1. Executive summary

The International Center for Tropical Agriculture (CIAT) was commissioned by the directorate for Intellectual Property of the National Federation of Colombian Coffee Growers (FNC) to lead a second phase, after a pilot study had been conducted successfully, to support the implementation of denomination of origin for coffee. The study was to be implemented jointly with colleagues from CENICAFE and the FNC quality department. The objective of the study was to identify the causal but regionally-changing relationships between quality characteristics of the coffee product and the characteristics of the environment where it is grown. The coffee-growing departments of Huila (northern areas), Tolima (southern areas), Magdalena, César, Santander and Santander de Norte were selected for the Phase 2 study.

The rationale behind the study was to correlate environmental data held by CIAT with quality data of samples of the coffee product collected from farms during the 2007/08 harvest. To this end, in early 2007 a field survey was designed on the basis of prior knowledge from similar studies, and with input from the pilot phase conducted in 2006. Technical staff of the regional FNC offices identified the participating farms with the aim of including farms that were accessible and that covered the range of conditions that represent the coffee-growing environments of each department. The survey was implemented in three-step procedure, which included a project socialization phase, a preliminary round of advisory and informative field visits, and the actual field sampling.

A total of 481 farms were sampled in the Huila/Tolima Region and about 415 in the Northern Region. Each farm was geo-referenced to facilitate the analysis of spatial correlation. To reduce variability within the data, product samples were processed in a mobile unit that standardizes harvest and post-harvest processes.

Characterization of product quality was conducted at the FNC headquarters (physical and sensorial analyses) and at the FNC research center CENICAFE in Chinchina (near infrared spectrographic analyses for biochemical characteristics). Soil samples were also obtained in each farm. Also, descriptions for the agronomic practices used by farmers were collected by interviews on most farms. The information about the agronomic practices will provide valuable background information and context for the information compiled and analyzed in this study. All information including data files, the maps generated and the documents have been packaged and provided in a CD format to the partners of the study. In addition, a training workshop was conducted by CIAT for CENICAFE colleagues.

First, we analyzed the consistency of the coffee cuppers. The data of one cupper were identified as particularly suitable for our analyses. In a second step, the varieties Caturra and Colombia and Tipica,which were included in the analyses of physical and sensory characteristics, showed major differences at the departmental level due to environment by genotype interactions (G*E). For this reason, we ruled out the generation of unique quality profiles based on the departmental boundaries. We generated new spatial domains based of an innovative approach that integrated the formal analytical knowledge produced by the Phase 2 study with informal knowledge about coffee quality held by experts of the FNC. Six final spatial domains were generated, which substantially reduced the influence of the G*E interactions. Therefore, in the remaining analyses, data from all varieties were pooled on the bases of these new spatial domains.

Third, we quantified the environmental differences between the domains. These were generally statistically significant for all characteristics, including the number of dry months, annual precipitation and diurnal temperature range. We had observed the influence of these three characteristics on coffee product quality in the studies conducted in the first phase. In addition to the data we had already used already in the first phase, we also had included cloud frequency as a new source of spatial information. We also found significant similarity between the climate patterns and the conceptual spatial maps held by FNC experts on coffee quality. This is indeed a very important point as we use long-term simulated data of climate, while the knowledge held by experts also represents knowledge that has been accumulated over many years.

We found significant differences in product quality between the spatial domains, principally related to physical, biochemical and sensorial characteristics. Domain I coffees are characterized by low acidity and high body. Their fragrance and aroma exhibits a nutty character, as does their flavor and therefore the overall impression. Domain II coffees tend to be balanced with medium body and acidity, and they have a moderate level of sweetness. Their fragrance and aroma is characterized by nutty and chocolaty notes, which in the flavor are complemented by caramel tones. This leads to coffees that overall can be considered as having sweet, nut–caramel notes, with little astringency or off flavors. Domain III coffees are generally characterized by low acidity and medium body. Their sweetness level tends to be low. The fragrance and aroma has sometimes astringency notes, while the flavor can be chocolaty or nutty in some subregions. Coffees from domain III have the tendency to exhibit astringency and herbal nuances in the flavor as well. Domain IV coffees have medium to high levels of body and sweetness. Their acidity tends to be moderate, in some cases high, while the fragrance and aroma is characterized by sweet, fruity notes. These are generally reflected in the flavor, and complemented by sweet caramel nuances. This leads to an overall profile that is fruity and sweet, acidity in some cases may be citric. Domain V coffees are considered as having high levels of sweetness and acidity combined with medium body. They demonstrate fruity and floral fragrance and aroma, very often combined with sweet caramel notes, and a citric fragrance. The flavor fully reflects the fragrance and aroma expressions. It leads to coffees with overall profiles of high quality that are characterized by sweetness, and fruity and citric acidity, with clear caramel notes. The FNC experts stressed the high quality of these coffees. The spatial probability analyses we conducted with the actual samples confirmed the high potential in this domain for outstanding coffees, however, we also found that the existing potential is not realized fully in this region. This led to sensory descriptions based on samples from this study that under estimate the real potential of this domain. Domain VI coffees are balanced with medium levels of acidity, body and sweetness. They have sweet notes in fragrance and aroma, often accompanied by herbal tones. The flavor reflects these, but also shows sweet caramel aspects. Overall, these are sweet, fruity coffees that may have a herbal off taste.

Next we found that the differences in product quality were not random but have a clear spatial structure. We were able to show that some of these spatial structures in the quality data are related to those found in the environmental data. The cup profiles in both regions, i.e. the Huila/Tolima Region and the Northern Region are influenced strongly by climatic conditions. In the Final Scores of the sensory characteristics in the Northern Region, we found, for example,e that certain ranges of solar radiation and cloud frequency have a major positive impact. Acidity on the other hand is driven by altitude, cloud frequency and dry months per year. Final Score in the Huila/Tolima Region is positively influenced by certain altitude ranges and annual rainfall, while acidity in the Huila/Tolima region benefits from specific solar radiation values, diurnal temperature range and dry months per year. We recommend strengthening the knowledge about the cause–effect relationships with additional spatial probability analyses for other key characteristics of coffee quality. Furthermore, in the Northern Region we observed a strong influence of soil conditions on coffee quality. This provides an opportunity for management to have a direct impact on coffee quality.

Lastly we identified areas in Colombia and elsewhere in the world that are similar to the areas we analyzed in this study. The similarity analyses were conducted only on the basis of climate data and soils information at a coarse scale and it is important to note that we did not include altitude Within Colombia, there are some areas homologous to some parts of the six domains, while others are unique. We found some degree of similarity with other countries outside of Colombia in South America, specifically Brazil, and in Africa and less so in Asia. Domains IV and V tend to have most similarity with regions in other countries, except in Asia where domains II and III tend have more homologous areas. The degrees of similarity with other areas in and outside of Colombia are small, however, and are expected to decrease further once we can incorporate topography into the similarity analyses.

This study could not include temporal aspects on quality. It is well know from the wine industry that between-season differences in climate can have profound impacts on product quality. While these quality impacts are sometimes important in wine, they do not prevent product quality maintaining the expression of key characteristic traits. We do not have enough knowledge to prove this for the coffee sector.

In summary, we documented clear relationships between growing environment and characteristics of product quality. We believe that these provide ample evidence to build a strong case supporting any application procedure for regionally-based denomination of origin for the areas in which the study was carried out.

Table of Contents

Table of Figures

Table of Tables

2. Background of Denominations of Origin

Geographical indications (GIs) and the more demanding denominations of origin (DOs) are known more familiarly as labels of origin. They have often been used with wine and spirits, but are also applied other foods (e.g. cheeses, meat products, oils, or nuts). GIs are usually state-granted product protection schemes, which are increasingly recognized as a way for small and poor producers to escape the "commodity treadmill". Producers located in a GI area are only allowed to use the label if they follow the requirements for certification. GIs hold the potential of re-linking particular products to the social, cultural and environmental aspects of particular places, distinguishing them from mass produced goods (Barham 2003). Porter (1986) identifies differentiation as one of four key marketing strategies. Besides product branding, differentiation can be based on quality as stated by Cormorèche (1994, cited in Barjolle *et al.* 1997): "Next to the trend/cost quantity/quality competition, and to the brand competition, there exists a competition by the quality linked to a know-how, a region, a production basin, which is consistent with a logic where the price parameter is more flexible". There is growing consumer demand for GI products (Marsden *et al.* 2000, Murdoch *et al.* 2000, Van der Ploeg and Renting 2000, Van der Ploeg *et al.* 2000). Gilg (1996, p. 71) estimates that niche or specialty products including GI products could account for as much as 30 percent of overall food sales due to their higher value.

Four key international agreements and some specific European regulations address GIs. The agreements include the Paris Convention for the Protection of Industrial Property, the Madrid Agreement for the Repression of False or Deceptive Indications of Source on Goods, the Lisbon Agreement for the Protection of Denominations of Origin (DOs) and their International Registration, and the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) of 1994. The last agreement relates to the portion of the General Agreement on Tarrifs and Trade (GATT) dealing with intellectual property addressing geographical indications in Annex 1C, article 22(1). The agreement defines a GI as: indications which identify a good as originating in the territory of a [m]ember [country], or a region or locality in that territory, where a given quality, reputation or other characteristic of the good is essentially attributable to its geographical origin.

Avelino *et al.* (2006) wrote (translated from Spanish): "The GI is, nevertheless, much more than a simple stamp of the origin of a product (the classic Made in ...) that shows no relationship between the region of production and the characteristics of the product. For the name of a region to be eligible to be protected by a GI, the reputation or the quality, or characteristics of the product using the GI must be derived from its geographical origin (Bérard *et al.* 2001). In other words, the GI is a qualified seal of origin. According to article 2 of the Lisbon Agreement on the Protection of Denominations of Origin and its International Registry (1958), DO means 'the geographical denomination of a country, a region, or a locality that serves to designate an original product whose quality or characteristics must be [owed] exclusively or essentially to the geographic entity comprised of the natural and human factors.' For this reason the DO is a more demanding form of GI: in the case of a DO the origin in its widest sense, including the methods of production, must strongly explain the quality or the characteristics of the product (Falcetti 1994, Bertozzi 1995, Salette *et al.* 1998; Barham 2003), and not just one quality or characteristic of the product as in the case of the GI."

In its council regulation No 2081/92, the European Union declared two kinds of protection for local foods and food products, namely the "Protected Geographical Indication" (PGI) and the "Protected Designation of Origin" (PDO). The PGI status applies to agricultural products or foods that originate in a region, specific place, or country, and that possess a specific quality, reputation or other characteristic of that geographical area. The PDO status is applied to products that originate in a specific region, place, or country, and have qualities or characteristics that are essentially or exclusively due to a particular geographical environment. As such, the PDO is similar to the general definition of GIs in TRIPS, while the PDO status resembles that of a denomination of origin stipulated in the Lisbon Agreement.

Undoubtedly there is a strong link between these modern day agreements and the concept of *terroir* that is intimately linked with French wine production. Historically, *terroir* refers to an area or terrain, usually rather small, whose soil and microclimate impart distinctive qualities to wines produced there. The *terroir* concept viewed production as a complex undertaking translating local ecology through management so that products display their qualities to best advantage. A great deal of knowledge about the local terrain is needed for success, as well as respect for local natural conditions that can be expressed through the wine. The modern *Appellation d'Origine Controlée* (Name with a Controlled Origin, AOC) is built on the *terroir* concept and has been evolving in recent years along with EU recognition of labels of origin. Presence of an AOC label on a product reflects the association of product and its region of origin in three main categories: association with natural environment factors, with human factors (particular techniques and know-how), and with history. Although an AOC product must incorporate all three aspects of its *terroir* the natural and human factors are decisive and the tie to nature figures most prominently in determining a product's tie to its *terroir* (Barham 2003).

Avelino *et al.* (2006) wrote (translated from Spanish): "From this point of view, European regulation number 2081/92 regarding the protection of GIs and DOs of agricultural products and food products is of interest. The regulation stipulates in article 4 the requirements that must be fulfilled in requests for PGIs and PDOs. They must contain at least the following elements together with what is called the list of conditions:

1. The name of the product and the region of origin for which protection is required,
2. A description of the product and its physical, chemical, microbiological and organoleptic characteristics,
3. The geographical boundaries of the area from which the product comes,
4. A description of the method by which the product is produced: production norms,
5. The factors responsible for relating the product to the region where it is produced,
6. The procedure(s) for product [quality] control,
7. Labelling details,
8. Possible requirements that must be complied with according to EU and/or national regulations.

This list shows an application for a PGA or a PDO demands a profound technical study. One of the most important points in the list of requirements is undoubtedly point number 5. [The applicant] must demonstrate the relationship between the area of origin and the product, its quality or its reputation. The lack of an objective relationship is the principal reason for rejection of applications for PGIs and PDOs in Europe. (Information supplied by Antonio Berenguer of the General Directorate for Commerce of the European Commision.) Figure 1 summarises the strategies as functions of the reputation [of the product] and the objectives sought."

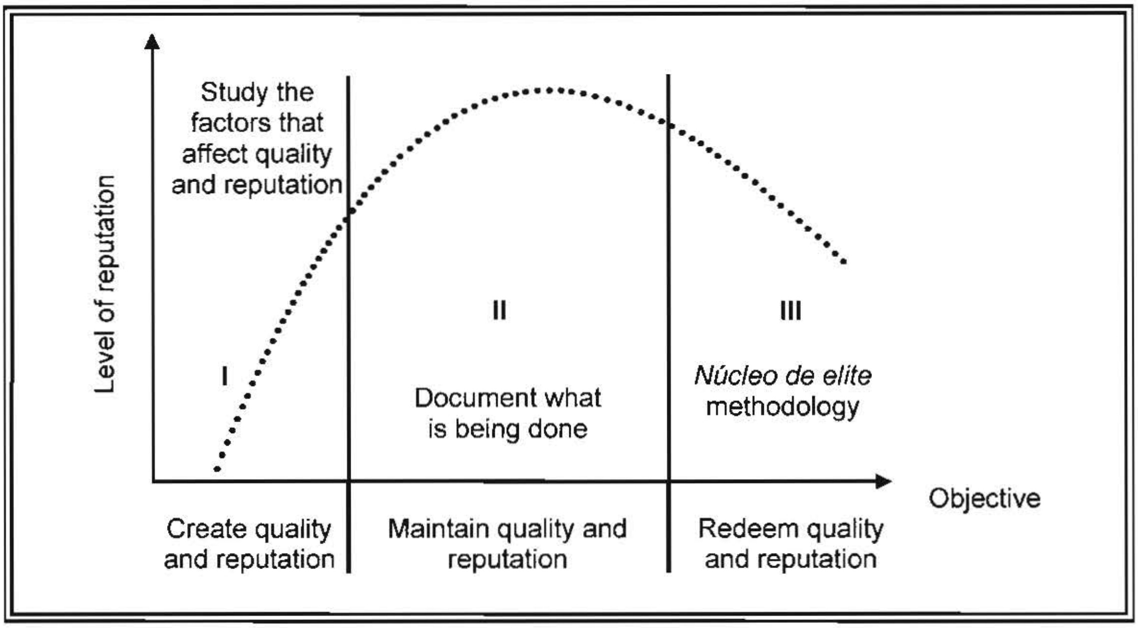


Figure 1. Some possible strategies for creating a Geographic Indication or a Denomination of Origin as a function of the level of reputation and the objective (redrawn from Avelino et al. 2006).

The work presented here contributes to the process of establishing DOs for coffee in Colombia by attempting to identify the relationships between coffee quality in the Nariño and Cauca Departments and their production environment.

3. Study objective

3a. General Objective

Conduct a case-study to identify production system, growing conditions and product quality relations that can support the implementation of concepts of DOs for coffee grown in the Departments of Huila (Northern part), Tolima (Southern part), Santander, North Santander, Cesar and Magdalena (Figure 2).

3b. Specific Objectives:

1. Obtain information on and establish a GIS database of production factors (topography, climate, soils, socio-economic, agronomic) for the Departments of Huila, Tolima, Santander, North Santander, Cesar and Magdalena;
2. Describe spatial variation in the climatic, topographic and soils factors;
3. Obtain information on and establish a GIS database of coffee quality information (physical and organoleptic characteristics) for the Departments of Huila, Tolima, Santander, North Santander, Cesar and Magdalena;
4. Describe spatial variation in the climatic, topographic and soils factors;
5. Conduct spatial correlation analyses between production factors and coffee quality information, so that specific production factors (soils, climate, altitude socio-economic or agronomic) can be identified that demonstrate a link between the territory under study and the characteristics of the coffee;
6. Summarize the results and recommend further actions in a final report;
7. Create GIS databases for the case-study area in ESRI Arc Info Format; and

8. Create a data set of product quality information.

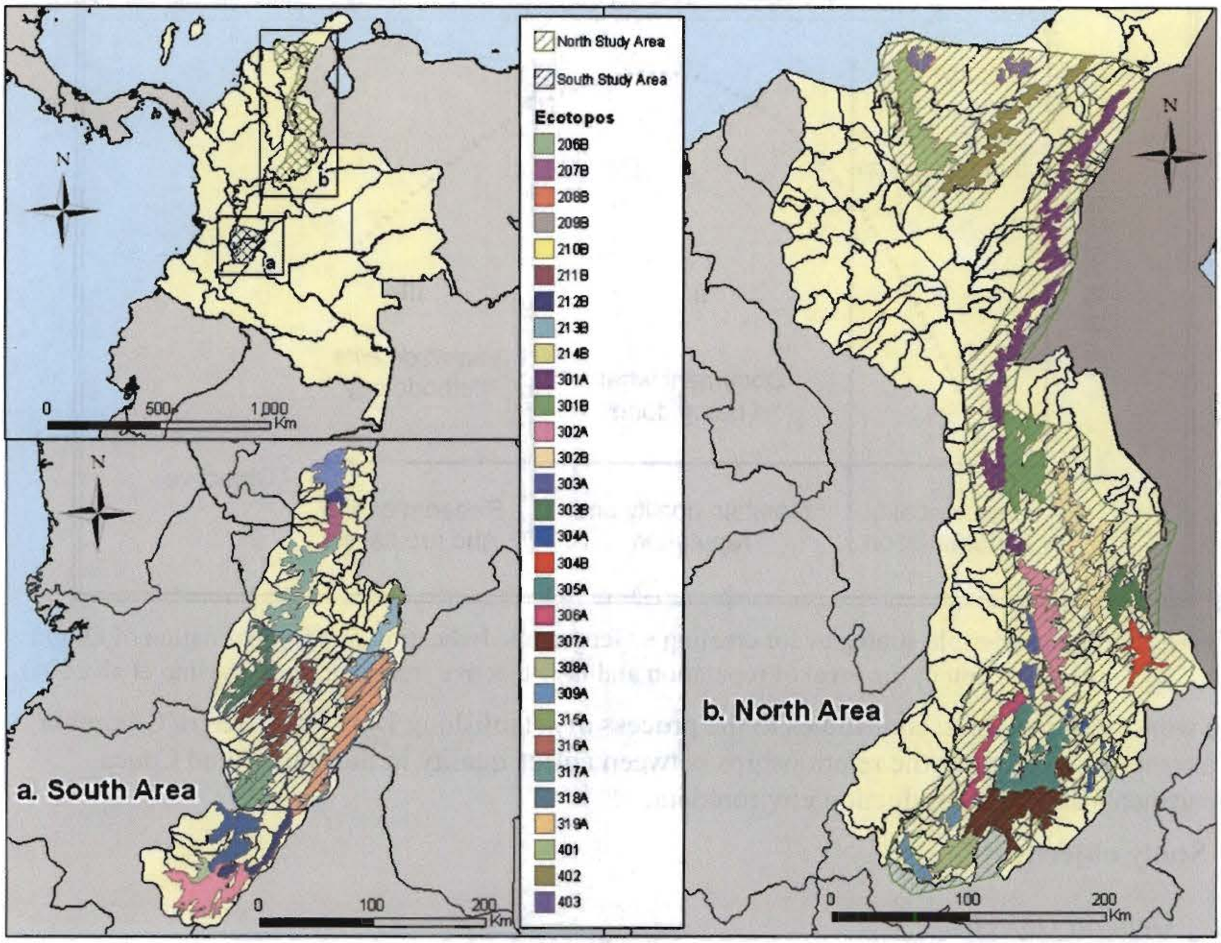


Figure 2. Maps of the two zones (a) Huila/Tolima (b) Santander, Santander Norte, César, and Magdalena.

4. Conceptual outline of study: Spatial epidemiology

Recent advances in both the availability of environmental availability and analytical methods have created new opportunities for investigators to improve on the traditional reporting of relationships between environment and coffee quality. It is now possible to study how these these relationships change as scale increases from local scale (Avelino *et al.* 2005, Vaast *et al.* 2006) to regional scale.

To this end we adapted the analytical framework of spatial epidemiology. Spatial epidemiology is the description and analysis of geographically-indexed health data with respect to demographic, environmental, behavioral, socioeconomic, genetic, and infectious risk factors (Elliott and Warwick 2004). In our case we derive spatial patterns of coffee quality from an analysis of geographically-indexed data on product quality from farms and environmental data (topography, climate and soils) that are not only available at each farm location but also the whole area under consideration (Figure 3).

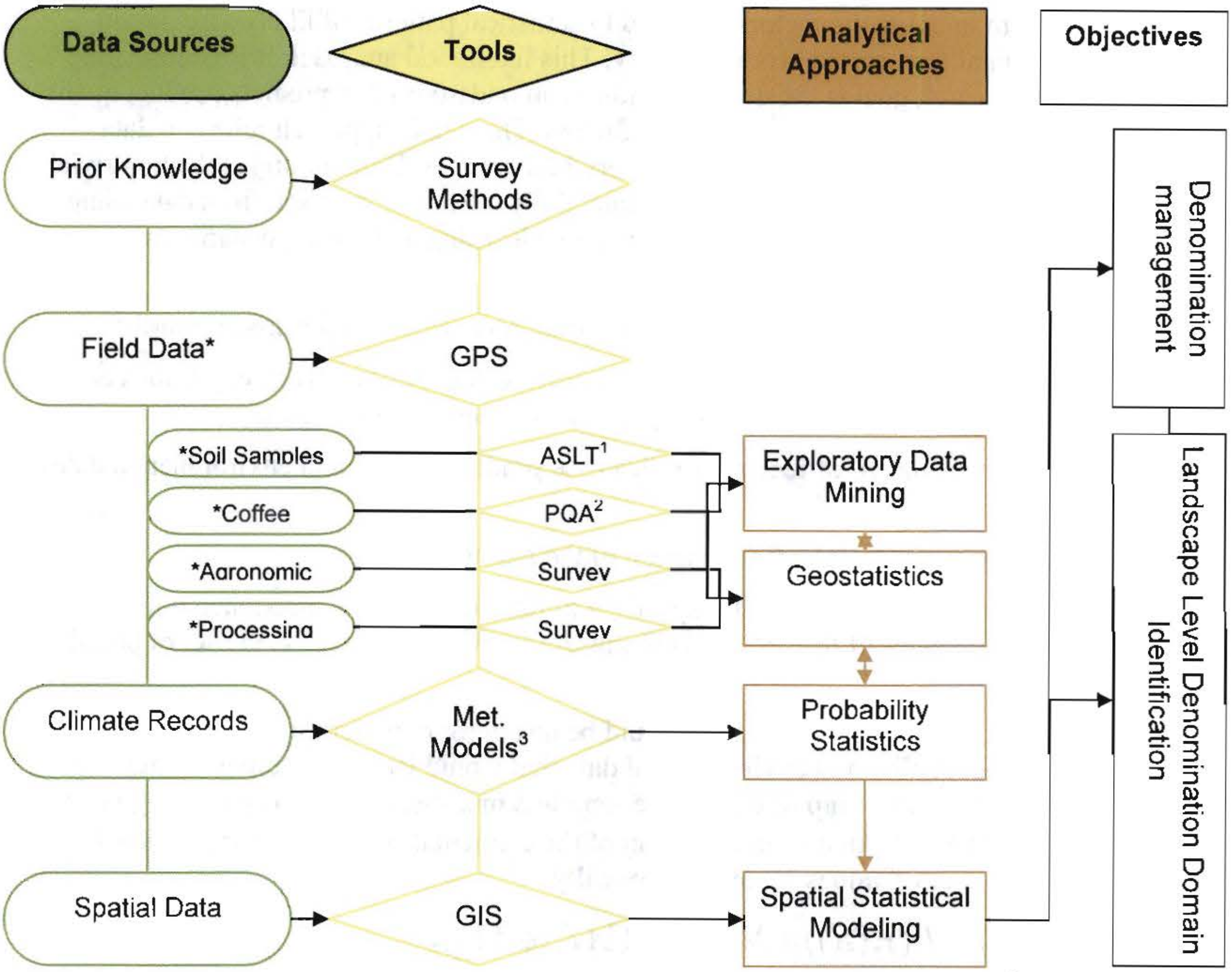
We first attempted to uncover the factors that lead to statistical patterns of likely association between environmental factors and product quality. This likelihood approach is predominantly statistical and our research goal is the determination of an odds-ratio that predicts product quality outcomes, given presumed causal environmental factors. The whole approach relies on data mining, which is the process of knowledge discovery in databases. Data mining is the nontrivial extraction of implicit, previously unknown, and potentially useful information from data using multidimensional data visualization techniques, machine learning and standard statistical methods.

De Nadai Fernandes *et al.* (2002) provide a useful formal representation for this approach:

> Let S be a coffee sample belonging to an unknown category of interest (e.g. Cauca or Nariño); CAT_i be a category i from Ω_{CAT}, a set of categories of interest;
>
> E(S) be a vector (size n) of features (in this case product quality and environmental data) in sample S;
>
> R(E(S)) be a (multivariate) $n * m$ function of the features;
>
> Nj be a particular subset j of the R^m space (m-dimensional space), associated with function R and vector of features E. This subset j is an element of Ω_N, the set of possible subsets.

A first issue of interest here is whether there would be an appropriate definition of a function R(E(S)) of the product quality and environmental data that would lead to meaningful information on the probabilities of membership of this coffee sample S in a specific category CAT_i, given the observation that $R(E(S)) \in N_j$, that is, the function of the elemental concentrations presented a value within the category of values N_j, or, algebraically:

$$\Pr\left[S \in CAT_i \mid R(E(S)) \in N_j\right], \qquad CAT_i \in \Omega_{CAT} \tag{1}$$



1Analytical Soil Laboratory Techniques 2Product Quality Analyses 3Meteorological Models

Figure 3. Conceptual framework of the project.

Equation (1) can be represented (under Bayes' rule) by:

$$\frac{\Pr\left[R(E(S))\in N_j \mid S\in CAT_i\right]\Pr\left[S\in CAT_i\right]}{\sum_{CAT_i\in\Omega_{CAT}}\Pr\left[R(E(S))\in N_j \mid S\in CAT_i\right]\Pr\left[S\in CAT_i\right]},\quad CAT_i\in\Omega_{CAT} \qquad (2)$$

In Equation (2) the last term in the numerator represents the *a priori* probability of having the coffee sample *S* classified in CAT_i without any information on product quality and environmental data. A comprehensive database with information for each coffee sample, including product quality and environmental data and agronomic data, might facilitate the estimation of the conditional probabilities in Equation (2). There might be more than one function R(E(S)) of the product quality and environmental data that provide useful information under the framework discussed so far. In this situation, aspects of the effectiveness of information generation can be taken into consideration to facilitate the selection.

We used graphical analyses as well as parametric and nonparametric methods to investigate the feasibility and nature for discrimination of coffee regions based on environmental and data of product quality. These techniques included multivariate data visualization methods (descriptive statistics, clustering, principal components, discriminant analyses, geostatistics), and data-driven Bayesian analyses. Once causal relationships were determined at point locations and the role of environmental factors elucidated, we used spatial modeling with Bayesian probability statistics to transfer the functional causal relationships to areas where we only had information on environmental factors but no data on product quality.

5. Methods and data generation

5a. Field sampling design and preparatory consultations

Ideally we would have been able to apply some kind of probability sampling, either model- or design-based approaches (e.g. Dobermann and Oberthür 1997, Brus and De Gruijter 1997) on which to implement collection of coffee bean and soil samples. Probability sampling means that every element, or sampling unit, of a population, has a known probability of being included in the survey sample. Once the population is defined, sampling sites are selected randomly. We used a variant of this approach, called stratified random sampling, which is typically used to assure that smaller groups are adequately represented in a sample.

Given logistical contrainst (e.g. travel time, harvest period, farm accessibility) we had to set an upper limit for sampling sites. The difficult issue in purposive quota sampling is to decide upon the specific characteristics on which the quota will be based. We used our prior knowledge generated by similar work (Läderach *et al.* 2006). We identified a number of key environmental factors that are linked to coffee liquor characteristics including altitude of the growing area, diurnal temperature range, annual average temperature, dew point, and annual average precipitation. Some of these characteristics are obviously highly correlated. Altitude and diurnal temperature range are less correlated in the target region.

The actual sampling was based on the files held by the FNC within each Department, which identify the Colombian farms on which coffee is grown. Within each Department, sampling sites were placed in four groups using cluster classification based on the environmental factors of annual precipitation, mean annual temperature, mean diurnal temperature range, the number of dry months and mean annual solar radiation. In the Huila/Tolima Region, the cluster classification was done separately for each Department. In the Northern Region, the first cluster classification was done for all six Departments. These clusters were then subdivided into three sub-areas, North, Center and South, and a new cluster classification was done for each sub-area. Table 2 provides details about the clusters.

Sampling was only implemented in regions that were identified as coffee growing zones by the FNC *ecotopo* maps. The number of sampled sites was restricted by logistical constraints. We defined in discussion with the regional FNC offices that the maximum farms that could be sampled were about 500 for the Huila/Tolima area and about 400 for the Northern region. We then assigned proportionally to the number of farms listed in the FNC files sampling sites to each cluster. A random number generator was used to locate the specified number of farms within each climate cluster. Technical field staff of the regional FNC offices helped to locate the farms in each cluster. The field staff was then instructed to identify the farms so that they were representative of the predominating production systems. The latter aspect is important in order to reduce heterogeneity in cropping systems mainly with respect to shade coverage (see for example Vaast *et al.* 2006). If farms were not representative, not existing any more or simply not accessible nearest available farms were sampled. Table 1 provides details about the number of sampled farms. Figure 4 – Figure 8 show the spatial distribution of sampling units, and sampled farms.

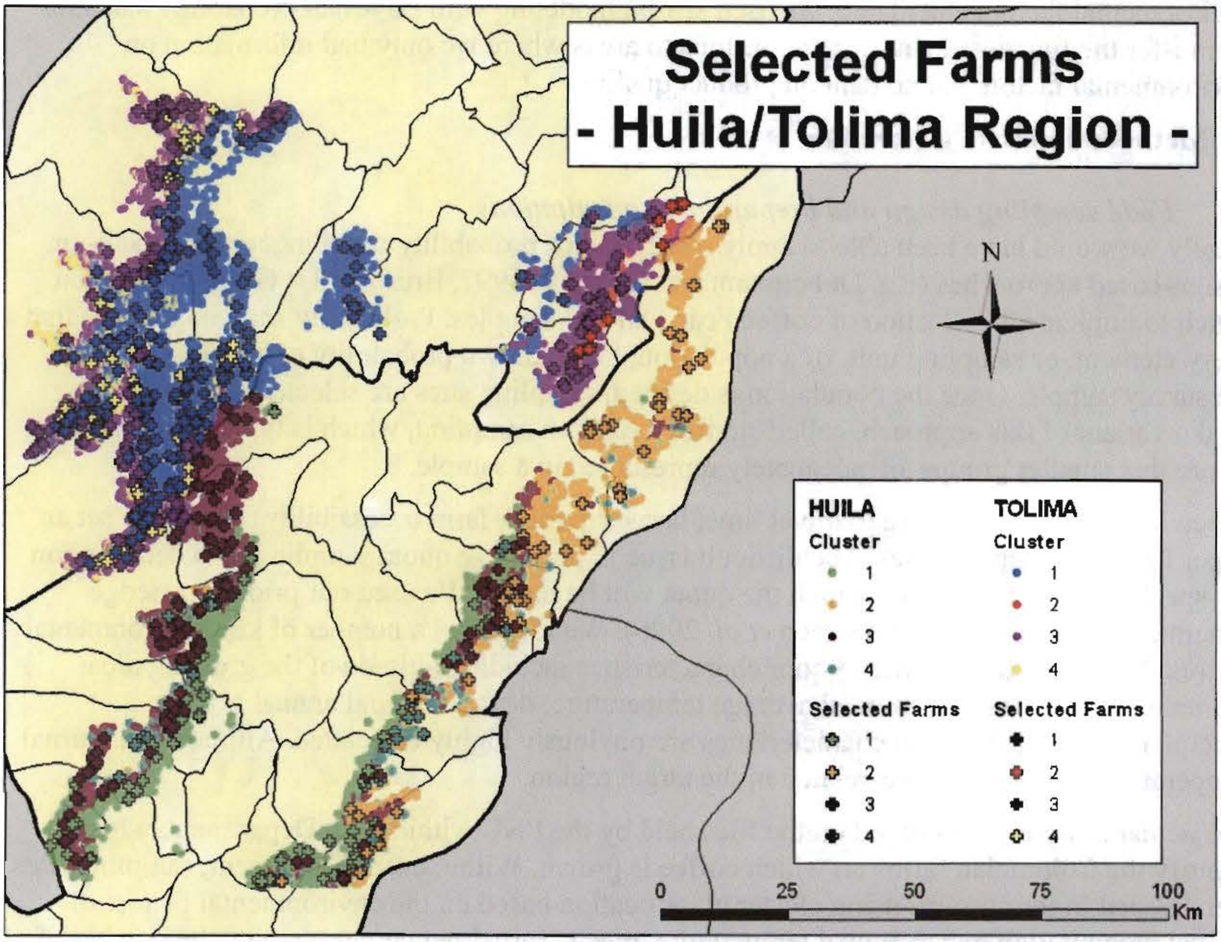


Figure 4. Map of the Huila/Tolima Region showing the environmental site clusters and the selected farms.

To prepare the actual field sampling work, between 10 and September 17, 2007 the project staff held meetings with extension agents in the coffee-growing municipalities of Santander, Norte de Santander, César-Guajira and Magdalena. In the meetings a brief description of the proposed designation of origin for coffee in Colombia was given, the mobile processing and georeferencing units explained, and it was shown how to take soil samples, and which analyses they would be subjected to. The procedures of how to ship samples to Cenicafé were discussed and agreed upon, and a timeline and access routes to carry out the field sampling were identified. A specific work plan was designed for each week of the sampling. Finally central sites for collection and processing of samples were identified within each district. Procedures for receiving any telephoned information from extension agents and heads of each area were agreed upon. The technicians verified the selected farms and if necessary reported any changes in accordance with the clusters classifications. As well, they visited all farms before sampling, and stimulated social interactions with the farmers. Samples were taken during the harvest by the field team, and then carried to the central processing site for wet processing and drying.

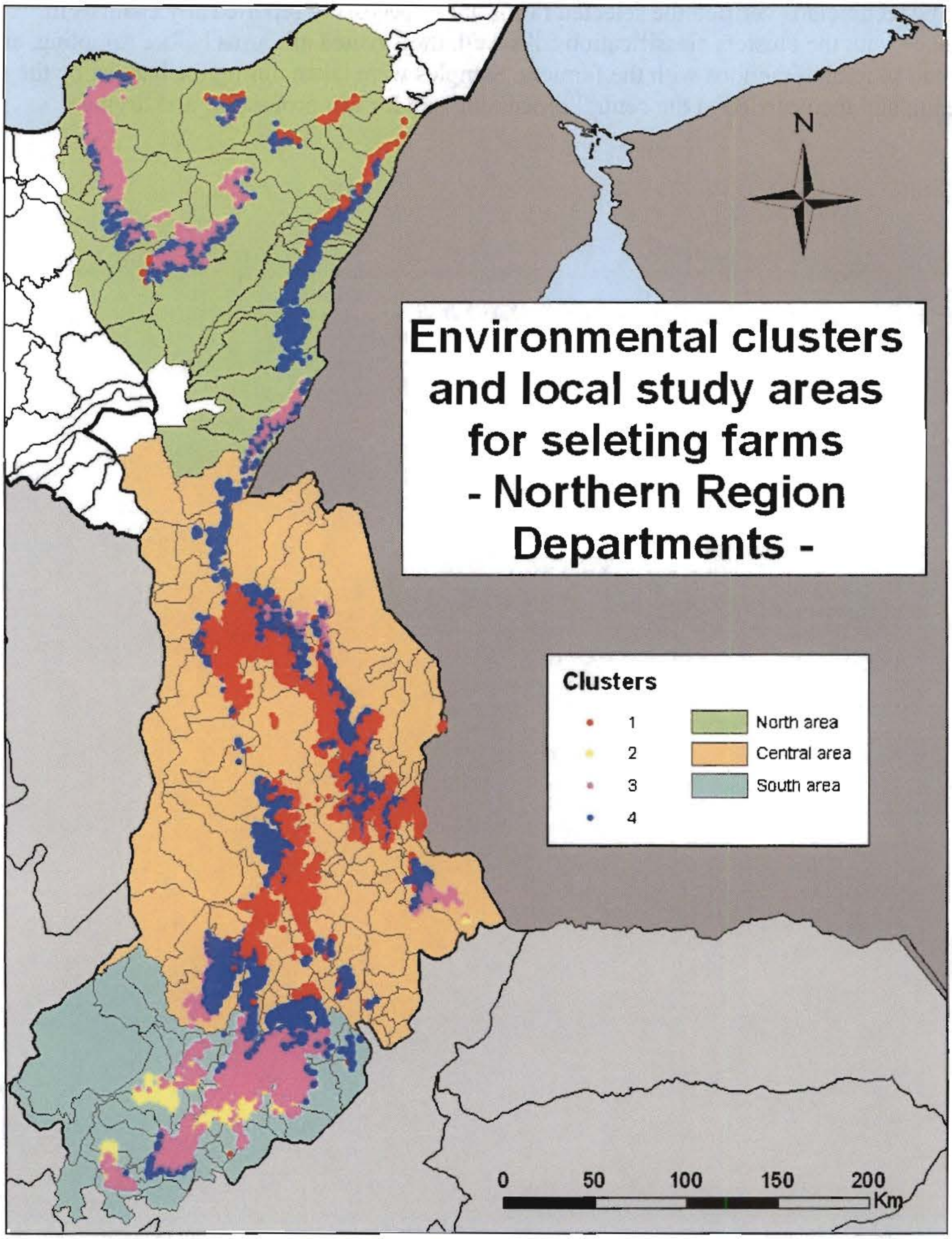


Figure 5. Map of the Northern Region Departments (Santander, Santander Norte, César, and Magdalena) showing altitude/diurnal temperature range clusters and the three selected zones.

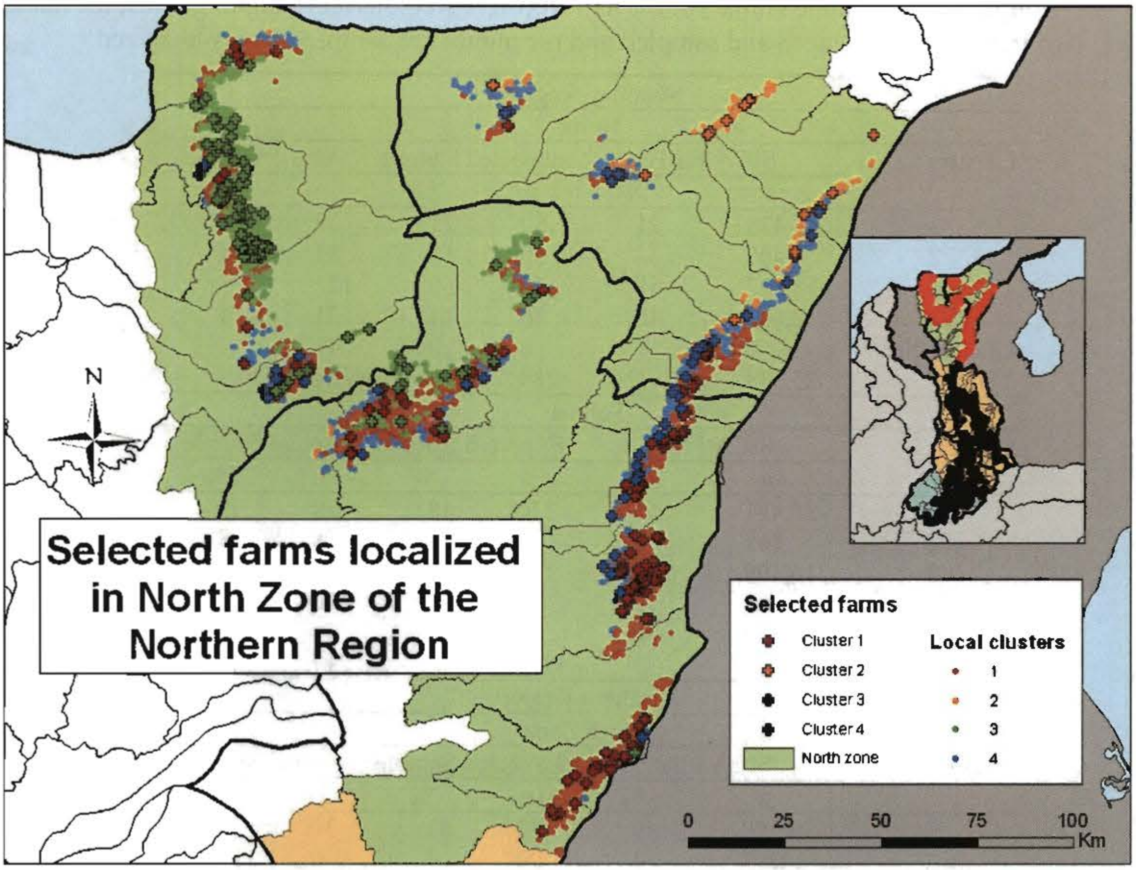


Figure 6. Local clusters and the farms selected within each in the North Zone of the Northern Region Departments (Santander, Santander Norte, César, and Magdalena).

Table 1. Sampling points in the Huila/Tolima and Northern Regions and sampling units, the number of farms geo-referenced and sampled and the number of samples finally analyzed.

Southern Region						
Huila						
Cluster	SICA FILE		Expected Sample		Actual Sample	
	n	%	n	%	n	%
1	5,425	21	60	21	38	17
2	4,405	17	50	18	32	14
3	13,092	51	140	50	122	55
4	2,607	10	30	11	31	14
Unclassified[1]	0	0	0	0	41	-
Total	25,529		280		264	
Tolima						
Cluster	SICA FILE		Expected Sample		Actual Sample	
	n	%	n	%	n	%
1	21,481	50	230	48	78	45
2	261	1	15	3	8	5
3	16,109	37	170	36	77	45
4	5,364	12	60	13	10	6
Unclassified[1]	0	0	0	0	44	-
Total	43,215		475		217	
Northern Region						
North Zone						
Cluster	SICA FILE		Expected Sample		Actual Sample	
	n	%	n	%	n	%
1	9,206	49	130	43	30	36
2	793	4	20	7	0	0
3	6,298	33	100	33	54	64
4	2,619	14	50	17	0	0
Unclassified[1]					52	-
Total	18,916		300		136	
Central Zone						
Cluster	SICA FILE		Expected Sample		Actual Sample	
	n	%	n	%	n	%
1	41,330	52	150	50	40	40
2	24,088	30	90	30	40	40
3	14,015	18	50	17	19	19
4	23	0	10	3	0	0
Unclassified[1]					27	-
Total	79456		300		126	
South Zone						
Cluster	SICA FILE		Expected Sample		Actual Sample	
	n	%	n	%	n	%
1	13,246	26	80	27	40	29
2	70	0	10	3	0	0
3	29,597	58	160	53	83	60
4	8,012	16	50	17	16	12
Unclassified[1]					14	-
Total	50,925		300		153	

[1] It was impossible to classify some of the samples in any cluster because either there were no geographic coordinates, or the climate data did not fit with any cluster

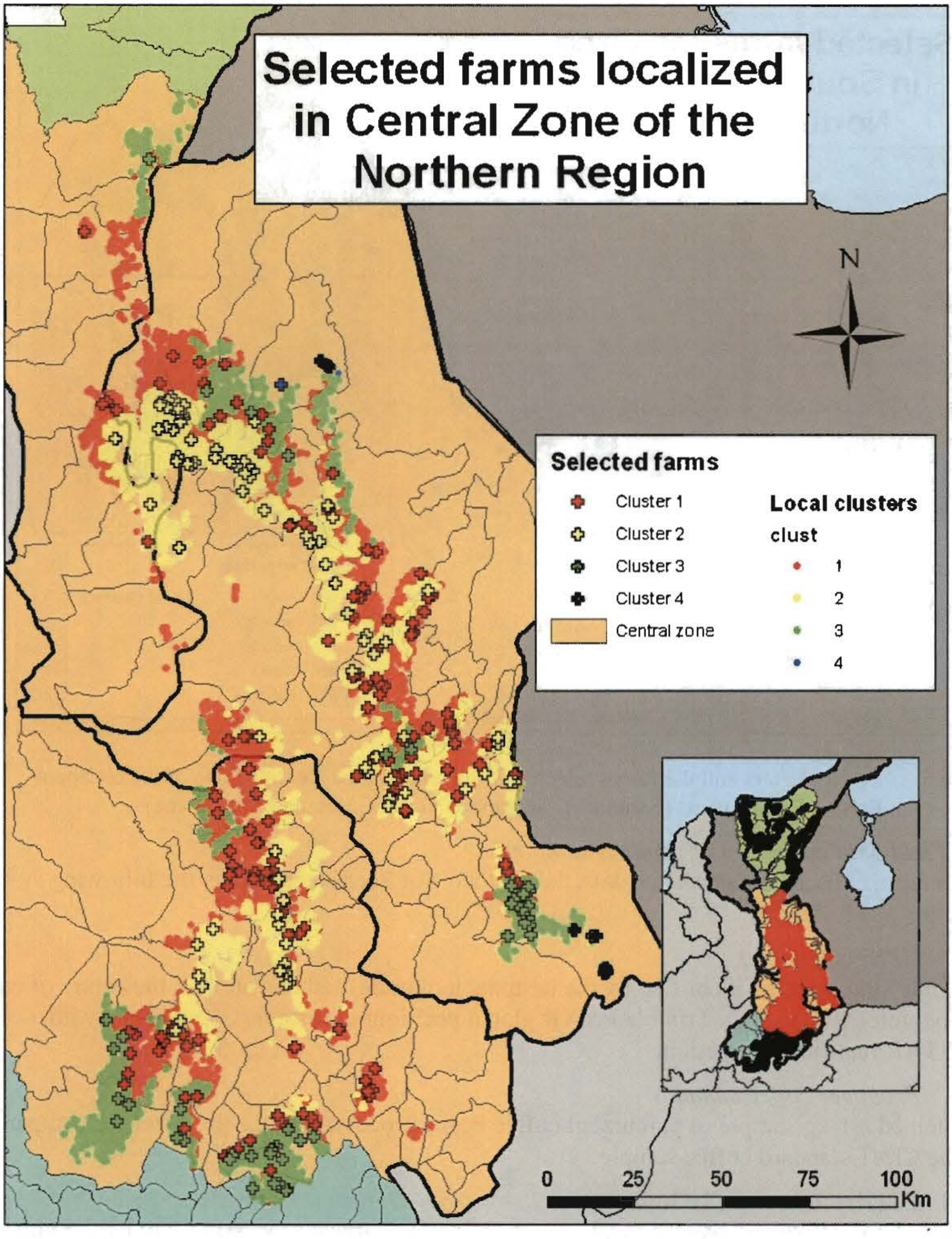


Figure 7. Local clusters and the farms selected within each in the Central Zone of the Northern Region Departments (Santander, Santander Norte, César, and Magdalena).

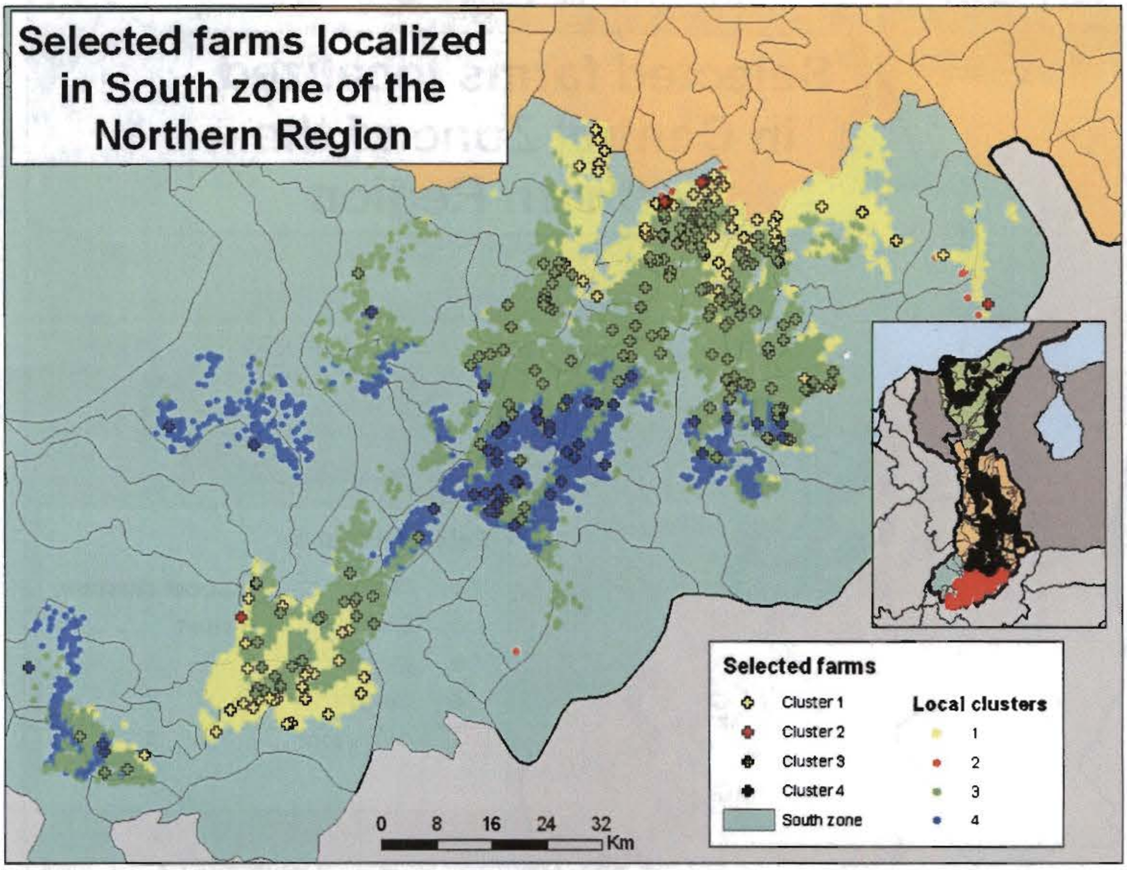


Figure 8. Local clusters and the farms selected within each in the Central Zone of the Northern Region Departments (Santander, Santander Norte, César, and Magdalena).

5b. Field data collection of product samples

In each farm, chosen in accordance with the protocols of Section 5a above, the following procedure was followed:

5b (i). GPS coordinates

All sample sites were all identified by the latitude, longitude, and elevation in the centre of each mapping unit (MU) using a Trimble ProXR global positioning system (GPS) device with OmniSTAR real-time correction.

5b (ii). Producer coffee samples

We obtained a 1-kg sample of parchment coffee beans processed by the producer for comparison with the CIAT standard coffee sample.

5b (iii). CIAT standard coffee samples

Samples of coffee were delivered to CIAT's mobile post-harvest processing unit by the farmers immediately after harvesting. Damaged, green and infected berries, stones, leaves and other artifacts were removed before de-pulping and removal of mucilage in a J.M. Estrada Model 100 unit. The samples were fermented separately for 5 hours in buckets, and then dried in a metal closet with four floors of drawers (compartments), each of which are perforated on the bottom. Air is heated to 40°C with gas and blown into the bottom of the closet. The hot air ascends through the closet passing through the beans, drying them and leaving at the top.

The dryer has the capacity to process 24 times 1-1.5 kg samples at the same time. The most recent samples were placed in the top compartment and were moved down when new samples were added, emulating the process of industrial dryers. Samples were dried until they reached a humidity of 10% to 12%, which took between 14 and 16 hours depending on the size of the beans. The dried samples were then placed in sealed plastic bags and stored at 18° C until the cupping process.

Figure 9 and Figure 10 track the field sampling progress over time.

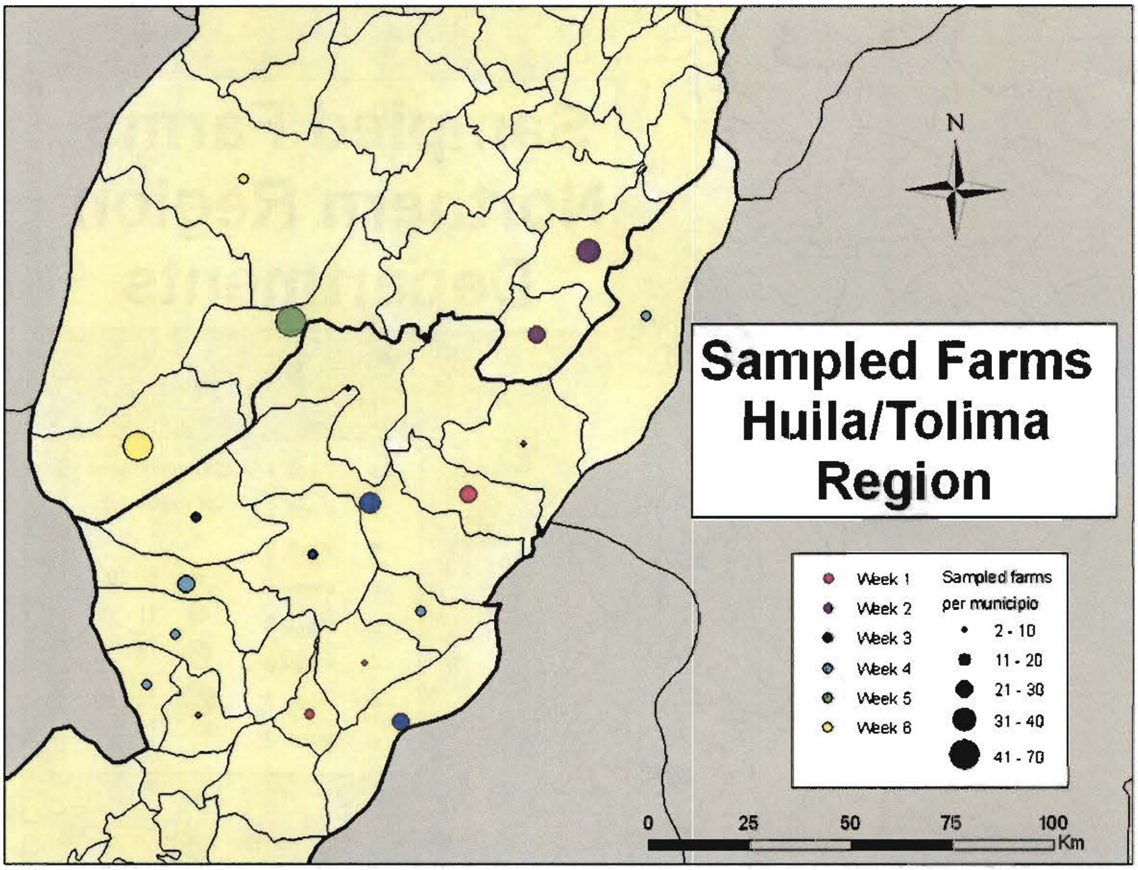


Figure 9. Progress on sampling by municipality in the Huila/Tolima Region.

5b (iv). Soil samples

Soil samples were taken from the same fields from which the coffee samples came using the CENICAFE soil sampling protocols.

5b (v). Farmer interview

We developed a questionnaire to capture of information on farmers'crop management. The questionnaire consists of three parts, including the administrative and geographic description of the farm with its area and personal information about the grower; details of field management practices including varieties; planting dates, system and distance; pruning and ratooning; shade management including, shade trees and planting distances; fertilization; and disease and pest management. When researchers or members of the grower associations visited the growers' farms to explain the project and collect the samples, they assisted the farmer in filling out the questionnaire.

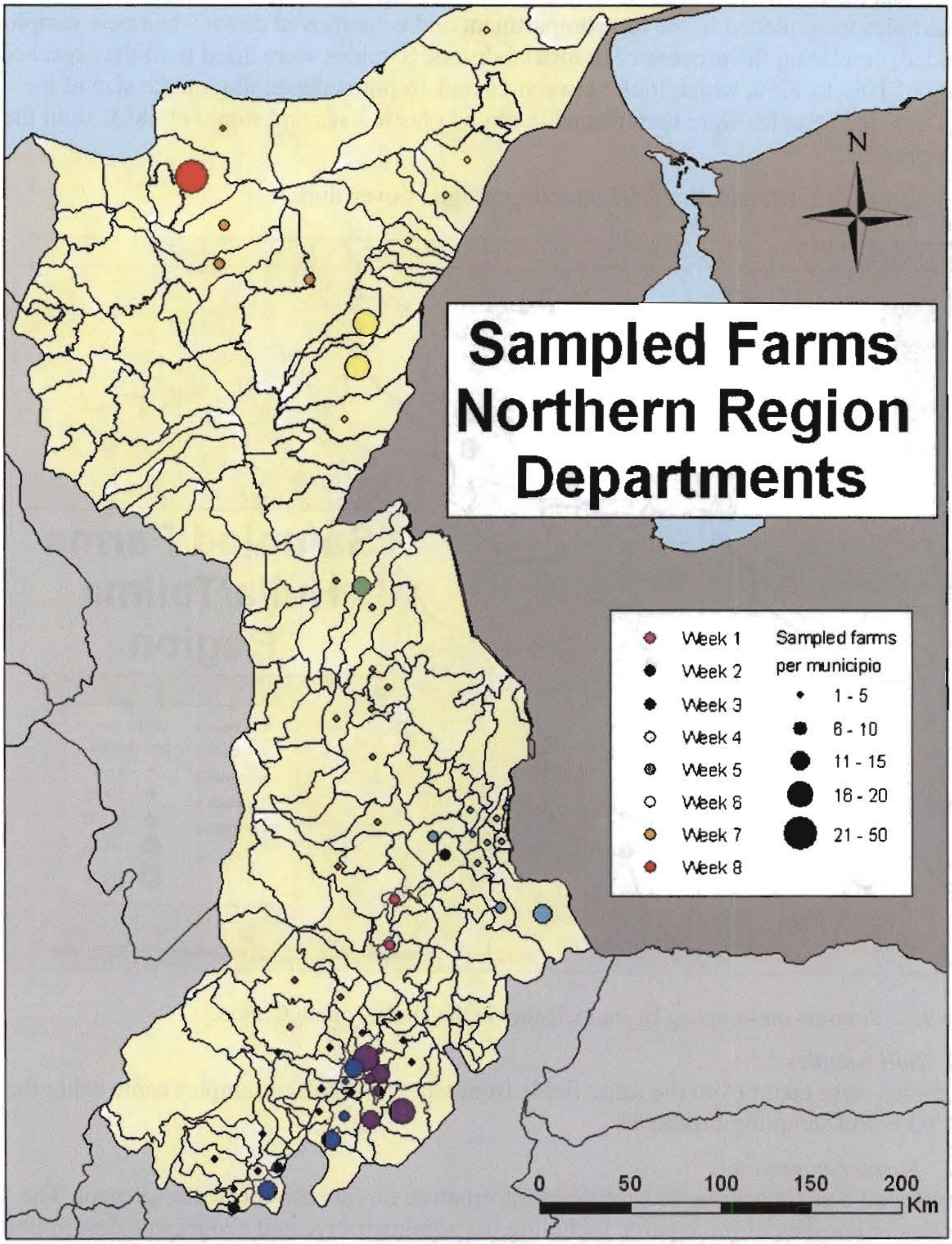


Figure 10. Progress on sampling by municipality in the the Northern Region Departments (Santander, Santander Norte, César, and Magdalena).

Table 2. Cluster analysis of sampling points for environmental data, soils, NIRS, coffee quality and agronomic data from farmers' interviews.

Southern Region						
Huila						
Cluster	Total	Environmental data	Soil	NIRS	Sensorial Physical	Agronomic
1	38	38	26	35	35	0
2	32	32	25	29	29	0
3	122	122	89	110	110	10
4	31	31	20	31	31	2
Unclassified	41	14				1
Tolima						
Cluster	Total	Environmental data	Soil	NIRS	Sensorial Physical	Agronomic
1	78	78	54	70	70	6
2	8	8	3	5	5	0
3	77	77	65	67	67	0
4	10	10	9	9	9	0
Unclassified	44	15				0
Northern Region						
North zone						
Cluster	Total	Environmental data	Soil	NIRS	Sensorial Physicall	Agronomic
1	30	30	25	26	26	29
2	0	0	0	0	0	0
3	54	54	47	47	47	17
4	0	0	0	0	0	0
Unclassified	52	52				44
Central zone						
Cluster	Total	Environmental data	Soil	NIRS	Sensorial Physical	Agronomic
1	40	40	37	37	37	40
2	40	40	37	37	37	40
3	19	19	19	19	19	19
4	0	0	0	0	0	0
Unclassified	27	27				27
South zone						
Cluster	Total	Environmental data	Soil	NIRS	Sensorial Physical	Agronomic
1	40	40	38	38	38	40
2	0	0	0	0	0	0
3	83	83	81	81	81	83
4	16	16	15	15	15	16
Unclassified	14	14				14

5b (vi). Expert knowledge

Coffee quality experts working for a long time for the FNC were asked to outline their understanding of spatial domains with distinct coffee profiles in the two regions. Therefore, three experts interactively draw spatial domains according to their best knowledge onto provided maps of the regions. These initial boundaries were then discussed in a group session and refined where necessary. Boundaries were then digitzed in the GIS lab of CIAT, in order to be used with the other available spatial information. Experts were also asked to populate matrices illustrating similarity between the outlined expert knowledge domains. Finally, the experts were asked to assign coffee quality descriptors to each of their identified expert knowledge domains.

The expert knowledge domains are of particular value for the process of denomination of origin as they constitute the accumulated long term evidence about the spatial relationships between coffee quality and specific regions. Also, they particularly lend themselves for the joint analyses with the used climate data in this study, which also describe the predominating climate based on long term observations.

The outlined domains were then analyzed jointly with other sources of information, specifically with spatial clusters of soils and climate that were generated using spatial cluster analyses. This visual interpretation of diverse sources of information permitted the identifaction of final analyses domains that were considered suitable for the process of implementing and defending denominations of origin.

5c. Environmental and product data analyses

5c (i). Point based analyses

Descriptive statistics

Data were assessed for normality using histograms and the *W* value of the Shapiro-Wilk test, which allowed a direct comparison of the distribution's fit (Starr *et al.* 1992). Summary statistics were computed for the all the data. Various multivariate analyses, including cluster analyses, principle component analyses, regression, and discriminant analysis were applied as described below. Hair *et al.* (1992) give a more general treatment of these techniques. Oberthür *et al.* (2000) give more detailed information on the application of discriminant analysis to local soil knowledge.

Note on data scale

All the environmental data and the biochemical and physical information on product quality used in this study were measured on an interval or ratio scale. However, information of sensorial quality was measured on a quasi-interval scale, that is product qualifications were made on a scale of 0 to 10 with increments of 0.5 giving a range of 21 points available. While such data are now commonly used in similar studies (Decazy *et al.* 2003, Avelino *et al.* 2005, Vaast et al. 2006), some may question the validity of using these data in parametric statistical methods. The sensorial data described here are analogous to a Likert scale (completely agree, strongly agree, agree, etc.), which are commonly analyzed using interval procedures. In considering ordinal Likert scale items, in a review of the literature Jaccard and Wan (1996) conclude, "for many statistical tests, rather severe departures (from intervalness) do not seem to affect Type I and Type II errors dramatically." Therefore, provided the scale item has at least five, and preferably seven categories, the assumption of normal distribution, required for many tests, may be assumed to be valid. Conversely, as the number of points decreases, it will be more likely that the distribution departs from the assumption of normality.

Principal components analysis

Principal component analysis (PCA) is a multivariate technique for examining relationships among several quantitative variables. Given a data set with p numeric variables, one can compute p principal components. Each principal component is a linear combination of the original variables, with coefficients equal to the eigenvectors of the correlation or covariance matrix.

PCA was originated by Pearson (1901). Principal components have a variety of useful properties (Rao 1964, Kshirsagar 1972): The eigenvectors are orthogonal, so the principal components represent jointly perpendicular directions through the space of the original variables. The principal component scores are jointly uncorrelated.

The first principal component has the largest variance of any unit-length linear combination of the observed variables. The jth principal component has the largest variance of any unit-length linear combination orthogonal to the first j-1 principal components. The last principal component has the smallest variance of any linear combination of the original variables. The scores on the first j principal components have the highest possible generalized variance of any set of unit-length linear combinations of the original variables. The first three principal components typically capture the majority of the variance in the data set under consideration, and are generally used in our analyses for interpretation.

Cluster analyses

The purpose of cluster analysis is to place objects into groups or clusters suggested by the data, not defined *a priori*, such that objects in a given cluster tend to be similar to each other in some sense, and objects in different clusters tend to be dissimilar.

Each observation begins in a cluster by itself. The two closest clusters are merged to form a new cluster that replaces the two old clusters. Merging of the two closest clusters is repeated until only one cluster is left.

The data representations of objects to be clustered also take many forms. The most common are:

A square distance or similarity matrix, in which both rows and columns correspond to the objects to be clustered. A correlation matrix is an example of a similarity matrix.

A coordinate matrix, in which the rows are observations and the columns are variables. The observations, the variables, or both may be clustered.

Any generalization about cluster analysis must be vague because a vast number of clustering methods have been developed in several different fields, with different definitions of clusters and similarity among objects.

The various clustering methods differ in how the distance between two clusters is computed. We used Ward's minimum-variance method where the distance between two clusters is the ANOVA sum of squares between the two clusters added up over all the variables. At each generation, the within-cluster sum of squares is minimized over all partitions obtainable by merging two clusters from the previous generation. The sums of squares are easier to interpret when they are divided by the total sum of squares to give proportions of variance (termed squared semipartial correlations).

Discriminant analyses

The purpose of discriminant analysis is to find a mathematical rule, or *discriminant function*, for guessing to which class an observation belongs, that is to say, discriminant analysis is used to classify observations into two or more known groups on the basis of one or more quantitative variables.

Classification can be done by either a parametric or a nonparametric method. A parametric method is appropriate only for distributions that are approximately normal within each class. The method generates either a linear discriminant function (the within-class covariance matrices are assumed to be equal) or a quadratic discriminant function (the within-class covariance matrices are assumed to be unequal).

When the distribution within each group is not assumed to have any specific distribution or is assumed to have a distribution different from the multivariate normal distribution, nonparametric methods can be used to derive classification criteria.

The performance of a discriminant function can be evaluated by estimating error rates (probabilities of misclassification). Error count estimates and posterior probability error rate estimates are evaluated. The error rates are also estimated by cross validation.

The linear discriminant function we used is:

$$\text{Constant} = -0.5\overline{X_j'}\,\text{COV}^{-1}\,\overline{X_j} \quad \text{Coefficient Vector} = \text{COV}^{-1}\,\overline{X_j} \tag{3}$$

5c (ii). Spatial analyses

Bayesian probability analyses

Various modeling approaches exist to identify suitable niches for specific crops, and one such approach has been used to create a spatial decision support system (SDSS), that is, a software tool based in geographical information science, which can assist users in decision-making. The tool, crop niche selection in tropical agriculture (CaNaSTA), was initially developed to suggest niche forage species to smallholder farmers in the tropics.

The engine used to develop CaNaSTA is Bayesian probability modeling. Bayesian methods provide a "formalism for reasoning under conditions of uncertainty, with degrees of belief coded as numerical parameters, which are then combined according to rules of probability theory" (Pearl 1990). A simple Bayesian model defines prior and conditional probability distributions and combines these to calculate posterior probabilities for each possible outcome. The probability distributions may be derived from data, set by experts or defined from a combination of data and expert opinion.

The CaNaSTA algorithm (O'Brien 2004) creates conditional probability tables of all predictor variables against response variable categories. In the case of coffee, predictor variables include climate and topographic factors and the response variable sensorial, fiscal or biochemical quality attributes. The primary model output is a discrete probability distribution at each location. A certainty value is also associated with each location, derived from the number of occurrences in the trial data with a particular combination of predictors and responses.

The probability distribution consists of the probability that the response variable is in each potential state. This information can be used to create maps showing the most likely response value ('Most likely'). The values in the probability distribution can also be weighted to produce a suitability value ('Score'). Finally, the certainty value can also be displayed as a map ('Certainty'), and can assist in the interpretation of the results. Once locations have been identified where a particular response is likely, further analysis can be carried out to determine which predictor variables are important. These driving factors can be either positive or negative, and can help with the analysis of specific conditions required for specialty coffee.

Calculating posterior probability distribution

A 'prior probability' is an initial estimate that may be modified once more information becomes available. If Y is a response variable, then the prior probability of Y is denoted $P(Y)$. 'Joint probability' refers to the probability of two events occurring together, such as a species thriving in a location with certain biophysical conditions. This is denoted by $P(X, Y)$, where X is a predictor variable (e.g., "rainfall is low") and Y is a response variable (e.g., "quality is high"). 'Conditional probability' is the probability of a response variable being in a given state, given that a predictor variable is a particular state, and is denoted $P(Y \mid X)$.

Conditional probability can be calculated from prior and joint probability:

$$P(Y \mid X) = \frac{P(Y,X)}{P(X)} \tag{12}$$

It can be shown that posterior probability can be calculated from conditional and prior probabilities:

$$P(Y \mid X^1, X^2, ..., X^n) \propto P(Y) \prod_k \left(\frac{P(Y \mid X^k)}{P(Y)} \right) \tag{13}$$

where X^k is the k^{th} predictor variable ($k = 1 .. n$).

For simplicity the left-hand side of equation 2, the posterior probability distribution $P(Y \mid X^1, X^2, ..., X^n)$ can be written as $(y_1, y_2, \ldots, y_m)$, $\Sigma y_j = 1$, where y_j is the probability that the response variable y will be in class j.

Score

The score metric is a weighted average of $y_1, y_2, \ldots, y_m$, devised as a way of displaying the entire probability distribution in summary in one map. The assumption is that the classes are ordinal, and class j is ranked higher than class $j - 1$ ($2 \leq j \leq m$).

The score s is calculated as follows:

$$\begin{aligned} w_i &= \frac{i-1}{n-1} \\ s &= \sum w_i y_i \end{aligned} \tag{14}$$

where n is the total number of response classes, w_i is the weight for the i^{th} class and y_i is the posterior probability value of the i^{th} class.

For example, for a response variable with four categories and probability distribution (0.2, 0.4, 0.3, 0.1), score s = 0*0.2 + 1/3*0.4 + 2/3*0.3 +1*0.1 = 0.433.

Certainty

Each conditional probability distribution is assigned a certainty value of 'low', 'medium' or 'high'. When calculating posterior probability, these are assigned the values 0, 1 and 2 respectively, and simply averaged over predictor variables to produce a combined certainty value. In general, if there are few data points in the input data in a given predictor variable class, certainty for all locations falling in this class will be low.

Driving factors

Once a probability surface has been created, it can be further analysed to identify driving factors. Analysis of driving factors attempts to identify the variable classes that disproportionately contribute to high values in the probability surface (positive driving factors) and low values in the probability surface (negative driving factors).

A sample of size n is taken from a region of interest and sorted by response value so that three sets can be obtained:

N = the set of all elements in the sample (size n)
$Q1$ = the set of elements in the upper quartile, ranked on response (size $n(Q1) = n/4$)
$Q4$ = the set of elements in the lower quartile, ranked on response (size $n(Q4) = n/4$)

For each predictor variable, the following can be calculated:

$n(x_i)$ = the number of elements in N that are in category i for predictor variable x
$n(x_i, Q1)$ = the number of elements in $Q1$ that are in category i for predictor variable x
$n(x_i, Q4)$ = the number of elements in $Q4$ that are in category i for predictor variable x

Then category i for predictor variable x is considered a positive driving factor if:

$$\frac{n(x_i, Q1)}{n(Q1)} \Big/ \frac{n(x_i)}{n} \geq c \tag{15}$$

and is considered a negative driving factor if:

$$\frac{n(x_i, Q4)}{n(Q4)} \Big/ \frac{n(x_i)}{n} \geq c \tag{16}$$

where c (> 1) is a user-defined threshold, with default value of 2.0.

Although the default is upper quartile and lower quartile (25%), this value can also be user-defined. For example, if there are $n = 100$ locations in the sample, of which $n(x_i) = 20$ are in predictor variable class i, and there are $n(Q1) = 25$ locations in the upper quartile, of which $n(x_i, Q1) = 15$ are in predictor variable class i, then the left-hand side of equation 15 evaluates to 3.75 and class i is therefore a positive driving factor.

Homologue analyses

FloraMap™ was developed to predict where wild plant germplasm could be expected to grow satisfactorily. It is an algorithm for mapping the distribution of plants and other organisms in the wild. It works on the premise that we know nothing about the organism other than the geographic location of a set of points where it was collected in the wild. From these calibration data we fit a climate probability model. This approach has had considerable success, and is being used widely. However, it has some major drawbacks for many applications: it requires a calibration set, it only works on climate, and it has not been used successfully on cultivated crops where the farmer alters the environment.

So, what do we do for those who ask the simple question, "Where else in the world is like my plot of land?" We have no calibration set. We do not know what species we are considering. We do not have an algorithm for predicting the probabilities of relevant soil characteristics. The question may be simple, but the answer is not. Homologue has been developed to cope with the complexities of this simple question. Homologue uses the basic algorithm of FloraMap, generalized to fit a range of generic species designated by the user. It incorporates statistical probability calculations for the mapping of soil characteristics. If we know where else in the world is like my plot of land, we can infer, from the agricultural practices there, what may be applicable to my plot.

The Homologue extension of the FloraMap algorithm (Jones *et al.* 1997; Jones and Gladkov 1999) is in two parts. The first generalizes the FloraMap algorithm so that it can generate a climate probability distribution from a single point, the second incorporates the probability of finding a soil with characteristics defined by the user. Note that FloraMap did not include soil characteristics because it was impossible to draw conclusions about the soil on which a given organism in the calibration set was found. Homologue relies on the investigator having enough knowledge of the point s/he is trying to match to be able to provide data on soils.

Spatially-interpolated climate surfaces are now available for many areas. These usually handle long-term climate normals interpolated over a digital elevation model (DEM) by various methods (Jones 1991, Hutchinson 1997). Pixel size depends on the underlying elevation model. It may be as little as 90 m (Jones 1996), which results in a massive data set, or 10 minutes of arc (about 18 km), which is as large as is practicable in many instances. In the latter case, the normal elevation model is the National Oceanographic and Atmospheric Administration (NOAA) TGPO006 (NOAA 1984). We have produced interpolated data sets at CIAT for the tropics using data from about 10,000 stations for Latin America, 7000 for Africa and 4500 for Asia. Each set of surfaces consists of the monthly rainfall totals, monthly average temperatures, and monthly average diurnal temperature range. This gives 36 climate variates in three groups of 12.

We use a simple interpolation algorithm based on the inverse square of the distance between the station and the interpolated point. For each interpolated pixel we find the five nearest stations. Then the inverse distance weights are calculated and applied to each monthly value of the data type being interpolated. Thus, for five stations with data values x and distances from the pixel distance d:

$$x_{pixel} = \sum_{1=1}^{5} d_i^{-2} \times \sum_{1=1}^{5} \frac{x_i}{d_i^{-2}} \tag{17}$$

Temperature data are standardized to the elevation of the pixel in the DEM using a lapse rate model (Jones, 1991). Using this simple interpolation has various advantages. First, it is the fastest of all the common methods. Second, it puts the interpolated surface exactly through each station point, because the weight *1/(d(I)**2)* becomes infinite as d approaches zero. Third, the interpolation is highly stable in areas of sparse data. It approaches the mean of the nearest stations when they all become equally distant. Fourth, it is relatively stable against errors in station elevation; only the local region of that station is affected. On the other hand, laplacian spline techniques and co-Kriging both propagate these errors more extensively. This is one advantage of using a proven lapse rate model instead of fitting a local one, as do both of these latter techniques.

The climatic events that occur through the year, such as summer/winter and start/finish of the rainy season, are of prime importance when comparing one climate with another. Unfortunately, they occur at different dates in many climate types. The most obvious case is where climates are compared between points in the Northern and Southern Hemispheres, but more subtle differences can be seen in climate event timing throughout the tropics. What we need is a method of eliminating these differences to allow us to make comparisons free of these annual timing effects.

The FloraMap probability algorithm is based on a principal components analysis of a large set of geographic coordinates from a germplasm, museum, or other collection where the original collection points have been noted. It works for just about any organism in the wild where the distribution is influenced mainly by climate (a very common occurrence). The algorithm can be seen as two separate parts. A principal components analysis that breaks down the climate data into orthogonal components, and a probability calculation, from these components, that compares any given pixel on a map to the fitted probability model.

The operation can be illustrated in two dimensions as follows. Figure 11 shows a scatterplot of two variates, x and y, quite highly correlated and therefore not at all independent. For any change in x, we would expect a change in y. However, we can find two new axes, α and β, such that they are not correlated, and that the variance accounted for in the first of the new axes is maximized.

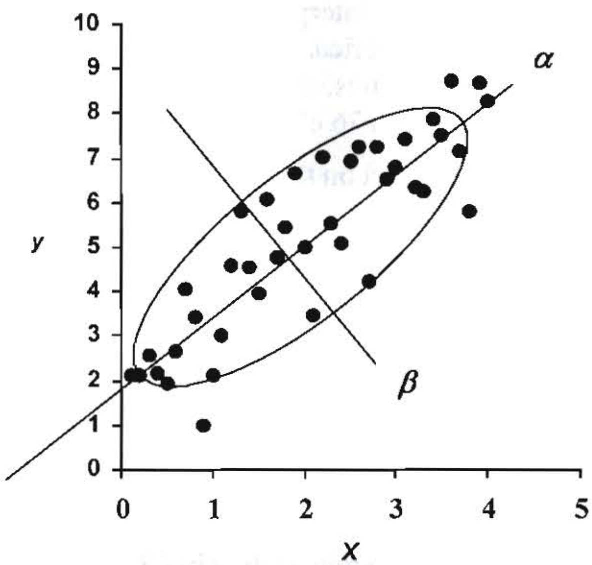


Figure 11. Illustrating the rotation of correlated variables x and y to the orthogonal variables α and β.

In this case, $\alpha = 0.454x + 0.891y$, and $\beta = 0.891x - 0.454y$. These new axes are orthogonal and uncorrelated. Movement along the α axis does not imply any movement at all along the β axis. The component α accounts for 95.6% of the original variance, β merely 4.4%. The trick to this linear transform is to calculate the eigenvalues and eigenvectors of the variance-covariance matrix of the system of variates. In FloraMap's case, this is a 36 x 36 matrix of climate variates.

In matrix notation, we need to find a matrix **Q** and a diagonal matrix **Λ** such that:

$$\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q} = \operatorname{diag}\lambda = \Lambda \tag{18}$$

Where **A** is our variance-covariance matrix. The matrix **Λ,** composed of the elements λ, is the diagonal matrix of the eigenvalues, which in our case hold the variance of the eigenvectors. The matrix **Q** is a symmetric matrix, which holds the eigenvectors as both rows and columns. The eigenvectors have two highly useful properties, one of which has been mentioned above—they are linearly independent of each other. The second useful property is that an eigenvector multiplied by any scalar is still an eigenvector.

A principal components analysis (PCA) can be performed on the sums of squares and cross products (SSCP) matrix, the variance-covariance matrix, or on the correlation matrix of a group of variates. In FloraMap, we use the variance-covariance matrix by standardizing the variates before we calculate the SSCP. But, we differ from many standard analyses in that our data have a structure that we want to preserve rather than standardize completely. The data are actually three groups of 12 values for different climate variables—rainfall, temperature, and diurnal temperature range. We want to conserve this difference to allow the user to apply weight across the board for the climate variables, for example, by increasing the importance of rainfall over that of temperature. In addition, the information across the 12 monthly values is of critical interest and we do not wish to standardize it away. We therefore standardize all rainfall values by the common variance for rainfall, and so forth.

Once we have found the matrices **Λ** and **Q**, we can describe the system of climate variates in terms of the principal components and their variances (eigenvectors and eigenvalues). We can choose a subset of the components (because the eigenvectors are independent), and we can scale them individually (because multiplying or dividing by a constant does not change the eigenvector's properties). This last point is important because this is exactly what we want to do to calculate the probabilities.

The normal probability density function for a single variate is given by:

$$z = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \tag{19}$$

From the integral of this function we can estimate the probability of observing a point drawn from this population.

Traditionally, we look at the probability that a point might lie further from the origin than the point in question. Also, we usually estimate the distribution parameters from the sample that we are investigating. Because of this, we use other statistics such as Student's *t* test to estimate the probability. In FloraMap, we make a simplifying assumption that the calibration set for the germplasm in question will contain sufficient points so that estimating from the sample will be equivalent to knowing the population parameters. This would not be true for small calibration sets, and even less for the Homologue case of a single point. We therefore assume a large calibration set and use the variance supplied by the user in the form of the expected adaptation range.

Soils data for Homologue

There are two main sources of soils information that are uniform, compatible and world wide. These are the World Inventory of Soils Emission Potentials Database, WISE (Batjes and Bridges 1994, Batjes 1995), and the FAO Soils Map of the World at 1:5,000,000 (FAO 1995). The FAO soils map gives mapping units that include a number of different soil types. Although these are not mapped in the sense that we know where they are, there are basic rules as to what percentage of the soil unit each soil type covers (see FAO 1978). Within a climate pixel there may be varying proportions of a number of mapping units. Figure 12 and Figure 13 illustrate how the probabilities of encountering a given soil within a climate pixel are calculated.

The map legend (FAO 1974) gives descriptive and some quantitative characteristics of the soil types, but for quantitative data with some idea of the variance of each soil characteristic within a soil type we must turn to the WISE database. We extracted data for 11 measurable soil characteristics from over 3000 profiles in the database. These were transformed to normality where necessary, and we calculated means and variances for those with sufficient profiles within a soil type. This left considerable gaps in the table, and these were filled where possible by regression on known characteristics. Where even this failed, we used analogies with known soil types to complete the table. Fortunately, this was the case for only a few less common soils.

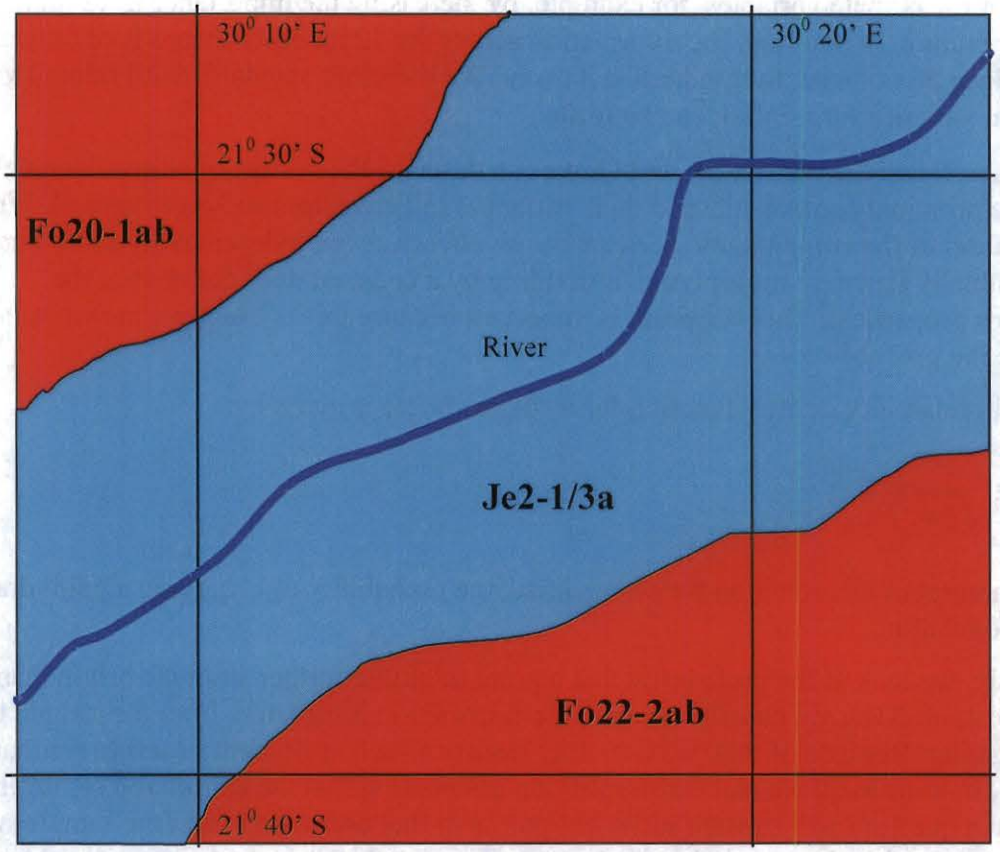


Figure 12. Schematic diagram to show a climate pixel with different proportions of three mapping units.

Fo20-1ab (10%)	=	Fo 70% Grade 2	⟶	Fo	17%
		I 30% Non Ag			
Je2-2ab (66%)	=	Je 70% Grade 1	⟶	Je	46.2%
		Zg 30% Non Ag			
Fo22-2ab (20%)	=	Fo 50% Grade 2			
		Af 20% Grade 2	⟶	Af	4%
		Ao 20% Grade 2	⟶	Ao	4%
		10% Non Ag			
River (4%)	=	Non Ag			

Figure 13. Schema for calculating probability areas of individual soils within the illustrated climate pixel. The left hand column shows the pixel percentages for each mapping unit. The center column gives the member soils and their nominal percentages in each mapping unit. The result is shown on the right.

Once we have the variances, we can calculate the probability that a given characteristic falls above, below, or between two values. However, what we need is the probability of encountering this condition within the climate pixel, and this involves combining the probabilities of finding the characteristic within a soil type, and the probability of finding the soil in the pixel. Thus we need to construct, for each pixel, the probability integral over all soils, for each soil characteristic. Figure 14 shows how this is done.

Each probability integral is then scaled and compressed into eight bytes, and stored in a composite structure including all the soil characteristics for each pixel.

Calculating the soil probability in Homologue

The soil characteristics taken from the WISE database files are shown in Table 3. Although a number of important properties are missing from this list, notably phosphorus content, it was deemed expedient to get Homologue working as a demonstration model with at least a viable list of quantitative soil characteristics. Those that are more difficult to standardize and obtain sufficient data for can be added later as the data become available.

As Figure 15 shows, although some soil characteristics are distributed relatively independently between the 3000 profiles, many are not. This is hardly surprising between pH measures of different soil types. We wish, however, to create a generalized algorithm that can cope with ANY selection of soil parameters that can be entered in the future. Therefore we must cope with the vagaries of the cross correlations. The simple way to do this is to extract the eigenvalues and eigenvectors of this correlation matrix. However, the user can select any subset of these characteristics (or none at all). We therefore have to delay taking the eigenvectors and eigenvalues until that selection has been made and then take them from the subset of the correlation matrix at the time of the analysis. This yields a probability problem in a varying number of dimensions depending on the set chosen, and the problem can be solved in exactly the same way as in the climate section, using Equations 17 and 19.

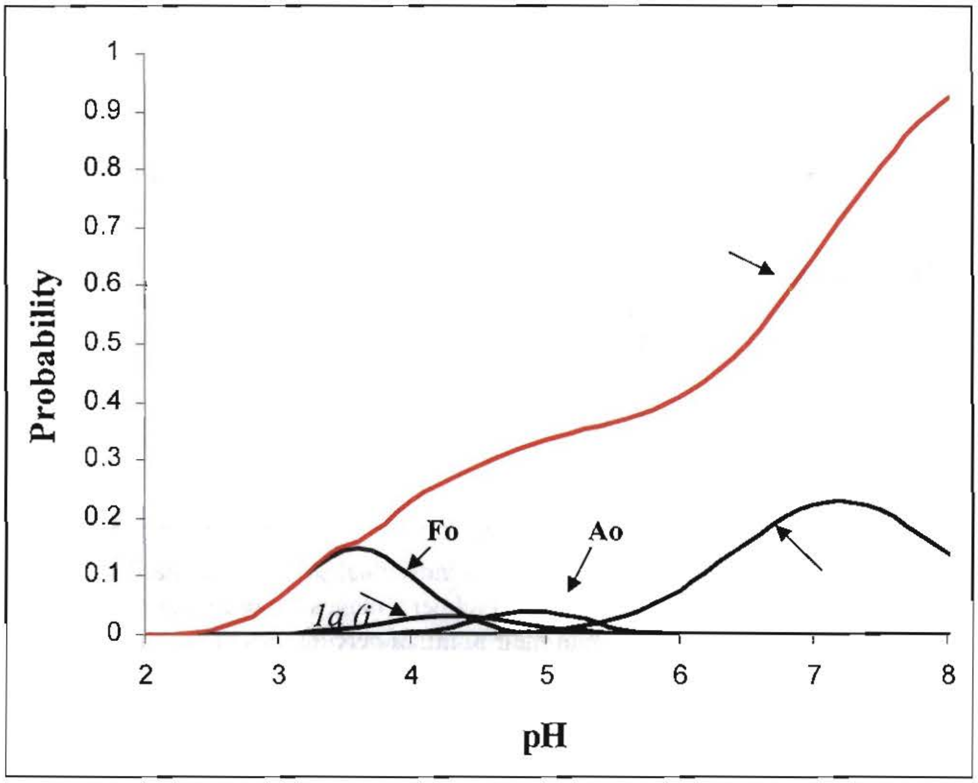


Figure 14. Illustration of how the probability densities of individual soils in Figures 7 and 8 are scaled and added to produce the overall pixel probability integral.

Table 3. Soil characteristics taken from the WISE database files.

Characteristic	Details
Depth	Soil depth to the C-horizon.
C	Total soil carbon content.
N	Total soil nitrogen
pH	pH in water
pH KCl	pH in potassium chloride
pH $CaCl_2$	pH in calcium chloride
CeC	Cation exchange capacity
Sand	Percentage sand
Silt	Percentage silt
Clay	Percentage clay
RD	Rooting depth

Depth	1.00										
C	-0.10	1.00									
N	-0.13	0.89	1.00								
pH	-0.39	-0.34	-0.27	1.00							
pH KCl	-0.33	-0.19	-0.14	0.92	1.00						
pH $CaCl_2$	-0.39	-0.30	-0.27	0.97	0.90	1.00					
CeC	-0.34	0.66	0.70	0.11	0.12	0.10	1.00				
Sand	-0.11	-0.41	-0.31	-0.03	-0.13	-0.02	-0.53	1.00			
Silt	-0.12	0.11	0.09	0.36	0.40	0.37	0.32	-0.64	1.00		
Clay	0.26	0.42	0.32	-0.32	-0.23	-0.33	0.37	-0.67	-0.14	1.00	
RD	0.35	-0.07	-0.09	-0.02	0.01	-0.02	-0.06	-0.13	0.12	0.05	1.00

Figure 15. The correlation matrix for 11 selected soil characteristics.

One further complication remains before we can come to combining the probabilities. Homologue allows the user to choose between using the actual probability of a characteristic falling between limits, and the mere fact that the probability exceeds a threshold value. In the latter case, the probability is evaluated and set to 1 if it exceeds the threshold, or to 0 if it falls below the threshold. In this manner, the orthogonalization and combination algorithm can operate transparently on probabilities regardless of the option selected.

Probability combination proceeds by forming the probit transforms of the probabilities (substituting a very small number for 0 and 0.99999 for 1). We then use the probits to calculate the orthogonal scores from the eigenvalues and eigenvectors; from these we calculate a radial distance in N-space, where N is the number of characteristics selected and apply Equations 14 or 16 to determine the overall probability of finding a soil with the selected characteristics in the given ranges.

5c (iii). Approach to regional analyses

In contrast to the preliminary study, which covered the two adjacent Departments in southern Colombia, Cauca and Nariño, the present study covered a broad swathe of eastern central Colombia. It included six Departments, southern Tolima and northern Huila, and Magdalena, César (grouped with areas from south-western Guajira by the FNC), Santander, and Santander Norte. These were grouped into two sets for data analysis, on a regional basis Tolima / Huila Region and the Northern Region comprising Magdalena, Cesar, Santander, and Santander Norte (Figure 16).

The primary reason for making the two Regional groupings was because the harvest in each is in distinctly different seasons (Figure 16). Furthermore, consultations with regional experts in the preparatory phase also revealed there were likely to be differences in the mix of varieties, especially more traditional varieties in the Northern Region, and differences in agronomic and post-harvest management.

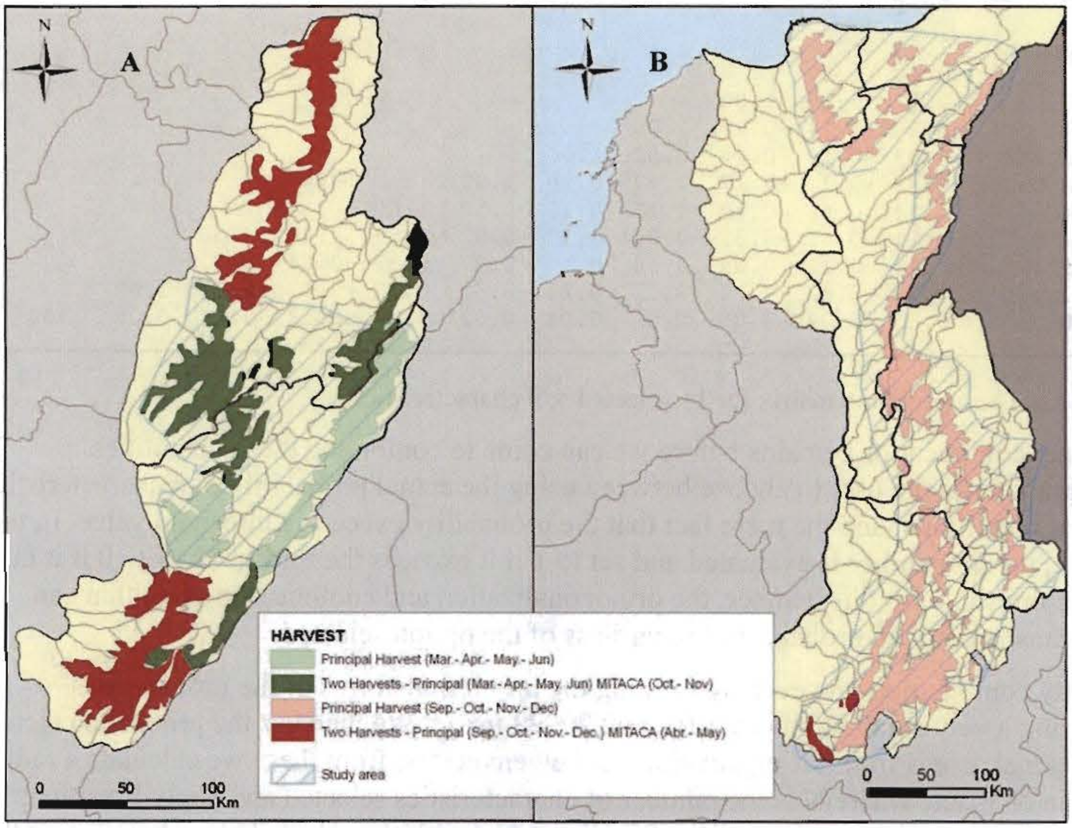


Figure 16. Maps showing the harvest seasons in (A) Huila/Tolima Region, and (B) Northern Region, which includes the four Departments of Magdalena, César, Santander, and Santander Norte.

5d. Environmental information

5d (i). Topography

Terrain attributes such as elevation, aspect and slope (Figure 17)were generated and mapped from the digital elevation model (DEM) of the shuttle radar topography mission (SRTM) using geographical information systems (GIS) methodology. The DEM is a raster file containing only spatial elevation data in a regular gridded pattern. The SRTM is a joint project between http://www.nasa.gov/http://www.nga.mil/ the National Geospatial-Intelligence Agency (NGA) and the National Aeronautics and Space Administration (NASA). In February 2000, the space shuttle produced, by radar interferometry, digital topographic data for 80% of the Earth's land surface. The data are of very good quality with 90 m resolution.

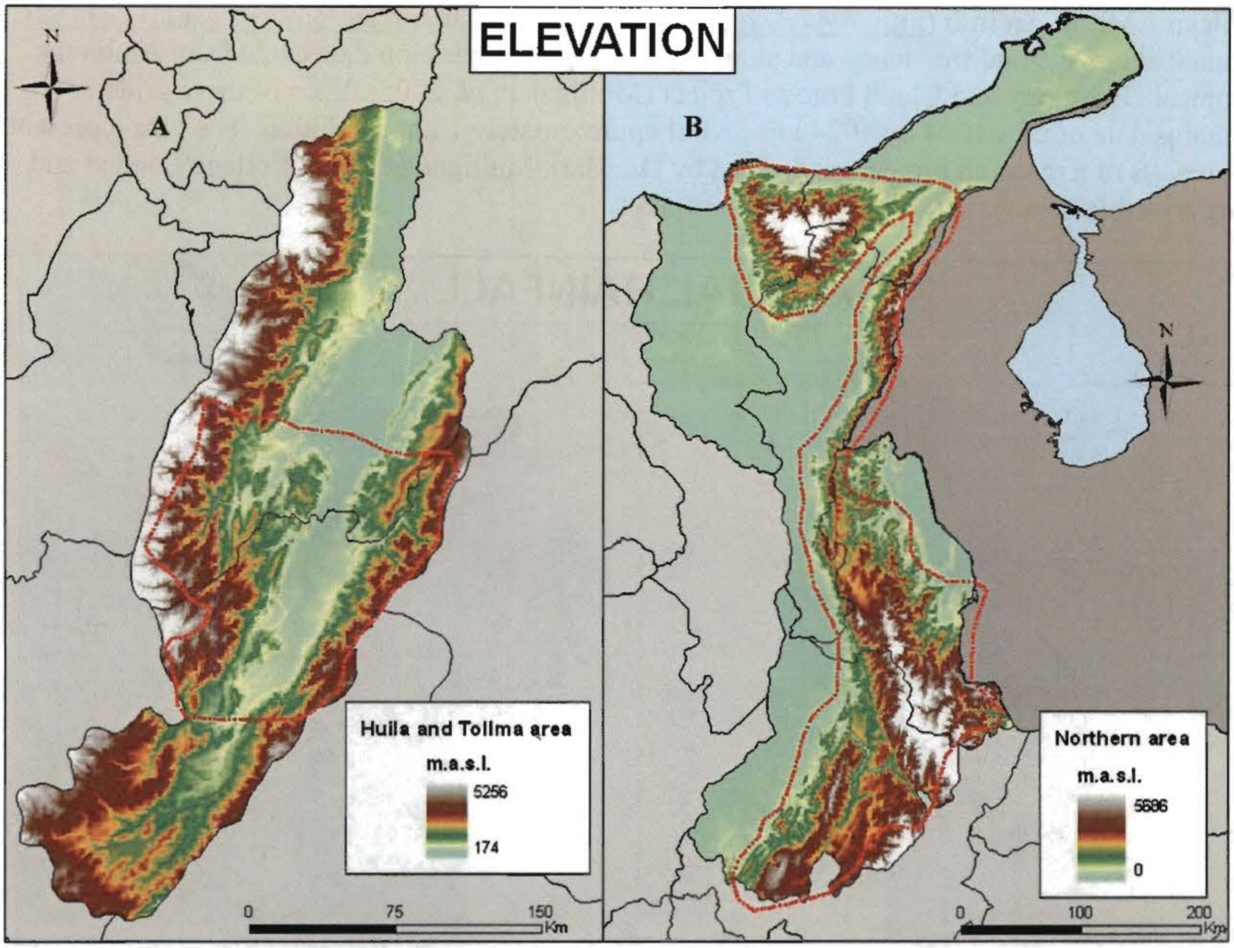


Figure 17. Elevation in (A) Huila/Tolima Region, and (B) Northern Region, which includes the four Departments of Magdalena, César, Santander, and Santander Norte.

5d (ii). Climate

Climate data were generated using WorldClim and MarkSim data. WorldClim (Hijmans *et al.* 2005) is a global database of climate variables in grid format with a spatial resolution (cell size) of 30 arc seconds, about 1 km at the Equator. The data layers were generated on 1 km^2 resolution through interpolation of average monthly climate data from 15,000 to 47,000 weather stations during the years1950 to 2000. Variables extracted from WorldClim were monthly total precipitation, and monthly mean, minimum, and maximum temperature. Annual average precipitation, annual average temperature, and dry month per year were then generated. Average annual precipitation (Figure 18) was obtained by summing all monthly total precipitations, average annual temperature by averaging the monthly mean temperatures, and dry months were defined as months with less than 90 mm of precipitation.

Annual average diurnal temperature range was calculated from WorldClim. As relative humidity varies diurnally and also between seasons, we mapped dew point. Dew point is the temperature at which air becomes saturated and produces dew and is a direct measure of the absolute amount of water vapour in the air. Dew point maps were calculated by the method of Linacre (Linacre 1977) from the WorldClim dataset. Mean annual insolation, which is the solar radiation that reaches the surface of the earth, was calculated from the SRTM topography data (MJ m^{-2} d^{-1})

with an AML in ArcInfo (http://www.wsl.ch/staff/niklaus.zimmermann/programs/aml1_2.html). Annual average cloud frequency and annual total evapotranspiration data came from Ambiotek Tropical Hydrology and Cloud Forests Project (Mulligan *et al.* 2005). Each of these grids contains data of tiles 1024 by 1024 km eachof approximately 1 km resolution. The data represent the results of a research project carried out by Dr. Mark Mulligan at Kings College London and are derived from a variety of original data sources.

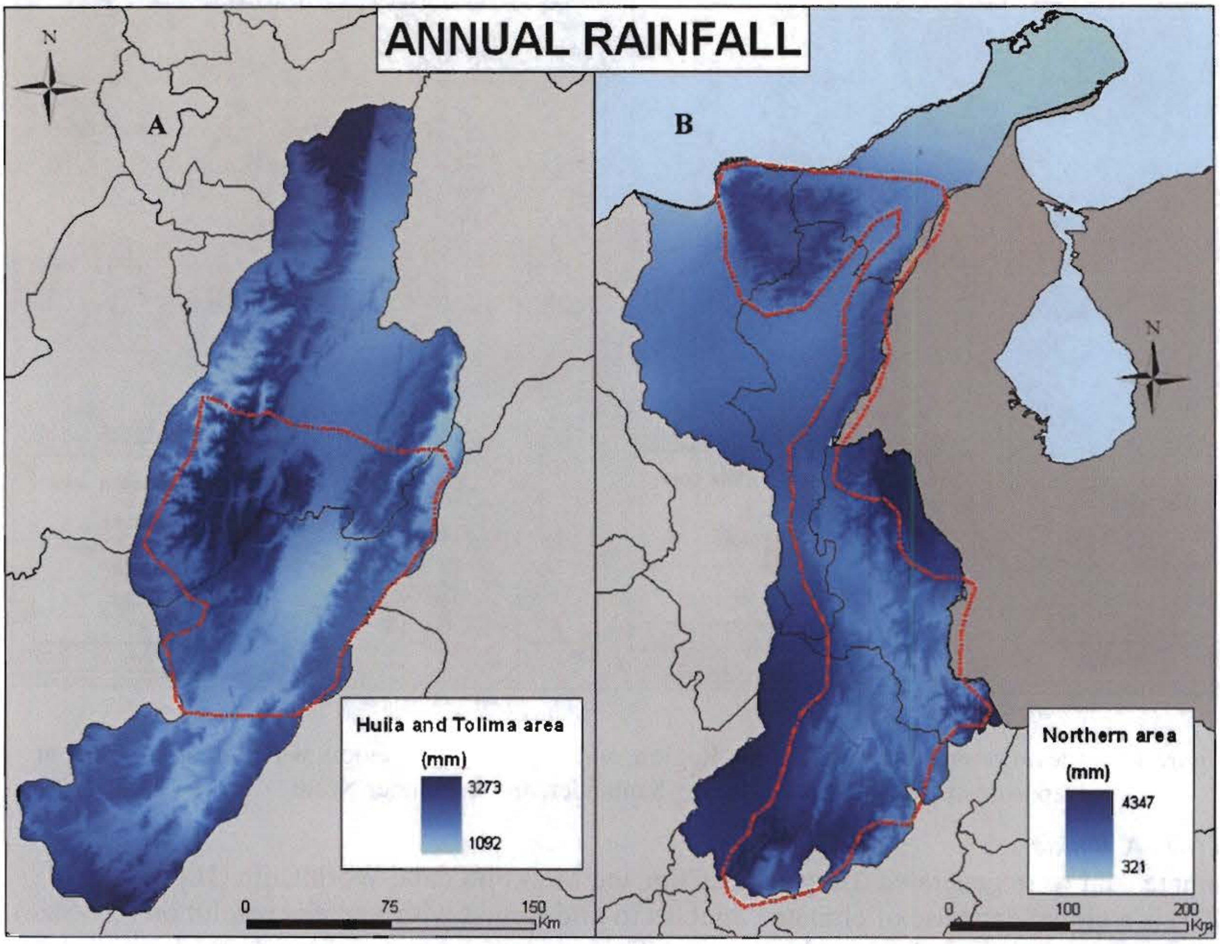


Figure 18. Mean annual precipitation in (A) Huila/Tolima Region, and (B) Northern Region, which includes the four Departments of Magdalena, César, Santander, and Santander Norte.

Rain days per year were estimated by Marksim using the WorldClim data as input for each cell. MarkSim uses a third-order Markov function to generate rainfall data(Jones and Thornton 2000, Jones *et al.* 2002). Annual average diurnal temperature range was calculated from WorldClim. As relative humidity varies diurnally and also between seasons, we mapped dew point. Dew point is the temperature at which air becomes saturated and produces dew and is a direct measure of the absolute amount of water vapour in the air. Dew point maps were calculated by the method of Linacre (Linacre 1977)from the WorldClim dataset. Mean annual insolation, which is the solar radiation that reaches the surface of the earth, was calculated from the MarkSim daily data (MJ m^{-2} d^{-1}) with an AML in ArcInfo (http://www.wsl.ch/staff/niklaus.zimmermann/programs/aml1_2.html).

5d (iii). Soils

Fertility analyses (pH, organic matter, P K, Ca, Mg, Al, Fe, Mn, Zn, Cu, cation exchange capacity and texture) on 347 soil samples from Cauca and Nariño Departments were carried out in the Cenicafe laboratory according to standard methods (Table 4).

Table 4. Methods used for characterization of soil fertility.

Fertility characteristic	Method
pH	Potentiometric in water 1:1
Aluminium	Yuan – atomic absorption
Organic matter	Walkey - Black colorimetric
Nitrogen	Calculated
Phosphorus	Bray II colorimetric
Potasium	Ammonium acetate - Atomic absorption
Calcium	Ammonium acetate - Atomic absorption
Magnesium	Ammonium acetate - Atomic absorption
Sodium	Ammonium acetate - Atomic absorption
Hierro	Ammonium acetate - Atomic absorption
Manganese	Ammonium acetate - Atomic absorption
Zinc	Ammonium acetate - Atomic absorption
Copper	Ammonium acetate - Atomic absorption
Cation exchange capacity	Ammonium acetate - Atomic absorption
Texture	Bouyoucos

5e. Product quality analyses

5e (i). Physical analyses

Physical assessment of samples of coffee samples was carried out using standard procedures. The protocol followed is detailed (in Spanish) in Appendix I.

5e (ii). Biochemical analyses

Bean samples were prepared using healthy ripe cherries collected during the harvest peak. The cherries were processed by the wet method (pulping, fermentation and drying detailed in Section 5b above) to obtain approximately 100 to 250 g of green coffee beans. Defective beans in the samples of green coffee were discarded.

Biochemical analyses of green bean samples were performed by near infrared spectroscopy (NIRS). NIR reflectance spectra were collected using a scanning monochromator NIRsystems spectrophotometer (model 6500, Perstrop Analytical Inc, 1201 Tech Road, Silver Spring, MD 20904, USA) driven by ISISCAN v.2.71 and the mathematical processing by WINISI III (v.1.50e) software (Intrasoft Intl., LLC, RD109, Sellers Lane, Port Matilda, PA 16870, USA). The analyses were performed on green coffee (3 g) after grinding to pass a 1.0 mm sieve. For each sample, a NIR spectrum was acquired in reflectance (R) mode, where R represented reflectance energy, in the 900–2500 nm range in 2 nm steps (Downey and Boussion 1996). The log (1/R) absorbance spectrum was obtained by the mean of these measurements and compared with the reference. The mean quadratic error estimated from two sub-samples (two distinct samplings of the same sample) based on the raw spectrum (log 1/R) was under 300 μabs; this error was below the manufacturer's specifications and indicated satisfactory repeatability of the spectral measurement. Given these results, a single spectrum was acquired per sample.

Data processing

For NIRS, the methodology used by Downey and Boussion (1996) was applied to all the samples. Chemometric processing consisted initially of a principal component analysis (PCA) based on second derivatives of the spectra on the 900 nm to 2500 nm segment. Chemometric processing then consisted of factorial discriminant analysis (FDA). For the experiment 63 principal components (PC), representing 99.99% of total variability, were used. Squared Mahalanobis distances were calculated basing the calculation on the coordinates of the individuals on the principal components, using SAS 8.0 software.

The prediction values (in percent of dry matter) for caffeine, trigoneline, chlorogenic acids, caffeoylquinic acid (CQA), sucrose, lipids, palmitic acid, stearic acid, oleic acid, linoleic acid, linolenic acid; araquidic acid, behenic acid were done using the calibration equations developed by Cenicafe from laboratory reference data. The quantification techniques used for caffeine and trigonelline were HPLC, for lipids: gravimetric, palmitic acid, stearic acid, oleic acid, linoleic acid, linolenic acid, araquidic acid, behenic acid: gas chromatography with an SID detector and for chlorogenic acid the UV-VIS detector was used.

5e (iii). Organoleptic analyses

Tasting, or cupping was carried out using standard procedures. The protocol followed is detailed (in Spanish) in Appendix 7.

5f. Overall approach to the analyses

The nethods described above were applied ro the data to achieve the following:

1. Identify the most reliable and consistent cuppers. Only results of these cuppers will be used in the further analyses.
2. Identify the most appropriate spatial analyses domain for which the relationships between coffee quality on one side, and environmental and production system characteristics on the other side are analyzed. Such domains should reduce as much as possible the environment by genotype interactions, in order to permit the generalization of a single quality profile for each identified domain so as to keep the denomination implementation procedures simple.
3. Understand the spatial relationships between coffee quality on one side, and environmental and production system characteristics on the other side for each identified domain.
4. Identify the most important environmental factors that impact on key coffee quality characteristics.
5. Provide recommendation as to how unique the identified spatial domains are if compared to other coffee growing regions.
6. Provide recommendations as to how the spatial domains can be used and potentially modified within the process of a denomination of origin.

6. Results

6a. Are the cuppers consistent?

In the preliminary study in the Cauca and Nariño Departments, there were problems with the sensory analysis in that the only 33 to 61% of almost 500 samples were correctly scored as determined by discriminant analysis. In this regard, trained sensory specialists typically classify over 80% of samples correctly when their scores are submitted to discriminant analysis. As a result, we were forced to rely on the data of just one the sensory specialist.

Descriptive data from each of four cuppers for each region individually and and with all data including all varieties and both processing methods were tabulated in their original and standardized to a mean of 0 and a standard deviation of 1 (Table 5 and Table 6). The mean data are broadly similar for all four cuppers, but once again only cupper 21 was consistent.

Table 5. Descriptive statistics for all cupping data for the Huila and Tolima areas in their original form and analyzed as standardized data. The data are for the four cuppers that took part in all cupping sessions.

Variable	Original data					Standardized data				
	Mean	Min[1]	Med[2]	Max[3]	SD[4]	Mean	Min	Med	Max	SD
					Cupper 17					
Fragrance and aroma	5.74	1.0	6.00	9.0	1.75	-0.01	-2.82	0.14	1.92	1.03
Flavor	5.45	0.5	5.50	9.0	2.12	-0.02	-2.42	0.00	1.70	1.02
Acidity	5.57	1.0	5.75	9.0	1.93	-0.03	-2.41	0.07	1.76	1.01
Body	5.62	1.0	6.00	8.5	1.84	-0.04	-2.58	0.17	1.54	1.01
Sweetness	5.36	1.0	5.50	9.0	1.92	-0.04	-2.31	0.03	1.85	1.00
					Cupper 18					
Fragrance and aroma	5.78	1.0	6.00	9.0	1.75	0.01	-2.82	0.14	1.92	1.04
Flavor	5.47	1.0	6.00	9.0	2.11	-0.01	-2.17	0.24	1.70	1.02
Acidity	5.62	1.0	6.00	6.5	1.93	0.00	-2.41	0.20	2.02	1.01
Body	5.76	1.0	6.00	9.0	1.83	0.04	-2.58	0.17	1.82	1.01
Sweetness	5.47	1.0	5.50	9.0	1.93	0.02	-2.31	0.03	1.85	1.01
					Cupper 19					
Fragrance and aroma	5.82	1.0	6.00	9.0	1.77	0.03	-2.82	0.14	1.92	1.05
Flavor	5.51	1.0	6.00	9.0	2.07	0.01	-2.17	0.24	1.70	1.00
Acidity	5.61	1.0	6.00	9.0	1.95	-0.01	-2.41	0.20	1.76	1.02
Body	5.72	1.0	6.00	9.0	1.85	0.01	-2.58	0.17	1.82	1.02
Sweetness	5.50	1.0	6.00	9.0	1.98	0.03	-2.31	0.29	1.85	1.03
					Cupper 21					
Fragrance and aroma	5.72	1.0	6.00	9.0	1.48	-0.03	-2.82	0.14	1.92	0.88
Flavor	5.55	0.5	6.00	9.0	1.97	0.03	-2.42	0.24	1.70	0.96
Acidity	5.70	1.0	6.00	9.0	1.87	0.04	-2.41	0.20	1.76	0.97
Body	5.68	1.0	6.00	9.0	1.75	-0.01	-2.58	0.17	1.82	0.96
Sweetness	5.42	1.0	5.50	9.0	1.85	-0.01	-2.31	0.03	1.85	0.96

Table 6. Descriptive statistics for all cupping data for the Northern area in their original form and analyzed as standardized data. The data are for the four cuppers that took part in all cupping sessions.

Variable	Original data					Standardized data				
	Mean	Min	Med	Max	SD	Mean	Min	Med	Max	SD
					Cupper 17					
Fragrance and aroma	5.27	0.0	5.50	9.5	1.34	0.05	-3.62	0.21	3.00	0.94
Flavor	5.25	0.0	5.50	9.5	1.66	0.01	-2.94	0.14	2.39	0.93
Acidity	5.35	0.0	5.50	9.5	1.53	0.03	-3.19	0.12	2.52	0.92
Body	5.24	0.5	5.50	9.0	1.59	0.01	-2.70	0.16	2.16	0.91
Sweetness	5.28	0.0	5.50	9.5	1.57	0.02	-3.05	0.14	2.47	0.91
					Cupper 18					
Fragrance and aroma	5.12	0.0	5.00	9.5	1.50	-0.06	-3.62	-0.14	3.00	1.05
Flavor	5.20	0.0	5.00	10.0	1.84	-0.02	-2.94	-0.14	2.67	1.03
Acidity	5.26	0.5	5.50	10.0	1.71	-0.02	-2.89	0.12	2.82	1.03
Body	5.14	0.5	5.50	9.5	1.76	-0.05	-2.70	0.16	2.45	1.01
Sweetness	5.25	0.0	5.50	9.5	1.75	0.00	-3.05	0.14	2.47	1.02
					Cupper 19					
Fragrance and aroma	5.06	0.0	5.00	10.0	1.66	-0.10	-3.62	-0.14	3.35	1.16
Flavor	5.17	0.0	5.00	10.0	1.95	-0.04	-2.94	-0.14	2.67	1.10
Acidity	5.20	0.0	5.50	10.0	1.88	-0.06	-3.19	0.12	2.82	1.13
Body	5.17	0.0	5.00	10.0	1.94	-0.03	-2.99	-0.13	2.73	1.11
Sweetness	5.16	0.0	5.00	10.0	1.98	-0.05	-3.05	-0.15	2.76	1.15
					Cupper 21					
Fragrance and aroma	5.35	1.0	5.00	8.5	1.16	0.11	-2.93	-0.14	2.30	0.81
Flavor	5.35	1.0	5.50	10.0	1.65	0.06	-2.38	0.14	2.67	0.93
Acidity	5.40	1.0	5.00	10.0	1.51	0.06	-2.59	-0.18	2.82	0.91
Body	5.34	1.0	5.50	10.0	1.68	0.07	-2.41	0.16	2.73	0.96
Sweetness	5.32	1.0	5.00	10.0	1.55	0.04	-2.47	-0.15	2.76	0.90

We carried out discriminant analyses (Table 7 and Table 8) with both the original data and the standardized data, both of which gave the same results. In the ideal case, applying, for example, the discriminant function derived for cupper 17 to the data would classify all predicted samples as cupped by cupper 17. Lower numbers of correctly classified samples results in lower percentages of cupping accuracy and is an indicator for low cupping consistency by the respective cupper.

Table 7. The number of correctly classified samples when the linear discriminant functions were used with the original data for each of the four cuppers to group the samples.

	Number and % of samples correctly predicted from cupper to cupper									
Cupper	17	%	18	%	19	%	21	%	TOTAL	%
17	168	22.76	212	28.73	108	14.63	250	33.88	738	100
18	141	18.95	311	41.80	126	16.94	166	22.31	744	100
19	122	16.58	251	34.10	169	22.96	194	26.36	736	100
21	126	16.94	149	20.03	95	12.77	374	50.27	744	100
TOTAL	557	18.80	923	31.16	498	16.81	984	33.22	2962	100

Table 8. The number of correctly classified samples when the linear discriminant functions were used with standardized data for each of the four cuppers to group the samples.

	Number and % of samples correctly predicted from cupper to cupper									
Cupper	17	%	18	%	19	%	21	%	TOTAL	%
17	111	14.27	159	20.44	234	30.08	274	35.22	778	100
18	82	10.54	191	24.55	245	31.49	260	33.42	778	100
19	90	11.57	173	22.24	270	34.70	245	31.49	778	100
21	89	11.44	136	17.48	217	27.89	336	43.19	778	100
TOTAL	1115	35.83	372	11.95	659	21.18	966	31.04	3112	100

The conclusion of this analysis is that only up to 50 percent of the over 3100 samples were correctly scored. In this regard, trained sensory specialists typically classify over 80% of samples correctly when their scores are submitted to discriminant analysis (Table 9). Of the four cuppers here only cupper 21 demonstrated acceptable cupping consistency and therefore all subsequent analyses were done on scores of cupper 21.

Table 9. The number of correctly classified samples from a BENCHMARK cupping that we conducted in our research work recently for some of our cuppers. As above, linear discriminant functions were used for each of the cuppers to group the samples.

Cupper	Number and % of samples correctly predicted from cupper to cupper							
	1	%	2	%	3	%	Total	%
1	175	87.7	13	6.5	13	6.5	201	100
2	20	7.1	243	86.2	19	6.7	282	100
3	35	11.3	54	17.4	222	71.4	311	100
Total	230	29.0	310	39.0	254	32.0	794	100

6b. Are the encountered varieties different?

6b (i). What varieties do we find, where?

Unlike the preliminary study in Cauca and Nariño Departments, in this extensive survey we found a substantial number of sites where growers produce the Tipica variety (Table 10). Specifically, in the northern region the Typica variety comprises 25-35% of the total crop. The Caturra variety dominates in most Departments, except in Santander and Santander Norte. In the northern part of the region surveyed, Colombia and Caturra varieties dominate in Santander and Santander Norte Departments. There is a noteworthy clustering of the Colombia and Tipica varieties in the Santander Norte and in Huila and Tolima Departments in the south eastern region. In total, seven different varieties were encountered (Table 10).

In general in the northern region, Caturra and Tipica varieties dominate, albeit without clear clustering at the local level. In contrast, in the Huila/Tolima region, there is a clear cluster of Tipica in the south-eastern part of Tolima, and another cluster, albeit less clear, with the variety Colombia in the south-western part of Tolima.

Table 10. The varieties of coffee grown in the sites surveyed and the number of sites in each of six Departments of Colombia with each variety.

Variety/Department	Huila	Tolima	Santander	Santander Norte.	César	Magdalena	Total
Caturra	125	100	55	26	27	36	369
Colombia	85	65	134	40	14	13	351
Tipica	19	32	5	19	24	22	121
Maragojipe	0	0	0	0	0	1	1
San Bernardo	1	1	0	0	0	1	3
Tabi	0	0	0	0	2	0	2
Castillo	0	1	3	0	1	0	5
Total	230	199	197	85	68	73	852

Only the dominant varieties, Caturra, Colombia and Tipica, will be included in the analyses. Figure 18 and Figure 20 show the distribution of these three dominant varieties within the two geographical regions.

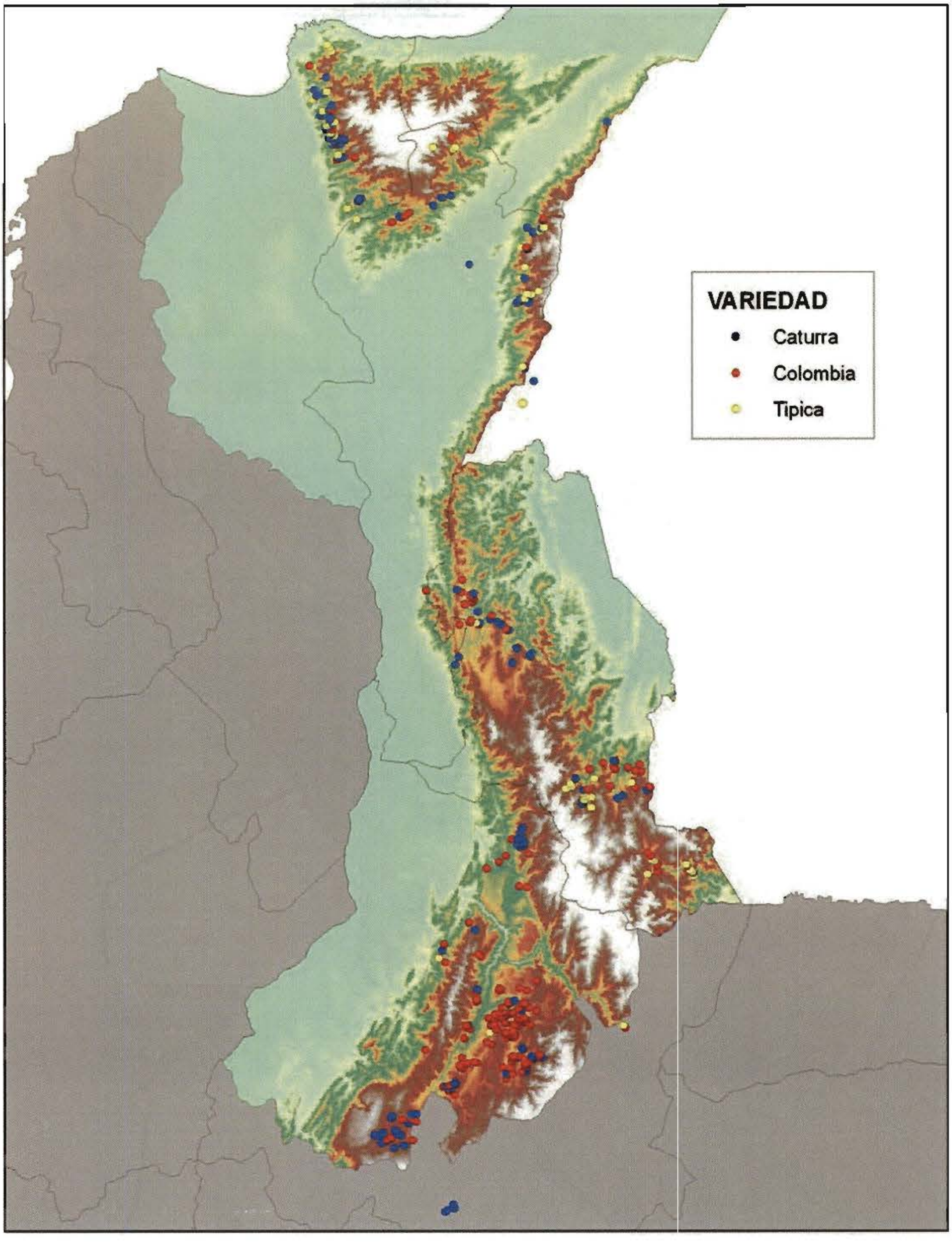


Figure 19. Distribution of the three dominant varieties, Caturra, Colombia, and Tipica, in the Northern Region.

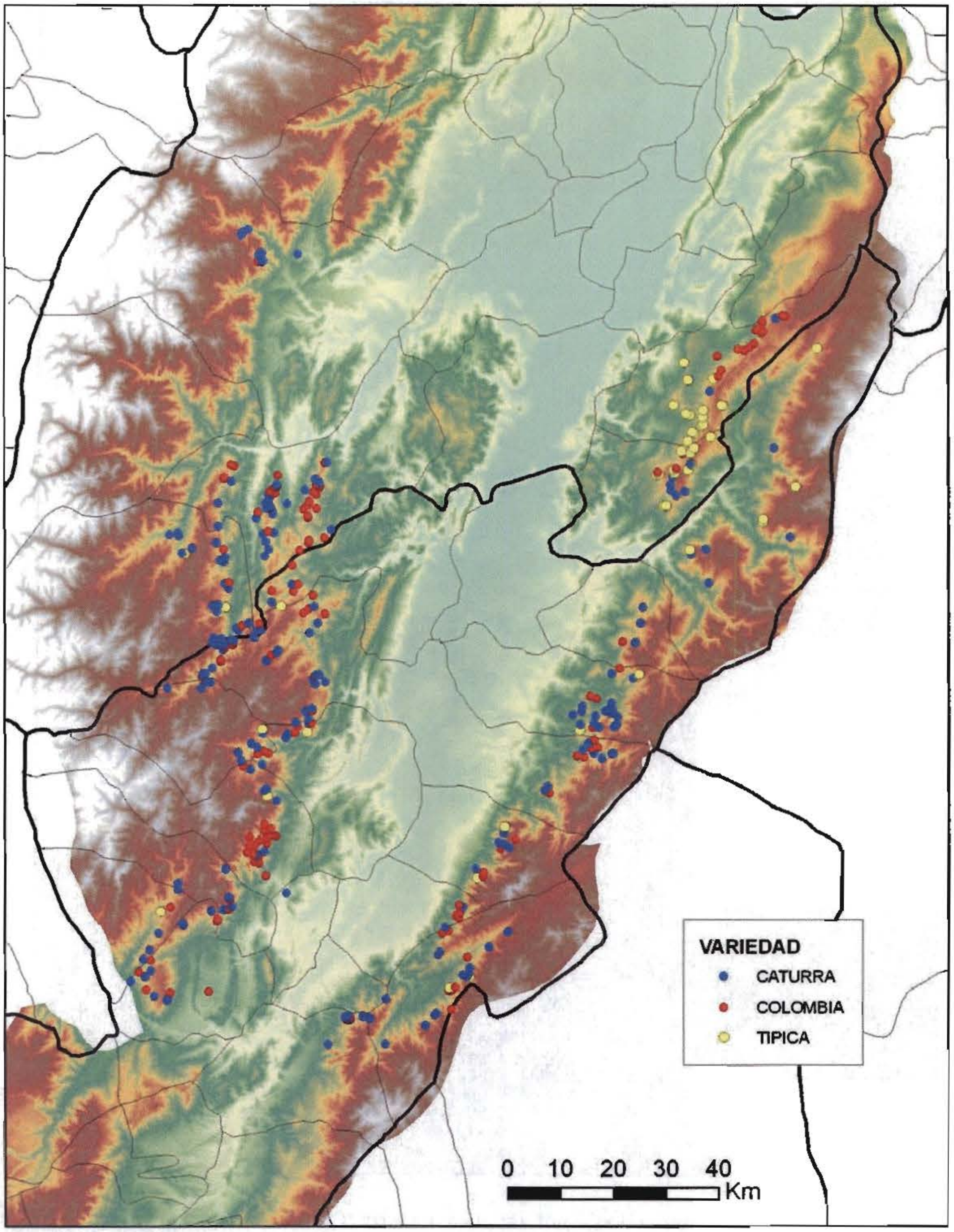


Figure 20. Distribution of the three dominant varieties, Caturra, Colombia, and Tipica, in the Huila/Tolima region.

6b (ii). Bean characteristics for each Department (physical, biochemical, sensory)

In Huila Department, as expected, Colombia variety has the highest percentage of grain in the larger size classes, contrasted with Tipica, which has the lowest percentage of grain in these classes (Table 11). Yield factors become very unsatisfactorily for large size classes for Tipica. In contrast, variety Colombia maintains stable yield factors across size classes. In contrast to its performance in Huila Department, the Tipica variety in Tolima achieves reasonable yield factors due to its good performance in size classes 16 and 17.

Table 11. Means of physical characteristics of the Caturra. Colombia, and Tipica coffee varieties grown in six Colombian Departments. Table A1.1 – Table A1.6 in Appendix 1 have a complete description of the data.

Screen size	Variety	Huila	Tolima	Santander	Santander Norte	César	Magdalena
12	Caturra	1.2	0.5	0.3	0.7	0.3	0.2
	Colombia	0.6	0.6	0.4	1.0	0.6	0.4
	Tipica	3.2	0.7	0.3	0.3	0.3	0.3
13	Caturra	3.6	2.3	1.3	3.2	1.9	1.5
	Colombia	2.3	1.8	1.6	4.1	2.7	2.2
	Tipica	6.7	2.3	1.0	1.4	1.8	1.4
14	Caturra	10.5	8.4	6.9	9.6	6.8	6.3
	Colombia	7.5	5.9	7.3	13.2	10.6	9.9
	Tipica	13.5	8.0	5.5	7.2	7.1	6.4
15	Caturra	28.6	23.0	17.9	25.8	20.2	19.2
	Colombia	21.3	17.6	18.9	29.4	27.0	24.3
	Tipica	31.9	25.2	15.4	18.1	18.5	17.1
16	Caturra	58.7	61.8	59.3	59.7	60.6	61.6
	Colombia	52.9	47.7	55.3	58.6	65.5	61.1
	Tipica	60.9	67.7	62.0	63.7	52.1	55.9
17	Caturra	59.6	64.8	72.0	63.9	71.5	74.9
	Colombia	60.9	62.2	66.5	59.7	61.5	68.6
	Tipica	52.4	64.4	79.8	78.0	72.5	72.5
18	Caturra	37.5	38.6	46.9	41.0	44.6	43.2
	Colombia	57.0	61.3	55.1	36.5	38.7	41.9
	Tipica	29.4	29.3	42.9	38.5	54.1	49.1
Yield factor							
YF 13	Caturra	86.6	86.1	85.4	86.3	85.2	84.7
	Colombia	85.6	85.6	85.2	86.6	85.0	84.6
	Tipica	88.7	86.8	84.8	84.7	84.2	84.5
YF 14	Caturra	88.4	87.1	85.9	87.7	86.0	85.3
	Colombia	86.6	86.4	85.9	88.7	86.1	85.5
	Tipica	97.8	87.8	85.2	85.3	85.0	85.1
YF 15	Caturra	94.3	91.2	89.0	92.5	89.1	88.1
	Colombia	90.3	89.3	89.3	95.9	91.1	89.9
	Tipica	128.9	91.7	87.6	88.5	88.3	87.9

In Santander Department, Tipica variety achieves the best yield factors due to its good performance in size classes 16 and 17. The Colombia variety has highest values in the size class 18, but is not as good in the other Departments. In Santander de Norte, the Tipica variety achieves the best yield factors compared with its performance in other Departments. The Caturra variety has the highest values in size class 18. In contrast, the Colombia variety performs poorly in Santander Norte. As in César Department, the Tipica variety achieves the best yield factors and has highest values in the large size classes in Magdalena Department. The Colombia variety also performs poorly in the Magdalena Department.

The Bonferoni multivariate tests draw together the results of all the forgoing information for the different Departments within the two regions (Table 12 and Table 13). In summary:

- Colombia and Caturra perform excellently in Tolima, Huila, and Santander. Tipica performs reasonably well in Tolima, and well in Santander and Santander Norte. Tipica is really outstanding in César and Magdalena.
- Varieties do differ from one another within the same Department, except in Huila and Santander (as far as the size yield factors are concerned), although this may be partly because there is very little Tipica grown in these Departments
- In the Huila /Tolima Region, varieties do not differ much, however, the same varieties show important differences in many places in the Northern Region.

Table 12. Bonferoni multivariate tests for selected physical characteristics, omparing the coffee varieties Caturra. Colombia, and Tipica within each of six Departments in Colombia.

Department	Huila			Tolima			Santander			Santander Norte			César			Magdalena		
Size/ Yield factor	Cat[1]	Col[2]	Tip[3]	Cat	Col	Tip	Cat	Col	Tip	Cat	Col	Tip	Cat	Col	Tip	Cat	Col	Tip
Size 17	A[4]	A	B	A	A	A	B	B	A	B	B	A	A	B	A	A	B	A
Size 18	B	A	B	A	B	C	A	A	A	A	A	A	A	A	A	B	B	A
YF 13	A	A	A	AB	B	A	A	A	A	A	A	B	AB	A	B	A	A	A
YF 14	A	A	A	AB	B	A	A	A	A	A	A	B	AB	A	A	A	A	A
YF 15	A	A	A	AB	B	A	A	A	A	A	B	C	B	A	B	B	A	B

[1] Cat = Caturra variety; [2] Col = Colombia variety; [3] Tip = Tipicia variety.
[4] Within each physical characteristic for each Department, varieties with the same letter do not differ significantly (P<0.05).

Table 13. Bonferoni multivariate tests for selected physical characteristics, comparing the coffee varieties Caturra, Colombia, and Tipica between six Departments in Colombia, grouped into two regions.

Huila/Tolima Region						
Variety	Caturra		Colombia		Tipica	
Size/Yield factor	Huila	Tolima	Huila	Tolima	Huila	Tolima
Size 17	B[1]	A	B	A	A	A
Size 18	B	A	A	A	B	A
YF 13	A	A	A	A	A	A
YF 14	A	A	A	A	A	A
YF 15	A	A	A	A	A	A

Northern Region												
Variety	Caturra				Colombia				Tipica			
Size/Yield factor	San[2]	SnN[3]	Ces[4]	Mag[5]	San	SnN	Ces	Mag	San	SnN	Ces	Mag
Size 17	A	B	A	A	A	B	BA	A	A	A	A	A
Size 18	A	A	A	A	A	B	AB	AB	A	B	AB	A
YF 13	A	A	B	B	B	A	B	C	A	AB	B	B
YF 14	B	A	BC	C	BC	A	B	C	A	A	A	A
YF 15	B	A	B	B	B	A	B	B	AB	A	AB	B

[1] Within each physical characteristic for each variety, Departments with the same letter do not differ significantly (P<0.05).
[2] San = Santander; [3] SnN = Santander Norte; [4] Ces = César; [5] Mag = Madalena.

The biochemical analyses (Table 14) are very interesting showing clear differences between varieties for some characteristics, and also clear differences between departments within the same variety, which demonstrates clear genotype by environment interaction. This is observable for chlorogenic acids, CQA total and sucrose. It is also interesting that Tipica variety tends to have higher values in both caffeine and trigonelline than other varieties. Furthermore, the values of both oleic acids and linolenic acids differ between the Huila / Tolima Region and the Northern Region.

The Bonferoni analyses show that the three varieties are particularly dissimilar in Tolima, Magdalena, Santander and Santander Norte Departments for most biochemical characteristics, while in Huila Department, they are dissimilar only for some characteristics (Table 15). Only in César Department are all three varieties similar in their biochemical characteristics.

There are also substantial differences between the departments for most biochemical characteristics, especially trigonelline, which is dissimilar between varieties in most Departments except in César.

Caturra variety tends to have more different values between Huila and Tolima Departments than do the other two varieties (Table 14 and Table 16). Tipica variety is has most similarities between the two Departments. Trigonelline, oleic acid and araquidic acid are always different between the two Departments. Comparing the Departments in the Northern Region, the three varieties tend to have fairly different biochemical characteristics (Table 14 and Table 16). The Tipica variety is the least dissimilar of the three within these Departments. Therefore, as seen for the physical characteristics, we find clear indications of genotype by environment interactions.

Table 14. Means of biochemical data of the Caturra. Colombia, and Tipica coffee varieties in six Departments of Colombia. Table A2.1 – Table A2.6 in Appendix 2 have a complete description of the data.

Biochemical charateristic	Variety	Huila	Tolima	Santander	Santander Norte	César	Magdalena
Caffeine	Caturra	1.4	1.3	1.3	1.3	1.3	1.4
	Colombia	1.4	1.3	1.4	1.3	1.3	1.4
	Tipica	1.4	1.5	1.4	1.4	1.4	1.5
Trigonelline	Caturra	1.0	0.9	0.9	1.0	0.9	0.9
	Colombia	0.9	0.8	0.9	1.0	0.9	0.9
	Tipica	1.1	1.1	1.0	1.1	1.0	1.0
Chlorogenic. acid	Caturra	6.7	6.7	6.5	6.8	6.8	6.5
	Colombia	6.6	6.6	6.4	6.8	6.8	6.4
	Tipica	6.7	6.2	6.6	6.6	6.6	6.2
CQA total	Caturra	5.6	5.9	5.6	5.8	5.8	6.0
	Colombia	5.6	5.8	5.6	5.7	5.8	5.9
	Tipica	5.6	5.6	5.7	5.7	5.8	5.7
Lipid	Caturra	17.8	17.6	18.3	19.3	18.8	18.2
	Colombia	17.3	16.4	17.9	19.4	18.1	18.0
	Tipica	17.6	17.3	19.4	19.8	18.7	18.3
Palmitic acid	Caturra	35.9	35.9	35.5	36.6	35.5	35.3
	Colombia	36.0	35.7	36.0	36.4	35.7	35.3
	Tipica	36.4	34.9	36.0	36.6	35.5	35.4
Estearic acid	Caturra	7.4	7.9	8.1	8.4	8.5	8.0
	Colombia	7.2	7.3	7.9	8.5	8.2	7.9
	Tipica	7.2	6.9	8.2	8.0	8.1	7.8
Oleic acid	Caturra	9.8	11.7	13.8	14.8	14.4	14.7
	Colombia	10.3	11.4	14.5	14.8	15.2	15.0
	Tipica	9.6	11.5	13.7	13.8	14.8	14.7
Linoleic acid	Caturra	41.1	40.5	39.7	37.6	38.6	38.9
	Colombia	40.6	40.8	38.6	37.6	38.0	38.5
	Tipica	40.6	40.8	38.6	38.8	38.3	38.4
Linolenic acid	Caturra	1.8	1.7	1.7	1.7	1.6	1.6
	Colombia	1.8	1.7	1.7	1.6	1.6	1.6
	Tipica	1.8	1.7	1.6	1.7	1.6	1.6
Araquidic acid	Caturra	2.1	2.3	2.1	2.2	2.4	2.4
	Colombia	2.2	2.3	2.2	2.4	2.4	2.5
	Tipica	2.2	2.4	2.4	2.3	2.5	2.6
Behenic acid	Caturra	0.5	0.5	0.5	0.4	0.5	0.5
	Colombia	0.5	0.5	0.4	0.5	0.5	0.5
	Tipica	0.5	0.6	0.5	0.5	0.5	0.6
Sucrose	Caturra	5.0	5.4	5.3	5.2	5.1	5.3
	Colombia	5.3	5.7	5.4	5.3	5.3	5.5
	Tipica	4.7	5.1	5.2	5.2	5.2	5.1

Table 15. Bonferoni multivariate tests for biochemical characteristics, omparing the coffee varieties Caturra. Colombia, and Tipica within each of six Departments in Colombia.

Department	Huila			Tolima			Santander			Santander Norte			César			Magdalena		
Biochemical characteristic	Cat[1]	Col[2]	Tip[3]	Cat	Col	Tip	Cat	Col	Tip	Cat	Col	Tip	Cat	Col	Tip	Cat	Col	Tip
Caffeine	A[4]	A	A	B	B	A	B	A	A	B	B	A	A	A	A	B	B	A
Trigonelline	B	C	A	B	B	A	B	B	A	B	B	A	A	A	A	B	B	A
Chlorog. acid[5]	A	A	A	A	A	B	AB	B	A	A	A	A	A	A	A	A	A	A
CQA total	A	A	A	A	A	B	A	A	A	A	A	B	A	A	A	A	B	B
Lipid	A	A	A	A	B	A	B	B	A	B	B	A	A	B	A	A	A	A
Palmitic acid	A	A	A	A	A	A	B	A	A	A	A	A	A	A	A	A	A	A
Estearic acid	A	A	A	A	B	C	A	A	A	AB	A	B	A	B	B	A	A	A
Oleic acid	B	A	B	A	A	A	B	A	AB	AB	A	B	B	A	AB	A	A	A
Linoleic acid	A	B	AB	A	AB	B	A	B	B	B	B	A	A	A	A	A	A	A
Linolenic acid	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Araquidic acid	A	A	A	B	B	A	B	B	A	A	A	A	A	A	A	B	AB	A
Behenic acid	A	B	A	B	B	A	A	A	A	A	A	A	A	A	A	B	B	A
Sucrose	B	A	C	B	A	C	A	A	B	A	A	A	A	A	A	B	A	B

[1] Cat = Caturra variety; [2] Col = Colombia variety; [3] Tip = Tipica variety.
[4] Within each biochemical characteristic for each Department, varieties with the same letter do not differ significantly ($P<0.05$).
[5] Chlorogenic acid

Table 16. Bonferoni multivariate tests for biochemical characteristics, comparing the coffee varieties Caturra, Colombia, and Tipica between six Departments in Colombia, grouped into two regions.

Huila / Tolima Region						
Variety	Caturra		Colombia		Tipica	
Biochemical characteristic	Huila	Tolima	Huila	Tolima	Huila	Tolima
Caffeine	A[1]	B	A	A	B	A
Trigonelline	A	B	A	B	B	A
Chlorog. acid[2]	A	A	A	A	A	B
CQA total	A	B	B	A	A	A
Lipid	A	A	A	B	A	A
Palmitic acid	A	A	A	A	A	A
Estearic acid	B	A	A	A	A	A
Oleic acid	B	A	B	A	B	A
Linoleic acid	A	A	A	A	A	A
Linolenic acid	A	B	A	A	A	A
Araquidic acid	B	A	B	A	B	A
Behenic acid	A	A	A	A	A	A
Sucrose	B	A	B	A	A	A

Northern Region												
Variety	Caturra				Colombia				Tipica			
Biochemical characteristic	San[2]	SnN[3]	Ces[4]	Mag[5]	San	SnN	Ces	Mag	San	SnN	Ces	Mag
Caffeine	B	C	A	A	A	B	B	A	B	B	B	A
Trigonelline	C	A	A	A	C	A	B	B	A	A	A	A
Chlorog. acid[2]	B	A	A	A	B	A	AB	B	AB	A	BC	C
CQA total	C	B	A	A	A	A	A	A	A	A	A	A
Lipid	C	A	A	A	B	A	B	B	B	A	C	C
Palmitic acid	B	A	A	A	B	A	B	B	AB	A	B	B
Estearic acid	B	A	B	B	C	A	B	B	B	A	AB	B
Oleic acid	B	A	A	A	B	A	A	A	B	A	A	A
Linoleic acid	A	C	A	A	A	B	B	B	A	B	B	B
Linolenic acid	A	A	A	A	A	B	AB	AB	A	A	A	A
Araquidic acid	B	A	A	A	B	A	AB	A	B	AB	AB	A
Behenic acid	B	AB	A	A	B	AB	AB	A	A	A	A	A
Sucrose	A	CB	A	A	A	A	A	A	A	A	A	A

[1] Within each biochemical characteristic for each variety, Departments with the same letter do not differ significantly ($P<0.05$).
[2] Chlorogenic acid
[3] San = Santander; [4] SnN = Santander Norte; [5] Ces = César; [6] Mag = Madalena.

There are no statistically significant differences in sensory characteristics between the three varieties in Tolima and Santander Norte, and only a few significant differences in sensory characteristics between them in Huila and Magdalena (Table 17 and Table 18). There are substantial significant differences between the varieties for their sensoric characteristics in all other departments.

In Colombia variety in César Department and Tipica variety in Santander Department outscored the other two varieties on many characteristics, although in Santander there was only a small number of Tipica samples included in the analyses.

In the southern Huila/Tolima Region, both Caturra and Colombia varieties from Huila have significantly higher scores for many sensory characteristics than those of the same varieties grown in Tolima (Table 17 and Table 19). In contrast, Tipica variety tends to have similar sensory characteristics in both Departments except for the clean cup characteristic, which is significantly higher in Huila compared with Tolima.

In the Northern Region, Caturra variety has no significant statistical differences in sensory characteristics between Departments, although the means do in fact indicate differences. Both Colombia and Tipica varieties are significantly better in one Department when compared with the others: Colombia variety is superior in César Department and Tipica variety is superior in Santander Department. With regard to the latter, however, we must keep in mind that there was only a small number of Tipica samples from Santander Department included in the analyses.

Table 17 shows differences in many characteristics, less for uniformity than the others, between varieties within one department, and between departments for one specific variety. This indicates high spatial and inter-varietal variability, which we have seen already for physical and biochemical characteristics.

The general conclusion from these analyses of physical, biophysical and sensoric characteristics is that the adminstriative spatial units of Departments are not the most appropriate spatial domain for the implementation of denominations of origin. If Departments were chosen by the FNC to implement the denomination, it is highly recommended and probably unavoidable that the denomination should be implemented for each variety. This obviously implies considerable complexity in the administrative procedures required. This fact had already been highlighted during the Phase 1 of this project, although the differences between varieties within and between the Departments of Cauca and Nariño were not as large as we have found them to be in Phase 2.

Table 17. Means of sensoric data of the Caturra. Colombia, and Tipica coffee varieties in six Departments of Colombia. Table A3.1 – Table A3.6 in Appendix 3 have a complete description of the data.

Sensoric charateristic	Variety	Huila	Tolima	Santander	Santander Norte	César	Magdalena
Fragrance and aroma	Caturra	5.6	5.0	5.1	5.0	5.1	5.1
	Colombia	6.0	5.4	5.6	5.4	5.9	5.2
	Tipica	5.8	5.5	6.0	5.2	5.0	5.2
Flavor	Caturra	5.6	4.6	4.7	4.8	5.3	4.8
	Colombia	6.3	5.1	5.4	5.2	6.5	5.0
	Tipica	5.2	5.0	7.4	5.0	5.0	5.9
Aftertaste	Caturra	5.6	4.5	4.4	4.5	4.8	4.8
	Colombia	6.3	5.0	5.3	4.9	6.7	5.3
	Tipica	5.1	4.9	5.7	4.8	5.3	5.2
Acidity	Caturra	5.8	5.1	5.1	5.2	5.5	4.8
	Colombia	6.6	5.5	5.5	5.5	6.3	5.3
	Tipica	5.3	5.3	6.2	5.4	5.1	5.9
Body	Caturra	5.7	5.0	4.6	4.8	5.0	4.9
	Colombia	6.4	5.2	5.5	5.3	6.3	5.0
	Tipica	5.5	5.1	5.9	5.0	5.2	5.5
Balance	Caturra	5.8	4.7	4.6	4.7	4.9	4.9
	Colombia	6.5	5.3	5.4	5.0	6.9	4.9
	Tipica	5.3	5.1	6.2	4.9	5.1	5.7
Uniformity	Caturra	7.4	7.0	7.0	6.2	6.7	6.6
	Colombia	7.5	7.2	7.0	6.6	7.8	5.8
	Tipica	7.1	7.2	6.0	7.1	7.3	6.4
Clean cup	Caturra	6.0	4.9	5.0	4.5	5.4	5.1
	Colombia	6.7	5.4	6.1	5.1	6.4	5.0
	Tipica	5.8	5.1	5.6	4.9	5.1	5.9
Sweetness	Caturra	5.5	5.0	5.0	4.7	5.5	4.7
	Colombia	6.2	5.2	5.3	5.8	6.1	5.0
	Tipica	4.9	5.0	6.7	5.7	5.0	5.8
Overall	Caturra	5.7	4.7	4.7	4.7	5.4	5.0
	Colombia	6.5	5.1	5.8	5.2	6.7	5.2
	Tipica	5.3	5.1	5.6	4.9	5.2	5.6
Final Score	Caturra	58.8	50.6	50.30	49.2	53.8	50.6
	Colombia	65.1	54.3	56.7	54.0	65.4	51.7
	Tipica	55.2	53.5	61.3	53.0	53.3	57.0

Table 18. Bonferoni multivariate tests for sensoric characteristics, omparing the coffee varieties Caturra. Colombia, and Tipica within each of six Departments in Colombia.

Department	Huila			Tolima			Santander			Santander Norte			César			Magdalena		
Sensoric characteristic	Cat[1]	Col[2]	Tip[3]	Cat	Col	Tip	Cat	Col	Tip	Cat	Col	Tip	Cat	Col	Tip	Cat	Col	Tip
Fragrance and aroma	A[4]	A	A	A	A	A	B	B	A	A	A	A	A	A	A	A	A	A
Flavor	A	A	A	A	A	A	B	B	A	A	A	A	B	A	B	A	A	A
Aftertaste	A	A	A	A	A	A	B	B	A	A	A	A	B	A	B	B	A	AB
Acidity	AB	A	B	A	A	A	B	B	A	A	A	A	B	A	B	A	A	A
Body	A	A	A	A	A	A	B	B	A	A	A	A	B	A	B	A	A	A
Balance	A	A	A	A	A	A	B	B	A	A	A	A	B	A	B	A	A	A
Uniformity	A	A	A	A	A	A	A	A	A	A	A	A	B	A	B	A	A	A
Clean Cup	A	A	A	A	A	A	B	B	A	A	A	A	A	A	A	A	A	A
Sweetness	AB	A	B	A	A	A	B	B	A	A	A	A	A	A	A	B	A	A
Overall	A	A	A	A	A	A	B	AB	A	A	A	A	A	B	AB	A	A	A
Final Score	A	A	A	A	A	A	B	B	A	A	A	A	B	A	AB	A	A	A

[1] Cat = Caturra variety; [2] Col = Colombia variety; [3] Tip = Tipica variety.
[4] Within each sensoric characteristic within each Department, varieties with the same letter do not differ significantly (P<0.05)

Table 19. Bonferoni multivariate tests for sensoric characteristics,comparing the coffee varieties Caturra, Colombia, and Tipica between six Departments in Colombia, grouped into two regions.

Huila/Tolima Region						
Variety	Caturra		Colombia		Tipica	
Sensoric characteristic	Huila	Tolima	Huila	Tolima	Huila	Tolima
Fragrance and aroma	A[1]	A	A	B	A	A
Flavor	A	B	A	B	A	A
Aftertaste	A	B	A	B	A	A
Acidity	A	A	A	B	A	A
Body	A	A	A	B	A	A
Balance	A	B	A	B	A	A
Uniformity	A	B	A	A	A	A
Clean Cup	A	B	A	B	A	B
Sweetness	A	A	A	B	A	A
Overall	A	B	A	B	A	A
Final Score	A	B	A	B	A	A

Northern Region												
Variety	Caturra				Colombia				Tipica			
Sensoric characteristic	San[2]	SnN[3]	Ces[4]	Mag[5]	San	SnN	Ces	Mag	San	SnN	Ces	Mag
Fragrance and aroma	A	A	A	A	B	B	A	B	A	B	B	B
Flavor	A	A	A	A	BA	BA	A	B	A	B	B	B
Aftertaste	A	A	A	A	B	B	A	BA	A	A	A	A
Acidity	A	A	A	A	B	B	A	B	A	A	B	A
Body	A	A	A	A	A	A	A	A	A	A	A	A
Balance	A	A	A	A	B	B	A	B	A	B	B	B
Uniformity	A	A	A	A	B	BC	A	C	B	A	B	B
Clean Cup	A	A	A	A	A	A	A	A	A	B	BA	BA
Sweetness	A	A	A	A	A	A	A	A	A	B	B	B
Overall	A	A	A	A	BA	B	A	B	A	B	BA	BA
Final Score	A	A	A	A	BA	B	A	B	A	B	B	B

[1] Within each sensoric characteristic for each variety, Departments with the same letter do not differ significantly ($P<0.05$).
[2] San = Santander; [3] SnN = Santander Norte; [4] Ces = César; [5] Mag = Madalena.

6c. *Identifying new spatial units for the denomination of origin*

6c (i). *Expert knowledge domains and environmental clusters*

We consulted three coffee-quality experts of the FNC to identify domains that each had a unique coffee quality profile. According to the accumulated knowledge of these experts, 13 domains could be recognized. Of these 13 domains, only two were identified for the Huila/Tolima Region, but there were 11 for the Northern Region (Figure 21 and Figure 22). From here on we refer tp these as "expert domains".

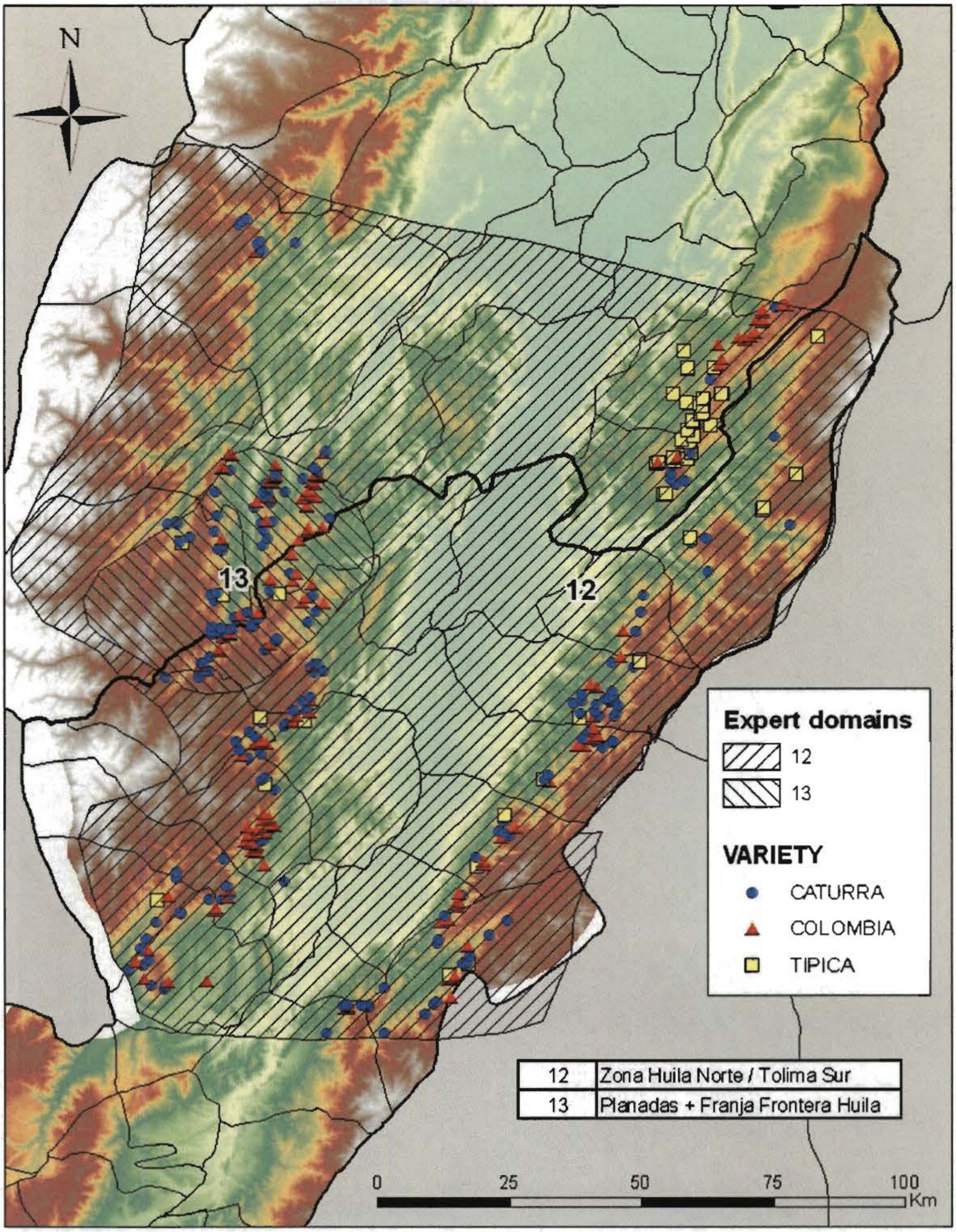


Figure 21. Distribution of original expert domains with sampling points (showing variety) in the Huila/Tolima Region.

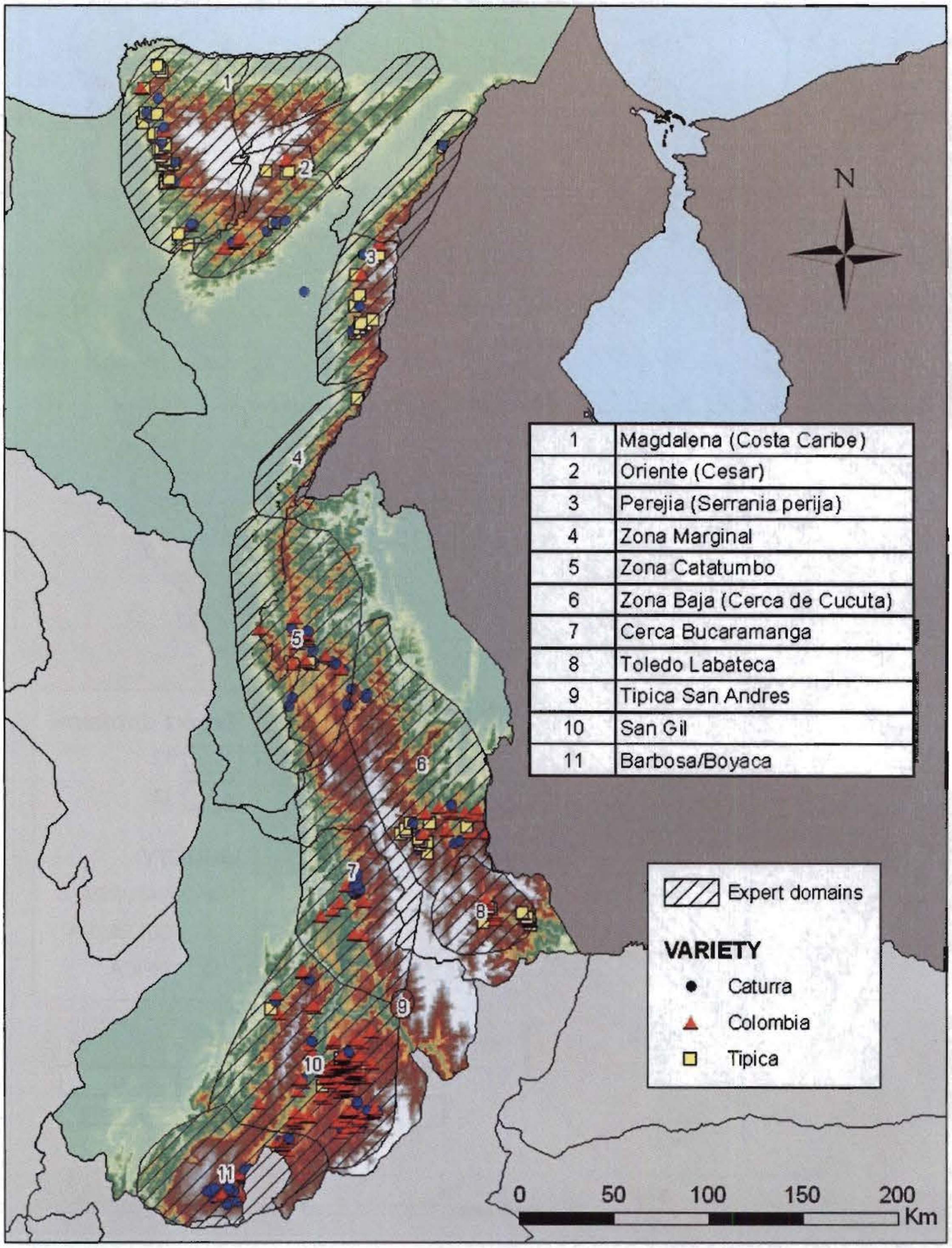


Figure 22. Distribution of original expert domains with sampling points (showing variety) in the Northern Region.

These expert domains included a large part of the project zone in Huila/Tolima with the exception of the small inclusion of Planadas and the Franja Frontera Huila. It is interesting to note that the northeast corner in the Huila/Tolima Region with predominantly Tipica variety was not considered as having a unique quality profile. Apart from this northeast corner, the remainder of the Huila/Tolima Region is dominated by a mixture of the Caturra and Colombia varieties, interspersed only occasionally with variety Tipica.

In the northern Region, the 11 expert domains included Magdalena (Costa Caribe), Oriente (César), Perejia (Serrania perija), a marginal zone, the Catatumbo zone, the Zona Baja (cerca de Cucuta), the zone close to Bucaramanga, Toledos Labateca, the Tipica zone of San Andrés, the San Gil region, and finally the Barbosa/Boyaca area.

The northernmost zones of Magdalena (Costa Caribe), Oriente (César), and Perejia (Serrania perija) have an abundance of the Tipica variety, with some Caturra but very little of the Colombia variety. The marginal zone was, except for one site, not sampled. The Catatumba zone, Bucaramanga, San Gil and Barbosa/Boyaca are all dominated by a mix of Caturra and Colombia, with Colombia very dominant in San Gil and Caturra strongly represented in Barbosa/Boyaca. The Zona Baja and Toledo Labateca have a high occurance of the Tipica variety, while San Andrés was not sampled. In any case, San Andrés is a very small area compared to the other zones.

These expert domains were then analyzed to determine their similarity with the spatial distribution of the generated climate clusters and soils clusters (Table 20 and Table 21) for the Tolima/Huila Region and the Northern Region separately Figure 23 Figure 26). As in the first study, climate has a stronger influence than soil, although soil remains important in some cases in determining the clusters, There is a good degree of similarity between the combined climate and soil clusters and the expert domains.

The similarity in the Huila/Tolima region between the climate clusters and the expert domains is indeed striking: Climate clusters 2 and 3 almost exclusively occur in the large expert domain Huila/Tolima. The smaller zone of Planadas / Franja Frontera Huila is dominated by the climate clusters 1, 4 and 5. In contrast, the soils clusters are distributed across the two Huila/Tolima expert domains.

There is also a similar pattern in the distribution of climate clusters and expert domains in the Northern Region, although the match is not as clear cut as in the Huila/Tolima Region. We find that Magdalena (Costa Caribe) is dominated by climate clusters 2 and 3. The Oriente (César) and Perejia (Serrania perija) zones are dominated by climate clusters 1 and 4. The expert domain of Catatumbo, the domain Zona Baja (cerca de Cucuta), and the zone close to Bucaramanga are all clearly dominated by the climate cluster 5. The remaining expert domains of Toledos Labateca, of San Andres, the San Gil region, and the region of Barbosa / Boyaca are dominated by climate clusters 1, 2, and 4.

Table 20. Mean climate and physical characteristics for each of the five dominant clusters in the Huila/Tolima Region and the Northern Region. Table A4.1 and Table A4.2 in Appendix 4 have a complete description of the data.

		Huila/Tolima Region			
Cluster number	1	2	3	4	5
n	41	114	112	68	48
Site characterisitic					
Annual precipitation (mm)	2460	1700	1750	2170	2190
Annual evaporation (mm)	870	880	820	880	860
Mean dewpoint (°C)	14.0	14.7	13.4	15.1	13.0
Number of dry months	1.0	1.8	1.3	1.9	1.1
Mean temperature (°C)	20.3	20.3	19.0	21.3	19.2
Mean diurnal temperature range (°C)	10.4	10.0	10.0	10.5	10.3
Mean daily solar radiation (MJ m^{-2} d^{-1})	23.8	23.6	23.3	23.5	23.2
Mean cloud cover (%)	94.8	95.9	96.1	95.4	95.3
Altitude (masl)	1660	1460	1740	1400	1800
Aspect (compass °)	162	231	180	178	203
Slope (°)	17.7	17.4	20.8	19.5	21.4
		Northern Region			
Cluster number	1	2	3	4	5
n	118	86	38	62	79
Site characterisitic					
Annual precipitation (mm)	2260	2560	2430	1990	1400
Annual evaporation (mm)	960	1130	1320	1030	990
Mean dewpoint (°C)	13.1	14.0	16.5	15.5	14.2
Number of dry months	1.7	1.9	3.7	3.1	4.4
Mean temperature (°C)	19.1	19.9	21.8	21.2	19.6
Mean diurnal temperature range (°C)	10.6	10.6	10.0	10.5	9.9
Mean daily solar radiation (MJ m^{-2} d^{-1})	24.1	23.5	22.7	23.1	23.1
Mean cloud cover (%)	94.5	89.6	81.4	91.1	92.6
Altitude (masl)	1710	1520	1040	1210	1440
Aspect (compass °)	195	209	230	196	178
Slope (°)	12.5	15.6	21.2	16.4	19.3

Table 21. Mean soil characteristics for each of the five dominant clusters in the Southern Region and the Northern Region. Table A5.1 – Table A5.4 in Appendix 5 have a complete description of the data.

Huila/Tolima Region					
Cluster number	1	2	3	4	5
n	204	57	38	9	7
Soil characterisitic					
pH	5.3	4.7	4.5	4.8	5.7
N (%)	0.3	0.3	0.4	0.4	0.4
OM (%)o	6.9	8.4	9.7	10.0	10.1
K (cmol/kg)	0.3	0.4	0.3	2.1	0.5
Ca (cmol/kg)	6.6	5.0	3.8	3.5	12.5
Mg (cmol/kg)	1.8	1.2	1.0	1.7	1.8
Al (cmol/kg)	1.1	2.6	4.1	3.0	0.7
CEC (cmol/kg)	15.9	18.3	20.1	17.6	20.3
P (mg/kg)	28.3	29.8	40.6	77.1	584.7
Fe (mg/kg)	267.5	602.9	925.2	1356.0	400.7
Mn (mg/kg)	68.1	43.0	48.6	58.8	71.4
Zn (mg/kg)	3.6	4.7	5.2	11.8	15.4
C (mg/kg)u	2.8	2.7	3.1	4.2	5.4
Clay (%)	30.7	40.0	40.6	37.4	27.6
Silt (%)	26.5	24.2	25.4	25.1	30.1
Sand (%)	42.9	35.7	34.1	38.8	42.0
Northern Region					
Cluster number	1	2	3	4	5
n	84	86	145	59	17
Soil characterisitic					
pH	4.9	4.4	5.2	4.7	4.2
N (%)	0.3	0.4	0.3	0.3	0.4
OM (%)o	7.5	8.6	6.8	7.8	9.7
K (cmol/kg)	0.3	0.3	0.3	0.3	0.2
Ca (cmol/kg)	4.4	2.9	5.9	4.1	1.9
Mg (cmol/kg)	1.0	0.7	1.2	0.9	0.5
Al (cmol/kg)	1.8	4.1	1.1	2.7	5.5
CEC (cmol/kg)	16.2	18.8	14.9	16.7	21.8
P (mg/kg)	17.3	25.5	25.9	16.9	17.4
Fe (mg/kg)	375.6	924.0	219.0	595.7	1405.8
Mn (mg/kg)	48.5	23.1	82.8	47.8	18.8
Zn (mg/kg)	3.6	5.3	3.5	5.7	5.6
C (mg/kg)u	3.3	3.7	2.9	11.6	3.2
Clay (%)	37.2	44.0	32.0	40.4	48.6
Silt (%)	26.5	25.9	27.3	26.5	23.6
Sand (%)	36.4	30.1	40.8	33.1	27.5

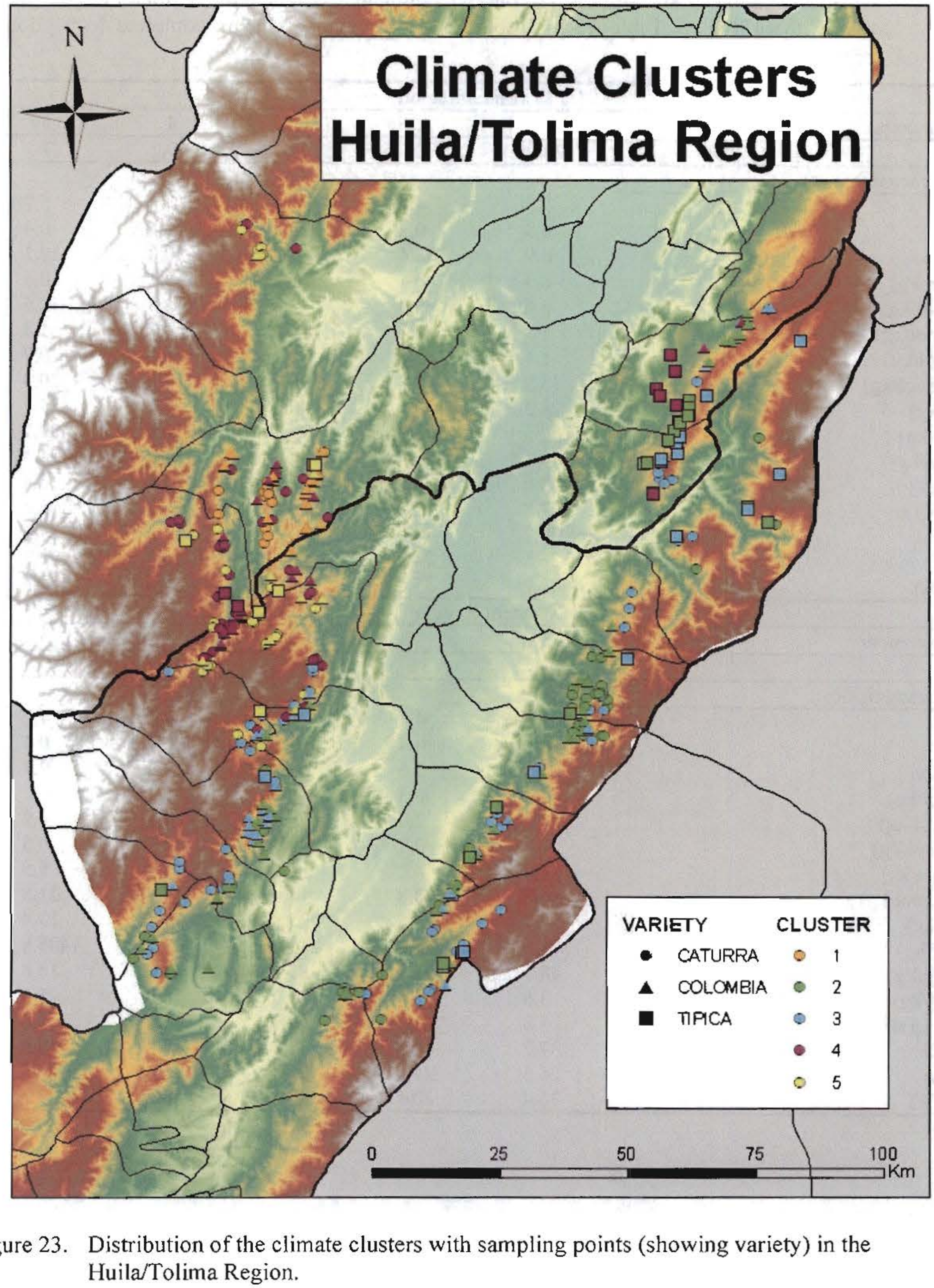


Figure 23. Distribution of the climate clusters with sampling points (showing variety) in the Huila/Tolima Region.

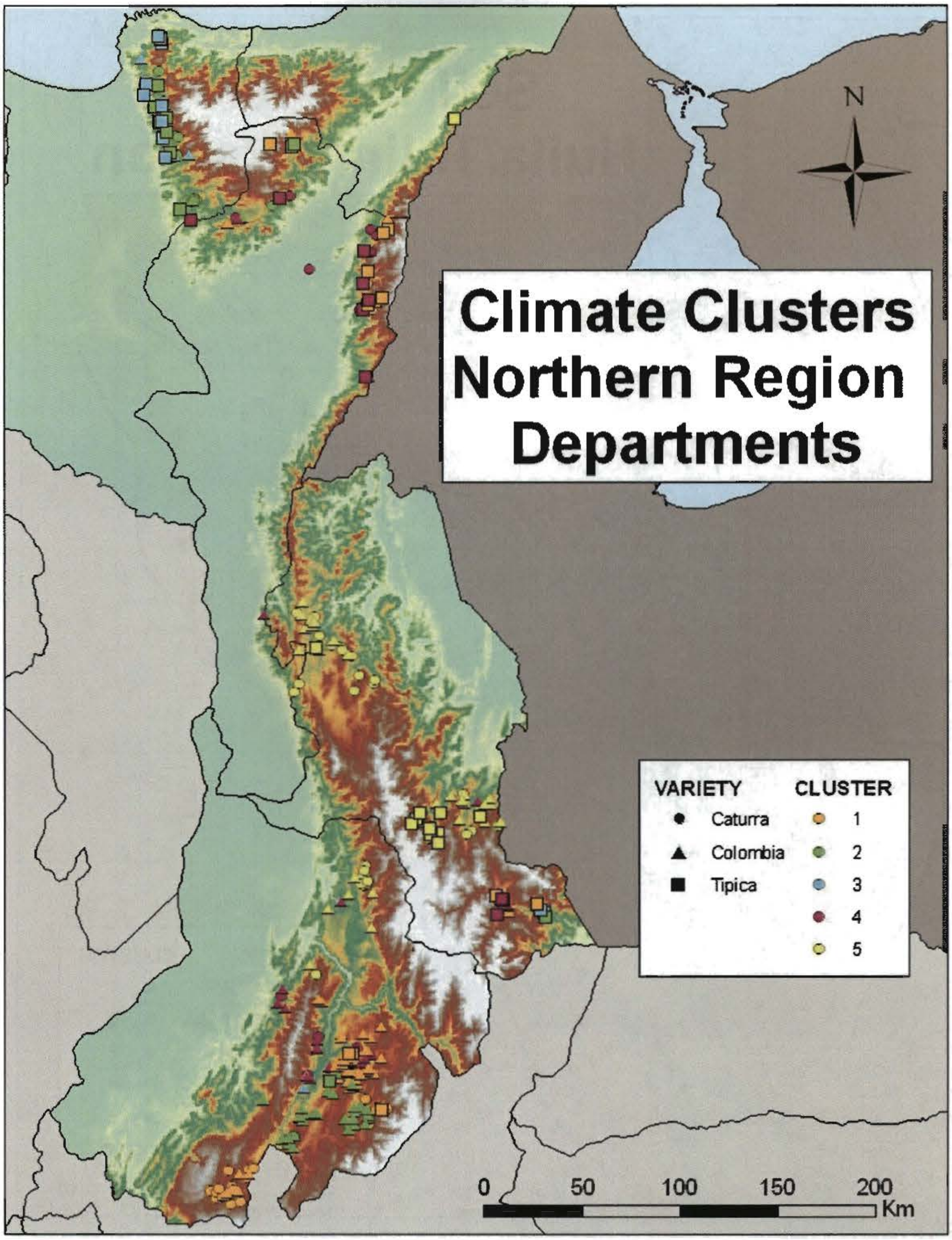


Figure 24. Distribution of the climate clusters with sampling points (showing variety) in the Northern Region.

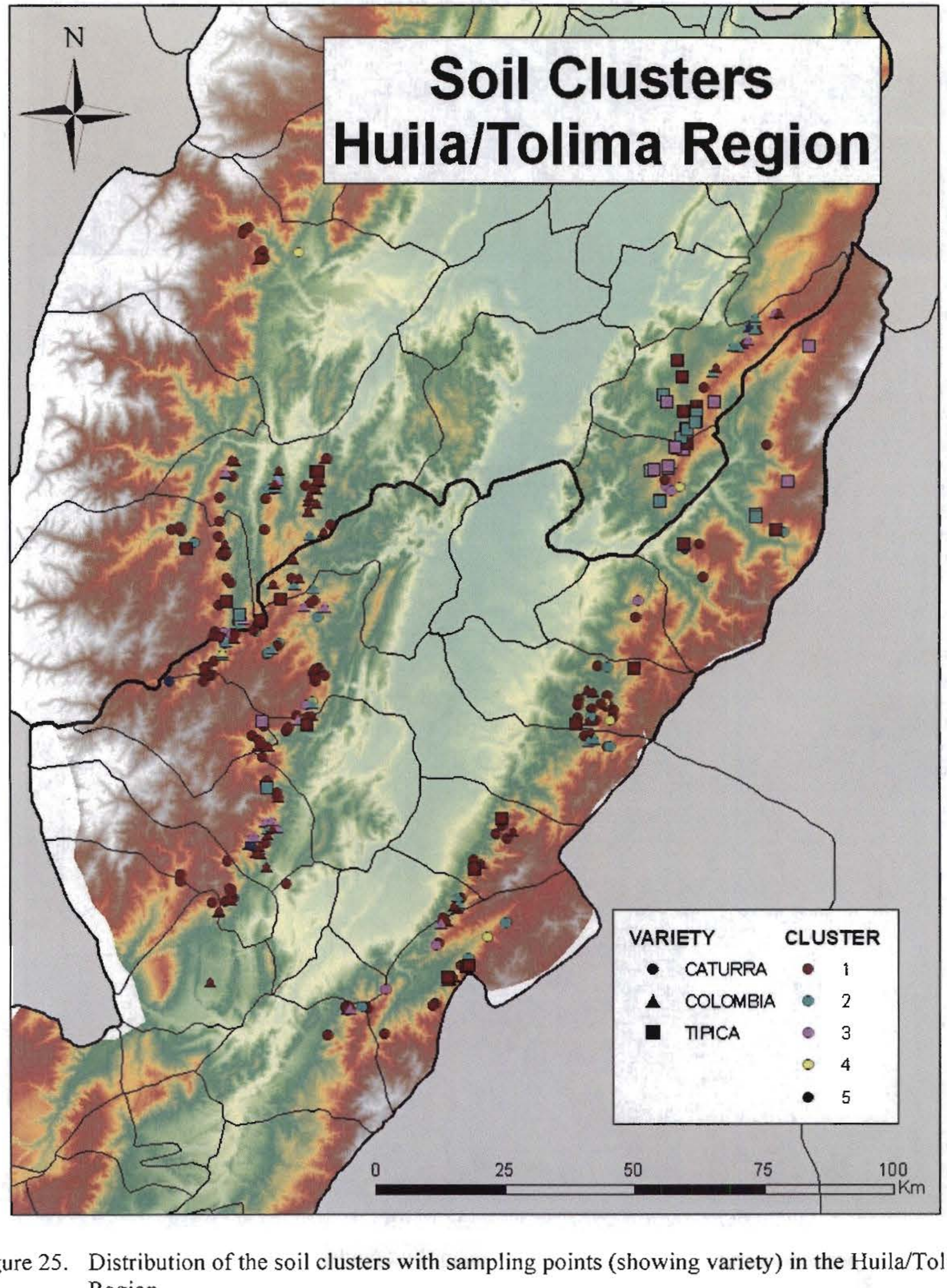


Figure 25. Distribution of the soil clusters with sampling points (showing variety) in the Huila/Tolima Region.

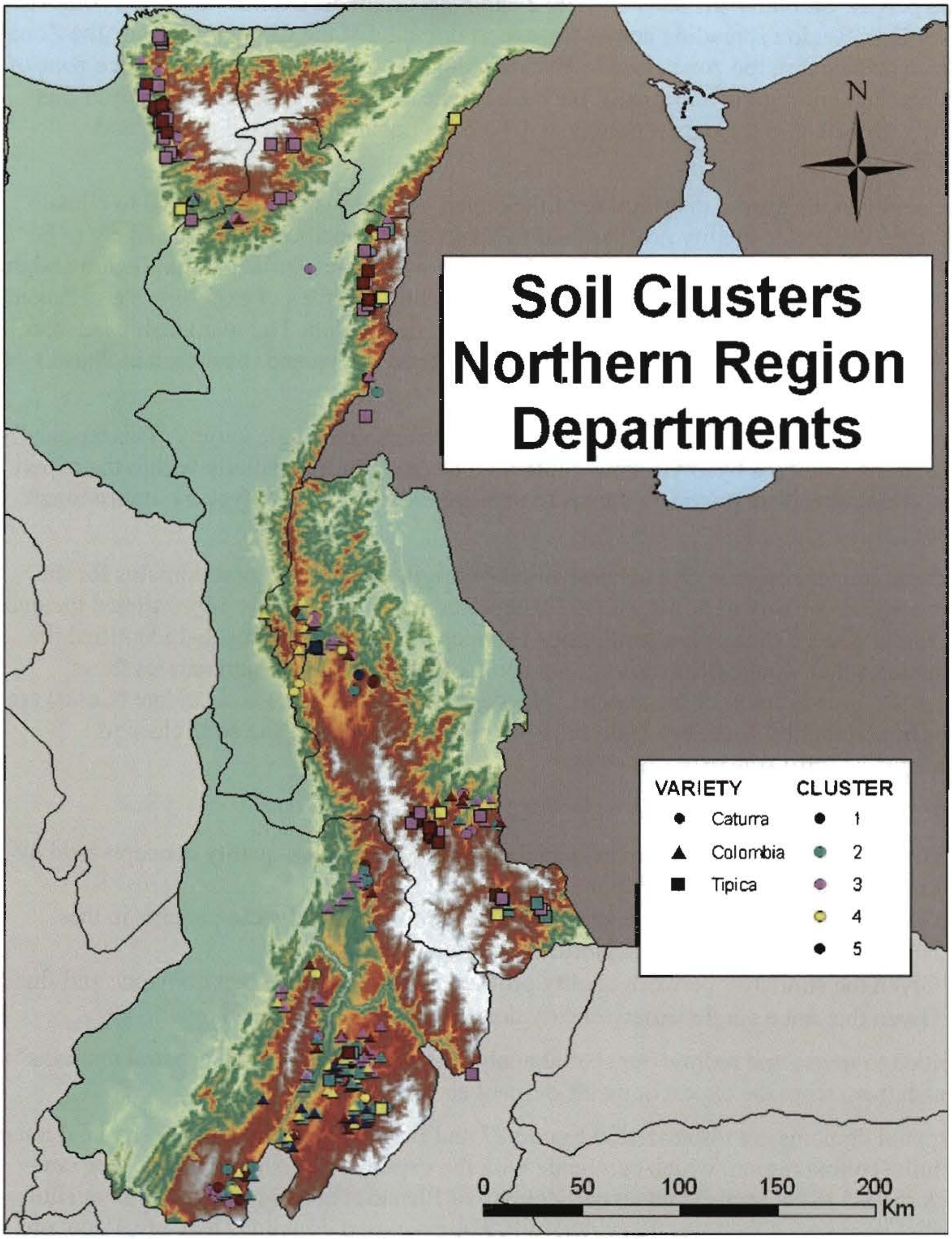


Figure 26. Distribution of the soil clusters with sampling points (showing variety) in the Northern Region.

The soils clusters in the Norther Region are also spread across all expert domains, except that we observe a prevalence of soil clusters 1 and 3 in the northern part of the Northern Region, dominating the expert domains of Magdalena (Costa Caribe), Oriente (Cesar), and Perejia

(Serrania perija). Similarly the soils clusters 2 and 5 occur almost exclusively in the southern part of the Northern Region spreading across the expert domains of the Catatumbo zone, the Zona Baja (cerca de Cucuta), the zone close to Bucaramanga, Toledos Labateca, the Tipica zone of San Andres, San Gil region, and finally Barbosa/Boyaca. We also note that soil cluster 2 is specifically prevalent in the expert domains of Toledos Labateca, San Gil region, and Barbosa/Boyaca.

It can be said that the spatial distribution of the expert domains is strongly linked to climate patterns, ergo the coffee quality profiles held by experts for specific regions are likely to be dependent on climate characteristics. This holds true for both the Huila/Tolima Region and the Northern Region. In the Northern Region the coffee quality profiles of experts are also linked to soil patterns, and specifically so in the northern part of this region. The relationship between climate characteristics and coffee quality has already been shown and illustrated in Phase 1 of this project.

Most importantly, the expert domains are not dominated by one single variety. This bascically implies that the genotype by environment interactions are reduced spatially within the expert domains, which therefore provide a means to manage the issue of genotype by environment interaction.

However, an implementation of a denomination of origin within 11 expert domains for the Northern Region is likely to be very complex undertaking. We therefore copnsultated the quality experts again to see if there were similarities between the quality profiles that identified the expert domains that would allow us to group some of them. Table 26 summarizes these assessments. In the opinion of the experts, Magdalena (Costa Caribe) and Oriente (Cesar) are similar. The Catatumbo zone, the Zona Baja (cerca de Cucuta), and the zone close to Bucaramanga are also similar.

Summarizing:

- Given the strong relationship in spatial distribution between quality concepts held by experts, the climate and partly the soils;
- Given the identified spatial distribution of predominating climate patterns in the Huila/Tolima Region and the Northern Region;
- Given the similarity between quality profiles from different expert domains; and finally
- Given that not a single variety clearly dominates any one expert domain;

We decided to revise and redraw our spatial analyses units, which we call "spatial domains" to distinguish them from the expert domains defined above.

These spatial domains are visualized in Figure 27 and Figure 28. There are two spatial domains in the Huila/Tolima region, which conincide with the expert knowledge domains. The new spatial domain 5 is the same as the expert domain of Planadas and the Franja Frontera Huila. The new spatial domain 6 consitutes the widerHuila/Tolima expert domain. There are 4 domains in the Northern region: Spatial domain 1 is equal to the expert domain of Magdalena (Costa Caribe); spatial domain 2 brings together the Oriente (Cesar) and Perejia (Serrania perija); spatial domain 3 joins the Catatumbo zone, the Zona Baja (cerca de Cucuta), and the zone close to Bucaramanga; while spatial domain 4 is made up by the expert domains of Toledos Labateca, San Gil region, and the region of Barbosa/Boyaca.

To consolidate the argument for these spatial domains further, we first computed cluster analyses using the results of both the sensoric analyses and the NIRS analyses to generate 5 sensoric clusters and 5 NIRS clusters for both the Huila/Tolima Region and the Northern Region. We then summarized the distribution of varities, and as well of climate, soils, NIRS and sensoric clusters within the new spatial domains. The results of this interpretation are presented in Table 22Table 25. The key observations are that:

- Specifically the distribution of sampling points in the climate clusters is captured well by the new spatial domains, for example the new spatial domain 4 has a total 182 sampling sites, and 100 of these sampling sites are classified within the climate cluster 1 and 48 belong to climate cluster 2 (Table 22);
- The variability within the soils clusters is also well captured by the new spatial domains. For example, spatial domain 3 has 82 sampling sites, of which more than 50 belong to just two soils' clusters (Table 23);
- While expectedly not as obvious as in the case of soils and climate clusters, also the variability in the distribution of sampling sites in the NIRS and sensoric clusters is well captured by the new domains, in most cases half the points within one spatial domain belong to only two NIRS or sensoric clusters, respectively (Table 24 and Table 25).

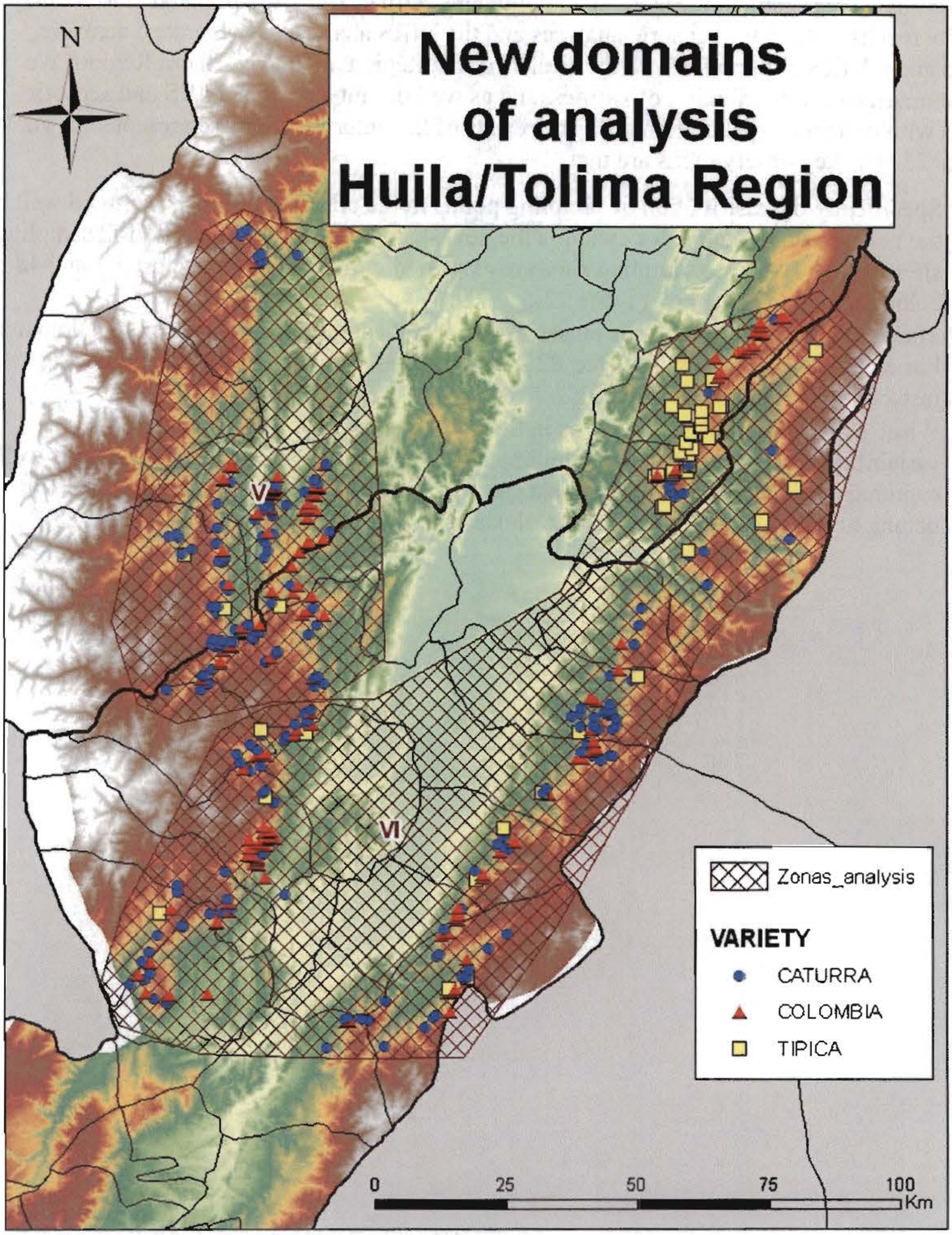


Figure 27. Distribution of the re-drawn new spatial domains with sampling points (showing variety) in the Huila/Tolima Region.

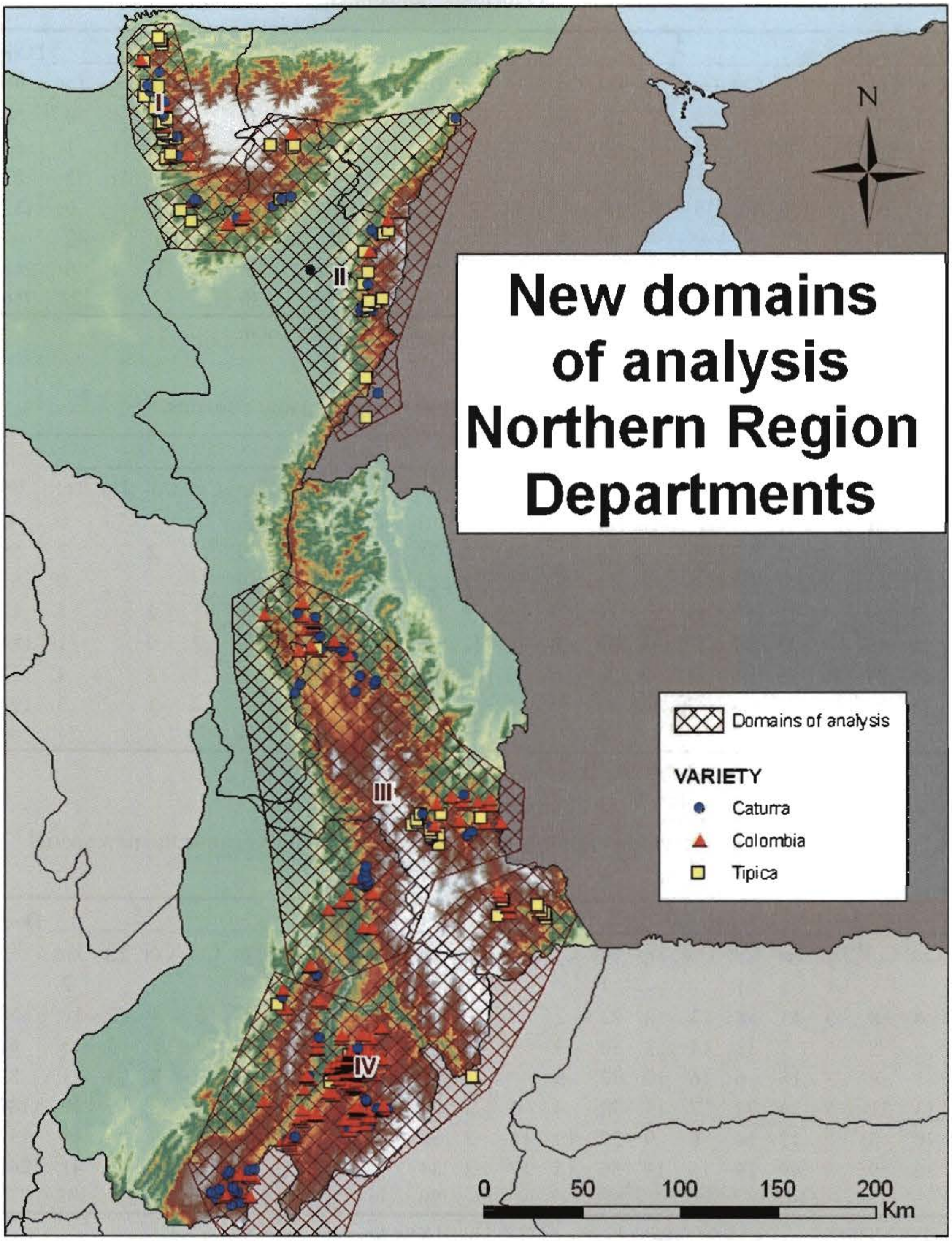


Figure 28. Distribution of the re-drawn new spatial domains with sampling points (showing variety) in the Northern Region.

Table 22. Matrices of varieties within climate clusters against the new spatial domains.

Cluster	1				2				3				4				5				Domain
Domain	Cat[1]	Col[2]	Tip[3]	Tot[4]	Cat	Col	Tip	Tot	Cat	Col	Tip	Tot	Cat	Col	Tip	Tot	Cat	Col	Tip	Tot	total
I				0	14	5	7	26	15	7	10	32				0				0	58
II	7	4	7	18	7	2	3	12				0	16	5	9	30			1	1	61
III				0				0				0		4		4	34	33	11	78	82
IV	24	71	5	100	10	36	2	48		4	2	6	5	20	3	28				0	182
V	18	22	1	41				0	3			3	37	17	2	56	29	9	4	42	142
VI				0	47	49	18	114	61	32	16	109	3	2	7	12	4	1	1	6	241
Total				159				200				150				130				127	766

[1] Cat = Caturra variety; [2] Col = Colombia variety; [3] Tip = Tipica variety. [4] Tot = Total.

Table 23. Matrices of varieties within soil clusters against the new spatial domains.

Cluster	1				2				3				4				5				Domain
Domain	Cat[1]	Col[2]	Tip[3]	Tot[4]	Cat	Col	Tip	Tot	Cat	Col	Tip	Tot	Cat	Col	Tip	Tot	Cat	Col	Tip	Tot	total
0		1		1	1			1	1			1	2	1		3				0	6
I	3	6	2	11	7	10	1	18	8	4	6	18	3	6		9		2		2	58
II	11	4	2	17	3	6	2	11	12	8	4	24	4	5		9				0	61
III	4	9	5	18	5	15		20	9	15	8	32	1	6	1	8	1	2	1	4	82
IV	15	15	7	37	11	21	4	36	28	30	12	70	5	19	6	30	2	9		11	184
V	45	24	10	79	3	14	7	24	6	3	4	13	2			2	1	3		4	122
VI	75	38	12	125	14	13	6	33	7	13	5	25	5	1	1	7	2	1		3	193
Total				288				143				183				68				24	706

[1] Cat = Caturra variety; [2] Col = Colombia variety; [3] Tip = Tipica variety. [4] Tot = Total.

Table 24. Matrices of varieties within clusters of sensorial characteristics against the new spatial domains.

Cluster	1				2				3				4				5				Domain
Domain	Cat[1]	Col[2]	Tip[3]	Tot[4]	Cat	Col	Tip	Tot	Cat	Col	Tip	Tot	Cat	Col	Tip	Tot	Cat	Col	Tip	Tot	total
0		1	1	2	1			1		1		1				0		2		2	6
I	4	8	3	15	8	13	2	23	2			2	3	1	2	6	6	4	2	12	58
II	2	5		7	13	14	3	30	3			3	2	4	2	8	10	2	3	15	63
III	1	14	3	18	6	16	10	32	5			5	5	9	2	16	2	8	1	11	82
IV	13	21	5	39	24	37	11	72	4	3	2	9	4	8	3	15	19	23	7	49	184
V	10	5	2	17	14	6	9	29	19	15	2	36	16	19	7	42	8	8	2	18	142
VI	36	9	4	49	29	16	11	56	15	17	2	34	33	21	7	61	22	16	3	41	241
Total				147				243				90				148				148	776

[1] Cat = Caturra variety; [2] Col = Colombia variety; [3] Tip = Tipica variety. [4] Tot = Total.

Table 25. Matrices of varieties within clusters of NIRS characteristics against the new spatial domains.

Cluster	1				2				3				4				5				Domain total
Domain	Cat[1]	Col[2]	Tip[3]	Tot[4]	Cat	Col	Tip	Tot	Cat	Col	Tip	Tot	Cat	Col	Tip	Tot	Cat	Col	Tip	Tot	
0	1	1		2	2			2	1	1		2				0				0	6
I	6	8	3	17	2	3	2	7	4	8	1	13	6	1	3	10	3	8		11	58
II	5	7	5	17	5	6		11	4	3		7	9	7	1	17	7	1	3	11	63
III	7	14	6	27	3	8	3	14	4	13	1	18	4	6	5	15	3	5		8	82
IV	19	32	7	58	15	13	4	32	9	20	4	33	8	17	9	34	10	12	5	27	184
V	8	15	1	24	14	14		28	28	10	16	54	14	4	3	21	3	10	2	15	142
VI	49	28	5	82	37	21	2	60	24	7	10	41	16	6	3	25	9	17	7	33	241
Total				227				154				168				122				105	776

[1] Cat = Caturra variety; [2] Col = Colombia variety; [3] Tip = Tipica variety. [4] Tot = Total.

Table 26. Matrix of the similarity of coffee quality with expert domains.

Legend
0 = Not similar
1 = Similar

		Norte Santander			Sierra Nevada			Santander			Bolivar		Huila/Tolima	
		Zona Catatumbo	Toledo Labateca	Zona Baja (Cerca de Cucuta)	Magdalena (Costa Caribe)	Oriente (Cesar)	Perejia (Serrania perija)	Tipica San Andres	Cerca Bucaramanga	San Gil	Barbosa/Boyaca	Zona Marginal	Planadas + Franja Frontera[1]	Zona Huila Norte/Tolima Sur
Norte Santander	Zona Catatumbo		0	1	0	0	0	0	0	0	0	0	0	0
	Toledo Labateca			0	0	0	0	0	0	0	0	0	0	0
	Zona Baja (Cerca de Cucuta)				0	0	0	0	1	0	0	0	0	0
Sierra Nevada	Magdalena (Costa Caribe)					1	0	0	0	0	0	0	0	0
	Oriente (Cesar)						0	0	0	0	0	0	0	0
	Perejia (Serrania perija)							0	0	0	0	0	0	0
Santander	Tipica San Andres								0	0	0	0	0	0
	Cerca Bucaramanga									0	0	0	0	0
	San Gil										0	0	0	1
	Barbosa/Boyaca											0	1	0
Bolivar	Zona Marginal												0	0
Huila/Tolima	Planadas + Franja Frontera[1]													0
	Zona Huila Norte/Tolima Sur													

[1] Franja Frontera Huila

6c (ii). Bean characteristics (physical, biochemical, sensory, and expert opinilon) within spatial domains.

In this section we analyze the most important physical, biochemical and sensory characteristics for each of the six new spatial domains. The information is summarized in Table 27 Table 33

As far as the physical characteristics are concerned, there are statistically signifcant differences for the yield factors in the Northern Region and for the large screen sizes in the Huila/Tolima region. The spatial domain V tends to have better results tan the spatial domain VI, both in the Huila/Tolima Region. In the Northern Region, spatial domains II and IV stand out, but we also note very positive values in spatial domain I for screen size 17, that is for large beans.

Most of the biochemical characteristics do have statistically significant differences between spatial domains, except for sucrose in the Northern Region and trigonelline in the Huila/Tolima Region. The differences beween the means, however, tend generally to be small. One notable exception in the Huila/Tolima Region is that sucrose is substantially higher in the spatial domain V compared with spatial domain VI. There are also notable differences in chlorogenic acid between the four spatial domains in the Northern Region.

The differences in the means of the sensoric attributes are substantial, and exist for most of the spatial domains. The Bonferoni tests show that differences betweeen most attributes are statistically significant. It is noteworthy that spatial domain VI has much higher values than spatial domain V in the Huila/Tolima region, although coffees from spatial domain V are generally considered superior to those of spatial domain VI. This point is discussed in the conclusions and recommendations section. It is also interesting that flavor, aftertaste, body, clean cup and the overall rating are relatively low in spatial domain III in the Northern Region.

In conclusion, when coffee quality in one spatial domain differs from that in the other spatial domains, it thereby establishes a key requirement for denomination of origin. To consolidate the argument further, we asked the FNC coffee-quality experts to define the quality profiles in more detail. The experts described the quality attributes for the intial domains that they had defined. These attributes can readily be interpreted in a meaningful manner for the six new spatial domains (Table 33):

Spatial domain I: Coffees are characterized by low acidity and high body.Their fragrance and aroma exhibits a nutty character, as does their flavor and therefore the overall impression. The expert opinions are nicely confirmed by the sensoric cupping data.

Spatial domain II: Coffees from this domain tend to be balanced with medium body and acidity, and they have a moderate level of sweetness. Their fragrance and aroma is characterized by nutty and chocolaty notes, which in the flavor are complemented by caramel tones. This leads to coffees that overall can be considered as having sweet, nut–caramel notes, with little astringency or off flavors. Generally, the cupping data are in line with this assessment.

Spatial domain III: These coffees are generally characterized by medium acidity and medium body. Their sweetness level tends to be low. The fragrance and aroma has sometimes astringency notes, while the flavor can be chocolaty or nutty in some subregions. Coffees have the tendency to exhibit astringency and herbal nuances in the flavor as well. These assessments are reflected by the cupping notes for spatial domain 3.

Spatial domain IV: Coffees of this domain have medium to high levels of body and sweetness. Their acidity tends to be moderate, in some cases high, while the fragrance and aroma is characterized by sweet, fruity notes. These are generally reflected in the flavor, and complemented by sweet caramel nuances. This leads to an overall profile that is fruity and sweet, acidity in some cases may be citric. Off flavors are seldom found. These expert assessments are squarely confirmed by the cupping scores, which rank these coffees highly.

Spatial domain V: These coffees are considered as having high levels of sweetness and acidity combined with medium body. They demonstrate fruity and floral fragrance and aroma, very often combined with sweet caramel notes, and a citric fragrance. The flavor fully reflects the fragrance and aroma expressions. It leads to coffees with high overall quality profiles that are characterized by sweetness, and fruity and citric acidity, with clear caramel notes. Surprisingly, the cupping scores do not reflect this assessment. This is discussed in the Conclusions and recommendations.

Spatial domain VI. These are balanced coffees with medium levels of acidity, body and sweetness. They have sweet notes in fragrance and aroma, often accompanied by herbal tones. The flavor reflects these, but also shows sweet caramel aspects. Overall, these are sweet, fruity coffees that may have a herbal off taste. Cupping scores are suprisingly high for these coffees.

In summary, the new spatial domains capture the quality differences between coffees in a very meaningful way. They tend to achieve this independent of the coffee variety. Each spatial domain has usually at least two dominant varieties, yet the differences in quality attributes are consistent across the spatial domains. We have therefore established reasonable spatial domains for the denomination of origin. In the next step we shall attempt to consolidate the definition of the spatial domains with análisis of the environmental data.

Table 27. Means of physical characteristics of the coffee varieties grown in six spatial domains.Table A6.1 in Appendix 6 has a complete description of the data.

Physical characteristic	Domain					
	Northern Region				Huila/Tolima Region	
	I	II	III	IV	V	VI
YF 13	84.6	84.9	86.0	85.3	85.8	86.6
YF 14	85.2	85.7	87.5	86.0	86.7	88.7
YF 15	88.4	89.1	92.7	89.4	90.1	95.9
Screen size 17	72.8	70.5	61.5	70.0	65.6	58.5
Screen size 18	43.8	47.3	46.1	49.1	49.3	41.2

Table 28. Bonferoni multivariate tests for selected physical characteristics, comparing yield factors and screen sizes across the six spatial domains.

Physical characteristic	Domain					
	Northern Region				Huila/Tolima Region	
	I	II	III	IV	V	VI
YF 13	C	BC	A	B	A	A
YF 14	B	B	A	B	A	A
YF 15	B	B	A	B	A	A
Screen size 17	A	A	B	A	A	B
Screen size 18	A	A	A	A	A	B

Table 29. Means of selected biochemical characteristics of the coffee varieties grown in six spatial domains. Table A7.1 in Appendix 7 has a complete description of the data.

Biochemical characteristic	Domain					
	Northern Region				Huila/Tolima Region	
	I	II	III	IV	V	VI
Chlorogenic. acid	6.4	6.7	6.7	6.5	6.6	6.6
Caffeine	1.4	1.4	1.3	1.4	1.3	1.4
Trigonelline	1.0	0.9	1.0	0.9	0.9	1.0
Sucrose	5.3	5.2	5.3	5.3	5.7	5.0

Table 30. Bonferoni multivariate tests for selected biochemical characteristics, compared across the six spatial domains.

Biochemical characteristic	Domain					
	Northern Region				Huila/Tolima Region	
	I	II	III	IV	V	VI
Chlorogenic. acid	A	C	C	B	B	A
Caffeine	B	B	A	C	B	A
Trigonelline	B	A	A	B	A	A
Sucrose	A	A	A	A	A	B

Table 31. Means of sensory characteristics of the coffee varieties grown in six spatial domains.Table A8.1 and Table A8.2 in Appendix 8 have a complete description of the data.

Sensory characteristic	Domain					
	Northern Region				Huila/Tolima Region	
	I	II	III	IV	V	VI
Fragrance and aroma	5.2	5.2	5.2	5.5	5.2	5.7
Flavor	5.4	5.1	4.9	5.3	4.9	5.7
Aftertaste	5.3	5.0	4.5	5.2	4.8	5.7
Acidity	5.4	5.3	5.3	5.4	5.3	5.9
Body	5.4	5.1	4.9	5.3	5.1	5.8
Balance	5.4	5.1	4.7	5.3	5.0	5.9
Uniformity	6.4	7.0	6.6	7.0	7.1	7.4
Clean cup	5.6	5.2	4.8	5.8	5.2	6.1
Sweetness	5.3	5.2	5.4	5.3	5.1	5.6
Overall	5.5	5.3	4.9	5.5	5.0	5.8
Final Score	54.8	53.6	51.2	55.5	52.6	59.6

Table 32. Bonferoni multivariate tests for selected sensory characteristics, compared across the six spatial domains.

Sensory characteristic	Domain					
	Northern Region				Huila/Tolima Region	
	I	II	III	IV	V	VI
Fragrance and aroma	A	A	A	A	B	A
Flavor	A	A	A	A	B	A
Aftertaste	A	AB	B	AB	B	A
Acidity	A	A	A	A	B	A
Body	A	A	A	A	B	A
Balance	A	AB	B	A	B	A
Uniformity	B	A	B	A	B	A
Clean cup	A	AB	B	A	B	A
Sweetness	A	A	A	A	B	A
Overall	A	AB	B	A	B	A
Final Score	A	A	A	A	B	A

Table 33. Matrices of product quality characteristics as described by the FNC quality experts' opinion.

Spatial domain		I	II			III			IV				V	VI
Department		Sierra Nevada			Bol[1]	Sant[2]	Norte Santander		Santander				Huila/ Tolima	
Indications: 1 = low 2 = medium 3 = high Adjectives: 1 = yes 0 = no		Magdalena (Costa Caribe)	Oriente (César)	Perejia (Serrania perija	(Almost none)	Cerca Bucaramanga	Zona Catatumbo	Zona Baja (Cerca de Cucatá	Toledo Labateca	Tipica San Andrés	San Gil	Barbosa/Boyaca	Planadas + Franja frontera Huila	Zona Huila norte/ Tolima sur
	Acidity	1	2	2		2	1	1	2	2	2	3	3	2
	Body	3	2	2		2	2	2	3	2	2	2	2	2
	Sweetness	1	1	2		1	1	1	3	2	2	3	3	2
	Clean Cup	1	2	3		1	2	2	3	2	3	2	3	2
Fragrance and aroma	Floral	0	0	1		0	0	0	0	0	1	0	1	0
	Fruity	0	0	0		0	0	0	1	1	1	1	1	0
	Herbal	0	0	0		1	0	0	0	0	1	0	0	1
	Nutty	1	1	1		0	0	1	0	0	0	0	0	0
	Sweet	0	0	1		1	0	0	1	1	1	1	1	1
	Caramel	0	1	0		0	0	0	1	0	1	0	1	0
	Chocolate	0	1	1		0	0	0	1	0	0	0	0	0
	Citric	0	0	0		0	0	0	0	0	0	1	1	0
	Astringent	0	0	0		1	1	0	0	0	0	0	0	0
	Immature	0	0	0		1	1	0	0	0	0	0	0	0
Flavor	Floral	0	0	1		0	0	0	0	0	0	0	1	0
	Fruity	0	0	0		0	0	0	0	1	1	1	1	0
	Herbal	0	0	0		1	0	1	0	0	1	0	0	1
	Nutty	1	1	1		0	0	1	0	0	0	0	0	0
	Sweet	0	0	1		0	0	0	1	1	1	1	1	1
	Caramel	0	1	0		0	0	0	1	0	0	1	1	1
	Chocolate	0	0	0		1	1	0	1	0	0	0	0	0
	Citric	0	0	0		0	0	0	0	0	0	1	1	0
	Astringent	0	0	0		1	1	1	0	0	1	0	0	0
	Immature	0	0	0		1	0	0	0	0	0	0	0	0
	Add other attribute	Soft					Dirty							
Overall	Floral	0	0	1		0	0	0	0	0	1	0	0	0
	Fruity	0	0	0		0	0	0	0	1	1	1	1	1
	Herbal	0	0	0		1	0	0	0	0	1	0	0	1
	Nutty	1	1	1		1	0	1	0	0	0	0	0	0
	Sweet	0	0	1		0	0	0	1	1	1	1	1	1
	Caramel	0	1	0		0	0	0	0	0	0	1	1	0
	Chocolate	0	0	0		0	0	0	1	0	0	0	0	0
	Citric	0	0	0		0	0	0	0	0	0	1	1	0
	Astringent	0	0	0		1	1	1	0	0	0	0	0	0
	Immature	0	0	0		1	1	0	0	0	0	0	0	0
	Add other attribute								Soft			Bala-nced		

[1] Bol = Bolivar, [2] Sant = Santander

6c (iii). Environmental characteristics (soils, climate, topography) within spatial domains
The information about the environmental characteristics wihtin spatial domains I - VI is summarized in Table 34 to Table 39. It becomes immediately obvious that the information presented illustrates great differences for the climate, soils and topography between the spatial domains.

There are partly substantial differences in the means of key climate characteristics such as annual total evaporation, cloud frequency, annual rainfall, dew point, diurnal temperaure range, and dry months per year in the four spatial domains of the Northern Region (Table 34). For example in the four spatial domains of the Northern Region, dry months per year ranges from 1.3 in spatial domain IV to 4.0 dry months in spatial domain I. Annual rainfall ranges from 1410 mm in spatial domain 3 to 2390 mm per year in spatial domain I. Annual average cloud frequency indicates high cloud coverage for spatial domain 4 (96%) and relatively low cloud frequency for spatial domain 1 (79%). The Huila/Tolima Region shows less drastic differences in the mean values between the two spatial domains, which are mainly manifested in difference in rainfall, 2290 mm per year in spatial domain V compared with 1740 mm per year in spatial domain VI (Table 34).

The Bonferoni tests show statistically significant differences in both the Northern Region and the Huila/Tolima Region for all characteristics except for solar radiation and dry months per year in the Huila/Tolima Region (Table 35).

The same picture emerges for soil attributes with substantial differences between the mean values for spatial domains for most analyzed characteristics. The differences in the mean value are very obvious for the P – Fe complex, and for the micro nutrients Mn, Zn and Cu. In the Northern Region we also observe large differences in mean values for soil texture, with for example the soils of spatial domain I having only 27% sand content compared with 45% sand content in spatial domain IV (Table 36).

Overall, differences in soil characteristics tend to be more pronounced in the Northern Region than in the Huila/Tolima Region. This is also clearly reflected in the Bonferoni tests, which demonstrate statistically significant differences for most soil attributes in the Northern Region, but only for about 40% of them in the Huila/Tolima Region (Table 37).

As expected, the topography differs less than do the soils and climatic characteristics (Table 38 and Table 39). But there are still substantial differences, for example between the spatial domains in the Northern Region for aspect and altitude. Aspect ranges from 175 degrees in spatial domain III to 229 degrees in spatial domain I, and altitude ranges from 1250 m in spatial domain I to 1600 m in spatial domain IV. Differences in the Huila/Tolima Region are less profound and are manifested mainly in aspect values, with 174 degrees in the spatial domain V as compared to 209 degrees in spatial domain VI.

In summary, we conclude that the observations about the differences in environmental conditions reflect those found for product quality, with some substantial divergence between spatial domains. Also, it is notable that soils are more variable in the Northern Region than in the Huila/Tolima Region. Extending this argument, it becomes clear that the product quality differences in the Huila/Tolima are largely driven by differences in climatic conditions. On the other hand, it appears that soil conditions also influence coffee quality in the Northern Region, together with the climate. In the next sections we shall attempt to investigate this aspect further. But we can already state that the differences in coffee quality assocoiated with differences in environmental characteristics between the new spatial domains fully justifies using them.

Table 34. Means of climate characteristics of the six spatial domains.Table A9.1 and Table A9.2 in Appendix 9 have a complete description of the data.

Climate characteristic	Domain					
	Northern Region			Huila/Tolima Region		
	I	II	III	IV	V	VI
Annual rainfall (mm)	2390	2040	1410	2340	2290	1740
Annual total evaporation (mm yr^{-1})	1350	1080	990	970	880	850
Dew point temperature (°C)	15.4	14.2	14.4	13.6	14.2	14.1
Average temperature (°C)	20.8	19.9	19.8	19.6	20.4	19.7
Diurnal temperature range (°C)	9.9	10.5	9.9	10.6	10.5	10.0
Dry months per year	4.0	4.0	4.3	1.3	1.4	1.5
Annual average cloud frequency (%)	79	86	93	96	95	96
Solar radiation (MJ m^{-2} d^{-1})	22.9	22.1	23.1	24.2	23.4	23.5

Table 35. Bonferoni multivariate tests for selected climate characteristics, comparing the six spatial domains.

Climate characteristic	Domain					
	Northern Region			Huila/Tolima Region		
	I	II	III	IV	V	VI
Annual rainfall (mm)	A	B	C	A	A	B
Annual total evaporation (mm yr^{-1})	A	B	C	C	A	B
Dew point temperature (°C)	A	BC	B	C	A	A
Average temperature (°C)	A	B	B	B	A	B
Diurnal temperature range (°C)	C	B	C	A	A	B
Dry months per year	A	A	A	B	A	A
Annual average cloud frequency (%)	A	C	B	A	B	A
Solar radiation (MJ m^{-2} d^{-1})	B	C	B	A	A	A

Table 36. Means of soils characteristics of the six spatial domains. Table A10.1 – Table A10.3 in Appendix 10 have a complete description of the data.

Soil characteristic	Domain					
	Northern Region				Huila/Tolima Region	
	I	II	III	IV	V	VI
pH	5.2	5.1	4.9	4.6	5.2	5.0
N (%)	0.3	0.3	0.3	0.4	0.3	0.3
Organic matter (%)	6.3	7.9	6.4	8.5	7.7	7.6
K (cmol/kg)	0.2	0.3	0.3	0.3	0.4	0.3
Ca (cmol/kg)	5.6	6.8	3.7	3.7	6.2	5.9
Mg (cmol/kg)g	1.0	1.4	1.0	0.8	1.8	1.5
Al (cmol/kg)	0.7	1.4	1.6	3.4	2.1	1.6
CEC (cmol/kg)	15.3	16.9	13.3	18.4	17.8	16.5
P (mg/kg)	18.8	41.0	20.7	17.9	33.0	50.6
Fe (mg/kg)	264.6	449.2	457.5	639.8	382.2	479.2
Mn (mg/kg)	47.2	115.6	76.5	27.1	67.3	57.0
Zn (mg/kg)	1.4	4.2	4.9	5.1	3.4	5.2
Cu (mg/kg)	3.5	11.4	3.2	3.2	2.5	3.2
Sand (%)	27.0	33.6	31.5	45.0	38.4	30.8
Silt (%)	27.5	31.6	22.7	26.3	28.4	24.5
Clay (%)	45.9	35.0	45.7	28.7	33.5	44.7

Table 37. Bonferoni multivariate tests for selected soils characteristics, comparing the six spatial domains.

Soil characteristic	Domain					
	Northern Region				Huila/Tolima Region	
	I	II	III	IV	V	VI
pH	A	AB	B	C	A	B
N (%)	B	A	B	A	A	A
Organic matter (%)	B	A	B	A	A	A
K (cmol/kg)	B	A	A	AB	A	A
Ca (cmol/kg)	A	A	B	B	A	A
Mg (cmol/kg)g	B	A	B	B	A	A
Al (cmol/kg)	C	B	B	A	A	A
CEC (cmol/kg)	B	AB	C	A	A	B
P (mg/kg)	B	A	B	B	A	A
Fe (mg/kg)	C	B	B	A	B	A
Mn (mg/kg)	C	A	B	A	A	A
Zn (mg/kg)	B	A	A	A	B	A
Cu (mg/kg)	A	A	A	A	A	A
Sand (%)	C	B	B	A	A	B
Silt (%)	B	A	C	B	A	B
Clay (%)	A	B	A	C	B	A

Table 38. Means of topography characteristics of the coffee varieties grown in six spatial domains. Table A11.1 in Appendix 11 has a complete description of the data.

Topographic characteristic	Domain					
	Northern Region				Huila/Tolima Region	
	I	II	III	IV	V	VI
Aspect (°)	229	221	175	193	174	209
Slope (°)	20.4	20.4	19.1	11.9	20.2	18.7
Altitude (masl)	1250	1430	1410	1600	1600	1590

Table 39. Bonferoni multivariate tests for selected topographic characteristics, comparing the six spatial domains.

Topographic characteristic	Domain					
	Northern Region				Huila/Tolima Region	
	I	II	III	IV	V	VI
Aspect (°)	C	B	B	A	A	A
Slope (°)	A	A	A	b	A	A
Altitude (masl)	A	AB	C	BC	B	A

6d. Relationships between environment and bean characteristics

6d (i). Correlation between environment and bean characteristics

Correlations between coffee quality characteristics and environment were analyzed by visualizing the relationships between the principal components of the principal component analyses on both soils and climate as related to the coffee quality characteristics. Table 40 – Table 42 summarize the findings.

The correlation coefficients overall are generally low to moderate. However, it is necessary to take into account that these coefficients are based on principal components that summarize the individual environmental characteristics. Moreover, the coffee quality information represents data from a commercial production environment as opposed to controlled experiments. For these reasons even moderate correlation coefficients are highly likely to represent real trends.

There are several patterns that emerge from these analyses. Most striking is the fact that correlations coefficients in the Northern Region indicate an impact of both climate and soils on coffee quality characteristics. In contrast, in the Huila/Tolima Region only climate has a discernable impact (Table 40).

In the Northern Region in both spatial domains I and II there are clearly identifiable relationships between coffee quality and soil characteristics. In spatial domain I, principal component (PC) 2 for soil mainly affects the biochemical quality characteristics of the coffee, while PC 3 mainly affects physical coffee quality characteristics (Table 41). In spatial domain I, PC 2 is mainly dominated by a pH–Fe–P–soil texture complex, while. PC 3 is dominated by an organic matter–texture–copper complex. In spatial domain II, soils PCs 1 and 3 affect the sensoric characteristics, while soils PC 3 affects physical coffee quality characteristics. PC 1 is mainly dominated by a pH–base ion complex. PC 3 in this spatial domain is dominated by a texture–P influence.

In spatial domain I, climate PCs 1 and 3 are correlated with the biochemical characteristics of coffee quality and with sensoric/physical characteristics, respectively. Climate PC 1 in this spatial domain is influenced by temperature, dew point, and altitude. Climate PC 3 in this spatial domain is strongly driven by an aspect–solar radiation complex. In spatial domain II, climate PC 2 negatively influences the sensoric characteristics of coffee quality. This component is strongly dominated by precipation and evaporation. In spatial domain III, climate PC 1, which is a temperature–dewpoint–altitude complex, positively impacts the sensoric and physical characteristics of coffee quality. In spatial domain IV, climate PC 1, a complex of temperature, dewpoint and altitude, also positively impacts the sensoric characteristics of coffee quality.

Table 40. Contributions of the first three principal components of soils and of climate selected physical characteristics, comparing yield factors and screen sizes across the six spatial domains.

Physical charateristic	Domain/contributor	Soil characteristics			Climate characteristics		
	Principal component	1	2	3	1	2	3
	Northern Region						
Yield factor 13	I	0.049	0.359	0.203	-0.487	-0.169	0.011
	II	0.165	0.024	0.199	-0.055	-0.077	-0.168
	III	-0.149	-0.02	-0.002	-0.037	-0.043	0.086
	IV	-0.018	0.163	0.012	-0.125	-0.118	0.227
Yield factor 14	I	-0.044	-0.107	-0.206	0.388	-0.006	0.151
	II	-0.267	-0.372	-0.001	0.055	0.238	0.177
	III	0.026	0.2	-0.014	0.086	-0.211	-0.22
	IV	-0.279	-0.141	-0.124	0.167	0.011	0.087
Yield factor 15	I	-0.167	0.134	0.173	0.096	0.099	0.065
	II	0.051	-0.139	-0.083	-0.124	-0.131	-0.028
	III	0.069	0.025	-0.088	0.062	-0.039	0.008
	IV	-0.27	0.07	-0.086	0.266	-0.191	0.353
Screen size 17	I	-0.117	-0.05	-0.059	-0.008	0.292	-0.21
	II	-0.069	0.035	-0.088	0.139	0.061	-0.205
	III	-0.062	0.055	-0.158	0.271	0.122	0.112
	IV	0.061	-0.005	0.071	0.114	0.016	-0.133
Screen size 18	I	0.213	0.297	-0.071	-0.25	-0.167	-0.11
	II	0.055	0.267	-0.228	-0.033	0.013	-0.08
	III	0.003	-0.064	-0.013	0.131	0.181	-0.032
	IV	0.133	0.167	0.099	-0.093	-0.045	-0.023
	Huila/Tolima Region						
Yield factor 13	V	0.072	0.086	-0.038	-0.135	0.227	0.022
	VI	-0.066	-0.124	0.16	-0.12	0.035	-0.122
Yield factor 14	V	0.062	-0.073	-0.155	0.242	-0.157	0.105
	VI	0.042	-0.024	-0.014	0.223	-0.284	-0.1
Yield factor 15	V	-0.104	0.062	0.053	0.002	-0.022	-0.147
	VI	-0.019	-0.004	0.106	0.13	-0.447	-0.231
Screen size 17	V	-0.247	-0.045	0.014	0.182	-0.152	0.122
	VI	0.055	0.045	-0.061	0.172	0.247	0.232
Screen size 18	V	0.036	-0.039	-0.14	-0.148	-0.174	-0.15
	VI	-0.134	-0.122	0.066	0.036	-0.105	-0.041

Table 41. Contributions of the first three principal components of soils and climate to the sensory characteristics of coffee from the four spatial domains of the Northern Region.

Sensory charateristic	Domain/contributor	Soil characteristics			Climate characteristics		
	Principal component	1	2	3	1	2	3
Fragrance ans aroma	I	-0.162	-0.145	-0.3	0.279	-0.041	0.036
	II	0.126	0.013	-0.033	0.022	-0.08	0.145
	III	0.064	0.045	-0.025	0.083	-0.082	0.074
	IV	-0.151	-0.031	-0.042	0.124	0.183	0.078
Flavor	I	0.015	-0.045	0.064	0.093	0.146	0.17
	II	0.202	0.061	-0.118	0.024	-0.16	-0.186
	III	-0.039	-0.075	-0.159	0.203	0.035	0.189
	IV	-0.153	-0.012	-0.065	0.202	0.106	0.053
Aftertaste	I	-0.062	-0.074	0.029	0.005	0.126	0.257
	II	0.183	0.049	-0.179	0.126	-0.108	-0.01
	III	0.035	0.021	-0.217	0.221	0.11	0.08
	IV	-0.166	-0.005	-0.074	0.201	0.064	0.034
Acidity	I	-0.022	-0.012	0.018	0.039	0.244	0.199
	II	0.315	0.066	-0.263	0.145	-0.177	-0.11
	III	-0.102	0.001	-0.101	0.004	-0.02	0.118
	IV	-0.043	0.037	-0.015	0.112	-0.006	-0.031
Body	I	0.075	-0.011	0.028	0.051	-0.017	0.177
	II	0.304	-0.038	-0.185	0.131	-0.202	0.003
	III	0.041	0.075	-0.287	0.211	0.047	0.138
	IV	-0.193	-0.02	-0.041	0.227	0.07	0.034
Balance	I	-0.052	-0.024	0.092	0.088	0.151	0.21
	II	0.117	0.014	-0.28	0.114	-0.044	0.026
	III	0.083	0.045	-0.2	0.216	0.072	0.099
	IV	-0.193	-0.018	-0.094	0.225	0.106	0.04
Uniformity	I	-0.174	-0.103	-0.09	-0.018	0.021	0.102
	II	0.037	0.11	0.048	-0.105	0.016	0.09
	III	0.055	0.026	-0.176	-0.016	-0.001	0.104
	IV	0.021	-0.072	-0.073	0.012	-0.043	-0.018
Clean cup	I	0.297	0.143	-0.036	0.011	-0.177	-0.05
	II	0.235	-0.084	-0.216	0.157	-0.208	0.023
	III	-0.045	0.074	-0.073	0.238	-0.002	0.151
	IV	-0.095	0.024	-0.081	0.137	0.174	-0.037
Sweetness	I	-0.063	0.181	0.15	-0.07	0.216	0.239
	II	0.296	0.244	-0.092	0.028	-0.234	-0.07
	III	-0.149	-0.012	-0.098	0.097	-0.051	0.112
	IV	-0.147	0.022	0.055	0.105	0.012	0.09
Overall	I	-0.012	-0.005	0.002	0.051	0.086	0.22
	II	0.269	-0.096	-0.209	0.178	-0.203	0.031
	III	0.161	0.127	-0.136	0.291	0.077	0.116
	IV	-0.131	-0.011	-0.077	0.208	0.119	-0.036
Final score	I	0.008	-0.001	-0.053	0.017	-0.059	0.251
	II	0.25	0.035	-0.192	0.107	-0.169	-0.011
	III	0.016	0.043	-0.188	0.205	0.03	0.148
	IV	-0.154	-0.008	-0.063	0.192	0.099	0.023

Table 42. Contributions of the first three principal components of soils and climate to the sensory characteristics of coffee from the two spatial domains of the Huila/Tolima Region.

Sensory charateristic	Domain/contributor	Soil characteristics			Climate characteristics		
	Principal component	1	2	3	1	2	3
Fragrance and aroma	V	-0.054	0.094	0.029	0.177	0.031	0.101
	VI	-0.071	0.151	0.038	0.204	0.017	0.026
Flavor	V	-0.058	0.137	0.022	0.218	0.026	0.014
	VI	-0.173	0.115	0.058	0.202	-0.016	0.062
Aftertaste	V	-0.076	0.182	0.035	0.222	-0.021	0.028
	VI	-0.064	0.151	0.063	0.200	0.035	0.000
Acidity	V	-0.045	0.000	-0.107	-0.087	-0.008	-0.024
	VI	-0.059	0.150	0.027	0.188	0.048	0.043
Body	V	-0.171	0.141	0.058	0.175	-0.028	0.017
	VI	-0.036	0.167	0.051	0.200	0.032	0.011
Balance	V	-0.092	0.152	0.037	0.204	0.015	0.032
	VI	-0.156	0.054	0.005	0.112	0.149	-0.055
Uniformity	V	-0.134	0.071	0.038	0.066	0.221	-0.021
	VI	-0.148	0.069	0.036	0.078	0.238	-0.048
Clean cup	V	-0.128	0.114	0.046	0.064	0.227	-0.048
	VI	-0.174	0.037	0.052	0.017	0.247	-0.065
Sweetness	V	-0.122	0.042	0.042	0.056	0.256	-0.040
	VI	-0.162	0.135	-0.183	-0.055	0.115	0.012
Overall	V	-0.143	0.087	-0.010	0.024	0.256	-0.075
	VI	-0.149	0.015	0.056	0.044	0.230	-0.054
Final score	V	-0.128	0.050	0.025	0.043	0.234	-0.034
	VI	-0.157	0.071	0.022	0.053	0.246	-0.049

In spatial domain 5 in the Huila/Tolima Region, climate PC 1 impacts positively on the sensoric attributes of coffee quality. PC 1 in spatial domain 5 is a temperature–dew point–altitude complex combined with the number of dry months. Spatial domain 6 in the Huila/Tolima Region exhibits a positive relationship between climate PC 2 and sensoric characteristics of coffee quality, and a negative relationship between PC 2 and the biochemical characteristics of coffee quality. PC 2 in spatial domain 6 is dominated by a complex of aspect–solar radiation–slope, and by the range of diurnal temperature.

6d (ii). Spatial distribution of relationships between environment and bean characteristics
The spatial distribution of relationships between environment and coffee quality characteristics using the CaNaSTA procedure is illustrated using a subset of the sensoric characteristics of coffee quality. Please note, that the colored areas in the Figures may go beyond the actual coffee growing regions. Therefore, it is important to take into the account the actual distritbution of coffee growing areas (e.g. for example by using the SICA data base). Please note also, that this section presents analyses that have been conducted for the complete Northern Region and the complete Huila/Tolima Region, respectively. For a more domain specific interpretation it is indispensible to also include the domain specific analyses in the interpretation. Figures for these analyses are included in the Appendix. Figure 29 – Figure 44, and Figure A12. 1 Figure A12. 24 in Appendix 12, which give a visual representation of the probability analyses. The figures show pairs of maps, the first of which shows the most likely class for the particular sensoric characteristic at each location. The second map of the pair shows the relative suitability of sites within the area covered by the analysis for growing coffee with a high value for the particular sensoric characteristic (see Section 5c (ii) on page 21). Obviously the two indicators are related.

The maps also show the superimposed boundaries of both the expert domains and the spatial domains, respectively.

We emphasize that no classification system will be able to capture and explain all the variability within a specific, classified area. However, we do expect that a suitable system of classification will capture the major patterns. In our case, we expect to display in a meaningful and useful manner the spatial distribution of coffee quality to underpin the implementation of a system of denomination of origin.

Final Score (Figure 29 – Figure 32):

In the Northern Region, there are two spatial domains that are fairly homogenous (spatial domains I and III), and two (spatial domains II and IV) that show a heterogenous distribution of the sensory characteristic of final score. Spatial domain I is dominated by lower ranked final score results, but spatial domain III indicates a high potential for good final score values. Spatial domains II and IV contain areas with both high and low values for final score. The difference appears to be that in spatial domain II the changes between areas with high values and low values areas are rather gradual, whereas in the spatial domain IV there is frequently close juxtapositions of very high and very low areas. These findings are also confirmed in the map that outlines the suitability for high final scorevalues for the Nothern Region: three out of the four domains clearly have a high potential to obtain excellent final score values, but particularly so in spatial domain III. However, as later also described for domain V, the cupping data indicate lower sensory quality for domain V than the Canasta analyses indicate in their predictions. This clearly indicates the situation that we have relatively high potential for quality in the area, which however is no yet being realized by coffee growers.

The spatial domains for the Huila/Tolima Region contain both areas that have potential for low and high Final score values. Spatial domain V contains proportionally a higher percentage of areas of high final score than does spatial domain VI. This is not unexpected as much of spatial domain VI comprises marginal areas at low altitude. Spatial domain VI does, however, have high potential in the northeastern and southwestern parts of the domain.

The maps for the individual spatial domains (Appendix 12) basically provide more detail than the composite maps. This is important to note, as the compound maps were generated by CaNaSTA for the whole region, whereas the individual maps have been generated separately for each spatial domain.

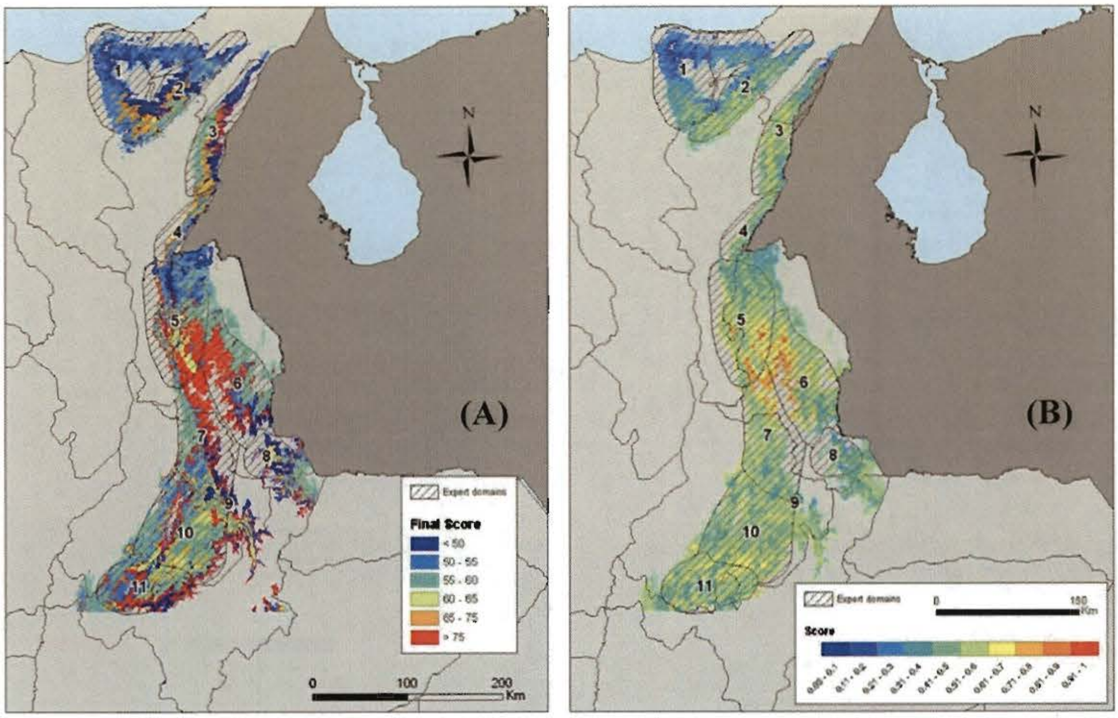


Figure 29. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Final score compared with the expert domains defined by FNC expert opinion in the Northern Region.

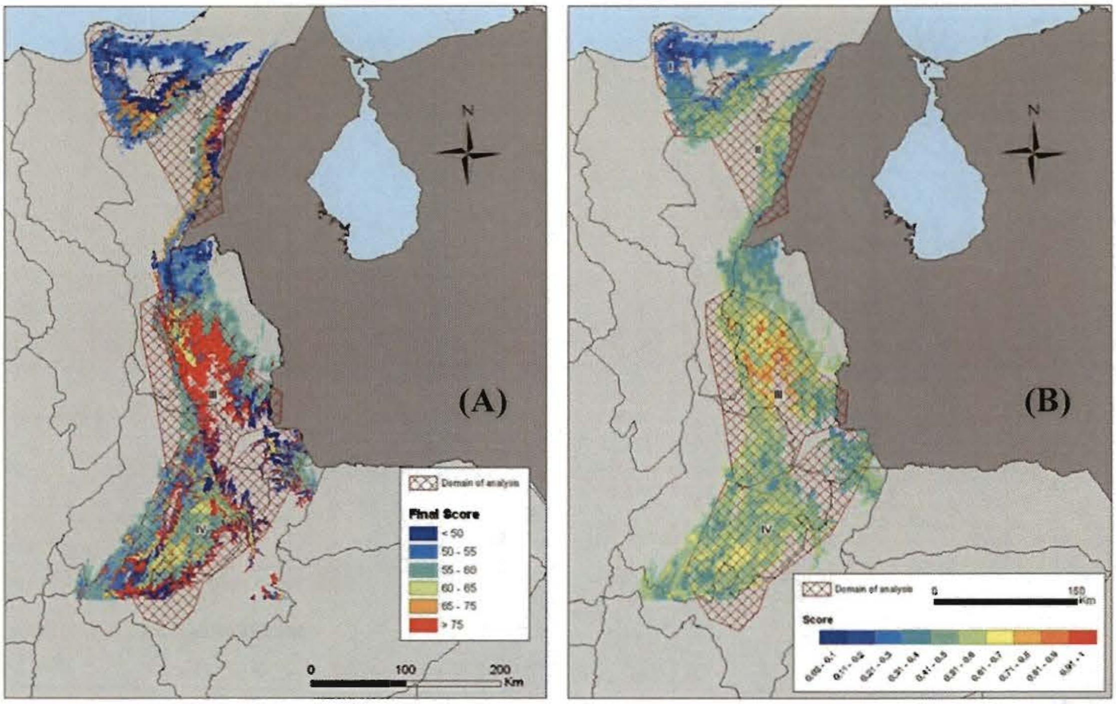


Figure 30. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic of Final score compared with the four identified spatial domains in the Northern Region.

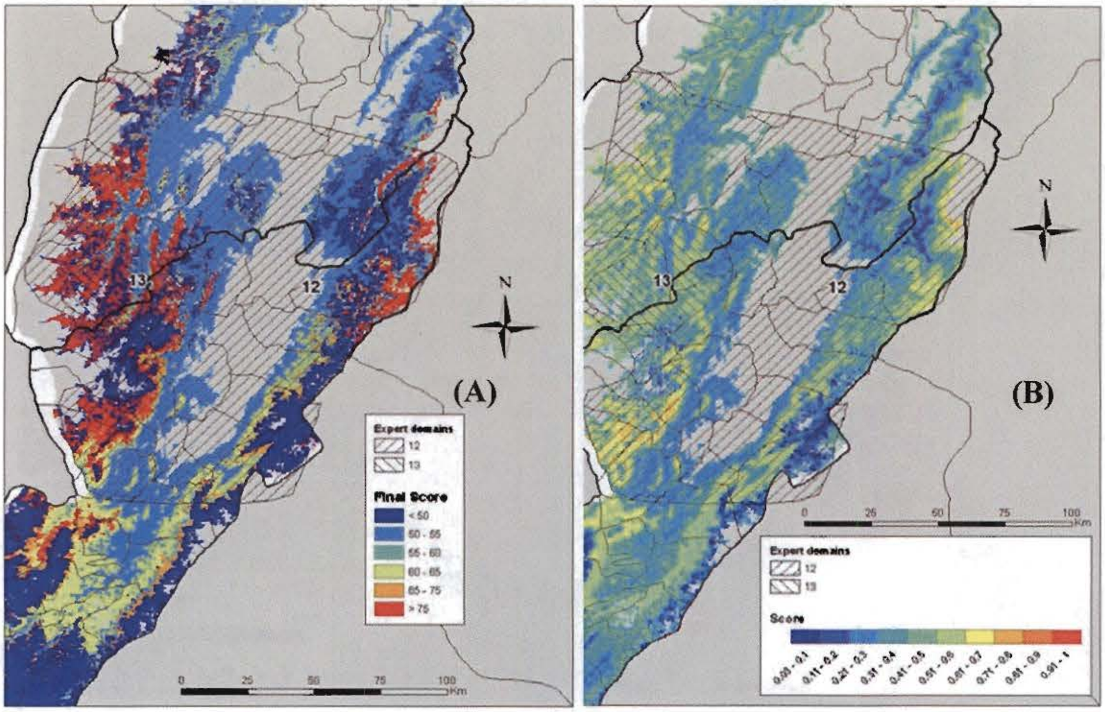


Figure 31. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic of Final score compared with the expert domains defined by FNC expert opinion in the Huila/Tolima Region.

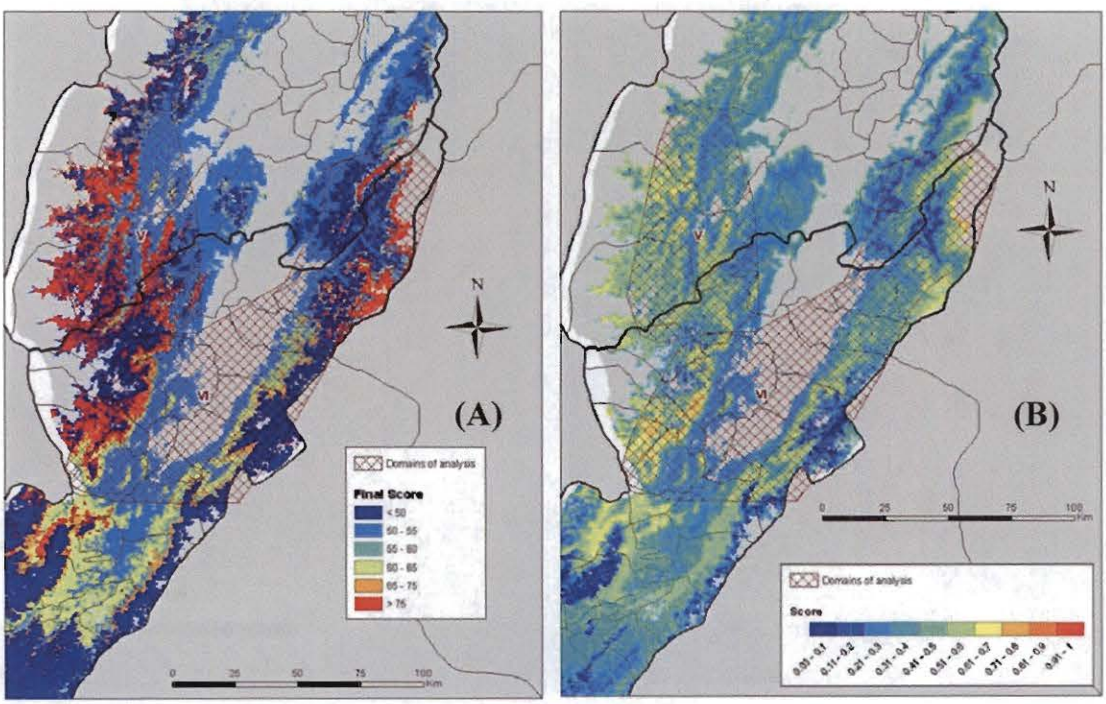


Figure 32. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic of Final score compared with the two identified spatial domains in the Hula/Tolima Region.

Fragrance and aroma (FA, Figure 33 – Figure 36):

Generally FA increases in the Northern Region from North to South. Only spatial domain IV has large areas in the highest FA class. Still, reasonable zones with high FA values are found within the spatial domain III. Spatial domains I and II only have very limited occurrence of favorable areas for high FA. In spatial domain II these areas are concentrated along the eastern slopes of the mountain range.Spatial domain III shows has most areas with high FA values in the center of the domain. Spatial domain IV is actually dominated by areas with favorable values for FA. There are less favorable zones only in the central part of this spatial domain.

Spatial domains V and VI in the Huila/Tolima Region show different patterns for FA. Spatial domain V is characterized by a pattern of highly heterogenous distribution of FA, which range from very favorable zones to unfavorable zones, all very close by one another. Spatial domain VI on the other hand shows more gradual changes, and in general less spatial variation in FA especially in the southern part of the spatial domain.

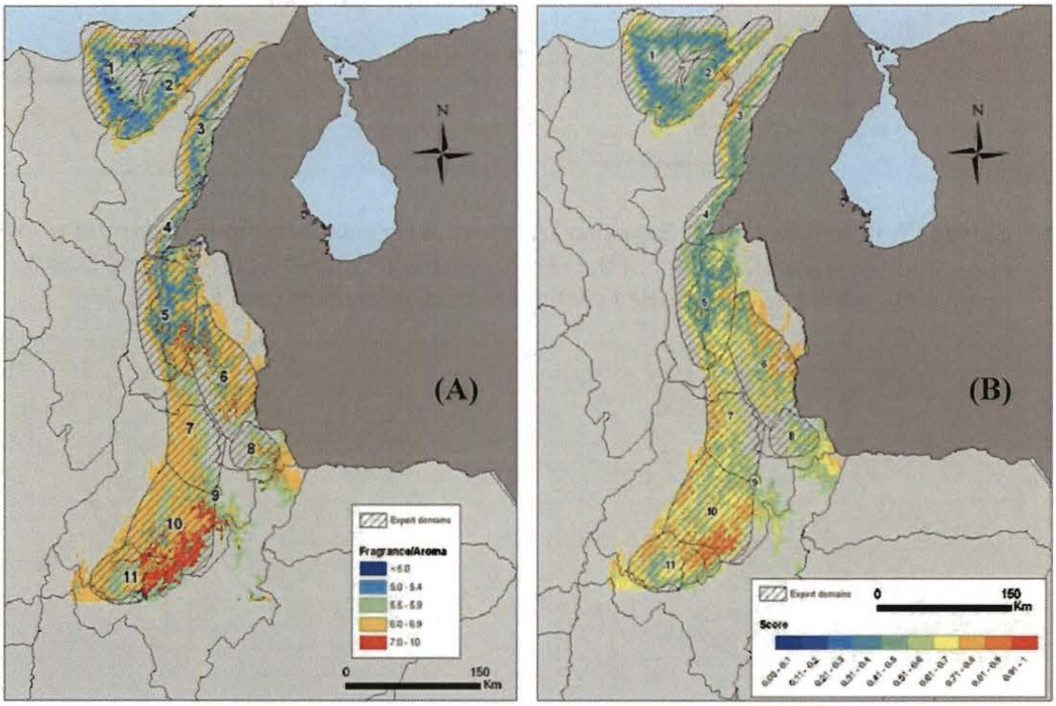


Figure 33. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Fragrance and aroma compared with the expert domains defined by FNC expert opinion in the Northern Region.

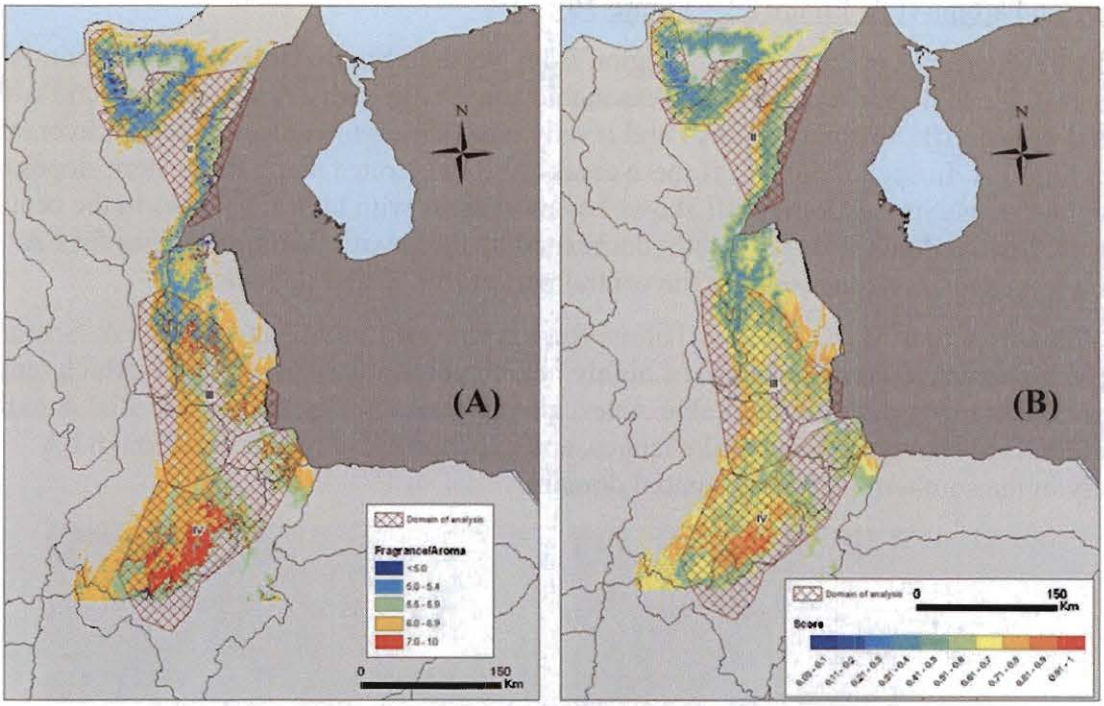


Figure 34. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic of Fragrance and aroma compared with the four identified spatial domains in the Northern Region.

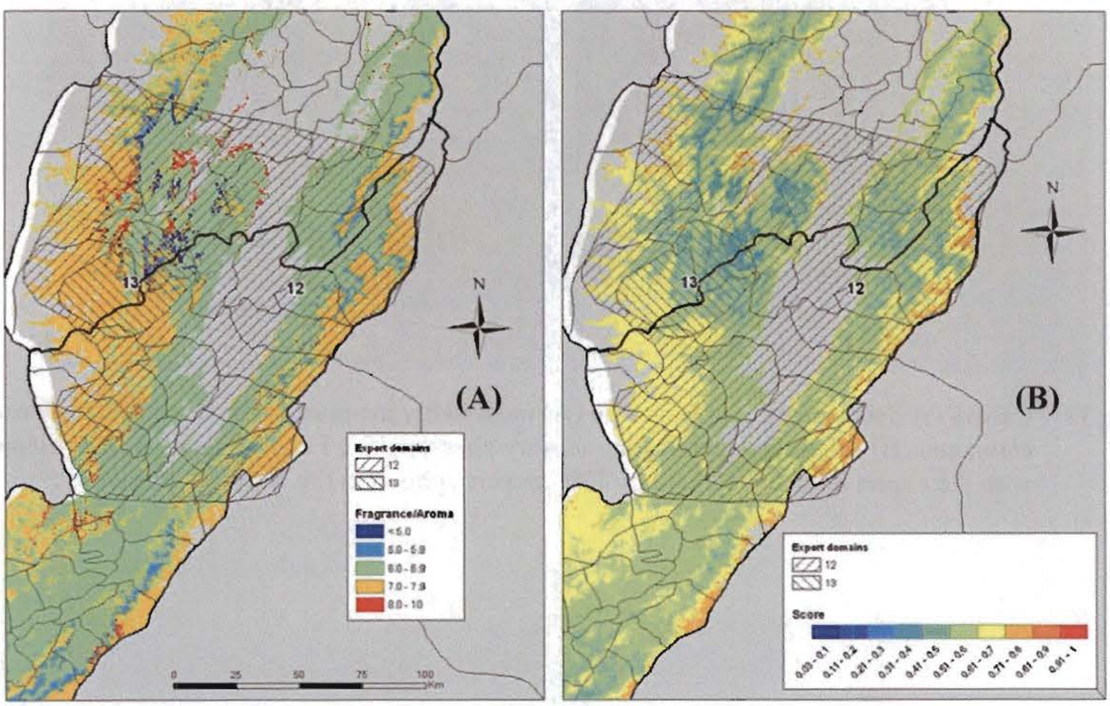


Figure 35. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic of Fragrance and aroma compared with the expert domains defined by FNC expert opinion in the Huila/Tolima Region.

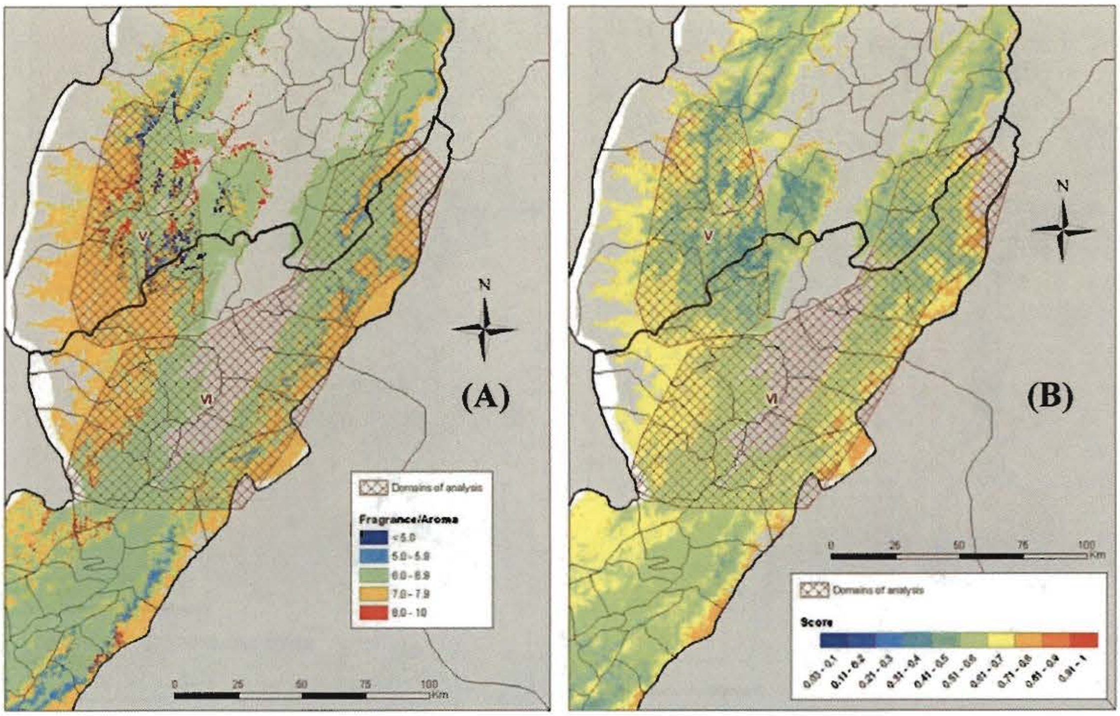


Figure 36. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic of Fragrance and aroma compared with the two identified spatial domains in the Hula/Tolima Region.

Acidity (Figure 37 – Figure 40):

Considering the distribution for the whole Northern Region, acidity tends to be highest in the central part of this Region, specifically so in spatial domains II and III. The northern tip of the region has very low acidity values, while the southern extension has reasonable areas with high acidity although areas with lower value dominate. Spatial domain I only has in its southern tip some potential for high acidity. Spatial domain II has fairly extensive areas with potential for higher acidity values, usually these are distributed in a North–South direction. Spatial domain III is dominated by areas with high acidity potential, with a particular prevalence in the central part. Spatial domain IV has moderately large areas with high acidity. These are, however, concentrated in the eastern and northern parts of the domain.

The Huila/Tolima Region displays a heterogenous spatial distribution pattern for acidity values. The largest proportion of areas with potential for higher acidity is in the Norther part of the Huila/Tolima Region. Spatial domain V tends to have proportionally less area with low potential for high acidity than spatial domain VI. The whole east northern area of spatial domain VI has very high potential for acidity, while the southern part is dominated by areas with low potential, with local islands with high acidity.

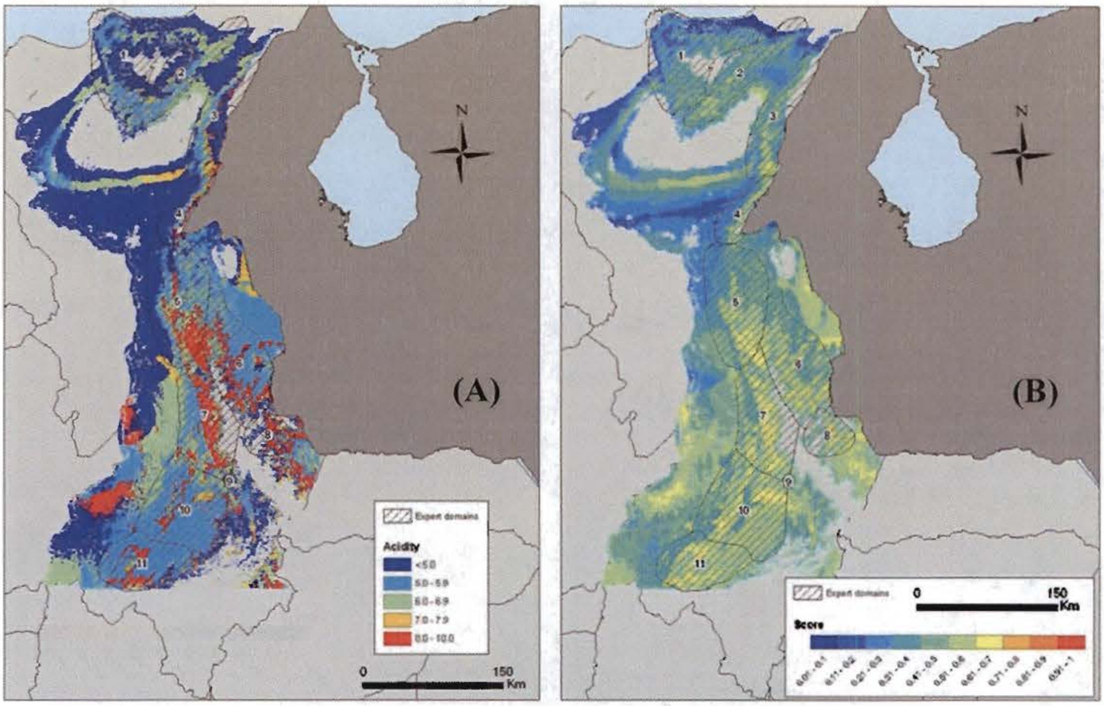


Figure 37. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Acidity compared with the expert domains defined by FNC expert opinion in the Northern Region.

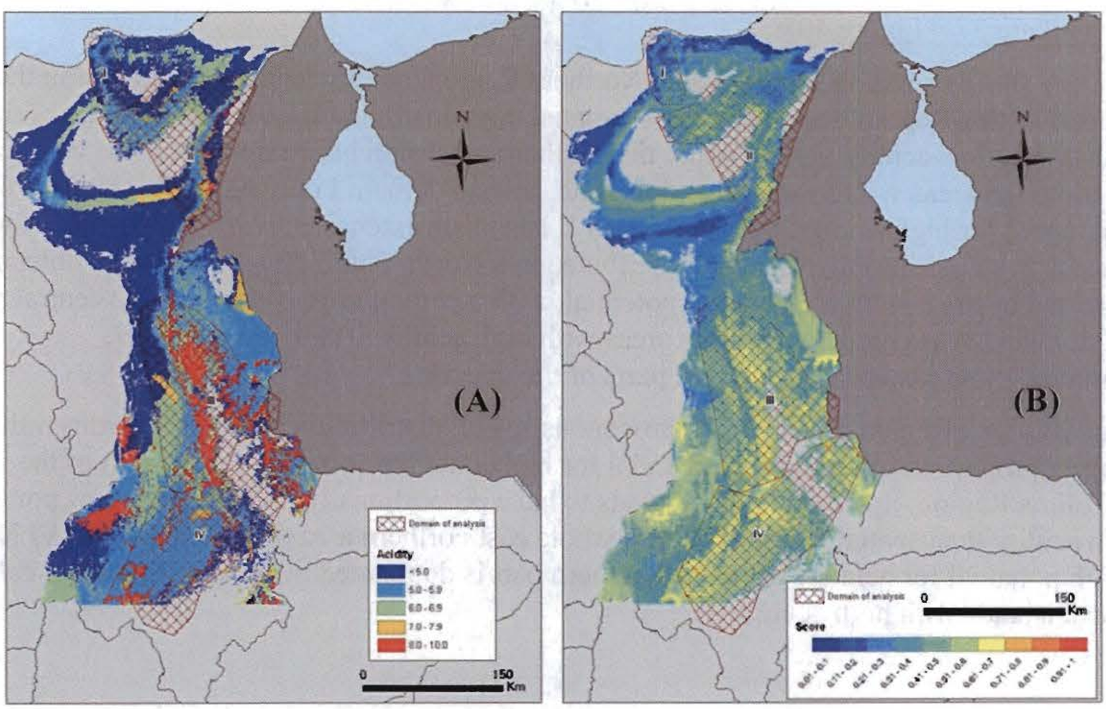


Figure 38. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic of Acidity compared with the four identified spatial domains in the Northern Region.

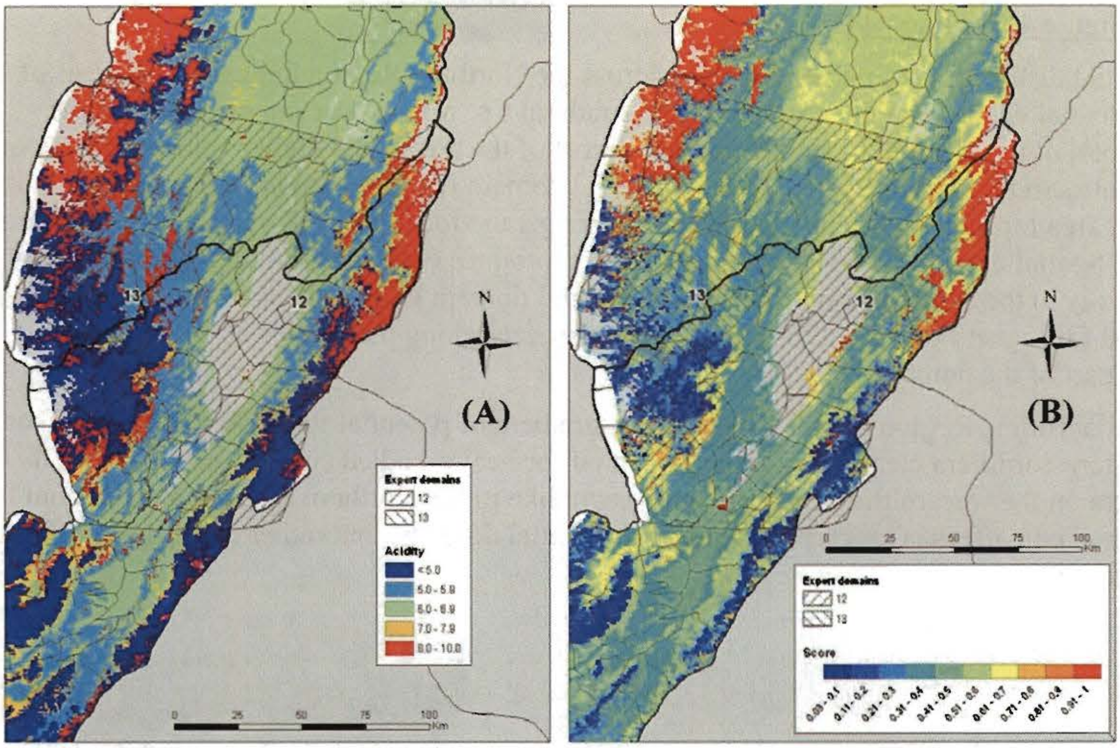


Figure 39. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic of Acidity compared with the expert domains defined by FNC expert opinion in the Huila/Tolima Region.

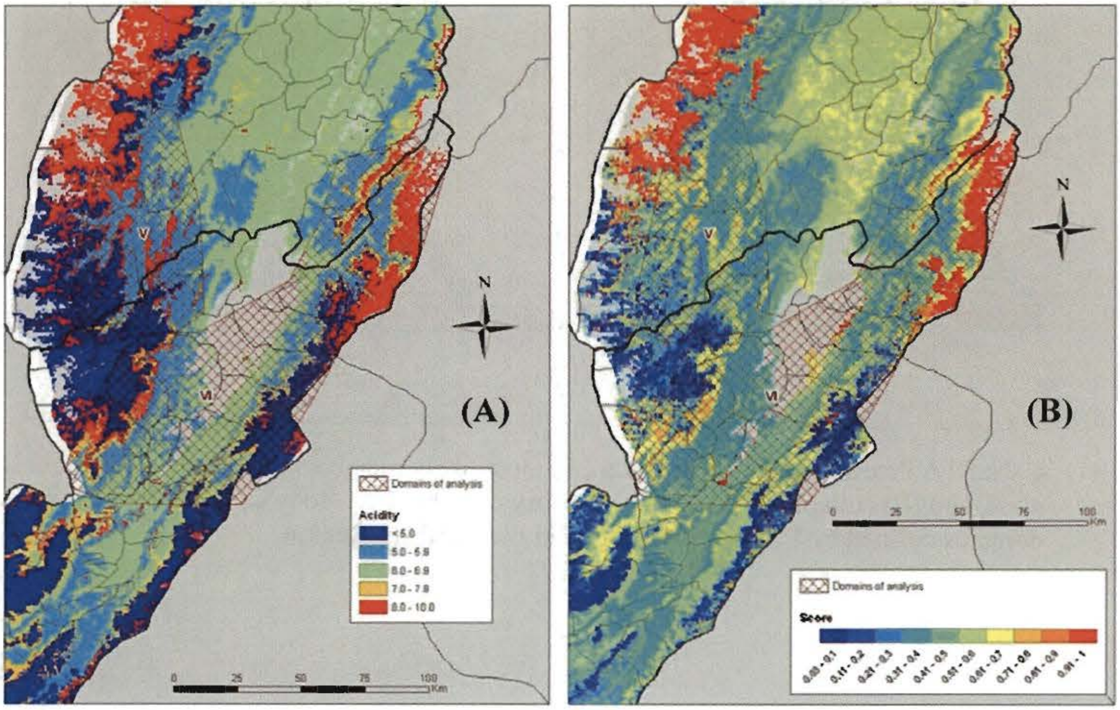


Figure 40. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic of Acidity compared with the two identified spatial domains in the Hula/Tolima Region.

Body (Figure 41 – Figure 44):

The spatial distribution of values for body across the Northern Region indicates that areas of high potential occur from the northern to the southern tip, with a large coherent area in the central part of the Region. In spatial domain I most of the areas with higher values for body are in the northern part of this spatial domain. Spatial domain II is likely to have proportionally the smallest area for high values for body values amongst the four spatial domains in the Northern Region. Spatial domain III has large areas that can produce coffees with high values for body, specifically in the center and north of center. Spatial domain IV has also moderate areas with potential for higher bodied coffees, most of which occur along the fringes of the central and eastern part of the domain.

The Huila/Tolima Region shows a west to east gradient of potential for higher bodied coffees. The eastern cordillera clearly has higher potential for heavy-bodied coffees than the central cordillera on the west of the Region. Therefore, unlike in the Northern Region, the potential for higher bodied coffees is less represented by the spatial domains, but rather by the geography of the cordilleras.

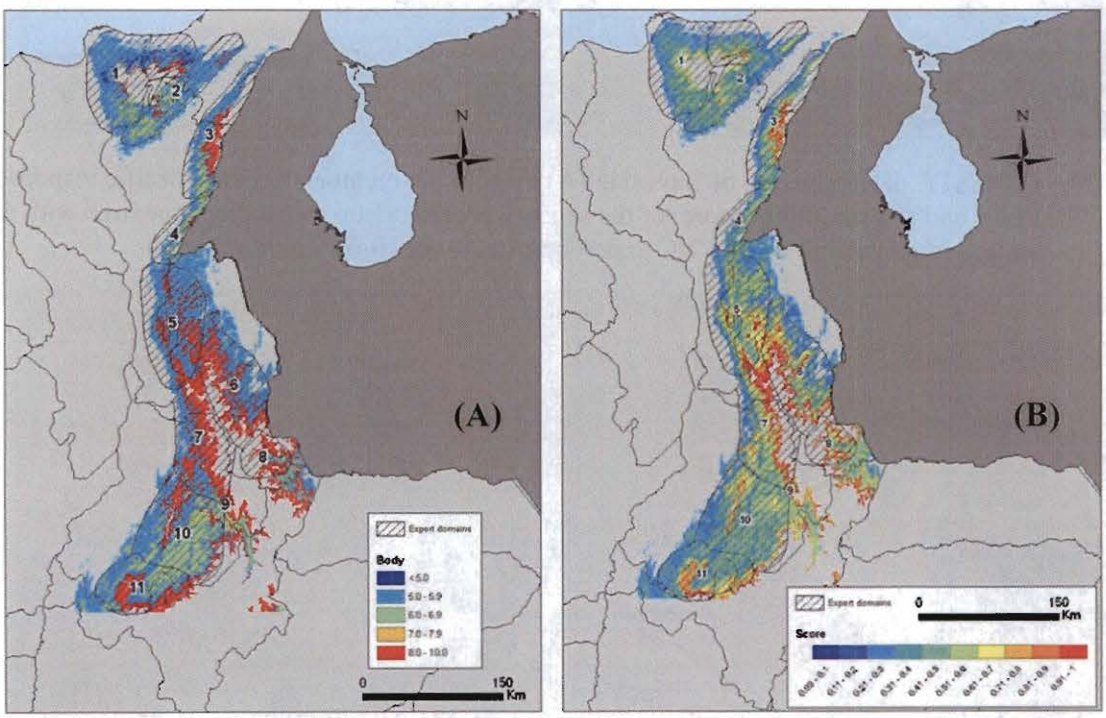


Figure 41. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Body compared with the expert domains defined by FNC expert opinion in the Northern Region.

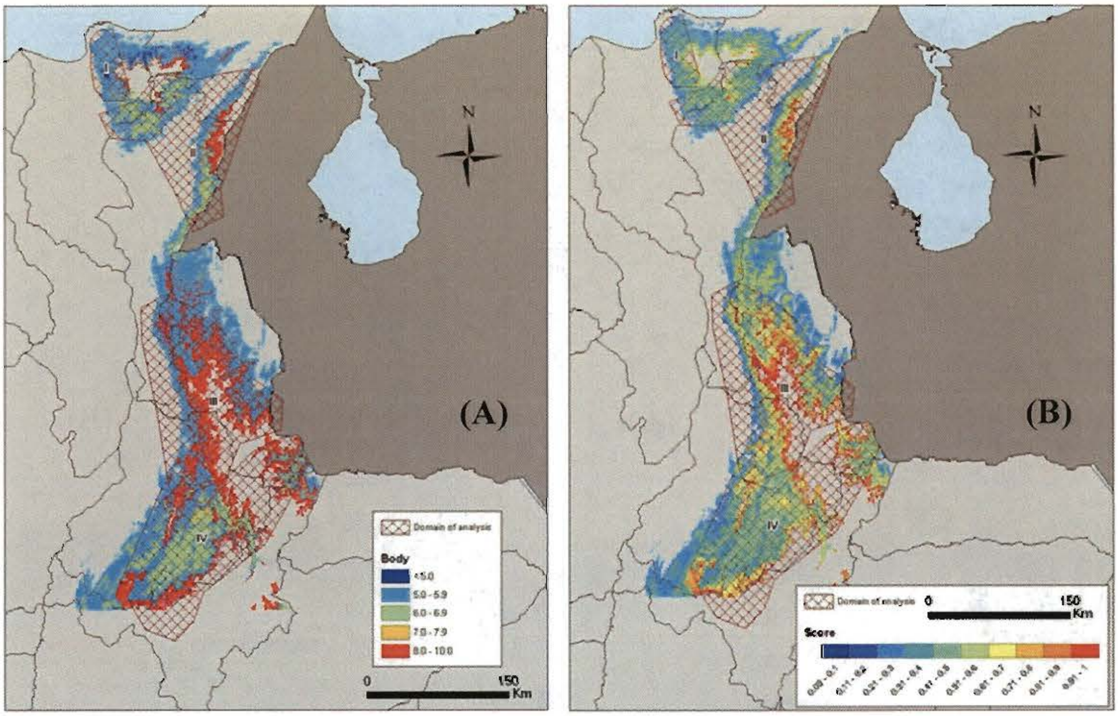


Figure 42. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic of Body compared with the four identified spatial domains in the Northern Region.

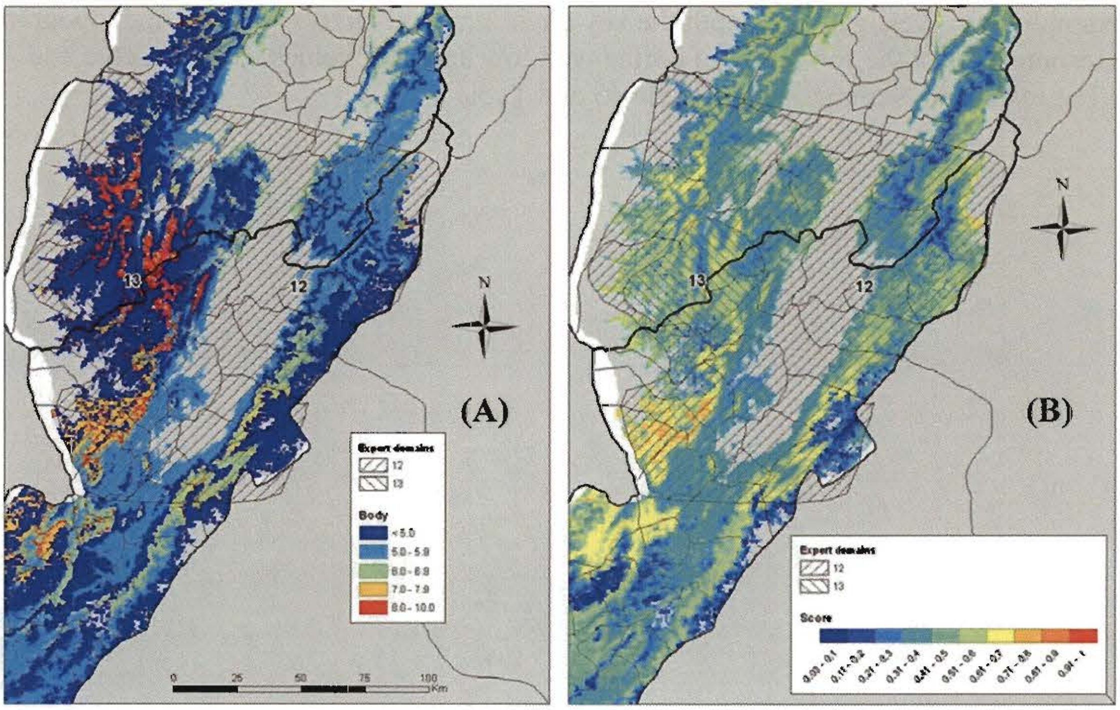


Figure 43. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic of Body compared with the expert domains defined by FNC expert opinion in the Huila/Tolima Region.

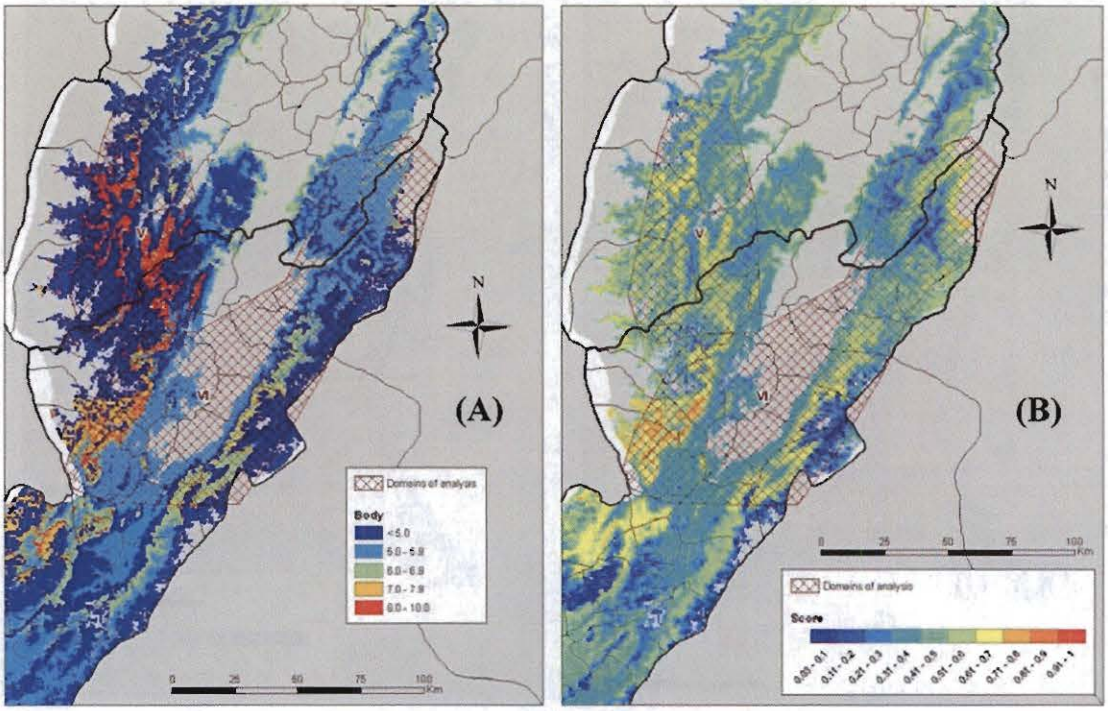


Figure 44. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic of Body compared with the two identified spatial domains in the Hula/Tolima Region.

As an example, we investigated in depth the key environmental factors (climate and topography) that are responsible for the formation of coffee with low and high values for final score and acidity. The results are summarized in Table 43 and Table 44.

Table 43. Climatic and topographic factors that influence the quality characteristics of final score and acidity in the Northern Region.

Positive influence		
Factors	Range	Importance
Final score		
Solar radiation (MJ m^{-2} d^{-1})	19 – 20	2.09
Annual average cloud frequency (%)	87 – 90	2.04
Acidity		
Altitude (m)	1321 – 1600	3.41
Average temperature (°C)	18.6 – 20.2	3.36
Altitude (m)	1601 – 1890	3.17
Average temperature (°C)	17.1 – 18.5	2.91
Altitude (m)	1891 – 2170	2.41
Average temperature (°C)	20.3 – 21.8	2.3
Average temperature (°C)	15.4 – 17.0	2.26
Annual average cloud frequency (%)	91 – 94	2.25
Annual total evaporation (mm yr^{-1})	741 – 885	2.17
Dry months per year	3.0 – 3.2	2.07
Annual average cloud frequency (%)	95 – 98	2.01
Negative influence		
Factors	Range	Importance
Final score		
Annual average cloud frequency (%)	75 – 78	3.82
Annual total evaporation (mm yr^{-1})	1321 – 1470	2.59
Diurnal temperature range (°C)	9.1 – 9.4	2.18
Acidity		
Annual average cloud frequency (%)	75 – 78	2.6

Across the Northern Region, final score is positively affected by specific values for solar radiation and by high cloud frequency. Lower values for final score are mainly driven by high annual total evaporation, low cloud frequency and high diurnal temperature range. Acidity is positively influenced by a range of factors, most notably including altitude (1300–1890 m), and average temperature (17.1–20.2 degrees centigrade). Low cloud frequency tends to reduce acidity.

Across the Huila/Tolima Region, final score is positively influenced by altitudes between 1575 and 1800 m and annual rainfall values between 1550 and 1750 mm. High average temperatures and low altitude values reduce the final score values. Acidity is positively affected by solar radiation, diurnal temperature range and dry months per year. Low altitudes and both low and high average temperatures reduce acidity.

Table 44. Climatic and topographic factors that influence the quality characteristics of Final score and Acidity in the Huila/Tolima Region.

Factors	Range			Importance
Positive influence				
Final score				
Altitude (m)	1575	-	1800	2.08
Annual rainfall (mm)	1550	-	1750	2.00
Acidity				
Solar radiation (MJ m-2 d-1)	20	-	20	2.60
Diurnal temperature range (°C)	9.0	-	9.3	2.51
Dry months per year	5.0	-	5.2	2.34
Negative influence				
Factors	Range			Importance
Final score				
Average temperature (°C)	23.6	-	25.05	3.15
Altitude (m)	675	-	900	2.59
Acidity				
Altitude (m)	900	-	1125	3.11
Average temperature (°C)	16.1	-	17.5	2.60
Average temperature (°C)	22.1	-	23.5	2.38

6d (iii). Climatological uniqueness of analyses domains

For each of the six spatial domains we conducted a climatic and geographic similarity analysis: For each spatial domain, we determined the similarity within the rest of Colombia, within Latin America, Asia, and Africa. Please note that these analyses do not include topography (i.e. most importantly does not consider altitude) as a factor. If we were to include topography, similarity with regions elsewhere would further decrease and the rareness of Colombian coffee growing regions even further increase. Figure 45–Figure 56 summarize the analyses visually, and Table 45 details the results.

Spatial domain I: This domain is unique. There are only very few areas in Colombia and in Latin America (Brazil, Panama) that have similar climatic conditions. In Africa, there are some small, similar regions in Tanzania, Angola, Congo and Cameroon. In Asia, only southeast China and some small zones in Southeast Asia are identified as similar to spatial domain I. Mostly the similarities are low.

Spatial domain II: There are a few areas with similar climatic conditions in Colombia, in Peru, in Brazil and as well in Central America. But the extent of these zones is fairly small. There are no similar areas in Africa, except for the eastern coast of Madagascar. Some areas of China and Central Indonesia show some similarity with spatial domain II.

Spatial domain III: This spatial domain too is unique. There is almost no similar areas within Colombia, and just very few, and small areas in Brazil, Paragay, Venezuela and Central America that show similarity at low probability levels. There is also only very limited similarity between spatial domain III and Africa and Asia, with a few areas in Uganda / Democratic Republic of Congo, and in China.

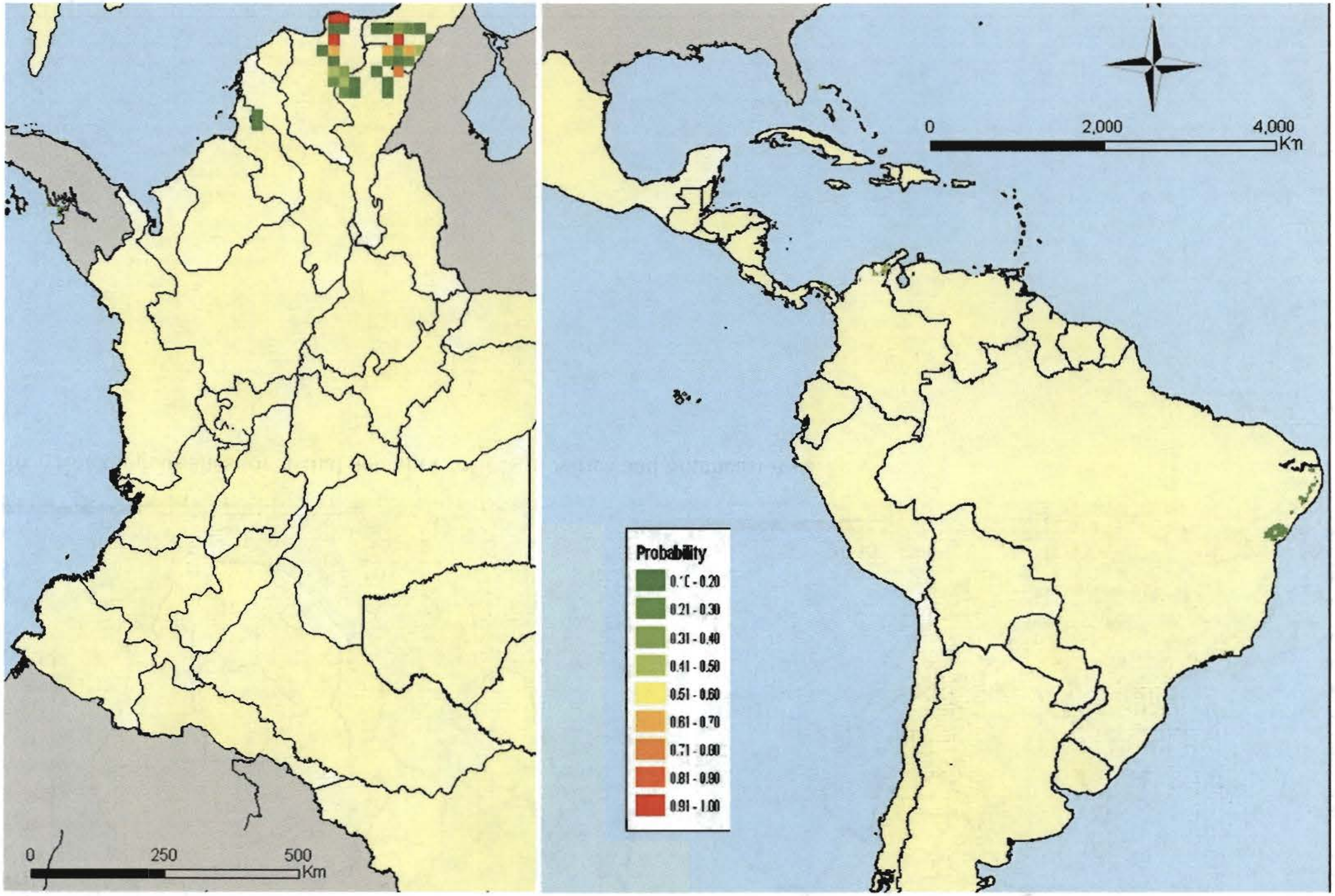


Figure 45. Occurrence of spatial domain I coffees and their homologous sites in Colombia and their homologous sites elsewhere in South America.

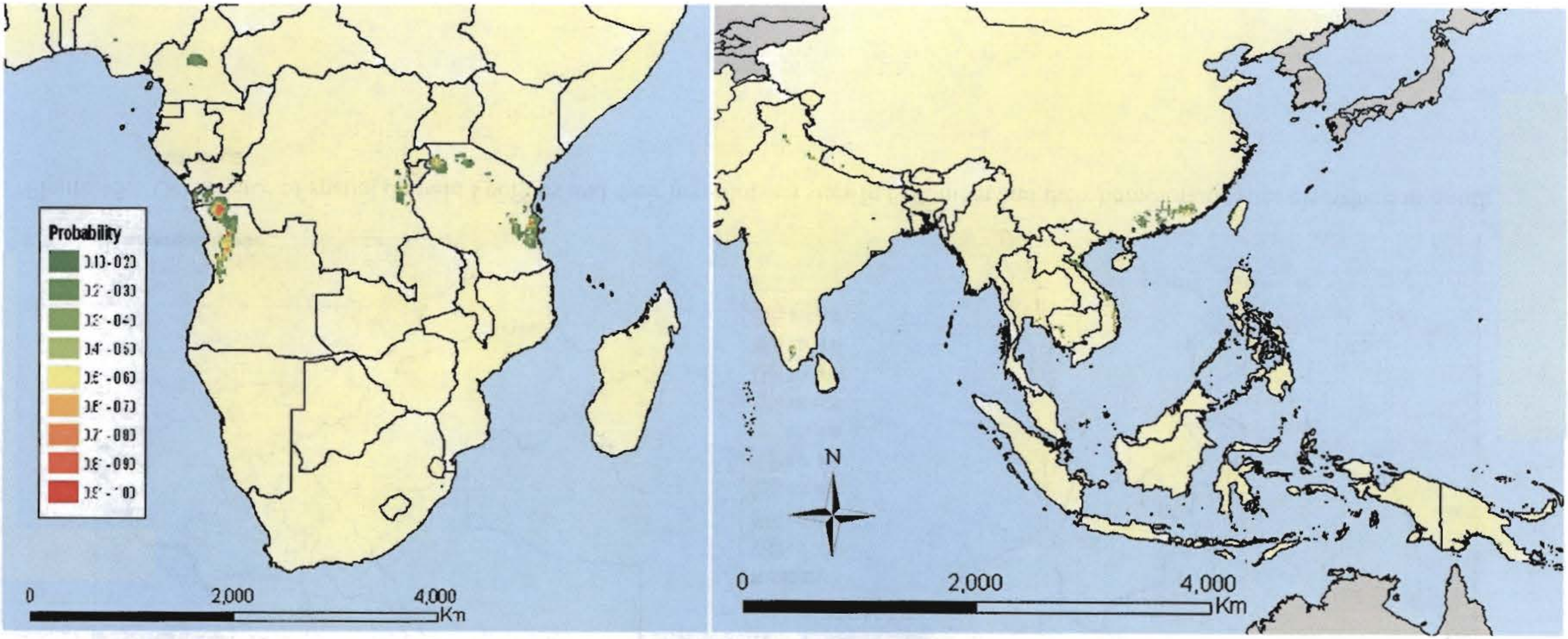


Figure 46. Homologous sites for spatial domain I coffees in Africa and Southeast Asia.

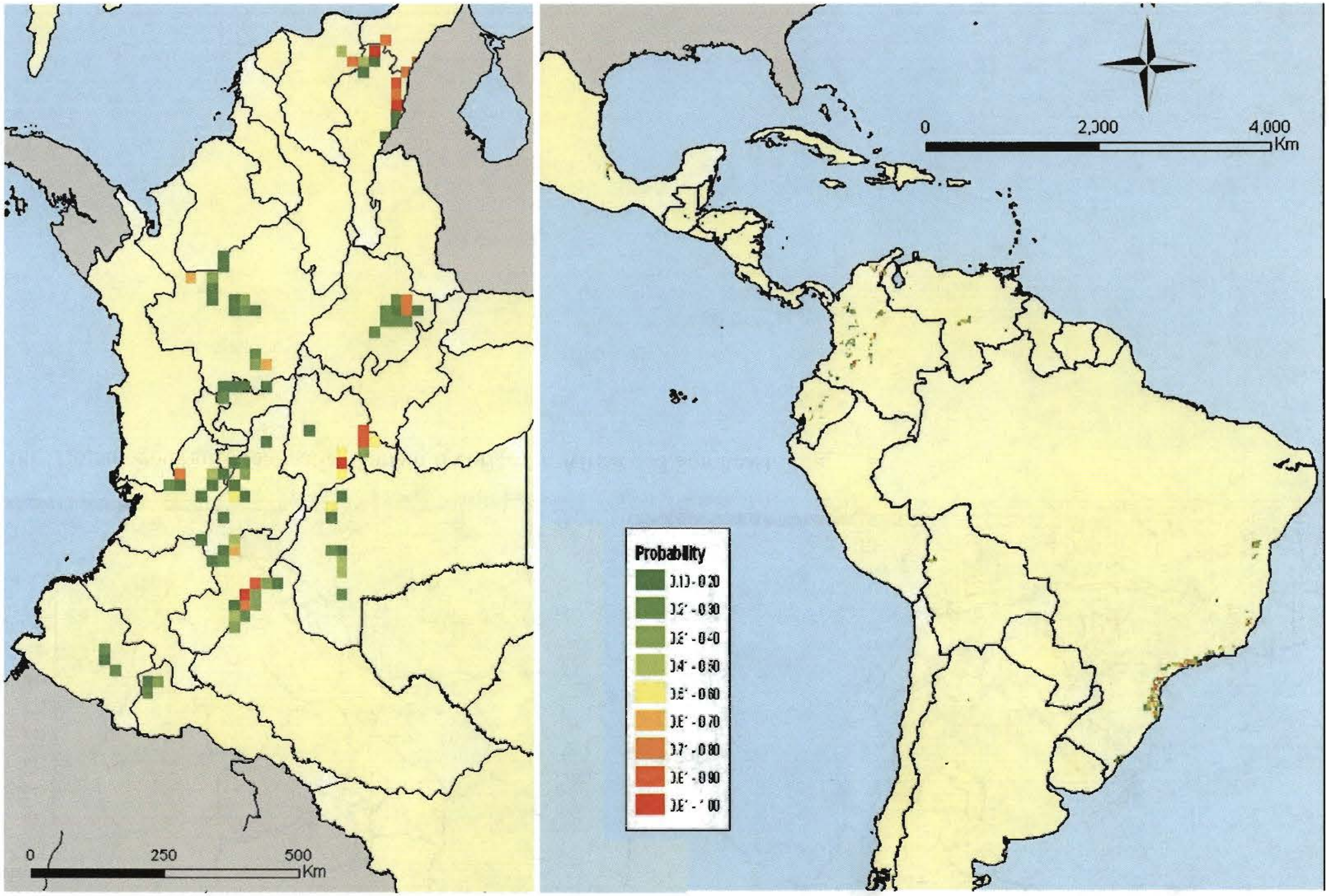


Figure 47. Occurrence of spatial domain II coffees and their homologous sites in Colombia and their homologous sites elsewhere in South America.

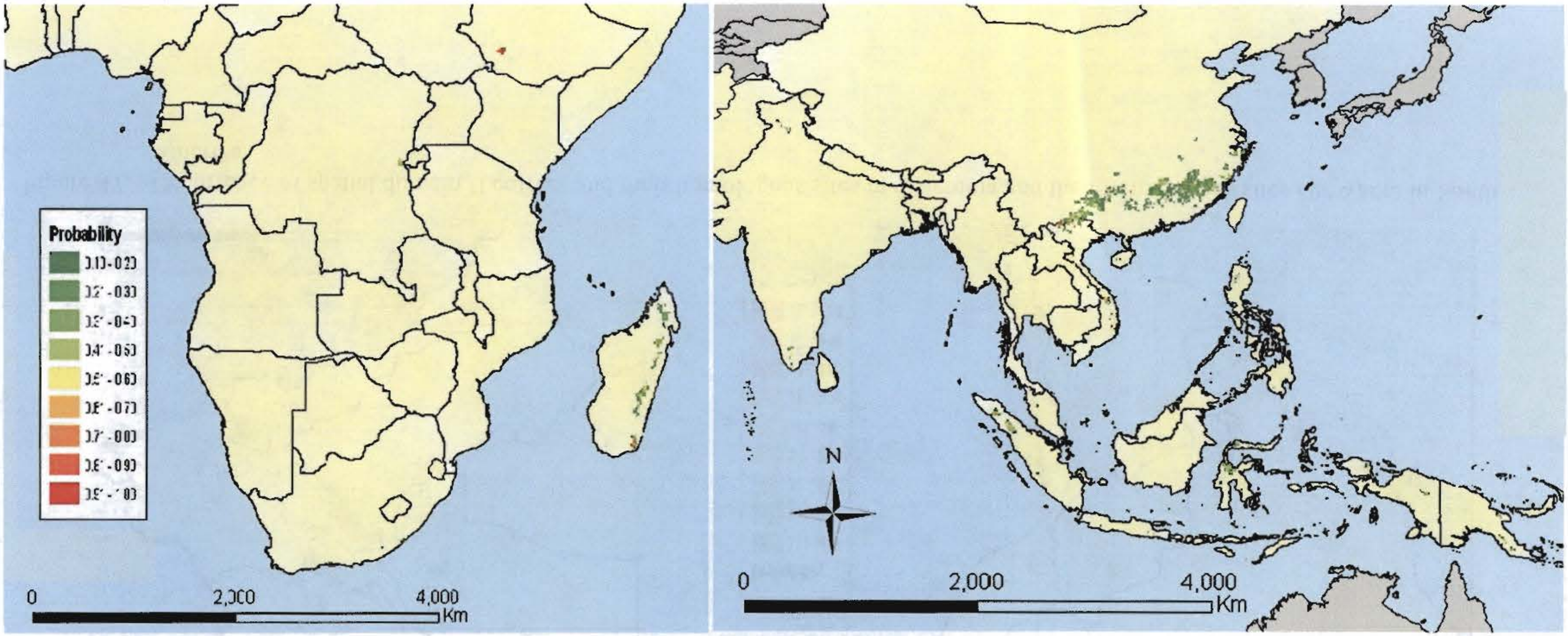


Figure 48. Homologous sites for spatial domain II coffees in Africa and Southeast Asia.

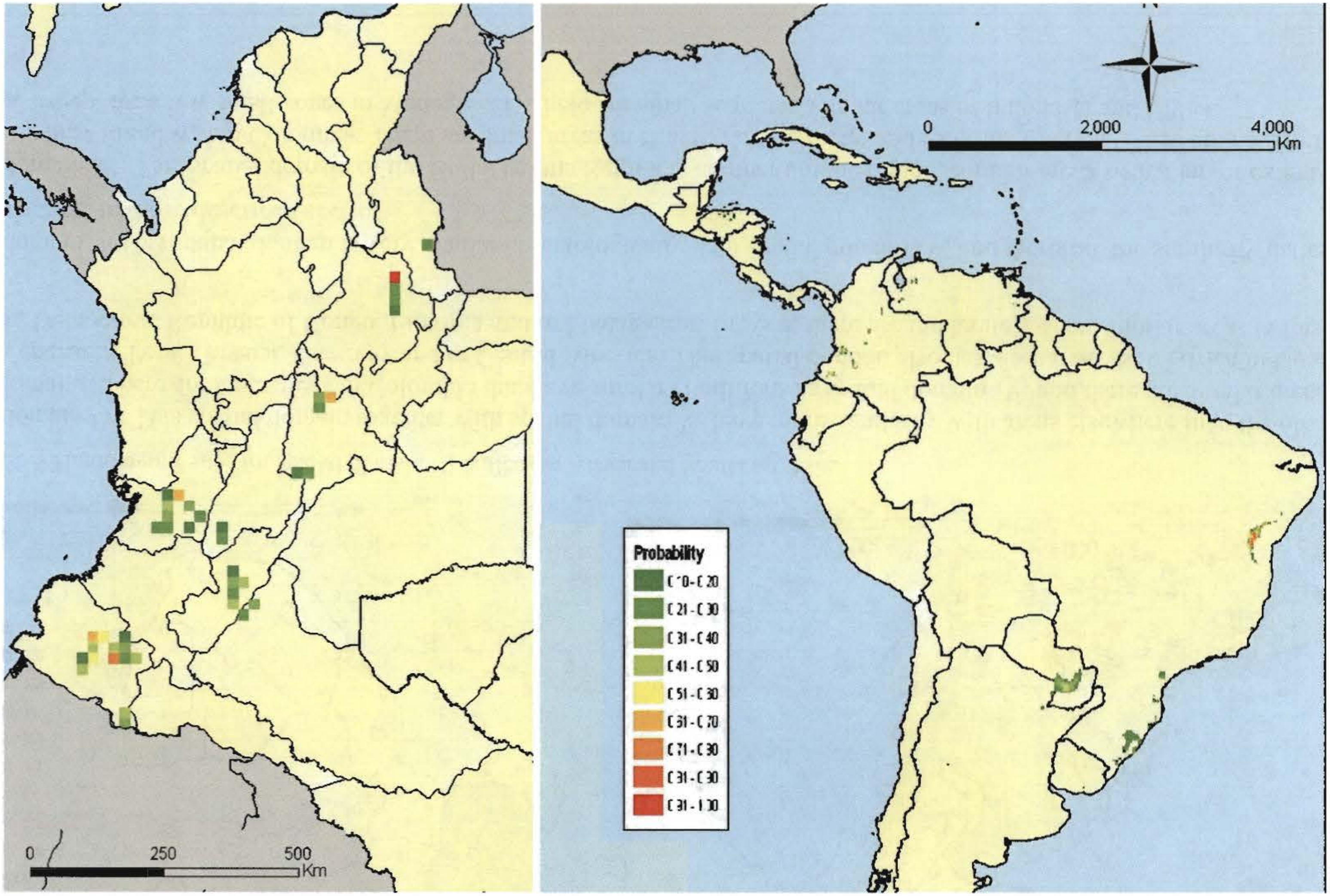


Figure 49. Occurrence of spatial domain III coffees and their homologous sites in Colombia and their homologous sites elsewhere in South America.

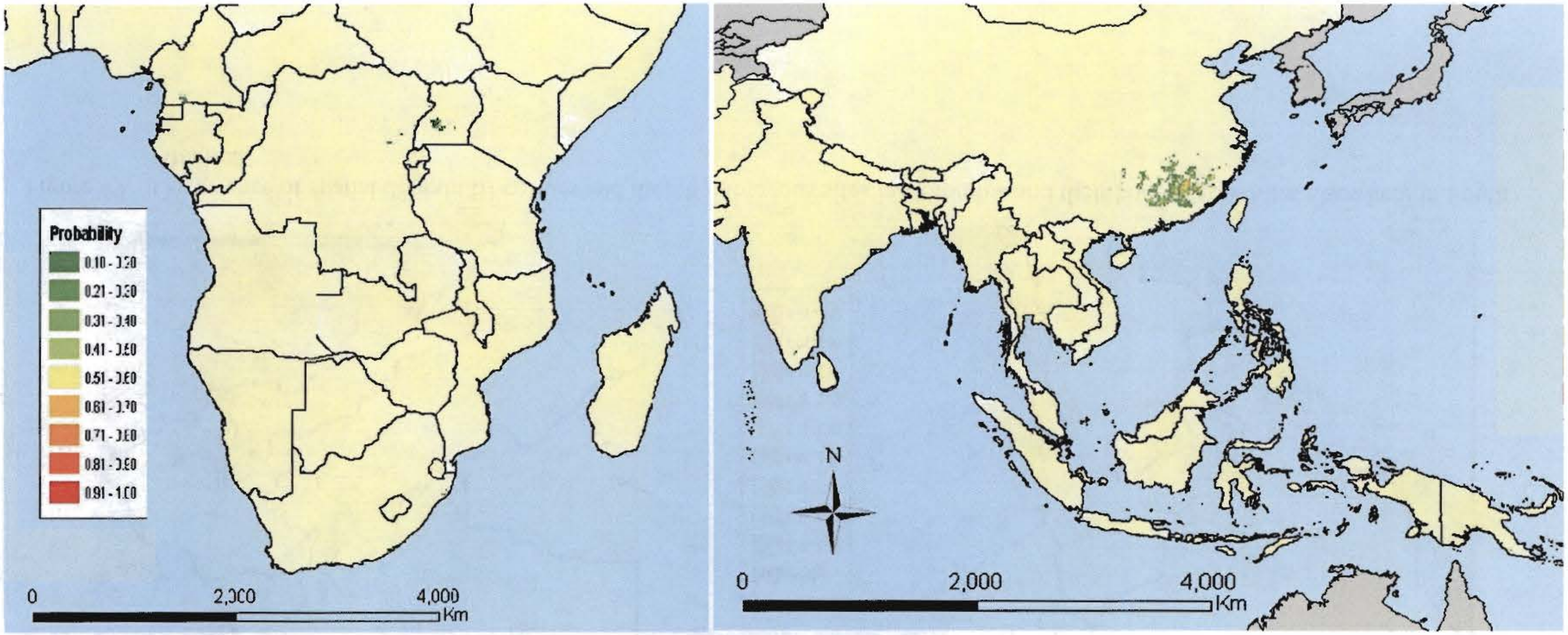


Figure 50. Homologous sites for spatial domain III coffees in Africa and Southeast Asia.

Spatial domain IV: This spatial domain together with spatial domain V, have more similarity with areas elsewhere than the other spatial domains. There are large areas in Colombia that have similar conditions as spatial domain IV, and there are similar areas in Brazil, Venezuela, Peru, Guyana, Paraguay and in Central America. This spatial domain also has similar areas in Africa in Uganda, Tanzania, Democratic Republic of Congo, Ethiopia and in Madagascar. In Asia, there are moderately-sized similar areas in East Timor.

Spatial domain V This spatial domain is very similar climatologically with spatial domain IV, and therefore the similarity patterns are almost identical to those described above.

Spatial domain VI: This spatial domain of the Huila/Tolima Region is another unique region. Similar areas with a larger extent are principally only found within Colombia. There are small areas in Brazil, Peru, Urugay and Central America. There are no similar areas in Africa, except for a few small zones in Madagascar. There are small scattered similar areas in Indonesia and China.

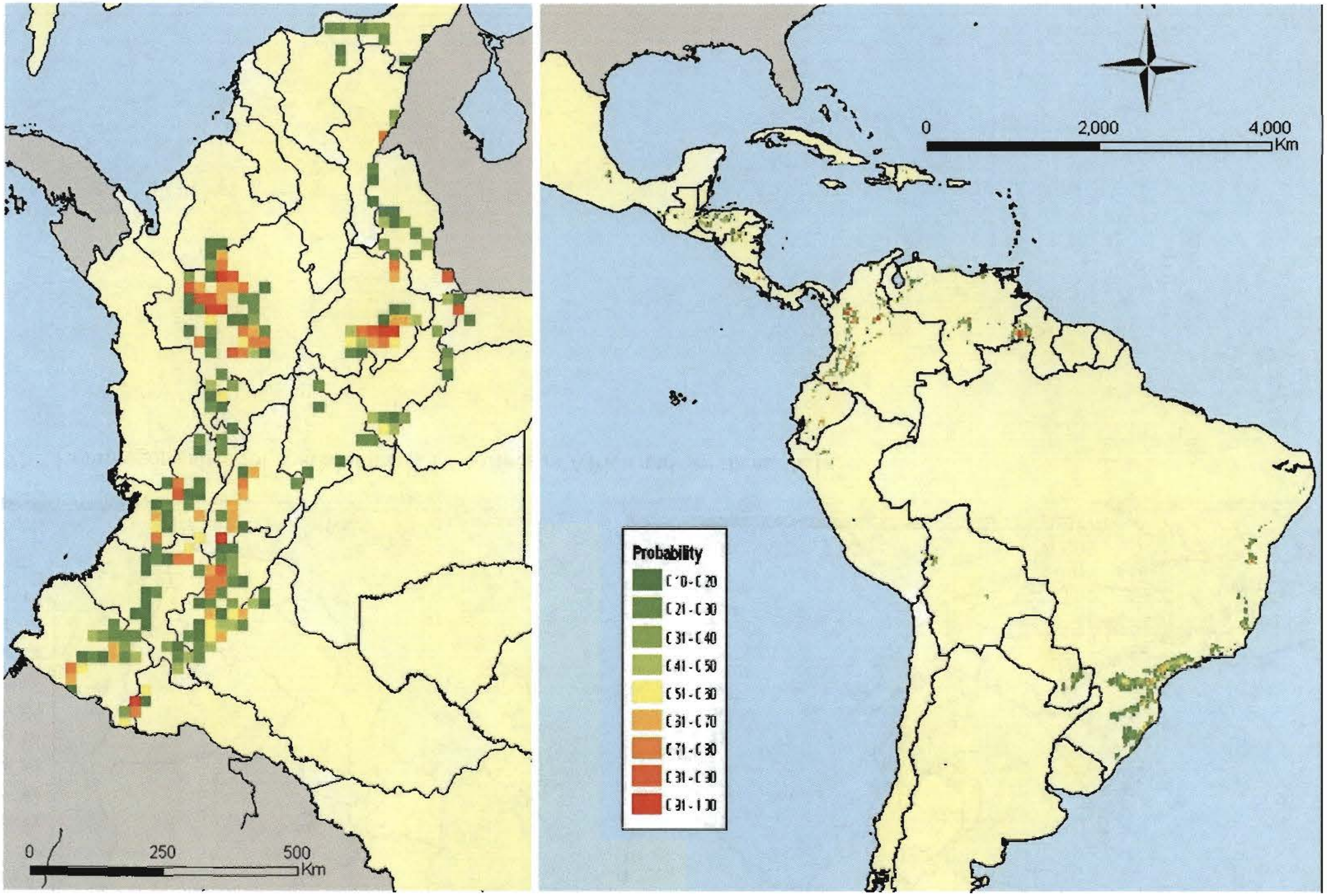


Figure 51. Occurrence of spatial domain IV coffees and their homologous sites in Colombia and their homologous sites elsewhere in South America.

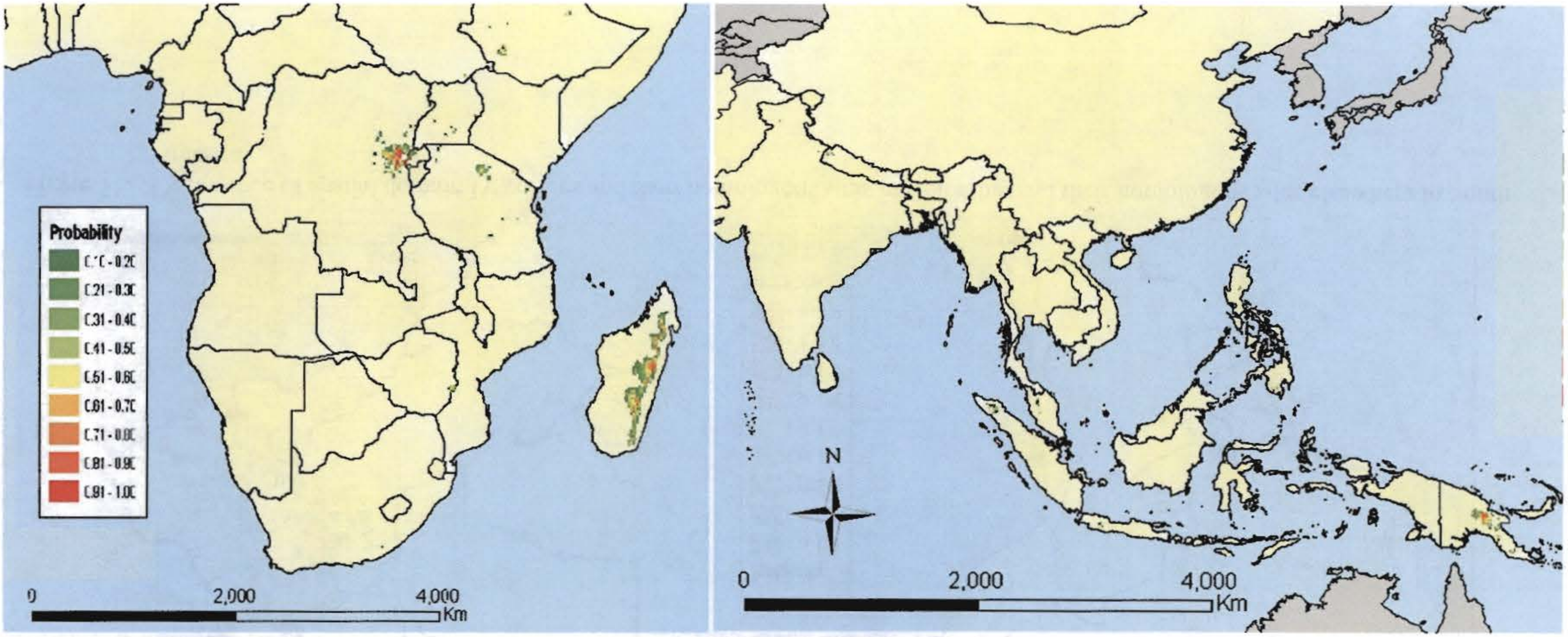


Figure 52. Homologous sites for spatial domain IV coffees in Africa and Southeast Asia.

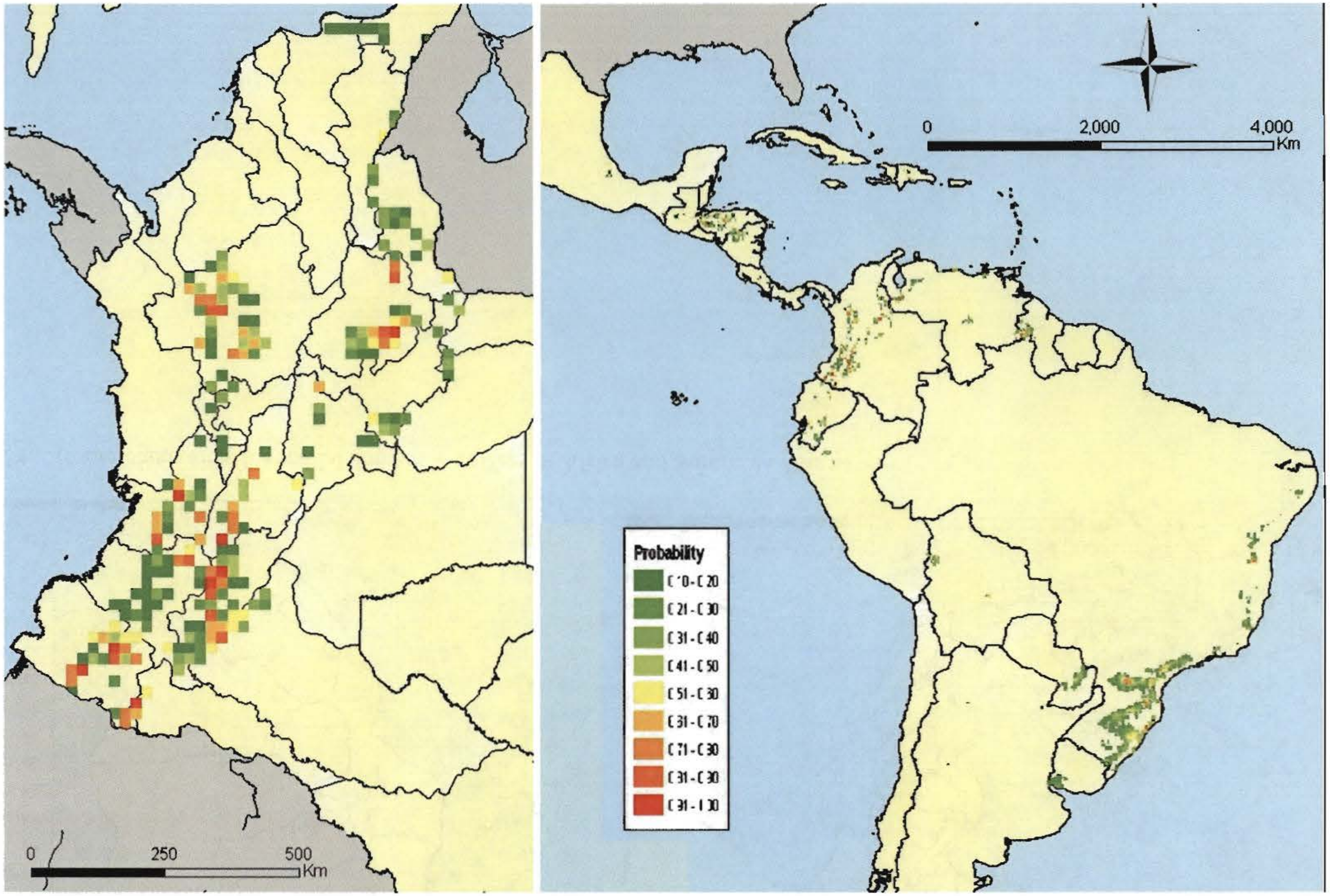


Figure 53. Occurrence of spatial domain V coffees and their homologous sites in Colombia and their homologous sites elsewhere in South America.

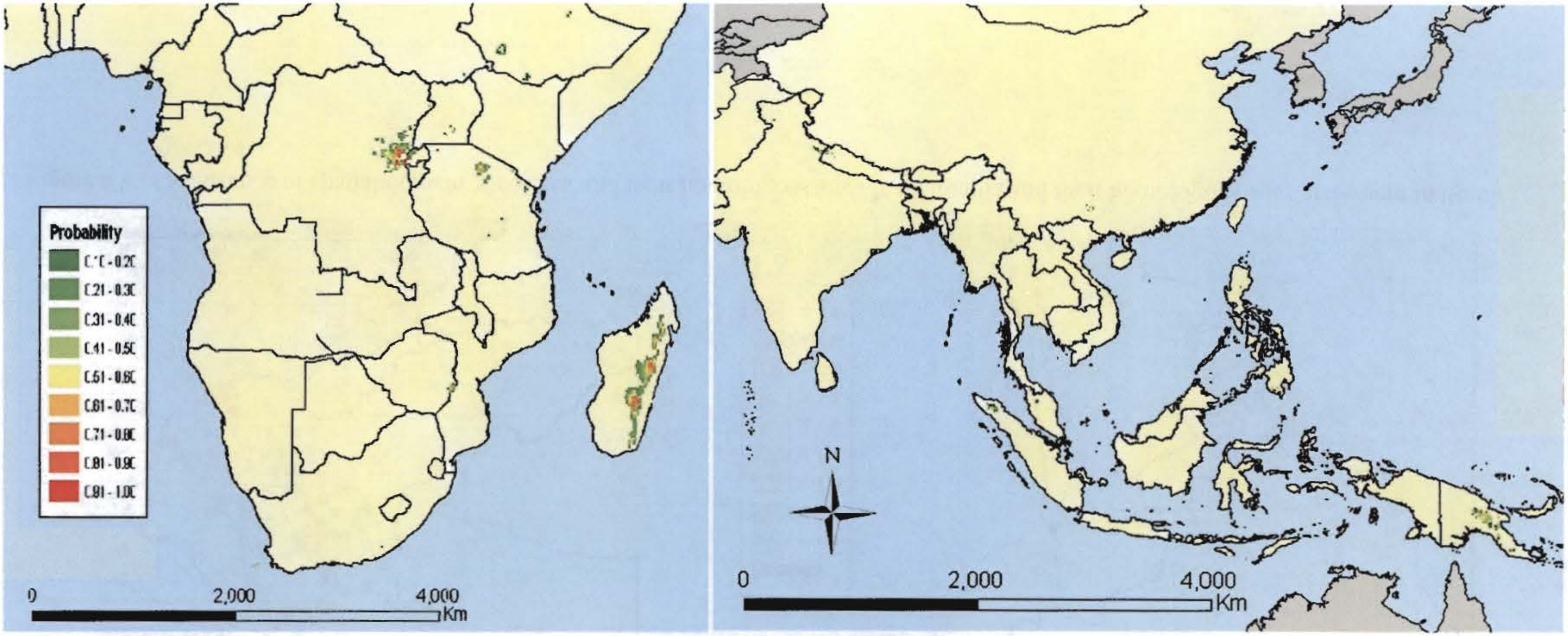


Figure 54. Homologous sites for spatial domain V coffees in Africa and Southeast Asia.

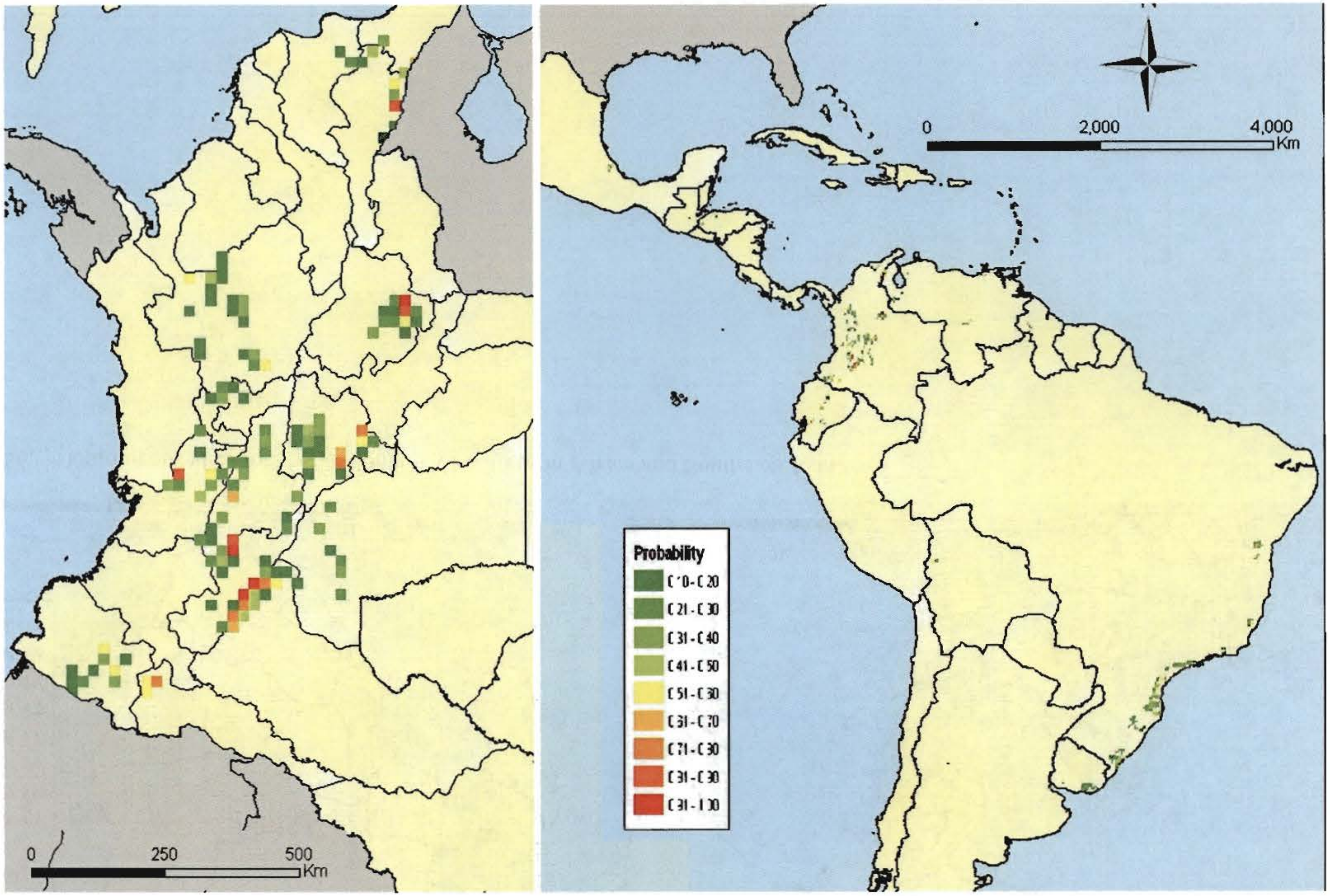


Figure 55. Occurrence of spatial domain VI coffees and their homologous sites in Colombia and their homologous sites elsewhere in South America.

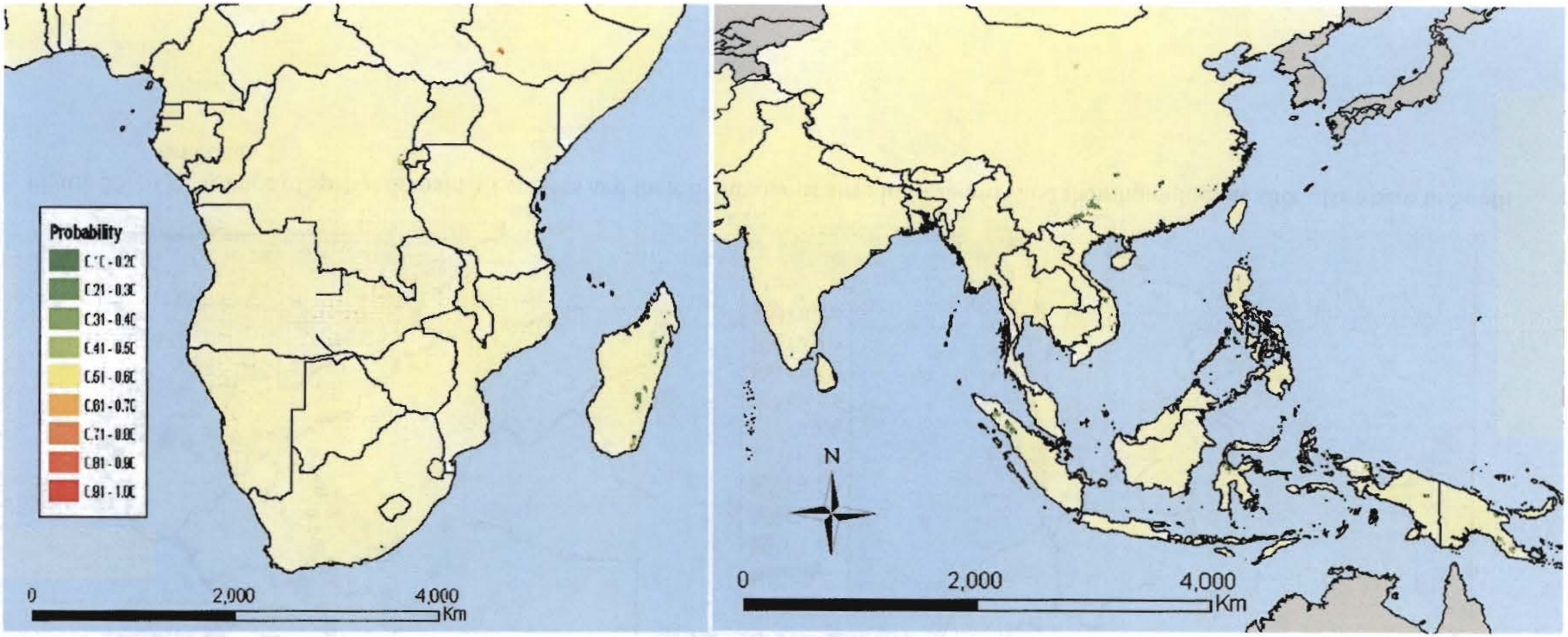


Figure 56. Homologous sites for spatial domain VI coffees in Africa and Southeast Asia.

Table 45. Areas elsewhere in Colombia, in South America, inAfrica, and in Asia estimated by Homologue to have homologous climates to the Domains in Colombia identified in the study.

Domain of analysis	Probability	Area (km^2)			
		Colombia	South America	Africa	Asia
I	0.1 – 0.5	9,400	56,600	161,200	93,100
	0.5 – 1.0	2,750	3,300	32,100	6,600
II	0.1 – 0.5	24,720	86,000	31,200	279,900
	0.5 – 1.0	8,640	39,500	9,400	17,700
III	0.1 – 0.5	11,950	114,700	15,100	133,400
	0.5 – 1.0	2,400	13,700	0	14,000
IV	0.1 – 0.5	53,200	353,900	353,900	40,000
	0.5 – 1.0	26,400	76,200	76,200	11,000
V	0.1 – 0.5	52,190	449,900	449,900	42,600
	0.5 – 1.0	25,260	83,900	83,900	6,900
VI	0.1 – 0.5	35,960	125,200	125,200	66,500
	0.5 – 1.0	9,990	18,300	18,300	3,100

7. Conclusions and recommendations

7a. Thematic issues: Can the differences in growing environment and product quality justify denominations of origin in the Colombian Huila, Tolima, César, Magdalena, Santander and Santander de Norte departments?

- The regional differences in environmental and product-quality characteristics between the coffee-growing areas in the two Colombian departments of Huila (northern part) and Tolima (southern part), and are often statistically significant. Considering the magnitude of the differences in sensorial characteristics, they are also likely to be important in commercial terms. However, in contrast to the first Phase of this study where we found in no significant differences between the varieties in the departments of Cauca and Nariño, here we observed large genotype by environment interactions within the boundaries of administrative departments. In the context of a denomination of origin (DO), this complicates substantially the establishment of a unique common coffee quality profile for the departments under consideration. On occasions we found that the three main varieties, (Caturra, Colombia and Tipica) all had different quality profiles within one department. We therefore suggest not to use administrative boundaries as the spatial units for a denomination of origin but to implement a new spatial delimitation designed to reduce the genotype by environment interactions.
- This situation is not unexpected. In the first Phase study, the Bolívar district of southern Cauca was environmentally and in terms of product quality more similar to those of Nariño than of Cauca. We therefore recommended that the Bolívar district be included in the Nariño DO. . Also, although the Inzá district of Cauca should be part of the Cauca DO we argued that it would merit an immediate recognition as sub-denomination of Cauca. Similarly, the growing regions along the Pacific slopes of the Western Cordillera in Cauca had environmental and product qualities distinct from the other Cauca regions and merited recognition as a sub-denomination.

- We propose here to design new spatial delimitations based on integrating both formal and informal knowledge about coffee quality. Formal knowledge is provided by the data collected and analyzed in this study, complemented by informal knowledge contributed by three FNC experts oncoffee quality. Analyzing both sources of information in a spatially explicit approach permitted us to generate six new spatial units, which we call *spatial domains*. Two of these spatial domains are in the Huila/Tolima Region and four are in the Northern Region.
- These six new spatial domains integrate the 13 domains delimited by the quality experts with the information provided by climate and soils data. The regional differences in environmental and product-quality characteristics between the coffee-growing areas in these spatial domains are generally statistically significant. Considering the magnitude of the differences in sensorial characteristics, they are also likely to be important in commercial terms. These differences of the environmental and product-quality characteristics are not random. Moreover, the spatial structures in the data of both the environmental and the product-quality characteristics are correlated.
- We therefore believe that this study provides ample evidence to justify a denomination of origin (DO) based on these six spatial domains: There are two spatial domains in the Huila/Tolima region, which conincide with the expert knowledge domains. The new spatial domain V is the same as the expert domain of Planadas and the Franja Frontera Huila. The new spatial domain VI constitutes the wider Huila/Tolima expert domain. There are four domains in the Northern region: Spatial domain I is equal to the expert domain of Magdalena (Costa Caribe); spatial domain II brings together the Oriente (Cesar) and Perejia (Serrania perija); spatial domain III joins the Catatumbo zone, the Zona Baja (cerca de Cucuta), and the zone close to Bucaramanga; while spatial domain IV is made up by the expert domains of Toledos Labateca, San Gil region, and the region of Barbosa/Boyaca. These six domains would form the main denomination units. There is however opportunity to further subdivide these units, if commercial opportunities would require. Considering the generated formal and informal knowledge, a subdivisions could include Oriente (Cesar) and Perejia (Serrania perija), Toledos Labateca, and the zone of Tipica San Andres. We had no samples for coffee quality in this analysis for the latter zone.
- Spatial domain I: Coffees are characterized by low acidity and high body. Their fragrance and aroma exhibits a nutty character, as does their flavor and therefore the overall impression. The expert opinions are nicely confirmed by the sensoric cupping data.
- Spatial domain II: Coffees from this domain tend to be balanced with medium body and acidity, and they have a moderate level of sweetness. Their fragrance and aroma is characterized by nutty and chocolaty notes, which in the flavor are complemented by caramel tones. This leads to coffees that overall can be considered as having sweet, nut--caramel notes, with little astringency or off flavors.
- Spatial domain III: These coffees are generally characterized by medium acidity and medium body. Their sweetness level tends to be low. The fragrance and aroma sometimes has astringency notes, while the flavor can be chocolaty or nutty in some subregions. Coffees have the tendency to exhibit astringency and herbal nuances in the flavor as well. The region has higher potential (see Canasta) than is currently being realized by growers.

- Spatial domain IV: Coffees of this domain have medium to high levels of body and sweetness. Their acidity tends to be moderate, in some cases high, while the fragrance and aroma is characterized by sweet, fruity notes. These are generally reflected in the flavor, and complemented by sweet caramel nuances. This leads to an overall profile that is fruity and sweet, in some cases may be citric acidity.
- Spatial domain V: These coffees are considered to have high levels of sweetness and acidity combined with medium body. They demonstrate fruity and floral fragrance and aroma, very often combined with sweet caramel notes, and a citric fragrance. The flavor fully reflects the fragrance and aroma expressions. It leads to coffees with high overall quality profiles that are characterized by sweetness, and fruity and citric acidity, with clear caramel notes. Surprisingly, the cupping scores do not reflect this assessment (see below). The region has higher potential (see Canasta) than is currently being realized by growers.
- Spatial domain VI. These are balanced coffees with medium levels of acidity, body and sweetness. They have sweet notes in fragrance and aroma, often accompanied by herbal tones. The flavor reflects these, but also shows sweet caramel aspects. Overall, these are sweet, fruity coffees that may have a herbal off taste.
- The spatial domain V is generally considered as a region that produces outstanding coffees with an international reputation amongst specialty-coffee experts. The results of the analyses of the formal quality information provided by the sensory assessments do not support this. There can be various reasons for this anomoly, including deviations during the growing season from the average climate. This is not uncommon in many wine denominations of origin. In a more detailed analysis, we found that domain V has an equal proportion of high quality coffee samples as domain VI, however, the proportion of lower quality coffees is much higher, even though the potential for good quality coffees is immense in this spatial domain. Figure 57 visualizes this situation for the example of Final Score. The background is the likely suitability for coffees with a high Final Score, the white points represent low quality coffees, and the dark red points represent high quality coffees. It becomes obvious that the domain 5 has high potential for quality coffees that is currently not being realized.

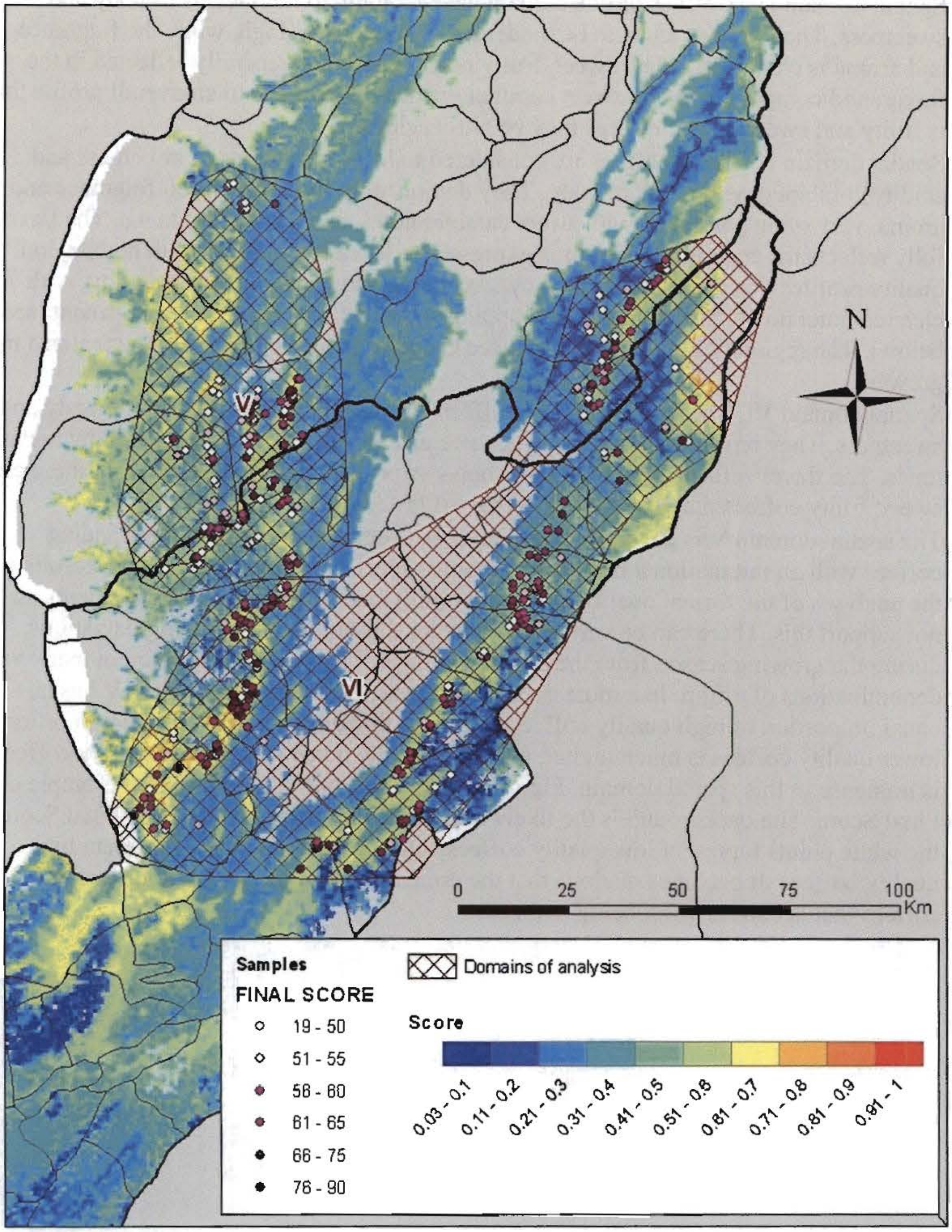


Figure 57. CaNaSTA determination of Bayesian suitability score of the sensory characteristic of Final score overlain by the two identified spatial domains in the Hula/Tolima Region. The colour of the dots indicate the level of the Final score obtained by each sample.

- The cup profiles in both the Huila/Tolima Region and the Northern Region are influenced strongly by climatic conditions. Each spatial domain tends to have its unique climate profile. In the assessed example of Final Score in the Northern Region, for example, we find that certain ranges of solar radiation and cloud frequency have a major positive impact. Acidity on the other hand is driven by altitude, cloud frequency and dry months per year. Final Score in the Huila/Tolima Region is positively influenced by certain altitude ranges and annual rainfall, while acidity benefits from specific solar radiation values, diurnal temperature range and dry months per year. Additionally, in the Northern Region we observe a strong influence of soil conditions on coffee quality. This provides an opportunity for management to actively impact on sensory coffee quality, specifically in domain II by means of management of the soil base complex, applying nitrogen, and targeting of areas with heavier soil texture. In domain I soil management can be used to improve the physical quality of coffee by addressing the phosphorus iron complex, and by targeting areas with higher than average soil acidity and lighter than average soil texture.

- Colombian coffee growers are indeed in a unique position to implement a denomination of origin for their Huila/Tolima Region and Northern Region growing areas in sense that (a) climate (in interaction with management) drives to a large degree the expression certain quality characteristics, and (b) the climate profiles for all the spatial domains are internationally rather rare. This implies firstly that there is a cause – effect relationship between spatial domain and coffee quality, and secondly that there are globally only few areas where one can expect to find similar cause – effect relationships.

7b. Methodological issues: What other steps should be included in the process of implementing denominations of origin?

- The current extent of the six spatial domains includes most of the commercially important coffee growing areas in Colombia. In fact, it is likely that they include more areas than are strictly useful for the denomination of origin. The delimitation has been conducted on an *ad hoc* basis, and is only an approximation of the final units for the denomination of origin. We therefore strongly suggest that the outlines of the units should be refined using available sources information such as the FNCSICA data base, the FNC map of ecotopos and the altitude maps to which CIAT has access. Also, we recommend that the process include the commercial and quality departments of the FNC, as well as the regional FNC offices.
- Due to time limitations, we only partially implemented some useful analyses in order to generate the minimum knowledge base that permits the implementation of denominations of origin. Some of these analyses can be expanded to strengthen further an application for the denominations of origin. These analyses most notably include a complete spatial probability analysis of quality driving factors, which currently has only been conducted for selected examples of sensory characteristics. We furthermore strongly recommend including an analysis of the complete set of agronomic information that was collected by the CENICAFE team.
- There is an obvious impact of short-term climatic changes on the characteristics of product quality, for which the wine industry is a classic example. It was impossible to quantify this impact in this short-term study. It is important to note that temporary variability in quality due to year-to-year differences usually does not prevent the implementation of denominations of origin. For this reason, it is of outmost importance to resample at least a carefully-selected sub-set of farms on a regular basis to quantify the

magnitude of temporal variability and to define the climatic representativeness of a particular growing season. (This has already been recommended for Phase 1.)

- All the analyses presented here are based on samples for product quality that had standardized harvest and post-harvested processing. It is therefore likely that samples collected directly from growers will exhibit somewhat different cup profiles from those presented here. This is normal and expected. However, it is imperative to conduct a comparative analysis of growers' and standardized samples to understand whether and how much some of the production processes need to be adjusted to comply with the requirements that will be stipulated in the DO regulatory documents. (This has already been recommended for Phase 1.)
- Considering the likely impact of climate change on coffee quality, we strongly recommend a full analysis of the expected impacts that global climate change will have on the ability of Colombian growing regions to maintain over the coming decades quality profiles that are defined now.

8. References

Avelino,J., Barboza,B., Araya,J.C., Fonseca,C., Davrieux,F., Guyot,B. and Cilas,C. (2005). Effects of slope exposure, altitude and yield on coffee quality in two altitude terroirs of Costa Rica, Orosi and Santa Maria de Dota. *Journal of the Science of Food and Agriculture* **85:**1869-1876.

Avelino, J., Román, D., Romero, S. and Fonseca, C. (2006). *Indicaciones Geograficas y Denominaciones de Origen: Algunos Fundamentos y Metodologías con Ejemplos de Costa Rica sobre Café.* Instituto del Café de Costa Rica: San José, Costa Rica. 36 pp.

Barham, E. (2003). Translating terroir: the global challenge of French AOC labeling. *Journal of Rural Studies* **19:**127–138:

Batjes, N.H. (1995). *World Inventory of Soils Emission Potentials: WISE2.1 – Daatabase User's Manual and Coding Protocols.* Technical Paper 26, International Soils Reference and Information Center (ISRIC), Wageningen, the Netherlands.

Batjes, N.H. and Bridges, E.M. (1994). Potential emissions of radiator fleet for active trace gases from soil to atmosphere was special reference to methane: Development of a global database (WISE). *Journal of Geophysical Research* **99(D8):**16,479-16,489.

Barjolle, D. and Lehmann, B. in collaboration with Chappuis, J.-M. and Dufour, M. (1997). Protected designation of origin and institutions (France, Spain and Italy). Typical and traditional productions: Rural effect and agro-industrial problems 52nd EAAE Seminar, Parma, June 19-21 1997.

Bérard, L., Beucherie, O., Fauvet, M., Marchenay, P. and Monticelli, C. (2001). Les acteurs historiques, culturels, économiques et environnementaux dans la dé limitation des zones IGP. *Ingénieries* **28:**49-63.

Bertozzi L. (1995). Designation of origin: quality and specification. *Food Quality and Preference* **6:**143-147.

Brus, D.J. and De Gruijter, J.J. (1997). Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). *Geoderma*, **80:**1-44.

Cormorèche, P. (1994) in "La qualité alimentaire", Chambres d'Agriculture, supplement to number 821, April 1994.

De Nadai Fernandes, E.A., Tagliaferro, F.S., Azevedo-Filho, A. and Bode, P. (2002). Organic coffee discrimination with INAA and data mining/KKD techniques: new perspectives for coffee trade. *Accreditation and Quality Assurance* **7:**378-387.

Decazy, F., Avelino, J., Guyot, B., Perriot, J.J., Pineda, C. and Cilas, C. (2003). Quality of different Honduran coffees in relation to several environments. *Journal of Food Science* **68**:2356-2361.

Deutsch, C.V. and Journel, A.G. (1992). *GSLIB. Geostatistical Software Library and Users' Guide.* Oxford University Press: New York.

Dobermann, A. and Oberthür, T. (1997). Fuzzy mapping of soil fertility - A case study on irrigated rice land in the Philippines. *Geoderma*, **77**:317-339.

Downey G, Boussion J (1996) Authentification of coffee bean variety by near-infrared reflectance spectroscopy of dried extract. *Journal of Science in Agriculture* **71**:41-49

Elliott, P.and Warwick, D. (2004). Spatial epidemiology: Current approaches and future challenges. *Environmental Health Perspective* **112**:998-106.

Falcetti M. 1994. The « terroir ». What is the « Terroir »? Why study it? Why teach it? *Bulletin de l' OIV (International Vine and Wine Office)*, **67**:246-275.

FAO (Food and Agriculture Organization) (1974). *FAO-UNESCO soil map of the world 1:5,000,000.* Volume 1, Legend.UNESCO and FAO: Paris, France and Rome, Italy.

FAO 1978

FAO (Food and Agriculture Organization) (1995). *Digital soil map of the world and derived soil properties (version 3.5).* Land and Water and Digital Media Series 1. FAO: Rome, Italy.

Gilg, A. (1996). *Coutryside Planning: The First Half Century.* 2nd Edition. Routledge: New York.

Hair, J.F., Anderson, R.E., Tatham, R.L. and Black, W.C. (1992). *Multivariate Data Analysis with Readings.* Macmillan Publishing Company: New York.

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

Hutchinson, M.F. (1997). *ANUSPLIN Version 3.2 Users Guide.* The Australian National University, Centre for Resource and Environmental Studies: Canberra. 39pp.

Isaaks, E.H. and Srivastava, R.M. (1989). *Applied Geostatistics.* Oxford University Press, New York.

Jaccard, J.and Wan, C.K. (1996). *LISREL: Approaches to Interaction Effects in Multiple Regressions.* Sage Publications: Thousand Oaks, California.

Jones, P.G. (1991). The CIAT climate database version 3.41. Machine readable data set. Centro Internacional de Agricultura Tropical (CIAT): Cali, Colombia.

Jones, P.G. (1996). Climate database, Haiti. Machine readable data set. Centro Internacional, de Agricultura Tropical (CIAT): Cali, Colombia.

Jones, P.G., Galwey, N., Beebe, S.E. and Tohme, J. (1997). The use of geographical information systems in biodiversity exploration and conservation. *Biodiversity and Conservation* **6**:947-958.

Jones, P.G. and Gladkov, A. (1999). FloraMap: *A Computer Tool for Predicting the Distribution of Plants and Other Organisms in the Wild: Version 1, 1999.* CD-ROM with 87-page indexed manual. CIAT CD-ROM Series. Centro Internacional de Agricultura Tropical (CIAT): Cali, Colombia.

Jones, P.G.; Thornton, P.K. (2000). MarkSim: Software to generate daily weather data for Latin America and Africa. *Agronomy Journal* **92**:445-453.

Jones, P.G.; Thornton, P.K.; Díaz, W.; Wilkens, P.W.; Jones, A.L. (2002). *MarkSim: A Computer Tool that Generates Simulated Weather Data for Crop Modeling and Risk Assessment: Version 1.* Centro Internacional de Agricultura Tropical: Cali, Colombia. 87pp.

Kshirsagar, A.M. (1972). *Multivariate Analysis.* Marcel Dekker: New York.

Läderach, P., Oberthür, T., Niederhauser, N., Usma, H., Collet, L. and Pohlan, H.A.J. (2006). Café especial: Factores, dimensiones e interacciones. In: *El Cafetal del Futuro: Realidades y Visiones* (Pohlan, H.A.J., ed.). Shaker Verlag: Germany. Pp. 141-160.

Linacre, E. (1977). A simple formula for estimating evaporation rates in various climates, using temperature data alone. *Agricultural Meteorology* **18**:409-424.

Marsden, T., Banks, J. and Bristow, G. (2000). Food supply chain approaches: exploring their role in rural development. *Sociologia Ruralis* **40**:424-438.

Mulligan, M. and Burke, S.M. (2005). Global cloud forests and environmental change in a hydrological context. DFID FRP Project ZF0216. Final Report. URL: http://www.ambiotek.com/cloudforests/

Murdoch, J., Marsden, T., banks, J. 2000. Quality, nature and embeddedness: Some theoretical considerations in the context of the food sector. *Economic Geography* **76**:107-125

NOAA (1984). National Oceanographic and Atmospheric Administration TGP-006 D. Computer compatible tape. Boulder, Colorado.

Oberthür, T., Goovaerts, P. and Dobermann, A. (1999). Mapping soil texture classes using field texturing, particle size distribution and local knowledge by both conventional and geostatistical methods. *European Journal of Soil Science* **50**:457-479.

Oberthür, T., Dobermann, A. and White, P. (2000). The rice soils of Cambodia. II. Statistical discrimination of soil properties by the Cambodian agronomic soil classification system (CASC). *Soil Use and Management* **16**:20-26.

O'Brien, R. (2004). *Spatial Decision Support for Selecting Tropical Crops and Forages in Uncertain Environments.* PhD thesis, Department of Spatial Sciences, Curtin University of Technology, Perth. 278pp.

Pearl, J. (1990). Bayesian decision methods. In: *Readings in Uncertainty Reasoning* (Shafer, G. and Pearl, J., eds.). Morgan Kaufmann: San Mateo, California. Pp. 345-352.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* **6**:559 -572

Porter, M. (1986). *L'avantage Concurrentiel.* InterEditions: Paris.

Rao, C.R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhya* A **26**:329 -358.

Salette, J., Asselin, C. and Morlat, R. (1998). The relationships between « terroir » and product: an analysis of the « terroir-vine-wine » system and its analogous application to other products. *Sciences des Aliments* **18**:251-265.

Starr, J.L., Parkin, T.B. and Meisinger, J.J. (1992). Sample size consideration in the determination of soil nitrate. *Soil Science Society of America Journal* **56**:1824-1830.

Vaast, P., Bertrand, B., Perriot, J.-J., Guyot, B. and Genard, M. (2006). Fruit thinning and shade improve bean characteristics and beverage quality of coffee (*Coffea arabica* L.) under optimal conditions. *Journal of the Science of Food and Agriculture* **86**:197-204.

Van der Ploeg, J.D. and Renting, H. (2000). Impact and potential: A comparative review of European rural development practices. *Sociologia Ruralis* **40**:329-343.

Van der Ploeg, J.D., Renting, H., Brunori, G., Knickel, K., Mannion, J., Marsden, T., de Roest, K., Sevilla-Guzman, E., Ventura, F. 2000. Rural development: from practices and policies towards theory. *Sociologia Ruralis* **40**:391-408.

Webster, R. and Oliver, M.A. (1990). *Statistical Methods in Soil and Land Resource Survey.* Oxford University Press: Oxford.

9. Appendices

List of Appendices

Appendix 1: Tables of descriptive statistics of size characteristics.

Table A1.1 Size characteristics of coffee sampled in Huila department (in grams per size class), and yield factors for size classes <=13, 14 and => 15.

Screen Size	Variety	N	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Skewness
12	Caturra	123	1.2	0.0	0.2	0.4	1.1	15.0	3.9
	Colombia	81	0.6	0.0	0.2	0.4	0.8	4.7	3.0
	Tipica	20	3.2	0.0	0.2	0.5	1.0	44.6	4.3
13	Caturra	123	3.6	0.0	0.9	1.6	4.7	19.7	2.1
	Colombia	81	2.3	0.1	0.8	1.3	3.0	15.7	2.8
	Tipica	20	6.7	0.0	1.0	2.4	4.3	68.5	4.0
14	Caturra	123	10.5	0.9	4.1	6.8	16.1	46.2	1.5
	Colombia	81	7.5	0.6	3.4	5.8	10.4	27.9	1.3
	Tipica	20	13.5	0.2	3.0	7.6	20.1	43.0	1.1
15	Caturra	123	28.6	4.6	13.8	23.5	40.2	87.2	0.9
	Colombia	81	21.3	3.8	11.4	17.1	31.1	61.6	0.9
	Tipica	20	31.9	2.7	14.2	27.9	50.3	78.0	0.5
16	Caturra	123	58.7	16.3	49.1	58.2	71.4	94.3	-0.4
	Colombia	81	52.9	13.9	39.0	55.1	66.6	94.0	0.0
	Tipica	20	60.9	4.4	47.5	67.0	81.5	94.3	-0.7
17	Caturra	123	59.6	13.4	45.9	66.4	76.9	98.1	-0.5
	Colombia	81	60.9	17.5	54.2	61.9	71.4	98.5	-0.5
	Tipica	20	52.4	1.6	28.9	52.7	77.4	96.7	0.0
18	Caturra	123	37.5	2.1	15.4	33.2	51.9	164.1	1.5
	Colombia	81	57.0	4.8	26.9	50.9	86.3	141.6	0.5
	Tipica	20	29.4	0.2	6.4	17.3	43.4	114.6	1.4
Yield factor									
YF 13	Caturra	123	86.6	72.3	85.0	86.2	87.7	101.8	0.8
	Colombia	81	85.6	82.6	84.5	85.5	86.5	89.7	0.5
	Tipica	20	88.7	82.9	84.7	85.9	89.4	135.0	4.1
YF 14	Caturra	123	88.4	72.9	85.4	86.9	89.5	115.0	2.3
	Colombia	81	86.6	83.1	85.1	86.3	87.5	97.6	1.8
	Tipica	20	97.8	82.9	85.0	86.7	91.2	286.4	4.4
YF 15	Caturra	123	94.3	79.5	87.4	90.4	97.4	165.1	3.1
	Colombia	81	90.3	84.2	86.7	89.1	92.2	112.0	2.1
	Tipica	20	128.9	83.0	86.4	91.0	99.0	777.8	4.4

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Table A1.2. Size characteristics of coffee sampled in Tolima department (in grams per size class), and yield factors for size classes <=13, 14 and => 15.

Screen Size	Variety	N	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Skewness
12	Caturra	83	0.5	0.0	0.2	0.4	0.6	3.2	0.5
	Colombia	57	0.6	0.0	0.2	0.3	0.7	3.9	0.6
	Tipica	29	0.7	0.0	0.3	0.6	0.9	2.8	0.7
13	Caturra	83	2.3	0.0	0.8	1.6	3.1	9.0	2.3
	Colombia	57	1.8	0.1	0.5	1.1	2.1	9.8	1.8
	Tipica	29	2.3	0.1	0.9	2.0	3.5	6.3	2.3
14	Caturra	83	8.4	0.9	5.1	7.6	10.4	29.6	8.4
	Colombia	57	5.9	0.5	1.6	3.9	8.2	22.8	5.9
	Tipica	29	8.0	2.2	4.5	7.0	11.3	18.9	8.0
15	Caturra	83	23.0	3.9	12.8	19.0	32.9	65.4	23.0
	Colombia	57	17.6	3.5	7.5	13.1	24.5	72.5	17.6
	Tipica	29	25.2	6.5	13.6	22.7	37.1	57.9	25.2
16	Caturra	83	61.8	10.1	50.7	62.4	78.0	95.4	61.8
	Colombia	57	47.7	14.4	30.1	45.5	61.0	92.3	47.7
	Tipica	29	67.7	33.9	51.8	71.5	86.4	95.4	67.7
17	Caturra	83	64.8	22.8	54.4	67.7	77.3	95.5	64.8
	Colombia	57	62.2	25.6	53.7	64.3	72.5	106.5	62.2
	Tipica	29	64.4	24.7	50.1	68.9	77.8	96.0	64.4
18	Caturra	83	38.6	5.7	18.0	29.4	49.8	173.2	38.6
	Colombia	57	61.3	3.3	29.4	62.5	87.4	121.5	61.3
	Tipica	29	29.3	5.5	13.7	25.9	43.0	69.1	29.3
Yield factor									
YF 13	Caturra	83	86.1	69.0	85.0	86.1	87.1	107.2	86.1
	Colombia	57	85.6	82.0	84.5	85.7	86.8	89.8	85.6
	Tipica	29	86.8	82.9	84.7	85.4	86.9	110.4	86.8
YF 14	Caturra	83	87.1	70.3	85.6	86.9	88.5	108.5	87.1
	Colombia	57	86.4	82.4	84.7	86.2	87.8	93.1	86.4
	Tipica	29	87.8	83.3	85.2	86.8	88.7	111.0	87.8
YF 15	Caturra	83	91.2	72.1	88.1	90.4	93.5	114.6	91.2
	Colombia	57	89.3	83.4	86.0	87.5	91.8	104.7	89.3
	Tipica	29	91.7	84.8	87.6	90.2	93.4	113.9	91.7

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Table A1.3. Size characteristics of coffee sampled in Santander department (in grams per size class), and yield factors for size classes <=13, 14 and => 15.

Screen Size	Variety	N	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Skewness
12	Caturra	54	0.3	0.0	0.1	0.2	0.4	0.9	0.9
	Colombia	131	0.4	0.0	0.1	0.3	0.5	2.5	2.3
	Tipica	5	0.3	0.1	0.1	0.2	0.3	0.8	1.8
13	Caturra	54	1.3	0.0	0.5	1.1	2.0	4.2	1.0
	Colombia	131	1.6	0.0	0.4	0.9	2.0	9.6	2.1
	Tipica	5	1.0	0.1	0.1	0.8	1.1	3.0	1.6
14	Caturra	54	6.9	0.7	3.8	5.7	8.4	18.8	1.1
	Colombia	131	7.3	0.3	2.2	5.3	11.0	27.3	1.2
	Tipica	5	5.5	2.5	2.8	5.6	6.1	10.3	0.9
15	Caturra	54	17.9	2.6	11.9	17.6	22.5	35.3	0.3
	Colombia	131	18.9	1.1	8.6	15.7	27.2	55.2	0.7
	Tipica	5	15.4	6.0	13.0	13.3	21.8	23.0	0.2
16	Caturra	54	59.3	14.5	48.1	62.2	69.6	92.3	0.5
	Colombia	131	55.3	9.1	40.6	56.8	70.4	104.6	0.2
	Tipica	5	62.0	38.3	53.8	60.6	75.1	82.2	0.3
17	Caturra	54	72.0	49.6	63.4	72.4	80.5	95.0	0.1
	Colombia	131	66.5	26.7	57.6	68.1	76.8	102.4	0.3
	Tipica	5	79.8	63.4	67.1	82.3	84.4	101.6	0.5
18	Caturra	54	46.9	14.4	30.1	42.9	55.4	143.6	1.7
	Colombia	131	55.1	3.6	26.4	45.4	75.7	164.4	1.0
	Tipica	5	42.9	15.6	32.8	39.4	44.1	82.4	1.1
Yield factor									
YF 13	Caturra	54	85.4	68.7	84.7	85.7	86.6	88.8	-4.8
	Colombia	131	85.2	82.7	84.5	85.1	85.9	88.3	0.4
	Tipica	5	84.8	83.6	84.0	84.1	85.6	86.7	0.9
YF 14	Caturra	54	85.9	69.6	85.2	86.1	87.3	89.3	-4.2
	Colombia	131	85.9	82.8	84.9	85.5	86.8	92.8	1.3
	Tipica	5	85.2	84.0	84.0	84.2	86.0	88.0	1.3
YF 15	Caturra	54	89.0	74.9	86.9	88.6	90.7	97.4	-0.7
	Colombia	131	89.3	83.1	86.0	87.9	91.4	107.6	1.5
	Tipica	5	87.6	85.2	85.2	86.3	88.6	92.8	1.4

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Table A1.4. Size characteristics of coffee sampled in Santander Norte department (in grams per size class), and yield factors for size classes <=13, 14 and => 15.

Screen Size	Variety	N	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Skewness
12	Caturra	23	0.7	0.0	0.2	0.4	0.9	3.6	2.4
	Colombia	40	1.0	0.0	0.2	0.5	1.0	9.4	3.6
	Tipica	19	0.3	0.0	0.1	0.2	0.5	0.9	1.2
13	Caturra	23	3.2	0.4	0.6	1.9	4.8	14.9	2.0
	Colombia	40	4.1	0.2	1.1	2.3	5.3	17.3	1.7
	Tipica	19	1.4	0.4	0.8	1.0	1.9	4.1	1.7
14	Caturra	23	9.6	1.7	5.2	7.4	12.7	29.9	1.5
	Colombia	40	13.2	2.4	5.6	10.6	18.6	37.2	1.0
	Tipica	19	7.2	1.3	4.7	7.3	8.9	14.4	0.5
15	Caturra	23	25.8	6.9	15.3	23.3	38.3	54.3	0.5
	Colombia	40	29.4	8.3	18.1	29.2	37.7	69.9	0.6
	Tipica	19	18.1	6.0	11.9	16.2	22.2	46.0	1.5
16	Caturra	23	59.7	23.3	46.8	62.1	71.3	91.1	0.3
	Colombia	40	58.6	20.2	48.9	61.5	70.6	83.6	0.6
	Tipica	19	63.7	47.4	49.8	57.7	73.7	93.3	0.7
17	Caturra	23	63.9	32.9	50.2	70.4	77.1	93.8	0.2
	Colombia	40	59.7	15.7	50.0	58.1	77.9	88.8	0.4
	Tipica	19	78.0	48.7	69.5	79.2	87.0	101.7	0.2
18	Caturra	23	41.0	9.8	24.1	37.4	53.7	97.2	0.8
	Colombia	40	36.5	4.4	18.8	32.5	46.8	122.0	1.3
	Tipica	19	38.5	8.6	26.8	40.9	52.7	62.1	0.2
Yield factor									
YF 13	Caturra	23	86.3	84.2	84.9	86.0	87.4	90.0	0.9
	Colombia	40	86.6	83.9	85.3	86.5	87.3	96.5	2.5
	Tipica	19	84.7	82.8	84.2	84.5	85.5	87.6	0.9
YF 14	Caturra	23	87.7	84.5	85.4	86.4	89.3	97.4	1.5
	Colombia	40	88.7	84.5	86.1	87.3	89.4	109.3	2.8
	Tipica	19	85.3	82.9	84.6	85.0	85.8	89.4	1.4
YF 15	Caturra	23	92.5	85.8	87.6	89.2	95.5	116.9	2.0
	Colombia	40	95.9	86.7	88.9	92.1	98.2	140.6	2.3
	Tipica	19	88.5	83.4	87.4	87.9	88.7	96.3	1.3

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Table A1.5. Size characteristics of coffee sampled in César department (in grams per size class), and yield factors for size classes <=13, 14 and => 15.

Screen Size	Variety	N	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Skewness
12	Caturra	25	0.3	0.0	0.10	0.2	0.4	1.8	2.7
	Colombia	10	0.6	0.0	0.2	0.5	0.9	1.9	1.4
	Tipica	20	0.3	0.0	0.1	0.2	0.4	0.9	1.3
13	Caturra	25	1.9	0.2	0.8	1.2	2.2	12.2	3.3
	Colombia	10	2.7	0.4	0.5	2.5	4.0	7.2	1.0
	Tipica	20	1.8	0.2	0.4	1.0	2.8	6.7	1.3
14	Caturra	25	6.8	0.9	4.8	5.7	7.8	19.6	1.5
	Colombia	10	10.6	3.2	5.7	9.1	16.3	21.9	0.6
	Tipica	20	7.1	1.1	2.6	5.3	10.7	21.6	1.2
15	Caturra	25	20.2	4.1	12.6	18.2	27.4	56.2	1.3
	Colombia	10	27.0	14.7	17.6	26.6	34.2	44.3	0.3
	Tipica	20	18.5	4.5	8.1	14.3	28.0	44.1	0.8
16	Caturra	25	60.6	20.1	47.9	63.7	73.1	88.9	0.5
	Colombia	10	65.5	53.2	58.6	64.4	68.9	88.4	1.1
	Tipica	20	52.1	25.7	32.0	49.0	69.6	94.5	0.5
17	Caturra	25	71.5	35.3	65.8	70.9	78.2	100.2	0.4
	Colombia	10	61.5	45.4	51.8	57.2	73.4	79.7	0.3
	Tipica	20	72.5	30.1	64.6	74.2	87.0	98.0	0.9
18	Caturra	25	44.6	7.3	28.2	39.5	56.5	113.1	1.0
	Colombia	10	38.7	12.4	28.0	36.1	49.0	73.7	0.6
	Tipica	20	54.1	6.4	27.9	58.7	81.1	105.3	0.0
Yield factor									
YF 13	Caturra	25	85.2	82.8	84.1	85.0	86.1	87.8	0.3
	Colombia	10	85.0	83.2	84.2	85.0	85.8	86.6	0.0
	Tipica	20	84.2	83.0	83.5	84.0	84.7	87.5	1.6
YF 14	Caturra	25	86.0	83.0	84.6	85.7	87.3	90.0	0.5
	Colombia	10	86.1	83.4	84.4	86.0	86.8	89.8	0.6
	Tipica	20	85.0	83.3	83.7	84.4	85.5	89.6	1.5
YF 15	Caturra	25	89.1	83.4	86.7	88.5	89.7	100.3	1.4
	Colombia	10	91.1	85.8	86.9	89.7	94.0	101.1	0.9
	Tipica	20	88.3	83.8	85.1	86.9	89.7	100.7	1.4

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Table A1.6. Size characteristics of coffee sampled in Magdalena department (in grams per size class), and yield factors for size classes <=13, 14 and => 15.

Screen Size	Variety	N	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Skewness
12	Caturra	35	0.2	0.0	0.10	0.2	0.3	1.1	2.1
	Colombia	13	0.4	0.0	0.10	0.3	0.6	1.2	1.1
	Tipica	18	0.3	0.0	0.10	0.3	0.4	0.7	0.4
13	Caturra	35	1.5	0.1	0.70	1.2	1.7	4.6	1.5
	Colombia	13	2.2	0.3	0.70	2.7	3.2	4.6	0.1
	Tipica	18	1.4	0.0	0.80	1.2	1.8	4.3	1.1
14	Caturra	35	6.3	2.4	4.10	5.7	8.4	11.2	0.5
	Colombia	13	9.9	3.6	6.00	11.7	14.2	17.4	0.0
	Tipica	18	6.4	0.6	3.00	7.1	7.9	12.8	0.1
15	Caturra	35	19.2	7.0	13.70	19.5	23.4	50.7	2.1
	Colombia	13	24.3	9.7	16.60	25.5	30.1	40.6	0.0
	Tipica	18	17.1	2.5	11.60	18.0	22.5	28.5	0.6
16	Caturra	35	61.6	35.6	51.40	62.9	69.7	97.8	0.4
	Colombia	13	61.1	33.1	54.40	64.8	71.7	76.2	0.9
	Tipica	18	55.9	13.3	40.60	62.6	66.9	84.3	0.9
17	Caturra	35	74.9	39.3	70.80	73.3	80.9	92.7	1.3
	Colombia	13	68.6	51.9	60.30	65.9	71.8	94.9	0.9
	Tipica	18	72.5	36.5	68.90	72.4	81.8	98.1	0.8
18	Caturra	35	43.2	8.3	32.20	41.2	51.9	93.5	0.7
	Colombia	13	41.9	16.1	30.00	35.7	57.3	87.8	0.8
	Tipica	18	49.1	20.5	34.50	43.2	63.4	110.0	1.1
Yield factor									
YF 13	Caturra	35	84.7	82.9	83.80	84.5	85.3	87.6	0.8
	Colombia	13	84.6	83.3	83.50	84.3	85.3	86.4	0.3
	Tipica	18	84.5	82.6	84.10	84.6	85.1	86.0	-0.4
YF 14	Caturra	35	85.3	83.6	84.30	85.1	85.9	88.8	0.9
	Colombia	13	85.5	83.4	84.40	85.2	86.5	88.3	0.7
	Tipica	18	85.1	82.6	84.20	85.3	85.8	87.1	-0.4
YF 15	Caturra	35	88.1	85.5	86.30	88.1	88.9	93.9	1.0
	Colombia	13	89.9	85.1	86.70	90.3	92.0	96.8	0.5
	Tipica	18	87.90	83.7	86.70	88.3	89.2	92.5	-0.3

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Appendix 2: Tables of descriptive statistics of biochemical characteristics

Table A2.1. Descriptive statistics for biochemical data of the Caturra, Colombia, and Tipica coffee varieties in Huila Department.

Biochemical charateristic	Variety	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Skewness
Caffeine	Caturra	1.4	1.1	1.3	1.4	1.4	1.6	-0.1
	Colombia	1.4	1.1	1.3	1.4	1.4	1.6	0.1
	Tipica	1.4	1.2	1.3	1.4	1.5	1.6	0.6
Trigonelline	Caturra	1.0	0.7	0.9	0.9	1.0	1.3	0.7
	Colombia	0.9	0.7	0.8	0.9	1.0	1.2	0.4
	Tipica	1.1	0.8	1.0	1.1	1.1	1.3	0.0
Chlorogenic. acid	Caturra	6.7	5.1	6.3	6.7	7.1	8.3	-0.3
	Colombia	6.6	5.0	6.2	6.6	7.0	8.1	0.2
	Tipica	6.7	5.9	6.4	6.8	7.1	7.9	0.1
CQA total	Caturra	5.6	4.9	5.4	5.6	5.8	6.6	0.4
	Colombia	5.6	5.0	5.5	5.6	5.7	6.3	0.5
	Tipica	5.6	5.0	5.3	5.5	5.9	6.1	0.2
Lipid	Caturra	17.8	14.3	16.7	18.0	18.8	20.5	-0.4
	Colombia	17.3	13.0	16.1	17.5	18.4	20.6	-0.4
	Tipica	17.6	15.7	16.8	17.7	18.6	19.1	-0.4
Palmitic acid	Caturra	35.9	33.8	35.1	35.8	36.4	39.8	1.2
	Colombia	36.0	34.0	35.2	35.9	36.7	40.8	1.0
	Tipica	36.4	34.3	35.9	36.5	37.1	38.6	-0.2
Estearic acid	Caturra	7.4	6.2	7.0	7.4	7.7	9.0	0.0
	Colombia	7.2	5.6	6.9	7.3	7.6	8.3	-0.4
	Tipica	7.2	6.2	6.6	7.3	7.6	8.1	-0.1
Oleic acid	Caturra	9.8	3.0	8.6	9.9	10.9	14.5	-0.3
	Colombia	10.3	3.2	9.4	10.1	11.2	13.6	-0.9
	Tipica	9.6	7.3	8.4	10.0	10.9	11.2	-0.5
Linoleic acid	Caturra	41.1	35.0	39.8	41.4	42.5	46.4	-0.2
	Colombia	40.6	36.3	39.2	40.6	41.9	44.7	-0.1
	Tipica	40.6	36.6	39.7	40.5	41.6	44.4	0.0
Linolenic acid	Caturra	1.8	1.4	1.6	1.8	1.9	2.1	-0.1
	Colombia	1.8	1.5	1.6	1.7	1.8	2.2	0.4
	Tipica	1.8	1.5	1.7	1.8	1.9	2.1	-0.1
Araquidic acid	Caturra	2.1	1.4	1.9	2.1	2.3	2.7	0.0
	Colombia	2.2	1.5	2.0	2.2	2.3	2.7	-0.3
	Tipica	2.2	1.7	2.0	2.1	2.5	2.6	0.0
Behenic acid	Caturra	0.5	0.3	0.5	0.5	0.6	0.7	-0.1
	Colombia	0.5	0.3	0.5	0.5	0.6	0.7	-0.4
	Tipica	0.5	0.4	0.5	0.5	0.6	0.7	-0.1
Sucrose	Caturra	5.0	3.7	4.5	5.0	5.3	6.7	0.2
	Colombia	5.3	3.9	5.0	5.4	5.7	6.4	-0.2
	Tipica	4.7	3.5	4.4	4.6	4.9	5.8	0.2

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Table A2.2. Descriptive statistics for biochemical data of the Caturra, Colombia, and Tipica coffee varieties in Tolima Department.

Biochemical charateristic	Variety	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Skewness
Caffeine	Caturra	1.3	1.1	1.3	1.3	1.4	1.6	-0.3
	Colombia	1.3	1.1	1.3	1.3	1.4	1.6	0.0
	Tipica	1.5	1.2	1.4	1.5	1.6	1.8	-0.3
Trigonelline	Caturra	0.9	0.6	0.8	0.9	0.9	1.1	0.1
	Colombia	0.8	0.5	0.8	0.8	0.9	1.1	-0.4
	Tipica	1.1	0.7	1.0	1.1	1.2	1.4	-0.6
Chlorogenic. acid	Caturra	6.7	5.1	6.3	6.7	7.1	8.0	-0.3
	Colombia	6.6	5.0	6.2	6.6	7.0	7.9	-0.2
	Tipica	6.2	5.2	5.8	6.2	6.4	7.4	0.2
CQA total	Caturra	5.9	5.3	5.7	5.9	6.2	6.9	0.4
	Colombia	5.8	5.1	5.6	5.8	6.1	6.7	0.5
	Tipica	5.6	5.2	5.4	5.6	5.8	6.8	1.6
Lipid	Caturra	17.6	14.3	16.8	17.7	18.6	19.9	-0.5
	Colombia	16.4	11.4	15.3	16.5	17.8	19.8	-0.8
	Tipica	17.3	14.8	16.6	17.0	18.2	19.5	0.0
Palmitic acid	Caturra	35.9	31.3	35.1	36.0	36.9	38.5	-0.9
	Colombia	35.7	32.0	35.0	35.6	36.4	39.6	0.2
	Tipica	34.9	31.7	33.8	35.0	35.8	38.9	0.1
Estearic acid	Caturra	7.9	6.1	7.6	7.9	8.2	9.1	-0.1
	Colombia	7.3	5.5	6.9	7.4	7.8	9.0	-0.1
	Tipica	6.9	5.6	6.3	6.7	7.7	8.7	0.6
Oleic acid	Caturra	11.7	5.5	10.8	11.7	12.8	15.1	-0.9
	Colombia	11.4	5.2	9.9	11.8	12.9	16.1	-0.8
	Tipica	11.5	8.4	10.5	11.1	12.3	16.1	0.8
Linoleic acid	Caturra	40.5	34.1	39.2	40.5	41.8	45.9	-0.1
	Colombia	40.8	35.4	39.9	40.9	42.0	45.3	-0.5
	Tipica	40.8	36.5	38.9	41.6	42.5	44.1	-0.5
Linolenic acid	Caturra	1.7	1.4	1.6	1.7	1.8	2.1	0.2
	Colombia	1.7	1.4	1.6	1.7	1.8	2.1	-0.1
	Tipica	1.7	1.4	1.7	1.7	1.8	2.1	0.3
Araquidic acid	Caturra	2.3	1.7	2.1	2.3	2.5	2.9	-0.1
	Colombia	2.3	1.8	2.1	2.3	2.4	2.9	0.2
	Tipica	2.4	2.0	2.2	2.4	2.6	2.9	0.2
Behenic acid	Caturra	0.5	0.3	0.4	0.5	0.6	0.7	0.3
	Colombia	0.5	0.3	0.4	0.5	0.5	0.7	0.9
	Tipica	0.6	0.4	0.5	0.6	0.7	0.7	0.1
Sucrose	Caturra	5.4	3.5	5.2	5.5	5.8	6.8	-0.5
	Colombia	5.7	4.0	5.3	5.7	6.2	6.8	-0.3
	Tipica	5.1	3.9	4.4	4.8	5.6	7.3	1.0

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Table A2.3. Descriptive statistics for biochemical data of the Caturra, Colombia, and Tipica coffee varieties in Santander Department.

Biochemical charateristic	Variety	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Skewness
Caffeine	Caturra	1.3	1.1	1.3	1.3	1.4	1.5	0.1
	Colombia	1.4	1.2	1.3	1.4	1.5	1.6	0.3
	Tipica	1.4	1.1	1.4	1.4	1.5	1.6	1.7
Trigonelline	Caturra	0.9	0.7	0.9	0.9	0.9	1.1	0.2
	Colombia	0.9	0.7	0.8	0.9	0.9	1.1	0.0
	Tipica	1.0	1.0	1.0	1.0	1.1	1.1	0.3
Chlorogenic. acid	Caturra	6.5	5.4	6.2	6.5	6.9	8.0	0.2
	Colombia	6.4	4.7	5.9	6.4	6.9	7.8	0.0
	Tipica	6.6	6.2	6.3	6.6	6.7	6.9	0.1
CQA total	Caturra	5.6	5.0	5.5	5.6	5.8	6.4	0.2
	Colombia	5.6	4.7	5.4	5.6	5.9	6.3	0.1
	Tipica	5.7	5.3	5.5	5.7	5.8	6.1	0.4
Lipid	Caturra	18.3	16.4	17.6	18.1	19.0	20.7	0.6
	Colombia	17.9	15.4	17.3	17.9	18.6	20.2	0.1
	Tipica	19.4	18.4	18.8	19.8	19.8	19.9	0.7
Palmitic acid	Caturra	35.5	32.7	35.1	35.5	36.1	37.2	0.5
	Colombia	36.0	34.3	35.3	36.0	36.6	38.5	0.3
	Tipica	36.0	34.5	35.4	36.0	36.8	37.1	0.4
Estearic acid	Caturra	8.1	6.9	7.8	8.1	8.5	9.1	0.0
	Colombia	7.9	6.6	7.6	7.8	8.1	9.0	0.1
	Tipica	8.2	7.3	7.6	8.1	8.7	9.4	0.5
Oleic acid	Caturra	13.8	10.7	13.1	14.0	14.6	16.2	0.5
	Colombia	14.5	11.9	14.0	14.5	15.1	17.0	0.3
	Tipica	13.7	12.7	13.3	13.8	14.2	14.7	0.4
Linoleic acid	Caturra	39.7	35.5	38.2	39.8	40.8	44.4	0.2
	Colombia	38.6	34.2	37.5	38.5	39.7	42.4	0.1
	Tipica	38.6	36.3	38.3	38.4	39.3	40.6	0.3
Linolenic acid	Caturra	1.7	1.5	1.7	1.7	1.8	2.0	0.3
	Colombia	1.7	1.2	1.6	1.7	1.8	2.2	0.1
	Tipica	1.6	1.6	1.6	1.6	1.7	1.7	0.5
Araquidic acid	Caturra	2.1	1.7	2.0	2.1	2.3	2.5	0.0
	Colombia	2.2	1.6	2.0	2.2	2.3	3.0	0.1
	Tipica	2.4	2.2	2.3	2.4	2.4	2.5	0.5
Behenic acid	Caturra	0.5	0.3	0.4	0.5	0.5	0.6	0.1
	Colombia	0.4	0.2	0.4	0.5	0.5	0.7	0.2
	Tipica	0.5	0.5	0.5	0.5	0.5	0.6	1.2
Sucrose	Caturra	5.3	4.5	5.1	5.3	5.6	6.1	0.2
	Colombia	5.4	4.4	5.0	5.3	5.6	6.9	0.6
	Tipica	5.2	4.9	5.2	5.2	5.3	5.3	1.0

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Table A2.4. Descriptive statistics for biochemical data of the Caturra, Colombia, and Tipica coffee varieties in Santander Norte Department.

Biochemical charateristic	Variety	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Skewness
Caffeine	Caturra	1.3	1.2	1.2	1.3	1.4	1.5	0.4
	Colombia	1.3	1.1	1.3	1.3	1.4	1.6	0.5
	Tipica	1.4	1.3	1.3	1.4	1.4	1.5	0.1
Trigonelline	Caturra	1.0	0.8	0.9	1.0	1.1	1.2	0.2
	Colombia	1.0	0.7	0.9	1.0	1.0	1.2	0.2
	Tipica	1.1	0.9	1.0	1.1	1.2	1.2	0.1
Chlorogenic. acid	Caturra	6.8	6.0	6.3	6.9	7.2	7.8	0.1
	Colombia	6.8	5.6	6.4	6.7	7.2	7.8	0.2
	Tipica	6.6	5.1	5.9	6.8	7.1	7.9	0.5
CQA total	Caturra	5.8	5.4	5.6	5.8	6.0	6.5	0.7
	Colombia	5.7	5.1	5.4	5.7	6.0	6.5	0.2
	Tipica	5.7	5.0	5.4	5.6	5.9	6.5	0.1
Lipid	Caturra	19.3	17.4	18.7	19.2	20.2	20.8	0.2
	Colombia	19.4	17.0	18.9	19.4	19.9	21.1	0.4
	Tipica	19.8	18.4	19.4	19.7	20.4	21.2	0.1
Palmitic acid	Caturra	36.6	34.9	35.8	36.6	37.4	38.3	0.1
	Colombia	36.4	34.7	35.9	36.4	36.9	38.1	0.2
	Tipica	36.6	34.9	35.6	36.5	37.6	38.2	0.0
Estearic acid	Caturra	8.4	7.2	8.0	8.4	8.6	9.5	0.2
	Colombia	8.5	6.9	8.3	8.5	8.8	9.6	0.6
	Tipica	8.0	7.5	7.5	8.0	8.4	9.0	0.4
Oleic acid	Caturra	14.8	12.9	14.2	14.7	15.4	16.7	0.1
	Colombia	14.8	12.6	14.1	14.7	15.6	16.9	0.2
	Tipica	13.8	11.9	12.9	13.9	14.6	15.6	0.2
Linoleic acid	Caturra	37.6	34.7	36.1	37.5	39.2	41.5	0.5
	Colombia	37.6	34.1	35.9	37.6	39.2	42.1	0.1
	Tipica	38.8	36.1	37.8	38.5	39.6	41.9	0.6
Linolenic acid	Caturra	1.7	1.5	1.7	1.7	1.9	1.9	0.5
	Colombia	1.6	1.4	1.6	1.7	1.7	1.9	0.5
	Tipica	1.7	1.4	1.5	1.7	1.8	1.9	0.2
Araquidic acid	Caturra	2.2	1.7	2.0	2.2	2.4	2.8	0.0
	Colombia	2.4	1.9	2.2	2.4	2.5	2.9	0.1
	Tipica	2.3	2.0	2.1	2.3	2.4	2.7	0.2
Behenic acid	Caturra	0.4	0.3	0.4	0.4	0.5	0.6	0.4
	Colombia	0.5	0.3	0.4	0.5	0.5	0.6	0.2
	Tipica	0.5	0.4	0.4	0.5	0.5	0.6	0.5
Sucrose	Caturra	5.2	4.1	4.8	5.0	5.6	6.6	0.6
	Colombia	5.3	4.0	4.9	5.3	5.8	6.2	0.2
	Tipica	5.2	4.3	5.0	5.2	5.4	5.9	0.6

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Table A2.5. Descriptive statistics for biochemical data of the Caturra, Colombia, and Tipica coffee varieties in César Department.

Biochemical charateristic	Variety	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Skewness
Caffeine	Caturra	1.3	1.1	1.3	1.3	1.4	1.6	0.0
	Colombia	1.3	1.1	1.2	1.3	1.4	1.5	0.3
	Tipica	1.4	1.1	1.3	1.4	1.5	1.6	0.4
Trigonelline	Caturra	0.9	0.6	0.8	0.9	1.0	1.1	0.3
	Colombia	0.9	0.8	0.9	0.9	1.0	1.1	0.3
	Tipica	1.0	0.8	1.0	1.0	1.1	1.2	0.3
Chlorogenic. acid	Caturra	6.8	5.3	6.4	6.9	7.2	8.2	0.0
	Colombia	6.8	6.0	6.2	6.6	7.5	7.9	0.7
	Tipica	6.6	5.7	6.0	6.6	7.1	7.4	0.1
CQA total	Caturra	5.8	5.3	5.6	5.9	6.0	6.5	0.0
	Colombia	5.8	5.2	5.4	5.6	6.2	6.6	0.4
	Tipica	5.8	5.4	5.7	5.8	6.0	6.3	0.5
Lipid	Caturra	18.8	16.0	18.2	18.9	19.3	20.9	0.3
	Colombia	18.1	16.2	17.2	18.0	18.4	20.7	0.9
	Tipica	18.7	17.1	18.2	18.7	19.4	20.2	0.2
Palmitic acid	Caturra	35.5	33.7	35.0	35.5	36.1	36.8	0.2
	Colombia	35.7	34.4	35.3	35.7	36.1	37.3	0.4
	Tipica	35.5	33.5	35.1	35.7	36.0	36.9	0.7
Estearic acid	Caturra	8.5	7.7	8.1	8.5	8.7	9.6	0.1
	Colombia	8.2	7.4	7.9	8.2	8.5	9.2	0.4
	Tipica	8.1	7.0	7.9	8.2	8.4	9.1	0.5
Oleic acid	Caturra	14.4	13.0	13.8	14.3	15.0	16.0	0.3
	Colombia	15.2	14.4	14.5	15.1	15.6	16.3	0.3
	Tipica	14.8	12.6	13.7	14.8	16.0	17.8	0.2
Linoleic acid	Caturra	38.6	35.1	37.7	38.9	39.3	40.3	1.1
	Colombia	38.0	36.2	37.3	38.1	39.2	39.3	0.3
	Tipica	38.3	33.4	36.6	38.8	39.9	41.0	0.7
Linolenic acid	Caturra	1.6	1.4	1.5	1.6	1.7	1.9	0.6
	Colombia	1.6	1.4	1.6	1.7	1.7	1.8	0.8
	Tipica	1.6	1.4	1.5	1.6	1.7	1.8	0.2
Araquidic acid	Caturra	2.4	2.0	2.3	2.4	2.5	2.7	0.2
	Colombia	2.4	2.1	2.3	2.4	2.5	2.7	0.1
	Tipica	2.5	2.2	2.4	2.5	2.6	3.0	0.9
Behenic acid	Caturra	0.5	0.4	0.5	0.5	0.5	0.6	0.3
	Colombia	0.5	0.4	0.4	0.5	0.5	0.6	0.2
	Tipica	0.5	0.5	0.5	0.5	0.6	0.7	0.1
Sucrose	Caturra	5.1	4.4	4.8	5.1	5.3	5.9	0.4
	Colombia	5.3	4.8	5.1	5.3	5.6	5.9	0.2
	Tipica	5.2	4.5	4.8	5.1	5.6	6.3	0.7

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Table A2.6. Descriptive statistics for biochemical data of the Caturra, Colombia, and Tipica coffee varieties in Magdalena Department.

Biochemical charateristic	Variety	N	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Skewness
Caffeine	Caturra		1.4	1.2	1.3	1.4	1.5	1.7	0.2
	Colombia		1.4	1.3	1.4	1.4	1.5	1.6	0.1
	Tipica		1.5	1.4	1.4	1.5	1.6	1.7	0.3
Trigonelline	Caturra		0.9	0.8	0.9	0.9	0.9	1.1	0.3
	Colombia		0.9	0.7	0.9	0.9	1.0	1.1	0.5
	Tipica		1.0	0.8	0.9	1.0	1.1	1.2	0.6
Chlorogenic. acid	Caturra		6.5	5.4	5.9	6.4	7.0	8.0	0.6
	Colombia		6.4	5.3	6.0	6.6	6.7	7.3	0.6
	Tipica		6.2	4.4	5.7	6.2	6.7	7.7	0.3
CQA total	Caturra		6.0	5.2	5.7	6.1	6.3	6.7	0.3
	Colombia		5.9	5.1	5.7	5.9	6.1	6.3	0.8
	Tipica		5.7	4.9	5.6	5.8	5.9	6.2	1.1
Lipid	Caturra		18.2	16.3	17.6	18.0	18.8	20.8	0.4
	Colombia		18.0	16.4	17.0	17.7	18.3	20.3	0.7
	Tipica		18.3	16.4	17.3	18.3	19.0	20.2	0.1
Palmitic acid	Caturra		35.3	33.1	34.7	35.2	35.9	36.5	0.6
	Colombia		35.3	34.3	34.9	35.3	35.7	36.2	0.0
	Tipica		35.4	34.0	34.8	35.3	36.0	36.9	0.2
Estearic acid	Caturra		8.0	6.8	7.6	7.9	8.5	8.9	0.1
	Colombia		7.9	7.0	7.5	8.0	8.2	8.9	0.1
	Tipica		7.8	6.9	7.4	7.7	8.0	9.2	1.0
Oleic acid	Caturra		14.7	12.8	14.0	14.7	15.4	17.1	0.3
	Colombia		15.0	13.6	14.6	14.7	15.5	17.3	1.2
	Tipica		14.7	12.1	14.1	14.9	15.9	16.2	0.8
Linoleic acid	Caturra		38.9	35.5	37.8	39.0	39.8	42.0	0.2
	Colombia		38.5	35.5	38.4	38.7	39.5	39.8	1.4
	Tipica		38.4	34.6	37.0	38.2	40.1	41.6	0.0
Linolenic acid	Caturra		1.6	1.4	1.5	1.6	1.7	1.9	0.4
	Colombia		1.6	1.4	1.6	1.6	1.7	1.8	0.3
	Tipica		1.6	1.3	1.5	1.6	1.6	1.7	0.3
Araquidic acid	Caturra		2.4	2.1	2.3	2.4	2.6	2.8	0.1
	Colombia		2.5	2.3	2.4	2.5	2.6	2.8	0.3
	Tipica		2.6	2.3	2.5	2.6	2.6	2.9	0.2
Behenic acid	Caturra		0.5	0.4	0.5	0.5	0.6	0.6	0.2
	Colombia		0.5	0.5	0.5	0.5	0.6	0.6	0.9
	Tipica		0.6	0.5	0.5	0.6	0.6	0.7	0.1
Sucrose	Caturra		5.3	4.2	4.9	5.3	5.6	6.6	0.2
	Colombia		5.5	4.6	5.2	5.4	5.7	6.5	0.3
	Tipica		5.1	4.5	4.8	5.0	5.3	6.1	0.8

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Appendix 3: Tables of descriptive statistics of sensoric characteristics

Table A3.1. Descriptive statistics for sensoric data of the Caturra, Colombia, and Tipica coffee varieties in Huila Department.

Sensorial charateristic	Variety	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Skewness
Fragrance and aroma	Caturra	5.6	1.0	5.0	6.0	6.5	8.5	-1.3
	Colombia	6.0	1.0	5.0	6.0	7.0	8.5	-1.1
	Tipica	5.8	3.0	5.8	6.0	6.5	7.0	-1.5
Flavor	Caturra	5.6	1.0	5.0	6.0	7.0	9.0	-0.5
	Colombia	6.3	1.0	5.5	6.5	7.5	9.0	-0.7
	Tipica	5.2	2.0	4.0	5.0	6.5	8.5	-0.1
Aftertaste	Caturra	5.6	1.0	4.5	5.5	7.0	9.0	-0.5
	Colombia	6.3	2.0	5.5	6.5	7.5	9.0	-0.6
	Tipica	5.1	2.0	4.0	5.0	6.0	8.5	0.2
Acidity	Caturra	5.8	1.0	5.0	6.0	7.0	9.0	-0.7
	Colombia	6.6	2.0	5.5	7.0	8.0	9.0	-0.6
	Tipica	5.3	2.0	4.5	5.3	6.0	8.0	-0.3
Body	Caturra	5.7	1.0	5.0	6.0	7.0	9.0	-0.7
	Colombia	6.4	2.0	5.5	6.5	7.5	8.5	-0.8
	Tipica	5.5	2.0	5.0	5.5	6.5	8.0	-0.4
Balance	Caturra	5.8	1.0	4.5	6.0	7.0	9.0	-0.5
	Colombia	6.5	2.0	5.5	6.5	7.5	9.0	-0.7
	Tipica	5.3	2.0	4.3	5.3	6.0	8.5	-0.1
Uniformity	Caturra	7.4	3.0	7.0	7.5	8.0	9.0	-1.4
	Colombia	7.5	4.0	7.0	7.5	8.0	9.0	-1.3
	Tipica	7.1	4.0	7.0	7.0	7.5	8.5	-1.7
Clean cup	Caturra	6.0	1.5	5.0	6.0	7.5	9.0	-0.5
	Colombia	6.7	2.5	6.0	7.0	8.0	9.0	-0.5
	Tipica	5.8	3.0	5.0	5.8	6.3	8.5	-0.1
Sweetness	Caturra	5.5	1.0	4.5	6.0	7.0	9.0	-0.7
	Colombia	6.2	2.0	5.0	6.5	7.0	9.0	-0.6
	Tipica	4.9	1.0	3.8	5.0	5.8	8.5	-0.2
Overall	Caturra	5.7	1.0	5.0	6.0	7.0	9.0	-0.6
	Colombia	6.5	2.0	5.5	6.5	8.0	9.5	-0.6
	Tipica	5.3	2.5	4.0	5.0	6.3	9.0	0.4
Final Score	Caturra	58.8	18.5	51.0	61.0	72.0	89.5	-0.6
	Colombia	65.1	25.0	57.0	65.0	75.0	86.5	-0.6
	Tipica	55.2	26.0	49.8	54.8	61.0	82.5	-0.2

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Table A3.2. Descriptive statistics for sensoric data of the Caturra, Colombia, and Tipica coffee varieties in Tolima Department.

Sensorial charateristic	Variety	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Skewness
Fragrance and aroma	Caturra	5.0	1.0	3.5	5.5	6.5	8.0	-0.5
	Colombia	5.4	1.0	5.0	6.0	6.5	8.0	-0.9
	Tipica	5.5	1.0	5.0	5.5	6.5	8.0	-1.2
Flavor	Caturra	4.6	1.0	2.5	5.0	6.0	9.0	0.2
	Colombia	5.1	1.0	4.0	5.0	7.0	8.5	-0.4
	Tipica	5.0	1.0	4.5	5.0	6.0	9.0	-0.1
Aftertaste	Caturra	4.5	1.0	2.0	5.0	6.0	8.5	0.2
	Colombia	5.0	1.0	4.0	5.0	6.5	8.5	-0.2
	Tipica	4.9	1.0	4.0	5.0	5.5	8.5	-0.1
Acidity	Caturra	5.1	1.0	3.0	5.0	6.5	8.5	-0.1
	Colombia	5.5	1.0	5.0	6.0	7.0	9.0	-0.5
	Tipica	5.3	2.0	4.5	5.0	6.0	8.5	0.1
Body	Caturra	5.0	1.0	3.0	5.0	6.5	8.5	-0.1
	Colombia	5.2	1.0	4.0	5.0	7.0	8.0	-0.4
	Tipica	5.1	2.0	4.0	5.0	6.0	8.5	0.5
Balance	Caturra	4.7	2.0	3.0	5.0	6.5	9.0	0.2
	Colombia	5.3	1.0	4.0	5.5	7.0	8.5	-0.3
	Tipica	5.1	2.0	4.5	5.0	6.0	8.5	0.2
Uniformity	Caturra	7.0	4.0	6.0	7.0	8.0	9.0	-1.0
	Colombia	7.2	3.0	6.5	7.5	8.0	9.0	-1.4
	Tipica	7.2	4.0	7.0	7.5	8.0	8.5	-1.4
Clean cup	Caturra	4.9	1.0	3.0	5.0	7.0	9.0	0.1
	Colombia	5.4	1.0	4.0	5.5	7.0	9.0	-0.3
	Tipica	5.1	1.0	4.5	5.0	5.5	8.5	0.3
Sweetness	Caturra	5.0	1.0	3.0	5.0	6.5	8.5	-0.1
	Colombia	5.2	1.0	4.0	5.5	6.5	8.5	-0.5
	Tipica	5.0	1.0	4.5	5.0	5.0	8.0	0.0
Overall	Caturra	4.7	1.0	2.5	5.0	6.5	10.0	0.2
	Colombia	5.1	1.0	4.0	5.0	7.0	8.0	-0.4
	Tipica	5.1	2.0	4.5	5.0	6.0	8.5	0.1
Final Score	Caturra	50.6	21.0	34.0	51.0	63.0	84.5	0.2
	Colombia	54.3	19.0	48.0	54.0	68.0	81.0	-0.3
	Tipica	53.5	22.0	49.0	50.5	59.0	83.5	0.1

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Table A3.3. Descriptive statistics for sensoric data of the Caturra, Colombia, and Tipica coffee varieties in Santander Department.

Sensorial charateristic	Variety	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Skewness
Fragrance and aroma	Caturra	5.1	2.0	3.5	5.5	6.5	8.5	0.4
	Colombia	5.6	3.0	5.0	5.5	6.5	8.0	0.0
	Tipica	6.0	4.0	6.0	6.0	6.0	8.0	0.0
Flavor	Caturra	4.7	1.0	3.0	5.0	6.0	9.0	0.1
	Colombia	5.4	2.0	5.0	5.5	6.0	10.0	0.3
	Tipica	7.4	5.0	6.5	7.0	8.5	10.0	0.3
Aftertaste	Caturra	4.4	1.0	3.0	4.8	6.0	8.5	0.0
	Colombia	5.3	2.0	5.0	5.5	6.0	10.0	0.1
	Tipica	5.7	4.0	4.0	6.0	6.5	8.0	0.2
Acidity	Caturra	5.1	1.0	4.0	5.0	6.0	8.5	0.1
	Colombia	5.5	2.0	5.0	5.0	6.0	10.0	0.5
	Tipica	6.2	5.0	5.0	6.0	7.0	8.0	0.5
Body	Caturra	4.6	1.0	3.0	5.0	6.0	8.0	0.2
	Colombia	5.5	2.0	5.0	6.0	6.0	10.0	0.2
	Tipica	5.9	4.0	5.0	6.0	6.5	8.0	0.2
Balance	Caturra	4.6	2.0	3.0	5.0	6.0	8.0	0.0
	Colombia	5.4	2.0	5.0	5.5	6.0	8.0	0.5
	Tipica	6.2	4.0	5.0	6.5	7.0	8.5	0.0
Uniformity	Caturra	7.0	4.0	7.0	7.5	8.0	8.5	1.3
	Colombia	7.0	3.0	6.5	7.0	8.0	10.0	0.9
	Tipica	6.0	4.0	5.0	6.0	7.0	8.0	0.0
Clean cup	Caturra	5.0	1.0	3.0	5.0	7.0	8.0	0.1
	Colombia	6.1	2.0	5.0	6.0	7.0	9.0	0.4
	Tipica	5.6	3.0	4.0	6.0	7.0	8.0	0.2
Sweetness	Caturra	5.0	1.0	4.0	5.0	6.0	8.0	0.7
	Colombia	5.3	2.0	5.0	5.0	6.0	10.0	0.3
	Tipica	6.7	6.0	6.0	6.5	6.5	8.5	1.9
Overall	Caturra	4.7	1.0	3.0	5.0	6.0	8.0	0.1
	Colombia	5.8	2.0	5.0	6.0	7.0	10.0	0.2
	Tipica	5.6	3.0	4.0	6.0	7.0	8.0	0.2
Final Score	Caturra	50.30	21.5	36.0	52.3	61.5	81.5	0.0
	Colombia	56.7	29.0	50.0	57.0	63.5	83.0	-0.1
	Tipica	61.3	46.0	51.0	61.0	67.0	81.5	0.6

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Table A3.4. Descriptive statistics for sensoric data of the Caturra, Colombia, and Tipica coffee varieties in Santander Norte Department.

Sensorial charateristic	Variety	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Skewness
Fragrance and aroma	Caturra	5.0	2.5	3.5	5.0	6.0	8.5	0.4
	Colombia	5.4	2.0	4.5	5.3	6.5	8.5	0.1
	Tipica	5.2	3.0	4.0	5.5	6.0	8.0	0.0
Flavor	Caturra	4.8	2.0	3.0	5.0	5.5	10.0	1.2
	Colombia	5.2	2.0	4.0	5.3	6.0	10.0	0.5
	Tipica	5.0	2.0	4.0	5.0	6.0	10.0	0.9
Aftertaste	Caturra	4.5	2.0	3.0	5.0	5.0	10.0	1.0
	Colombia	4.9	2.0	3.8	5.0	6.0	10.0	0.4
	Tipica	4.8	2.0	2.0	4.5	6.0	10.0	0.8
Acidity	Caturra	5.2	2.0	5.0	5.0	6.0	10.0	0.7
	Colombia	5.5	3.0	4.3	5.0	6.0	10.0	0.9
	Tipica	5.4	3.0	4.0	5.0	6.0	10.0	1.1
Body	Caturra	4.8	2.0	3.0	5.0	5.0	10.0	1.4
	Colombia	5.3	2.0	4.0	5.0	6.0	10.0	0.7
	Tipica	5.0	2.0	3.0	5.0	6.0	10.0	0.5
Balance	Caturra	4.7	2.0	3.0	5.0	5.5	10.0	1.0
	Colombia	5.0	2.0	3.8	5.0	6.0	10.0	0.2
	Tipica	4.9	2.0	3.0	5.0	6.0	10.0	0.7
Uniformity	Caturra	6.2	2.0	5.0	6.0	7.0	10.0	0.3
	Colombia	6.6	4.0	6.0	6.5	7.3	10.0	0.2
	Tipica	7.1	4.0	7.0	7.0	8.0	10.0	0.7
Clean cup	Caturra	4.5	2.0	3.0	5.0	5.5	8.5	0.2
	Colombia	5.1	2.0	3.3	5.3	6.5	8.0	0.3
	Tipica	4.9	2.0	3.0	5.5	6.5	7.5	0.4
Sweetness	Caturra	4.7	2.0	4.0	5.0	6.0	8.5	0.0
	Colombia	5.8	3.0	5.0	5.5	6.8	10.0	0.9
	Tipica	5.7	3.0	4.0	5.5	7.0	10.0	0.7
Overall	Caturra	4.7	2.0	2.5	5.0	5.5	10.0	0.9
	Colombia	5.2	2.0	3.5	5.5	6.5	10.0	0.1
	Tipica	4.9	2.0	3.0	5.5	6.5	7.5	0.4
Final Score	Caturra	49.2	29.0	35.5	51.0	56.5	84.5	0.8
	Colombia	54.0	33.5	41.5	55.5	64.0	82.5	0.2
	Tipica	53.0	29.5	39.5	52.5	63.0	82.0	0.3

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Table A3.5. Descriptive statistics for sensoric data of the Caturra, Colombia, and Tipica coffee varieties in César Department.

Sensorial charateristic	Variety	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Skewness
Fragrance and aroma	Caturra	5.1	2.0	4.0	5.0	6.5	8.5	0.2
	Colombia	5.9	5.0	5.5	5.8	6.5	7.0	0.3
	Tipica	5.0	3.0	4.3	5.0	5.5	7.0	0.0
Flavor	Caturra	5.3	2.0	4.0	6.0	6.0	10.0	0.4
	Colombia	6.5	4.5	5.0	6.0	7.5	10.0	1.0
	Tipica	5.0	3.0	3.8	5.0	5.8	8.0	0.5
Aftertaste	Caturra	4.8	2.0	3.0	5.0	6.0	9.0	0.0
	Colombia	6.7	4.0	5.5	6.5	7.0	10.0	0.7
	Tipica	5.3	2.0	3.3	5.0	6.5	10.0	0.8
Acidity	Caturra	5.5	3.0	5.0	5.5	6.0	8.5	0.5
	Colombia	6.3	4.0	5.0	6.0	7.0	10.0	0.9
	Tipica	5.1	2.0	4.5	5.0	5.0	10.0	1.5
Body	Caturra	5.0	3.0	4.0	5.0	6.0	8.5	0.2
	Colombia	6.3	4.0	5.0	6.0	7.0	10.0	1.1
	Tipica	5.2	2.0	4.0	5.0	6.0	10.0	0.8
Balance	Caturra	4.9	2.0	3.0	5.0	6.0	9.0	0.0
	Colombia	6.9	4.0	5.0	7.0	7.5	10.0	0.4
	Tipica	5.1	2.0	3.0	5.0	6.0	10.0	1.0
Uniformity	Caturra	6.7	3.0	6.5	7.0	7.5	9.0	1.1
	Colombia	7.8	5.0	7.0	8.0	8.0	10.0	0.1
	Tipica	7.3	3.0	6.8	7.8	8.0	10.0	0.8
Clean cup	Caturra	5.4	2.0	3.0	6.0	7.5	9.0	0.3
	Colombia	6.4	3.0	5.5	6.8	7.5	8.0	1.1
	Tipica	5.1	2.0	4.0	5.0	6.3	8.0	0.1
Sweetness	Caturra	5.5	3.0	4.5	6.0	6.5	8.5	0.2
	Colombia	6.1	3.5	5.0	6.0	7.0	8.0	0.3
	Tipica	5.0	2.0	4.0	5.0	5.8	10.0	1.4
Overall	Caturra	5.4	2.0	3.0	6.0	7.5	9.0	0.3
	Colombia	6.7	3.0	5.5	6.8	8.0	10.0	0.2
	Tipica	5.2	2.0	4.0	5.0	6.0	10.0	0.6
Final Score	Caturra	53.8	31.0	42.5	52.5	63.0	88.0	0.2
	Colombia	65.4	42.5	52.0	71.8	75.5	78.5	-0.8
	Tipica	53.3	28.5	40.0	51.3	65.3	80.5	0.3

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Table A3.6. Descriptive statistics for sensoric data of the Caturra, Colombia, and Tipica coffee varieties in Magdalena Department.

Sensorial charateristic	Variety	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Skewness
Fragrance and aroma	Caturra	5.1	3.5	4.5	5.0	5.5	7.5	0.5
	Colombia	5.2	4.0	4.5	5.0	6.0	7.0	0.7
	Tipica	5.2	3.5	5.0	5.0	6.0	6.5	0.3
Flavor	Caturra	4.8	1.0	4.0	5.0	5.5	8.0	0.5
	Colombia	5.0	3.0	4.0	5.0	5.0	8.0	0.6
	Tipica	5.9	4.0	5.0	5.5	6.5	10.0	1.4
Aftertaste	Caturra	4.8	1.0	3.0	5.0	5.5	10.0	0.6
	Colombia	5.3	2.0	4.0	5.0	5.5	10.0	0.9
	Tipica	5.2	3.0	4.0	5.0	6.0	8.0	0.5
Acidity	Caturra	4.8	2.0	4.0	5.0	5.5	10.0	0.6
	Colombia	5.3	3.0	5.0	5.0	5.5	10.0	2.1
	Tipica	5.9	4.0	5.0	5.3	6.5	10.0	1.3
Body	Caturra	4.9	2.0	4.0	5.0	6.0	10.0	0.2
	Colombia	5.0	3.0	5.0	5.0	6.0	7.0	0.6
	Tipica	5.5	4.0	5.0	5.0	6.0	8.0	0.7
Balance	Caturra	4.9	2.0	3.0	5.0	6.0	10.0	0.4
	Colombia	4.9	2.0	5.0	5.0	6.0	7.0	0.8
	Tipica	5.7	4.0	5.0	5.5	6.5	10.0	1.6
Uniformity	Caturra	6.6	2.0	6.0	6.5	8.0	10.0	0.8
	Colombia	5.8	3.0	5.0	6.0	6.0	8.0	0.4
	Tipica	6.4	4.0	5.5	6.5	7.5	8.5	0.3
Clean cup	Caturra	5.1	1.0	3.0	5.5	6.0	8.5	0.2
	Colombia	5.0	2.0	4.0	5.0	5.5	7.0	0.9
	Tipica	5.9	4.0	5.5	6.0	7.0	8.0	0.1
Sweetness	Caturra	4.7	2.0	4.0	5.0	5.0	10.0	0.7
	Colombia	5.0	3.0	4.5	5.0	6.0	7.5	0.1
	Tipica	5.8	4.0	5.0	5.0	7.0	10.0	1.3
Overall	Caturra	5.0	1.0	3.0	5.0	6.0	10.0	0.2
	Colombia	5.2	2.0	4.5	5.0	5.5	10.0	1.2
	Tipica	5.6	3.5	4.0	5.5	7.0	8.0	0.1
Final Score	Caturra	50.6	18.5	39.0	51.0	57.0	80.5	0.0
	Colombia	51.7	37.0	47.5	51.0	57.0	71.5	0.3
	Tipica	57.0	43.5	50.0	54.0	66.5	75.0	0.6

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Appendix 4: Tables of descriptive statistics of climate characteristics across clusters.

Table A4.1. Descriptive statistics for climate data of the five clusters in the Huila/Tolima Region.

Climate charateristic	Cluster	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Kurtosis	Skewness
Annual precipitation	1	2460	2340	2430	2450	2490	2520	-0.2	-0.4
(mm)	2	1700	1380	1590	1740	1800	1950	-0.8	-0.5
	3	1750	1340	1690	1730	1780	2080	2.1	0.6
	4	2170	1860	2070	2160	2290	2420	-0.9	-0.1
	5	2190	1850	2110	2210	2280	2420	0.3	-0.7
Annual evaporation	1	870	780	830	860	910	990	-0.6	0.5
(mm)	2	880	720	820	870	940	1070	-0.8	0.1
	3	820	700	780	810	870	980	-0.5	0.4
	4	880	720	830	880	950	1020	-0.8	0.1
	5	860	690	800	870	900	1000	-0.4	-0.1
Mean dewpoint	1	14.0	13.2	13.8	14.0	14.2	14.9	-0.3	0.3
(°C)	2	14.7	12.8	14.2	14.7	15.1	18.9	4.8	1.3
	3	13.4	12.1	12.9	13.3	13.8	14.9	-0.3	0.3
	4	15.1	13.1	14.6	15.1	15.6	17.9	0.9	0.7
	5	13.0	10.2	12.6	13.1	13.6	14.4	1.5	-1.1
Number of dry months	1	1.0	1.0	1.0	1.0	1.0	2.0	17.8	4.4
	2	1.8	0.0	1.0	2.0	2.0	4.0	-0.2	0.1
	3	1.3	0.0	0.0	2.0	2.0	4.0	-0.5	0.0
	4	1.9	1.0	2.0	2.0	2.0	2.0	5.3	-2.7
	5	1.1	0.0	1.0	1.0	1.0	2.0	1.8	1.1
Mean temperature	1	20.3	19.5	20.0	20.2	20.6	21.2	-0.4	0.2
(°C)	2	20.3	18.2	19.7	20.1	20.8	25.0	5.4	1.6
	3	19.0	17.3	18.4	19.1	19.5	21.1	-0.1	0.3
	4	21.3	19.2	20.8	21.2	21.8	24.1	1.9	0.9
	5	19.2	16.0	18.9	19.3	19.8	20.8	2.5	-1.3
Mean diurnal	1	10.4	10.2	10.3	10.4	10.4	10.6	0.5	0.4
temperature range	2	10.0	9.0	9.7	10.0	10.3	11.0	-0.9	-0.1
(°C)	3	10.0	8.9	9.5	10.0	10.5	10.7	-1.2	-0.5
	4	10.5	9.4	10.5	10.6	10.7	10.8	1.9	-1.8
	5	10.3	9.9	10.2	10.4	10.5	10.8	-0.4	-0.2
Mean daily solar	1	23.8	21.0	23.0	24.0	25.0	25.0	-0.9	-0.5
radiation	2	23.6	19.0	23.0	24.0	25.0	25.0	1.3	-1.1
($MJ\ m^{-2}\ d^{-1}$)	3	23.3	21.0	22.0	23.0	25.0	25.0	-1.2	-0.3
	4	23.5	21.0	23.0	23.0	24.5	25.0	-0.5	-0.2
	5	23.2	19.0	22.0	23.0	24.0	25.0	0.6	-0.5
Mean cloud cover	1	94.8	94.0	95.0	95.0	95.0	95.0	0.6	-1.6
(%)	2	95.9	95.0	95.0	96.0	97.0	97.0	-1.5	0.1
	3	96.1	95.0	96.0	96.0	97.0	97.0	-1.1	-0.1
	4	95.4	94.0	95.0	95.0	96.0	97.0	0.4	1.0
	5	95.3	94.0	95.0	95.0	96.0	97.0	0.8	0.6
Altitude	1	1660	1550	1600	1680	1700	1790	-1.0	0.1
(masl)	2	1460	680	1400	1490	1550	1660	8.6	-2.3
	3	1740	1500	1670	1730	1820	2000	-0.3	0.1
	4	1400	860	1340	1430	1500	1610	3.2	-1.6
	5	1800	1610	1700	1770	1900	2240	0.7	0.9
Aspect	1	162	9	79	124	262	359	-1.3	0.3
(compass °)	2	231	1	168	267	300	359	-0.2	-0.8
	3	180	4	66	190	277	356	-1.4	-0.1
	4	178	8	98	161	275	357	-1.2	0.1
	5	203	15	109	202	296	358	-1.3	-0.1
Slope	1	17.7	1.4	9.3	17.4	26.3	36.4	-1.2	0.1
(°)	2	17.4	3.7	10.7	16.3	24.2	42.4	-0.5	0.4
	3	20.8	5.3	12.7	20.6	28.0	40.0	-1.1	0.1
	4	19.5	5.8	13.2	19.3	25.1	35.2	-0.8	0.1
	5	21.4	2.1	16.1	22.0	26.7	38.5	0.0	0.0

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Table A4.2. Descriptive statistics for climate data of the five clusters in the Northern Region.

Climate charateristic	Cluster	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Kurtosis	Skewness
Annual precipitation	1	2260	1800	2210	2300	2370	2550	-0.1	-0.8
(mm)	2	2560	2220	2420	2490	2730	2840	-1.2	0.1
	3	2430	2030	2340	2390	2460	2980	1.4	1.2
	4	1990	1290	1880	2030	2130	2310	0.9	-1.0
	5	1400	1220	1330	1390	1450	1680	0.0	0.2
Annual evaporation	1	960	690	900	960	1020	1150	0.2	-0.1
(mm)	2	1130	840	1010	1120	1280	1400	-1.1	0.0
	3	1320	990	1240	1360	1450	1600	-0.3	-0.6
	4	1030	780	930	1060	1110	1200	-0.8	-0.5
	5	990	790	920	980	1040	1230	-0.1	0.3
Mean dewpoint	1	13.1	3.5	12.6	13.1	13.8	15.6	26.1	-3.7
(°C)	2	14.0	12.2	13.4	14.0	14.6	16.7	0.3	0.2
	3	16.5	13.5	15.7	16.6	17.4	19.4	-0.2	-0.1
	4	15.5	13.4	14.3	15.2	16.0	21.4	3.2	1.6
	5	14.2	11.3	13.2	14.3	15.0	17.3	-0.3	0.2
Number of dry months	1	1.7	0.0	1.0	1.0	2.0	4.0	-0.1	1.2
	2	1.9	0.0	0.0	1.0	4.0	4.0	-2	0.1
	3	3.7	1.0	4.0	4.0	4.0	5.0	2.8	-1.9
	4	3.1	1.0	3.0	3.0	4.0	5.0	-0.3	-0.8
	5	4.4	3.0	4.0	4.0	5.0	6.0	-0.8	0.3
Mean temperature	1	19.1	8.8	18.5	19.2	19.9	21.6	22.5	-3.3
(°C)	2	19.9	18.1	19.5	19.9	20.3	21.7	0.0	-0.2
	3	21.8	19.1	21.2	22.0	22.6	24.5	-0.3	-0.2
	4	21.2	18.6	20.2	20.9	22.0	27.6	3.6	1.5
	5	19.6	16.4	18.6	19.5	20.5	22.6	-0.3	0.2
Mean diurnal	1	10.6	9.1	10.3	10.6	11.0	11.4	-0.1	-0.4
temperature range	2	10.6	9.0	10.1	10.7	11.0	11.4	-0.3	-0.6
(°C)	3	10.0	9.1	9.5	10.2	10.3	10.8	-1.1	-0.4
	4	10.5	9.1	10.2	10.5	10.8	11.6	1.2	-0.7
	5	9.9	9.3	9.7	9.9	10.1	10.4	-0.3	-0.1
Mean daily solar	1	24.1	17.0	24.0	25.0	25.0	25.0	7.2	-2.4
radiation	2	23.5	18.0	23.0	24.0	25.0	25.0	3.8	-1.8
($MJ\ m^{-2}\ d^{-1}$)	3	22.7	18.0	21.0	23.0	24.0	25.0	0.0	-0.5
	4	23.1	18.0	22.0	24.0	25.0	25.0	-0.3	-0.8
	5	23.1	16.0	23.0	23.0	24.0	25.0	6.5	-1.9
Mean cloud cover	1	94.5	84.7	94.4	96.0	96.7	97.6	1.4	-1.7
(%)	2	89.6	76.7	81.1	95.4	96.7	97.5	-1.7	-0.4
	3	81.4	74.3	77.9	79.9	81.9	96.0	1.1	1.5
	4	91.1	84.3	87.2	90.4	96.0	97.4	-1.7	0.1
	5	92.6	80.3	89.4	93.8	95.7	96.8	0.8	-0.7
Altitude	1	1710	1330	1600	1690	1770	3360	28.5	3.9
(masl)	2	1520	1160	1420	1500	1600	1870	0.3	0.4
	3	1040	660	900	1050	1200	1320	-0.7	-0.4
	4	1210	90	1110	1230	1400	1530	5.3	-1.9
	5	1440	980	1310	1460	1560	1840	-0.1	-0.4
Aspect	1	195	1	81	223	301	358	-1.5	-0.2
(compass °)	2	209	3	117	252	301	359	-1.1	-0.5
	3	230	2	153	243	323	357	-0.5	-0.6
	4	196	22	92	189	300	357	-1.3	0.0
	5	178	7	92	178	271	351	-1.3	-0.1
Slope	1	12.5	0.6	7.2	10.4	16.4	40.1	1.2	1.2
(°)	2	15.6	2.5	8.9	13.4	22.5	41.4	-0.1	0.8
	3	21.2	1.0	15.4	21.4	29.5	42.4	-0.3	-0.1
	4	16.4	1.1	10.4	15.8	21.9	30.2	-0.8	0.0
	5	19.3	5.3	13.8	19.4	24.4	35.3	-0.5	-0.1

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Appendix 5: Tables of descriptive statistics of soils' characteristics across clusters.

Table A5.1. Descriptive statistics for data of soil macronutrients, aluminium and cation exchange capacity of the five clusters in the Huila/Tolima Region.

Soil charateristic	Cluster	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Kurtosis	Skewness
pH	1	5.3	4.0	4.8	5.3	5.7	7.1	-0.4	0.0
	2	4.7	3.7	4.4	4.7	4.9	6.2	0.9	0.7
	3	4.5	3.8	4.2	4.4	4.7	6.0	1.4	1.4
	4	4.8	4.0	4.2	4.4	5.2	6.6	1.7	1.4
	5	5.7	4.3	4.8	5.5	6.9	7.2	-1.8	0.2
Nitrogen (%)	1	0.3	0.1	0.2	0.3	0.3	0.7	2.9	1.2
	2	0.3	0.2	0.3	0.3	0.4	0.6	0.6	0.7
	3	0.4	0.3	0.3	0.4	0.5	0.7	0.1	0.8
	4	0.4	0.2	0.3	0.3	0.4	0.7	0.9	1.4
	5	0.4	0.3	0.4	0.4	0.4	0.5	1.9	1.0
Organic matter (%)	1	6.9	2.0	5.0	6.5	8.1	23.7	7.4	2.0
	2	8.4	4.1	6.8	8.1	9.7	15.9	1.1	1.0
	3	9.7	5.5	7.2	9.3	11.9	19.3	0.8	1.0
	4	10.0	5.2	6.1	7.4	9.1	22.8	1.8	1.6
	5	10.1	7.4	9.0	9.9	10.9	14.1	1.9	1.0
Potassium (cmol/kg	1	0.3	0.1	0.2	0.2	0.4	1.2	4.0	1.7
	2	0.4	0.1	0.3	0.3	0.4	0.8	0.9	0.9
	3	0.3	0.1	0.2	0.3	0.5	0.8	-0.1	0.8
	4	2.1	0.1	0.3	0.4	0.6	14.8	8.9	3.0
	5	0.5	0.2	0.2	0.5	1.0	1.0	-1.7	0.4
Calcium (cmol/kg)	1	6.6	0.2	3.4	6.2	9.2	20.7	-0.1	0.6
	2	5.0	0.2	1.6	3.8	8.1	16.7	0.5	1.1
	3	3.8	0.3	1.1	1.9	4.7	14.1	1.1	1.5
	4	3.5	0.2	1.0	2.4	4.1	13.1	4.8	2.1
	5	12.5	2.8	4.1	6.0	20.3	30.0	-1.0	0.8
Magnesium (cmol/kg)	1	1.8	0.1	0.9	1.6	2.5	5.7	0.6	0.9
	2	1.2	0.2	0.5	1.0	1.8	3.8	0.9	1.0
	3	1.0	0.2	0.5	0.8	1.1	5.6	13.2	3.2
	4	1.7	0.2	0.4	0.6	1.2	8.7	6.7	2.5
	5	1.8	0.5	1.6	1.8	2.2	2.9	1.9	-0.5
Aluminium	1	1.1	0.1	0.2	0.3	1.4	9.8	7.6	2.5
	2	2.6	0.1	0.6	1.9	3.6	9.2	0.7	1.1
	3	4.1	0.1	1.5	3.8	5.9	14.0	1.3	1.0
	4	3.0	0.2	0.4	3.9	4.3	8.5	0.4	0.8
	5	0.7	0.1	0.1	0.2	1.2	2.9	3.9	2.0
Cation exchange capacity (cmol/kg)	1	15.9	6.0	12.0	16.0	19.0	38.0	2.1	0.9
	2	18.3	10.0	16.0	18.0	20.0	32.0	0.4	0.6
	3	20.1	10.0	15.0	20.0	23.0	44.0	4.6	1.5
	4	17.6	12.0	14.0	15.0	20.0	25.0	-1.2	0.5
	5	20.3	11.0	18.0	21.0	23.0	27.0	2.2	-1.0

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Table A5.2. Descriptive statistics for data of soil phosphorus, iron, micronutrients, and soil components of the five clusters in the Huila/Tolima Region.

Soil charateristic	Cluster	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Kurtosis	Skewness
Phosphorus (mg/kg)	1	28.3	1.0	5.0	12.5	32.5	228.0	7.5	2.5
	2	29.8	1.0	3.0	12.0	37.0	144.0	2.0	1.7
	3	40.6	2.0	6.0	14.0	34.0	253.0	4.2	2.2
	4	77.1	1.0	2.0	28.0	100.0	251.0	-0.1	1.2
	5	584.7	273.0	310.0	430.0	810.0	1364.0	2.6	1.7
Iron (mg/kg)	1	268	67	191	259	332	498	-0.7	0.4
	2	603	484	546	594	662	790	-0.5	0.5
	3	925	797	845	893	1007	1168	-0.5	0.8
	4	1356	1232	1263	1324	1455	1480	-2.3	0.1
	5	401	183	189	333	573	840	1.3	1.3
Manganese (mg/kg)	1	68.1	4.0	32.0	54.5	92.5	353.0	5.9	2.0
	2	43.0	4.0	17.0	29.0	59.0	152.0	0.9	1.3
	3	48.6	6.0	11.0	24.5	60.0	232.0	3.5	2.1
	4	58.8	7.0	15.0	21.0	33.0	204.0	0.7	1.6
	5	71.4	21.0	32.0	64.0	80.0	199.0	4.3	1.9
Zinc (mg/kg)	1	3.6	0.0	1.0	2.0	4.0	31.0	16.3	3.4
	2	4.7	1.0	2.0	4.0	7.0	19.0	4.0	1.7
	3	5.2	1.0	2.0	4.0	7.0	17.0	1.5	1.5
	4	11.8	2.0	4.0	5.0	6.0	65.0	8.7	2.9
	5	15.4	5.0	9.0	13.0	24.0	32.0	0.0	0.9
Copper (mg/kg)	1	2.8	0.0	1.0	2.0	4.0	19.0	9.3	2.3
	2	2.7	0.0	1.0	2.0	3.0	32.0	35.5	5.5
	3	3.1	0.0	1.0	2.0	4.0	30.0	23.9	4.6
	4	4.2	1.0	2.0	2.0	3.0	18.0	7.0	2.6
	5	5.4	2.0	3.0	4.0	8.0	12.0	0.8	1.2
Sand (%)	1	30.7	12.0	21.5	29.0	38.0	63.0	-0.3	0.7
	2	40.0	15.0	32.0	40.0	48.0	64.0	-0.5	-0.1
	3	40.6	25.0	31.0	37.0	51.0	65.0	-1.1	0.4
	4	37.4	18.0	32.0	38.0	49.0	53.0	-1.0	-0.1
	5	27.6	15.0	21.0	26.0	36.0	39.0	-1.3	0.0
Silt (%)	1	26.5	12.0	21.0	26.0	31.0	49.0	0.3	0.6
	2	24.2	12.0	18.0	23.0	28.0	47.0	0.5	1.0
	3	25.4	15.0	20.0	23.5	29.0	47.0	0.5	1.0
	4	25.1	19.0	19.0	23.0	29.0	34.0	-1.6	0.4
	5	30.1	19.0	27.0	29.0	37.0	43.0	0.5	0.5
Clay (%)	1	42.9	10.0	29.0	43.0	55.0	71.0	-1.0	-0.1
	2	35.7	15.0	25.0	32.0	44.0	68.0	-0.3	0.6
	3	34.1	18.0	22.0	32.0	44.0	58.0	-1.2	0.3
	4	38.8	14.0	28.0	44.0	53.0	58.0	-1.3	-0.4
	5	42.0	23.0	32.0	39.0	58.0	60.0	-1.2	0.2

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Table A5.3. Descriptive statistics for data of soil macronutrients, aluminium and cation exchange capacity of the five clusters in the Northern Region.

Soil charateristic	Cluster	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Kurtosis	Skewness
pH	1	4.9	4.0	4.5	4.8	5.2	6.3	0.6	0.7
	2	4.4	3.7	4.2	4.3	4.5	6.0	3.7	1.7
	3	5.2	4.1	4.9	5.1	5.6	7.4	1.7	1.0
	4	4.7	3.9	4.3	4.6	4.9	6.5	1.9	1.2
	5	4.2	4.0	4.1	4.2	4.3	4.7	0.7	1.1
Nitrogen (%)	1	0.3	0.1	0.3	0.3	0.4	0.6	0.7	0.8
	2	0.4	0.2	0.3	0.3	0.4	0.6	-0.1	0.7
	3	0.3	0.1	0.2	0.3	0.3	0.7	1.3	1.1
	4	0.3	0.2	0.3	0.3	0.4	0.5	-0.8	0.0
	5	0.4	0.3	0.3	0.4	0.4	0.6	0.0	0.7
Organic matter (%)	1	7.5	2.8	5.6	6.9	8.4	18.4	2.9	1.6
	2	8.6	4.3	6.8	7.9	10.0	17.1	0.7	0.9
	3	6.8	1.2	4.7	5.8	7.9	20.5	2.8	1.6
	4	7.8	3.9	5.9	7.4	9.5	13.1	-0.8	0.2
	5	9.7	5.5	8.0	9.0	11.1	16.8	0.4	0.8
Potassium (cmol/kg	1	0.3	0.1	0.2	0.2	0.3	1.4	11.2	2.8
	2	0.3	0.1	0.2	0.2	0.3	1.4	21.8	3.9
	3	0.3	0.1	0.2	0.2	0.3	1.9	15.7	3.5
	4	0.3	0.1	0.2	0.2	0.3	1.2	15.0	3.2
	5	0.2	0.1	0.2	0.2	0.3	0.6	6.3	2.1
Calcium (cmol/kg)	1	4.4	0.1	1.3	4.1	6.4	15.7	0.6	0.9
	2	2.9	0.2	0.8	1.7	2.9	20.2	6.0	2.3
	3	5.9	0.2	2.3	4.9	7.7	23.6	1.5	1.3
	4	4.1	0.2	1.1	3.6	5.6	14.3	0.8	1.2
	5	1.9	0.4	0.9	1.3	2.6	6.7	4.1	1.8
Magnesium (cmol/kg)	1	1.0	0.1	0.3	0.9	1.4	7.1	13.7	3.1
	2	0.7	0.1	0.2	0.4	0.6	9.1	30.5	5.0
	3	1.2	0.1	0.4	0.8	1.4	10.0	14.8	3.3
	4	0.9	0.1	0.3	0.6	1.5	3.6	1.1	1.1
	5	0.5	0.1	0.3	0.4	0.6	1.5	5.4	2.0
Aluminium	1	1.8	0.1	0.5	1.3	3.0	7.8	1.4	1.1
	2	4.1	0.2	2.3	4.1	5.3	12.9	1.1	0.9
	3	1.1	0.1	0.2	0.5	1.3	13.0	25.2	4.2
	4	2.7	0.1	0.7	2.8	4.0	9.6	0.6	0.9
	5	5.5	1.9	3.9	5.6	7.1	10.8	0.1	0.6
Cation exchange capacity (cmol/kg)	1	16.2	8.0	13.0	16.0	19.0	30.0	0.3	0.6
	2	18.8	10.0	15.0	17.5	22.0	35.0	0.3	0.8
	3	14.9	4.0	11.0	14.0	18.0	34.0	1.0	0.8
	4	16.7	9.0	13.0	16.0	20.0	33.0	0.8	0.8
	5	21.8	10.0	18.0	20.0	25.0	39.0	2.3	0.9

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Table A5.4. Descriptive statistics for data of soil phosphorus, iron, micronutrients, and soil components of the five clusters in the Northern Region.

Soil charateristic	Cluster	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Kurtosis	Skewness
Phosphorus (mg/kg)	1	17.3	1.0	3.0	6.0	13.0	133.0	6.7	2.7
	2	25.5	1.0	3.0	5.0	11.0	291.0	9.1	3.1
	3	25.9	1.0	4.0	7.0	16.0	272.0	10.2	3.2
	4	16.9	1.0	4.0	5.0	14.0	135.0	6.5	2.7
	5	17.4	2.0	4.0	5.0	13.0	120.0	7.6	2.8
Iron (mg/kg)	1	376	291	341	372	415	473	-0.8	0.2
	2	924	703	808	905	1035	1245	-0.7	0.5
	3	219	51	176	231	267	433	-0.1	-0.2
	4	596	472	524	614	654	715	-1.3	-0.2
	5	1406	1278	1318	1370	1451	1645	-0.2	1.0
Manganese (mg/kg)	1	48.5	2.0	12.0	32.5	74.0	191.0	0.7	1.1
	2	23.1	2.0	5.0	9.0	17.0	226.0	10.7	3.1
	3	82.8	1.0	21.0	42.0	112.0	498.0	4.8	2.1
	4	47.8	0.0	13.0	38.0	79.0	138.0	-0.7	0.7
	5	18.8	2.0	5.0	16.0	23.0	85.0	8.0	2.5
Zinc (mg/kg)	1	3.6	0.0	1.0	2.0	4.5	13.0	0.8	1.3
	2	5.3	0.0	2.0	3.0	4.0	59.0	19.3	4.1
	3	3.5	0.0	1.0	2.0	4.0	26.0	9.0	2.7
	4	5.7	0.0	2.0	3.0	5.0	59.0	20.0	4.2
	5	5.6	1.0	3.0	4.0	7.0	33.0	14.0	3.6
Copper (mg/kg)	1	3.3	0.0	1.0	3.0	5.0	17.0	6.6	2.0
	2	3.7	0.0	1.0	2.0	4.0	51.0	34.4	5.3
	3	2.9	0.0	1.0	2.0	4.0	12.0	2.7	1.4
	4	11.6	0.0	1.0	2.0	4.0	475.0	57.4	7.5
	5	3.2	1.0	1.0	1.0	4.0	11.0	1.0	1.4
Sand (%)	1	37.2	17.0	29.0	35.0	45.0	67.0	-0.1	0.7
	2	44.0	20.0	36.0	44.0	51.0	68.0	-0.5	0.2
	3	32.0	5.0	24.0	31.0	41.0	69.0	0.5	0.5
	4	40.4	17.0	33.0	41.0	49.0	63.0	-0.6	-0.3
	5	48.6	22.0	37.0	48.0	60.0	69.0	-0.4	-0.3
Silt (%)	1	26.5	15.0	21.0	25.0	31.5	49.0	0.3	0.7
	2	25.9	12.0	21.0	25.0	29.0	51.0	1.1	0.9
	3	27.3	12.0	22.0	27.0	31.0	48.0	-0.1	0.5
	4	26.5	13.0	21.0	27.0	30.0	49.0	0.5	0.5
	5	23.6	13.0	19.0	23.0	29.0	37.0	-0.2	0.4
Clay (%)	1	36.4	8.0	30.0	37.0	47.0	66.0	-0.3	-0.1
	2	30.1	10.0	22.0	30.0	35.0	59.0	0.1	0.5
	3	40.8	10.0	30.0	41.0	51.0	82.0	-0.6	0.1
	4	33.1	14.0	22.0	34.0	42.0	63.0	-0.6	0.4
	5	27.5	14.0	18.0	26.0	34.0	52.0	-0.2	0.7

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Appendix 6: Table of descriptive statistics of selected physical characteristics across spatial domains.

Table A6.1. Descriptive statistics for data of selected physical characteristics, comparing yield factors and screen sizes across the six spatial domains.

Physical characteristic	Domain	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Kurtosis	Skewness
				Northern Region					
Yield factor 13	I	84.6	82.6	83.7	84.5	85.3	87.6	0.5	0.6
	II	84.9	82.8	83.8	84.7	85.6	87.8	-0.6	0.5
	III	86.0	83.1	84.7	85.6	86.9	96.5	8.0	2.2
	IV	85.3	68.7	84.5	85.2	86.1	88.8	48.3	-4.9
Yield factor 14	I	85.2	82.6	84.3	85.1	85.8	88.8	0.7	0.7
	II	85.7	83.0	84.2	85.3	86.7	90.0	-0.3	0.7
	III	87.5	83.1	85.0	86.4	88.4	109.3	11.7	2.9
	IV	86.0	69.6	85.0	85.7	86.9	92.8	21.3	-2.1
Yield factor 15	I	88.4	83.7	86.5	88.2	89.9	96.8	1.1	0.9
	II	89.1	83.4	86.1	88.4	90.0	101.1	1.3	1.3
	III	92.7	83.4	87.5	89.7	94.9	140.6	9.3	2.7
	IV	89.4	74.9	86.7	88.4	90.9	107.6	3.0	1.2
Screen size 17	I	72.8	36.5	68.5	72.6	80.2	98.1	1.8	-0.6
	II	70.5	30.1	61.2	72.1	80.9	100.2	0.2	-0.5
	III	61.5	15.7	50.2	62.6	74.4	93.8	-0.1	-0.4
	IV	70.0	26.7	59.9	71.1	80.3	102.4	0.1	-0.4
Screen size 18	I	43.8	8.3	30.8	41.5	52.5	110.0	1.5	1.1
	II	47.3	6.4	28.2	44.3	63.4	113.1	-0.3	0.6
	III	46.1	4.4	21.5	36.1	58.7	164.4	1.7	1.4
	IV	49.1	3.6	26.7	43.3	61.7	163.1	1.9	1.3
				Huila/Tolima Region					
Yield factor 13	V	85.8	82.0	84.8	85.6	86.8	107.2	46.3	5.2
	VI	86.6	69.0	84.8	86.1	87.5	135.0	55.5	5.4
Yield factor 14	V	86.7	82.2	85.2	86.4	87.6	108.5	27.2	3.8
	VI	88.7	70.3	85.3	86.9	89.0	286.4	186.6	12.9
Yield factor 15	V	90.1	82.7	87.1	88.8	92.1	114.6	4.6	1.7
	VI	95.9	72.1	87.3	90.4	94.8	777.8	216.8	14.4
Screen size 17	V	65.6	22.8	55.7	67.9	76.6	106.5	-0.1	-0.3
	VI	58.5	1.6	46.6	61.9	72.3	98.5	-0.4	-0.5
Screen size 18	V	49.3	4.8	23.3	39.2	66.5	173.2	1.3	1.1
	VI	41.2	0.2	15.4	33.4	59.3	138.2	0.2	0.9

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Appendix 7: Table of descriptive statistics of selected biochemical characteristics across spatial domains.

Table A7.1. Descriptive statistics for data of selected biochemical characteristics, compared across the six spatial domains.

Biochemical charateristic	Domain	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Kurtosis	Skewness
			Northern Region						
Chlorogenic. acid	I	6.4	4.4	5.9	6.4	6.8	8.0	0.2	0.0
	II	6.7	5.3	6.2	6.6	7.2	8.2	-0.5	0.3
	III	6.7	5.1	6.3	6.7	7.2	7.8	-0.3	-0.2
	IV	6.5	4.7	6.0	6.5	6.9	8.0	-0.1	0.0
Caffeine	I	1.4	1.2	1.4	1.5	1.5	1.7	-0.2	-0.1
	II	1.4	1.1	1.3	1.3	1.4	1.7	0.3	0.0
	III	1.3	1.1	1.2	1.3	1.4	1.5	-0.4	0.0
	IV	1.4	1.1	1.3	1.4	1.4	1.6	0.2	-0.3
Trigonelline	I	1.0	0.8	0.9	0.9	1.0	1.2	-0.4	0.4
	II	0.9	0.6	0.8	0.9	1.0	1.2	-0.4	0.0
	III	1.0	0.7	0.9	1.0	1.1	1.2	-0.5	-0.1
	IV	0.9	0.7	0.8	0.9	1.0	1.2	-0.2	0.1
Sucrose	I	5.3	4.2	4.9	5.3	5.5	6.6	0.1	0.5
	II	5.2	4.4	4.8	5.1	5.5	6.3	-0.1	0.6
	III	5.3	4.0	4.9	5.1	5.6	6.9	0.2	0.5
	IV	5.3	4.4	5.1	5.3	5.6	6.4	0.0	0.4
			Huila/Tolima Region						
Chlorogenic. acid	V	6.6	5.0	6.2	6.6	7.0	8.0	0.0	-0.2
	VI	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Caffeine	V	1.3	1.1	1.3	1.3	1.4	1.6	0.9	0.0
	VI	1.4	1.1	1.3	1.4	1.4	1.8	0.4	0.3
Trigonelline	V	0.9	0.5	0.8	0.9	0.9	1.1	0.8	0.0
	VI	1.0	0.7	0.9	1.0	1.1	1.4	-0.2	0.3
Sucrose	V	5.7	4.3	5.3	5.6	6.0	7.3	0.0	0.2
	VI	5.0	3.5	4.5	5.0	5.4	6.8	-0.3	0.1

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Appendix 8: Tables of descriptive statistics of organoleptic characteristics across spatial domains.

Table A8.1. Descriptive statistics for data of selected organoleptic characteristics, compared across the four spatial domains in the Northern Region.

Sensory charateristic	Domain	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Kurtosis	Skewness
Fragrance ans aroma	I	5.2	3.5	4.5	5.0	5.5	7.5	0.5	0.4
	II	5.2	2.0	4.5	5.0	6.0	8.5	0.1	0.0
	III	5.2	2.0	4.0	5.0	6.0	8.5	-0.1	0.0
	IV	5.5	2.0	5.0	5.5	6.5	8.5	0.3	-0.4
Flavor	I	5.4	2.0	5.0	5.0	6.0	10.0	2.2	0.5
	II	5.1	1.0	4.0	5.0	6.0	10.0	0.1	0.4
	III	4.9	2.0	4.0	5.0	6.0	10.0	1.1	0.9
	IV	5.3	1.0	5.0	5.5	6.0	10.0	0.4	-0.4
Aftertaste	I	5.3	2.0	4.5	5.0	6.0	10.0	1.8	0.8
	II	5.0	1.0	3.0	5.0	6.0	10.0	-0.3	0.5
	III	4.5	2.0	3.0	5.0	6.0	10.0	0.8	0.7
	IV	5.2	1.0	4.0	5.5	6.0	10.0	0.3	-0.2
Acidity	I	5.4	2.0	5.0	5.0	6.0	10.0	2.5	0.9
	II	5.3	2.0	4.0	5.0	6.0	10.0	1.6	0.9
	III	5.3	2.0	4.0	5.0	6.0	10.0	1.1	0.6
	IV	5.4	1.0	5.0	5.0	6.0	10.0	1.5	0.2
Body	I	5.4	2.0	5.0	5.0	6.0	10.0	2.3	0.1
	II	5.1	2.0	4.0	5.0	6.0	10.0	0.5	0.5
	III	4.9	2.0	3.0	5.0	6.0	10.0	1.1	0.9
	IV	5.3	1.0	5.0	5.5	6.0	10.0	0.8	-0.4
Balance	I	5.4	2.0	5.0	5.3	6.0	10.0	2.5	0.4
	II	5.1	2.0	3.0	5.0	6.0	10.0	-0.1	0.6
	III	4.7	2.0	3.0	5.0	6.0	10.0	0.6	0.7
	IV	5.3	2.0	4.0	5.5	6.5	8.5	-0.4	-0.5
Uniformity	I	6.4	2.0	5.5	6.3	7.0	10.0	1.3	-0.2
	II	7.0	2.0	6.5	7.3	8.0	10.0	1.4	-0.9
	III	6.6	2.0	6.0	6.6	8.0	10.0	0.4	-0.2
	IV	7.0	3.0	6.5	7.0	8.0	10.0	1.4	-0.9
Clean cup	I	5.6	2.0	5.0	5.5	6.0	8.5	0.2	-0.3
	II	5.2	1.0	3.0	5.0	7.0	9.0	-1.1	-0.2
	III	4.8	2.0	3.0	5.0	6.5	8.5	-1.2	0.0
	IV	5.8	1.0	5.0	6.0	7.0	9.0	-0.5	-0.5
Sweetness	I	5.3	2.0	4.5	5.0	6.0	10.0	2.4	0.8
	II	5.2	2.0	4.0	5.0	6.0	10.0	0.2	0.5
	III	5.4	2.0	4.5	5.0	6.5	10.0	0.8	0.4
	IV	5.3	1.0	5.0	5.0	6.0	10.0	1.4	-0.1
Overall	I	5.5	2.0	4.5	5.5	6.5	10.0	0.5	0.3
	II	5.3	1.0	3.5	5.0	7.0	10.0	-0.8	0.1
	III	4.9	2.0	3.0	5.0	6.5	10.0	-0.5	0.4
	IV	5.5	1.0	4.0	6.0	6.5	10.0	-0.4	-0.3
Final score	I	54.8	27.5	50.0	53.3	60.5	80.5	0.5	0.0
	II	53.6	18.5	40.0	52.0	65.5	88.0	-0.8	0.1
	III	51.2	29.0	38.5	51.0	61.5	84.5	-0.6	0.4
	IV	55.5	21.5	49.5	56.8	63.5	83.0	-0.1	-0.3

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Table A8.2. Descriptive statistics for data of selected organoleptic characteristics, compared across the two spatial domains in the Huila/Tolima Region.

Sensory charateristic	Domain	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Kurtosis	Skewness
Fragrance and aroma	V	5.2	1.0	4.0	5.8	6.5	8.0	-0.4	-0.7
	VI	5.7	1.0	5.0	6.0	6.5	8.5	2.0	-1.3
Flavor	V	4.9	1.0	3.0	5.0	6.5	9.0	-1.0	-0.1
	VI	5.7	1.0	5.0	6.0	7.0	9.0	-0.1	-0.5
Aftertaste	V	4.8	1.0	3.0	5.0	6.5	8.5	-1.0	0.0
	VI	5.7	1.0	5.0	6.0	7.0	9.0	-0.2	-0.4
Acidity	V	5.3	1.0	4.0	5.0	7.0	9.0	-0.7	-0.3
	VI	5.9	1.0	5.0	6.0	7.0	9.0	0.2	-0.6
Body	V	5.1	1.0	4.0	5.0	6.5	8.5	-0.8	-0.3
	VI	5.8	1.0	5.0	6.0	7.0	9.0	-0.1	-0.5
Balance	V	5.0	1.0	3.0	5.0	7.0	9.0	-1.1	-0.1
	VI	5.9	1.0	5.0	6.0	7.0	9.0	-0.3	-0.5
Uniformity	V	7.1	3.0	6.5	7.0	8.0	9.0	1.1	-1.2
	VI	7.4	4.0	7.0	7.5	8.0	9.0	2.0	-1.2
Clean cup	V	5.2	1.0	3.0	5.0	7.0	9.0	-1.0	-0.1
	VI	6.1	1.5	5.0	6.0	7.5	9.0	-0.5	-0.4
Sweetness	V	5.1	1.0	4.0	5.0	6.5	8.5	-0.8	-0.3
	VI	5.6	1.0	5.0	6.0	7.0	9.0	0.2	-0.6
Overall	V	5.0	1.0	3.0	5.0	6.5	9.5	-1.1	-0.1
	VI	5.8	1.0	5.0	6.0	7.0	10.0	-0.2	-0.5
Final score	V	52.6	19.0	36.0	53.5	66.5	84.5	-1.0	-0.1
	VI	59.6	18.5	51.5	60.0	71.0	89.5	-0.1	-0.5

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Appendix 9: Tables of descriptive statistics of climatic characteristics across spatial domains.

Table A9.1. Descriptive statistics for data of selected climate characteristics, compared across the four spatial domains in the Northern Region.

Climate charateristic	Domain	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Kurtosis	Skewness
Annual rainfall (mm)	I	2390	2030	2370	2430	2460	2480	3.4	-1.9
	II	2040	1290	2000	2040	2180	2420	1.8	-1.0
	III	1410	1220	1330	1400	1460	1760	1.1	0.8
	IV	2340	920	2220	2350	2530	2980	4.1	-1.3
Annual total evaporation (mm yr^{-1})	I	1350	1120	1290	1330	1390	1600	0.2	0.4
	II	1080	690	1000	1090	1160	1270	1.3	-1.0
	III	990	790	920	990	1050	1230	-0.2	0.3
	IV	960	580	900	970	1030	1160	1.4	-0.7
Dew point temperature (°C)	I	15.4	12.9	14.4	15.0	16.5	19.4	0.0	0.6
	II	14.2	3.5	13.3	14.2	14.9	21.4	8.1	-1.1
	III	14.4	11.3	13.2	14.3	15.3	19.8	1.9	0.9
	IV	13.6	5.4	12.9	13.6	14.3	17.5	7.9	-1.5
Average temperature (°C)	I	20.8	18.1	19.7	20.4	21.7	24.5	-0.3	0.5
	II	19.9	8.8	19.0	19.9	20.8	27.6	7.9	-1.1
	III	19.8	16.4	18.7	19.7	20.9	25.1	1.4	0.8
	IV	19.6	11.3	19.0	19.7	20.4	22.9	9.1	-2.0
Diurnal temperature range (°C)	I	9.9	9.0	9.5	9.9	10.2	10.7	-1.0	-0.1
	II	10.5	9.1	10.3	10.5	10.8	11.6	2.8	-0.7
	III	9.9	9.3	9.7	9.9	10.1	10.4	-0.5	0.0
	IV	10.6	9.1	10.3	10.7	11.0	11.4	0.2	-0.7
Dry months per year	I	4.0	4.0	4.0	4.0	4.0	5.0	58.0	7.6
	II	4.0	3.0	4.0	4.0	4.0	5.0	9.8	-0.5
	III	4.3	1.0	4.0	4.0	5.0	6.0	0.2	-0.1
	IV	1.3	0.0	1.0	1.0	2.0	8.0	9.9	2.6
Annual average cloud frequency (%)	I	79.2	74.3	77.9	79.0	81.1	81.9	0.1	-0.7
	II	86.0	80.3	85.1	86.5	87.2	87.6	2.6	-1.3
	III	92.7	88.4	89.4	93.8	95.7	96.8	-1.7	-0.2
	IV	96.1	91.9	95.4	96.0	97.0	97.9	1.1	-1.0
Solar radiation (MJ m^{-2} d^{-1})	I	22.9	19.0	22.0	23.0	24.0	25.0	-0.3	-0.3
	II	22.1	17.0	21.0	22.0	24.0	25.0	0.0	-0.7
	III	23.1	16.0	23.0	23.0	24.0	25.0	6.1	-1.8
	IV	24.2	18.0	24.0	25.0	25.0	25.0	10.2	-2.8

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Table A9.2. Descriptive statistics for data of selected climate characteristics, compared across the two spatial domains in the Huila/Tolima Region.

Climate charateristic	Domain	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Kurtosis	Skewness
Annual rainfall (mm)	V	2290	2050	2160	2280	2420	2520	-1.2	0.0
	VI	1740	1340	1650	1740	1810	2040	0.2	-0.1
Annual total evaporation (mm yr^{-1})	V	880	740	820	880	920	1020	-0.8	0.2
	VI	850	690	800	840	900	1070	-0.5	0.4
Dew point temperature (°C)	V	14.2	11.1	13.6	14.1	14.8	17.9	1.7	0.3
	VI	14.1	10.2	13.2	14.1	14.8	18.9	1.4	0.5
Average temperature (°C)	V	20.4	17.2	19.7	20.3	21.0	24.1	1.3	0.3
	VI	19.7	16.0	18.9	19.6	20.4	25.0	2.0	0.6
Diurnal temperature range (°C)	V	10.5	9.9	10.3	10.4	10.6	10.8	-0.5	-0.2
	VI	10.0	8.9	9.5	10.0	10.5	11.0	-1.1	-0.3
Dry months per year	V	1.4	1.0	1.0	1.0	2.0	2.0	-1.9	0.3
	VI	1.5	0.0	1.0	2.0	2.0	4.0	-0.2	0.0
Annual average cloud frequency (%)	V	95.1	94.0	95.0	95.0	95.0	96.0	1.3	0.2
	VI	96.0	95.0	95.0	96.0	97.0	97.0	-1.3	-0.1
Solar radiation ($MJ\ m^{-2}\ d^{-1}$)	V	23.4	19.0	23.0	23.0	24.0	25.0	0.4	-0.4
	VI	23.5	19.0	23.0	24.0	25.0	25.0	-0.3	-0.7

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Appendix 10: Tables of descriptive statistics of soil characteristics across spatial domains.

Table A10.1. Descriptive statistics for data of soil macronutrients, aluminium and cation exchange capacity, compared across the four spatial domains in the Northern Region.

Soil charateristic	Domain	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Kurtosis	Skewness
pH	I	5.2	4.6	5.0	5.1	5.4	6.7	5.0	1.5
	II	5.1	3.9	4.4	5.2	5.7	7.1	-0.5	0.2
	III	4.9	3.7	4.6	4.9	5.3	6.6	-0.1	0.3
	IV	4.6	3.8	4.2	4.5	4.8	7.4	5.6	2.1
N (%)	I	0.3	0.1	0.2	0.3	0.3	0.5	0.4	0.5
	II	0.3	0.2	0.3	0.3	0.4	0.7	0.9	0.8
	III	0.3	0.1	0.2	0.3	0.3	0.6	1.5	1.0
	IV	0.4	0.1	0.3	0.3	0.4	0.6	0.0	0.3
Organic matter (%)	I	6.3	1.2	5.0	6.2	7.3	13.9	1.9	0.9
	II	7.9	3.2	5.6	7.1	9.8	20.5	2.7	1.3
	III	6.4	2.8	4.6	6.1	7.4	15.5	2.7	1.4
	IV	8.5	1.8	6.2	8.0	10.1	18.5	0.9	0.9
K (cmol/kg)	I	0.2	0.1	0.1	0.2	0.2	0.7	5.8	2.2
	II	0.3	0.1	0.2	0.3	0.4	1.1	4.9	1.8
	III	0.3	0.1	0.2	0.2	0.4	1.9	11.7	3.3
	IV	0.3	0.1	0.2	0.2	0.3	1.5	15.4	3.3
Ca (cmol/kg)	I	5.6	1.5	3.8	5.2	6.4	17.8	5.7	1.8
	II	6.8	0.1	1.7	5.7	10.4	19.7	-0.5	0.7
	III	3.7	0.2	1.1	3.0	5.2	20.2	6.0	1.9
	IV	3.7	0.2	0.8	1.9	5.2	23.6	3.9	1.9
Mg (cmol/kg)g	I	1.0	0.2	0.6	0.9	1.4	3.6	4.2	1.5
	II	1.4	0.1	0.4	1.1	2.2	5.8	1.8	1.2
	III	1.0	0.1	0.4	0.8	1.2	9.1	25.8	4.3
	IV	0.8	0.1	0.2	0.4	0.9	10.0	22.3	4.2
Al (cmol/kg)	I	0.7	0.1	0.3	0.6	1.0	2.1	0.7	1.0
	II	1.4	0.1	0.2	0.4	2.3	6.5	1.7	1.6
	III	1.6	0.1	0.3	1.2	2.5	8.2	2.9	1.5
	IV	3.4	0.1	1.3	3.1	5.0	13.0	1.3	1.0
CEC (cmol/kg)	I	15.3	7.0	12.0	15.0	18.0	34.0	5.4	1.3
	II	16.9	10.0	13.0	16.0	20.0	28.0	-0.5	0.5
	III	13.3	4.0	11.0	12.5	16.0	27.0	0.5	0.5
	IV	18.4	6.0	14.0	18.0	22.0	39.0	0.5	0.6

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Table A10.2. Descriptive statistics for data of soil phosphorus, iron, micronutrients, and soil components, compared across the four spatial domains in the Northern Region.

Soil charateristic	Domain	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Kurtosis	Skewness
P (mg/kg)	I	18.8	1.0	3.0	6.0	15.0	193.0	10.6	3.2
	II	41.0	1.0	5.0	13.0	43.0	272.0	4.3	2.1
	III	20.7	1.0	5.0	7.0	18.0	180.0	8.3	2.8
	IV	17.9	1.0	3.0	5.0	10.0	291.0	19.1	4.3
Fe (mg/kg)	I	265	80	182	249	315	612	0.8	0.9
	II	449	51	231	322	637	1609	2.1	1.5
	III	458	100	260	358	555	1591	2.6	1.6
	IV	640	70	341	619	885	1645	-0.6	0.5
Mn (mg/kg)	I	47.2	5.0	27.0	40.0	68.0	136.0	1.1	1.0
	II	115.6	4.0	35.0	112.0	155.0	458.0	3.0	1.4
	III	76.5	1.0	27.0	58.0	100.0	498.0	10.3	2.6
	IV	27.1	0.0	4.0	9.5	24.0	385.0	25.5	4.4
Zn (mg/kg)	I	1.4	0.0	1.0	1.0	2.0	7.0	6.9	2.1
	II	4.2	0.0	1.0	2.0	7.0	26.0	8.2	2.5
	III	4.9	1.0	2.0	3.0	6.0	30.0	9.7	2.7
	IV	5.1	0.0	2.0	3.0	5.0	59.0	23.7	4.5
Cu (mg/kg)	I	3.5	0.0	2.0	3.0	5.0	12.0	2.4	1.2
	II	11.4	0.0	2.0	3.0	5.0	475.0	60.5	7.8
	III	3.2	0.0	1.0	2.0	5.0	17.0	6.6	2.1
	IV	3.2	0.0	1.0	2.0	4.0	55.0	58.0	7.1
Sand (%)	I	27.0	8.0	22.0	28.0	32.0	45.0	0.6	-0.1
	II	33.6	13.0	23.0	33.0	43.0	60.0	-0.9	0.2
	III	31.5	5.0	26.0	31.0	37.0	51.0	0.5	-0.2
	IV	45.0	14.0	37.0	45.0	52.0	69.0	-0.3	0.0
Silt (%)	I	27.5	17.0	23.0	27.0	31.0	41.0	-0.4	0.2
	II	31.6	13.0	25.0	30.0	38.0	49.0	-0.9	0.1
	III	22.7	12.0	17.0	22.0	27.0	41.0	0.0	0.7
	IV	26.3	13.0	21.0	25.0	30.0	51.0	0.5	0.6
Clay(%)	I	45.9	24.0	39.0	47.0	53.0	63.0	-0.8	-0.1
	II	35.0	14.0	27.0	34.0	40.0	67.0	0.6	0.6
	III	45.7	22.0	36.0	46.0	54.0	82.0	-0.1	0.3
	IV	28.7	8.0	19.0	26.0	34.0	66.0	0.4	0.7

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Table A10.3. Descriptive statistics for data of selected soil characteristics, compared across the two spatial domains in the Huila/Tolima Region.

Soil charateristic	Domain	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Kurtosis	Skewness
pH	V	5.2	4.1	4.7	5.1	5.7	7.2	0.0	0.5
	VI	5.0	3.7	4.5	5.0	5.5	6.4	-0.9	0.1
N (%)	V	0.3	0.2	0.3	0.3	0.4	0.7	1.8	1.1
	VI	0.3	0.1	0.2	0.3	0.4	0.7	1.6	0.9
Organic matter (%)	V	7.7	3.2	5.6	7.1	8.9	22.8	5.0	1.7
	VI	7.6	2.0	5.4	7.2	9.2	23.7	4.2	1.5
K (cmol/kg)	V	0.4	0.1	0.2	0.3	0.4	14.8	118.8	10.8
	VI	0.3	0.1	0.2	0.3	0.4	1.3	4.6	1.8
Ca (cmol/kg)	V	6.2	0.2	1.5	5.0	9.6	30.0	1.9	1.1
	VI	5.9	0.2	2.4	5.2	8.2	20.7	0.9	1.0
Mg (cmol/kg)g	V	1.8	0.1	0.6	1.5	2.5	8.7	3.9	1.5
	VI	1.5	0.2	0.8	1.3	2.0	5.7	2.6	1.4
Al (cmol/kg)	V	2.1	0.1	0.2	0.8	3.6	9.8	1.1	1.4
	VI	1.6	0.1	0.2	0.7	2.4	14.0	7.9	2.4
CEC (cmol/kg)	V	17.8	7.0	15.0	18.0	20.0	34.0	0.9	0.5
	VI	16.5	6.0	12.0	16.0	20.0	44.0	3.4	1.2
P (mg/kg)	V	33.0	1.0	3.0	5.5	15.0	1364.0	89.0	8.9
	VI	50.6	1.0	9.0	20.0	64.0	810.0	33.0	5.0
Fe (mg/kg)	V	382	67	216	298	470	1480	3.9	1.8
	VI	479	114	226	364	666	1474	0.6	1.1
Mn (mg/kg)	V	67.3	4.0	25.0	53.0	98.0	238.0	1.4	1.3
	VI	57.0	4.0	21.0	41.0	70.0	353.0	7.6	2.3
Zn (mg/kg)	V	3.4	0.0	1.0	2.0	4.0	65.0	74.7	7.9
	VI	5.2	0.0	2.0	4.0	6.0	32.0	9.7	2.7
Cu (mg/kg)	V	2.5	0.0	1.0	2.0	3.0	18.0	11.7	2.6
	VI	3.2	0.0	1.0	2.0	4.0	32.0	29.0	4.6
Sand (%)	V	38.4	18.0	29.0	37.5	48.0	63.0	-0.9	0.1
	VI	30.8	12.0	21.0	29.0	38.0	65.0	-0.1	0.7
Silt (%)	V	28.4	18.0	25.0	29.0	31.0	44.0	0.2	0.4
	VI	24.5	12.0	19.0	22.0	29.0	49.0	0.6	1.1
Clay (%)	V	33.5	10.0	24.0	32.0	43.0	68.0	-0.7	0.3
	VI	44.7	14.0	32.0	46.0	58.0	71.0	-1.1	-0.2

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Appendix 11: Table of descriptive statistics of selected topographic characteristics across spatial domains.

Table A11.1. Descriptive statistics for data of selected topographic characteristics, compared across the six spatial domains.

Topographic charateristic	Domain	Mean	Min[1]	LQ[2]	Med[3]	UQ[4]	Max[5]	Kurtosis	Skewness
				Northern Region					
Aspect (°)	I	229	2	153	253	304	356	-0.5	-0.7
	II	221	5	150	264	325	359	-1.1	-0.5
	III	175	7	90	176	268	351	-1.3	0.0
	IV	193	1	88	215	296	357	-1.4	-0.2
Slope (°)	I	20.4	1.0	12.6	20.9	27.5	36.4	-0.6	-0.2
	II	20.4	1.1	15.0	20.0	25.5	40.1	0.3	0.0
	III	19.1	5.3	13.7	19.2	24.1	35.3	-0.6	0.0
	IV	11.9	0.6	7.1	10.1	14.7	42.4	3.3	1.6
Altitude (masl)	I	1250	660	1050	1270	1420	1810	-0.3	-0.2
	II	1430	90	1230	1430	1630	3360	7.5	1.1
	III	1410	470	1280	1440	1550	1840	3.3	-1.3
	IV	1600	670	1460	1600	1730	3100	7.6	1.3
				Huila/Tolima Region					
Aspect (°)	V	174	8	95	153	271	359	-1.3	0.2
	VI	209	1	122	240	295	359	-1.0	-0.5
Slope (°)	V	20.2	1.4	13.8	20.4	26.1	38.5	-0.6	-0.1
	VI	18.7	3.7	11.2	17.8	25.5	42.4	-0.9	0.3
Altitude (masl)	V	1600	860	1490	1610	1730	2240	1.9	-0.4
	VI	1590	680	1480	1590	1720	2020	1.7	-0.6

[1] Min = minimum; [2] LQ = lower quartile; [3] Med = median; [4] UQ = upper quartile; [5] Max = maximum.

Appendix 12: CaNaSTA analysis of sensory characteristics by spatial domain.

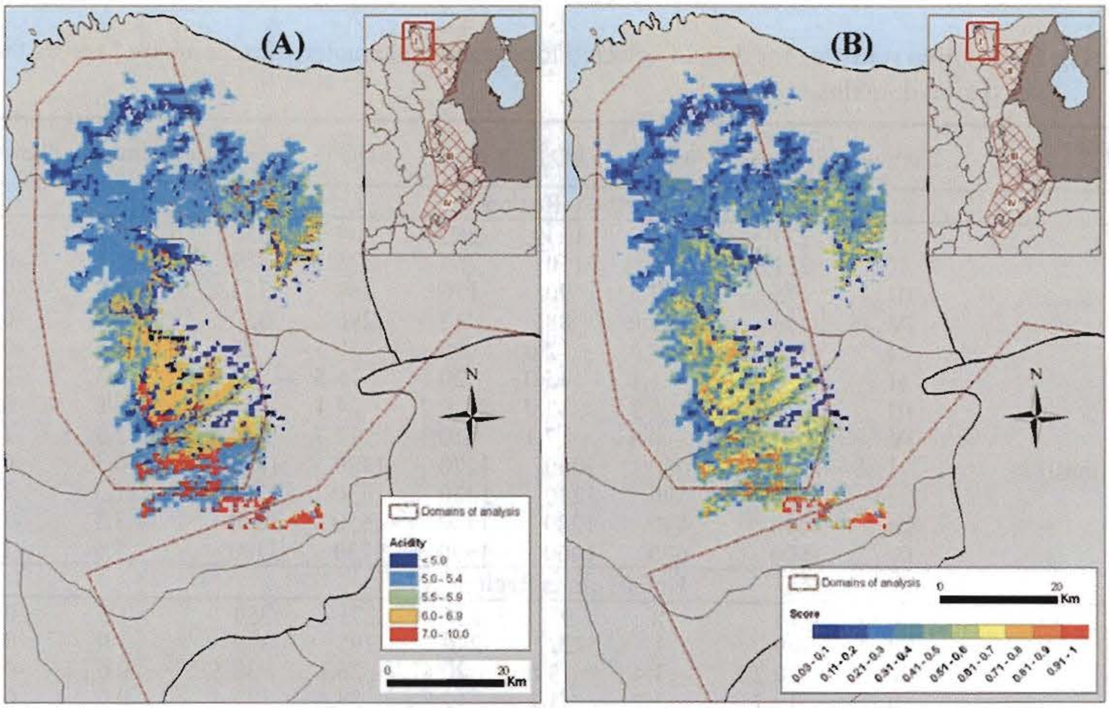


Figure A12. 1. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Acidity in domain I of the Northern Region.

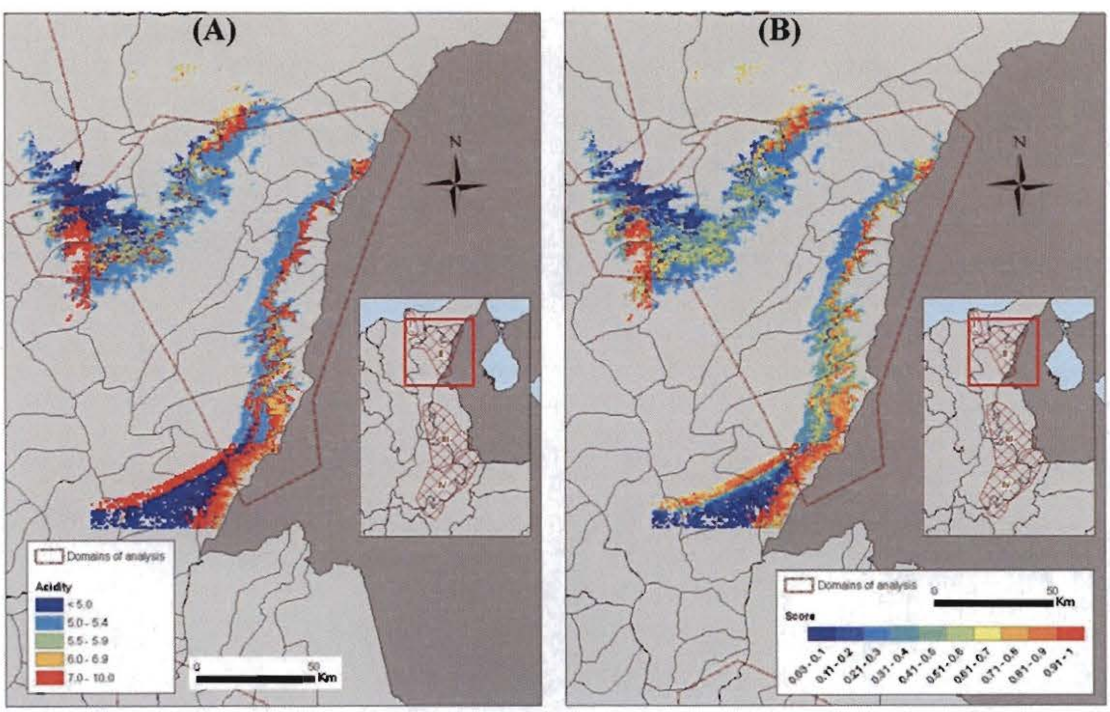


Figure A12. 2. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Acidity in spatial domain II of the Northern Region.

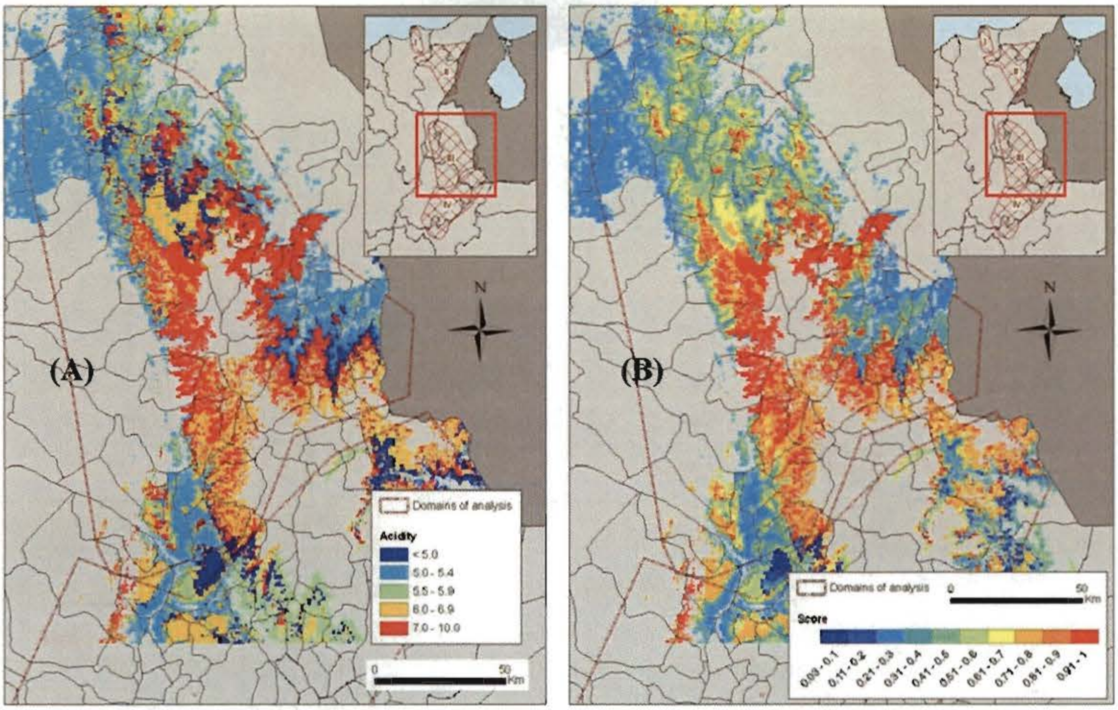


Figure A12. 3. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Acidity in spatial domain III of the Northern Region.

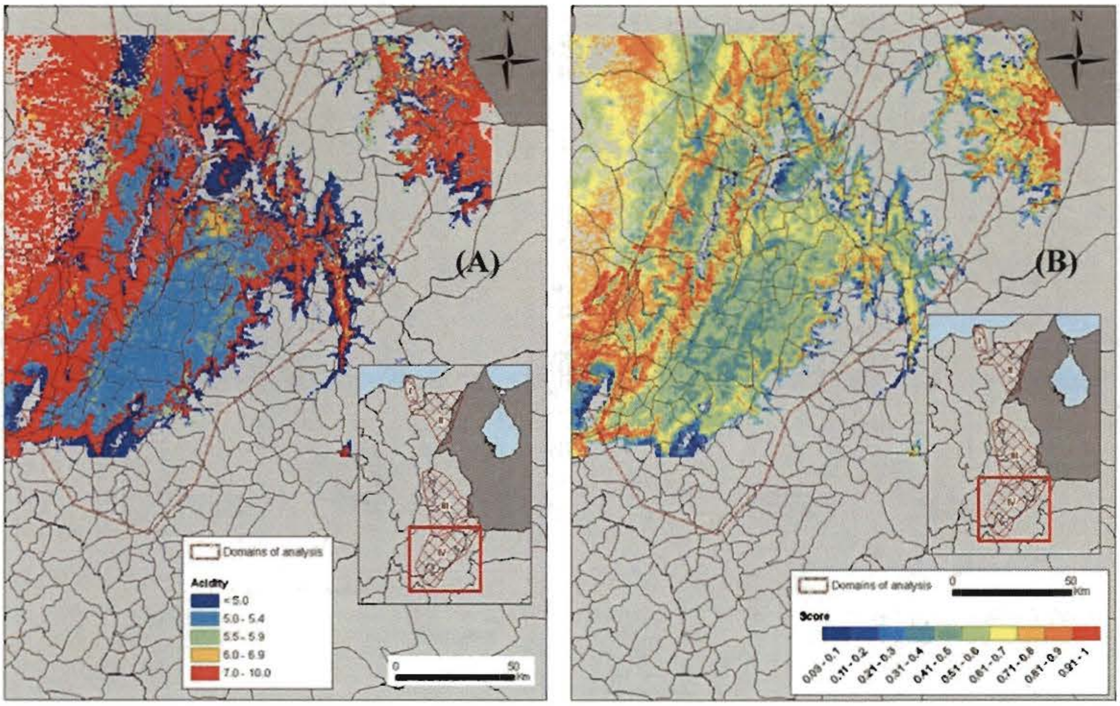


Figure A12. 4. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Acidity in spatial domain IV of the Northern Region.

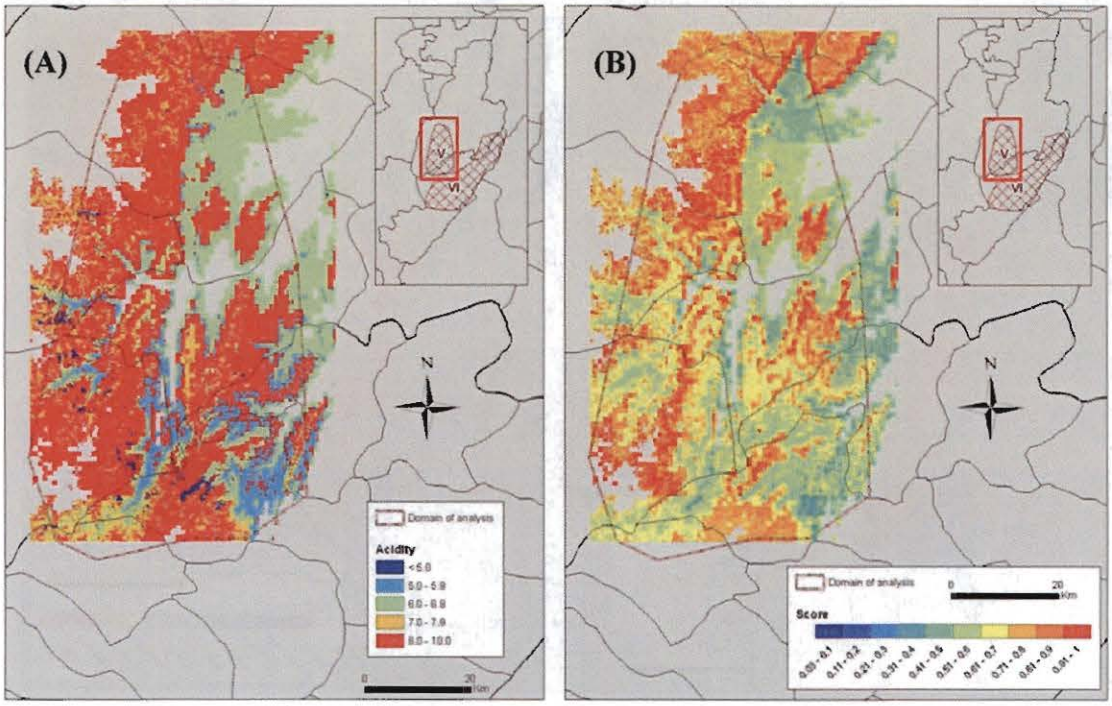


Figure A12. 5. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Acidity in spatial domain V of the Huila/Tolima Region.

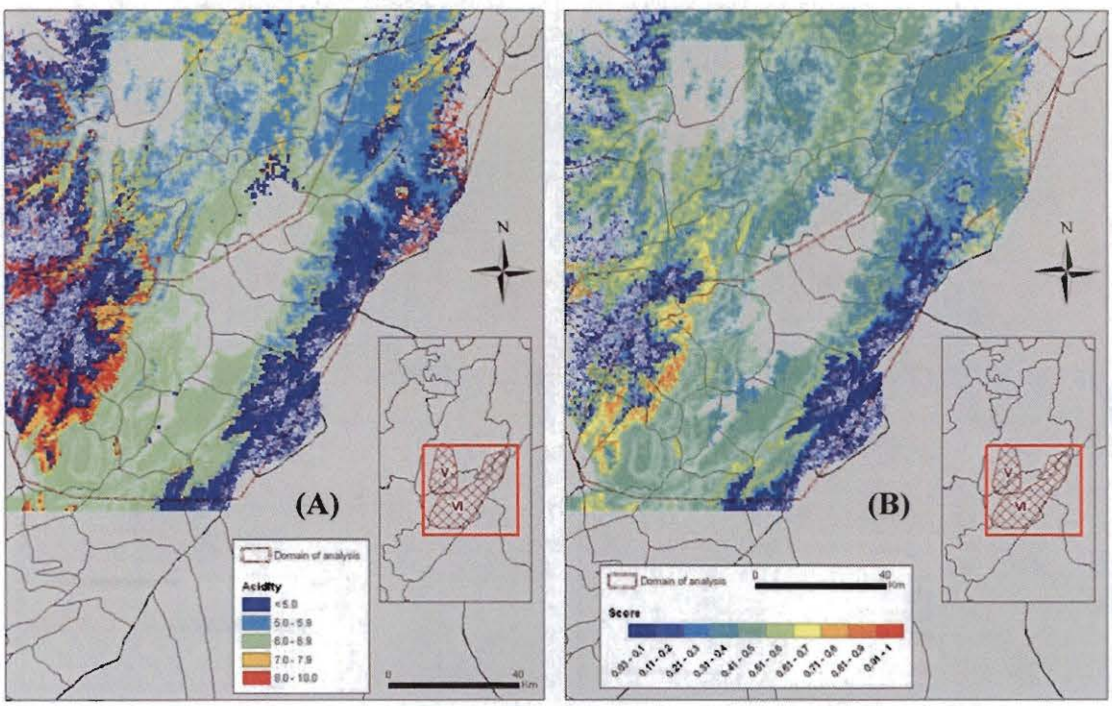


Figure A12. 6. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Acidity in spatial domain VI of the Huila/Tolima Region.

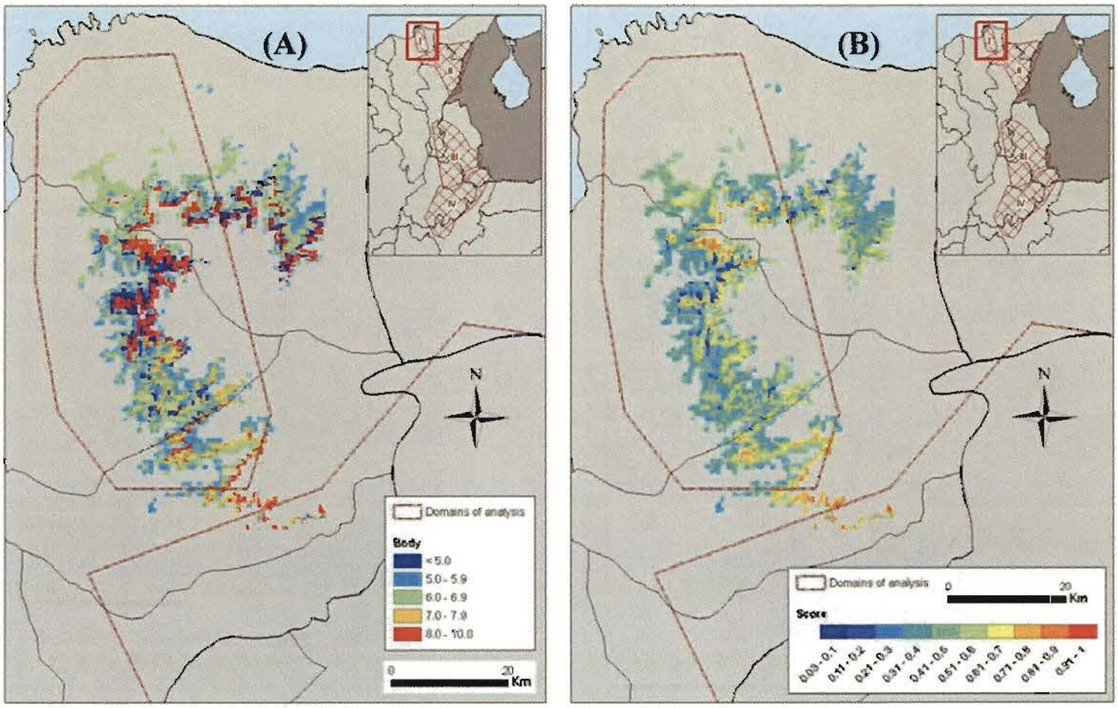


Figure A12. 7. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Body in domain I of the Northern Region.

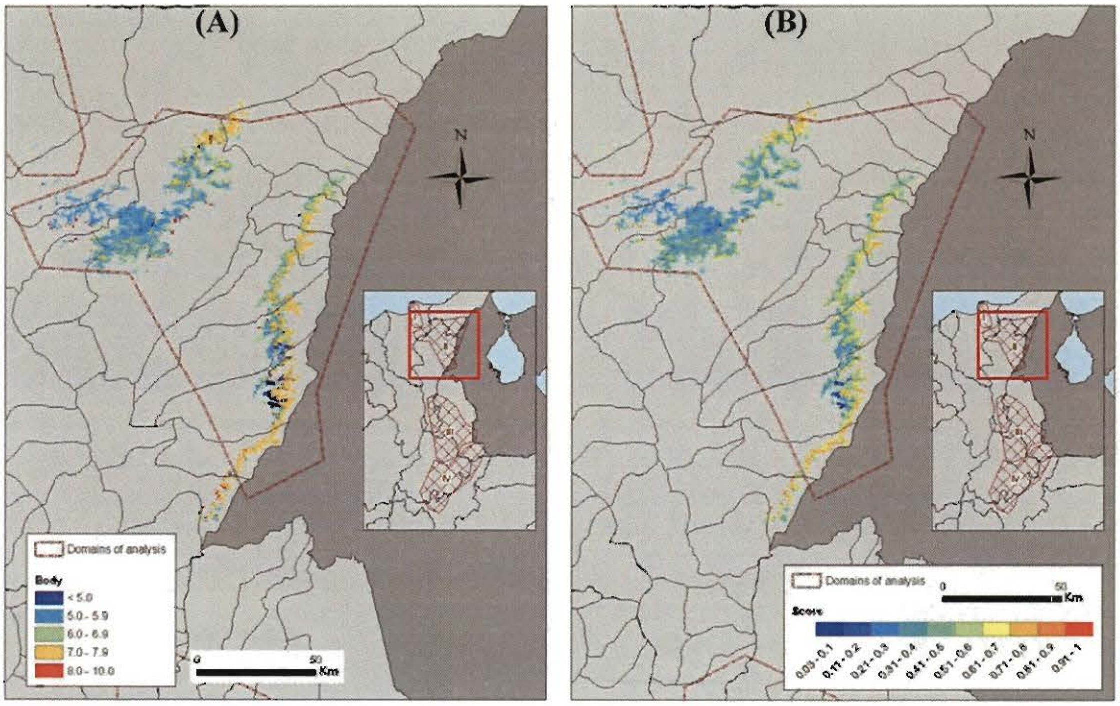


Figure A12. 8. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Body in spatial domain II of the Northern Region.

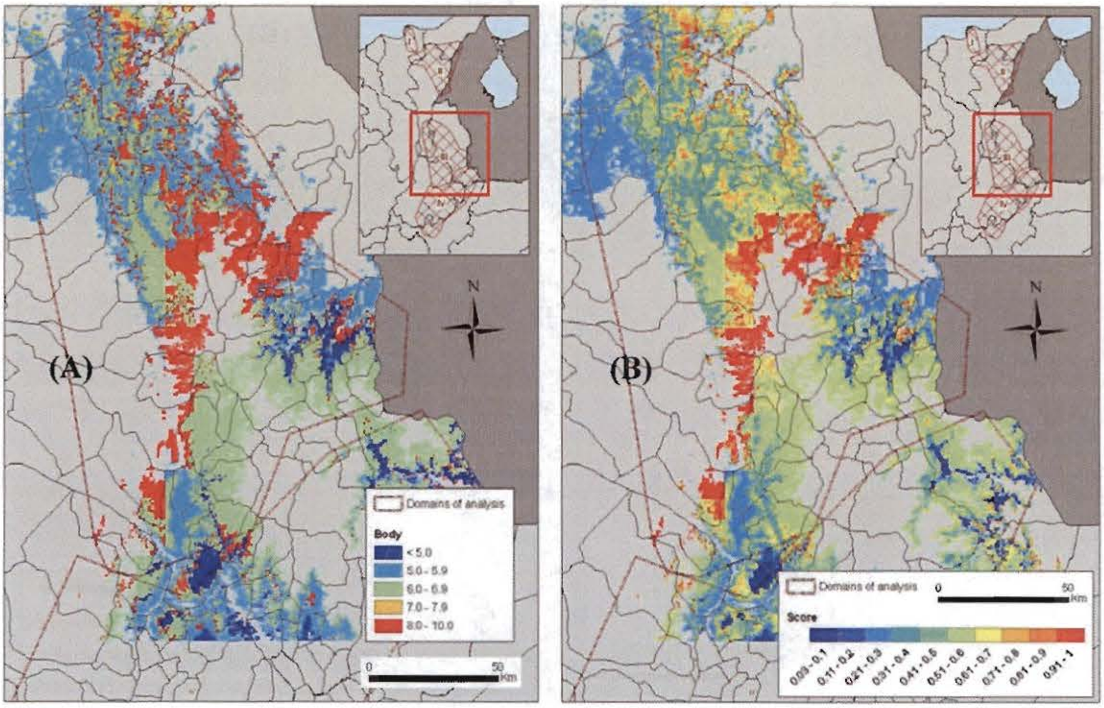


Figure A12. 9. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Body in spatial domain III of the Northern Region.

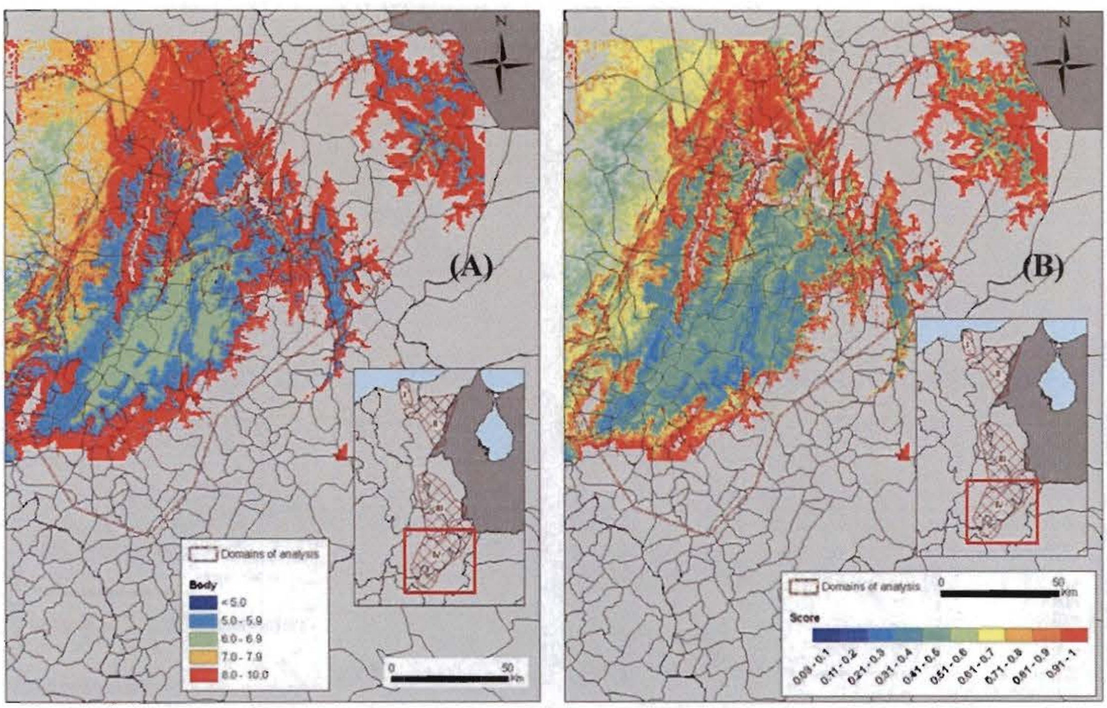


Figure A12. 10. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Body in spatial domain IV of the Northern Region.

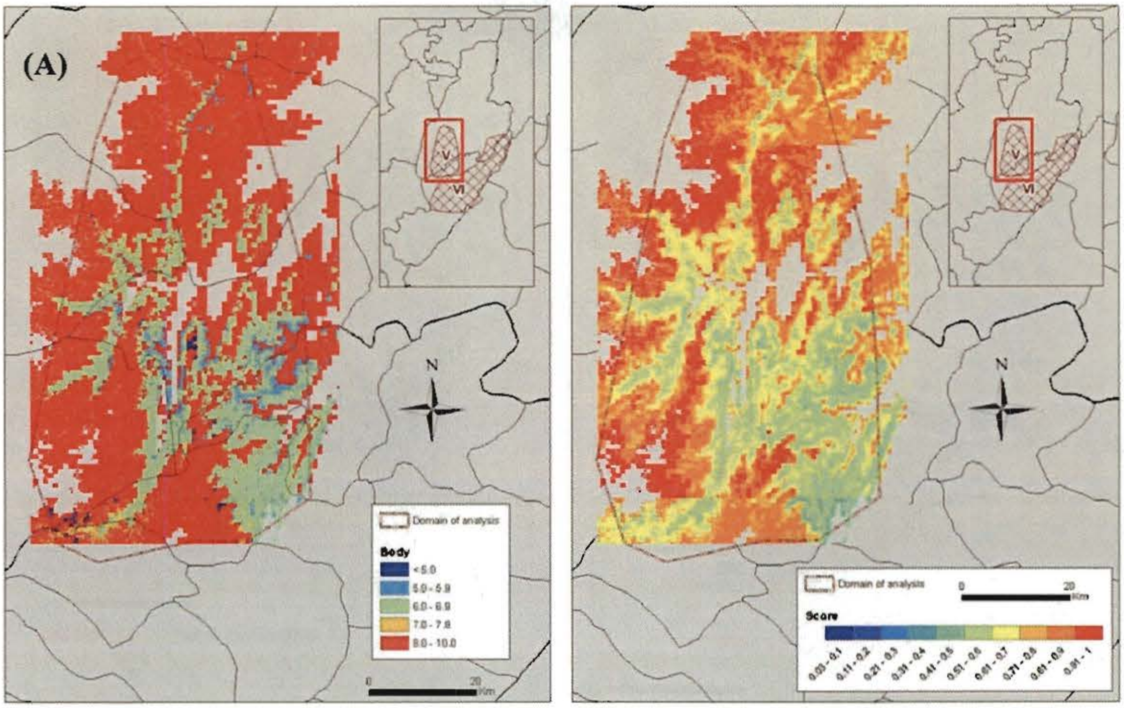


Figure A12. 11. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Body in spatial domain V of the Huila/Tolima Region.

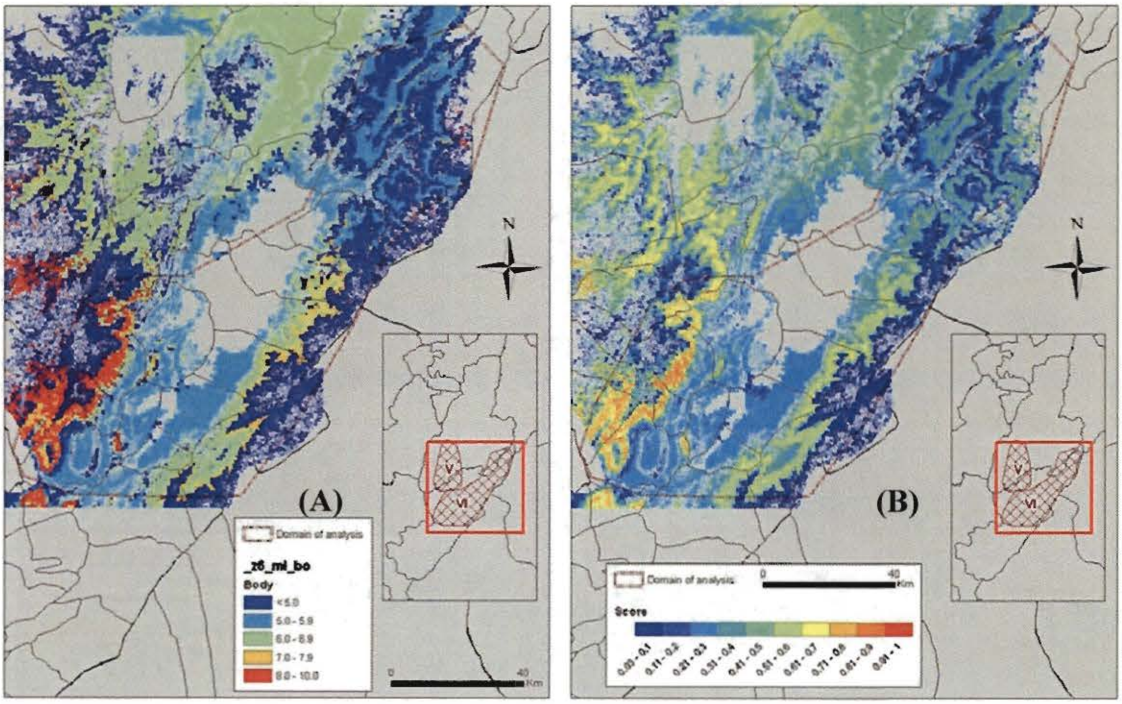


Figure A12. 12. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Body in spatial domain VI of the Huila/Tolima Region.

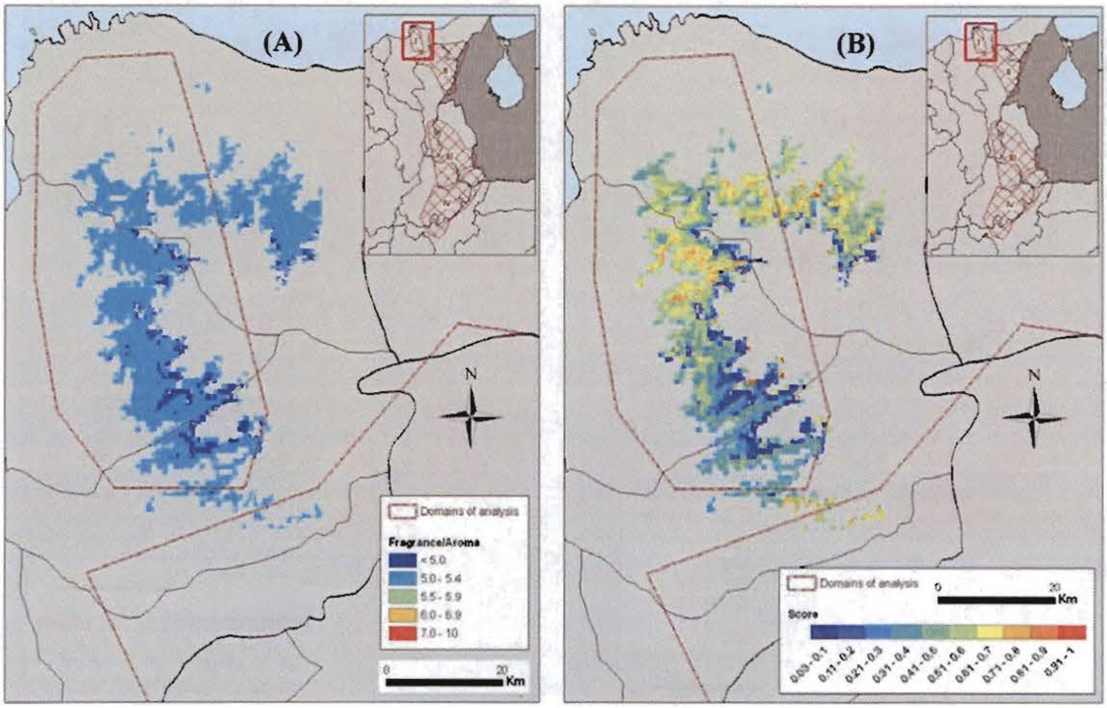


Figure A12. 13. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Fragrance and aroma in domain I of the Northern Region.

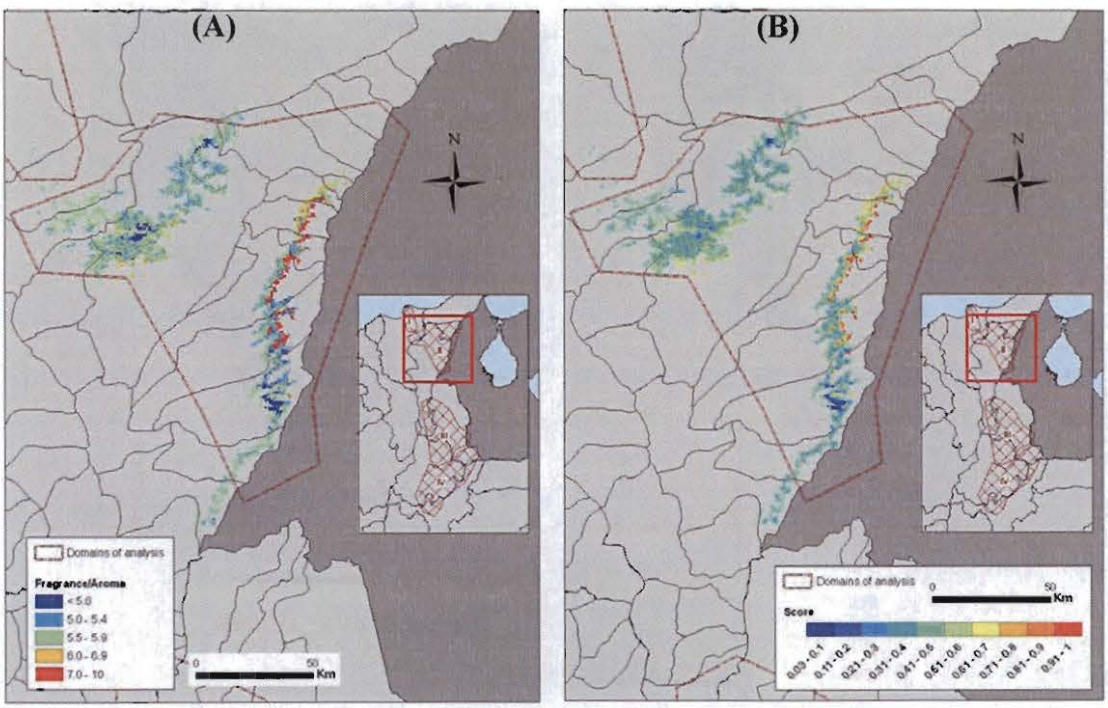


Figure A12. 14. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Fragrance and aroma in spatial domain II of the Northern Region.

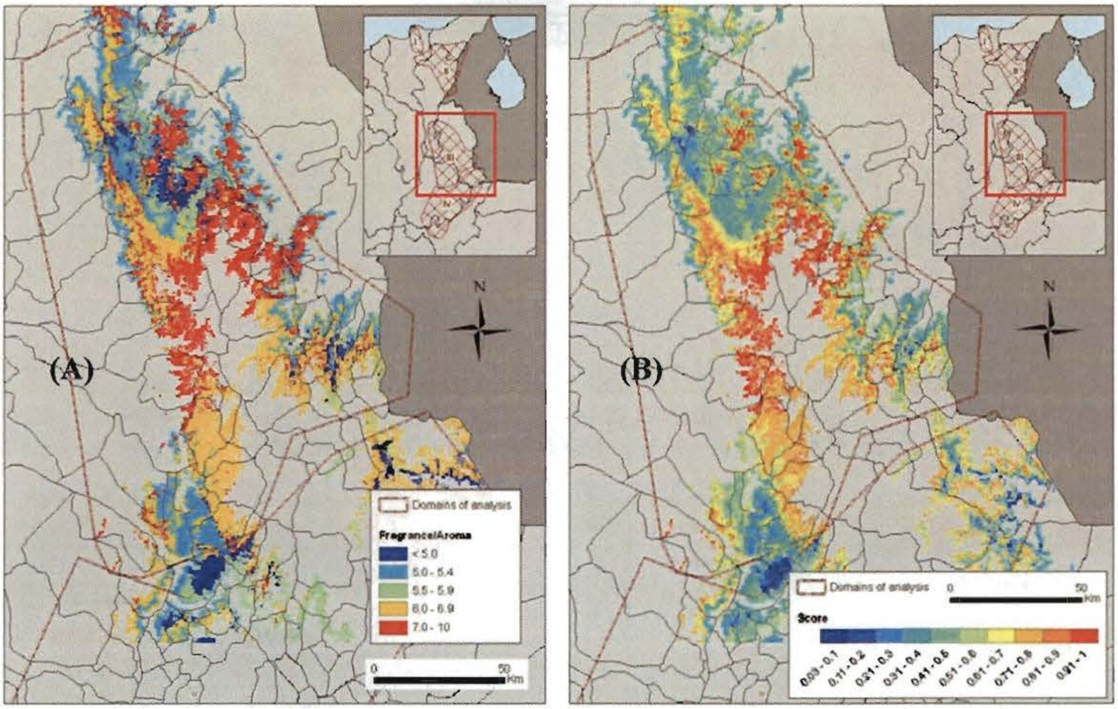


Figure A12. 15. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Fragrance and aroma in spatial domain III of the Northern Region.

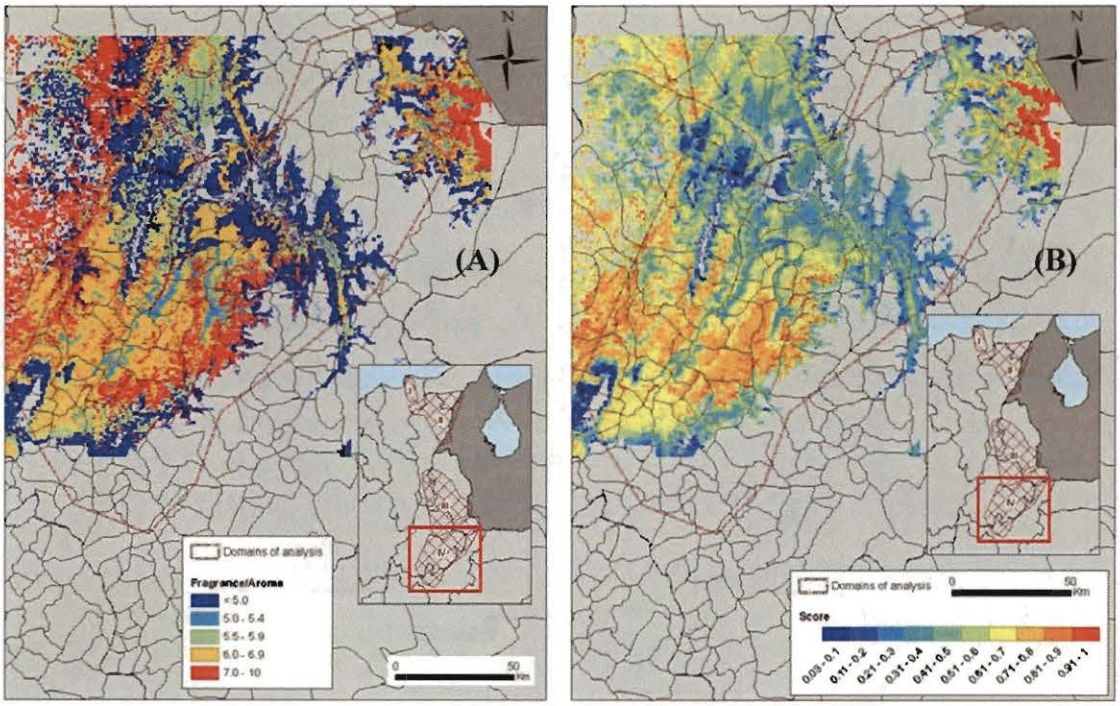


Figure A12. 16. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Fragrance and aroma in spatial domain IV of the Northern Region.

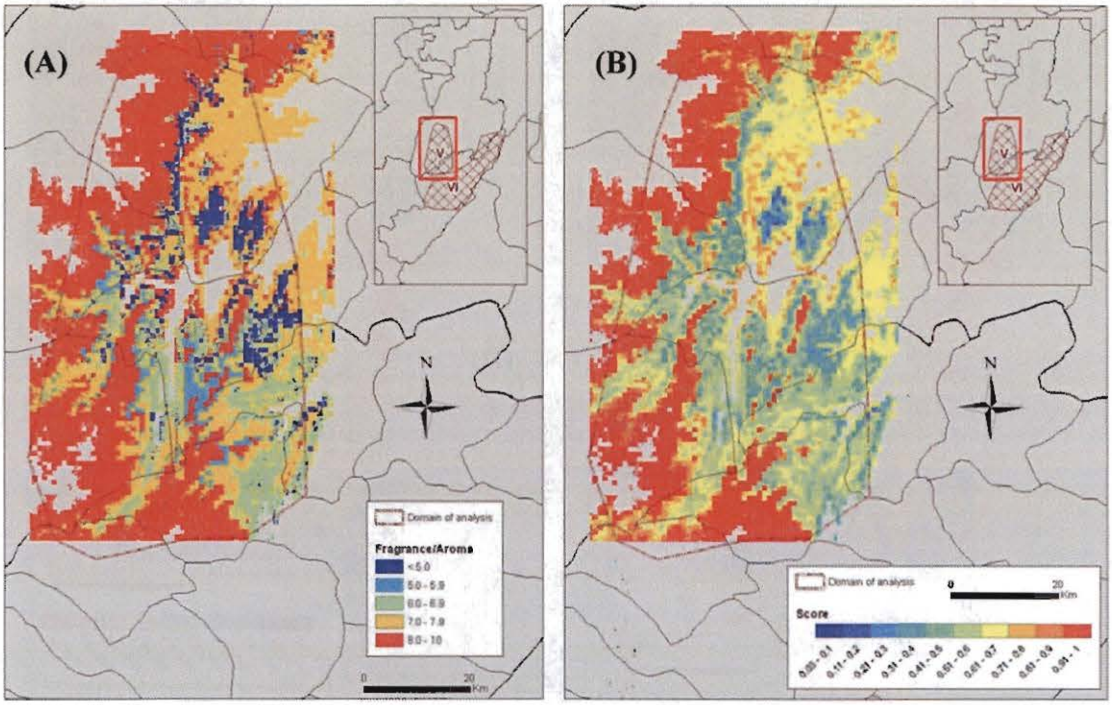


Figure A12. 17. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Fragrance and aroma in spatial domain V of the Huila/Tolima Region.

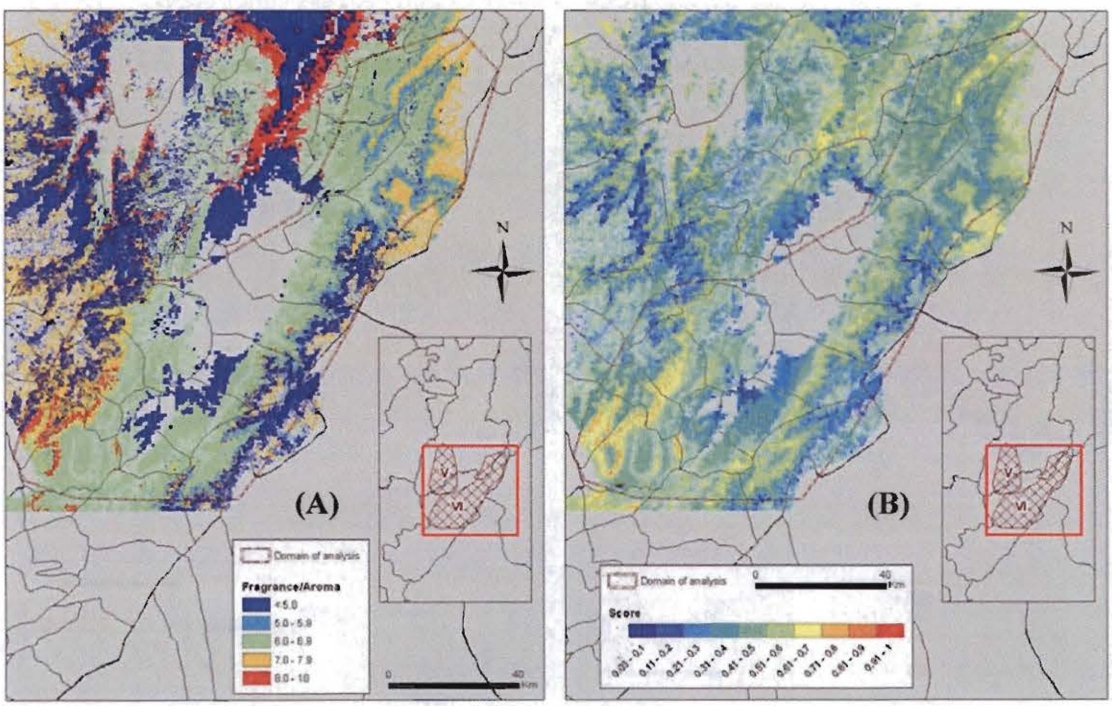


Figure A12. 18. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Fragrance and aroma in spatial domain VI of the Huila/Tolima Region.

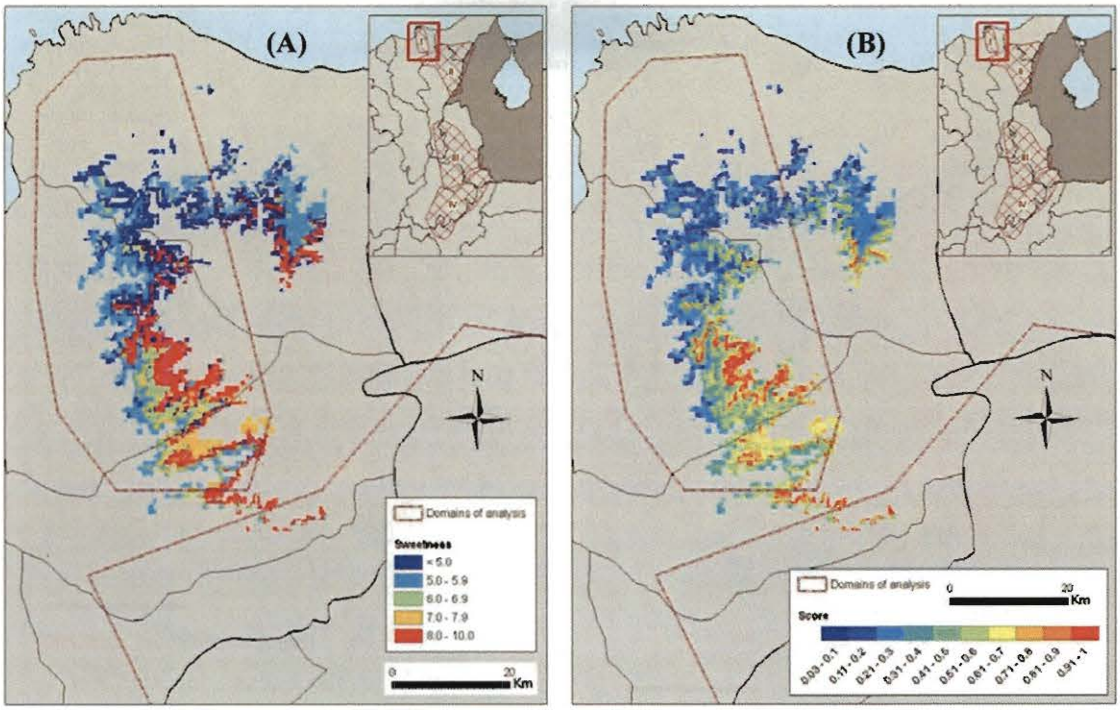


Figure A12. 19. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Sweetness in domain I of the Northern Region.

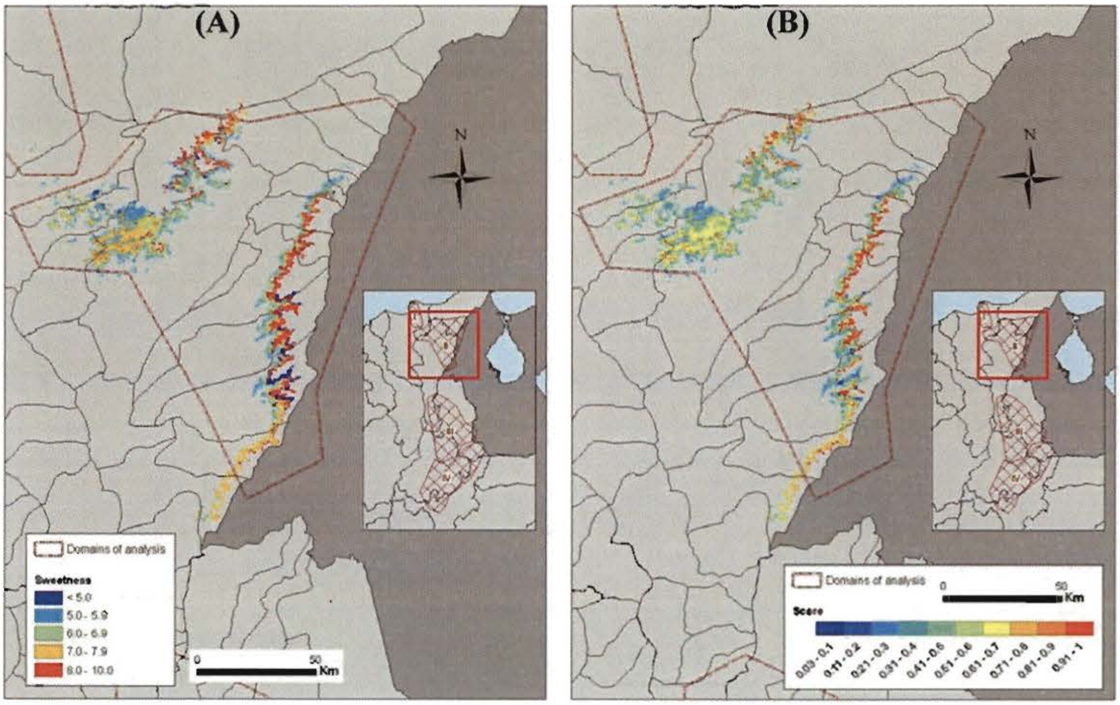


Figure A12. 20. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Sweetness in spatial domain II of the Northern Region.

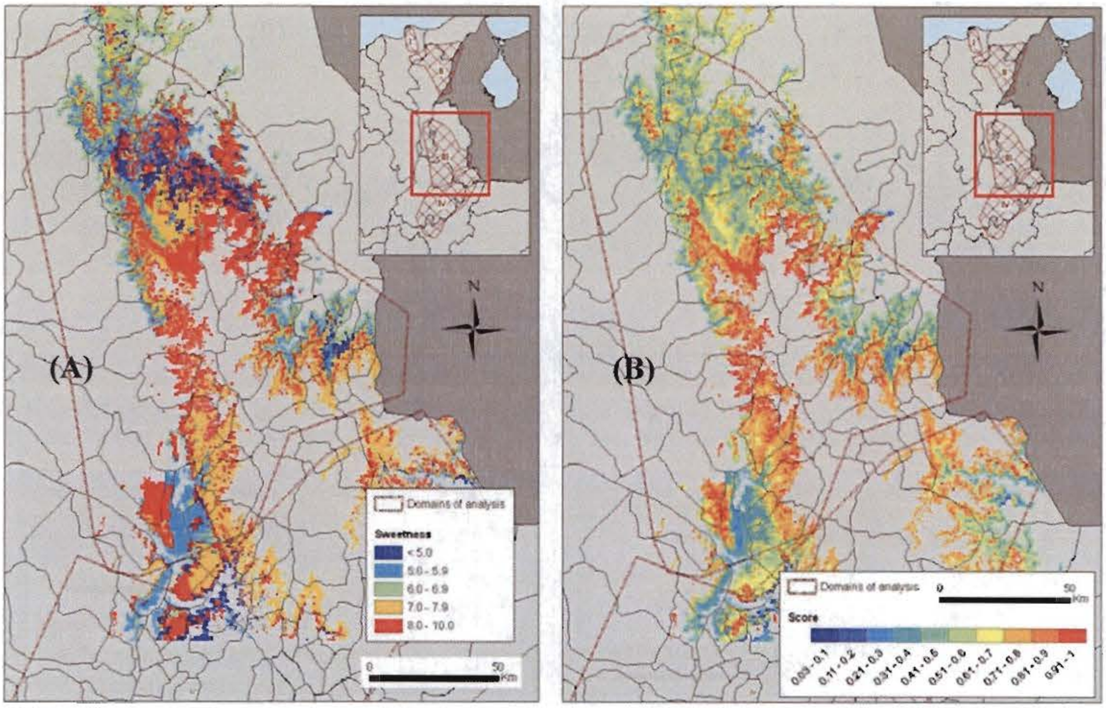


Figure A12. 21. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Sweetness in spatial domain III of the Northern Region.

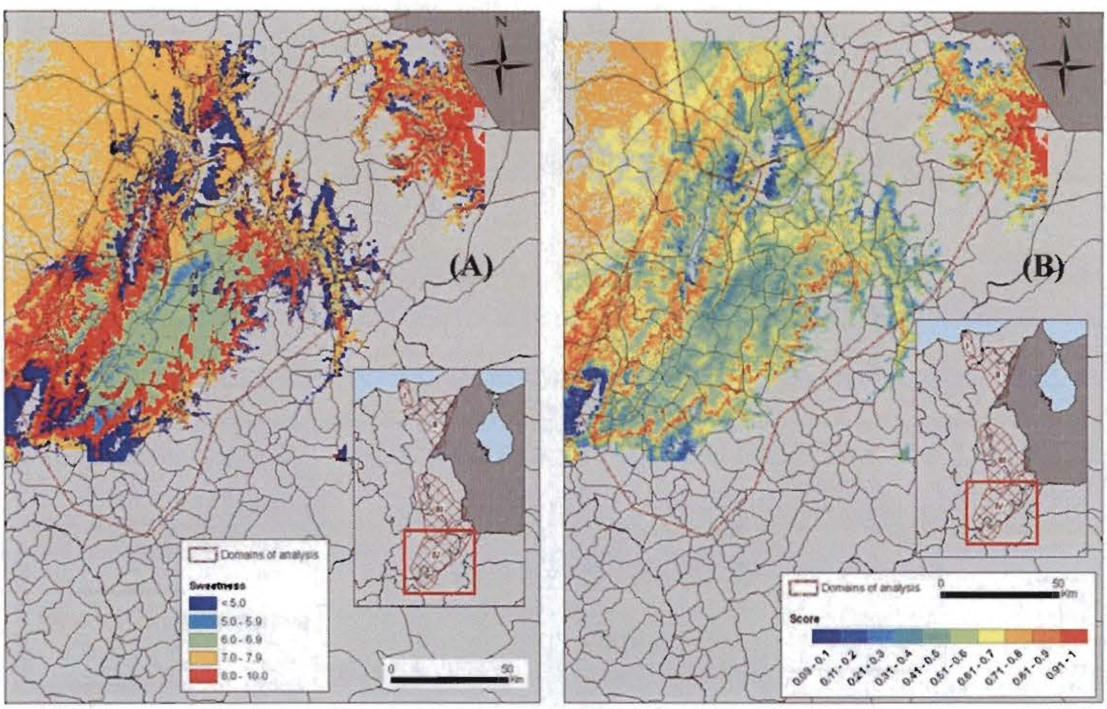


Figure A12. 22. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Sweetness in spatial domain IV of the Northern Region.

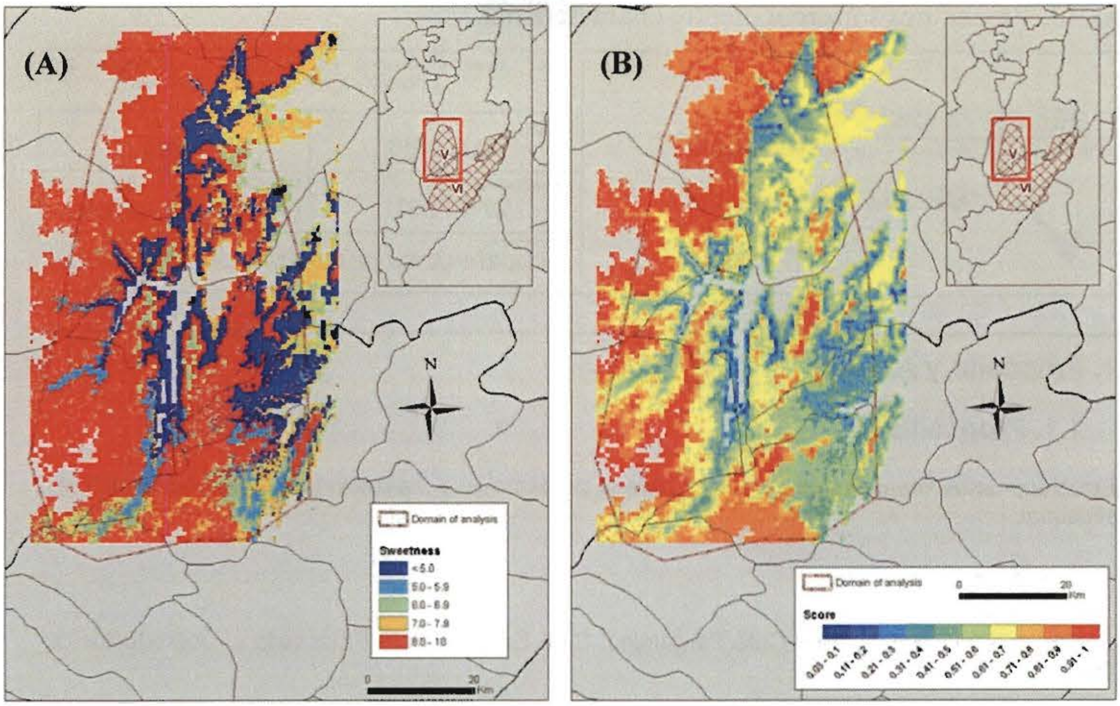


Figure A12. 23. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Sweetness in spatial domain V of the Huila/Tolima Region.

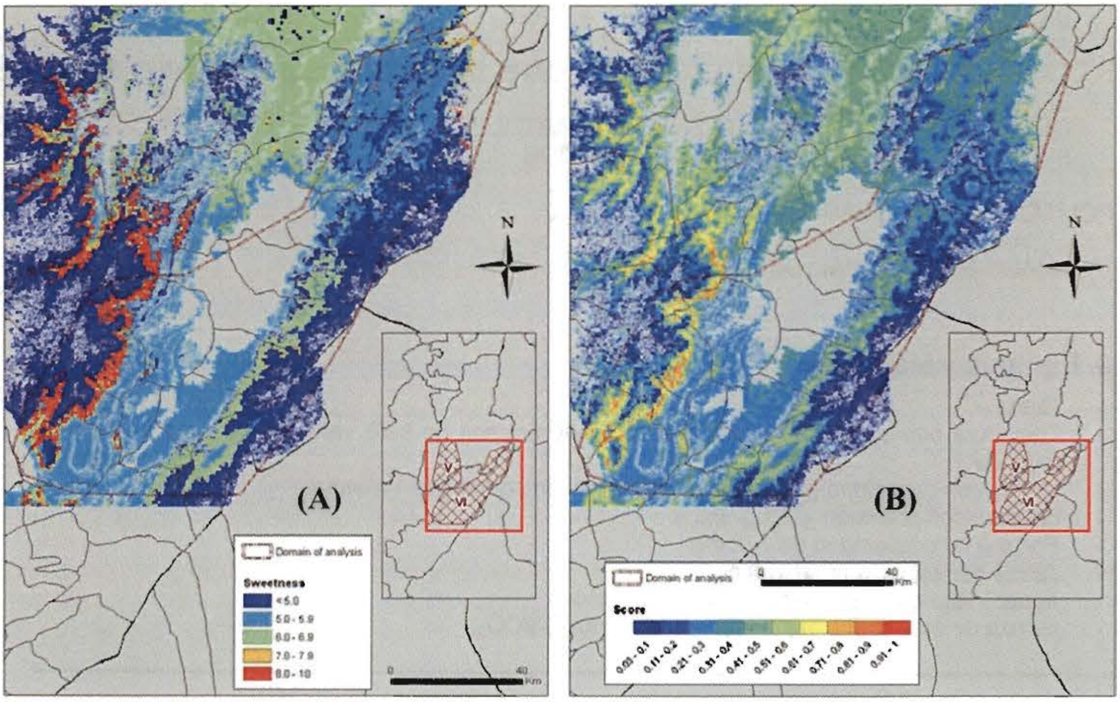


Figure A12. 24. CaNaSTA determination of Bayesian (A) most likely membership in a specific response class, and (B) suitability score of the sensory characteristic Sweetness in spatial domain VI of the Huila/Tolima Region.

Appendix 13: Procedures for organoleptic characteristics

almacafé almacenes generales de depósito de café s.a. Oficina de Calidad de Café	Título : CAFÉ PERGAMINO – CAFÉ EXCELSO – CAFÉ PROCESADO - COPRODUCTOS: ANÁLISIS SENSORIAL	
	Código : MLC-I-010	Página : 1 de 5
	Fecha Emisión : 27-04-2001	Revisión : 4 Fecha : 24-10-2006
	Responsable: COORDINADOR DE EVALUACIÓN Y CONTROL	

1. Propósito y Alcance

1.1. Propósito

Establecer unos lineamientos generales para la preparación y realización de la prueba de análisis sensorial.

1.2. Alcance

Oficina de Calidad de Café – Café Pergamino, Café Excelso, Café Procesado y, Coproductos de Trilla.

2. Referencias

REGISTROS RELACIONADOS

CAFÉ PERGAMINO – CAFÉ EXCELSO – CAFÉ PROCESADO: ANÁLISIS SENSORIAL BÁSICO	MLC-R-012
CAFÉ PERGAMINO – CAFÉ EXCELSO – CAFÉ PROCESADO: ANÁLISIS SENSORIAL COMPLETO	MLC-R-013
CAFÉ PERGAMINO: ANÁLISIS SENSORIAL PERFILES	MLC-R-014
SPECIALTY COFFEES: SENSORY EVALUATION	MLC-R-015

POLÍTICAS RELACIONADAS

ANÁLISIS SENSORIAL: DEFINICIONES	MRN-C-015

3. Equipos y Materiales

- Greca
- Tostadora con capacidad de tostar hasta 500 gramos de café verde con un máximo de 12 minutos
- Molino con capacidad para moler 100 gramos en no más de 1 minuto
- Balanza con precisión $\pm$ 0,1 gramos
- Pocillos con capacidad de 192 ml
- Jarras con capacidad de 480 ml
- Agua: - libre de cloro u otros sabores extraños
- dureza de muy suave hasta media (3,2 mmol/l $CaCO_3$)

	NOMBRE	CARGO	FECHA (dd/mm/aaaa)	FIRMA
Revisó:	Armando Cortés Z.	E. Eval. y Asignación	24-10-2006	[illegible]
Aprobó:	D. Rodrigo Alarcón S.	C. Eval. y Control	24-10-2006	[illegible]

4. Descripción de Actividades

El Analista de Calidades y/o el Técnico de Operación Logística y/o el Auxiliar II de Operación Logística y/o el Ayudante de Operaciones y/o Técnico de Evaluación y Control:

Una vez se tenga las muestras analíticas para análisis sensorial de café pergamino en almendra despasillada, de café excelso, de coproducto ó de café procesado, según el caso, se procede de la siguiente forma:

4.1. Relaciona la(s) muestra(s) en registro "Relación de Muestras para Análisis Sensorial – Resultados" MLC-R-045. Todas las pruebas rutinarias serán pruebas cerradas. Es decir, cada muestra estará codificada, para evitar errores por sesgo.

4.2. Toma aproximadamente 120 gramos de café de la muestra, utilizando la sonda de la tostadora.

4.2.1. Si el café es procesado, toma una muestra por cada caja o bolsa y sigue con el paso 4.10.

4.3. Introduce la muestra en el tambor de la tostadora cuando la máquina haya alcanzado una temperatura entre 200°C – 240°C.

4.4. Deja tostar de 7 a 12 minutos aproximadamente hasta alcanzar el grado definido de tostión.

4.5. Retira la muestra del tambor, después de completar el ciclo de tostión,

4.6. Descarga la muestra en las bandejas perforadas de enfriamiento.

4.6.1. La temperatura de los granos deberá descender a 37°C o menos.

4.7. Traslada los granos a una bandeja plástica, retirando la película plateada.

4.8. Purga el molino con aproximadamente 20 gramos del café tostado.

4.8.1. Repite este proceso entre muestras.

4.9. Muele aproximadamente 40 gramos de la muestra tostada, por cada preparación de taza en jarra y 20 gramos por cada preparación en pocillo.

4.9.1. Para café pergamino al recibo / control prepara cuatro (4) pocillos.

4.9.2. Para café pergamino optimización / reevaluación prepara ocho (8) pocillos.

4.9.3. Para café excelso optimización prepara dieciséis (16) jarras.

4.9.4. Para café excelso en origen prepara cuatro (4) pocillos.

4.9.5. Para café excelso en puerto prepara cuatro (4) pocillos.

4.9.6. Para café procesado prepara un (1) jarra.

4.9.7. Para cataciones de otros programas (como cafés especiales y atención de clientes), el número de pocillos a preparar es libre, según necesidad.

4.10. Desecha el sobrante de la muestra.

4.10.1. Si la catación es para análisis sensorial completo (perfilación, programas especiales, atención clientes, etc.) deja la muestra sobrante, no molida, en la bandeja plástica.

4.11. Pesa 33.5 gramos (7% peso/volumen) ó 13.9 gramos (7% peso/volumen) ó 28.5 gramos (5% peso/volumen), del café molido, según el caso.

4.11.1. Si el recipiente es una jarra y la catación es de café pergamino ó de café excelso, pesa 33.5 gramos.

4.11.2. Si el recipiente es un pocillo y la catación es de café pergamino ó de café excelso, pesa 13.9 gramos.

4.11.3. Si el recipiente es una jarra y el café a evaluar es procesado, pesa 28.5 gramos.

4.12. Deposita la muestra tostada y molida en el recipiente de catación. Los recipientes empleados para la evaluación sensorial (jarras o pocillos), deben estar libres de olores extraños y fisuras o roturas.

4.13. Adiciona al recipiente, agua a punto de ebullición, hasta la mitad de su capacidad.

4.13.1. Si el recipiente utilizado es una jarra, coloque la tapa inmediatamente después de adicionar agua, en caso de que se requiera evaluación del aroma. Para esto introduce la nariz y percibe el aroma de la muestra.

4.14. Agita la mezcla café-agua con una cuchara metálica, enjuagándola entre muestra y muestra (romper taza).

4.14.1. Para el "rompimiento de taza", se empleará una cuchara limpia y libre de olores o sabores extraños.

4.14.2. Se limpiará la cuchara empleada en el "rompimiento de taza" entre cada muestra presentada en una sesión. La limpieza de la cuchara se realizará con agua caliente depositada en un recipiente destinado para tal fin.

4.14.3. El agua empleada para la limpieza de la cuchara se debe cambiar cada 16 recipientes, independientemente de si corresponden a jarras / pocillos ó una / varias muestras.

4.14.4. Cuando se encuentre una sensación clara y definidamente fuerte y defectuosa en una de las muestras presentadas y a la cual se esta "rompiendo taza", se debe cambiar el agua de lavado de la cuchara que se esta empleando para tal fin.

4.15. Completa el volumen con agua hirviendo y deja sedimentar por lo menos cinco (5) minutos.

4.15.1. Si el recipiente utilizado es una jarra, se podrá servir en un pocillo parte de la infusión preparada.

4.15.2. Cuando sea necesario retirar el "sobrenadante ó espuma" en las muestras presentadas (cuando éstas se presentan en pocillo) se debe emplear dos recipientes: uno para depositar allí el "sobrenadante ó espuma" y otro para lavar entre cada recipiente donde se presentan las muestras, la cuchara con la cual se esta retirando el "sobrenadante ó espuma".

El Evaluador / Juez / Catador:

4.16. Efectúa la prueba de análisis sensorial, según Manual de Referencias y Normas "Análisis Sensorial: Definiciones" MRN-C-015.

4.16.1. Realiza la prueba vistiendo una blusa de laboratorio.

4.16.2. Se abstendrá de emplear en el momento de realizar las prueba, lociones, perfumes, labiales o cualquiera otro elemento de uso personal que afecte al gusto y/o olfato.

4.16.3. Mantendrá la adecuada y necesaria concentración para el eficiente desarrollo de la prueba.

4.17. Registra los resultados de la prueba de manera escrita en los formatos correspondientes a cada prueba realizada y los entregará al final de cada sesión de catación al líder del panel correspondiente.

4.17.1. Si la evaluación se realiza en laboratorios de puerto y/o regionales, utiliza el registro "Café Pergamino – Café Excelso – Café Procesado: Análisis Básico" MLC-R-012. Para evaluaciones detalladas utiliza "Café Pergamino – Café Excelso – Café Procesado: Análisis Sensorial Completo" MLC-R-013.

4.17.2. Si la evaluación se efectúa en laboratorio central, utiliza el registro "Café Pergamino – Café Excelso – Café Procesado: Análisis Sensorial Básico" MLC-R-012.

4.17.3. Si la evaluación es de perfiles, utiliza el registro "Café Pergamino: Análisis Sensorial Perfiles " MLC-R-014.

4.17.4. Si la evaluación se efectúa a cafés especiales, utiliza el registro "Specialty Coffees: Sensory Evaluation" MLC-R-015.

4.18. Dispone del sobrante del análisis realizado, que se encuentra en la "escupidera" empleada.

El Líder de Panel y/o Responsable del Programa:

4.19. Analiza resultado de panel y toma decisión según corresponda.

El Técnico de Operación Logística y/o el Auxiliar II de Operación Logística y/o el Ayudante de Operaciones y/o Técnico de Evaluación y Control:

4.20. Relaciona resultados de catación en registro ""Relación de Muestras para Análisis Sensorial – Resultados-" MLC-R-045.

NOTAS

1. **Se debe mantener la "sala de catación" libre de olores extraños ó ajenos al producto en análisis, al momento de la prueba.**

2. **Se debe mantener la "sala de catación" libre de ruidos extraños ó ajenos a la misma, al momento de la prueba.**

3. **En casos especiales (Pedido de Cliente) la relación peso/volumen (café/agua), puede variar.**

4. **El Evaluador puede a su criterio utilizar jarras o pocillos, en los casos en que se indica utilización de pocillos.**

5. **El número de pocillos a preparar expresado en este instructivo representa el mínimo, y en cualquier caso se pueden preparar más.**

6. **Los elementos empleados para las pruebas de análisis Sensorial (Jarras y pocillos) deben quedar limpios, por lo menos, antes de cada fin de semana laboral.**

FIN DEL PROCEDIMIENTO