Geographical Analyses to Explore Interactions between Inherent Coffee Quality and Production Environment

LÄDERACH, PETER*; VAAST, PHILIPPE/***; OBERTHÜR, THOMAS*; RACHEL, O'BRIEN****; LARA ESTRADA, LEONEL DEMOCRITO***; NELSON, ANDY *****

*International Center for Tropical Agriculture (CIAT), Cali, Colombia, South America, Recta Cali Palmira, KM 18, GIS & Land Use Project.
*Corresponding author: p.laderach@cgiar.org
**Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), France
*** Centro Agronómico Tropical de Investigación y Enseñaza (CATIE), Costa Rica
****Charles Sturt University, Australia
*****Joint Research Centre of the European commission, TP 440, Via Enrico Fermi1, I-21020 Ispra (Va), Italy

Abstract

In recent years, research has been focusing on the interactions of coffee cup quality and its production environment. Nevertheless, there has been a lack of methods and tools to extrapolate the findings of mainly controlled experiments to wider geographic areas. We present Spatial Decision Support (SDS) elements based on Bayesian Statistics and Geographically Weighted Regression (GWR) methodologies to identify niches, map quality response attributes, and determine their interactions with the environment. To do so, we use two case studies. With the data set of Cauca, Colombia, we introduce the niche concept and test the predictive approaches. With the data set of Nicaragua, we demonstrate the value of spatial analyses assessing the variability of coffee quality response variables and their determining environmental factors.

Résumé

Récemment, la recherche sur le café s'est concentrée sur les interactions entre la qualité en tasse du café et l'environnement. Cependant, il manquait des méthodes et des outils pour extrapoler les résultats de la plupart des expériences contrôlées à des aires géographiques plus étendues. Nous présentons un outil basé sur des méthodologies de statistiques de Bayes et de Régression Géographiquement Pondérée (GWR) pour cartographier les caractéristiques de qualité du café et pour identifier leurs interactions avec l'environnement. Pour cela, on utilise deux cas d'étude. Avec les données du département du Cauca au sud de la Colombie, nous introduisons le concept de niche et testons l'outil présenté. Avec les données du Nicaragua, nous démontrons la valeur des analyses spatiales pour évaluer la variabilité des caractéristiques de qualité du café et des facteurs environnementaux qui les dirigent.

Introduction

Usually, due to high resource inputs, agricultural research is conducted in few experimental sites, findings and generated knowledge is thereafter rolled out and applied to wide areas without taking into account the changes of the environment over space. The development of tools and methodologies for extrapolating findings that are site and environment specific is required. Spatial decision support tools can help to extrapolate findings and identify niches where a specific coffee trait is likely to be found. Niches are clusters of sites with environmental characteristics that favor product qualities of similar nature. Spatial decision support tools give insights on interactions between species performance and the environment. We use two case studies to demonstrate the utility of geographical analysis. With a data set of Cauca, Southern Colombia, we introduce the niche concept and test a Spatial Decision Support tool and with a data set from Nicaragua, we demonstrate the value of spatial analyses to assess the variability of coffee quality response variables and their determining environmental factors.

Methodology

A Spatial Decision Support (SDS) tool, that is, a software tool based in Geographical Information Science (GIS) to assist users in decision-making was developed. The tool, CaNaSTA (Crop Niche Selection in Tropical Agriculture) employs Bayesian Statistics. 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.

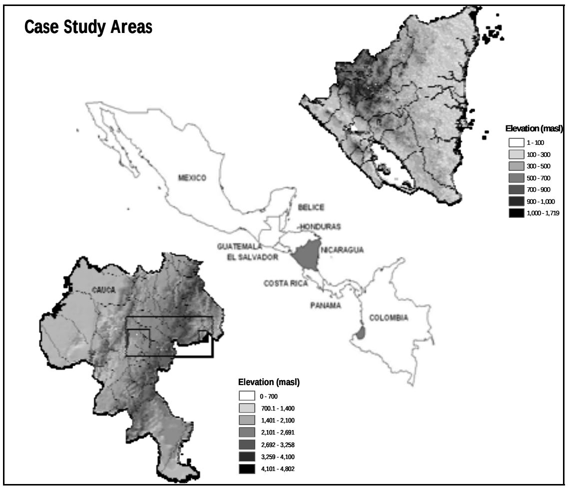
The CaNaSTA algorithm (O'Brien, 2004) creates conditional probability tables of all predictor variables against response variable categories. In the case of coffee, predictor variables include climate and topographic factors, and the response variable can contain sensorial, physical or biochemical quality attributes. The primary model output is a probability distribution at each location. 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, an indicator of reliability (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; a significance indicator is used to compare the importance of the factors. These factors can be either quality enhancing or reducing, and help with the analysis of specific conditions required for specific coffees.

Geographically weighted regression (GWR) assumes that "...the relationships between variables measured at different locations might not be constant over space" (Fotheringham, 2002). We use GWR to illustrate that the interaction between the environmental factors and coffee quality attributes vary in space. GWR is a spatial

statistical method employing moving windows for regression (Fotheringham, 2002) used to describe the spatial variability of coffee quality attributes.
Evidence data used for the predictions and analyses consisted of generated climatic factors with a resolution of 1 km and terrain attributes with a resolution of 90m. Climate layers were generated using WorldClim (Hijmans et al., 2005) and MarkSim (Jones and Thornton, 2000; Jones et al., 2002) data. WorldClim is a global database of climate variables in grid format. The data layers were generated through interpolation of average monthly climate data from 15,000 to 47,000 weather stations during the years 1950 to 2000. Variables generated from WorldClim are annual average precipitation, annual average temperature and dry month per year. MarkSim uses interpolated climate surfaces based on a third-order Markov function. Annual average diurnal temperature range and mean annual solar radiation were generated using MarkSim. Dew point maps were calculated by the method of Linacre (Linacre, 1977) from the WorldClim dataset. Terrain attributes such as elevation, aspect and slope were generated and mapped from the digital elevation model (DEM) of the Shuttle Radar Topography Mission (SRTM) using geographical information systems (GIS) methodology (Jarvis, 2004).

Case Studies

We present two case studies, one using a data set from the Colombian coffee growing department Cauca and a second one from the Nicaraguan coffee growing departments, Matagalpa, Jinotega, Nueva Segovia and Región Autónoma del Atlántico Norte.



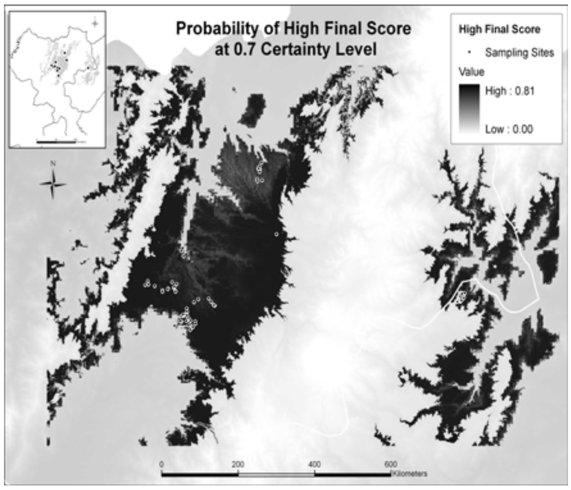


Figure 1 (left): Cauca and Nicaragua case study area. For Cauca the window of the "Driving Factor" analyses is shown: The large rectangle represents the entire study area (775866 ha), the medium the niche of El Tambo-Timbio (160765 ha), and the small the niche of Inzá (16005 ha).

Figure 2 (right): CaNaSTA "Score" analysis; the likelihood of high Final Scores at a 0.7 certainty level are shown. The areas of El Tambo-Timbio and Inzá are identified as potential high value coffee niches.

The Cauca case study data set consist of 88 sample points, 44 from sites in El Tambo-Timbio, 27 from Inzá and 17 from other municipalities. Coffea Arabica was harvested in May and June 2005 in homogeneous and geo-referenced plots of farmers, post-harvest processes were standardized with a mobile processing unit. The sensorial quality was assessed by an international cupping panel in August 2005. The sensorial evaluation was conducted according to SCAA standards and with an adapted cupping form assessing ten sensorial attributes: Fragrance/Aroma, Flavor, Aftertaste, Acidity, Sweetness, Body, Balance, Uniformity, Clean Cup, and Cuppers Score. The probability of High Final Score (Fig.2), the sum of all the sensorial attributes minus the defects, was predicted with CaNaSTA. To validate CaNaSTA three different tests and training sets were used with 25/75, 50/50 and 75/25 percent of the data accordingly to predict and test the model, respectively. Each set was repeated 10 times with randomly chosen predicting and testing sites. The performance of predicting the right quality classes was conducted only on the 70-90 quality class ranges due to the focus on high quality coffees. A quality class refers to 5 scores in the SCAA grading system. With the likelihood ratio chi-square test the dependency of the evidence and predictor scores was tested. The test is using a conformity matrix where the axes represent the evidence and predictor ranges accordingly and the matrix cells the agreement between them. “Driving Factor” analysis was then applied to determine the factors most impacting on sensorial coffee quality. The analysis was conducted on the two different environmental niches and on the entire Cauca data set (Fig. 1). The analysis extends of the Inzá niche is of 16 005 ha, the one of El Tambo-Timbio of 160 765 ha, and the Cauca area including all the sampled municipalities of Cauca of 775 866 ha.

The Nicaraguan data set consists of 67 sample points collected and analyzed by Lara-Estrada (2005). The samples, all of Caturra variety, were picked and processed in a standardized way and physical, bio-chemical and sensorial attributes were assessed. The sensorial attributes including Acidity, Body, Bitterness, Aroma, Cuppers preference, Flavor, were determined in the cupping lab of Atlantic SA in Nicaragua. Bio-chemical attributes including Cholorenic Acids, Caffeine, Fat Content, Trigonelline, and Sucrose Content, were assessed by Near Infrared Spectroscopy NIRS in CIRAD, France. “Score” analyses, an indicator of the likelihood to produce high quality, was conducted and combined with the “Certainty” surfaces. The cross correlations between pair surfaces were calculated (Nelson, 2004). GWR analyses were conducted to spatially quantify the impact of the environmental factors on quality attributes.

I Results: Niche concept

With the “Score” analyses (Fig. 2) the areas of El Tambo-Timbio and Inzá were identified as niches with high probability to produce specialty coffee. CaNaSTA was then validated, comparing prediction and evidence quality scores (Tab. 1), with the hypothesis being:

H_0 = Prediction and evidence scores are independent
H_1 = Prediction and evidence scores are dependant

In El Tambo-Timbio the p-value decreases from 6.2% to 1.9 % with increasing amount of prediction points (Tab.1). With the 25/75 set, the H_0 is being accepted; prediction and evidence scores are independent. For the 50/50 and 75/25 sets, H_1 can be accepted at a 5.2 % confidence interval, prediction and evidence scores are dependent. For Inzá the same is true, with the exception that H_1 for the 50/50 set can only be accepted at a 8.2 % confident level, which might be due to little evidence data. When analyzing the entire area no pattern is distinguishable, 50 and 75 percent of the data points are apt to predict the niches at a confidence level of 5.6 and 13% accordingly. These results make a lot of sense, when we recall the CaNaSTA methodology that uses site data and its environmental factor combination to predict entire areas. In the case of Cauca, this implies that sites from Inzá are used to predict qualities in El Tambo-Timbio and vice versa. Therefore, it becomes obvious that the niches cannot be identified at a low confidence level, but the methodology still serves for a general delimitation of niches that can thereafter be refined by adapting the analyses window to niche scale.

Table 1: P values of the likelihood ratio chi-square for the entire area and for the two niches.

	25 / 75	**50 / 50**	**75 / 25**
Cauca	0.43	0.056	0.13
El Tambo-Timbio	0.062	0.051	0.019
Inzá	0.86	0.081	0.014

To illustrate the site specificity of the interactions of environmental factors with quality, the "Driving Factor" analysis was run for the entire data set and for the two niches (Tab.2 and 3) separately. For the entire data set, only 1 enhancing and 3 reducing factors were identified, having a significance value > 2. As stated previously, by running a general analysis, we are predicting quality areas based on evidence data from distinct environmental conditions and insights of interactions with coffee quality are only of general nature. When analyzing niche by niche, a more detailed set of responsible factors can be obtained.

Table 2: Quality Enhancing Factors impacting on the Final Score attribute of the niches Inzá, El Tambo-Timbio and the whole Cauca sampling area. In parenthesis the significance indicator c is shown.

Quality Enhancing Factors	**Entire Data Set**	**Inzá**	**El Tambo-Timbio**
Altitude (masl)		1750 -1800 (2.02)	1652 -1725 (2.32)
			1725 -1798 (2.39)
Average Annual Dew Point (°C)		11.9-12.2 (2.43)	
		12.3-12.6 (2.07)	12.3 – 12.8 (2.38)
Average Annual Temperature (°C)		17.7 -18.1 (2.55)	17.8-18.9 (2.32)
		18 -18.4 (2.21)	
Average Annual Precipitation (mm)		1645 -1674 (2.2)	
	1760 – 1934 (2.31)	1587 -1616 (2.1)	

Table 3: Quality Reducing Factors impacting on the Final Score attribute of the niches Inzá, El Tambo-Timbio and the whole Cauca sampling area. In parenthesis the significance indicator c is shown.

Quality Reducing Factors	**Entire Data Set**	**Inzá**	**El Tambo-Timbio**
Altitude (masl)	1528 – 1623 (2.74)		
Slope (Grad)			34.5 – 40.9 (2.55)
		22.4 -25.6 (2.54)	21.6 – 27.9 (2.10)
Average Annual Dew Point (°C)	12.8 – 13.5 (2.4)	11.5 -11.9 (2.57)	14.3 -14.8 (2.00)
Average Annual Temperature (°C)		17.3 – 17.7 (2.47)	20 - 21 (2.02)
Average Annual Solar Radiation (Mj/m2/d)			21.8 – 22.3 (2.32)
Average Annual Precipitation (mm)	1133 – 1587 (2.78)		
Dry Month per Year (Month / Year)		3 (2.81)	

For both niches, altitude, average annual temperature and average annual dew point play an enhancing role for Final Score quality. The ranges are only slightly different, Inzá permitting lower temperatures and higher altitudes than El Tambo-Timbio. Average Annual Precipitation plays an important enhancing role in Inzá and for the entire data set of Cauca. Slope influences Final Score negatively in both niches. Dew Point ranges below and above the ones identified as enhancing Driving Factors impact negatively, the same is true for Average Annual Temperature. The optimal coffee growing annual average temperature in Inzá is somewhere between 17.7 and 18.4C whereas it is slightly higher (17.8 – 18.9C) for El Tambo-Timbio. The results demonstrate variability in the environmental factors that impact on Final Score and the necessity of assessing these factors according to their niches.

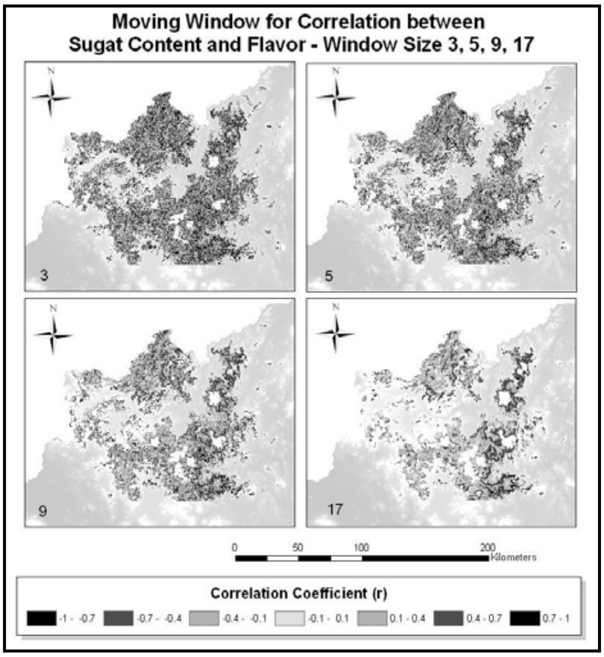
II Results: Variability in space

Recent studies show the interactions of environmental factors and coffee quality and the correlations between quality attributes for selected study sites: Vaast *et al.* (2005a) reported no differences in the caffeine content of high and low quality coffees; Avelino *et al.* (2005) did not find any strong correlation between sensorial characteristics and caffeine, trigonelline, fat, sucrose, cholorenic acids; Vaast *et al.* (2005a & b) showed that there is a strong relationship between high trigonelline content of coffee beans and higher bitterness and lower acidity of the coffee beverage; Decazy *et al.* (2003) found a positive relationship between bean sucrose content and coffee acidity and quality; Decazy *et al.* (2003) found that high bean fat content related to good acidity and beverage preference. Little work has been done on the spatial variability of coffee attributes and their interactions with the environment. The study of Lara-Estrada (2005) was used to put its data and results in a spatial perspective, and to add value to the study. The correlation of the "Score" response maps, for 10 different quality attributes, demonstrate the variability in correlation between the responses (Tab.4).

Table 4: Correlation coefficients of response variable pairs (Pref. = Preference, Caff= Caffeine, C.A. = Cholorenic Acids, F.C. = Fat Content, Suc=Sucrose, Trigo= Trigonelline)

	Acidity	Aroma	Bitter	Body	Flavor	Pref.	Caff	C. A.	F.C.	Suc.
Aroma	0.72									
Bitter	-0.22	-0.08								
Body	0.69	0.76	0.17							
Flavor	0.82	0.64	0.06	0.74						
Pref.	0.82	0.74	0.00	0.82	0.90					
Caff	-0.13	-0.03	0.16	0.07	0.07	0.02				
C. A.	0.08	0.08	0.26	0.25	0.12	0.17	0.50			
F. C.	-0.24	-0.21	-0.17	-0.24	-0.24	-0.22	0.28	-0.04		
Suc.	0.33	0.44	0.03	0.42	**0.28**	0.35	-0.10	0.35	-0.23	
Trigo	0.18	-0.04	-0.23	-0.06	0.09	0.03	-0.37	-0.57	-0.31	-0.27

A single figure averaging the correlation of a pair of variables is not always very meaningful. For example, sugar content and flavor are poorly correlated (r = 0.28); only when visualizing the r coefficient on a map, does the importance of the spatial variability become evident (Fig. 3) and highly correlated areas (up to r = 1 or r = -1) appear (Fig. 3). The spatial correlation window sizes of 3, 5, 9 and 17 grid cells used, translates into 9, 16, 65 and 94 ha analyzed window; these correspond to farm size (up to 9 ha), groups of farms (up to 16 ha), association (up to 65 ha), and micro-catchments (up to 294 ha). The different resolutions give insight on the scale where correlation patterns emerge, which is valuable information for coffee quality profile identifications and their marketing. The analysis also demonstrates to farmers' associations the strengths and weaknesses of their coffee qualities. Not only do quality responses vary in space, but also the environmental factors and especially their impact on quality. A GWR analysis on the overall importance of the environmental factors impacting on Flavor result in Flavor being significantly dependent on “Number of Average Annual Dry Month” (DM) and “Average Annul Diurnal Temperature Range” (DTR) at a confidence level of 5% and 1%, respectively.



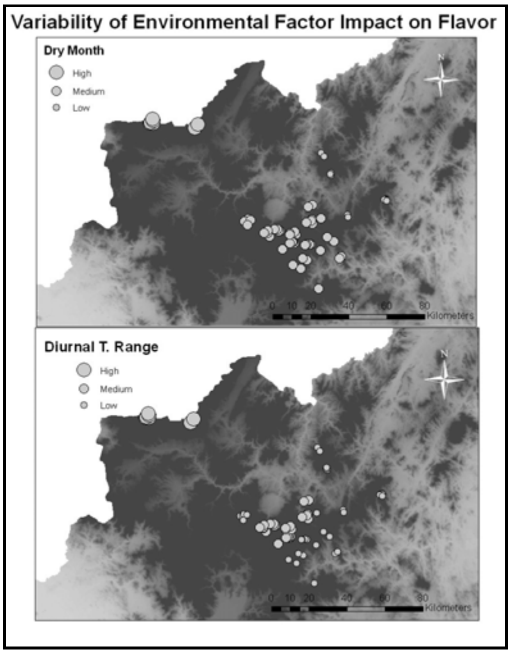


Figure 3 (left): The variability of the correlation between Sugar Content and Flavor at different moving window resolutions

Figure 4 (right): Variability of the impact on Flavor for two decisive environmental factors (dry months and diurnal temperature range). Bigger dots are representing a larger impact on Flavor than smaller dots.

Even tough these two environmental factors are significant for Flavor; their contribution to each site is distinct. Figure 4 shows the variable impact of DM and DTR on the Flavor quality. Bigger dots are representing a larger impact on Flavor than smaller dots. The impact of these factors is very heterogeneous in space but clusters and forms niches where DM or DTR are more decisive.

Conclusion

CaNaSTA predicts niches likely to produce high Final Score coffee quality values at p = 0.056 – 0.1 confidence level for a 800 000 ha area in Cauca using only quality data of 88 sites. Within pre-defined niches, quality classes can be predicted at p = 0.019 – 0.051 confidence levels for the El Tambo-Timbio niche (160 000 ha) and at p = 0.014 – 0.081 confidence levels for the Inzá niche (16005 ha). The ranges of the factor enhancing or reducing quality in the two niches of Cauca are very different depending on the environmental envelopes predominating in the niche. The importance and utility of SDS tools and Geographical Analyses for assessing the variability of environmental factors and causal quality responses is shown in a case study in Nicaragua. They are very powerful tools to extrapolate point information to surface information. Environmental factors and their impact on quality are very heterogeneous in space. Nevertheless Geographical Analyses allow the identification of niches with similar factor combination.

References

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

Decazy, F., J. Avelino, B. Guyot, J.J. Perriot, C. Pineda, and C. Cilas. 2003. Quality of Different Honduran Coffees in Relation to Several Environments. JFS, 68 (7):2356-2361.

Fotheringham, A.S., C. Brunsdon, M. Charlton. 2002. Geographically Weighted Regression the analysis of spatially varying relationships. John Wiley and & Sons Ltd. West Sussex, England.

Jarvis, A., J. Rubiano, Neslon, A. Farrow, A. Mulligan, M. 2004 Practical use of SRTM data in the tropics – Comparison with digital elevation models generated from cartographic data. Working Document no. 198. CIAT, Cali, Colombia.

Jones, P., and P. Thornton. 2000. MarkSim: Software to generate daily weather data for Latin America and Africa. . Agron. J. 93.

Jones, P.G., P.K. Thornton, W. Diaz, and P.W. Wilkens. 2002. MarkSim, Version 1. A computer tool that generates simulated weather data for crop modeling and risk assessment. CIAT CD-ROM series, CIAT, Cali, Colombia.

Hijmans, R.J., S.E. Cameron, J.L.Parra, P.G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J Climatol. 25

Läderach, P. T. Oberthur, N.Niederhauser, H. Usma, L.Collet, H.A.J. Pohlan. 2006. Café especial: Factores, dimensiones e interaciones, in El cafetal del futuro: Realidades y Visiones. (eds) Pohlan, J., L. Soto, J. Barrera, Aachen, Shaker Verlag, 2006, 462pp.

Lara-Estrada, L.D. 2005. Efectos de la altitud, sombra, producción y fertilización sobre la calidad del café (Coffea arabica L. var. Caturra) producido en sistemas agroforestales de la zona cafetalera nor-central de Nicaragua. Master thesis, Centro Agronómico Tropical de Investigación y Enseñanza. Costa Rica.106pp.

Nelson, A. 2004. The spatial analysis of socio-economic and agricultural data across geographical scales: Examples and applications in Honduras and elsewhere, PhD Thesis, University of Leeds, England. pp369.

Linacre, E. 1977. A simple formula for estimating evaporation rates in various climates, using temperature data alone. Agric. Meteorol. 18:409-424.

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 J. PEARL (eds.), Morgan Kaufmann, San Mateo, CA, pp. 345-352.

Vaast, P. Cilas, C. Perriot, J; Davrieux, J; Guyot, B; Bolaños, M. 2005a. Mapping of Coffee Quality in Nicaragua According to Regions. Ecological Conditions and Farm Management. In Proceedings of the 20th International Congress on Coffee Research (ASIC) Bangalore, India. p 842-850.

Vaast, P; Van Kanten, R; Siles, P; Dzib, B; Frank, N; Harmand, J; Genard, M. 2005b. Shade: A Key Factor for Coffee Sustainability and Quality. Proceedings of the 20th International Congress on Coffee Research (ASIC) Bangalore, India. p 887-896.