Statistics methods for the assessment of coffee yield losses caused by pests and diseases under different production situations

Internship report

Eugénie CLEMENT

Tutors: Rolando Cerda B.

Jacques Avelino

Clémentine Allinne

From 9rd March to 31st August 2015

Department GMM - School year 2014/2015

Acknowledgments

I wish my thanks particularly to Rolando Cerda who supervised me during all this internship, especially for his availability and his explanation on coffee process (some time repeated!). I would like to thanks warmly Jacques Avelino who gave me the opportunity to do this internship, for his availability and more especially to all the vocabulary needed working in agriculture learned from him. Finally I wish thanks to Clémentine Alline, to all the time spent with me and all the things she brought me.
All of them make me complete this internship in the best conditions. I sincerely thank them a lot for all the time spend on meeting and discussion, all the energy and dynamic they have transmitted to me and their continuous support.

To all CIRAD staff present in CATIE who was available to answer to all my questions and technician staff which always welcomed me with their smile.

I would like to thank the pedagogical team of the department GMM of INSA Toulouse, who was reactive to all my questions and shared with me their knowledge.

Finally, this research is supported by CASCADE project which I thank. CASCADE is "Ecosystem-based Adaptation for Smallholder Subsistence and Coffee Farming Communities in Central America", funded by the International Climate Initiative (ICI). The German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU) supports this initiative on the basis of a decision adopted by the German Bundestag.

Fomentado por el:

Ministerio Federal
de Medio Ambiente, Protección de la Naturaleza,
Obras Públicas y Seguridad Nuclear

**en virtud de una resolución del
Parlamento de la República Federal de Alemania**

Impossible not to thank all my companions of "Andino", who made me discover many Latin-American cultures and with which I spent a completely amazing time in Costa Rica.

E. Clément

Abstract

Costa Rica and Central America, in general, recently knew epidemic on coffee plant which has an impact on losses of production. The need to qualify and quantify yield losses caused by pests and diseases become urgent. Factors which influence production loss are numerous and their interactions are complex. In the context of the doctoral research of Rolando Cerda which aimed at assessing coffee plantation yield losses, my objective was to propose statistical methods which permit to explain the influence of production situation on pests and diseases incidence and on yield. A conceptual model was created, to illustrate this link and bring to light factors impacting yield losses of the coffee plant. This study is using data from 69 coffee plots in rural part of the area of Turrialba, Costa Rica where we can find different production situations. A qualitative and quantitative approach was conducted to provide strong understanding of all the relation. A first step was the treatment of data, and creates variables. For the qualitative approach, the use of PLS (Partial Least Squares) approach to select information before clustering was very important. Finally Fisher's exact test and correspondence analysis permit to resume all the interaction between profiles of combination of pests and diseases incidences with typologies such as soil, management, climate, coffee plant production characteristics, etc. It permits to synthesize the impact of production situation on injury profiles, which them self have an impact on yield losses.

Key-words

Clustering – Injury Profiles - PLS – Canonical analysis – Yield losses – conceptual model

Résumé

Le Costa Rica et l'Amérique centrale, plus généralement, ont connu de récentes épidémies sur les plantations de café qui ont engendré de conséquentes pertes sur la production. La nécessité de qualifier et quantifier les pertes de rendement causées par les maladies et ravageurs (bioagresseurs) devenait urgente. Les facteurs influençant les pertes sont nombreux et leurs interactions complexes. Dans le cadre d'une thèse sur l'évaluation des pertes de rendements, mon objectif a été de proposer des méthodes statistiques permettant de mettre en lien l'ensemble des facteurs influençant la présence de bioagresseurs. Un modèle conceptuel a d'abord été établi pour illustrer ces liens et mettre en évidence les facteurs qui agissent sur les pertes de rendement des caféiers. L'étude s'appuie ensuite sur une caractérisation de 69 parcelles en milieu rural dans la région de Turrialba, au Costa Rica, où l'on peut rencontrer différentes situation de production. Une approche à la fois quantitative et qualitative a été menée afin de balayer au mieux les relations existantes.

La méthodologie statistique établie comprend une première partie concernant le traitement des variables. Concernant l'approche qualitative, l'utilisation de l'approche PLS (Partial Least Squares) a montré des résultats intéressants en éliminant les informations non nécessaire à notre étude. Cette méthode nous permet dans un deuxième temps de construire des typologies pertinentes pour chacun des facteurs qui affectent les bioagresseurs. Enfin une approche multidimensionnelle nous permet de voir les interactions entre ces typologies (le sol, la gestion de la parcelle, le climat, etc.) et leurs influences sur les bioagresseurs, qui eux-mêmes ont un impact sur les pertes de rendements.

Table of contents

List of figures

List of tables

List of abbreviations

AUDPC: Area Under Disease Progress Curve
AUDPS: Area Under Disease Progress Stairs
CATIE: Centro Agronómico Tropical de Investigación y Enseñanza
CIRAD: Centre de coopération Internationale en Recherche Agronomique pour le Développement
CPPC: Coffee plant production characteristics
IP: Injury profiles
PCA: Principal Component Anlaysis
PLS: Partial Least Squares
sAUDPC: Standardized Area Under Disease Progress Curve

Introduction

Agriculture is a fascinating domain, to the heart of many challenges of tomorrow. Statistics showed all his interest; William Gosset, creator of Student test for barley culture is an example. Using my statistic skills to help and revolve agriculture question is the best motivation to me. Coffee culture is leaving a high crisis, and research is needed. Costa Rica is well-known for its coffee production and CIRAD researcher currently working in this subject as a response of sector's crisis.

Cropping system research approach is appreciated as the focus is not made on the plant itself but on the overall environment characteristics and interaction. This perspective needs to take into account various factors. A plot (agricultural site) is considered as the study entity and not the plant as classically. Due to the large amount of factors, multidimensional statistics methods are required. My contribution was to develop methods which permit to highlight relation between yield losses caused by pests and diseases and production situations. This present report presents the context of the study. Then it's necessary to present first the final biological results to understand secondly the statistic research done to access to the final result.

1. Context of my internship

In this first part I will present my employer, CIRAD, then the campus where my internship takes place. I will expose the coffee crisis context, which explains the needs and issues that this project tries to answer and the goal and the motivation of my work. Finally I expose my objectives and how I organized the work.

1.1. Agronomic environment

1.1.a. CIRAD: a French agricultural research for development

CIRAD is a French research organization founded in 1984, is a public industrial and commercial enterprise under the joint authority of the Ministry of Higher Education and Research and the Ministry of Foreign Affairs. CIRAD is the agricultural research and international cooperation organization working for the sustainable development of tropical and Mediterranean regions. CIRAD main objective and target is to build "an agriculture capable of feeding eight billion human beings by 2030, an agriculture that is faire and preserves the health and the environment of all" (CIRAD). From its regional offices, CIRAD conducts joint operations with more than 90 countries.

Short information with key figures: CIRAD has a staff of 1800, including 800 researchers. It has an operating budget of 203 million Euros. It comprises 37 research units organized in three scientific departments:

- **Biological Systems**
- **Performance of Tropical Production and Processing Systems**
- Environment and Societies

As well illustrated in my internship, some projects regroup different problematic. My internship takes part in different research units.

I.1.b. Host structure: CATIE, a famous campus of agronomy in Costa Rica

The Tropical Agricultural Research and Higher Education Center (Centro Agronómico Tropical de Investigación y Enseñanza or CATIE) is a regional center dedicated to research and graduate education in agriculture, and the management, conservation and sustainable use of natural resources. Since 1973, date of creation, CATIE has the important mission of development in Latin America and Caribbean, in order to enhance human wellbeing and reduce rural poverty, combining science, education and innovation. CATIE is organized in three main divisions:

- Education
- Research and Development
- Outreach and Development

CIRAD is one of international research centers which are associated to CATIE, working with the division research and development in the departments of climatic change and agroforestry, interacting with other researchers from different countries and other research centers.

I.1.a. Work team

Jacques Avelino, is part of the department Biological Systems, in the unit Pests and Diseases: Risk Analysis and Control. He is a plant pathologist expert in coffee rust. Clémentine Alline is part of the department Performance of Tropical Production and Processing Systems in the unit Tropical and Mediterranean Cropping System Functioning and Management. She is an agro-ecologist, expert in farming system research.
Rolando Cerda is an agroforestry researcher, specialized in the measurement of **ecosystem services**[1] provided by agroforestry systems such as coffee and cacao. He is doing a doctoral research for the purpose to assess yield and economic losses caused by pests and diseases in a range of production situations in coffee agroecosystems.
He is supervised by Clémentine Alline and Jacques Avelino, Three of them are my supervisors. As you can see our team of work is a multidisciplinary one.

[1] **Ecosystem services** regroup all advantages and services of ecosystem such as coffee yield, carbon emission…

[2] **Shade canopy** represent the shade above coffee plant

I.2. Actuality and evolution of coffee production need research

I.2.a. Context of coffee crisis in Latin America

Latin America and Caribbean have suffered of a significant decrease of coffee production for two years. Pests and diseases are the main cause (Baker 2014). At the same time, coffee's price decreased while production costs increased. In consequence, plantations are abandoned or receive less management (PROMECAFE 2013). Future predictions of coffee production are worrying, principally due to one disease: coffee leaf rust. Farmers react at this disease but others pests and diseases could take advantage of the situation and are also becoming more important. Climate change aggravates the situation with change in rainfall and temperatures. Coffee tree is a plant that grows traditionally in the understorey of mid-elevation forests. We speak about an important coffee crisis.

Pests and diseases appearance and severity in the field is influenced by various factors such as:

- **Shade canopy[2],** contributes positively or negatively to the plant, depend of many aspect (management, structure and type of shade canopy…).
- **Climate**, rain and temperature which is influenced by the altitude of the plantation.
- **Management** (fertilization, pruning, regulation of shade…). Management can modify the microclimate and the physiology of the crop (Avelino et al 2004).
- **Soil,** (Avelino et al 2011) found that physiological plant resistance against pests and diseases is influenced by nutriment in soil

In this study, we define injury profile as the combination of several injuries caused by a set of pests and/or diseases. These factors listed above, are a key aspect of the study as they strongly affect the **injury profile.**

To obtain/assure good general production (coffee, fruits…), a balance in diversified **agroecosystem[3]** is recommended (Cheatham et al 2009). Agroecosystem needs evolution to be less vulnerable, and less dependent on management which is actually the unique form to

[2] **Shade canopy** represent the shade above coffee plant

[3] **Agroecosystem** is an ecosystem under agricultural management, connected to other ecosystem

counter pests and disease (Avelino et al 2011). Current behavior shows the contrary because coffee area has decreased but production is intensified (Jha et al 2014). In order to bring forward diversified and balanced system, a better understanding of the relation between these factors and the injury profile will permits to develop adapted way of controlling pests and diseases attacks.

I.2.b. Research needs

Following the context explained above, regional coffee organizations are interested in new studies on the sustainability of coffee production. With the aim to help all actors of coffee, it is important to explore that alternative model of production exists with appropriate cost and benefits (PROMECAFE 2013). Such production alternative will allowed farmers to cope pest and disease outbreak.

On the other hand, Scientifics identified that we miss information about crop losses in order to identify management characteristics plot in agroforestry system[4] which permit to regulate pest, diseases and good yield (Savary et al 2006b, Avelino et al 2011). Coffee plant is perennial, which complicate its analysis. Crop losses are difficult to identify because of the existence of primary losses (for example infected fruits) and secondary looses (for example dead branches which will not product the following year). This lack of information cause slow financial help from international funding agencies (Avelino, personal communication). Assessment of crop losses of coffee plantation taking into account parameters of production situation does not exist nowadays.

To understand better all reactions present in a biological process, an approach of cropping system research is necessary. A cropping system research considers an agricultural plot where the entire ecosystem is considered. Cropping system research is an integrated approach. This approach requires accurate measurement in order to reflect as much as possible the phenomenon happening at agricultural fields that could interact with the topic. Experts from different subjects are needed because various scientific fields are taken in consideration. Confident statistics methods are needed to generate useful results for the coffee sector (producer and technicians). These problematic are an integral part of the doctoral of Rolando Cerda.

[4] **Agroforestry** is a system combine, harvestable trees or shrubs in crop or pastureland which permit to improve the production.

The research takes place in the area of Turrialba, Costa Rica, an important coffee growing area in Central America. In this region, coffee is being cultivated under different conditions: climate, altitude, types of shade and management intensities.

1.3. Goals of my internship

1.3.a. State of art

Due to the multi-disciplinary aspect that implies working in agricultures and the multi-criteria approach we established, we gathered high quantity of variables. Statistics become more complex because multivariate approach is rarely done in biology. Methods which permit to estimate yield losses, identify injury profiles and determine relation under different production situation are scarce for all crops in general.

Farming system research is a constraining and recent approach, which is looking for statistics methods which permit to synthesize all the information measured. Diverse methods are used according to the objectives of the analyses. To evaluate and understand yield losses, we should consider various pest and diseases together, but also all the characterization of the production situation itself. It is this multidimensional approach which is a rarely assessed. To summarize information and behavior of coffee plant, in our case a clustering approach is preferred and will be set up. Its main advantage is the rapidly detect of complex and, nonlinear relations among several variables. Clustering also permits to normalize variables (S. Savary, 1995) A similar study was done in Asia for the case of rice pest (S. Savary, 2000). Figure 1 present the process of this study.

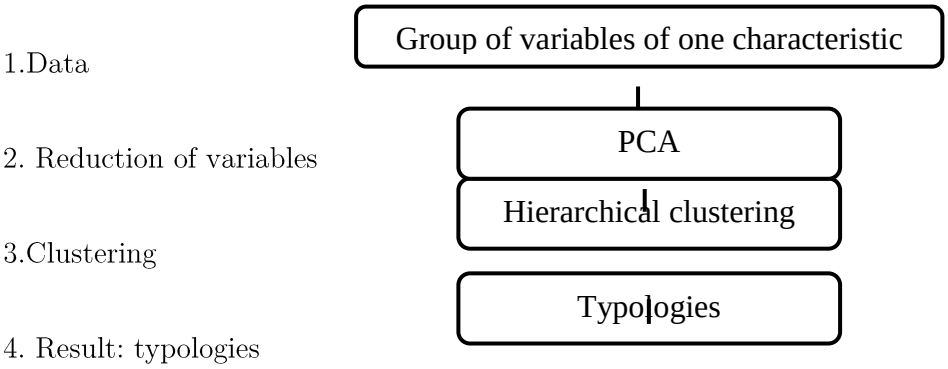


Figure 1 Process we can find in the literature

I.3.b. Aim of my work

I have the expectation to propose an alternative statistic method when classical methods are not possible or successful, for example to test the independence between two variables; in our case chi-square test is not possible. The main objective is to determine a methodology which permits to highlight relation between production situation of coffee plant and injury profiles and yield.
I create typology for different groups of variables (management, climate, etc.). Profiles and typologies represent the same thing: a group of agricultural plots with similar characteristics. In a way of clarity, I use the term profile for groups of pests and diseases and the term of typologies for others groups of variables.
My internship takes place after one year of study, only half of data of the project are collected. I will use this first year data set in my statistics exploration. Once all the exploration will be done, appropriate and optimal method will be use next year with the complete dataset.

During my internship I focused on two research questions of the doctoral research:

- What is the impact of production situation on injury profiles and yield formation?
- How do yield losses caused by injury profiles vary in function of production situations?

General objective:

Propose different statistic methods in order to assess coffee yield losses caused by injury profiles under different production situations, using data of one year of field measurements.

Specifics objectives:

- Organize the collected data into a data base.
- Test different statistic methods, with both qualitative and quantitative approaches, to evidence influence of situation production on the coffee production processes. In other words, illustrate relation presented in a conceptual model (Figure 3).
- Explain the advantages and disadvantages of the most appropriate and proposed methods.

I.3.c. Organization of my work

My first month was dedicated, to discover the environment, data, and contribute to scientific publications for an agronomy seminary in Montpellier in September 2015 (0, Annex C). To contribute to these publications I have manage a dataset called Colosse (presented in Annex A, as not part of my main objective).

As part of my main work, I organized data bases and created new variables (indicators). For example incidence of one pest is taken during all the year by branch, in our study we resume all this information with the value of sAUDPC, explain part II.2.b. Data bases evolved during all the internships depending on the evolution of the research. Then, I looked at the method of injury profiles' creation and explored several possibilities, on how to create adapted typology for each group of variables.

Finally, according to the result analysis, I focused on selecting and validating the statistical method which reflects the best agricultural practices and issues. During all the internship I strengthened my knowledge using scientific publications.

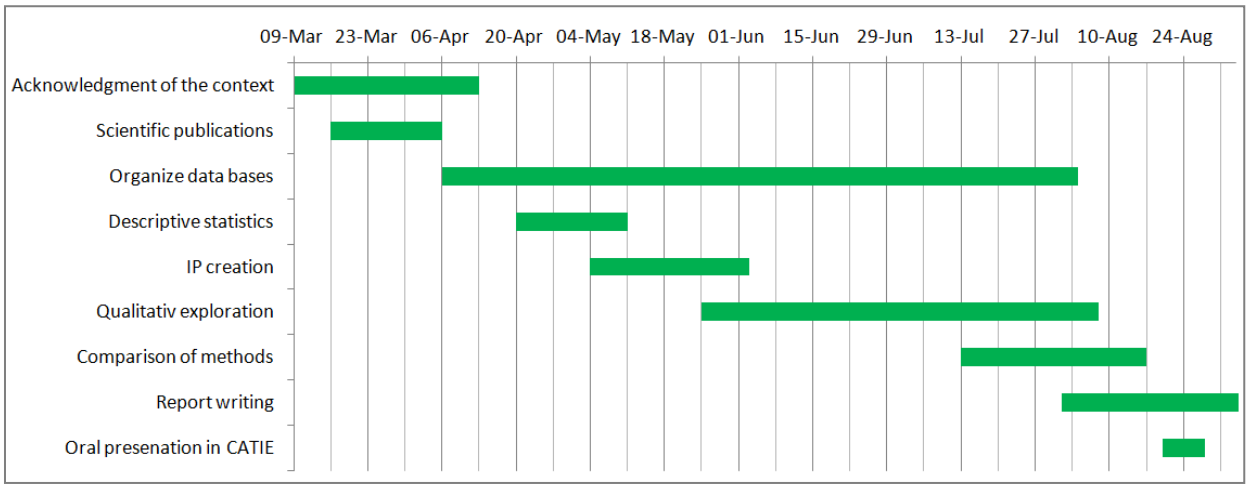


Figure 2 Gantt diagram of my work

II. How do production situation affects injury profiles?

In this part, I expose the direct biological conclusion learned from this work. I expose a conceptual model which permits to understand links in coffee production processes. From this conceptual model, I will present the methods used leading to final graphics which permit to relate effect of production situation on injury profiles.

II.1. Conceptual model

In order to understand my mission and the factor which affect coffee yield, I introduce a conceptual model for the assessment of yield losses of coffee (Figure 3). It permits to present relations between different factors which influence actual yield in the coffee plantation. Before going more into details, it is important to establish the definitions of several terms.

- **Attainable yield:** is the yield without being affected by yield reducing factors (especially pests and diseases) under all the possible control measurements, but limited by yield defining and limiting factors.
- **Actual yield:** is the yield that can be actually achieved using only the available resources (labor and inputs) of the farmer, generally affected mainly by pests and diseases (Savary et al 2006a; Savary and Willocquet 2014).
- **Yield loss:** is the difference between attainable yield and actual yield (Zadocks and Schein, 1979).

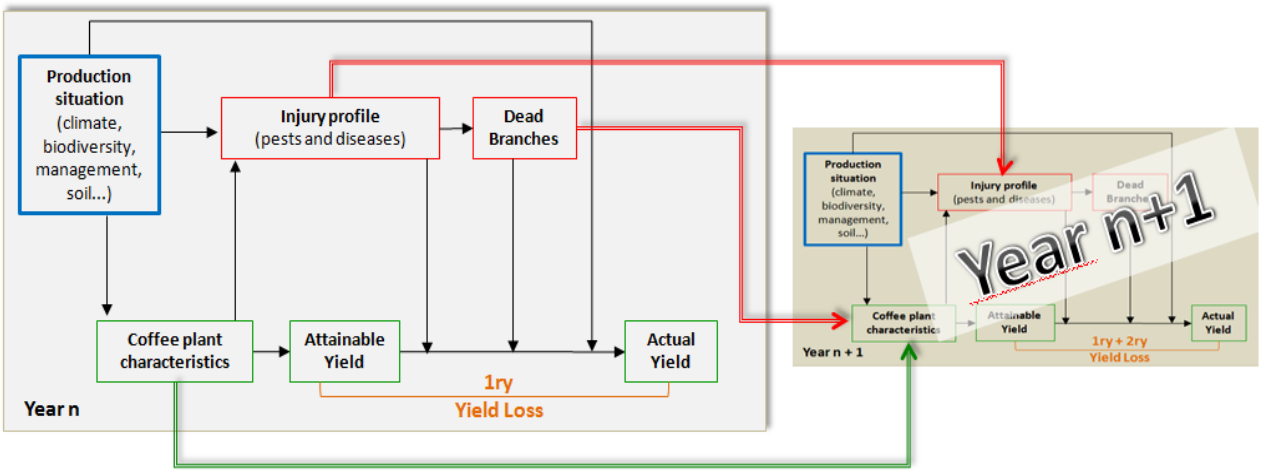


Figure 3 Conceptual model for the assessment of primary and secondary yield losses of coffee (R.Cerda, J.Avelino, C.Allinne, E.Clement, et al))

Explanation and illustration of conceptual model:

Every culture/study site has its own **production situation/context.** In our study, it is represented by the climate, soil, shade canopy (biodiversity) and management. We know that production situation influences pests and diseases, for example, one application of fungicide will have an impact on the presence of pests and diseases. Production situation has also an effect on coffee plant characteristics which include yield components as number of fruiting nodes. For example, components of the soil permit high growth of the plant; it also impacts deficiencies of leaves (nutritional status). Finally it affects directly yield, because of consequence of climate for example, climate events (storm, freeze, hard rains) which causes damages to the plants and fruit fall.

Pests and diseases are influenced by the coffee plant characteristics. Then pests and diseases influence the number of dead branches, subsequently both of them affect negatively yield.

Numbers of fruits by branch, number of productive branches are coffee plant production characteristics, part of coffee plant characteristics which determine attainable.

As I only have data from the first year of field measurement, secondary losses of pest and diseases outbreak could not be represented in this study but further work will done as part of Rolando Cerda doctoral research.

II.2. Data and statistics methods

II.2.a. A large source of information, plots from the CASCADE project

The set of agricultural fields selected for the project is part of a larger project called CASCADE (Central American Subsistence and Coffee farmer ADaptation based on Ecosystems). One part of this program focus on identifying and test the use of shade canopy as part of an overall adaptation strategy to help people to adapt to the adverse effects of climate change. This project takes places in three countries: Costa Rica, Honduras and Guatemala and several studies are being implemented.

The strategy for establishing this network was based on the selection of coffee plantations in different conditions of: climate (rain, temperature and altitude), shade canopy and management intensities of cropping systems. This network covers a wide range of production situations and permits the observation of a wide range of pest and diseases profiles.

Our part of the project CASCADE regroups 69 plots. At each plot, the study area is composed by 8 rows containing 15 coffee plants per rows. In this selected study area, 8 plants over 120

plants are selected for measurement. The selection is made with the aim to obtain a good representation of the plot.

Variables used in this study can be regroup in eight groups and cover two scales, the plant scale (number of fruiting nodes, presence of pests or diseases…) and the plot scale (composition of soil, temperature…) but mainly of my studies were doing by plot. A part of these variables are use to estimate coffee yield. Another part permits to describe production situation, and coffee plant to link with injury profiles and yield losses.

Pests and diseases variables

In our area of study we find 4 diseases:

- coffee rust (*Hemilia vastatrix* Berkeley and Broome),
- cercospora leaf spot (*Cercospora coffeicola* Berk. and Curtis),
- anthracnose (Colletotrichum spp.)
- a last category which regroup all others diseases of coffee.

We also measure the incidence of 2 pests:

- leaf miner (*Leucoptera coffeella* Guérin-Méneville)
- dieback (*Colletotrichum spp*).

Measurements of pests and diseases are taken five times per year, during different seasons (dry season, rainny season; before, during and at the end of the harvest). In 3 branches of each studied plant, the presence of pests and/or diseases is collected on each leave except for dieback which requires particular approach (more detailed are given in part II.2.b. for this variable).

Production situation variables

Production situation regroups various information which permit to describe the agricultural plot characteristics. They are regroup in 4 groups such as climate, agricultural management, shade canopy and soil. More details on the variables contained in each group are given below:

- **Soil,** one time a year, 11 chemical components of soil are measured (pH, Acidity, Calcium, Carbon, Nitrogen, etc.) Soil texture is also taken and express as a percentage of sand, silt and clay.

- **Climate**

In our study, climate is characterized by the following information of the plot:

- Altitude
- Rainfall
- Slope
- Orientation of slope
- Temperature
- Humidity

Temperature, rainfall and humidity are taken during all the year, hourly records; the mean value is calculated for different periods following the season.

- **Shade canopy**

It represents the variety and density of trees present in plots in addition of coffee plant. For four different groups of types of trees (fruit trees, timber trees, musaceas trees and services trees), we have the number of tree (density), the trunk basal area and the number of species.
We also calculate the Simpson index which reflects number of trees diversification (diversity index: 0 plot with high diversification, 1 plot with low diversity) and the Shannon index reflects proportion of trees diversification from 0 to 1 (0 only one specie is present, the more high is the index value, the more high is species diversity).

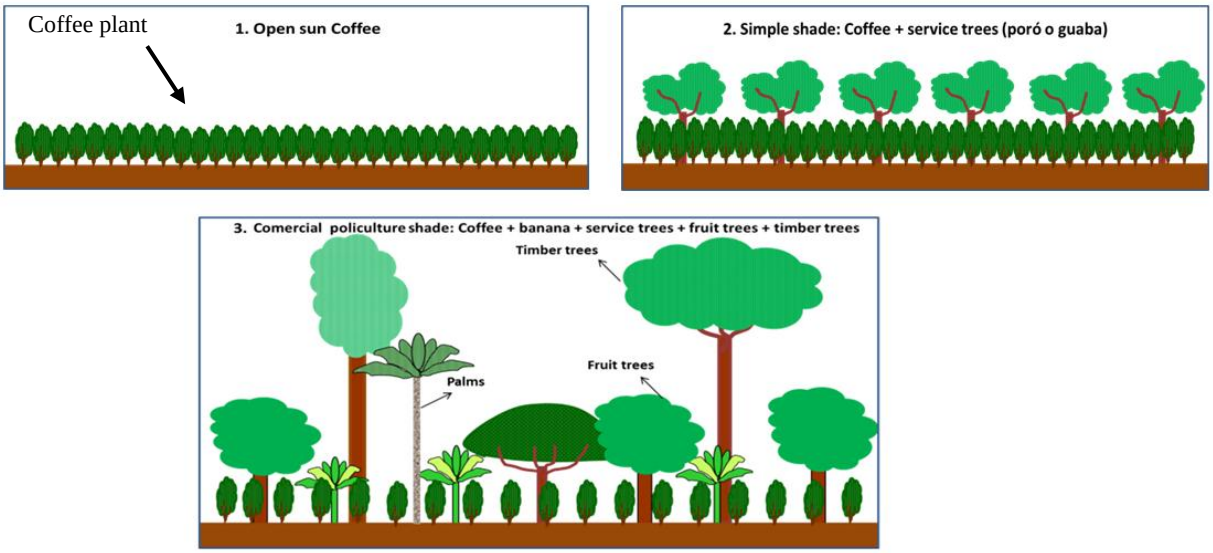


Figure 4 Types of shade representing the botanic complexity in the coffee research plots network

Figure 4 shows different type of shade canopy that we can find in our set of plots. Some have a full exposure to sunshine, others have simple shade from associated service trees and finally some have a high diversification of shade.

- **Management** is evaluated by the number of
  - applications (fertilizer, fungicide, herbicide)
  - weeding,
  - pruning of coffee
  - harvest of coffee berries
  - different practices applied
  - the sum of all the number of practices applied
  - the percentage of shade
  - the distance between coffee rows and plants
  - the orientation of rows

  Such information is collected at every plot, one time per year.

Coffee plants characteristics variables

- **Plant deficiency**

Plants with symptoms of nutrient deficiencies will be counted in the plot. Seven deficiencies are measured: calcium, iron, magnesium, phosphorus, zinc… We consider the maximum of the value between the two measurements taken, one at beginning and the other at the end of the harvest season.

- **Chemical analysis of coffee leaves**

At the end of the harvest season, several leaves are sampled from the selected plants. Leaf sampling happens in the high, middle and lower part along the plant's main stem. Laboratory analysis permits to know the leave content of 10 chemical elements (calcium, carbon, copper, iron, etc.).

- **Coffee plants production characteristics**
  - Height of productive stem
  - Number of productive branches
  - Number of fruiting nodes per branch
  - Number of stems per plant
  - Number of productive stems per plant
  - Number of fruits per node
  - Number of fruiting nodes per plant
  - Age

You can find in Annex E the detail (unit, mean and standard deviation) of all variables that I used in this study.

II.2.b. Creation of variables

- **DieBack**

DieBack is a representation of the progressive death caused by the combination of negatives factors, mainly pests and diseases. The scale is: 1, defoliation; 2, several dead branches; 3, dead branches, in the major part of the plant; 4, dead plant. From the contingency table of observations, we calculated a dieback index.

$$Ind\_DB = \frac{1N_1 + 2N_2 + 3N_3 + 4N_4}{4\,(N_1 + N_2 + N_3 + N_4)} \times 100 \qquad (1)$$

Where N_i represents the number of coffee plant corresponding to the scale i of DieBack

Table 1 Example of contingency table of DieBack observations

Die Back value	1	2	3	4
Number of plants	18	12	43	5

Observations from the example (cf.Table 1), gives a DieBack index of 61.21% which means that the plot is under a high disease attack. The lower is the index; the more healthy is the coffee plant in the plot.

- **Area under disease progress curve**

To compare the incidence of pests and diseases, we need to calculate an index which resumes the incidence of the disease. The area under the disease progress curve (AUDPC) permits to resume a series of data in only one value (Wilcoxon, 1975, Shaner and Finney, 1977). The AUDPC, an integral model, is useful to relate loss of diseases measurement over a specific period of crop growth. AUDPC can be estimated using the following equation:

$$AUDPC = \sum_{i=1}^{n-1} \frac{y_i + y_{i+1}}{2} \times (t_{i+1} - t_i) \qquad (2)$$

y_i: percentage of leaves with pest
t_i: time (in day) of the measurement
n: number of assessment times

To compare AUDPC of cases with different time of measurements, it is important to standardize them by the number of days.

Table 2 Values of rust incidence, 7 measurements

Date	18/7	13/8	17/9	18/10	20/11	23/12	27/1	AUDPC	sAUDPC
% of leaves with rust	26.93	23.12	17.55	11.64	10.51	13.41	12.84	3034.14	15.72

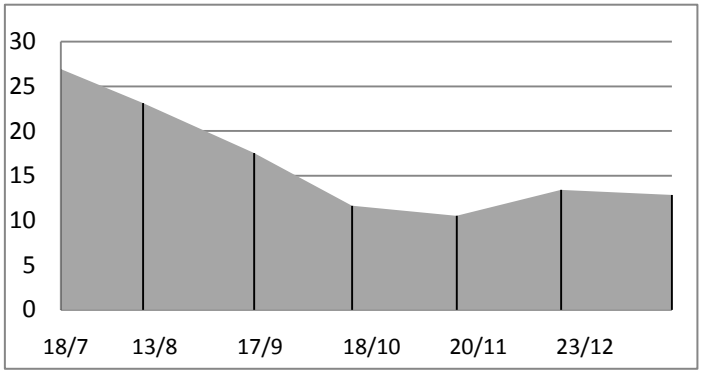


Figure 5 AUDPC of rust incidence

This example (cf. Table 2 Values of rust incidence, 7 measurementsTable 2 and Figure 5) presents the incidence of the rust at each measurement. We see that rust was more present in July. November is the period with less rust incidence occurs just before a small increase of rust incidence. This type of graph permits to compare incidence from different pests and diseases and their evolution during a given period.

With this measurement, it is important to take into account the first and the last observation can be underestimated; that is why the area under the diseases progress stairs (AUDPS) was also proposed. Although AUDPS success to outperform AUDPC, it can be less precise in some situations. (I. Simko, HP. Piepho et al, 2012). In our case AUDPC was selected because we have a comparatively large variance in the first and last observations, a situation where AUDPC is preferred.

II.2.c. Preliminary study: estimation of yield

To reduce confusion, harvest of a coffee plant represents the actual yield. Collecting data of coffee harvest is a very intensive work explaining that coffee yield is available only for few coffee plants (35). This data base will permit to estimate yield for all coffee plant of our study (expressed in gram). Secondly, creation of a yield model will allowed to evaluate the impact of pests, diseases and dead branches on actual yield and to estimate the attainable yield. As the conceptual model illustrates (Figure 3) pests and diseases, dead branches, coffee plants characteristics and production situation influence directly yield. Estimation of yield was part of a preview study; I have only used it and improve it with the collected data. For the coherence of the study, and to provide accurate estimation of yield losses I find it important to present it even if it was a small part of the work. In this model, the direct effect of production situation on yield cannot be taken into account because of its complexity. Effect of pests and diseases, dead branches and coffee plant productive characteristics are studied.

Method:

This model was established by using 8 variables of coffee plants production characteristics, 6 on pests and diseases and the number of dead branches. Effect of each farmer on the plot, is taken into account because plot is considered to have random effect in the model. A transformation of yield with square root is performed to obtain a validated model. Selection of variables is doing with the criteria of AIC (based on the likelihood function and the number of parameters).

Result:

5 variables were selected by the criteria of AIC, only one disease was present. Measuring one pest or disease incidence is a heavy work that is why we compared the complete model called, model 1, with the same model without the information of cercospora disease, called model 2. This comparison is to look at if the contribution of data of cercospora disease is important.

Table 3 Value and p-value of each component of model 1 and 2

	Model 1		Model 2	
	Value	p-value of anova	Value	p-value of anova
Intercept	10.19	<0.01	8.58	0.01
NPS	4.28	<0.01	3.88	<0.01
NFNode	2.07	0.02	1.95	0.03
NFN_Plant	0.03	<0.01	0.03	<0.01
DeadB	-0.20	0.02	-0.18	0.04
sAUDPC_cerc	-0.38	0.04	X	X

NPS: Number of productive stems, NFN_Plant : Number of fruiting nodes per plant, NFNode : Number of fruits per node, DeadB : Number of dead branches, sAUDPC_cerc: standardized Area Under the Disease Progress Curve of cercospora (% day^{-1})

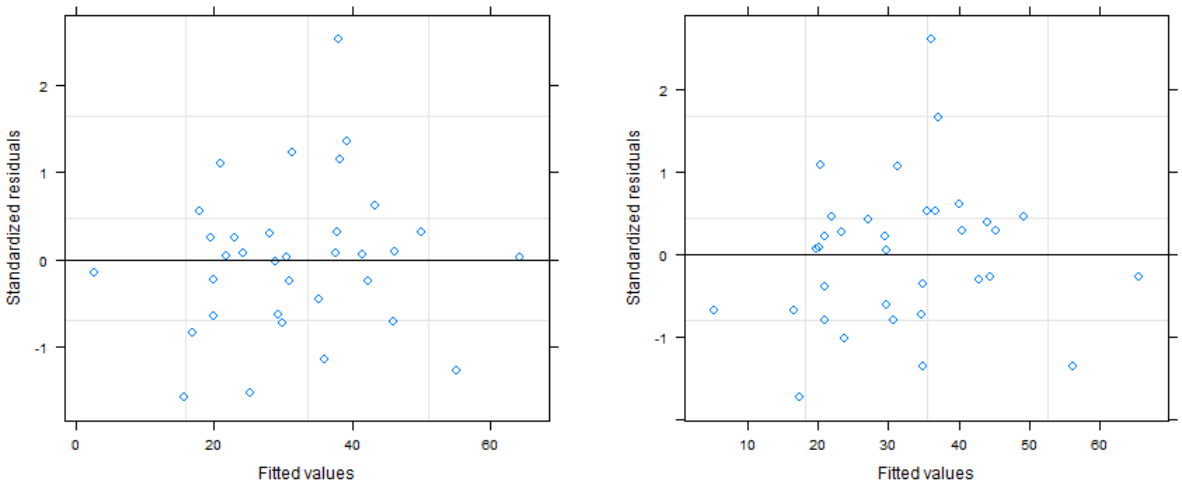


Figure 6 Representation of residuals of model 1 and model 2

Variables selected by AIC are significant (cf. Table 3), residuals are randomly distributed so both of model 1 and 2 are validated (cf. Figure 6).

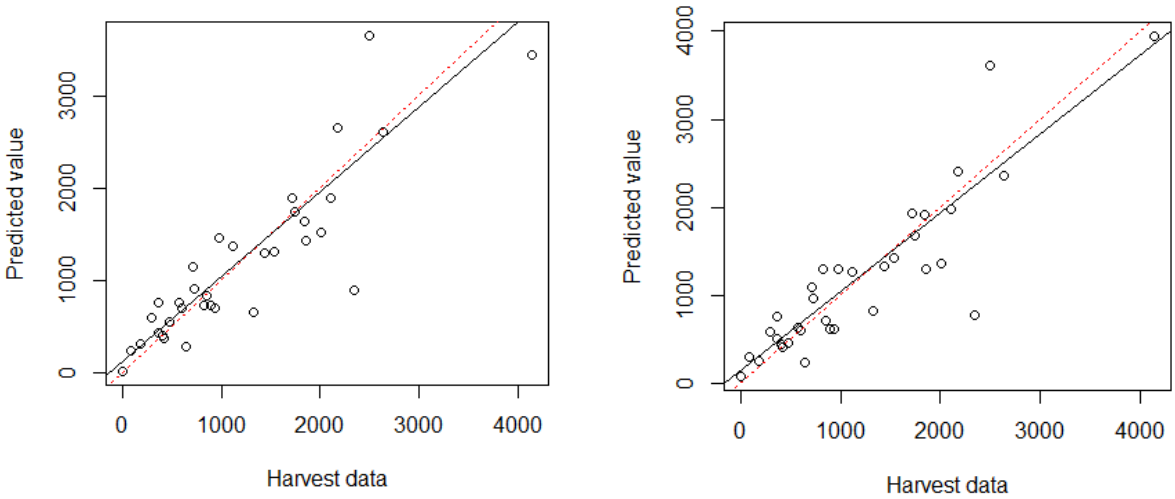


Figure 7 Representation of fitted values

Red line represent when observation is equal to fitted value. Black line represents the linear regression between observations and fitted values. These models slightly under estimate the yield for high value of harvest. In the other hand, low values of harvest are overestimated. Even if model 2 used fewer components, they both give similar predictions.

Table 4 Indicator for model 1 and 2

	AIC	BIC	R^2	RMSE	RRMSE
M1	225.94	238.16	0.77	429.24	0.36
M2	229.66	240.34	0.77	426.03	0.36

Model 2 has a similar AIC and BIC of model 1. In addition model 2 has a better root mean square error (RMSE). Thus, model 2 is chosen to estimate the actual yield. From this model remove the negative impact of dead branches, we obtain an estimation of attainable yield. Differences between attainable yield and actual yield give an estimation of yield losses.

Categorize of yield and yield losses is needed for the following of the study, especially when we will analyze the interaction between all typologies with injury profiles. It is clustering in 3 equitable groups, with rules present in Table 5.

Table 5 Cluster of actual yield and yield losses

Y1	yield_ha < 4000	YL1	yield losses < 1000
Y2	4000 < yield_ha < 8000	YL2	1000 < yield losses < 3000
Y3	8000 < yield_ha	YL3	3000 < yield losses

Comparison of yield between agricultural plots is done using the plot yield. Computation of the plot yield per hectare is done by multiplying the plant yield by the density of plants per plot. As cropping system research is always looking of simple indicator, a first conclusion can be done. The dead branches variable seems to resume the impact of incidence of pests and diseases on actual yield.

II.2.d. Statistics methods used

Mathematics details of statistic method are developed part 0 where the different methods are discussed.

Selection of variables

- **Injury profiles**

To construct injury profiles, we use components from PCA which represent more than 85% of the cumulative variability of pests and diseases.

- **Typologies**

To construct typology we perform the Partial Least Square Discriminant Analysis (PLS-DA) which put though injury profiles and variables from the typologies we want to create, for example 17 variables for climate. From this, we extract the 2 components, representative the group of variables, as the data base for the creation of typologies.

Clustering

To cluster plots we use hierarchical clustering. We use Euclidian distance between plots and a Ward method for the clustering. This method merges, at each step, the two plots which minimize the increase of the total within-cluster variance.

Analyze

To describe each typology, analyses of variance using linear models is done and the test Fisher's Least Significant Difference with $p < 0.05$ is perform to compare the three injury profiles.

To look at the link between typologies and injury profiles, Fisher's exact test is be performed, because Chi square test conditions are not reunited. Finally to observe all the combination of typology with injury profile, a graphic resulting from a correspondence analysis is perform.

II.3. Biological results with one year of data

This part presents a description of pests and diseases incidence then the construction of the injury profiles. Finally I expose preliminary conclusion, of agronomy point of view, we obtain with data of one year.

II.3.a. Descriptive statistics of pests and diseases

In our area of study we find many pests and diseases, only pests and diseases present in many plots were studied, which reduces the number to six pests and disease. We use sAUDPC to represent the incidence of each pest or disease per plot, which can be interpreted as the average value (expressed in percentage) of leaves attacked by the pest or/and disease during all the study. For the DieBack, this is more an index evaluated in percentage of the health of the plant. Detail explanation of the index calculation is presented part II.2.b.

Table 6 Summary (mean, range, etc.) of variable sAUDPC for each pests and diseases and DieBack index

	Rust	Cerc	Antr	Miner	Others	IndDB
Mean	60.83	4.88	1.56	1.14	9.29	45.80
Median	62.94	3.58	1.16	0.82	8.84	46.35
Sd	14.33	3.29	1.22	0.94	4.59	13.45
Max	88.23	13.80	5.27	4.66	25.72	95.57
Min	12.78	0.67	0	0	2.01	4.73

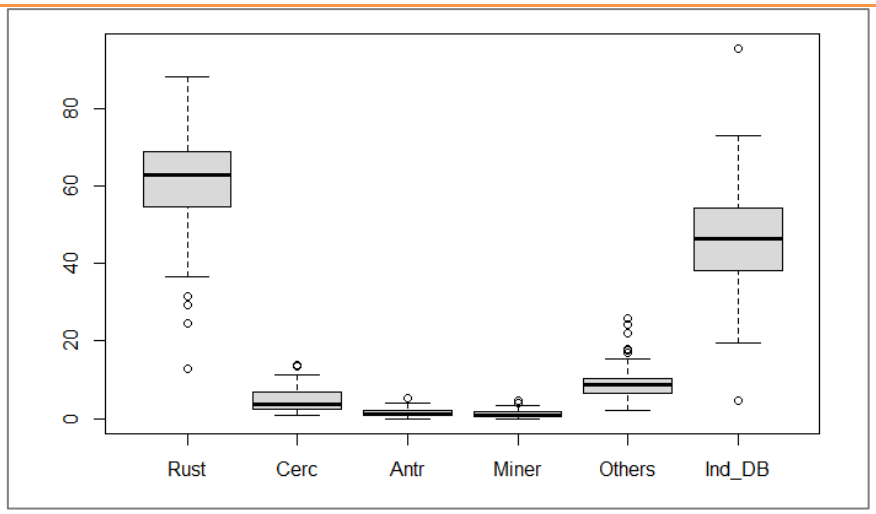


Figure 8 Distributional characteristics of each pests and diseases

As explained in the context, coffee rust is the main disease; and was present in every plot at all time along the study period, with different intensity. Dieback shows a high variability from 4.43% to 95.57%. That proves we have very different diseases prevalence in the study area. Incidence of all others diseases are more present than cercospora disease, anthracnose and miner. Figure 9 shows the geographical repartition of the 69 plots and the incidence of each pest or disease per plot. There repartition seems to be not correlated to the geographical situation, illustrated by the randomly repartition. We also remarked that the incidence of one pest or disease at one plot is not correlated to the incidence of another pest or disease at another one.

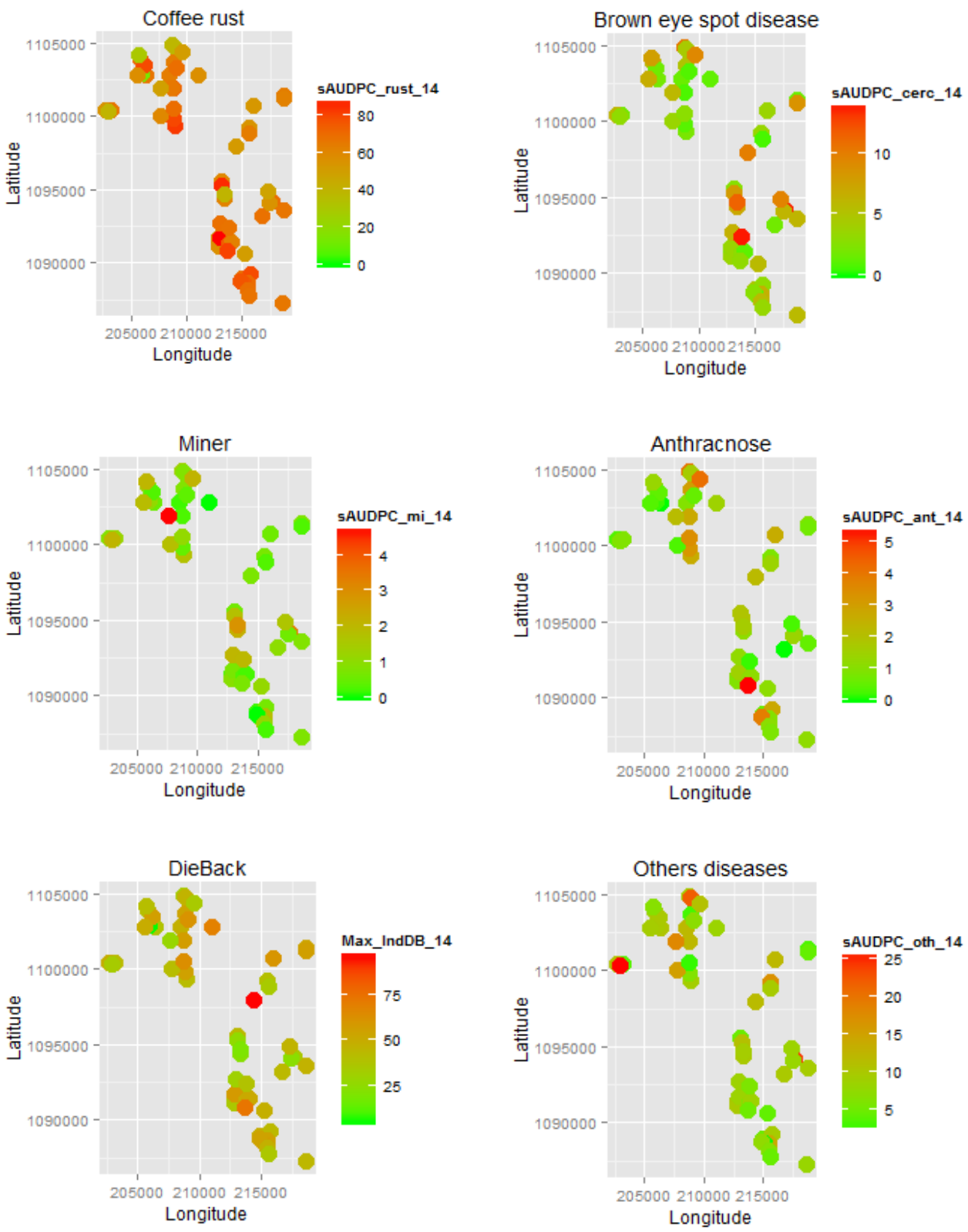


Figure 9 Geographical representation of pests and diseases incidence, format = UTM, units = meters

To look at the relation between pests and diseases, we calculate the Spearman correlation.

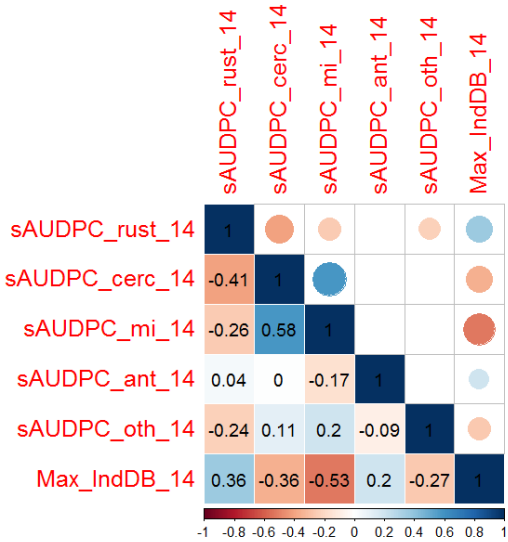


Figure 10 Matrix of correlations (Spearman) among pests and diseases

Above the diagonal, white cell mean there is no significant correlation. Anthracnose and others diseases doesn't have important correlation with others pests or diseases. Miner has a high negative correlation with DieBack, and a positive correlation with cercospora leave spot. Rust and cercospora leave spot are negatively correlated.

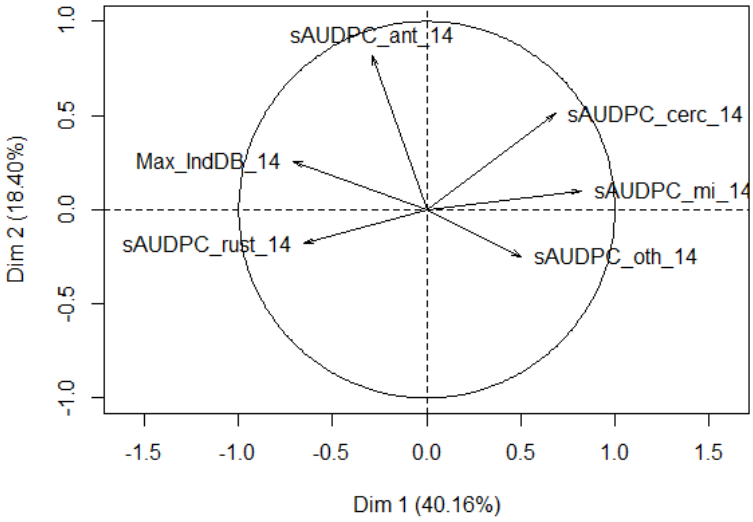


Figure 11 Representation of pests and diseases variables in the two first axis of the PCA

The two first axes of the principal component analysis (PCA) represent 58.56% of the total variability. Even if we have only six variables which represent pests and diseases, 2 axes are not sufficient to represent pests and diseases, 4 axes will be choosen. Coefficients of spearman correlation are confirmed. In this analysis, a group composed of cercopora, miner and others diseases is opposed to variables of coffee rust and DieBack. The next part presents, injury profiles which regroup plots with close behavior concerning the incidence of pests and diseases.

II.3.b. Clustering to create injury profiles

Clustering permits to describe our dataset and to illustrate link with variables of production situation. We use the four first components from the PCA as variables. Number of components is selected by the rule of keeping the number of component which permits to represent more than 85% of the variability. We use hierarchical cluster analysis as the method to regroup plots with close behavior. The analysis of the dendrogram (Annex D, Figure 17) suggests to represent our data set with 3 different injury profiles, we will called them IP1, IP2 and IP3.

Table 7 Means of pests and diseases incidence for each IP

Different letters along rows indicant significant differences among injury profiles (p<0.05)

	IP1	IP2	IP3
Number of plot	33	24	12
Rust	66.8 a	50.2 b	65.6 a
Cerc	3.4 b	7.6 a	3.5 b
Antr	1.0 b	1.4 b	3.1 a
Miner	0.7 b	2.0 a	0.5 b
Others	7.6 b	12.2 a	8.1 b
IndDB	46.8 b	36.6 c	60.4 a

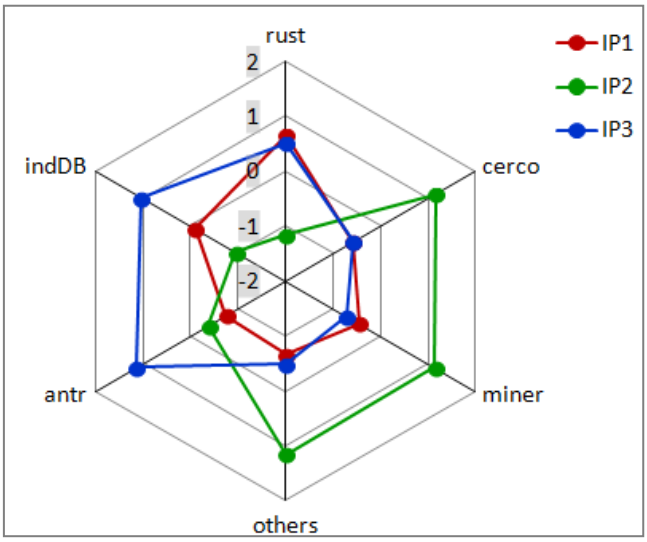


Figure 12 Radar chart representing difference of pests and diseases incidence for each profile

Description of injury profiles

Figure 12, represents the mean values standardized of the incidence of pest and disease for each profile. IP2 is markdown from IP1 and IP3. It regroups plots with few incidence of rust, and a few Dieback but plots of IP2 have high incidence of cercospora, miner and others disease. Difference between IP1 and IP3 are a high presence of anthracnose and dieback in plots of IP3.

Dead branches and yield according to injury profiles

Table 8 Means of different yield and dead branches by injury profiles

Different letters along rows indicant significant differences among injury profiles (p<0.05)

	IP1	IP2	IP3
Yield_ha	7558 b	5791 ab	4710 a
Yield losses	2184 b	689 c	4111 a
% yield losses	24% b	11% c	43% a
Dead branches	29.52 b	18.37 c	51.43 a
% dead branches	24.12 b	10.96 c	43.62 a

As the estimation of yield losses is approximate using the presence of dead branches, yield losses obviously follow the trend of dead branches, low presence of dead branches involve low loss. We can see that each injury profile have differences in yield and yield losses. IP2 has less loss than IP1 but has also a lower yield. IP3 seems the worse profile with a low yield and high yield looses.

II.3.c. Results of typologies

Description of each typology for each group of variables is not developed, only some characteristics will be introduced in the analysis. You can find a summary of each typology in Annex F.

Table 9 p-value of Fisher's exact test

	IP	Actual yield	Yield losses
Shade canopy	0,10	0.85	0.81
Climate	0,01	<0,01	0.02
Soil	0,04	0.38	0.51
Management	<0,01	0.58	<0,01
CPPC	<0,01	<0,01	0.05
Leaves	<0,01	0.04	<0,01
Deficiencies	<0,01	0.38	<0,01

Table 9 presents result of Fisher's exact test which tests the independence between injury profiles, actual yield and yield losses category's (detail Table 5) and typologies.

All typologies created are linked with injury profiles except typologies of shade canopy. Typologies of soil and typologies of shade canopy are independent with yield losses' categories. Finally, only typologies of climate, CPPC and chemical analysis of leaves are not independent with yield category.

Next step present the interaction between all typologies with injury profiles

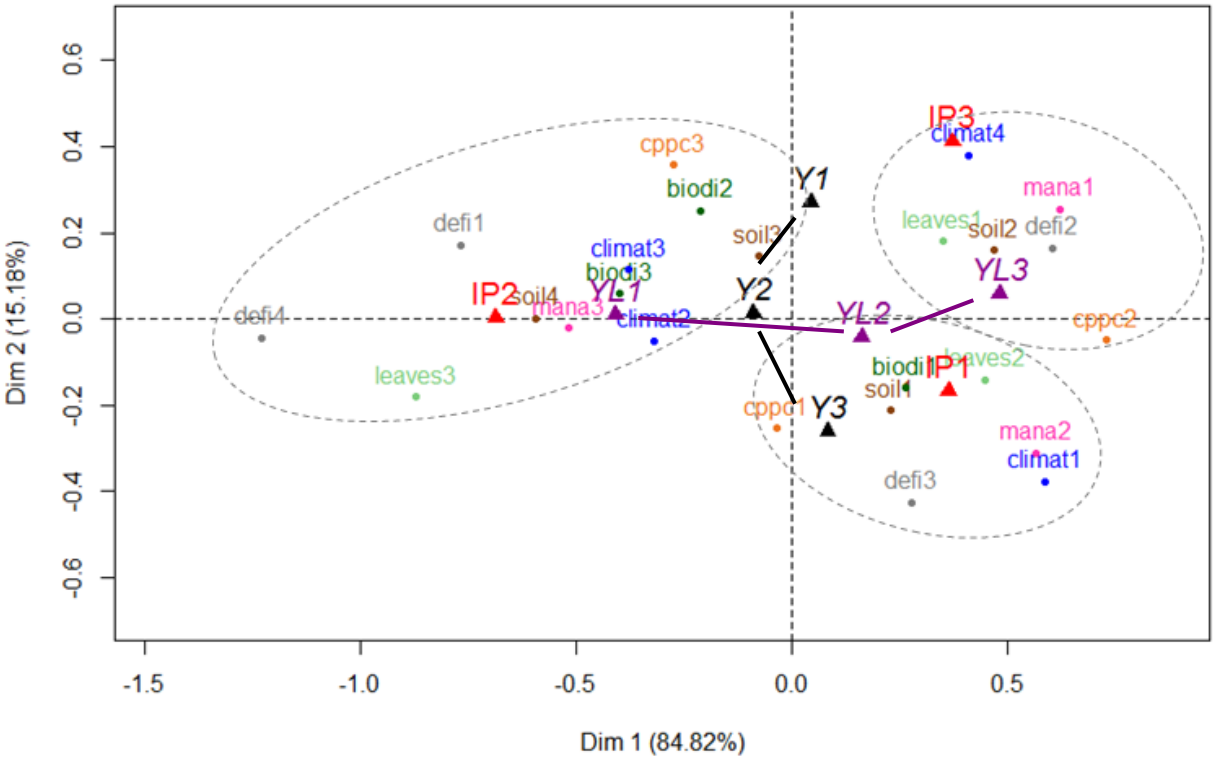


Figure 13 Correspondence analysis of injury profiles with typologies. Yield and yield losses has illustrative variables

- <u>Actives variables of correspondence analysis</u>

First axe separates IP2 with IP3 and IP1, the incidence of rust, cercospora, minador influence this axe and represent 85% of the variability. Axe 2 opposes IP3 and IP1, incidence of anthracnose has an influence in this axe.

- <u>Illustrative variables (yield and yield losses) of correspondence analysis</u>

Concerning illustrative variables, first axe represents an axe of yield losses, and second axe actual yield. Yield and yield losses are independent according to this representation.

- <u>General interpretation</u>

IP3 has high yield losses and a low yield, it has the worse situation. Typologies close to IP3 are climate characterize with high rainfall and temperature, in low altitude (climat4), low management (mana1) and old and high plants (cppc2).

IP1 has a good yield but at the same time important yield looses, this type of plot has a high management (mana2) and they are present in low altitude with low rain, high temperature (climat1) and young plant (cppc1).

IP2 has very few yield losses and mean yield. It regroups plots with an important distance of rows between plant, and a middle management (mana3). Climate is characterized by with lots of rain, in high altitude, with cold temperature (climat2 and climat3). Coffee plants are characterized by low number of fruiting node and fruit per node (cppc3).

Relation with deficiencies, analyze of leaves and soil should be done to precise this study and provide more accurate conclusion. Such analysis requires strong agronomical competences and was not developed in this report.

To conclude and making advices to farmers, costs and economic aspect need to be compare with yield. High yield with high production costs can be less profitable than a lower yield with low production costs. This will be done in the doctoral research of Rolando Cerda.

III. How to optimize the statics process to highlight the relation between production situation and injury profiles?

III.1. Introduction

To illustrate the influence of production situation and coffee plant characteristics on injury profiles many possibilities were investigated. A qualitative approach is set up. Three steps are important part of the process: data used to the clustering, method of clustering, investigating the link between typologies, injury profiles and their interactions.
For production situation, four typologies were created: shade canopy, management, climate and soil. Three typologies describe the coffee plant characteristics: plant nutrient deficiencies, chemical analysis of leaves and coffee plant production characteristics. The objective of creating typologies under the approach of this research was not doing classical clustering, but doing clustering and at the same time constructs typologies which maximize the relationship with injury profiles. That means to keep only information of the group of variables which showed significant relationships with incidence of pests and diseases.

Selection of data entering in the process of cluster is the key subject of this following part.

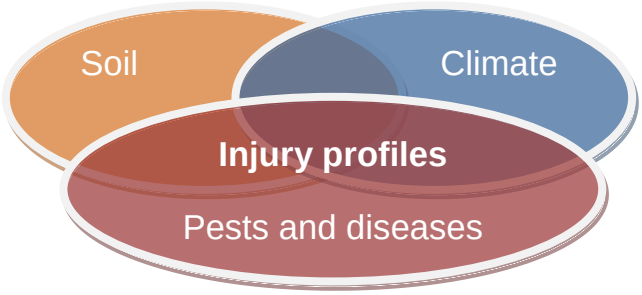


Figure 14 Illustration of the need of selection of information

Figure 14 illustrates, in a synthetic view, the fact that not all the information of each group of variables is necessary to explain the impact of pests and diseases. For this step many methods

were investigated. As an example, one investigation of selection of variables was done one by one. For each variables of each group, we tested if variable had a significant relation (anova test from linear model) with at least one pests or disease. From his result, we selected variables having significant relation. This method was not performing; selection didn't have biological meaning and relations were not significant.

In the following report, I present and compare five multivariable methods of information's selection.

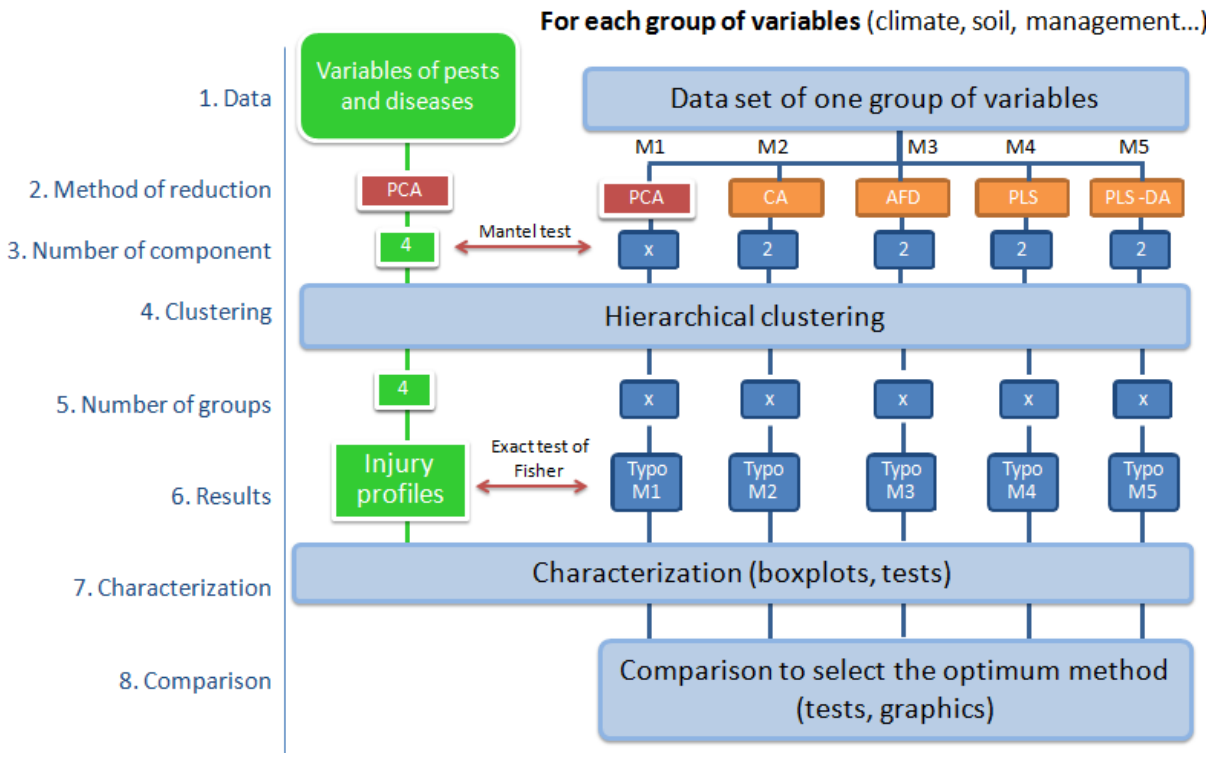


Figure 15 Graphical explanation of the research study

Figure 15 presents the different steps of the research study. Only methods of reduction of variables will be discussed in this report. In green we have the "classic" process of the creation of injury profiles, with the use of components from PCA as variables for the clustering. The construction of injury profiles will not be approach as we keep all information of pests and diseases the problem of selection information did not occur. The right part represent the process done for one group of variables (example: soil) and repeated for the each groups of variables.

Different methods of clustering were compared: k-means, Partitioning Around Medoids (PAM), hierarchical clustering. Relatively quickly, hierarchical clustering was the only one who gave coherent results with biology point of view.

How to optimize the statics process to highlight the relation between production situation and injury profiles?

Methodology to construct typologies can be resuming with the seven following steps:

Step 1: Which method should be used to create component in order to create typology maximizing link with IP?
Step 2: Mantel test permit to test the proximity of the variables used to clustering with pests and diseases
Step 3: How many components I keep?
Step 4: Selection of number of profiles/typologies
Step 5: Description of the new profiles/typologies
Step 6: Fisher's exact test to look at the relation with IP
Step 7: Characterization of profiles/typologies (boxplots, anova, LSD test, etc.)
Last step (not presented Figure 15) is how to assess the relation between typologies and IP, yield and yield losses. To visualize actual yield and yield losses in this graph, these quantitative variables are categorized in 3 clusters which regroup close number of plots (detail Table 5).

III.2. Statistic methods for reduction of variables

Five methods are comparing. Result from components from the PCA can be seen as the reference because we only reduce data and not doing a selection. Then, canonical analysis and PLS are perform and explain below. Both methods permit to put through 2 set of variables, X and Y. A summary of those methods is presented; they are extract from Tenehaus (1998) and Wikistat.

We choose, X as the small dataset, in our case, it is always the dataset representing pests and diseases: the 6 variables of incidence or the qualitative variable of injury profiles. Y represents a group of variables from production situation or coffee plant characteristics. We define p as the number of variables of X, q number of variables of Y.

III.2.a. Canonical Correlation analysis

Canonical Correlation analysis created by Hotelling, H. (1936) is the base of different multivariate analyzes which permits to put through 2 set of variables, X and Y. When X is equal to Y we find the PCA.

We define 2 subspaces of variables F_X generates by vector x_j $(j = 1, \dots, p)$ and F_Y generates by vector y_k $(k = 1, \dots, q)$. We note $R_{XX} = \frac{1}{n}X'X$ and $R_{YY} = \frac{1}{n}Y'Y$ the correlation matrix intra X and Y.

$R_{XY} = \frac{1}{n}X'Y$ represents the correlation matrix inter X and Y, as PCA perform to describe R_{XX} canonical analysis will permit to describe the correlation matrix $R_{XX} = \frac{1}{n}X'X$.

Canonical analysis consists to find a couple (V_1, W_1) which maximize the correlation between these 2 components. V_1 is a linear combination of variables of x_j. W_1 is a linear combination of variables of y_k. Then, the process finds a couple (V_2, W_2) which maximize the correlation between these 2 components. V_2 is a linear combination of variables of x_j but not correlated to V_1 . W_2 is a linear combination of variables of y_k, not correlated to W_1 and so on.

V^s variables form an orthonormal basis of F_X, similarly, W^s variables constitute an orthonornal system of F_Y. Variables V^s and W^s are called canonical variables. Finally we keep the first two variables, superiors dimensions concern low correlation which we don't want to represent. Moreover we have a small X with six variables that why we can choose only 2 dimensions. From them we obtain a representation of each observation in the new plan of 2 dimensions.

Canonical analysis needs the inversion of matrix X'X and Y'Y, that why it cannot be perform when number of observations is lower of number of variables.

- Different applications of canonical correlation analysis:

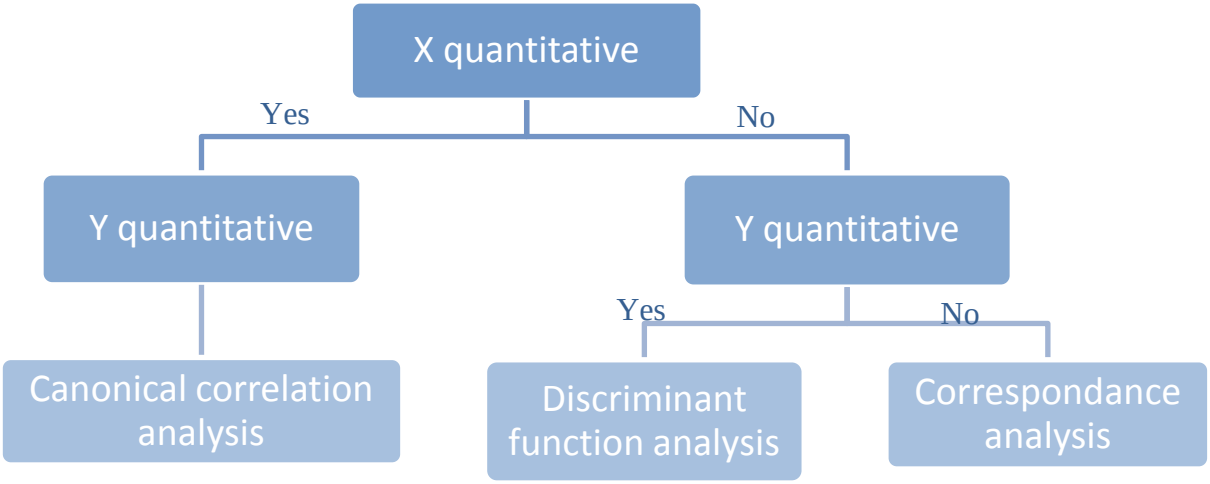


Figure 16 Summary of method base on canonical analysis

In the case where X is a qualitative variable, canonical correlation analysis is called discriminant function analysis. It's the same process, qualitative variable is transform as indicator. In the other case where X and Y are qualitative variables, we named it correspondence analysis.

Canonical analysis is perform with the MixOmics package in R. Discrimant function analysis is perform with the MASS package in R Correspondence analysis is perform with FactoMineR package in R.

III.2.b. Partial Least Squares regression

PLS is an old method (Wold, 1966). We focus on the regression version of PLS. Main advantages of PLS are the possibility to perform it with missing data and also when number of observations is lower than number of variables. PLS permits to modeling variables, X, by another dataset, Y, as canonical analysis. The principle is the same, find latent variables solution of:

Max $Cov\,(Xu, Yv)$ with ||u||=||v||=1

Canonical analysis maximizes the criterion of correlation and not covariance. PLS regression permits to consider that one group of variables, X, is explain by another one, Y.

PLS is based on NIPALS algorithm:

X and Y are two matrix with centered data

Initialization of ω_1 by the first column of Y

For h=1 to r **do**

While Convergence not achieve **do**

$u_h = X'\omega_h / \omega_h' \omega_h$

$u_h = u_h / u_h' u_h$ vector loading associated to X

$\varepsilon_h = X' u_h$ latent variable associated to X

$v_h = Y' \varepsilon_h / \varepsilon_h' \varepsilon_h$ loading vector associated to Y

$v_h = v_h / v_h' v_h$ latent variable associated to Y

$\omega_h = Y' v_h$

End while

$c_h = X'\varepsilon / \varepsilon'\varepsilon$ partial regression of X on ε

$d_h = Y'\omega / \omega'\omega$ partial regression of Y on ω

Residues $X \leftarrow X - \varepsilon c'$

Residues $Y \leftarrow Y - \omega d'$

End for

PLS-DA consider qualitative variable as indicator for each category, and consider them like quantitative variables. From this, we extract the 2 first components for each observation.

Caret package is used for PLS-DA and plsdepot package for pls regression.

To resume five different methods are perform to generate components in link with pests and diseases

- **M1:** components which represent more than 85% of the variance, from the PCA, to clustering.
- **M2:** the 2 first components from canonical analysis with pests and diseases
- **M3:** the 2 first components from canonical analysis with IP
- **M4:** the 2 first components of PLS with pests and diseases
- **M5:** the 2 first components of PLS-DA with IP

In the following part, acronym (M1, M2, M3, M4, M5) will be used to refer to each method.

III.3. Relations between variables

To synthesize the relation between two groups of variables, Mantel test was used. To test the relation between two categorical variables, we had not enough observations to exercise the famous Chi-square test explaining the use of Fisher's exact test. Finally to assess the interaction between typologies in a representation of injury profiles, a graphic of correspondence analysis is performed.

III.3.a. Mantel test

Distance data are used to cluster plots with similar characteristics. To look at the relation between two groups of distance data, we used the Mantel test. Mantel statistics (linear or monotonic) are appropriate to test hypotheses that only concern and can only be formulated in terms of distances. It is important to make the difference that testing the relationship between two variables and rectangular data tables is not equivalent to testing the relationship between distance matrices derived from them. (Legendre & Fortin, 2010).

The null hypothesis is that the observed relationship between the two distance matrices (**X** and **Y**) could have been obtained by any random arrangement in space. In other words, there isn't relation between the two matrices of distance.

The procedure of the test is given by:

1- Calculate, the indicator's correlation, r :

$$r = \frac{1}{n-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{(x_{ij} - \bar{x})}{s_x} \frac{(y_{ij} - \bar{y})}{s_y} \qquad (3)$$

Where:
n : *number of observations*
x_{ij} : *value of the distance between the observation i and j*
$\bar{x}$: *mean of the observation* $\boldsymbol{X}$
S_x: *standard deviation of* $\boldsymbol{X}$

2- Permute object in one matrix and calculate again r

3- Repeat the procedure 1 and 2, a number of times, (in our case we repeated 999 times)

According to the value of r obtained during the repetition a probability (p-value) is given to the first correlation r obtained. Like the permutation is doing randomly, is rarely to obtain same value.

III.3.b. Between injury profiles and typologies

- ### Fisher's exact test

Fisher's exact test permits to test the null hypothesis that rows and columns categories in a contingency table are independent. This test can be performed with few observations which is interesting in our study, to the contrary to Chi-square test where if 20% of the expected cell frequencies are under 5, it cannot be valid.

Individuals, in our study plots are describing classified in two categorical variables. We would like to study at, if similar repartition is obtained between the two categorical variables. The idea of this test is to condition according to the value of margin of the contingency table.
We have two variables **X** and **Y**, with, respectively, **I** and **J** observed categories. Now we will work with the contingency table (I x J) of this two categorical variables, where a_{ij} represents the number of observation in which $x = i$ and $y = j$.
N represents the total number of observations which is equal to the sum of R_i or the sum of C_j :

$$N = \sum_{i=1}^{I} R_i = \sum_{j=1}^{J} C_j$$

We calculate P, the conditional probability of getting the actual matrix given the particular row and column sums, given by:

$$P = \frac{(R_1!\ R_2!\ \dots R_I!)\ (C_1!\ C_2!\ \dots C_J!)}{N!\ \prod_{ij} a_{ij}\,!} \qquad (4)$$

It is a multivariate generalization of the hypergeometric probability function (Weisstein, Eric W).
Then, find all possible contingency table with the row and column sums of R_i and C_j. For each one, calculate P. To obtain the p-value of the test, firstly it is necessary to order contingency table according to measure of dependence. Secondly, we add P from tables that represent equal or greater deviation from independence than the observed table, which represent the p-value of the test.

- ### Correspondence analysis

To represent the interaction between all typologies and injury profiles, graphics from correspondence analysis are analyzed. It is the case of canonical correlation analysis performing with 2 qualitative variables.

III.4. Results: Comparison of methods

The objective of his part is to compare result of typologies obtain from five methods and look at the relation with the incidence of pests and diseases, more precisely injury profiles. Table 10 exposes a summary our results.

What we can except?

M1 keep all the information of data without selection of information, it will probably be the worst method to illustrate the influence of production situation on injury profiles. M1 keeps all the information of the sub database that is why we can expect a better inertia inter groups from this method.

Mantel test permits to look at the "scale" of pests and diseases and not at the scale of injury profiles. We use this test to look at if the matrix of plot distance of components used to cluster is correlated with the matrix of plot distance of pests and diseases. It permits to study the evolution of these components created. Components used in M2 and M4 are from methods that extract components in relation with variables of pests and diseases, we expected a better correlation.

Fisher's exact test is performed between created typology and IP. M3 and M5 are expected to give a better representation between typologies and injury profiles.

Finally for each method, we perform a correspondence analysis which permits to visualize the interaction between all typologies, in a perception of injury profile representation, graphics with typologies from M3 and M5 are expected to be better.

Focus on each statistical indicator

Concerning inertia, M1 does not give a better part of inertia inter cluster explained as expected. Result of inertia is very different for each typology. We do not consider this indicator in the following analyze.

Our expectations concerning the Mantel test are confirmed for the typology of coffee plant production characteristics, components of M2 and M4 are more correlated with the incidence of pests and diseases. Climate shows a counter-example. Mantel test results with M1 show that only 3 groups of variables have a significant distance matrix correlation with pests and diseases (climate, leaves and deficiencies). The necessity to use component which select information is confirm.

p-value of Fisher's exact test with IP is under 0.05 for almost all typologies from M2, M3, M4 and M5. Exceptions are 2 times (M2 and M3) for shade canopy and one time (M2) for soil and leaves. That prove the reduction of variable has a strong impact to typologies obtain.

Fisher's exact test show that typologies obtain of M3 and M5 are more linked with IP.

Table 10 Summary of results

	Method	% inertia inter explain	Mantel Test [correlation]	Mantel Test [p-value]	Fisher's exact test [p-value]		
					IP	Yield	YL
Shade	M1	21.61	0.02	0.35	0.29	0.87	0.53
	M2	24.84	0.03	0.23	0.04	0.43	0.02
	M3	14.87	-0.01	0.63	<0.01	0.53	0.40
	M4	27.68	0.08	0.02	0.4	0.54	0.23
	M5	22.03	0.01	0.39	0.1	0.85	<0.01
Climate	M1	36.25	0.31	0.01	0.32	0.26	0.05
	M2	15.60	0.04	0.15	<0.01	0.88	0.09
	M3	16.78	0.08	0.03	<0.01	0.19	0.22
	M4	57.03	0.33	0.01	0.02	0.02	0.02
	M5	51.13	0.36	0.01	0.01	<0.01	0.86
Soil	M1	33.01	0.07	0.06	0.37	0.02	0.52
	M2	16.60	-0.03	0.8	0.39	0.74	0.05
	M3	4.78	0.02	0.21	<0.01	0.69	0.55
	M4	37.42	0.05	0.05	<0.01	0.17	0.64
	M5	42.54	0.13	0.01	0.04	0.38	0.10
Management	M1	8.08	0.07	0.06	0.28	0.93	0.12
	M2	25.77	0.17	0.01	<0.01	0.01	<0.01
	M3	17.37	0.25	0.01	<0.01	0.63	0.01
	M4	22.16	0.22	0.01	<0.01	0.10	0.01
	M5	20.93	0.19	0.01	<0.01	0.58	0.90
CPPC	M1	29.62	0.03	0.28	0.04	0.01	0.19
	M2	3.94	0.11	0.04	<0.01	0.09	0.12
	M3	14.06	0.06	0.06	<0.01	0.11	0.02
	M4	17.92	0.10	0.03	<0.01	<0.01	<0.01
	M5	26.55	0.00	0.47	<0.01	<0.01	0.16
Leaves	M1	4.96	0.18	0.01	0.02	0.29	0.26
	M2	56.50	0.11	0.02	0.95	0.95	0.36
	M3	3.00	0.10	0.03	<0.01	0.98	0.37
	M4	12.04	0.20	0.01	<0.01	0.04	0.02
	M5	11.66	0.19	0.01	<0.01	0.08	0.01
Deficiencies	M1	26.21	0.10	0.03	<0.01	0.95	<0.01
	M2	33.40	0.12	0.03	<0.01	0.22	<0.01
	M3	26.74	0.18	0.01	<0.01	0.57	<0.01
	M4	27.16	0.16	0.01	<0.01	0.47	<0.01
	M5	26.69	0.18	0.01	<0.01	0.51	<0.01

Focus with illustrative variables

All the different typologies of deficiencies are strongly significant with categories of yield losses but not with yield. According to the method used to build component, relation with yield losses, and yield can be very different for the same group of variables and seems to follow random trends, as result of Fisher's exact test show.

Focus of each typology

Shade canopy is significant with IP (Fisher's exact test) only in M2 and M3 when Mantel test doesn't find a correlation of matrix distance. We are mistrusting once a relation between shade canopy and IP, in these two cases.

Climate and soil typologies from M4 and M5 have stronger relation with IP and pests and diseases. Management and deficiencies are completely significant with M2, M3, M4 and M5.

Typologies of leaves have a rarely behavior when we used M2.

Concerning typologies of coffee plant production characteristics, Mantel test show that matrix distance of components from M3 and M5 aren't correlated but typologies are linked with IP.

Conclusion

Typologies from M3 and M5 are the most interesting. We have separated this methods looking at the representation of correspondence analysis which show the interaction of all. Typologies are more regrouping and clear in graph from M5 (Figure 18). Biological meaning and interpretation was optimum in this representation. It shows a main difference of the representation of yield as illustrative variables. Typologies from M5; yield seems completely expressed by axis 2 (Figure 13). In the graph, from typologies of M3, yield has not a good representation. Finally typologies obtained by M5 are the optimum to illustrate influence of production situation and coffee plant characteristics on injury profiles.

Conclusion

Assessment of coffee yield looses is complex. Qualitative approach permitted to represent in a more easy way interactions between factors which influence injury profiles and yield. However, each step of the clustering needed to be investigated to maximize the result. Our research had showed that the part of selection of variables, more precisely of information, is the principal one, which permits to obtain optimum results.

In our case, to explain the influence of production situation on injury profiles, the use of PLS-DA before clustering was the optimum. This method permitted to eliminate useless information to explain injury profiles and obtain more significant typologies linked with injury profiles.

To follow, it will be interesting to not be restricted by linear method. PLS Spline can be a possibility as doing to explain key factors for the development of American leaf spot (Avelino, 2007).

Professional and personal balance

Professional

This experience permits me to apply and consolidated many methods learnt at school but I have also learnt new statistic methods. Once I have realized that statistics are a very young science which evolves very quickly. During all my working life, I will need to be up to date with the evolution of statistics methods. I have improved both my English and Spanish competences. The fact to be with non statistics specialist required from me a pedagogical speaking which I always consider very important. Finally, I have learnt a lot about coffee, from the plant to the roast but I still don't like the taste of it!

Personnel

My opinion about work in a research centre changed. I thought research is a solitary job. I was very surprise, to take advantage of a dynamic of a work team. I have realized that, the objective of my work, i.e. to bring solutions to coffee farmers, was a real motivation for me.
Six months out of France was a very good experience to learn from others cultures and to think about my future plan. I was in a campus, living in community where I learn a lot about myself and the others.

Both of personnel and professional point of view cannot expect better experience.

References

Avelino, J., Ten Hoopen, G. M., & DeClerck, F. (2011). Ecological mechanisms for pest and disease control in coffee and cacao agroecosystems of the neotropics. Ecosystem Services from Agriculture and Agroforestry Measurement and Payment. London: Earthscan, 91-117.

Avelino, J., Cabut, S., Barboza, B., Barquero, M., Alfaro, R., Esquivel, C., ... & Cilas, C. (2007). Topography and crop management are key factors for the development of American leaf spot epidemics on coffee in Costa Rica. Phytopathology, 97(12), 1532-1542.

Baker P. 2014. The 'Big Rust': an update on the coffee leaf rust situation. Disease & Pest Control. C&CI. Pp. 37-39

CIRAD, www.cirad.fr , 30/11/2009

Cooke, B. M., Jones, D. G., & Kaye, B. (Eds.). (2006). The epidemiology of plant diseases (Vol. 2). Dordrecht: Springer.

Hotelling, H. (1936). Relations between two sets of variates. Biometrika, 321-377.

Legendre, P., & FORTIN, M. J. (2010). Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. Molecular Ecology Resources, 10(5), 831-844.

PROMECAFE. 2013. Boletín No. 135. www.promecafe.org ; www.iica.int

Savary, S., Willocquet, L., Elazegui, F. A., Teng, P. S., Van Du, P., Zhu, D., ... & Srivastava, R. K. (2000). Rice pest constraints in tropical Asia: characterization of injury profiles in relation to production situations. Plant Disease, 84(3), 341-356.

Savary, S., Madden, L. V., Zadoks, J. C., & Klein-Gebbinck, H. W. (1995). Use of categorical information and correspondence analysis in plant disease epidemiology. Advances in botanical research, 21, 214-239.

Simko, I., & Piepho, H. P. (2012). The area under the disease progress stairs: calculation, advantage, and application. Phytopathology, 102(4), 381-389.

M. Tenenhaus, La régression PLS : théorie et applications, Technip,1998.

Weisstein, Eric W. "Fisher's Exact Test." From MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/FishersExactTest.html

Wikistat, Composantes principales et régressions PLS parcimonieuses

Wikistat, Analyse canonique des correlations (ACC)

H. Wold, Multivariate analysis, Academic Press, 1966.

Zadoks, J. C., & Schein, R. D. (1979). Epidemiology and plant disease management. Epidemiology and plant disease management.

Appendices

Annex A. A specific experiment to estimate primary and secondary losses of coffee: COLOSSE

In the context of the doctoral research a specific set of plot was needed in order to quantify crop losses. The CoLosses experiment was designed with the main objective to estimate primary and secondary losses of coffee. This is an experiment with duration of three years, where six different treatments consisting in a different sequences of fungicide application. It enables the estimation of attainable yields, actual yields and primary and secondary losses. I have analyse data from this experimentation, but in this present report I don't expose the work done. We have close descriptive variables such as the incidence of pests and diseases, the composition of leaves and soil, and the coffee plants production characteristics. As part of an agronomy seminary in Montpellier in September 2015, I have contribute to a scientific paper (see *0, Modeling attainable and yield losses due to pests and diseases to compare performance of coffee farming systems).*

Data from these two experimentations permit to fulfill at issues raised by the doctoral.

Annex B. Modeling attainable and yield losses due to pests and diseases to compare performance of coffee farming systems

AGRO 2015

5th International Symposium for Farming Systems Design 7-10 September 2015, Montpellier, France

Modeling attainable yield and yield losses due to pests and diseases to compare performances of coffee farming systems

Rolando Cerda [*±1,2], **Jacques Avelino** [2,3,4], **Esther Lechevallier** [1,5], **Louise Krolczyk** [3,6], **Charlie Mathiot** [2,6], **Eugénie Clément** [3,7], **Christian Gary** [8], **Philippe Tixier** [2,9] **& Clémentine Allinne** [1,2]

[1] CIRAD, UMR System, SupAgro Montpellier bat 27, 2 place Viala, 34060 Montpellier, France
[2] CATIE, Program of Sustainable Agriculture and Agroforestry, 7170 Turrialba, Costa Rica
[3] CIRAD, UR Bioagresseurs, TA A-106 – Avenue Agropolis, 34398, Montpellier, France
[4] IICA-PROMECAFE, AP 55, 2200 Coronado, San José, Costa Rica
[5] ENSAT-INP, F-31326 Castanet Tolosan, France
[6] AgroParisTech, F-75231 Paris, France
[7] INSA Toulouse 135, Avenue de Rangueil - 31077 Toulouse– France
[8] INRA, UMR System, SupAgro Montpellier bat 27, 2 place Viala, 34060 Montpellier, France
[9] CIRAD, UPR GECO, TA B-26 / PS4 - Boulevard de la Lironde - 34398 Montpellier, France

[*] Speaker
[±] Corresponding author: rcerda@catie.ac.cr

1 Introduction

The regulation service of pests and diseases (P&D) can be assessed in terms of avoided crop losses (Avelino *et al.*, 2011). This approach is necessary to compare the performance of agroecosystems with different sets of P&D (injury profiles). For perennial crops, such as coffee, the assessment of yield losses is difficult due to the biennial production cycle, the complexity of agroecosystems where these crops are grown, and the existence of primary losses (in the current year) and secondary losses (losses in the following years due to the physiological damages caused by P&D) (Zadoks & Schein, 1979). Through this work, we contribute to: i) a better understanding of the impact of P&D on primary and secondary coffee yield losses; and ii) build a conceptual model to identify the main factors that determine coffee yield losses.

2 Materials and Methods

Trial: coffee at full sun exposure (5000 plants ha^{-1}) where six treatments with different sequences of fungicide applications are compared, with a duration of three years (Fig 1). In each of four replicates (40 m^2), six coffee plants were marked for measurements: fruiting nodes, fruits per node, dead branches, P&D incidences and severity, and yields.

Studied variables: standardized area under the disease progress curve (sAUDPC) and severity of P&D, dead branches, and yield losses. Depending on the treatment, at the end of each year we can obtain attainable yields (Yatt) and actual yields (Yac), whose differences represent primary or secondary losses (Fig 1). *Yield loss (%) = ((Yatt – Yac)/Yatt)*100*

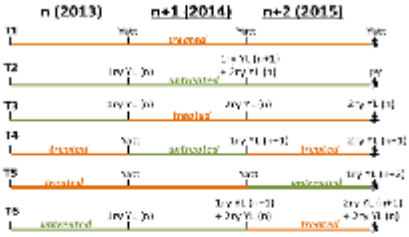
Statistical analysis: i) sAUDPC, severity, dead branches and yields were compared among treatments trough analyses of variance, with fruit load of the previous year and soil acidity as covariates (they characterize the production conditions of each plot), and LSD (Fisher) with $p<0.05$; ii) correlations (Spearman) between measured and studied variables of 2013 and 2014 to support the relationships presented in the conceptual model.

3 Results - Discussion

Diseases and yield losses. We observed significant effects of the treatments on several diseases in both years (Table 1). In the first year (2013), no difference of yield between treatments was found and therefore no yield loss was calculated. In the second year (2014), significant differences between treatments were observed for sAUDPC of coffee leaf rust and dead branches; and the yield of the treatments conducting to attainable yield (T1&T5) was different to others, making possible to calculate yield losses. The negative impacts of abandon fungicide application from one year to another (T4) can be worse than no fungicide application in two consecutive years (T2&T6), reflected by higher dead branches and similar yields. Both primary and secondary losses were high, showing a severe impact of diseases (Table 1). Although we applied fungicides, there was presence of diseases, which indicates that yield losses could be even higher.

Correlations and conceptual model. Several significant correlations were found between variables from one year to another and within the same year. The fruit load in 2013 influenced negatively the fruit load in 2014; higher yield components (fruits per node, fruiting nodes, fruit load) caused higher sAUDPC and severities of diseases in most cases; dead branches had positive correlations specially with severity of diseases, and dead branches of 2013 influenced negatively the yield components in 2014 (Fig. 2). These findings indicate that physiological aspects and impacts of diseases lead to yield losses in a given year and in next years. Based on that, we present a conceptual model, showing how different factors can influence the components of attainable yield, and how this last one can be reduced to actual yield, due to primary and secondary yield losses of coffee (Fig. 3).

AGRO 2015 — 5th International Symposium for Farming Systems Design — 7-10 September 2015, Montpellier, France



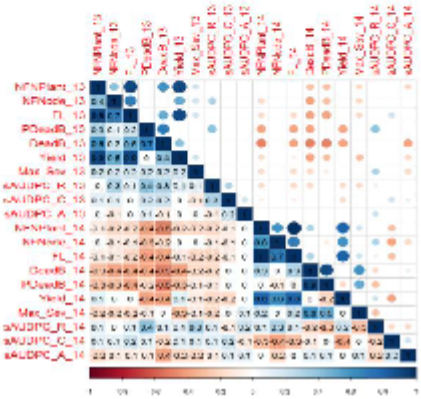
Yatt: attainable yield; YL: yield losses; PY: primitive yield

Fig. 1. Treatments for the assessment of yield losses.

Table 1. Effects of treatments on P&D and yield losses.

	Year 2013		Year 2014			
Variable	T1&T4&T5	T2&T3&T6	T1&T5	T2&T6	T3	T4
sAUDPC_R	24 a	34 b	34 a	41 b	40 ab	31 a
sAUDPC_C	24 a	29 b	22 a	24 a	24 a	27 a
sAUDPC_A	6 a	5 a	8 a	4 b	6 ab	8 a
Max_Sev	3.1 a	3.2 a	3.5 a	3.4 a	3.2 a	3.6 a
DeadB	37 a	54 b	28 ab	29 ab	13 a	44 b
Yield	2077 a	2267 a	3397 a	1823 b	2125 b	1931 b
Estimated Primary Losses = ((3397-1931)/3397)*100 = 43%						
Estimated Secondary Losses = ((3397-2125)/3397)*100 = 37%						
Estimated Primary and Secondary Losses = ((3397-1823)/3397)*100 = 46%						

See interpretation of Fig. 2. for the meaning of variables and their units



Interpretation. Bellow the diagonal there are the Spearman coefficients; above the diagonal: the darker and bigger the circles, the stronger the correlation (symmetric matrix). Blue color indicates positive correlation; orange color indicates negative correlation. Cells with white color indicate that there is no significant correlation.

NFNPlant: Number of fruiting nodes per plant
NFNode: Number of fruits per node
FL: Fruit load (=NFNPlant x NFNode)
PDeadB: Percentage of dead branches
DeadB: Number of dead branches
Yield: Coffee yield (grams of coffee cherries per plant)
Max_Sev: Maximum of severity of diseases in leaves (scale 0-6)
sAUDPC_R: standardized Area Under the Disease Progress Curve of coffee leaf rust (*Hemileia vastatrix*) (% day^{-1})
sAUDPC_C: of cercospora (*Cercospora coffeicola*) (% day^{-1})
sAUDPC_A: of anthracnoses (*Colletotrichum* spp.) (% day^{-1})
Numbers 13 and 14 represents the variables in 2013 and 2014, respectively

Note: we constructed the accumulated Areas Under the Disease Progress Curve, but, since there were differences in the total time of incidence measurements between 2013 and 2014, we standardized them per day.

Fig. 2. Matrix of correlations (Spearman) among variables that determine yields.

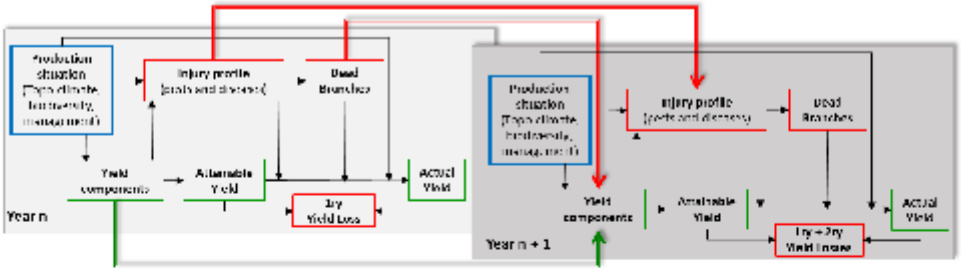


Fig. 3. Conceptual model to assess primary and secondary yield losses of coffee.

4 Conclusions

The negative impacts of diseases in coffee are not limited only to the year where they have developed. The high estimated secondary losses indicate that economic and technical measures to help coffee farmers to face phytosanitary issues (as coffee rust epidemic in 2012-13, Avelino et al., 2015) need to be continued on several years.

The proposed conceptual model shows the main factors that should be taken into account to assess primary and secondary yield losses of coffee. Based on this model, statistical models could be developed (to be finalized in 2015) to estimate attainable yields and yield losses, in order to assess the performance of different coffee farming systems.

Acknowledgments. We thank the INRA Smatch Metaprogramme that funds this experiment through the DAMAGE action.

References

Avelino, J., ten Hoopen, M., & DeClerck, F. (2011). Ecological mechanisms for pest and disease control in coffee and cacao agroecosystems of the Neotropics. In: DeClerck, F., Le Coq, J.F., Beer, J. (Eds.), Ecosystem services from agriculture and agroforestry measurement and payment. Earthscan, 91-117. London.

Avelino J, Cristancho M, Georgiou S, Imbach P, Aguilar L, Bornemann G, *et al.* (2015). The coffee rust crises in Colombia and Central America (2008-2013): impacts, plausible causes and proposed solutions. Food Security. 2015;DOI 10.1007/s12571-015-0446-9.

Zadoks, J.C. & Schein, R.D. (1979). Epidemiology and plant disease management. New York

Annex C. Ecosystem services provided by coffee agroecosystems across a range of topo-climatic conditions and management strategies

AGRO 2015

5th International Symposium for Farming Systems Design 7-10 September 2015, Montpellier, France

Ecosystem services provided by coffee agroecosystems across a range of topo-climatic conditions and management strategies

Rolando Cerda *±[1,2], **Clémentine Allinne**[1,2], **Louise Krolczyk**[3,4], **Charlie Mathiot**[2,3], **Eugénie Clément**[4,5], **Celia A. Harvey**[6], **Jean-Noel Aubertot**[7], **Philippe Tixier**[2,8], **Christian Gary**[9] & **Jacques Avelino**[2,4,10]

[1] CIRAD, UMR System, SupAgro Montpellier bat 27, 2 place Viala, 34060 Montpellier, France
[2] CATIE, Program of Sustainable Agriculture and Agroforestry, 7170 Turrialba, Costa Rica
[3] AgroParisTech, F-75231 Paris, France
[4] CIRAD, UR Bioagresseurs, TA A-106 – Avenue Agropolis, 34398, Montpellier, France
[5] INSA Toulouse 135, Avenue de Rangueil - 31077 Toulouse– France
[6] Conservation International, 2011 Crystal Drive Suite 500, Arlington, VA, USA
[7] INRA-Centre de Recherches de Toulouse, 31326 Castanet Tolosan Cedex, Toulouse, France
[8] CIRAD, UPR GECO, TA B-26 / PS4 - Boulevard de la Lironde - 34398 Montpellier, France
[9] INRA, UMR System, SupAgro Montpellier bat 27, 2 place Viala, 34060 Montpellier, France
[10] IICA-PROMECAFE, AP 55, 2200 Coronado, San José, Costa Rica

* Speaker
± Corresponding author: rcerda@catie.ac.cr

1 Introduction

There is an urgent need to ensure that farming systems not only provide high yields, but also the provision of ecosystem services (ES) on which agriculture and farmer households depend. We compared the provision of four ES in different types of coffee agroecosystems: i) regulation of pests and diseases (P&D); ii) provision of agroforestry products (coffee, bananas, fruits, timber); iii) maintenance of soil fertility; and iv) carbon sequestration. We provide key insights on how coffee agroecosystems could be most effectively managed to ensure the continued provision of ES.

2 Materials and Methods

We established a coffee research network (69 coffee plots) in Turrialba, Costa Rica for two years of field measurements (2014-2015). Coffee agroecosystems were selected according to the combination of three factors: i) Altitude: low (<850m.a.s.l.) and high (>850m.a.s.l.); ii) Shade: full sun coffee, simple shade (dominated by *Erythrina poeppigiana*) and diversified shade (musaceas, service trees, fruit trees and timber trees); iii) Management: low (few cropping practices and low inputs) and high (many cropping practices and high inputs). We calculated the areas under the disease progress curve (AUDPC) of P&D, registered the severity, and counted the number of dead branches. We also assessed the effectiveness of coffee agroecosystem in regulating P&D by estimating the coffee yield losses (=attainable yields minus actual yields; estimated by modelling). Yields, costs and incomes of agroforestry products were used to calculate economic indicators and to assess their overall contribution to farmer households (Cerda *et al.*, 2014). Soil fertility was determined by laboratory analysis. Above-ground biomass carbon was estimated with the use of allometric equations.
Statistical analysis: analyses of variance using general linear mixed models and the test LSD (Fisher) with $p < 0.05$ to compare the effect of the three factors (altitude, shade and management) and their interactions on the provision of ES.

3 Results - Discussion

The interaction of shade and management was the most important for explaining the regulation of P&D. Coffee leaf rust (*Hemileia vastatrix*), the severity of P&D attacks and the number of dead branches were higher in full sun coffee plantations with high management as well as in coffee under diversified shade with low management; indicating that none of those extremes are good for avoiding P&D. Coffee under diversified shade with high management showed fewer P&D impacts, suggesting that complex agroforestry systems can contribute to the regulation of P&D (Fig. 1).

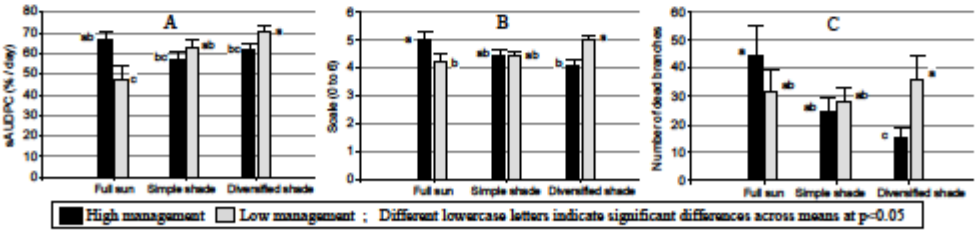


Fig. 1. A) Coffee leaf rust, B) Maximum severity, and C) Dead branches, according to shade and management.

Attainable yields and actual yields of coffee were similar among coffee in full sun with high management and coffee agroforestry systems, but coffee under diversified shade with high management tended to have the lowest yield losses (Fig 2). These results reinforce the idea that diversified shade systems can help to regulate P&D (Fig 1). These findings become important knowledge for the development of agroecosystems that are capable of balancing high yields and reduce the impacts of P&D (Avelino *et al*., 2011).

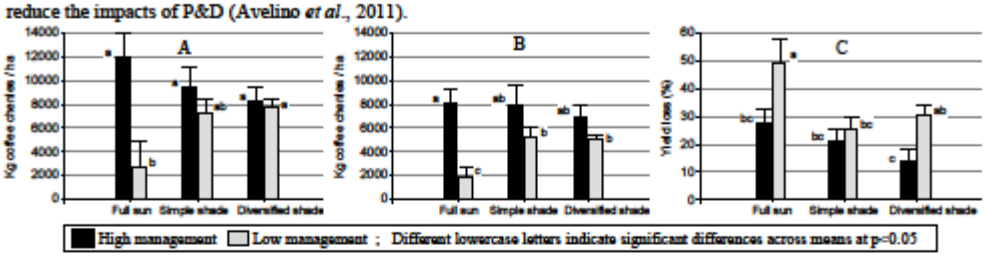
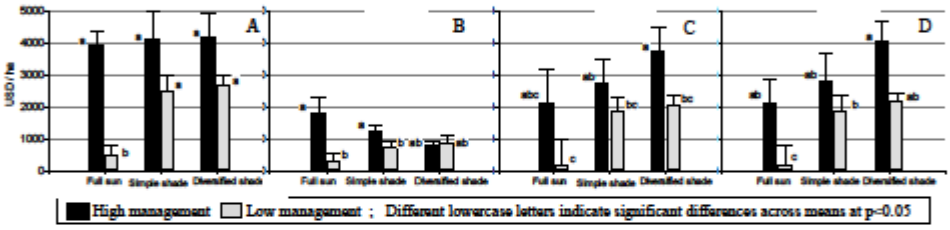


Fig. 2. A) Attainable yields, B) Actual yields, and C) Yield losses of coffee, according to shade and management.

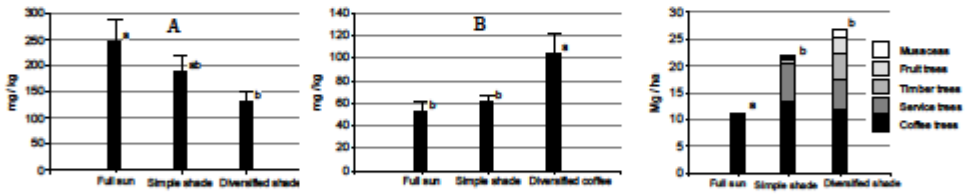
Only high coffee yields would not always be the best for farmer households. Cash costs of coffee in full sun were high and therefore its cash flow tended to be lower than in agroforestry systems. Besides, the agroforestry products of shade canopies increased even more the cash flow, the domestic consumption, and therefore, the family benefit (Fig 3).



Cash flow = Gross incomes – Cash costs; Family benefit = Cash Flow + value of domestic consumption of agroforestry products

Fig. 3. A) Gross income, B) Cash costs, C) Cash Flow, and D) Family Benefit, according to shade and management.

The sole effect of shade was the most important on soil and carbon. Most elements of soil fertility were better in coffee under diversified shade, as in the case of acidity and potassium (Fig 4), two key indicators of soil quality. Finally, agroforestry systems had at least double the above-ground carbon compared to coffee in full sun (Fig 5).



Different lowercase letters indicate significant differences across means at p<0.05

Fig 4. A) Acidity, and B) Potassium in soil. **Fig. 5.** Above-ground biomass carbon.

Conclusions

The provision of ES varies across different types of coffee agroecosystems. The best ES are provided by coffee agroforestry systems. Coffee farming systems should be designed with the inclusion of productive shade canopies and managed with constant cropping practices, trying to reduce as much as possible the cash costs for a higher family benefit; being also the best alternative to reduce yield losses.

Acknowledgements. This research is part of CASCADE project "Ecosystem-based Adaptation for Smallholder Subsistence and Coffee Farming Communities in Central America", funded by the International Climate Initiative (ICI).

References

Avelino, J., ten Hoopen, M., & DeClerck, F. (2011). Ecological mechanisms for pest and disease control in coffee and cacao agroecosystems of the Neotropics. In: DeClerck, F., Le Coq, J.F., Beer, J. (Eds.), Ecosystem services from agriculture and agroforestry measurement and payment. Earthscan, 91-117. London.

Cerda, R., Deheuvels, O., Calvache, D., Niehaus, L., Saenz, Y., Kent, J., Vilchez, S., Villota, A., Martinez, C. & Somarriba, E. (2014). Contribution of cocoa agroforestry systems to family income and domestic consumption: looking toward intensification. *Agroforestry Systems*, 88, 957-981.

Annex D. Figures of the clustering

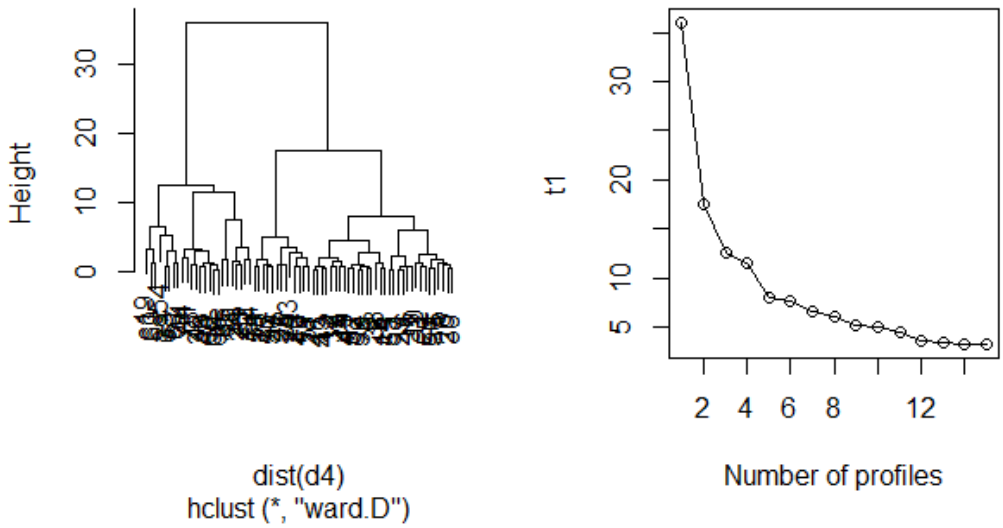


Figure 17 On the left part, cluster dendrogram, on the right, representation of the size of stems

Annex E. Description of variables

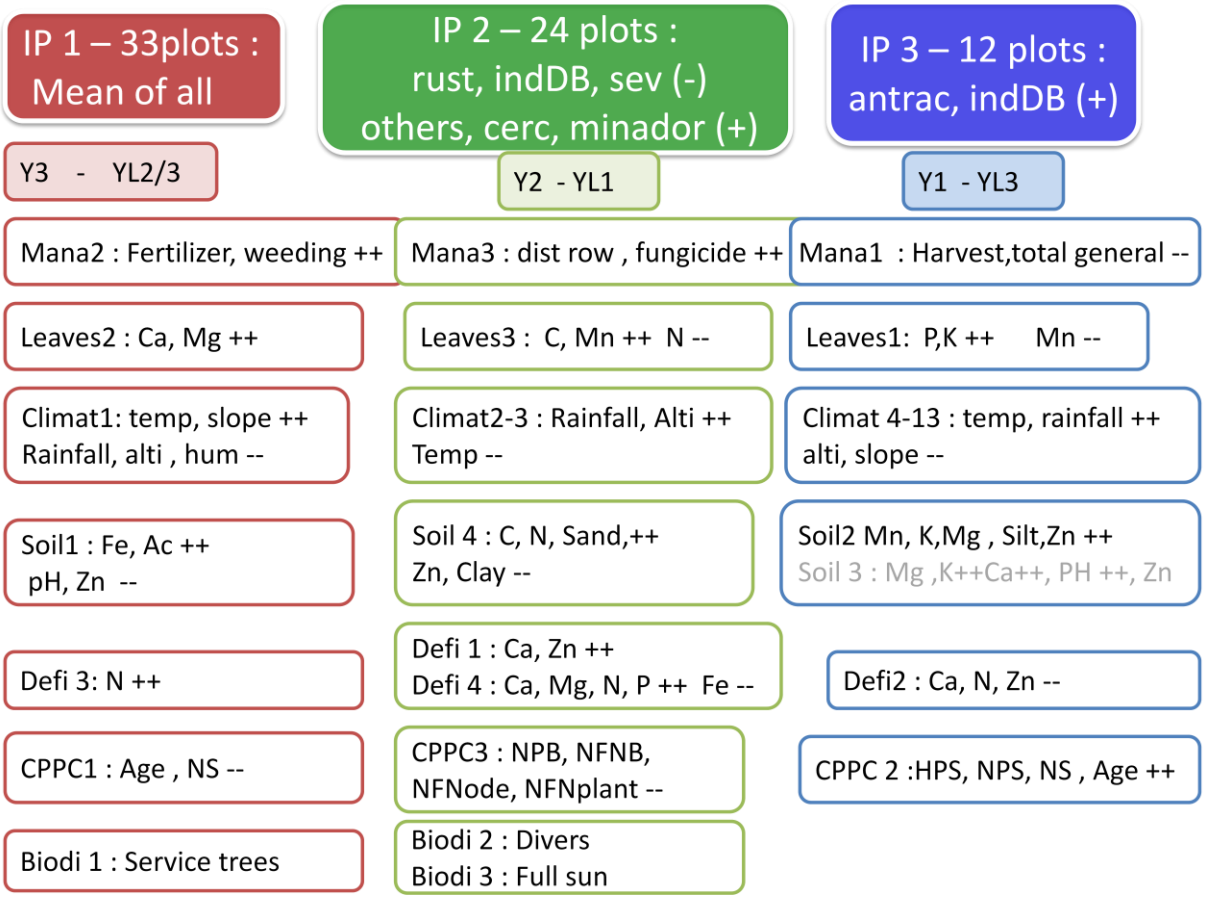
	Variables	Meaning	Units	Mean	Standard deviation
	Plot		-		
	District		-		
	Community		-		
	Owner		-		
	IndComunity		-		
CPPC	Age		years	17,00	9,50
	HPS	Height of productive stem	cm	205,99	28,52
	NPB	Number of productive branches	number	3,41	1,30
	NFNB	Number of fruiting nodes per branch	number	31,34	10,02
	NS	Number of stems per plant	number	6,44	2,18
	NPS	Number of productive stems per plant	number plant^{-1}	2,89	1,19
	NFNode	Number of fruits per node	number node^{-1}	3,45	1,45
	NFN_Plant	Number of fruiting nodes per plant	number plant^{-1}	508,57	299,42
	NFN_ha	Number of fruiting nodes per hectare	number ha^{-1}	2513,61	1421,31
	Fruit_load	Fruit load per plant	number coffee cherries plant^{-1}	2122,88	1831,90
	DeadB	Number of dead branches per plant	number plant^{-1}	27,48	25,68
Modelling	Yield_plant	Yield of coffee harvested in coffee plots	g coffee cherries plant^{-1}	907,31	618,14
	Ac_Yield_plant	Actual yield of coffee per plant (estimated by modelling)	g coffee cherries plant^{-1}	1315,08	896,38
	Ac_Yield_ha	Actual yield of coffee per hectare (estimated by modelling)	kg coffee cherries ha^{-1}	6406,65	4131,80
	Att_Yield_plant	Attainable yield of coffee per plant (estimated by modelling)	g coffee cherries plant^{-1}	1704,17	1075,74
	Att_Yield_ha	Attainable yield of coffee per hectare (estimated by modelling)	kg coffee cherries ha^{-1}	8433,60	5037,54
	YL_plant	Yield losses of coffee per plant	g coffee cherries plant^{-1}	389,09	429,01
	YL_ha	Yield losses of coffee per hectare	kg coffee cherries ha^{-1}	2026,95	2353,46
	YL_per	Percentage of yield losses of coffee	%	23,13	18,18
Soil	S_pH	pH in soil	-	4,76	0,40
	S_Ac	Acidity in soil	mg kg^{-1}	165,41	123,31
	S_Ca	Calcium in soil	mg kg^{-1}	930,38	1204,86
	S_C	Carbon in soil	%	4,29	2,63
	S_Cu	Copper in soil	mg kg^{-1}	18,42	6,85
	S_Fe	Iron in soil	mg kg^{-1}	257,24	133,63
	S_Mg	Magnesium in soil	mg kg^{-1}	159,78	193,89
	S_Mn	Manganesium in soil	mg kg^{-1}	31,28	27,48
	S_P	Phosphorus in soil	mg kg^{-1}	5,59	5,14
	S_K	Potasium in soil	mg kg^{-1}	67,68	58,99

	S_N	Nitrogen in soil	%	0,38	0,21
	S_Zn	Zinc in soil	mg kg^{-1}	2,57	1,41
	Sand	Percentage of sand in soil	%	55,65	14,83
	Silt	Percentage of silt in soil	%	21,93	7,11
	Clay	percentage of clay in soil	%	22,42	12,50
Climate	Altitude_quanti	Altitude of the plot in meters above sea level	m.a.s.l	877,09	125,63
	Slope	Slope of the coffee plot	%	26,01	18,07
	Orientation_slope_WE	Orientation of the slope with respect West-East	-	0,08	0,66
	Orientation_slope_NS	Orientation of the slope with respect North-South	-	-0,11	0,70
	Rainfall_jul-aug	Rainfall during July and August (rainy season) around coffee plot	mm	832,99	187,79
	Rainfall_sep-oct	Rainfall during September and october (less rainy season) around coffee plot	mm	636,67	91,44
	Rainfall_nov-jan	Rainfall during November and January (highest rainy season) around coffee plot	mm	1267,65	280,35
	T_dec-jan	Temperature during December and January (highest rainy season) around coffee plot	Celsius degrees	20,57	0,65
	T_feb-mar	Temperature during February and March (dry season) around coffee plot	Celsius degrees	20,86	0,81
	Hum_dec-jan	Relative humidity during December and January (highest rainy season) around coffee plot	%	92,22	2,91
	Hum_feb-mar	Relative humidity during February and March (dry season) around coffee plot	%	89,18	1,19
Deficiency	Max_def_Ca	Maximum percentage of plants with calcium deficience	%	19,71	26,51
	Max_def_Fe	Maximum percentage of plants with iron deficience	%	83,51	18,32
	Max_def_Mg	Maximum percentage of plants with magnesium deficience	%	4,81	10,15
	Max_def_N	Maximum percentage of plants with nitrogen deficience	%	32,17	23,67
	Max_def_P	Maximum percentage of plants with phosphorus deficience	%	3,78	8,23
	Max_def_K	Maximum percentage of plants with potassium deficience	%	1,13	3,31
	Max_def_Zn	Maximum percentage of plants with zinc deficience	%	10,10	15,69
	Max_def_B	Maximum percentage of plants with boron deficience	%	0,05	0,32
	Max_def_S	Maximum percentage of plants with sulfur deficience	%	0,01	0,10
	Max_def_Cu	Maximum percentage of plants with copper deficience	%	0,00	0,00
	Max_def_Mo	Maximum percentage of plants with molybdenum deficience	%	0,01	0,11
Leaves	L_Ca	Calcium in leaves	%	1,20	0,28
	L_C	Carbon in leaves	%	48,30	1,45
	L_Cu	Copper in leaves	mg kg^{-1}	21,09	10,26
	L_Fe	Iron in leaves	mg kg^{-1}	111,32	70,23

	L_Mg	Magnesium in leaves	mg kg^{-1}	0,47	0,17
	L_Mn	Manganesium in leaves	%	228,78	149,09
	L_P	Phosphorus in leaves	%	0,19	0,03
	L_K	Potasium in leaves	%	1,66	0,73
	L_N	Nitrogen in leaves	%	3,42	0,32
	L_Zn	Zinc in leaves	mg kg^{-1}	10,00	4,18
Pests and diseases	AUDPC_rust	Area under the disease progress curve of coffee leaf rust (Hemileia vastarix), no cumulative	% day	9990,77	2248,17
	sAUDPC_rust	Standardized Area under the disease progress curve of coffee leaf rust (Hemileia vastarix), no cumulative	% day day^{-1}	60,83	14,33
	AUDPC_cerc	Area under the disease progress curve of brown eye spot (Cercospora coffeicola) of coffee , no cumulative	% day	841,71	612,93
	sAUDPC_cerc	Standardized Area under the disease progress curve of brown eye spot (Cercospora coffeicola) of coffee, no cumulative	% day day^{-1}	4,88	3,29
	AUDPC_mi	Area under the disease progress curve of leaf miner (Leucoptera coffeella) of coffee , no cumulative	% day	195,90	176,87
	sAUDPC_mi	Standardized Area under the disease progress curve of leaf miner (Leucoptera coffeella) of coffee, no cumulative	% day day^{-1}	1,14	0,94
	AUDPC_ant	Area under the disease progress curve of anthracnoses (Colletotrichum spp.) of coffee , no cumulative	% day	261,74	212,69
	sAUDPC_ant	Standardized Area under the disease progress curve of anthracnoses (Colletotrichum spp.) of coffee, no cumulative	% day day^{-1}	1,56	1,22
	AUDPC_oth	Area under the disease progress curve of others diseases of coffee , no cumulative	% day	1557,37	820,88
	sAUDPC_oth	Standardized Area under the disease progress curve of other diseases of coffee, no cumulative	% day day^{-1}	9,29	4,59
	max_rust	Maximum percentage of leaf rust (Hemileia vastarix) incidence	%	77,90	15,04
	max_cerc	Maximum percentage of brown eye spot (Cercospora coffeicola) incidence	%	9,05	5,23
	max_als	Maximum percentage of american leaf spot (Mycena citricolor) incidence	%	0,22	0,99
	max_mi	Maximum percentage of leaf miner (Leucoptera coffeella) incidence	%	3,20	1,96
	max_ant	Maximum percentage of anthracnoses (Colletotrichum spp.) incidence	%	3,72	3,06
	max_oth	Maximum percentage of others diseases incidence	%	16,62	7,70
	max_leca	Maximum percentage of lecanicilium (Lecanicilium spp.) incidence	%	4,90	9,60
	Max_broca	Maximum percentage of coffee berry borer (Hypothenemus hampei) incidence	%	2,99	3,23
	Max_IndDB	Maximum percentage of the die back index	%	45,80	13,45

Shade canopy	Max_Sev	Maximum of severity (Scale from 0 to 6 ; the higher number, the higher severity)	-	4,52	0,88
	D_FruitTrees	Density of fruit trees per hectare	Individuals ha^{-1}	29,28	59,79
	D_TimberTrees	Density of timber trees per hectare	Individuals ha^{-1}	33,16	54,18
	D_MusaceasTrees	Density of musaceas trees per hectare	Individuals ha^{-1}	206,78	286,62
	D_ServiceTrees	Density of service trees	Individuals ha^{-1}	176,28	153,81
	A_FruitTrees	Trunk basal area of fruit trees	m^2 ha^{-1}	0,43	0,81
	A_TimberTrees	Trunk basal area of timber trees	m^2 ha^{-1}	1,62	2,63
	A_MusaceasTrees	Trunk basal area of musaceas trees	m^2 ha^{-1}	3,35	4,84
	A_ServiceTrees	Trunk basal area of service trees	m^2 ha^{-1}	5,28	5,10
	Sp_FruitTrees	Number of species of fruit trees per 1000 m^2	Number $1000m^{-2}$	1,12	1,41
	Sp_TimberTrees	Number of species of timber trees per 1000 m^2	Number $1000m^{-2}$	0,72	0,70
	Sp_MusaceasTrees	Number of species of musaceas trees per 1000 m^2	Number $1000m^{-2}$	0,91	0,72
	Sp_ServiceTrees	Number of species of service trees per 1000 m^2	Number $1000m^{-2}$	1,67	1,34
	H_canopy	Shannon index for plants in the shade canopy of coffee plots	-	0,76	0,55
	D_canopy	Simpson index for plants in the shade canopy of coffee plots	-	0,38	0,28
	Shade	Percentage of the shade cover provided by the shade canopy of coffee plots	%	17,77	14,12
Management	Dist_rows	Distance between coffee rows	cm	172,97	23,52
	Dist_plants	Distance between coffee plants	cm	119,70	17,86
	Weeding	Number of hand weedings per year	Number $year^{-1}$	1,29	1,28
	Harvest_coffee	Number of harvests of coffee berries per year	Number $year^{-1}$	10,33	2,39
	Fertilizer	Number of fertilizer applications per year	Number $year^{-1}$	1,41	0,90
	Fungicide	Number of fungicide applications per year	Number $year^{-1}$	2,10	1,78
	Herbicide	Number of herbicide applications per year	Number $year^{-1}$	1,80	1,31
	Hard_weeding	Number of hand hard weedings per year	Number $year^{-1}$	0,35	0,68
	Pruning_coffee	Number of prunings of coffee trees per year	Number $year^{-1}$	0,80	0,41
	Diversity_practices	Count of how many different practices are applied per year	Number $year^{-1}$	5,22	1,11
	Total_general	Sum of all the number of practices applied per year	Number $year^{-1}$	18,09	4,28
	Orientation_rows_WE	Orientation of the coffee rows with respect West-East	-	0,07	0,73
	Orientation_rows_NS	Orientation of the coffee rows with respect North-South	-	-0,09	0,68

Annex F. Brief presentation of each typologies



Annex G. Graphic of correspondence analysis of M3

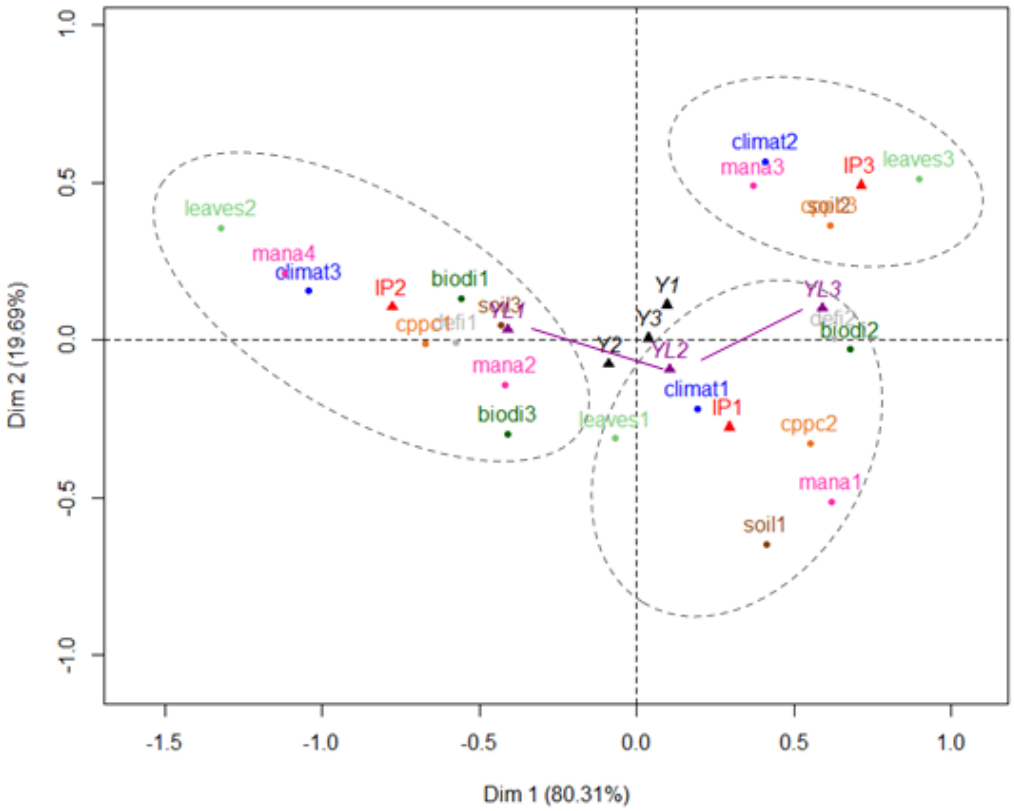


Figure 18 Correspondence analysis of injury profiles with typologies from M3. Yield and yield losses has illustrative variables