Journal of Agricultural Extension and Rural Development Vol. 4(7), pp. 141-146, 18 April, 2012
Available online http:// academicjournals.org/JAERD
DOI: 10.5897/JAERD11.095
ISSN- 2141 -2154 ©2012 Academic Journals

Full Length Research Paper

Discrimination of the cultivation systems for *coffea arabica* L. via the incidence of filamentous fungi using the zip model on the Bayesian approach

Vanessa Siqueira Peres da Silva[1], Juliana Garcia Cespedes[2], Marcelo Ângelo Cirillo[1], Fabiana Aparecida Couto[3] and Luís Roberto Batista[3]

[1]Department of Exact Sciences, Federal University of Lavras (UFLA), 37200-000, Lavras, MG, Brazil.
[2]Department of Science and Technology, Federal University of Are Paulo (UNIFESP), Are José of the Campos, SP, Brazil.
[3]Department of Food Sciences, Federal University of Lavras (UFLA), Lavras, MG, Brazil.

Accepted 3 February, 2012

Coffea arabica **beans can be produced when basically considering two system types, conventional and organic. The main difference lies in the fact that the conventional system sues chemical fertilizers and pesticides whereas, in the organic system the produce use inputs derived from organic matter. Naturally, the presence or absence of filamentous fungi occurs in both systems. In conventional farming, the fungi found are usually related to the production of mycotoxins whereas in organic coffees, there is still a lack of studies on the diversity of these fungi. With this motivation, this article proposes the use of a Zero-inflated Poisson model as an alternative to discriminate the organic and conventional growth systems in relation to the incidence of filamentous fungi in** ***C. arabica*** **beans using Bayesian inference techniques. The main advantage in favour of the use of this model is given in the update of the experimental results obtained in past experiments through the analysis of other collected coffee bean samples of the same species, allowing a more careful assessment of the production, coffee quality, and of the coffee products.**

Key words: *Coffea arabica*, cultivation system, ZIP model, Bayesian inference.

INTRODUCTION

Coffee production is basically done using the conventional and organic cultivation systems. The main difference between them lies in the fact that coffee production in the conventional system is done with the use of herbicides, pesticides and the application of inorganic nutrients, which renders this type of system as less environmentally appropriate. As regards organic farming system, Theodoro et al. (2003) mentions that no pesticides are used and that high-solubility fertilizers can be replaced by by-products from organic matter such as, for example, bio-fertilizers, coffee pulp and skin and animal-produced wastes. With these characteristics, the use of the organic system has seen a steady increase due to consumer demands for better quality products and a concern with the environment.

As regards coffee quality, it can be said that this is commercially determined by the physical characteristics of beans and the organoleptic features of beverage. However, the development of microbial infections on beans may compromise both its visual aspects such as the taste and aroma. Pasin et al. (2009) mentioned that amongst the micro-organisms associated with fruit and coffee beans, filamentous fungi represent the group that can cause the most damage, though according to Batista et al. (2009), Silva et al. (2008), Taniwaki et al. (2003), and Vilela et al. (2010), these micro-organisms can be isolated in the coffee beans and be associated with beverage quality and the risk of mycotoxin incidence

*Corresponding author. E-mail: phiriaustin923@gmail.com.

(Bokhari, 2007).

Given that the fungi form a large and diversified group of micro-organisms made by thousands of species (Hawksworth, 2004) and that they have great environmental and economic importance as they take part in environmental processes such as organic matter decomposition, nutrient cycling, biological disease and pest control (Moreira et al., 2008) many research works have been done and, amongst them, the concern with coffee quality has earned the spotlights, including the stages of bean production and processing, and studies aimed at genetic improvement that create more robustness in the plants, associated to maximum productivity (Illy and Viani, 1995; Carvalho et al., 1997). In this sense, several coffee cultivations, with important productive and vegetative characteristics have been developed, although, it should be pointed out that the fungi associated to the coffee beans can produce mycotoxins, and toxic substances that are highly hazardous to human health. Based on the foregoing, the high incidence of filamentous fungi in coffee beans has been reported in several studies such as those by Batista et al. (2003, 2009) and Silva et al. (2000, 2008). Kouba (2003) stated that organic foods would be more prone to mycotoxin contamination than conventional foods as they are not treated in the same manner as with anti-fungi agents, although, there are still no evidences that special foods are more prone to mycotoxin contamination than the conventional ones (FAO, 2000).

To set the statistical methodology that is to be employed in the research into context, the counting of these micro-organisms can be modelled via the Poisson model although, if one considers a sample expressive quantities of null values, the use of this distribution may be challenged, given the absence of knowledge regarding how much this model can support the number of these observations. For that, the recommendation is to use the Zero-inflated Poisson model, commonly referred to as ZIP model (Xie et al., 2001), where the adjustment is done in two steps: one for the null counts and another for those that are not null. The parameters to be estimated correspond to the proportion of zeros the sample may display, represented by p and the average rate of the Poisson distribution Poisson as defined by θ.

The applicability of this model to coffee quality is verified to infer the quantity of a particular fungus species in the coffee samples, such that the null value may indicate the absence of this fungus. The importance of this study gains sharper relief if the model proposed allows the researcher to update it with new experimental results. Based on this information, this work aims to propose the use of zero-inflated Poisson models in the Bayesian approach as an alternative to discriminate between conventional and organic coffee bean cultivation systems in relation to the incidence of filamentous fungi. With this purpose, the applicability of these models is given in the expectation of being recommended for the improvement of studies on coffee quality.

MATERIALS AND METHODS

To undertake this work, the description of the materials and methods was structured in the following manner: the first aspect described the procedure of obtaining the samples; the second aspect described the Zero-inflated Poisson model using Bayesian inference; the third aspect involved the definition of the prior distributions and the last part described the method of obtaining the posterior conditional distributions.

Sample obtaining

Coffee bean samples were collected from three locations in the Southern part of the state of Minas Gerais, Brazil. After that, the samples were transferred to the Mycology and Mycotoxin Laboratory at the Department of Food Sciences (DCA) and to the Mycology Laboratory at the Department of Phytopathology at the Federal University of Lavras (UFLA) in the city of Lavras, state of Minas Gerais – Brazil, where work was done to identify the fungi until the species. To isolate the fungi, a direct plating technique was used pursuant to the methodology as described by Sansom et al. (2004). Following the execution of this procedure, a count was started for the fungi, identified and otherwise, in the coffee samples evaluated, where the unidentified filamentous fungi were given the zero value. A high frequency of null values was observed in the sample, and thus, in line with the goal proposed, a Zero-inflated Poisson model was adjusted for the coffee bean samples that were grown in both conventional and organic systems using the Bayesian approach.

Zero-inflated Poisson model using Bayesian inference

In line with the definition of the Zero-inflated Poisson models (ZIP) of Johnson, et al. (1992) for estimating the parameters of the ZIP model the random variable *Y* was assumed, as described in Equation 1:

$$f(y;p,\theta)=\begin{cases} p+(1-p)\exp(-\theta), & se \quad y=0 \\ (1-p)\exp(-\theta)\dfrac{\theta^{y}}{y!}, & se\ y>0, \text{ where} \end{cases} \tag{1}$$

The percentage of zeros in the sample is represented by parameter $0<p<1$ and the mean rate for the standard Poisson distribution given by $\theta>0$. With this specification, the function of verisimilitude of the model with ZIP distribution is defined in Equation 2 by:

$$L(p,\theta \mid \underset{\sim}{y})=[p+(1-p)\exp(-\theta)]^{n_0}[(1-p)\exp(-\theta)\prod_{i=1}^{n}\frac{\theta^{y_i}}{y_i!}]^{(n-n_0)} \tag{2}$$

Prior distributions of the parameters p and θ

It is considered that the restriction $0<p<1$ and $\theta>0$ in the estimation of parameters p and θ is based on this information - prior to distributions, p and θ are respectively specified by Equations (3) and (4):

$$\pi(p)\sim Uniforme(a,b)\Rightarrow \pi(p)=\frac{1}{(b-a)} \tag{3}$$

$$\pi(\theta) \sim Gama(\alpha,\beta) \Rightarrow \pi(\theta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}\exp(-\beta\theta) \quad (4)$$

Conditional posterior distributions

Given Equations 2, 3 and 4, the joint posterior distribution is defined by:

$$\pi(p,\theta \mid \underset{\sim}{y}) \propto [p+(1-p)\exp(-\theta)]^{n_0}[(1-p)\exp(-\theta)\prod_{i=1}^{n}\frac{\theta^{y_i}}{y_i!}]^{(n-n_0)} \times \frac{1}{(b-a)}\frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}\exp(-\beta\theta)$$

Through the joint posterior distribution, the full conditional posterior distributions of p and θ are given in Equations (5) and (6):

$$\Rightarrow \pi(p \mid \theta, \underset{\sim}{y}) \propto \underbrace{[p+(1-p)\exp(-\theta)]^{n_0}[(1-p)\exp(-\theta)\prod_{i=1}^{n}\frac{\theta^{y_i}}{y_i!}]^{(n-n_0)}}_{\text{núcleo de uma } ZIP(p,\theta)} \quad (5)$$

$$\Rightarrow \pi(\theta \mid p, \underset{\sim}{y}) \propto \underbrace{\theta^{(\alpha^*-1)}\exp[(-\theta)^{\beta^*}]}_{\text{núcleo de uma } Gama(\alpha^*,\beta^*)} \quad (6)$$

It can be seen that the full conditional distributions have a closed form and can be summarized thus:

$$p \mid \theta, \underset{\sim}{y} \sim ZIP(p,\theta)$$

$$\theta \mid p, \underset{\sim}{y} \sim Gama(\alpha^*,\beta^*); \quad \begin{cases} \alpha^* = \sum_{i=1}^{n} y_i(n-n_0)+\alpha; \\ \beta^* = \beta+(n-n_0)-1 \end{cases}$$

Using the Gibbs sampling algorithm (Gelman et al., 2004) a sequence of samples was generated for the distribution of the joint probability, via the Monte Carlo methods with Markov chains with the purpose of obtaining an approximation of the margin distribution. This way, with the estimates of the ZIP model for the coffee samples in each cultivation system the models were compared, considering the approximation error between the frequencies observed and expected. This way, the results were discussed in comparison with other bibliographic references that allowed the indication of the level of adequacy in the formulation of the ZIP models in the Bayesian approach as an alternative to discriminating the coffee samples submitted to the two cultivation systems, based on the presence or absence of filamentous fungi.

Considerations

We would like to point out that counting of filamentous fungi is different from quantification of ochratoxin A. The toxin is onlyproduced by toxigenic fungi; on the other hand, the presence of fungi that grows on grain does not necessarily denote the existence of ochratoxin A. because there are times the fungus grows and does not produce ochratoxin A. And as such, the presence of the fungus does not necessarily indicate the presence of the toxin. We may also have a situation when the fungus develops, produces ochratoxin A. and because of some environmental stress or suffering in the processing of coffee beans is eliminated, that is to say, that in this sample, we could not identify the fungus, but in a liquid chromatography analysis, it will be possible to quantify the toxin. In this case, the absence of the fungus does not necessarily indicate the absence of the toxin. There is no microbiological method that is able to closely associate the presence of toxigenic fungi and mycotoxins in a single sample, even when using molecular biology. These techniques have identified genes involved in the production of toxins, but we do not know if these genes were expressed as to produce the toxin.

The great advantage of the methodology used in this study is that it allows the identification of the viable microorganisms, and not the dead ones.

RESULTS

The results are divided into various parts. The first part indicates the level of adequacy of the ZIP model adjusted via the Bayesian inference in both of the cultivation systems studied in this work. The second part of the results presents the validation of the model as an alternative to the discrimination of these systems, while the last part presents the discussion of the results, emphasizing the absence of filamentous fungi in agreement with other bibliographic references. It is based on this layout that the results are discussed; the discussion follows subsequently.

Adjustment and validation of the ZIP model in the Bayesian approach

Considering the count of filamentous fungi in coffee beans grown in conventional and organic cultivation systems, given the samples collected, as described in sample obtaining, the results described in Table 1 proved that via the frequency observed, the absence of filamentous fungi, represented by the frequency of values equal to zero is higher than the frequency observed in the other fungi. The occurrence of this result observed in both cultivation systems, suggested that the posterior distribution may supposedly present the existence of high deviations of symmetry and kurtosis that will certainly affect the quality of the adjustment of the ZIP model.

As a result of the sampling data described in Table 1, the use of the Bayesian methodology is based on the assumption of the prior distributions of *p* (5) and *θ* (6), considering the independence between the quantities of uncertainties which is characterized by U (0,1) and Gama (0,1; 0, 01) respectively. In keeping to these specifications, the Gibbs sampler was used where 60,000 iterations were processed, with the 10,000 beingdiscarded for the burn-in period of the chain, and to ensure the independence of the sample, a spacing was considered between the size 10 points sampled 10 ('thin'). This way, a sample of 5,000 was generated in the construction of the distributions of each parameter to be estimated. The posterior densities obtained for each parameter in both systems are shown in Figure 1A and B.

Table 1. Counts for the number of filamentous fungi in coffee samples grown in the conventional and organic systems.

Conventional		Organic	
No. of fungi	**Frequency observed**	**No. of fungi**	**Frequency observed**
0	73	0	62
83	16	130	27

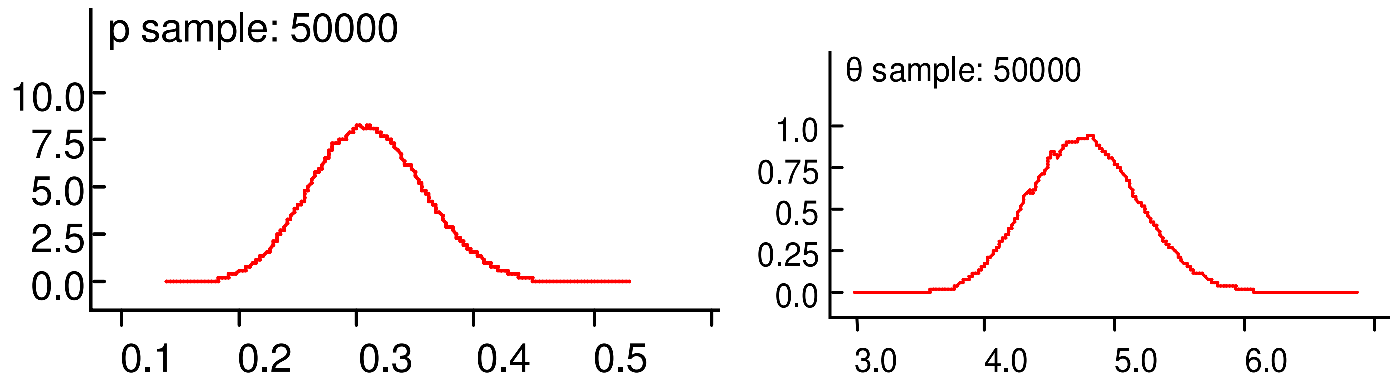

Figure 1A. Density of the posterior margin distribution for parameters *p* and *θ*, for samples collected in the organic cultivation system.

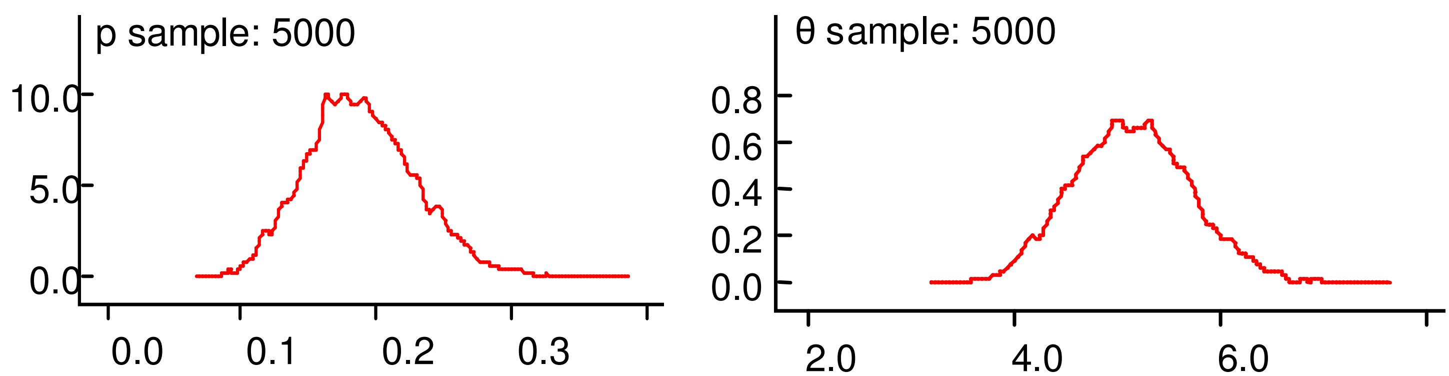

Figure 1B. Density of the posterior margin distribution for parameters *p* and *θ*, for samples collected in the conventional cultivation system.

Based on the approximate densities functions for each one of the adjusted parameters work was done to obtain the descriptive statistics as shown in Table 2.

The results described in Table 2 led to the finding that the Zero-inflated Poisson model can be adjusted to the sample data in both cultivation systems, that is, although, both systems presented high zero rates, the hypothesis of the effect of super-dispersion was discarded. Such a statement was corroborated with the verification that the values for standard deviations were below the values for posterior averages and thus, there was statistical evidence to believe that the zero-inflated Poisson model is coherent to adjust the count of filamentous fungi in the coffee bean samples as observed for the two cultivation systems. In line with the results shown in Figure 2A and B and the summary of the descriptive statistics shown in Table 2, the intervals of credibility, considering percentages of 2.5 and 97.5% are shown in Figure 2A and B.

Validation of the ZIP model in the discrimination of conventional and organic cultivation systems

Given the validation of the ZIP model in both cultivation systems and considering the posterior average for the

Table 2. A posterior summary of parameters *p* and *θ* for number of fungi grown in two cultivation systems.

Parameter	Mean	Standard deviation	2.5%	50%	97.5%
		Conventional farming			
p	0.189	0.040	0.116	0.186	0.274
θ	5.153	0.573	4.087	5.144	6.321
		Organic cultivation			
p	0.310	0.048	0.219	0.309	0.409
θ	4.770	0.429	3.967	4.760	5.649

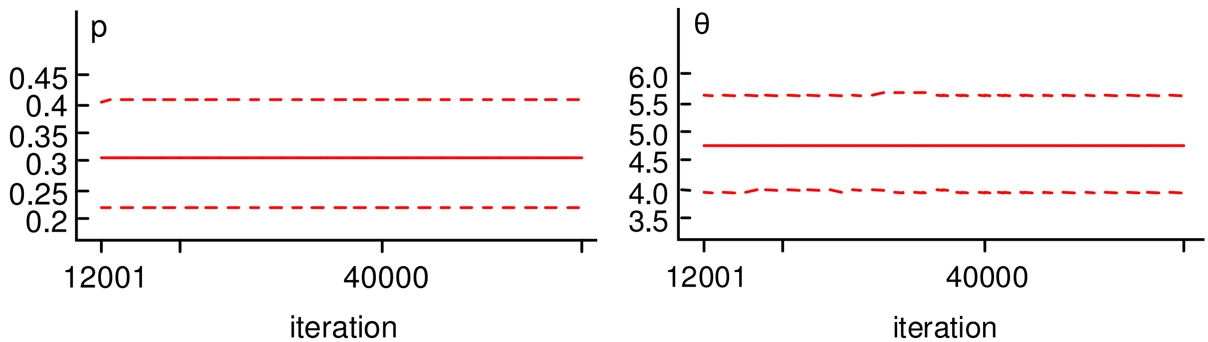

Figure 2A. 95% credibility interval for parameters p and θ for coffee samples grown in the organic system.

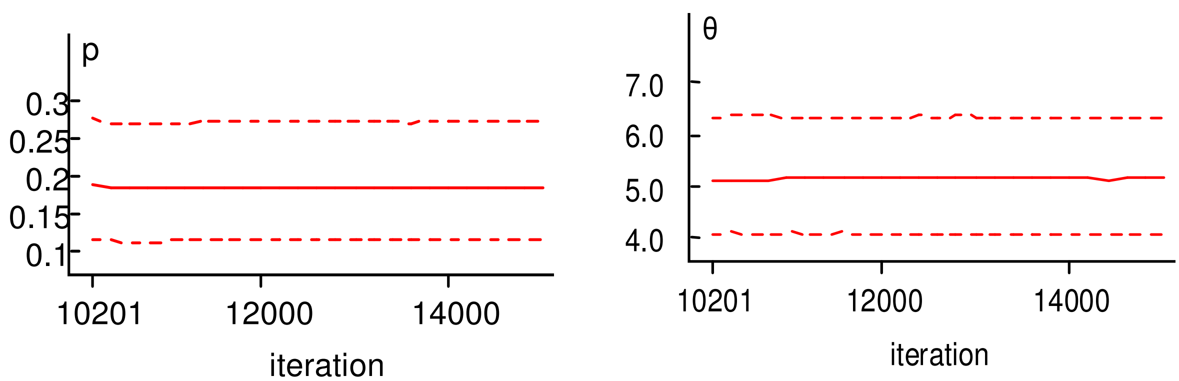

Figure 2B. 95% credibility interval for parameters p and θ for coffee samples grown in the conventional system.

distribution of mean rate *θ*, it was possible to obtain the approximation error between the frequencies observed and expected with the adjustment of the model. This way, assuming the Bayesian estimate for parameter *θ* in values 5,153 and 4,770 respectively for the coffee samples produced in the conventional and organic cultivation systems results were obtained as shown in Table 3.

The results shown in Table 3 suggested that the applicability of the Zero-inflated Poisson model in the Bayesian approach as proposed to discriminate the organic and conventional cultivation systems is viable. This recommendation is validated by the differences between the approximation errors for the frequencies observed and expected by the adjustment of the models such that, in conventional farming systems, for the quantity of zeros, that represent the absence of filamentous fungi, the error was significantly smaller in relation to the organic cultivation system. This discrimination was also in line with the work of Hole et al. (2005), where the majority of the studies on conventional and organic farming suggested there are sizeable

Table 3. Comparison of the adjustment of zero-inflated ZIP model in organic and conventional cultivation systems.

Nº of fungi	Frequency observed	Probability of occurrence	Frequency expected	Approximation error (%)
Conventional farming system				
0	73	0.820	59.860	18
83	16	0.180	2.880	82
Organic cultivation system				
0	62	0.700	43.400	30
130	27	0.300	8.100	70

differences between these two cultivation systems.

Naturally, there were also inconsistencies in these studies due to the complexity of the iterations and the large number of environmental variables, although, these inconsistencies indicated that the benefits to the biodiversity of the biological agriculture can vary according to factors such as location, climate, culture and type of species.

In the case of other statistical methodologies used, Bengtsson et al. (2005) did a meta-analysis on the biodiversity in organic and conventional farms, using 42 studies to compare these two cultivation systems. The results showed that the wealth of species was approximately 30% bigger in organic farms than in conventional ones.

However, based on the conformance of the experimental results discussed in this article with the agreements of the bibliographic references mentioned, it should be noted that the Zero-inflated Poisson model in the Bayesian approach was an alternative to differentiate the presence and absence of filamentous fungi in coffee bean samples, and was deemed as promising as it allows the researcher, at each sample observed, to update the model with prior experiments, adding the knowledge and refinement from the results to the quality of the coffee.

Conclusion

The Zero-inflated Poisson model in the Bayesian approach discriminated organic and conventional cultivation systems in relation to the presence or absence of filamentous fungi in coffee beans, with the advantage of giving the researcher the ability to update the knowledge on other samples evaluated, showing the importance of evaluating the production of the coffee, as well as its quality, and the coffee products.

REFERENCES

Batista LR, Chalfoun SM, Prado G, Schwan RF, Wheals AE (2003). Toxigenic fungi associated with processed (green) coffee beans (*Coffea arabica* L.). Int. J. Food Microbiol., 85(3): 293-300.

Batista LR, Chalfoun SM, Silva CF, Cirillo MA, Varga EA, Schwan RF (2009). Ochratoxin A in coffee beans (*Coffea arabica* L.) processed by dry and methods. Food Control. 20(9): 784-790.

Bengtsson J, Ahnstrom J, Weibull A (2005). The effects of organic agriculture on biodiversity and abundance: A meta-analysis. J. Appl. Ecol., 42(2): 261-269.

Bokhari FM (2007). Mycotoxins and toxigenic fungi in Arabic coffee beans in Saudi Arabic. Adv. Biol. Res.,1(1-2): 56-66.

Carvalho VD, Chagas SJR, Souza SMC (1997). Factors affecting the quality of coffee. Tell Agricultural, 18(187): 5-20.

Food and Agriculture Organization of the United Nations (FAO). Organic food and farming (2000). Myth and reality: Organic vs non-organic: the facts. Bristol House Soil Association, p. 32.

Gelman A, Carlin JB, Stern HS (2004). Bayesian data analysis. Chapman and Hall/CRC, London.Hawksworth DL (2004). Fungal diversity and its implications for genetic resource collections. Stud. Mycol., 50: 9-18.

Hole DG, Perkins AJ, Wilson JD, Alexanderd HI, Gricee PV, Evansa AD (2005). Does organic farming benefit biodiversity? Biol. Conserv., 122(1): 113-130.

Illy A, Viani R (1995). Espresso coffee: The chemistry of quality. Academic Press, San Diego.

Johnson N, Kotz S, Kemp AW (1992). Univariate Discrete Distributions. Wiley, New York, pp. 312-318.

Kouba M (2003). Quality of organic animal products. Livest. Prod. Sci. 80(1): 33-40.

Moreira FMS, Siqueira JO, Brussaard L (2008). Soil biodiversity in Brazilian ecosystems. Federal University Lavras, Lavras, p. 768.

Pasin LAAP, Almeida JR, Abreu MS (2009). Fungi associated with coffee beans (*Coffea arabica* L.) on five cultivars. Acta Bot. Bras., 23(4): 1129-1132.

Sansom RA, Houbraken JAMP, Kuijpers AFA, Frank JM, Frisvad JC (2004). New ochratoxin A or sclerotium producing species in *Aspergillus* section *Nigri*. Stud. Mycol., 50: 45-61.

Silva CF, Schwan RF, Dias ES, Wheals AE (2000). Microbial diversity during maturation and natural processing of coffee cherries of *Coffea arabica* L in Brazil. Int. J. Food Microbiol., 60(2-3): 251-260.

Silva CF, Batista LR, Schwan RF (2008). Incidence and distribution of filamentous during fermentation drying and storage of coffee (*Coffee Arabica* L) beans. Braz. J. Microbiol., 39(3): 238-240.

Taniwaki MH, Pitt JI, Teixeira AA, Iamanaka BT (2003). The source of ochratoxin A in Brazilian coffee and its formation in relation to processing methods. Int. J. Food Microbiol., 82(2): 173-179.

Theodoro CGC, Carvalho JG, Corrêa JB, Guimarães, JBC (2003). Nutritional diagnosis of organic coffee agroecosystems in the Minas Gerais state. Cienc. Agrotecnol., 27(6): 1222-1230.

Vilela DM, Pereira GV, Silva CF, Batista LR, Schwan RF (2010). Molecular ecology and polyphasic characterization of the microbiota associated with semi-dry processed coffee (*Coffea arabica* L.). Food Microbiol., 27(8): 1128-1135.

Xie M, He B, Goh TN (2001). Zero-inflated Poisson model in statistical process control. Comput. Stat. Data Anal., 38(2): 191-201.