiltration (Daniels & Hammer, 1992). Besides this, convex slopes are usually also dryer and thus more prone to decreased plant health and soil suppression.

In Figure 10 the scatterplot for temperature against disease incidence is shown. Temperature is important for plant growth in the subtropics, and might as well be important in the tropics for plant health and consequently also for disease resistance. However, it is questionable whether temperature also influences soil suppression (brake et al., 1995). Temperature has a quadratic relationship (R^2 = 0.206; p-value = 0.005) with a depression in disease incidence when the temperature is neither too high nor too low. Thus there seems to be an optimum in temperature for disease resistance.

It is expected that when temperature decreases, incidence can become higher. High temperatures are important for plant growth and development and under lower temperatures the plant health decreases and becomes thus more susceptible to the disease. Why the incidence increases while the temperature gets higher is not entirely clear, but it can be correlated to other processes at higher temperatures such as the fungi itself preferring higher temperatures for reproduction or a higher plant growth rate, inducing a higher plant stress and higher susceptibility to disease incidence.

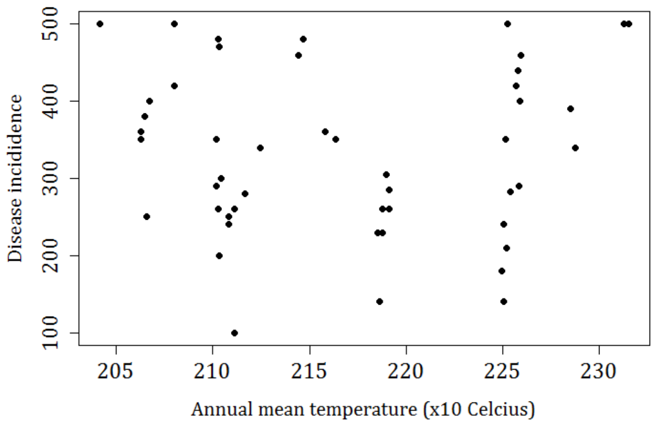


Figure 10
Scatterplot of temperature values on the x-axis and disease incidence on the Y-axis.

It is also remarkable that temperature has a relatively high R-squared, while the altitude, which is heavily correlated to temperature (-0.913), has not a high correlation. This is due to differences between altitude and temperature in the lower areas. When looking at all data points that have an altitude of 900 and higher the altitude-disease incidence relation also has an R-squared of 0.206, where disease incidence increases with a higher altitude.

A forward stepwise regression is also performed for all landscape variables in an attempt to find the best fitting model by combining different variables. Unfortunately for the landscape variables no good fit was found. The best model fit that was found was a combination of the variables slope and plan curvature. Slope had a significance code of 0.01 and plan curvature had a significance code of 0.1. The R-squared had a value of 0.153 with a p-value of 0.0219. With these results it does not seem that there is a good model fit for disease incidence when only using the landscape variables.

All landscape variables are assessed and even though some variables (profile curvature, temperature) appear to be connected to the disease incidence, the landscape variability in this area is high, thus limiting the results with only 49 observations. The results for all other landscape and climate variables are shown in Table 6.

The discussed results imply that only one landscape variable explains part of the variability in disease incidence. In this research many variables are considered and some variables might be more dominant to disease incidence than others and it is not known which are dominant. This does not mean that landscape does not influence disease incidence. Temperature, slope and aspect are important factors for banana growth and disease resistance in the subtropics (Robinson & Saúco, 2010; Pattison et al., 2014), but influence of these variables in the tropics, where temperature and precipitation are no limiting factors, is unknown.

The results for curvature and temperature shows that landscape variability can definitely not be ignored, but in this research the number of observations are too low to cover the high variability in the area. It is recommended to do further research on the relations between landscape variables and disease incidence, but more observations are needed as well as detailed, local landscape observations obtained in the field instead of via a DEM.

Table 6
Results of the statistical analysis of the data for all landscape and climate variables. The t-tests show to what extend the mean values between high and low disease incidence for each variable differ. The regression analysis shows the size of correlation present between disease incidence and each variable. The Q stands for a Quadratic regression and the L stands for a Linear regression. Whether a regression line is significant or not is indicated by a Y for Yes and a N for No. The meaning of the abbreviations is the same as in Table 3

	T-test	Regression analysis		
	P-value	R-squared	P-value	Significant (Y/N)
Altitude	0.286	0.052 (Q)	0.292	N
Slope	0.385	0.098 (L)	0.028	N
Aspect	0.132	0.028 (Q)	0.518	N
Prof_curv	**0.034**	0.019 (L)	0.351	N
Plan_curv	0.222	0.072 (Q)	0.180	N
Hillshade	0.065	0.093* (Q)	0.117	N
DisToRiv	0.747	0.020 (Q)	0.622	N
NDVI	0.298	0.019* (L)	0.349	N
Annual_prec	0.659	0.006 (Q)	0.867	N
Dryest_month	0.908	0.030 (Q)	0.491	N
Annual_temp	0.991	**0.206 (Q)**	**0.005**	N
Coldest_month	0.936	0.169 (Q)	0.014	N

* Outliers removed

Land management variables

The t-tests for land management show that there are significant differences between high and low incidence areas for number of leaves, crop cover, bare soil and distance between mats (p-value < 0.005). With the regression analysis plant density has a dominant relation with disease incidence (R^2 = 0.444; p-value = 0.000). Also the number of crops, proportion of other crops (crop cover) and the distance between mats have a high R-squared and will be discussed.

The number of leaves does show a significant difference (p-value = 0.000) between high and low incidence. A high disease incidence has on average 5.32 leaves per plant, while plants with low incidence have an average of 6.96. Number of leaves is an important indicator for the plant health, and often older leaves are removed from the plant to support the nutrient flow to the younger parts of the plant. The number of leaves however is not an explanatory variable for disease incidence. These results show a low number of leaves (average 5) is correlated to high incidence and high number of leaves (average 7) to a low incidence. Unfortunately the lower number of leaves is due to the disease and disease itself is not influenced by the number of leaves. Number of leaves is thus an important indicator for plant health, but cannot be adapted for disease control.

Crop cover also has a p-value lower than 0.005 (0.001) for the t-tests supporting the findings by Zhang et al. (2013), which showed the importance of intercropping for incidence of Fusarium wilt. Crop cover in this research is the proportion of other crops within the area, not taking the Gros Michel crop in account. The high incidence areas have an average proportion of secondary crops of 20.87 and the low incidence areas have an average of 54.42 (Figure 11). These results show that low incidence is correlated to a high proportion of crop cover on the farm. An important argument that supports this is the fact that cropping bananas in combination with a second crop, or intercropping, has a positive effect on the soil health and microbial activity and thus on the suppression of the disease in the soil. Resulting in a higher crop yield (Song et al., 2007). The t-test results for crop cover confirm the results found before in a field experiment by Zhang et al. (2013). Intercropping is thus an important variable to consider when improving disease control via soil management.

Both number of crops and crop cover also show a, even though not dominant, linear relationship with disease incidence (R^2 = 0.230, p-value= 0.000). However, these results are somehow contradictory. The regression coefficients shows a positive linear relationship, where high incidence is accompanied by high percentage of crop cover. The t-test however shows a clear high incidence with low crop cover percentages.

The correlation coefficients for number of crops and crop cover are biased due to the fact that many observations did not have any secondary crops and thus crop cover is 0%. Of all observations, 30% has a crop cover value of zero, which causes a bias in the regression analysis. When removing all zero values and running the regression again for crop cover the r-squared is 0.134 with a p-value of 0.03. These results are much lower than the original ones. For a t-test the presence of the zero's is visible in the median, but the effect is much smaller for a t-test than for a regression analysis. In Figure 11 the boxplot shows the clear difference between the high and low incidence values. The black lines show the medians of the two groups, while the red and the blue line represent the mean values for high and low incidence respectively.

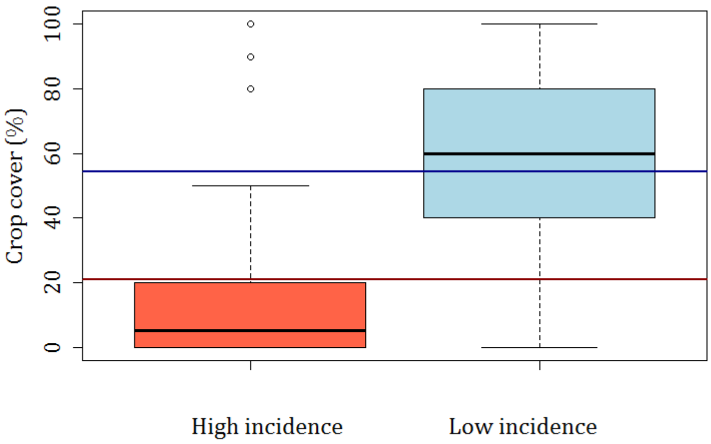


Figure 11
Boxplot for the proportion of secondary crops between the banana plants for high (red) and low (blue) incidence. The Y-axis shows the proportion of crop cover. The black lines are the median values, the outsides of the box are the lower and upper quartiles and the red and blue line are the mean values for high and low incidence respectively.

Beside crop cover also the proportion of bare soil shows a significant difference (p-value = 0.004) between high and low incidence areas (Figure 12). High incidence areas have an average of 30.13% bare soil and the low incidence areas have an average of 11.65% bare soil. This implies that a high percentage of bare soil around the banana plants is related to a higher disease incidence.

In literature the effect of green cover is shown to be important for a lower disease incidence of Fusarium wilt. These results do not show this for green cover, however they do show that low incidence has a lower percentage of bare soil than high incidence. Automatically the opposite result counts for the sum of percentages for green cover and brown cover that, together with bare soil, are always 100%. Green cover (grass and weeds between the plants) results in lower disease incidence because it stimulates microbial activity, improves the water holding capacity and reduces soil erosion. Brown cover (dead organic material such as banana leaves and mulch) mainly is important for the contribution of new organic material and the nutrient circulation in the soil. With these results it is not possible to say to what extend green cover is more important than brown cover. However, the results do show that the presence of bare soil should be avoided and be replaced by either green cover or brown cover. These results confirm, indirectly, the results Pattison et al. (2014) found during a field experiment in Australia.

Finally, the t-tests also shows a clear difference between high and low incidence areas for distance between mats (p-value = 0.001) Also a relation between disease incidence and distance between mats (R^2 = 0.306; p-value = 0.001) and a dominant relation between disease incidence and plant density (R^2 = 0.444; p-value = 0.000) was found in the regression analysis.

Plant density and distance between mats are negatively correlated to each other (Spearman = -0.69) and therefore these two variables are closely related. The difference between the plant density and distance between mats is present in the field, where density is estimated over the whole observation area, while distance between mats is variable even within the observation area and is thus harder to estimate.

If the density is high, the distance between mats is automatically small. High incidence areas are mainly where distance between mats is small (average = 2.76 m), while low incidence areas have a larger distance between mats (average = 4.67 m).

Mats that are close to each other induce a lot of

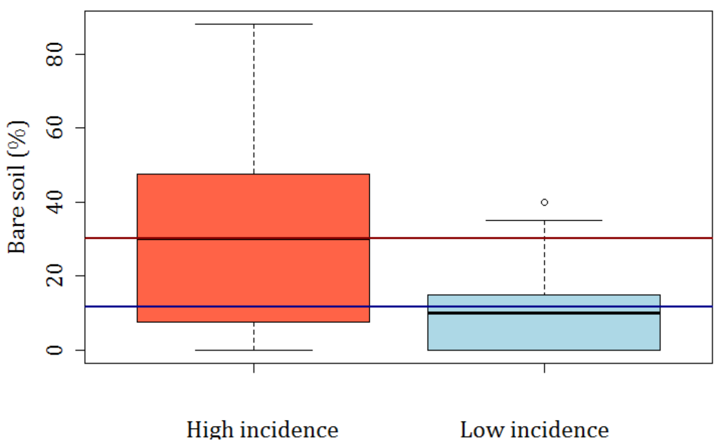


Figure 12
Boxplot for the proportion of bare soil for high (red) and low (blue) incidence. The Y-axis shows the proportion of bare soil. The black lines are the median values, the outsides of the box are the lower and upper quartiles and the red and blue line are the mean values for high and low incidence respectively.

competition between the banana plants. This competition can be in nutrient availability, water availability, rooting space in the soil and available sunlight (especially for the younger and smaller plants). Secondly, all these plants need water and nutrients and with a small distance between mats, soils can be depleted in essential elements for both plants and soil life. Above all, if the plants are closer to each other the chances that the plants infect each other with Fusarium wilt are higher than when plants have more individual space (Burdon & Chilver, 1982). A small distance between mats in the farm reduces plant health via competition, reduces soil suppression by soil depletion and increases the chances of infection.

The density of the banana plants in the field (mats/m^2) has a significant quadratic relationship (R-squared = 0.444; p-value = 0.000) with the disease incidence in the area showing partly the same results as for distance between mats. The scatterplot for disease incidence and plant density is shown in Figure 13. This quadratic relation shows a decreasing disease incidence with an increasing plant density until a depression is reached and the disease incidence increases again with even higher plant densities.

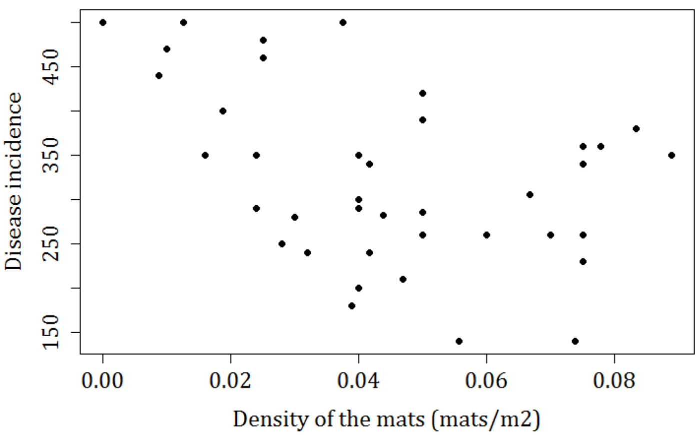


Figure 13
Scatterplot with plant density values on the x-axis and disease incidence on the Y-axis.

Opposite to a high density and small distance between mats, a low density also has a higher incidence. A reason for this is that microbial activity is lower, because there is less rooting activity and water movement in the soil. This can results in less fertile soils and thus a higher incidence. Also, plants that are far apart from each other are not able to shelter each other for damaging environmental condition such as heavy winds and surface runoff, creating soil erosion and thus resulting in a decreased plant health (Robinson & Saúco, 2010). However, it is also possible that this is not a cause-effect relation, but rather an effect-cause relation. Farmers want to have good production and if a crop (Gros Michel) is not producing enough due to the high disease incidence, they change their main crop from Gros Michel to another crop that does produce enough yield to maintain the farm. This results in less attention to the Gros Michel bananas and consequently less banana plants and a lower density. Whether a low density causes high incidence, or the opposite cannot be determined from these results, but the effects for plant density are visible and need to be considered when finding optimal conditions for banana cultivation in infected areas. These results also support the conclusions made in Burdon & Chilvers (1982) that Plant density cannot be neglected when looking at disease control.

For land management, number of shoots and tree cover appear not to be important for disease incidence. This is not remarkable and was in the line of expectation. Number of shoots is hard to use as an explanatory variable for disease incidence. The number of shoots is dependent on the age and the health of the plant, however most of the time shoots are manually removed from the plant by the farmer. This is done to create an optimal growth rate in the plant. If excessive shoots are cut off, these shoots don't need any additional nutrients and water to stay alive. These nutrients and this water can then be used for the primary shoots that will develop into a bunch caring plant. Tree cover is also less important. Tree cover is mainly present on the edges of the field, and therefore do not contribute to the microbial activity within the field. Green cover and brown cover also have no direct relations, but as discussed for bare soil, indirectly the sum of green and brown cover does make a difference for the disease incidence. The results of the statistical analysis for all variables are shown in Table 7.

For land management also a stepwise regression is executed. The stepwise regression for land management variables created are reasonably well model fit to explain the disease incidence with multiple variables. The obtained R-squared is 0.5369 with an accompanying p-value of 0.000. The following variables are included in this model, from low to high statistical significance; Green cover, Crop cover, distance between mats and number of leaves. The combination of these four variables explains over 50 % of the total variance in the disease incidence. This clearly shows how multiple variables can be of importance for the disease incidence, even though individually they do not show any relations or the reverse. It also clearly shows the complexity of all interactions that, together, determine the severity of the disease in the plants.

Table 7
Results of the statistical analysis of the data for all land management variables. The t-tests show to what extend the mean values between high and low disease incidence for each variable differ. The regression analysis shows the size of correlation present between disease incidence and each variable. The Q stands for a Quadratic regression and the L stands for a Linear regression. Whether a regression line is significant or not is indicated by a Y for Yes and a N for No. Density represents the plant density in mats/m², Dis_Mats stands for the distance between mats, Nr_leaves are the number of leaves, Nr_shoots are the number of shoots, Nr_crops the number of crops and crop cover stands for the proportion of other crops in the field (Table 1)

	T-test	Regression analysis		
Variables	P-value	R-squared	P-value	Significant (Y/N)
Density	0.070	**0.444* (Q)**	**0.000**	Y
Dis_mats	**0.001**	**0.306* (Q)**	**0.001**	N
Nr_leaves	**0.000**	0.149* **(L)**	0.010	N
Nr_shoots	0.832	0.052* **(L)**	0.138	N
Nr_crops	0.736	**0.246 (L)**	**0.000**	N
Crop cover	**0.001**	**0.230 (L)**	**0.000**	N
Tree cover	0.117	0.001 (L)	0.863	N
Green cover	0.081	0.045 (L)	0.143	N
Brown cover	0.887	0.084 (L)	0.044	N
Bare soil	**0.004**	0.203* (Q)	0.007	N

* Outliers removed

Thus also for land management variables interact with each other very much.

From the results it can be concluded that land management has an important contribution to the disease incidence. Crop cover, or better known as intercropping, shows a clear dependency for disease incidence. This variable has been research before in field experiments (Zhang et al., 2013) and this is the first time a field survey confirms these results. The same accounts for the dependency of soil cover (green cover + brown cover) that also showed dependency before in a field experiment. This is important information that can be extrapolated to the less variable large scale banana plantations. The high variability between land management of farms in this area was needed to obtain these results, which show dominant correlations between variables and disease incidence. This proves the importance of land management and thus the importance of changing land management to an optimal formula. The most significant results from the correlation analysis was with plant density, but also crop cover shows a correlation that backs up the results found in previous research. With these results changes can already be implemented, but further research should focus on obtaining more detailed observations both in areas with a high variability in land management and later also in areas with large scale banana plantations and low spatial variability.

Soil variables

The t-tests for the soil variables show that there is a distinction between high and low incidence for pH, Aluminium (Al), Acidity (Acid), potassium (K) and Phosphorus (P). Even though, no dominant relations are present, Acidity and Magnesium show evidence of a relation with disease incidence in the regression analysis.

The t-test for pH shows a significant difference between high incidence, where pH is low (5.10), and low incidence with a higher pH (5.54) (Figure 14). Acidic conditions in pH promote the growth of the Fusarium wilt fungi in the soil and increase the availability of toxic aluminium. However Peng et al., (1999) showed that, regardless of the increasing fungal growth under slightly acidic conditions, disease incidence is higher under alkaline conditions. The ideal pH for banana cultivation and plant health is between 5.5 and 6.0. The mean for low incidence is exactly between these two values (5.54). An optimum pH is wat is expected, since too low or too high pH causes toxic elements to become available and for other essential nutrients (Calcium, Magnesium) the availability decreases. pH is also the most researched soil characteristic for plant growth, plant health and nutrient availability. Therefore much is known about the optimal pH conditions in all agricultural crops, including bananas. pH is easy to adjust in the soil by adding either acidic or alkaline nutrients and is therefore very much controlled in all types of agricultural practices. It is known that some of the visited farms during the quick scan apply fertilizers to

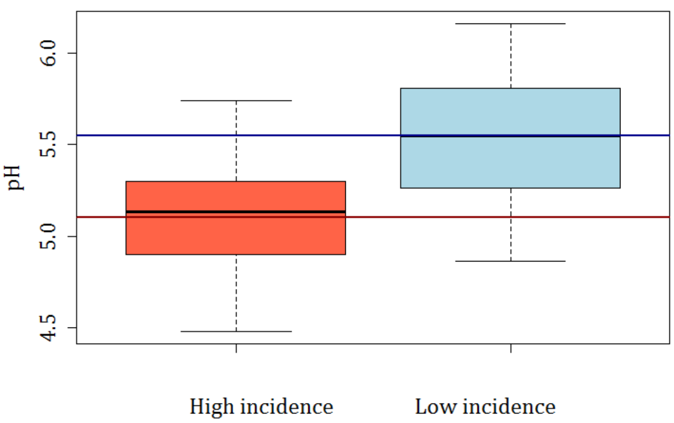


Figure 14
Boxplot for pH for high (red) and low (blue) incidence. The Y-axis shows the pH value. The black lines are the median values, the outsides of the box are the lower and upper quartiles and the red and blue line are the mean values for high and low incidence respectively.

improve soil conditions. pH Is the most important variable for farmers to look at, when applying fertilizer. pH Is also one of the main variables that was detected to be of importance for disease incidence in the previously executed pot- and field experiments by Segura, R (2015a). This research confirms the results from the previous pot- and field experiments.

pH is party correlated to Acidity and Aluminium (0.67). Aluminium and acidity both have a significant difference between values for high and low incidence (p-values = 0.001). Acidity and Aluminium area heavily correlated to each other (0.981) because a low acidity means that aluminium becomes available for plant uptake. This is a negative effect, since Aluminium is highly toxic to plants. The results clearly show this effect in both Aluminium and Acidity. High incidence locations have an average Aluminium concentration of 2.38 cmol(+)/L and an average Acidity of 2.95cmol(+)/L. For low incidence the average for Aluminium is 0.84 cmol(+)/L and for Acidity the average is 1.06 cmol(+)/L. The boxplot for Aluminium is shown in Figure 15. Besides the release of available Aluminium under high Acidity, a high Acidity is also harmful to banana plants because it can cause deficiencies is Calcium and Magnesium, which are essential nutrients for banana growth.

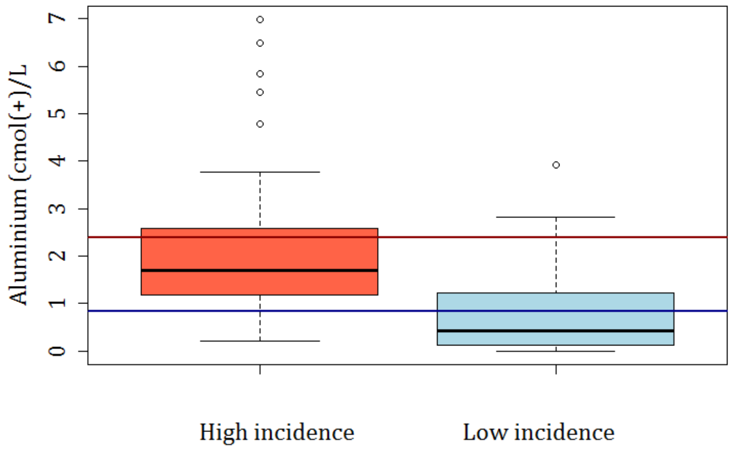


Figure 15
Boxplot for Aluminium for high (red) and low (blue) incidence. The Y-axis shows the Aluminium concentrations (cmol(+)/L). The black lines are the median values, the outsides of the box are the lower and upper quartiles and the red and blue line are the mean values for high and low incidence respectively.

Soil acidity also shows signs of a quadratic relation with disease incidence in the regression analysis ($R^2 = 0.221$; p-value = 0.003). High soil acidity increases the Fusarium wilt incidence in bananas and also promotes the growth of the fungi in the soil (Peng et al., 1999). In Figure 16 the scatterplot is shown. As visible, this relation shows that disease incidence exponentially decreases when acidity decreases.

Potassium and Phosphorous also have a clear distinction between high and low incidence areas. Potassium is the most important nutrient for the growth of banana and the production of bunches (Robinson & Saúco, 2010). Rishbeth (1957) and Huber et al. (2012) showed a connection between Potassium and Fusarium wilt of banana, where Potassium deficiencies are correlated to high incidence. These results show the same effect. High incidence (0.27) has a significant lower average for Potassium concentration than low incidence (0.51). It should be noted that most samples (84%) were below the minimum concentration for optimal banana cultivation (0.60 cmol(+)/L).

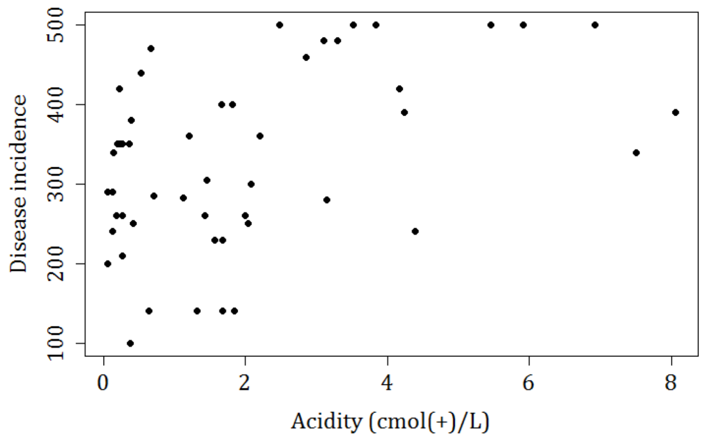


Figure 16
Scatterplot with Acidity concentrations on the x-axis and disease incidence on the Y-axis.

The same problem accounts for Phosphorous, where the optimal range is from 20-40 mg/L and 84% is below 20mg/L. Compared to Potassium or Nitrogen the Phosphorous requirements for bananas is by far not as high. This is because (i) plants accumulate Phosphorous over a long period, (ii) the redistribution within the plant is very effective and (iii) not so much Phosphorous is needed for bunch growth (Robinson & Saúco, 2010). This does not imply that Phosphorous is not an important nutrient for plant health and disease incidence. High incidence areas have a lower average concentration (5.21) than low incidence areas (16.04), however this is also partly caused by a few low incidence samples where Phosphorous concentrations are within the optimum range resulting in a much higher concentration than the other samples, thus having a big influence on the average values.

Magnesium did not show a significant result for the t-test, but does show a quadratic relation with disease incidence (R^2 = 0.309; p-value = 0.000). The highest incidence is located around a concentration of 3 cmol(+)/L and when Magnesium concentrations are either lower or higher the incidence decreases (Figure 17). This is weird and not expected at all. Magnesium deficiency is not good for banana plants (Robinson & Saúco, 2010) and it would be expected that low Magnesium concentrations would induce high incidence. This was also found in the pot- and field experiments executed by Segura, R (personal communication). To find a reason for this peculiar result, some extra analyses were performed. First of all, out of the 50 samples, 30 are below a concentration of 2.50 cmol(+)/L, which is the lower limit for optimal concentrations for banana cultivation.

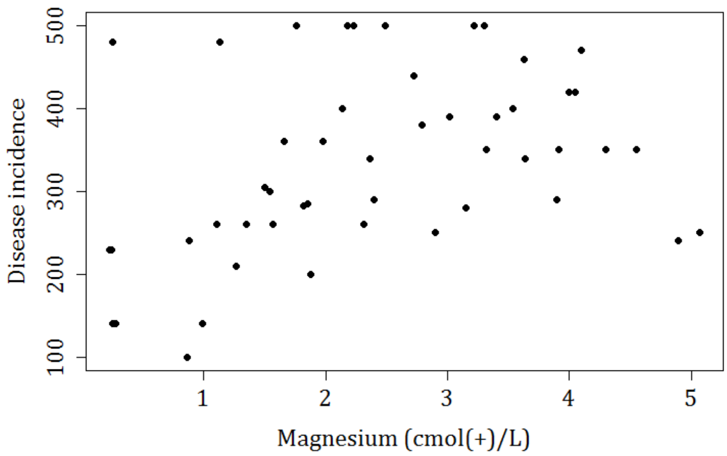


Figure 17
Scatterplot with Magnesium concentration on the x-axis and disease incidence on the y-axis.

In this case the t-test was also performed on the opposite site of the relation. It was tested whether there is a statistical difference in disease incidence for low and high magnesium concentrations. The distinction between high and low magnesium is set at 2.5 cmol(+)/L. As shown above, this value is the lower limit for optimal cultivation conditions. The results of this t-tests shows clearly that there is a difference between low and high magnesium concentrations (p-value = 0.015). In other words, a lower Magnesium concentration is related to a lower average disease incidence and high concentrations are related to higher disease incidence. Eventually the results for Magnesium show that lower Magnesium values are good for disease resistance, regardless of the fact that higher concentrations are recommended for optimal banana cultivation. A possible explanation for this behaviour can be caused by the important interactions between Calcium, Potassium and Magnesium in the soil. These three nutrients are all important for plant health, but they have an optimal balance in concentrations. If one nutrient is out of balance, the other two nutrients will change their availability as well. This unbalance in nutrient availability can cause plant stress and consequently also a higher susceptibility to disease incidence. (Ghorbani et al, 2008; Marschner, 2011).

What is surprising from these results is that Organic Matter (OM) does not appear to influence the disease incidence. This is unexpected since soil OM is an important indicator for soil conditions such as soil health, soil fertility and soil degradation (Segura et al., 2015b). Soil health is connected to microbial activity and nutrient availability, indirectly influencing respectively soil suppression and plant health. The collected soil samples show that the majority (92%) of all samples are above the optimum range of banana cultivation (2.0 - 4.0%.). This can both result in positive or negative conditions. It can be positive if the organic matter is not much higher that 4 % which makes for a fertile soil and possible soil suppression. However, excessive OM accumulation can be caused due to a low mineralization rate, thus implying low microbial activity and high conducive soils. At last, even though Organic Matter and Carbon proportions are considered as different variables, they are 1 to 1 correlated to each other. The results of the statistical analysis for all soil chemical variables are shown in Table 8.

The stepwise regression for soil chemical variables creates a model with an r-squared of 0.4487 and a p-value of 0.000. The variables contributing to this model from low to high significance are Magnesium, Carbon, Aluminium and Acidity. Also in this case a variable (Carbon) has an important contribution to the model, while in the individual regression no relations were found. Especially in the soil chemical variables it was expected that there are a lot of interactions between variables, and that is clearly visible here. This model combines 4 different variables and explains more than 44% of the variance by combining these variables.

From these soil chemical results it can be concluded that chemical soil components have an important contribution to disease incidence. The results found in this research for soil chemistry confirm the results found in previous research executed in pot- and field experiments. The variables that were found to be important all back up previous research and provide essential extra information about the influence of nutrient concentration specifically for Fusarium wilt of banana. The most remarkable findings are those of Magnesium. To find a clear explanation of this behaviour and implementing these results in soil management, extra research is needed on the relation between Magnesium, Fusarium wilt incidence and interactions within the soil. The results for Acidity/Aluminium and pH are not surprising, but nonetheless important. Here it is shown that also in a field survey these variables are important and need to be attended to in soil management. At last the results for Potassium and Phosphorous are proven to be important in a field survey and also these results can be applied in proper soil management practices. All these results can be directly implemented in improving soil management and improving Gros Michel production. Still it is advisable for extra research to continue on the interactions and relations between soil chemical properties and Fusarium wilt incidence. Not only for Acidity/Aluminium, Magnesium, pH, Potassium and Phosphorous but also for Manganese, Organic Matter and Nitrogen since these concentrations showed correlations with Fusarium wilt in previous research, however no similar results are found in this research.

Table 8
Results of the statistical analysis of the data for all soil chemical variables. The t-tests show to what extend the mean values between high and low disease incidence for each variable differ. The regression analysis shows the size of correlation present between disease incidence and each variable. The Q stands for a Quadratic regression and the L stands for a Linear regression. Whether a regression line is significant or not is indicated by a Y for Yes and an N for No. The meaning and units for all variables are described in Table 1

	T-test	Regression analysis		
Variables	P-value	R-squared	P-value	Significant (Y/N)
OM	0.486	0.174*(Q)	0.017	N
pH	**0.000**	0.036 (L)	0.185	N
Al	**0.001**	0.150 (Q)	0.022	N
Acid	**0.001**	**0.221** (Q)	**0.003**	N
Ca	0.225	0.118 (Q)	0.052	N
Mg	0.094	**0.309** (Q)	**0.000**	N
K	**0.002**	0.029*(Q)	0.518	N
P	**0.009**	0.023*(Q)	0.592	N
Fe	0.259	0.078*(Q)	0.173	N
Cu	0.355	0.009*(Q)	0.820	N
Zn	0.050	0.173*(Q)	0.015	N
Mn	0.109	0.173*(Q)	0.015	N
B	0.288	0.010*(Q)	0.794	N
C	0.485	0.174*(Q)	0.017	N
N	0.578	0.121 (Q)	0.048	N

***** Outliers removed

Microbial populations

For the microbial populations only the amount of fungi shows a relationship with disease incidence, even though it is not statistically convincing. This relation is a quadratic relation with a high disease incidence when fungal concentrations are either too high or too low. These fungi should not be confused with the fungus that is Fusarium wilt. These fungi have a different function and the reason why they induce soil suppression is because the growth of other fungi causes competition, antibiosis and parasitism for the fungus Fusarium wilt. In its turn, this results in lower Fusarium wilt concentrations and thus less incidence (Hornby, 1983). In this quadratic relation the disease incidence starts to decrease when fungi concentrations are increasing. Then an optimum is reached and after the optimum disease incidence increases with increasing fungi concentrations. However, fewer samples are on this site of the relation and is therefore less trustworthy. increase of disease incidence with increasing fungi amounts can be explained by the fact that, because a lot of different fungi are present, relatively also more Fusarium wilt fungus is present. However, it can also have another unknown reason or it might just be an inaccuracy.

What is unfortunately is that the other microbial variables do not show any correlation or difference between high and low incidence areas. These three were expected to have a relation with disease incidence due to the abundance of literature publications supporting these ideas (Scher & Baker, 1980; Hornby, 1983)

Also the stepwise regression does not show any results at all, implicating that variability in microbial populations cannot explain the observed variability in disease incidence for bananas. A possible and likely explanation for this is that the soil samples were not correctly treated during the collection for the measurements of these microbial populations. Since the study area was relatively far from the laboratories where the measurements were executed the soil samples were kept cool in an ice box for a maximum of four days. The icebox was however not fully functional and sitting in the back of the car, exposed to a lot of sunlight. The time it took for the samples to get to the lab and the treatment before, could have had an influence on the microbial populations in the soil. If this is the case, the measurements for microbial populations are inaccurate. This would explain why no real correlations are found, while it would be expected.

Overall, the influence of microbial populations on the supressiveness does not appear to be important with the results presented above, however, literature has shown enough evidence that microbial populations are very important for supressive soils. Therefore further research is recommended, with the advice to bring the soil samples to the lab on the same day as collected and to keep them in a cool and dark box between collection and delivery at the lab.

Table 9

Results of the statistical analysis of the data for all microbial populations. The t-tests show to what extend the mean values between high and low disease incidence for each variable differ. The regression analysis shows the size of correlation present between disease incidence and each variable. The Q stands for a Quadratic regression and the L stands for a Linear regression. Whether a regression line is significant or not is indicated by a Y for Yes and an N for No. The meaning and units for all variables are described in Table 1

	T-test	Regression analysis		
Variables	P-value	R-squared	P-value	Significant (Y/N)
Fungi	0.141	**0.238*(Q)**	**0.003**	N
Bacteria's	0.803	0.004*(Q)	0.924	N
Actinomi-cetes	0.087	0.076*(Q)	0.177	N

* Outliers removed

Interaction between all variables

To cover the interactions between the main sectors, first all Pearson correlation coefficients were calculated. Not many variables from different sectors were correlated with each other and again for the microbial populations not many correlations with other variables are present. This is remarkable, since it is always expected that microbial populations are very much influenced by many landscape, climate, land management and soil chemistry variables. The only correlations present, are correlations between the actinomicetes population and crop cover (-0.37), temperature (0.53) and Zinc (0.58). However these values are also relatively low for a Pearson correlation coefficient.

As a last action another stepwise regression was executed with variables from all sectors. To prevent overestimations not all variables are put together in one model. All the variables that were part of a stepwise regression model in a previous section were the input for this regression. The main focus of this last regression, was not necessarily to create the best fit, but rather to see which variables are the most dominant in the creation of a model. The model created has an r-squared of 0.647 and a p-value of 0.000. The variables part of this model from low to high significance are Carbon, Acidity, Magnesium, slope, plan curvature, crop cover, number of leaves and finally the distance between plants. What is visible, is that the three most important variables for this model fit are all three land management variables. It is also the case that landscape, land management and soil chemistry are all three included in this best model fit. This shows that all these sectors are important and have their own contribution in explaining variability in disease incidence. It should be noted that this does not mean that climate and microbial populations are not important. The main reason why their influence is smaller can simply be explained by the fact that fewer variables for these two sectors are included in this research.

Overall discussion of the results

The results show important information for changing soil management practices, but it also shows how complex the system is. It shows how important new research in this subject is, to improve soil management practices and reduce the damage Fusarium wilt can cause not only to the Gros Michel, but also to the Cavendish cultivar. This research gives a good overview of many variables that are thought to be important and some variables clearly show their importance. Variables such as curvature, temperature, plant density, crop cover, bare soil, pH, Acidity, Magnesium, Potassium, Phosphorous and fungi all appear to be important. Most of these variables (plant density, crop cover, bare soil, pH, Acidity, Magnesium, Potassium and fungi) showed their importance before in previous research, however never where these effects measured in a field survey. This research confirmed the results collected in previous pot- and field experiments on a farm scale level.

In the introduction three research questions where mentioned. The variability assessment, executed via a stratification map, clearly showed how high the spatial variability in the study area is. Especially altitude, temperature, slope and aspect were important variables for explaining the spatial variability, but also curvature and NDVI values are important to explain the landscape variability in the area, especially when assessing the variability for the purposes of banana cultivation. This high spatial variability can partly explain the variability in disease incidence as well. However the results from the quick scan show that landscape is not the only important source of variability important for disease incidence.

The quick scan was the main source of information and data for the research following the variability assessment. During the quick scan it was clear that not only landscape but also land management was very variable between farms. Despite the high variability it was also clear that production of Gros Michel is still very well possible, even though disease incidence is very variable. Both the variability in land management and variability in disease incidence was expected. The variability in disease incidence was important for a good execution of the quick scan, but the variability in land management between farms both is a challenge and a gift when executing a field survey. When doing a pot experiment or a field experiment it is relatively easy to keep all the factors constant except for those variables that are being researched. In a field survey this is not possible, making it extra challenging to find good results that are not indirectly connected to other variables. However, a high variability in land management did enable this research to look for those variables that were important for disease incidence in operating farms.

To capture this variability while also looking for dependencies between variables and disease incidence, the choice was made to cover many different variables and perform the same statistical analysis for each variable. The results support the results found in previous research during pot- and field experiments and provide new information on both expected and unexpected variable dependency. The variables that shows to be important for disease control in pot- and field experiments and that are confirmed in this research via a field survey are plant density, intercropping, green cover, pH, Acidity, Magnesium, Potassium and fungi populations. for the other variables, curvature, temperature and Phosphorous, importance is also shown and further research in this topic will find out whether these variables are also important in other banana producing regions.

These results are a confirmation of long suspected dependencies in a field survey. This can contribute to improving soil management and creating coexistence between Fusarium wilt and banana cultivation. This contributes to productive banana cultivation due to measures and practices that suppress the disease in either the soil or plant.

Implementation of results

With the help of these results land management and soil management practices can be adjusted to decrease disease incidence and increase the banana productivity.

The first results were the dependency of curvature and temperature on the disease incidence. These two variables are highly correlated to the landscape variability in the area. These results, together with the stratification map, can help the local farmers with location planning. It appears that concave slopes are less susceptible to disease incidence. Therefore, within the study area and only for Race 1 susceptible bananas, these results are a first step to optimal location planning for planting young bananas.

For land management it became clear that an optimum density is very important. In the future it is recommended to plant the young plants neither too far nor too close to each other. To obtain an optimum density value further research is needed, so that in the future this can be implemented in soil management.

These results also showed the negative effects bare soil has on disease incidence. Therefore it would be recommended to make sure that there is enough green cover between the plants and that dead leaves and dead organic material should be left on the soil to decompose. For the green cover it is important to make sure that no harmful weeds are introduced, but rather small plants and grasses that can do no harm.

Another option for improving disease incidence is a long time recommended way of cultivation; intercropping. The relations between disease incidence and the presence of a secondary crop (number of crops and crop cover) showed once again that intercropping does positively affect the suppression of Fusarium wilt. In the visited area most intercropping systems were with coffee, but also intercropping systems with other banana plants, plantains, citrus trees and palm trees are observed.

Another element of soil management which is very important for lowering the disease incidence is the use of fertilizers and creating an optimal nutrient balance in the soil. As Schneider & Ploetz (1990) mentioned, before application of fertilizers to control disease incidence, it is crucial to know what the desired ionic ratios are and to be aware of the current concentrations of nutrients, before adding fertilizer and risking ion toxicity. The addition of nutrients to control disease incidence is a delicate job. There are many interactions between the chemical components of the soil, and nutrient concentrations in the soil are not identical to the nutrient concentrations available for the host plant. Still the results from this research and the results from previous results all show that pH, Acidity, Magnesium and Potassium are all correlated to disease severity. The repetition in these results are a good sign and research in soil management can also start focussing on finding the optimal balance between these nutrients to suppress Fusarium wilt in both the soil and the plant.

The results for land management and soil conditions have an important contribution to improving soil management practices within the study area, but also outside the study area. The results are researched in Race 1 infected areas, but Race 4 is very similar to Race 1, and with some additional research it might be possible to implement the same measures for disease control in Race 1 as for Race 4. Segura, R (Personal communication, 2016) already proved during his pot-experiments that the disease reacts in similar ways to certain measures and tests for both Race 1 and Race 4. Therefore, after additional research, possible results that are found for disease incidence in Gros Michel are likely to be applicable for Race 4 infested areas as well. This means that these results are not only applicable in areas where Gros Michel are cultivated, but also in areas such as Asia where Race 4 already threatens the Cavendish cultivation. It can even be applicable in Latin America, for an optimal preparation for when Race 4 starts infecting these areas as well.

With these results small farms in the study area can be supported with the banana production. The landscape results can help with future planning of location for young bananas, but the results can also help the small farms to optimize their soil conditions. For the small farms within the study area a full chemical analysis of the soil is done. This, in combination of the results from this research, can be used to provide the farmers with detailed and farm adjusted advice on land management and nutrient application in the soil. For other small growing farms outside the study area the results from this research can also help in improving the soil conditions after a thorough analysis of the local soil conditions. This way small business-oriented soil management can be provided.

For the large scale plantations also the soil conditions need to be mapped before applying soil management. However land management practices also need to be taken into account and for large scale plantations it is advisable to do some additional research on the interactions between environmental variables.

For optimal adaptation of soil management practices and optimal disease control, further research is still needed. However, this research can now focus on the optimal conditions for disease control in a limited number of variables that showed to be dependent for disease incidence. This dependency was first shown in a pot experiment, later in a field experiment, and now also in a field survey.

Overall the land management on the farm and changing variables such as density and intercropping take time to implement, however for now, these options are the ones that appear to be most promising for the relative short term changes. Also the addition of extra nutrients can be implemented on a short term, however there is still a lot of uncertainty on the interactions within a soil. This research shows which variables are important, and which are not. This confirms those relations previously found in the pot- and field experiments. Therefore this research is the final step needed to change soil management practices in order to lower the disease incidence and obtain a higher productivity. As long as there is no resistant cultivar for Race 4 soil management provides a promising and sustainable alternative for disease control.

Methodological Discussion

In this chapter the methods used will be briefly discussed. The variability assessment, quick scan and statistical analysis will all be included. For the variability assessment the number of clusters will be discussed. For the quick scan the convenience sampling design, the method for observing disease incidence and missing variables will be discussed. This chapter will end with a brief discussion on the low number of data points and the lack of a normal distribution influencing the results for the statistical analysis.

The main point of discussion in the variability assessment is the number of clusters chosen for the cluster analysis. During the cluster analysis 12 different layers were used to create the cluster map. 25 Clusters out of 12 layers is relatively low and statistically it would have been better if there were at least 25 more clusters. However, the layers inserted in the cluster analysis were the 12 PCA layers. Here the first layers were of a higher importance than the last few layers. Therefore the 25 clusters created mainly covered those variables that were represented by the most important PCA layers, thus the most important variability is covered within the selected 25 clusters. A solution for the next time would be to only insert those numbers of layers that together explain around 80% of the variance. This reduces the number of layers for the input and thus also the number of clusters needed to capture all the variance.

Some elements of the execution of the quick scan also need to be discussed. First, in this research a convenience sampling design is used. In this research convenience sampling was the best option, however it does have some disadvantages that are clearly visible in the distribution of the data points. Convenience sampling often causes a clustered sampling pattern with a high variability in the distance between locations. This is also the case in this research and the clustering of locations can cause a bias in the data. In this research it is possible that the convenience sampling design caused a bias in the data, but this bias is not directly visible in the data, neither is it expressed in the results. In following research, not only the small growers, but also large scale plantations infecting Race 4 in Asia could be useful for research. In these areas, convenience sampling is not necessary, and therefore it is possible to use a grid or stratified random sampling pattern.

Another discussion point was already mentions in the previous chapter (Results & Discussion, Quick scan). The disease incidence in the plants was sometimes hard to measure due to differences in age between the plants and the presence of similar symptoms with another source than Fusarium wilt. Sometimes this complicated the estimation of disease incidence, but the fact that the disease incidence was measured as a continuous variable is a bigger advantage than a disadvantage. In this research the disease, as expressed in the plant, is used as a quantitative measure and no proxies such as Fusarium wilt concentrations or yield losses are used. For the regression analysis quantitative data is used (on a scale from 100-500) and for the t-tests a qualitative scale is used (either high or low incidence).

The quick scan was a method to observe a lot of data in a low number of days. This was very effective and enables this research to have a short period with field work, while still collecting enough data for good results. However, some variables were not taken into account, even though they are expected to be important for disease incidence as well. Physical soil variables, such as soil type, texture and clay content are not assesses in this research, because the collection of these data takes too much time. Unfortunately, physical soil characteristics are important for soil suppression and consequently also for disease incidence (Baker & Cook, 1974). In this area there are two main soil types (andosols, nitisols) but within these soil types a lot of spatial variability is expected, because of the size of the area. For determining the physical soil characteristics it is needed to auger in the soil and time is needed to make secure observations. The choice was made not to do this, because it is time consuming. Especially clay content is expected to be of importance (Stotzky & Martin, 1963; Dominguez et al., 2001). Overall, the lack of physical soil characteristics is an important limitation in this research, especially for location planning of young plants within the area.

Another variable not included in the quick scan, or rather soil samples, is the salinity of the soil. The most important contributor to soil salinity is Na^+ (Sodium), and this ion is not measured for the analysis. Other salty minerals are K^+, Ca^{2+}, Mg^{2+} and Cl^-. A high salinity causes the soil texture to become harder, limiting rooting depth for plants, microbial populations and water infiltration. All these effects do negatively influence the disease incidence. Salinity has shown to be of importance for Fusarium wilt in several occasions (Dominguez et al., 2003;), however in the study area the main soils are andosols and nitisols, which usually do not contain high salt concentrations. Therefore salinity might be a factor influencing disease incidence, but the expected concentrations for the study area in this research are not high enough for salinity to dominate the disease incidence.

The last soil chemical factor that is not included in this research but has also shown some signs of importance are the Ammonium (NH_4) and Nitrate (NO_3) concentrations in the soil (Schneider and Ploetz, 1990). It appears that NH_4 causes an increase in disease incidence, while NO_3 causes a decrease. In this research we did measure the Nitrogen concentrations and Ammonium and Nitrate fertilization are usually the main sources for Nitrogen. Therefore measuring the Nitrogen concentrations was satisfactory for this research.

This research might not have taken salinity, Ammonium and Nitrate into account, but it did take all the main macro and micro nutrients into account. This research focussed on confirming dependencies between environmental variables and disease incidence to improve soil management practices. The most important soil management variables that can be optimized are all the macro and micro nutrients. When applying fertilizer the balances between these macro and micro nutrients are the most important.

At last there is also the discussion about the statistical properties of the data. First of all, this research was limited in time and resources which resulted in 50 data points. This is not directly a problem, however a low number of data points does influence the R-squared and p-values. The R-squared values are lower and the p-values are higher, making it hard to get statistical significance. A second remark on the data is that the assumption is made that all the variables are normally distributed. This is however not entirely the case, thus influencing the results we have (as was clearly visible in Figure 11). A normal distribution is needed for the PCA, regression analysis and the t-tests. It is possible to execute statistical analysis that does not need a normal distribution, however the results and its effectiveness are often uncertain. Here it is decided to assume a normal distribution, partly because of the low number of data points. Even though theory in statistics has requirements like a normal distribution, in most research fields there is a complex reality (clustered sample locations, interactions between variables, high variability) where these requirements are seldom fulfilled.

Regardless of these shortcomings, this research had a goal before starting the research and this goal is achieved. The quick scan proved to be an effective method to collect as much data as needed in a short period of time. In 5 days 50 data points were collected and with the help of existing data and online resources this research got good results and was able to answer the research questions formulated in the beginning. These results are an important contribution to research in fusarium wilt of bananas and on how disease control can be improved.

Conclusion

This research aimed on proving the importance of the environment on the incidence of Panama disease. In previous research it was shown that disease incidence is dependent on land management and soil conditions. These results were found after the execution of pot- and field experiments, however data showing similar results in operating farms was not present. This research confirms the results previously found by executing a field survey.

Since this research focussed on a field survey, spatial variability in landscape and climate also needed to be accounted for. Therefore, this research first showed the high variability in landscape and climate present in the study area via a GIS analysis. During the fieldwork it also became clearly visible that disease incidence is highly variable. This provided opportunities for research, since this variability in disease incidence needs to be explained by something.

After establishing the variability in landscape and disease incidence, the environmental variables show results about which variables can explain this variability in disease incidence. For landscape and climate these variables turned out to be profile curvature and temperature. For land management variables such as plant density, intercropping and the presence of bare soil explain part of the variability in disease incidence. At last also for soil conditions variables could explain part of the variability; pH, Acidity, Magnesium, Potassium, Phosphorous and fungi populations.

The results for landscape and climate provide new information on disease incidence. It suggests that landscape and climate should be taken into account when planting bananas in a new location. This can be used for location planning in areas were landscape variability is high.

Most results for land management and soil conditions are a confirmation of the results found before in pot- and field experiments. These results are confirmed in this research via a field survey. This shows that the importance of land management and soil conditions is also applicable in operating banana farms. Therefore land management and soil conditions can be regarded as important factors in disease incidence and thus are important for disease control.

To use land management and soil conditions for disease control it is recommended to do further research both in pot- and field experiments and in field surveys. This research did not focus on the many interactions happening between disease incidence, land management and soil conditions. Therefore research in these interactions is recommended to optimize the use of land management and soil conditions for disease control.

Using land management and soil conditions for disease control is also known as soil management. When using soil management as a way of disease control, land management practices and soil conditions are sustainably manipulated to create optimal conditions for plant health and soil suppression. This research proved the importance of land management and soil conditions in a field survey, after this was already shown in pot- and field experiments. Therefore, soil management should be considered as a relevant and essential method for disease control.

This researched showed that disease incidence of Fusarium wilt is partly explained by the environment. This provides options for both planning and soil management. The next step is to implement the results by improving soil management practices, so that the damage of Panama disease can be tackled and reduced.

Acknowledgements

The collection of the field work data and the processing of the soil samples in the lab was carried out in collaboration with CORBANA, Corporación Bananera National, Guápiles, Costa Rica. The locations of the field work and related farms were provided by Bioversity International, Turrialba, Costa Rica. Big thanks is for Rafael Segura Mena, who supervised the research in Costa Rica, was a great support for knowledge during the field work and was the host during my stay in Costa Rica. I would also like to say thanks to Jorge Flores and Cindy Castillo who helped me finding the farms and provided expert knowledge on Panama disease during the fieldwork. Finally I would like to say thank to Jetse Stoorvogel on his advice and supervision during the research and critical look on the report and its content.

References

Alabouvette, C. (1999). Fusarium wilt suppressive soils: an example of disease-suppressive soils. *Australasian Plant Pathology*, 28(1), 57-64.

Amir, H., & Alabouvette, C. (1993). Involvement of soil abiotic factors in the mechanisms of soil suppressiveness to Fusarium wilts. *Soil Biology and Biochemistry*, 25(2), 157-164.

Baker, K. F., & Cook, R. J. (1974). *Biological control of plant pathogens*. WH Freeman and Company.

Beckman, C. H., & Ploetz, R. C. (1990). Host responses to the pathogen. *In Fusarium wilt of banana.* (pp. 93-105). APS Press.

Bhandari, A. K., Kumar, A., & Singh, G. K. (2012). Feature extraction using Normalized Difference Vegetation Index (NDVI): a case study of Jabalpur city. *Procedia Technology*, 6, 612-621.

Bouwmeester, H., Heuvelink, G. B. M., & Stoorvogel, J. J. (2016). Mapping crop diseases using survey data: The case of bacterial wilt in bananas in the East African highlands. *European Journal of Agronomy*, 74, 173-184.

Brake, V. M., Pegg, K. G., Irwin, J. A. G., & Chaseling, J. (1995). The influence of temperature, inoculum level and race of Fusarium oxysporum f. sp. cubense on the disease reaction of banana cv. Cavendish. *Crop and Pasture Science*, 46(3), 673-685.

Busacca, A. J., Cook, C. A., & Mulla, D. J. (1993). Comparing landscape-scale estimation of soil erosion in the Palouse using Cs-137 and RUSLE. *Journal of Soil and Water Conservation*, 48(4), 361-367.

Burdon, J. J., & Chilvers, G. A. (1982). Host density as a factor in plant disease ecology. *Annual review of phytopathology*, 20(1), 143-166.

Calberto, G., C. Staver and P. Siles. 2015. An assessment of global banana production and suitability under climate change scenarios, In: Climate change and food systems: global assessments and implications for food security and trade, Aziz Elbehri (editor). *Food Agriculture Organization of the United Nations (FAO),* Rome, 2015.

Cook, R. J., & Papendick, R. I. (1972). Influence of water potential of soils and plants on root disease. *Annual Review of Phytopathology*, 10(1), 349-374.

Daniels, R. B., & Hammer, R. D. (1992). *Soil geomorphology.* John Wiley & Sons.

Dijkshoorn, K., Huting, J., & Tempel, P. (2005). Update of the 1: 5 million Soil and Terrain Database for Latin America and the Caribbean (*SOTERLAC*).

Domınguez, J., Negrín, M. A., & Rodrıguez, C. M. (2001). Aggregate water-stability, particle-size and soil solution properties in conducive and suppressive soils to Fusarium wilt of banana from Canary Islands (Spain). *Soil Biology and Biochemistry*, 33(4), 449-455.

Domınguez, J., Negrın, M. A., & Rodrıguez, C. M. (2003). Evaluating soil sodium indices in soils of volcanic nature conducive or suppressive to Fusarium wilt of banana. *Soil Biology and Biochemistry*, 35(4), 565-575.

Fairbanks, D. H. (2000). Physio-climatic classification of South Africa's woodland biome. *Plant Ecology*, 149(1), 71-89.

FAOSTAT. (n.d.). Retrieved March 15, 2016, from http://faostat3.fao.org/browse/Q/QC/E.

Ghag, S. B., Shekhawat, U. K., & Ganapathi, T. R. (2015). Fusarium wilt of banana: biology, epidemiology and management. *International Journal of Pest Management*, 61(3), 250-263.

Ghorbani, R., Wilcockson, S., Koocheki, A., & Leifert, C. (2008). Soil management for sustainable crop disease control: a review. *Environmental Chemistry Letters*, 6(3), 149-162.

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis, 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965-1978.

Hornby, D. (1983). Suppressive soils. *Annual review of phytopathology*, 21(1), 65-85.

Huber, D., Römheld, V., & Weinmann, M. (2012). Relationship between nutrition, plant diseases and pests. *Marschners Mineral Nutrition of Higher Plants, third edition. Marschner*, P.(ed.), 283-298.

Kass, D. C. L., Jiménez, M. L., Kauffman, J. H., & Herrera Reyes, C. G. (1995). Costa Rica: Reference soils of the Turrialba valley and slopes of the Irazú volcano. Soil brief Costa Rica 2 (No. CATIE K19c). CATIE, Turrialba (Costa Rica). *International Soil Reference and Information Centre, Wageningen* (Países Bajos).

Knight, E. J., & Kvaran, G. (2014). LANDSAT–8 operational land imager design, characterization and performance. Remote Sensing, 6, 10286–10305.

Leal Filho, W., Adamson, K., Dunk, R. M., Azeiteiro, U. M., Illingworth, S., & Alves, F. (Eds.). (2016). Implementing Climate Change Adaptation in Cities and Communities: *Integrating Strategies and Educational Approaches*. Springer.

Li, A., Wang, A., Liang, S., & Zhou, W. (2006). Eco-environmental vulnerability evaluation in mountainous region using remote sensing and GIS—a case study in the upper reaches of Minjiang River, China. *Ecological Modelling*, 192(1), 175-187.

Mansfield, J. R., Sowa, M. G., Majzels, C., Collins, C., Cloutis, E., & Mantsch, H. H. (1999). Near infrared spectroscopic reflectance imaging: supervised vs. unsupervised analysis using an art conservation application. *Vibrational Spectroscopy*, 19(1), 33-45.

Marquardt, S. (2001). " Green Havoc": Panama Disease, Environmental Change, and Labor Process in the Central American Banana Industry. *The American historical review*, 106(1), 49-80.

Marschner, H. (2011). *Marschner's mineral nutrition of higher plants.* Academic press.

Mehlich, A. (1984). Mehlich 3 soil test extractant: A modification of Mehlich 2 extractant. *Communications in Soil Science & Plant Analysis,* 15(12), 1409-1416.

Metzger, M. J., Bunce, R. G. H., Jongman, R. H. G., Mücher, C. A., & Watkins, J. W. (2005). A climatic stratification of the environment of Europe. *Global ecology and biogeography*, 14(6), 549-563.

NASA JPL. (2009). ASTER Global Digital Elevation Model. NASA JPL. Retrieved from https://doi.org/10.5067/ASTER/ASTGTM.002

Navajothy, F. (2011). Spectroscopical differences among soil samples of selected susceptible and resistant varieties of banana in relation to wilt disease. *International Multidisciplinary Research Journal*, 1(3).

Ordonez, N., Seidl, M. F., Waalwijk, C., Drenth, A., Kilian, A., Thomma, B. P., Ploetz, R. C & Kema, G. H. (2015). Worse comes to worst: bananas and Panama disease—when plant and pathogen clones meet. *PLoS Pathog*, 11(11), e1005197.

Pattison, A. B., Wright, C. L., Kukulies, T. L., & Molina, A. B. (2014). Ground cover management alters development of Fusarium wilt symptoms in Ducasse bananas. *Australasian Plant Pathology*, 43(4), 465-476.

Palti, J. (1981). *Cultural practices and infectious crop diseases*. Berling: Springel-verslag

Pegg, K. G., & Langdon, P. W. (1986, October). Fusarium wilt (Panama disease): a review. *In Banana and Plantain Breeding Strategies: Proceedings of an International Workshop, Cairns, Australia* (pp. 119-123).

Peng, H. X., Sivasithamparam, K., & Turner, D. W. (1999). Chlamydospore germination and Fusarium wilt of banana plantlets in suppressive and conducive soils are affected by physical and chemical factors. *Soil Biology and Biochemistry*, 31(10), 1363-1374.

Pérez-Vicente, L. (2004). Fusarium wilt (Panama disease) of banana: an updating review of the current knowledge on the disease and its causal agent. *Orozco-Santos, M; Orozco-Romero; J; Velázquez-Monreal, J,* 1-16.

Ploetz, R. C. (2006). Panama disease, an old nemesis rears its ugly head: part 2, the cavendish era and beyond. *Plant Health Progress*, 1-17.

Ploetz, R. C. (2015a). Management of Fusarium wilt of banana: A review with special reference to tropical race 4. *Crop Protection*, 73, 7-15.

Ploetz, R. C. (2015b). Fusarium wilt of banana. *Phytopathology*, 105(12), 1512-1521.

Ploetz, R., Freeman, S., Konkol, J., Al-Abed, A., Naser, Z., Shalan, K., ... & Israeli, Y. (2015). Tropical race 4 of Panama disease in the Middle East. *Phytoparasitica*, *43*(3), 283-293.

Rishbeth, J. (1957). Fusarium Wilt of Bananas in Jamaica II. Some Apsects of Host-parasite Relationships. *Annals of Botany*, 21(2), 215-245.

Robinson, J. C., & Saúco, V. G. (2010). *Bananas and plantains* (Vol. 19). Cabi.

Scher, F. M., & Baker, R. (1980). Mechanism of biological control in a Fusarium-suppressive soil. *Phytopathology*, 70(5), 412-417.

Schneider, R. W., & Ploetz, R. C. (1990). Influence of mineral nutrition on Fusarium wilt: a proposed mechanism involving cell water relations. *In Fusarium wilt of banana*. (pp. 83-91). APS Press.

Segura, R., Stoorvogel, J. J., Garcia-Bastidas, F., Salacinas-Niez, M., Sandoval, J. A., & Kema, G. (2015a). Soil management as an effective strategy for crop disease management: the case of panama disease in banana. *In Book of Abstract of the Wageningen Soil Conference: Soil Science in a Changing World* (pp. 48-48).

Segura, R. A., Serrano, E., Pocasangre, L., Acuña, O., Bertsch, F., Stoorvogel, J. J., & Sandoval, J. A. (2015b). Chemical and microbiological interactions between soils and roots in commercial banana plantations (Musa AAA, cv. Cavendish). *Scientia Horticulturae*, 197, 66-71.

Senay, G. B., & Elliott, R. L. (2000). Combining AVHRR-NDVI and landuse data to describe temporal and spatial dynamics of vegetation. *Forest Ecology and management*, 128(1), 83-91.

Shen, Z., Ruan, Y., Wang, B., Zhong, S., Su, L., Li, R., & Shen, Q. (2015a). Effect of biofertilizer for suppressing Fusarium wilt disease of banana as well as enhancing microbial and chemical properties of soil under greenhouse trial. *Applied Soil Ecology*, 93, 111-119.

Shen, Z., Ruan, Y., Xue, C., Zhong, S., Li, R., & Shen, Q. (2015b). Soils naturally suppressive to banana Fusarium wilt disease harbor unique bacterial communities. *Plant and Soil*, 393(1-2), 21-33.

Smith, L. I. (2002). *A tutorial on principal components analysis.* Cornell University, USA, 51, 52.

Song, Y. N., Zhang, F. S., Marschner, P., Fan, F. L., Gao, H. M., Bao, X. G., ... & Li, L. (2007). Effect of intercropping on crop yield and chemical and microbiological properties in rhizosphere of wheat (Triticum aestivum L.), maize (Zea mays L.), and faba bean (Vicia faba L.). *Biology and Fertility of Soils*, 43(5), 565-574.

Stotzky, G., & Martin, R. T. (1963). Soil mineralogy in relation to the spread of Fusarium wilt of banana in Central America. *Plant and Soil*, 18(3), 317-337.

Stover, R. H. (1962). Fusarial wilt (Panama Disease) of bananas and other Musa species. *Fusarial wilt (Panama disease) of bananas and other Musa species.*

Sudarma, I. M., & Suprapta, D. N. (2011). Diversity of soil microorganisms in banana habitats with and without Fusarium wilt symptom. J. *ISSAAS* (17), 1, 147-159.

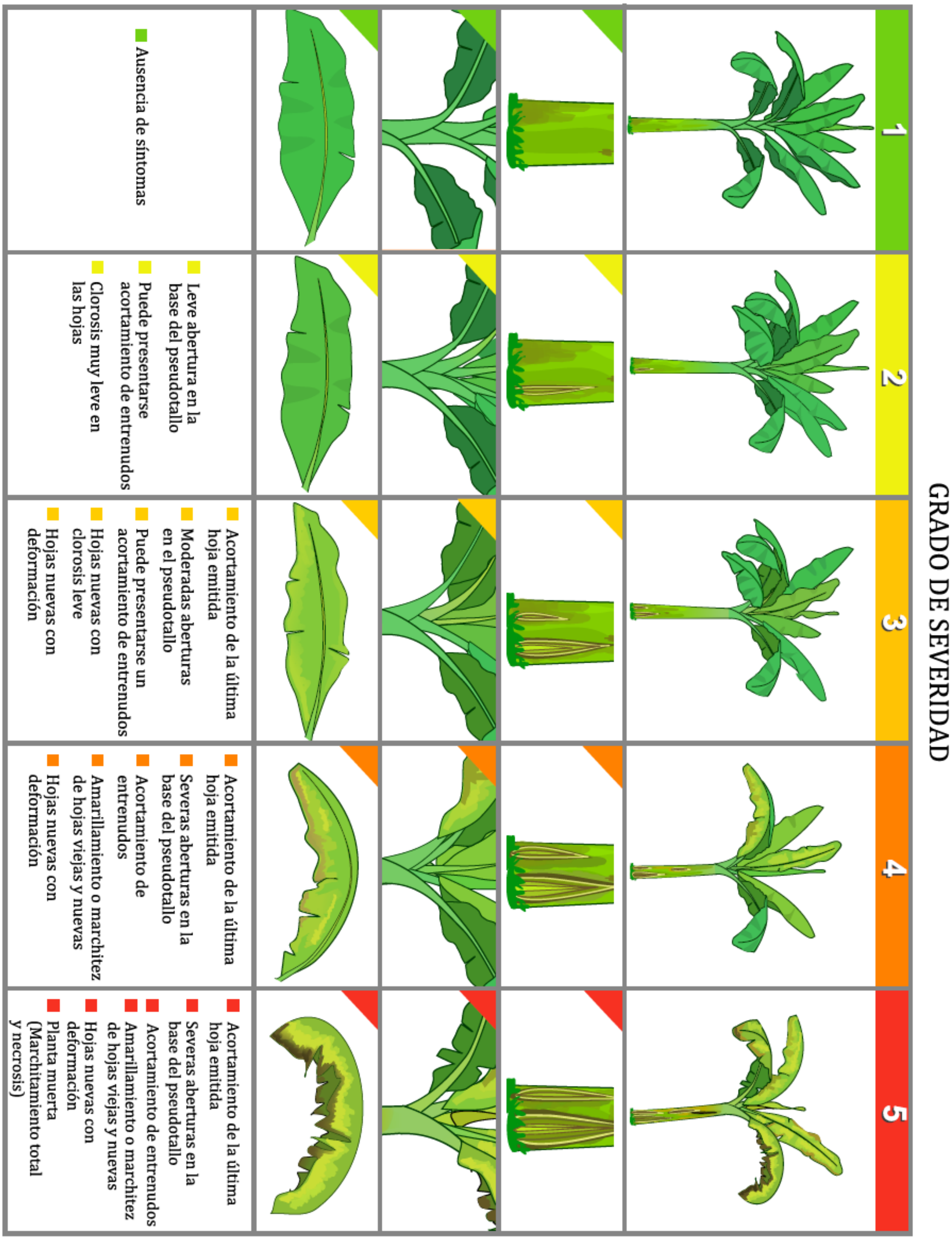
Tou, J. T., & Gonzalez, R. C. (1974). *Pattern recognition principles*.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3), 37-52.

Zhang, H., Mallik, A., & Zeng, R. S. (2013). Control of Panama disease of banana by rotating and intercropping with Chinese chive (Allium tuberosum Rottler): role of plant volatiles. *Journal of chemical ecology*, *39*(2), 243-252.

Appendix

Appendix 1: Grading system used for the observation of Fusarium wilt incidence in banana plants (source: CORBANA)



Appendix 2: Field form used for the Quick scan

Field form Quick scan MSc thesis Marianne Bosman

Date:	**Time:**
Location Number:	**Altitude:**
Longitude:	**Latitude:**

Mat characteristics

Mat density:	Nr of mats m^2	(m between plants)
Number of leaves per plant:		
Number of shoots per mat:		

Plant surroundings

Number of other crops:		**Name crops:**	
Proportion other crops:	(%)		
Name crop 1		**Proportion:**	(%)
Name crop 2		**Proportion:**	(%)
Proportion of tree cover:	(%)		
Green cover	(%)	**100%**	
Brown cover	(%)		
Bare soil	(%)		

Disease symptoms

Code	**% of plants**	**Remarks**
0		
1		
2		
3		
4		

Remainder information

Temperature:	(°C)
Soil moisture:	(m^3/m^3)

Remarks:

Appendix 3: Summary statistics for all principal component variables before and after standardization

		Original	Standardized			Original	Standardized
Altitude (m)				**Distance to River (m)**			
	Min	157.00	-2.03		Min	0.00	-1.19
	1st Qu	690.00	-0.68		1st Qu	107.00	-0.79
	Median	901.00	-0.15		Median	256.20	-0.22
	Mean	961.00	0.00		Mean	317.16	0.00
	3rd Qu	1150.00	0.48		3rd Qu	473.20	0.59
	max	2900.00	4.88		max	1651.24	5.01
	SD	397.00	1.00		SD	266.24	1.00
Slope (%)				**NDVI (-)**			
	Min	0.00	-1.64		Min	-0.09	-6.63
	1st Qu	8.51	-0.77		1st Qu	0.41	-0.35
	Median	14.76	-0.13		Median	0.46	0.20
	Mean	15.94	0.00		Mean	0.44	0.00
	3rd Qu	22.51	0.66		3rd Qu	0.49	0.65
	max	66.26	5.15		max	0.66	2.64
	SD	9.75	1.00		SD	0.08	1.00
Aspect (Radians)				**Annual precipitation (mm)**			
	Min	-1.00	-1.44		Min	2119.00	-2.75
	1st Qu	-0.71	-1.02		1st Qu	2818.00	-0.65
	Median	0.04	0.02		Median	2939.00	0.28
	Mean	0.03	0.00		Mean	3031.00	0.00
	3rd Qu	0.74	0.98		3rd Qu	3164.00	0.40
	max	1.00	1.35		max	4677.00	4.96
	SD	0.72	1.00		SD	332.00	1.00
Profile curvature (1/100 Meters)				**Annual temperature (°C)**			
	Min	-10.36	-19.00		Min	10.81	-4.89
	1st Qu	-0.29	-0.56		1st Qu	20.38	-0.40
	Median	0.00	-0.03		Median	21.56	0.16
	Mean	0.01	0.00		Mean	21.23	0.00
	3rd Qu	0.30	0.71		3rd Qu	22.56	0.62
	max	10.00	18.29		max	25.55	2.03
	SD	0.55	1.00		SD	2.13	1.00
Plan curvature (1/100 Meters)				**Precipitation driest month (mm)**			
	Min	-6.02	-13.16		Min	34.82	-2.61
	1st Qu	-0.23	-0.54		1st Qu	91.52	-0.77
	Median	0.00	-0.03		Median	108.86	-0.20
	Mean	0.02	0.00		Mean	115.16	0.00
	3rd Qu	0.29	0.59		3rd Qu	133.83	0.61
	max	4.58	9.95		max	210.37	3.09
	SD	0.46	1.00		SD	30.78	1.00
Hillshade (-)				**Temperature coldest month (°C)**			
	Min	66.00	-8.14		Min	9.96	-4.92
	1st Qu	225.00	-0.41		1st Qu	19.37	-0.42
	Median	239.00	0.32		Median	20.57	0.15
	Mean	232.49	0.00		Mean	20.25	0.00
	3rd Qu	247.00	0.71		3rd Qu	21.60	0.64
	max	254.00	1.05		max	24.57	2.07
	SD	20.46	1.00		SD	2.09	1.00