# Regression Methods for Evaluating Beekeeping Production

W. S.Tassinari[1], E. L. T. Peixoto[2], M. C. A. Lorenzon[2]

[1] Dep of Mathematics, Universidade Federal Rural do Rio de Janeiro, Rod. 465, km 7, Seropédica, CEP 723851-970, ERJ, Brasil, TF 55 021 37873975, [2] Dep of Animal Science, Universidade Federal Rural do Rio de Janeiro, lorenzon_ufrrj@yahoo.com.br.

**Abstract**

Brazilian beekeeping is a robust industry, making the country one of the globe's most important honey producers and it has developed from the Africanization of European honeybees. In order to analyse the overall situation, this research evaluated honey production using spatial and classic regression models.This study considered the beekeeping situation in the state of Rio de Janeiro, Brazil; data were obtained by a survey compiled in 2006 from 1418 beekeepers. The honey yield per colony per year is under the national average of 15 kg. There were a high honeybee population and a scarcity of melliferous flora in some regions, reducing the honey yield of the majority of counties to levels less than 9 kg per colony per year. Despite the difficulties of the beekeeping sector, the producers are motivated to raise bees regarding the success of the Rio de Janeiro honey market. The best statistiscal fit was a semiparametric spatial model based on the *thin plate spline approach.* The beekeeping sector needs support and research should include statistical analysis to ensure valid results.

Key words: bee, productivity, family farming, beekeeper

## Introduction

In Brazil, beekeeping has been developed from the Africanized honey bees and currently ranks as the 11$^{th}$ largest honey producer. The sector is facing important issues regarding the management of Africanized colonies, probably due to the lack of technical knowledge. Therefore, the introduction of modern apiculture techniques becomes imperative. It is important to note the principal constraints experienced by Brazilian beekeepers in some regions. Using classic linear and spatial regression models, this study aims to evaluate the beekeeping technology factors related to honey production and to identify possible spatial clusters. Given the complexity of the relations of interest, the choice of the models suitable for analysis is of much importance to reveal the real producing regions.

## Materials And Methods

This study considered the beekeeping situation in the state of Rio de Janeiro, in existence for more than 60 years, This state is located in the southeastern part of Brazil (22º54'23"S, 43º10'21"W), comprising 92 counties. The beekeeping data were obtained with a survey made up in 2006, when was interviewed 1418 beekeepers, distributed across 72 counties. At the time of this study, mostly beekeepers had 20 hives, which were not movable, they processed only honey and did not depend on agriculture. Rio de Janeiro has expansive wild vegetation, as it is not an agrarian state; the honey production is about 404 tons mostly from wild vegetation. Livestock represents the largest portion of agrarian production. Apiculture is not an important source of livehood, nonetheless, the consumption of beekeeping products in Rio de Janeiro is an integral part of the economy.

For statistical analysis we calculated the coefficients of variation, to check the distribution of the study variables (Bussab and Morettin, 2002). Then we adjusted the classic regression and the space regressions. Before the spatial model fit, it is recommended to measure spatial heterogeneity through spatial autocorrelation. The spatial autocorrelation of a particular attribute at any two locations depends on the lag distance between the two locations, but not the actual locations (Bailey and Gatrell, 1995). In this work, the Moran Index with simulations was performed to measure the residual autocorrelation (*R Development Core Team*, 2007). For spatial fit SAR and CAR models,

the weights for the adjacency and distance matrix ($w_{ij}$) were given by the weights $w_{ij}$ = 1 if *i* and *j* were adjacent, and 0 otherwise (Bailey and Gatrell, 1995). Akaike's Information Criterion (AIC) and the residuals sum squares (RSS) were used to measure the fit. Models with lower AIC and RSS values are considered to be better models in terms of balancing fit with parsimony (Cliff and Ord, 1981). The linear correlation between observed and predicted values was used to evaluate model consistency. All statistics analysis was done with the R Development Core Team, version 2.7.0 (2007).

**Results and Discussion**

This research reported the importance of including spatial effects in the regression models, even when the traditional linear regression model did not show a pattern of spatial dependence. This model facilitates the analysis of producing regions, which is not available with other models (Druck et al., 2004).

Despite regression analysis (classical and spatial) revealing that the data were well fitted, the semi-parametric model adjusted via thin plate spline (GAM) yielded the best overall fit (AIC = 590.62, SQR = 1865.81 and ρ = 0.886). GAM is an alternative to modeling nonlinear relationships that do not have a set. This type of model is based on non parametric functions, which the type of association is defined by the data.

It is important to remember that assumptions usually employed in the analysis hardly correspond to reality so accurate, even the most sophisticated model in question is used.

The predicted distribution of the beekeepers' average honey production across different counties in the state of Rio de Janeiro is presented in Figure 1. Most counties (37) demonstrated small-scale production (<14,44 kg), with a few (19) exhibiting large-scale production (>20.1 kg). The low honey yield per hive per year (13.64 kg) stresses the insufficient management conditions; this figure is under national average of 15 kg (*unpubl. data*) .

The GAM of spatial regression highlighted more counties engaged in large-scale production than the other models; the linear and SAR models were more similar in this regard. The all models stress the low honey production observed throughout most of the state of Rio de Janeiro problaby resulting from reduction in melliferous flora as well as inadequate technical support for beekeepers.

The high variability in production factors among different counties, the low average honey yield per hive and the large proportion of beekeepers reporting small yields indicate the deficits in familial farming system, suboptimal for beekeeping. The low number of counties reporting large-scale production (>20 kg) illustrates the difficulties in reaching this level of production. Therefore, the availability of melliferous flora in the state of Rio de Janeiro is limited.

This research allows two remarks regarding the beekeeping sector: 1) increasing losses from diseases in certain regions may represent important foci of infection. In 1992, only 2% of beekeepers reported losses due to bee disease, in 2006, this frequency increased to 8%; 2) nonetheless, these cases display improved honey yield, similar to the effect observed following a reduction in the number of hives. These findings indicate the scarcity of bee food, a reality that has forced beekeepers to move to the better areas in order to facilitate the improving of the management anywhere. It must be emphasised that Rio de Janeiro has abundant area to wild vegetation, without economic status as an agricultural producer. Livestock populations have been neglected for quite some time and are contributing to the deforestation of natural environments. These trends represent threats to Brazilian apiculture, dependent on native flora for survival.

The current state of beekeeping in Rio de Janeiro displays a troublesome lack of productivity, requiring thoughtful remediation. In this particular state, the beekeeping sector displayed robust growth during the 1950s and 1960s, but a lack of awareness regarding the limits of the field resulted in low profitability. Therefore, increasing numbers of beekeepers led to high losses, a low variety of bee products and declining honey yields, as reported in other Brazilian states with recently begun beekeeping efforts exhibiting quiet growth. This trend is typical for Brazilian agriculture; farmers receive inadequate financial support for expansion and development. Beekeeper education is essential to facilitate productive use of the resources available. Organized

beekeeping programs, as well as close and systematic monitoring, will be necessary to overcome the problems currently faced by the beekeeping sector.

## References


Brazilian Geography Institutte (FIBGE) (2005) Farming Survey, Rio de Janeiro, 1990-2005. [online] http: // www. sidra. ibge. gov.br /bda/tabela /protabl.asp?z= p&o=19&i=P.

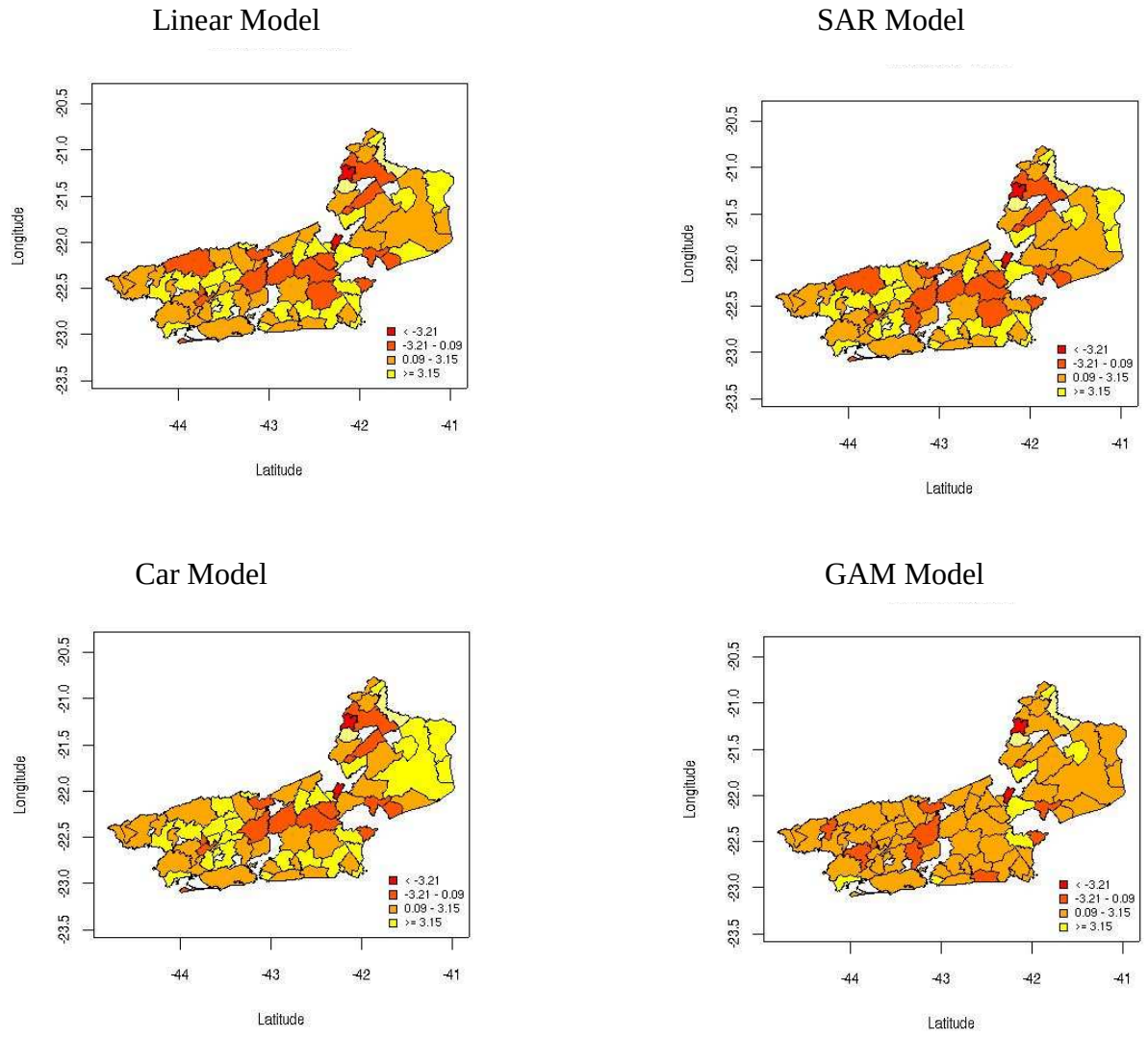
Venables W.N., Ripley B.D. (2004) Modern Applied Statistics with *S*, Springer, *G*ermany, Anselin, L., Under the hood: issues in the specification and interpretation of spatial regression models (forthcoming: Agricultural Economics ). [online] http: // agec 221. agecon. uiuc.edu/ users/ anselin/ papers/hood.pdf.

Bailey T.C., Gatrell A.C. (1995) Interactive spatial data analysis, Longman, England.

Cliff A.D., Ord J.K. (1981). Spatial processes: Models and applications. Pion Ltd., London.

R Development Core Team, (2007) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Austria. [online] http://www.R-project.org.

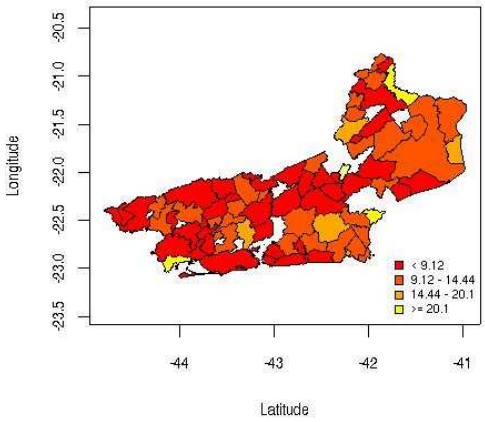
Druck S., Carvalho M.S., Câmara G., Monteiro A.V.M. (2004) Análise Espacial de Dados Geográficos, EMBRAPA, Brazil.

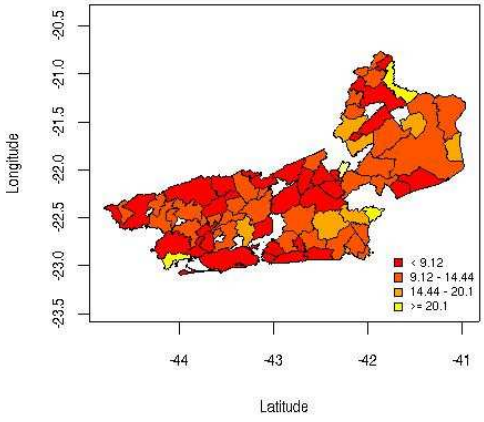

Source: MAPA, SESCOOP, FAERJ, Counties= 92.

**Figure 1**. Spatial distribution of the residuals from the regression models: spatial (SAR, CAR and GAM) and non-spatial (Linear).

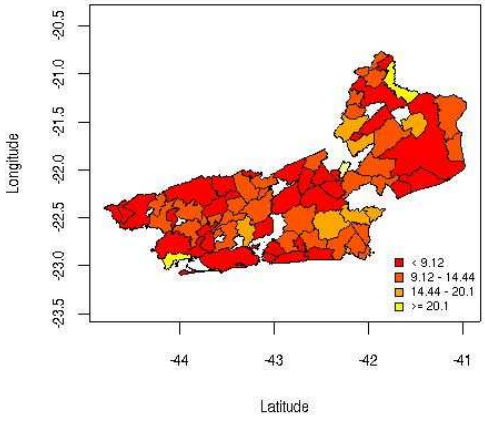
Linear Model



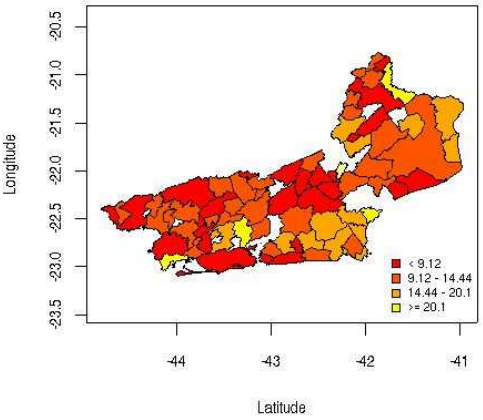

SAR Model




CAR Model




GAM Model




Source: MAPA, SESCOOP, FAERJ, Counties= 92.

**Figure 2**. Spatial distribution of the predicts values from honey yield per hive per year (kg), from Regression Models: spatial (SAR, CAR and GAM) and non-spatial (Linear).