Kernel Maps of the *Bactrocera carambolae* Occurrence in Amapá, Brazil

Paulina K. A. Santos[1], Marcos A. S. da Silva[2], José Victor T. A. Costa[3], Bruna H. S. A. da Silva[1], Cristiaini Kano[1], Márcia H. G. Dompieri[1]

[1]Embrapa Territorial
Av. Sd. Passarinho, 303 - Jardim Chapadão, 13070-115, Campinas, SP.

[2]Embrapa Coastal Tablelands
Av. Beira Mar, 3250, 49025-040, Aracaju, SE.

[3]Superintendência Federal da Agricultura do Amapá
Rua Tiradentes, Central, 68900-098, Macapá, AP

kayseane6@gmail.com
{marcos.santos-silva,marcia.dompieri,cristiaine.kano}@embrapa.br
jose.torres@agro.gov.br, bruna.silva@colaborador.embrapa.br

Abstract. *This work analyzes the spatial distribution of the Bactrocera carambolae occurrence between 2014 and 2019 in Amapá, Brazil. They monitored the quarantine pest present using traps distributed throughout the state. The technicians checked traps every fifteen days, recorded the occurrence, and sent the information to MAPA. To analyze the spatial distribution, we evaluated the degree of randomness of the traps using the non-homogeneous K function. Subsequently, we applied the quartic kernel function on the occurrences and the simulated occurrences, considering the non-homogeneous phenomenon and having the population per census tracts and highways as covariates. The results showed that we could use the method to monitor the pest on a regional scale.*

Resumo. *Este trabalho analisa a distribuição espacial da ocorrência de Bactrocera carambolae entre 2014 e 2019 no Amapá, Brasil. Os técnicos monitoraram a presença da praga quarentenária por meio de armadilhas distribuídas por todo o estado. Eles verificavam as armadilhas a cada quinze dias, registravam a ocorrência e enviavam as informações ao MAPA. Para analisar a distribuição espacial, avaliamos o grau de aleatoriedade das armadilhas usando a função K não homogênea. Posteriormente, aplicamos a função de kernel quártico nas ocorrências e nas ocorrências simuladas, considerando o fenômeno não homogêneo e tendo como covariáveis a população por setores censitários e rodovias. Os resultados mostraram que poderíamos usar o método para monitorar a praga em escala regional.*

1. Introduction

The pests are characterized as quarantine when there is a probability or possibility of entry, establishment, and dispersal in a particular region, and that brings with them significant economic risk [Fidelis et al. 2018, Silva et al. 2011]. In Amapá, the State Agricultural Defense and Inspection Agency (*Agência de Defesa e Inspeção Agropecuária do Estado*, DIAGRO) is responsible for the actions of combat, surveillance, prevention, and eradication of quarantine pests. The current programs on its official website involve efforts to control the quarantine pest Bactrocera carambolae, also known as the carambola fruit fly. Generally, pest detection, monitoring, and control methods consist of trapping and fruit sampling. Monitoring by traps on a regional scale aims to obtain information for developing plans for containment, suppression, and eradication of the pest [Silva et al. 2011].

Amapá divided its territory into zones from specific plans with different goals depending on the level of necessary pest control for effective control of the Bactrocera carambolae and the distribution of traps. They are: *Plano de Contenção*, *Plano de Erradicação do Sul*, and *Plano de Pós-Erradicação do Vale do Jari*. In 2017, a reformulation of these plans came into effect, keeping the same municipalities in the Plano de Contenção and *Plano de Pós-Erradicação do Vale do Jari*, the last renamed to *Plano de Erradicação*. *Plano de Erradicação de Bactrocera carambolae do Sul do Amapá* becomes *Plano de Supressão com vistas à Erradicação*.

This work aims to study the spatial pattern by kernel density maps of the occurrence of the pest Bactrocera carambolae from data obtained by the state monitoring carried out through traps installed throughout Amapá between the years 2014 to 2019. In the end, an inhomogeneous model is defined for the phenomenon, enabling the data obtained to be compared with simulations using the model to identify occurrences that are higher than expected. The following section presents a brief review of the subject, section 3 presents the studied area and the methods, section 4 shows the results and discussion, and section 5 the conclusions.

2. Related work

The majority of works on spatial sampling for pest monitoring address it at farm scale [Silva et al. 2019, Valente 2018, Dionísio et al. 2015, Matrangolo et al. 2014]. With the exception of Dionísio et al. [2015], who used geostatistical methods to analyze the spatio-temporal pattern of the occurrence of the pest Metamasius hemipterus in oil palm plantation, the others used the methods of analysis of the ratio of variance to mean, morisita index and analysis of the k parameter for the negative binomial probability distribution to assess whether the pests occur in an clustered pattern using the host as the unit of count. In all cases, even in those with low incidence, the authors registered a strong aggregation of pest occurrences.

Valente [2018] and Matrangolo et al. [2014] developed a sampling plan seeking to identify the minimum number of samples for the spatial analysis of pest occurrences, considering pest per host. In these cases, there was no need to consider the condition of complete spatial randomness of the traps, since no methods with this requirement were used. On the other hand, Dionísio et al. [2015] opted for the geostatistical analysis of the occurrences by randomly distributing 24 sampling units for analysis. The authors concluded that infestation occurs initially at the edges of the plantation, with subsequent dissemination to the entire area. The authors then suggest that the traps showed be placed on the edges of the plantation, for sampling and control of the pest.

Spatial trap distribution affects pest monitoring [Berec et al. 2015]. It could follow a regular, random, sparse, or even concentrated in a particular region with a greater probability of pest occurrence. However, we must distribute the traps randomly to analyze the pest's spatial distribution with kernel maps. The exploratory spatial distribution of the occurrence of the pest has been conducted by the use of the Kernel Density Estimator (KDE) because it is intuitive, simple to apply and generate a conservative estimation of hot-spots [Lin et al. 2011, Kitthawee and Dujardin 2010]. When conducting a KDE study, the key points are avoiding sampling bias (e.g., clustered traps tend to generate clustered hot spots) and the kernel function's bandwidth empirical definition [Wallner et al. 2014, Kumar et al. 2014].

3. Material and Methods

To assess whether the traps were spatially randomly distributed, we conducted the tests of Complete Spatial Randomness (CSR) and non-homogeneously with the support of the K function. We applied the tests by region covered by each control plan in the anthropized area for each monitoring period at intervals of 15 days for each year. The non-homogeneity tests considered the population by census sector and roads as covariates.

3.1. Hypothesis test of spatial non-homogeneity of traps and the K function

CSR test and inhomogeneous test assume basic conditions based in a poisson distribution: the events are independent and its distribution follows an intensity λ. In CSR, the intensity λ is constant in the entire area, however, in the inhomogeneous poisson process, intensity can vary spatially $f(\lambda)$, meaning that the events become more likely to happen in some areas than others according to some covariates [Waagepetersen 2008, Hahn et al. 2003].

Since the occurrence of the pest is related to regions of intense economic and social activity where there is consumption and transport of fruit [Silva et al. 2011], we have chose the the population by census sector and the set of roads in the state as covariates. When considering a covariate in a point event of coordinate (x, y), the intensity function resembles $\lambda(x, y) = exp(\alpha + \beta Z(x, y))$, where α and β are parameters to be adapted by the point event [Bivand et al. 2008]. The K function allows the evaluation of whether there is spatial dependence, from the observation of the density of points in various ranges of distances. The null hypothesis is that the distribution is random and independent. The K function is defined by $K(r) = \lambda^{-1}E(number\ of\ events\ contained\ within\ a\ distance\ of\ an\ arbitrary\ event\ )$, where λ is the intensity or intensity or the number of events expected per unit in the area [Bivand et al. 2008]. Considering the inhomogeneous test, where intensity is defined by a function, and edge effects, the estimated function is:

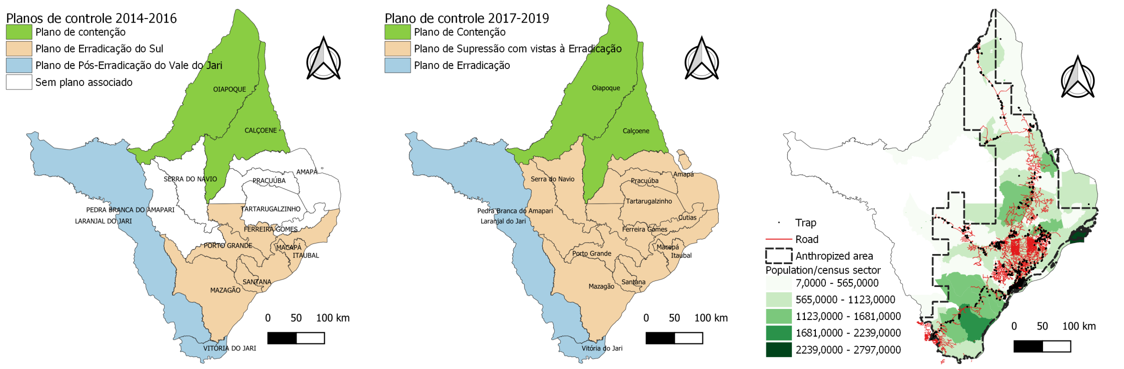
$$\hat{K}(r) = \frac{1}{A}\sum_{i}^{n}\sum_{j\neq i}^{n} w_{ij}^{-1}\frac{I_r(d_{ij})}{\lambda(x_i)\lambda(x_j),} \tag{1}$$

where A: area of the region; n: number of events; $I_r(d_{ij})$ follows as an indicator function regards to the distance between the points and the chosen distance r. w_{ij}^{-1} signaling the edge correction; $\lambda(x_i)$ and $\lambda(x_j)$ returning the intensity at that point mapped by the function $f(\lambda)$ [Bivand et al. 2008].

For inference and better visualization of the meaning of the resulting value of the formula, the value of $\hat{K}(r)$ under the assumption of complete randomness is generated for comparison, which results in $\hat{K}(r) = \pi r^2$. It is also common to include the lower and upper envelopes, created by several also random simulations of n events in region A. In this work we created confidence bands for 99 simulations and considering a confidence level of 98%. If the resulting value of $\hat{K}(r)$ is greater than the random simulation created and its envelope, $\hat{K}(r) > \pi r^2$, then the distribution exhibits clustering pattern. If it is smaller, $\hat{K}(r) < \pi r^2$, it exhibits dispersion pattern [Hohl et al. 2017].

3.2. Amapá's pest control plans

Amapá has divided its territory into three zones defined by its pest control plans. Fig. 1 shows these plans through two periods (2014-2016 and 2017-2019).



(a) Plans for 2014 to 2016 (b) Plans for 2017 to 2019 (c) Anthropized zone & covariates

Figure 1. Plans for the Bactrocera carambolae controle in Amapá, Brazil, for 2014-2016 (a), and 2017-2019 (b). Anthropidez and anthropizable zone used as the studied cut, traps spatial distribution, population densisity and roads (c). Source: municipality boundaries (IBGE 2019), population by census sector (IBGE Census 2010). Anthropized zone and roads (SEMA/AP). Maps in UTM/Zone 22N projection and SIRGAS 2000 datum, elaborated by the authors.

According to Silva et al. [2011], the pest enter the country mainly through imports of contaminated fruits and spread through transport to consumer centers. Thus, we identified at least two covariates that can influence the occurrence of the pest Bactrocera carambolae, changing the average spatial occurrence, the routes of transport of merchandises and the population per census sector (Fig. 1(c)). The analysis considered the anthropized and/or anthropizable area as the studied region (Fig 1(c)). The 16 municipalities have the Jackson and McPhail trap types, distributed in order to have a density of 0.4 Jackson traps/ha and 0.2 McPhail traps/ha monitored fortnightly. In regions not covered by the plans, local authorities and the population are responsible for signaling the appearance of focuses where the plague has not yet reached.

3.3. Kernel map

The kernel map can also validate the point distribution analysis by composing a surface proportional to the point intensity. Starting from a generic location of interest, a radius is defined and the points within this region are weighted. Therefore, it is possible to check the distribution of points in a visual way and analyze if the distribution presents intensity differences on its surface [Bivand et al. 2008]. The function chosen as estimator was the quartic function:

$$k(\frac{d_{is}}{r}) = \begin{cases} (1 - \frac{d_{is}^2}{r^2})^2, & if \;\; 0 < d_{is} \leq r \\ \quad 0 & , \;\; otherwise \end{cases} \tag{2}$$

where r is the bandwidth and $k(\frac{d_{is}}{r})$ is the weight of a point i at distance d_{is} to location of interest s.

The traps were distributed in a concentrated way, which impairs the interpretation of hot spots. However, we can consider that the phenomenon is non-homogeneous and that its occurrence depends, at least, on the population and the roads. Thus, one way to compensate for this aggregation of traps is to generate a false non-homogeneous distribution of traps. The model that represents the phenomenon is created by adjusting the data obtained from the 15-day window with covariates. The coefficients generated by the model identify the lambda function of the inhomogeneous process, and then the random point simulation is carried out. Then, we subtract the observed hot spot from this simulated one. The hot spot of this difference would be an approximation of the hot spot discounting the aggregation of traps.

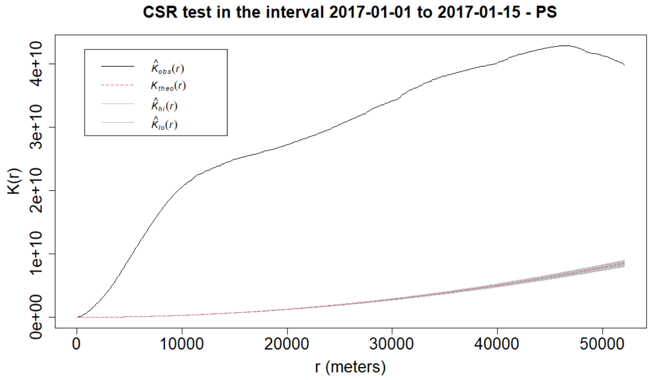
We performed the inhomogeneous model and the data simulation using the spatstat R package (version 3.0) and generated the maps with the SpatialKDE R package (version 0.8.2) in the R language (version 4.3.0).

4. Results and Discussion

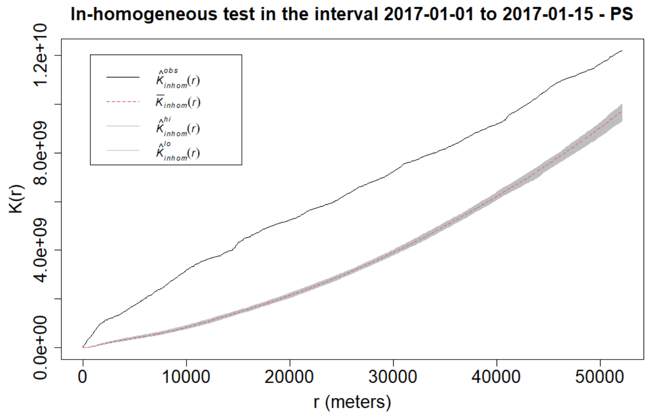
Fig. 2 shows the K curves for the CSR and inhomogeneous tests for a specific period of monitoring in the region covered by the *Plano de supressão com vistas à erradicação*. We conclude by the CSR test that the traps are clusterd (Fig. 2(a)), but if we consider that our phenomena is inhomogenous and considers population and roads as covariates the aggregation is significantly alleviated (Fig. 2(b)). Its suggests that the traps distribution followed a inhomogenous distribution.

Figs. 3-5 show the occurrences kernel maps of observed, simulated, and the difference between them for 15-days monitoring for the years 2019 (*plano de contenção*), 2014 (*plano de supressão*), and 2017 (*plano de erradicação*). From the observed and simulated kernel maps, we see differences between them, suggesting that in some regions, the occurrence exceeded expectations in all zones. We could not plot all maps, but this phenomenon is observed in all analyzed periods.

In the *plano de contenção* zone, considering the year 2019, the difference between observed and simulated shows slight variation, highlighting the border region with French Guiana as the main hot spot (Fig. 3(c)). In the *plano de supressão* region in 2014, there is a great difference between the observed and simulated maps in Macapá (Fig. 4(c)). In the region covered by the *plano de erradicação* region, 2017, the difference showed distinct regions where the observed exceeded the simulated (Fig. 5(c)). In general, as shown in the short literature review, the Bactrocera carambolae presents an aggregated spatial pattern of occurrence.



(a) CSR test: $\hat{K}_{obs}$ represents the observed, $\hat{K}_{theo}$ the theoretical curve, and superior and inferior bands ($\hat{K}_{hi}$ and $\hat{K}_{lo}$, respectively).



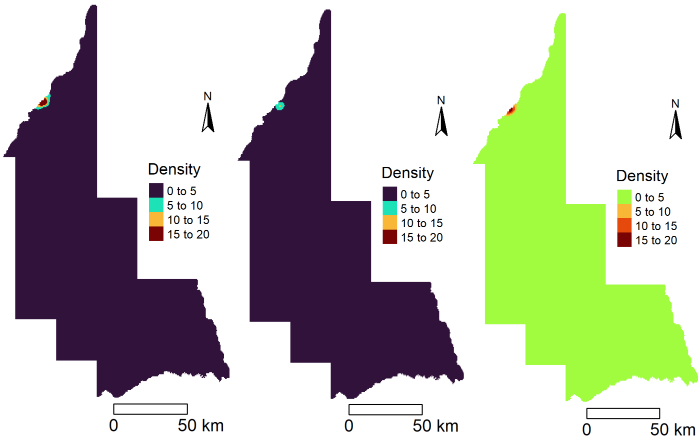
(b) K for inhomogeneous test: $\hat{K}^{obs}_{inhom}$ represents the observed, $\overline{\mathrm{K}}_{inhom}$ the mean curve, and superior and inferior bands ($\hat{K}^{hi}_{inhom}$ and $\hat{K}^{lo}_{inhom}$, respectively).

Figure 2. CSR versus in-homogeneous (population and roads as covariates) K function graphic for the *Plano de supressão com vistas à erradicação* for the period between 2017-01-01 to 2017-01-15. Source: elaborated by the authors.

The proposed method aims to eliminate the effect of the aggregate disposition of the traps by considering this aggregation as a consequence of the non-homogeneous process. In the difference, we identify hot areas beyond what we predicted. Economic activity and roads are crucial for pest establishment. In our case, we used the population by census sector as a proxy of economic activity [Wallner et al. 2014].

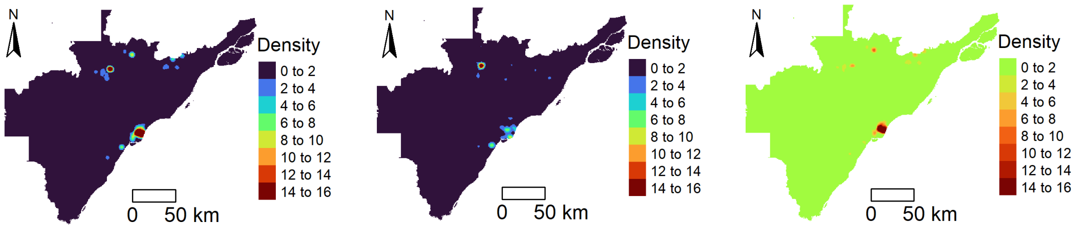
5. Conclusions

The kernel map produced by the difference between the observed and simulated maps revealed only certain areas with occurrence higher than expected, compensating the aggregated distribution of traps and suggesting an alternative for state monitoring of the pest. Future works may consider improving the method by considering more covariates for the non-homogeneous model, reducing the analyzed regions to cities or neighborhoods, evaluating other kernel functions, and producing marked kernel maps considering the number of captures in the traps.



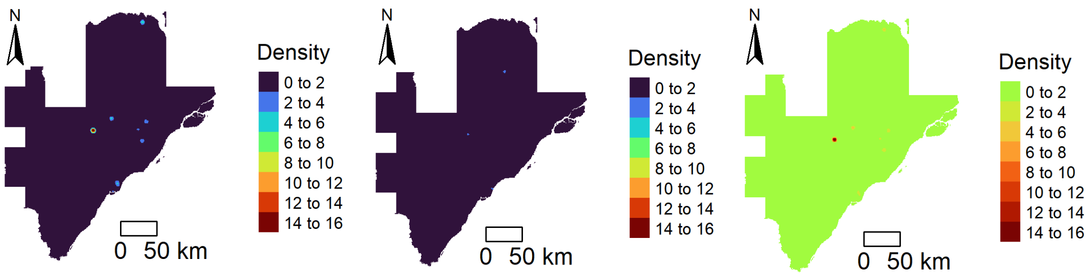
(a) Observed (O) (b) Simulated (S) (c) $Diff = O - S$

Figure 3. Kernel analysis for the *Plano de contenção* for the period 01-01-2019 to 15-01-2019



(a) Observed (O) (b) Simulated (S) (c) $Diff = O - S$

Figure 4. Kernel analysis for the *Plano de supressão com vistas a erradicação* for the period 01-01-2014 to 15-01-2014



(a) Observed (O) (b) Simulated (S) (c) $Diff = O - S$

Figure 5. Kernel analysis for the *Plano de erradicação do sul* for the period 01-12-2017 to 15-12-2017

Acknowledgments

This work was supported by the National Council for Scientific and Technological Development – CNPq, Brazil.

References

Berec, L., Kean, J. M., Epanchin-Niell, R., Liebhold, A. M., and Haight, R. G. (2015). Designing efficient surveys: spatial arrangement of sample points for detection of invasive species. *Biological Invasions*, 17:445–459.

Bivand, R. S., Pebesma, E., and Gómez-Rúbio, V. (2008). *Applied Spatial Data Analysis with R*. New York, USA.

Dionísio, L. F. S., Lima, A. C. S., Morais, E. G. F. d., Correia, R. G., Santos, A. V. F. d., and Ximenes, C. K. d. S. (2015). Distribuição espacial de metamasius hemipterus (coleoptera: Curculionidae) em plantio de dendê (elaeis guineensis jacq) em roraima. *Rev. Agroambiente On-Line*, 9(3):327–366.

Fidelis, E. G., Lohmann, T. R., da Silva, M. L., Parizzi, P., and Laranjeira, F. F., editors (2018). *Priorização de Pragas Quarentenárias ausentes no Brasil*. Brasília, DF.

Hahn, U., Jensen, E. B. V., Lieshout, M.-C. V., and Nielsen, L. S. (2003). Inhomogeneous spatial point processes by location-dependent scaling. *Advances in Applied Probability*, 35(2):319–336.

Hohl, A., Zheng, M., Tang, W., Delmelle, E., and Casas, I. (2017). *Spatiotemporal Point Pattern Analysis Using Ripley's K Function*, chapter VII, pages 155 – 176. CRC Press, 1st edition.

Kitthawee, S. and Dujardin, J.-P. (2010). The geometric approach to explore the bactrocera tau complex (diptera: Tephritidae) in thailand. *Zoology*, 113:243–249.

Kumar, S., Neven, L. G., and Yee, W. L. (2014). Evaluating correlative and mechanistic niche models for assessing the risk of pest establishment. *Ecosphere*, 5(7):86.

Lin, Y.-P., Chu, H.-J., Wu, C.-F., Chang, T.-K., and Chen, C.-Y. (2011). Hotspot analysis of spatial environmental pollutants using kernel density estimation and geostatistical techniques. *Int. J. Environ. Res. Public Health*, 8:75–88.

Matrangolo, C. A. R., Anjos, N. d., Leite, H. G., and Marcatti, G. E. (2014). Distribuição espacial dos danos de heilipodus naevulus em plantio de clones de eucalipto. *Arquivos do Instituto Biológico*, 81(2):119–125.

Silva, E. J. P. d., Souza, A. A. d., Oliveira, A. C. S. d., Neves, J. D. S. d., Santos Neto, A. L. d., Negrisoli Junior, A. S., and Souza, A. D. G. (2019). Distribuição espacial da broca-do-estipe-do-coqueiro. *Nativa*, 7(4):356.

Silva, R., Zucchi, R., and Llemos, W., editors (2011). *Moscas-das-frutas na Amazônia brasileira: diversidade, hospedeiros e inimigos naturais*. Macapá, Brasil, 1st edition.

Valente, F. I. (2018). *Distribuição espacial e plano de amostragem sequencial de Conotrachelus psidii (Marshall, 1922) (Coleoptera: Curculionidae) na cultura da goiabeira*. PhD thesis, Universidade Federal da Grande Dourados.

Waagepetersen, R. (2008). Estimating functions for inhomogeneous spatial point processes with incomplete covariate data. *Biometrika*, 95(2):351–363.

Wallner, A. M., Hamilton, G. C., Nielsen, A. L., Hahn, N., Green, E. J., and Rodriguez-Saona, C. R. (2014). Landscape factors facilitating the invasive dynamics and distribution of the brown marmorated stink bug, Halyomorpha halys (Hemiptera: Pentatomidae), after arrival in the United States. *Plos One*, 9(5):1–12.