ISSN 2175-6813

Revista Engenharia na Agricultura

Viçosa, MG, DEA/UFV - DOI: 10.13083/reveng.v30i1.12919 v.30, p.63-74, 2022

ANALYSIS OF PEST INCIDENCE ON APPLE TREES VALIDATED BY UNSUPERVISED MACHINE LEARNING ALGORITHMS

Eduardo Antonio Speranza[1]*, Célia Regina Grego[1] & Luciano Gebler[2]

1 - Embrapa Digital Agriculture, Campinas, São Paulo, Brazil

2 - Embrapa Grape and Wine, Bento Gonçalves, Rio Grande do Sul, Brazil

Keywords:
Pest management
Geostatistics
Unsupervised Machine Learning
Orchards
Apples

ABSTRACT

Integrated pest control is a practice commonly used in apple orchards in southern Brazil. This type of management is an important tool to help improve quality and increase yields. This study aimed to identify areas with higher and lower incidence of aerial pests in a commercial apple orchard, regarding data collected from three different crops using georeferenced traps. Geostatistical analyses were performed, based on the modeling of semivariograms and spatial interpolation using the kriging method; and clustering, based on specific unsupervised machine learning algorithms for count data. The algorithms were selected from measures of stability, connectivity and homogeneity, seeking to identify areas with different incidence of pests that could help farmer decision making regarding insect population control using pesticides. The geostatistical analysis verified the presence of individual pest infestations in specific sites of the study area. Additionally, the analysis using machine learning allowed the identification of areas with incidence above the average for all analyzed pests, especially in the central area of the map. The process of evaluation described in this study can serve as an aid for risk analysis, promoting management benefits and reducing cost in the farms.

Palavras-chave:
Manejo de Pragas
Geoestatística
Aprendizado de Máquina Não-Supervisionado
Pomares
Maçãs

RESUMO

O controle integrado de pragas é uma prática comumente utilizada em pomares de macieira no sul do Brasil. Esse tipo de manejo é uma importante ferramenta para ajudar a melhorar a qualidade e aumentar a produtividade. Este estudo teve como objetivo identificar áreas com maior e menor incidência de pragas aéreas em um pomar comercial de macieiras, a partir de dados coletados em três diferentes cultivos utilizando armadilhas georreferenciadas. Foram realizadas análises geoestatísticas, com base na modelagem de semivariogramas e interpolação espacial pelo método de krigagem; e agrupamento, com base em algoritmos de aprendizado de máquina não supervisionados específicos para dados de contagem. Os algoritmos foram selecionados a partir de medidas de estabilidade, conectividade e homogeneidade, buscando identificar áreas com diferentes incidências de pragas que pudessem auxiliar na tomada de decisão do agricultor quanto ao controle populacional de insetos utilizando agrotóxicos. A análise geoestatística verificou a presença de infestações individuais de pragas em locais específicos da área de estudo. Adicionalmente, a análise por aprendizado de máquina permitiu a identificação de áreas com incidência acima da média para todas as pragas analisadas, principalmente na área central do mapa. O processo de avaliação descrito neste estudo pode servir de auxílio para análise de risco, promovendo benefícios de manejo e redução de custos nas propriedades.

Received 29 Jul, 2021 • Accepted 1 Dec, 2021 • Published 27 Apr, 2022

Email (corresponding author)*: eduardo.speranza@embrapa.br

SECTION EDITOR IN CHARGE

Gabriel Araújo e Silva Ferraz

INTROUCTION

The commercial production of apple cultivars has increased significantly in recent years, especially in the states of Santa Catarina and Rio Grande do Sul, in Brazil. Brazilian production is self-sufficient to fulfil the domestic consumption and exports to highly demanding markets. In this scenario, apple-producing companies are moving towards efficient production with quality and sustainability, minimizing the negative effects on health and the environment by reducing the use of pesticides (FIORAVANÇO & DOS SANTOS, 2013).

The increase in yield and demand for fruit quality created the need of a system to assist in the management of the areas, resulting in the development of the Integrated Apple Production (IAP). IAP is a quality certification program that required the implementation of a system to support decision-making on the use of pesticides based on pest monitoring (KOVALESKI & RIBEIRO, 2002). In this context, in the states of Santa Catarina and Rio Grande do Sul, which currently account for about 97% of the Brazilian apple production (LAZZAROTTO, 2018), most companies adopt the IAP system (PROTAS & SANHUEZA, 2002). The IAP system is considered even if, due to the cost of the third-party certification phase, few producers complete the process with the certificate, despite following the entire technical protocol.

AIP provides a dynamic update on the incidence of pests in the crop area, favoring the control of populations, and requires farmers to store their monitoring data for a minimum period of five years, if there is a need for extra audits by the program. Based on this requirement, farmers have the availability of geographic and temporal data sheets in their orchards (FIORAVANÇO & DOS SANTOS, 2013). These are normally used weekly to manage pests during the crop year, but due to lack of tools or guidance, they end up without further use, being stored both physically (field notebooks) and virtually (electronic spreadsheets).

Installing sampling traps in specific locations in each crop allows the assessment of pest incidence in an apple-growing field. Information over the years can identify the most susceptible areas and establish the suitable management to control the pest population. Consequently, these actions can favor investment in areas where it is possible to obtain higher quality fruits.

There are numerous reports in the literature on the use of machine learning algorithms to predict pest and disease infestation in agriculture. In fruit growing, especially apples, these investigations are still recent. Boniecki *et al*. (2015) developed a model based on artificial neural networks that allows the identification of six distinct pest species of apple trees, based on the input of 16 variables related to color and 7 variables related to shape, obtained through images (photos). Wrzesień *et al*. (2019) used a supervised machine-learning algorithm, relying on the input of standard meteorological data, to simulate the results of physical sensors installed in trees to help to detect apple scab disease. Despite providing promising results in the simulation to replace physical sensors, the experiments were performed on a single tree, thus further studies are needed so the application can gain scale. Brunelli *et al*. (2019) developed an internet of things (IoT) and machine learning system for detecting the apple moth *(Cydia pomonella*) pest in orchards. The sensor developed is powered by solar energy and installed on the trees, which can collect and pre-process images of the insects, allowing the *in situ* identification of the occurrence of the pest and sending messages to the farmer via data network. Due to the low energy and cost of components, analyses carried out showed that the system is viable and easily scalable to be installed in different places in the orchard.

Besides the above applications, we can obtain spatial pest distribution maps from data collected from traps installed at specific locations in the crop. As these data are not previously labeled, the classification of these maps, in relation to different pest incidence classes, can be performed using unsupervised machine learning algorithms, also known as clustering. These algorithms have as main objective to cluster data samples in a natural way, they use the knowledge intrinsic in the data and heuristics to allocate more similar sample pairs in relation to the application domain in the same cluster; and pairs of less similar samples into distinct clusters (JAIN & DUBES, 1988; WITTEN *et al.*, 2011).

The objective of this study was to identify, from the incidence of different types of aerial pests over three crop years in a commercial apple orchard located in Vacaria, Rio Grande do Sul, Brazil, areas with different levels of susceptibility to pests. In this respect, we used methods to identify the spatial variability and specific clustering algorithms for count data.

MATERIALS AND METHODS

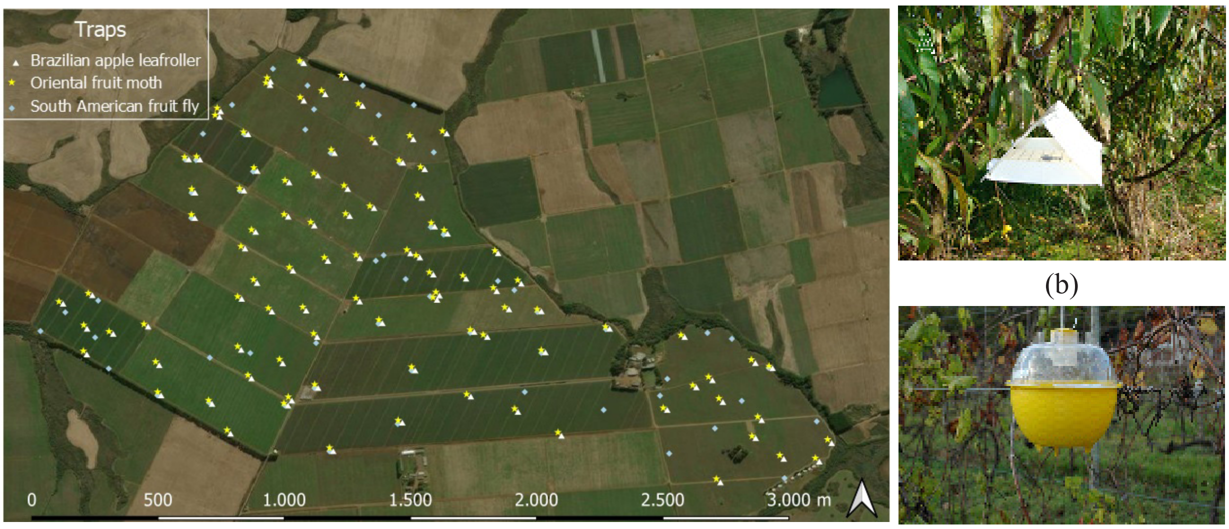
Study area and data collection

The dataset used in this work were obtained from a commercial apple orchard (*Malus domestica*) belonging to Agropecuária Schio LTDA (Vacaria, Rio Grande do Sul, Brazil: 50’49’13’’O; 28’28’43’’S). The area comprised about 305 ha planted with cultivars Gala, Fuji and their respective clones. The company follows specific Integrated Pest Management (IPM) standards, applying pesticides only where the level of pests obtained, according to the information collected from the traps, is high enough to justify the operation. Two hundred forty-eight traps georeferenced using GPS RTK (Figure 1a) were available to capture the following pests: Brazilian apple leafroller (*Bonagota Salubricola*) and Oriental fruit moth *(Grapholita molesta*), with 94 traps each (Figure 1b); and South American fruit fly (*Anastrepha fraterculos*), with 60 traps (Figure 1c).

The occurrence counts performed in the traps referenced in Figure 1a refer to the 2011/2012, 2012/2013 and 2013/2014 crop years. Candeia *et al*. (2016) performed a binarization of this same dataset in their work, assigning value 1 to samples with a number above the minimum and zero for samples with a number below the minimum number of occurrences defined for the use of pesticides, according to the norms of the IPM used by the company. However, in this study, we used the absolute count to enable the application of data clustering algorithms in the analyses.

Geostatistical Analysis

First, using pest incidence data, we performed a geostatistical analysis to verify the existence of spatial dependence by calculating and adjusting the scaled semivariograms to the best correlation functions. The scaling of semivariograms was according to Vieira *et al*. (2010), with the purpose of modeling all semivariograms on the same semivariance and distance scale for all variables having the same measurement unit and were sampled in the same study area. The parameters nugget effect (Co), structural variance (C1) and range (a) were used in the interpolation by



(a) (b) (c)

Figure 1. Study area (a) and location of traps for pest species Brazilian apple leafroller (represented by triangles), Oriental fruit moth (represented by stars) and South American fruit fly (represented by diamonds); and types of traps used in the field to control Brazilian apple leafroller, Oriental fruit moth (b) and South American fruit fly (c)

kriging. The scaled semivariograms were fitted to the spherical model, which is more suitable for the dataset, presenting two variance structures to express the spatial dependence: in the first, there was an increase in semivariance as a function of the distance to range; and in the second, a semi-variance plateau was reached. From the fitted semivariograms, the dataset were interpolated using ordinary kriging with the adjustment parameters. Maps for each season were constructed using the QGIS software (QGIS, 2021).

Unsupervised machine learning

The geostatistical analysis performed allowed to identify the presence of spatial variability in the study area in relation to the incidence of pests for the data collected in the three referenced crop years. The analysis using machine learning allowed to verify possibilities of partitioning the study area into areas with higher and lower incidence of pests, taking as input the data collected in the three crop years.

This work uses samples of a specific type of data as input, as well as in the geostatistical analysis, specific clustering algorithms were used for count data (non-continuous). The *optClust* algorithm, a package of the same name and available in the R software (SEKULA *et al*., 2017), was used to identify which would be the ideal algorithm to generate incidence maps for each pest. This function allows verifying, from the data and statistical techniques that identify stability, connectivity and homogeneity indices, which is the best algorithm for clustering the samples, considering sixteen clustering algorithms. Of this set, six algorithms are specific to count data and were selected for this work, based on the definition of the *countData=TRUE* parameter. In addition to the algorithm, this function also suggests an optimal amount of clusters to classify the dataset. In this case, we considered two to five clusters.

Because this is a statistical and non-deterministic methodology, thirty runs of the *optCluster* function were performed for each pest incidence dataset. Thus, the choice of the combination algorithm and number of clusters, for each case, regarded the number of times this combination appears among the thirty tests performed. The deterministic annealing algorithms (UEDA & NAKANO, 1998) and simulated annealing (VAN LAARHOVEN & AARTS, 1987), both using a negative binomial model for data counting and available in the *MBCluster.Seq* package of the R software (SI *et al*., 2014), were those showing the greatest adherence to the data sets used. These algorithms are variants of the Expectation Maximization (EM) algorithm (MOON, 1996), and because they use traditional models of probabilistic distributions for count data, they are suitable for simulating clusters for pest incidence datasets. However, although these algorithms start with random centroids, at some point the result may converge to a local rather than a global minimum. In this way, we executed each algorithm chosen for each data set at least ten times and selected the result that presented the smallest sum of squared errors, that is, the smallest sum of the distances of each sample in relation to the center of the cluster to which it was associated.

The comparison of the maps obtained per crop using geostatistics techniques with the final pest incidence map obtained from cluster analysis used the Kappa coefficient statistics (MCHUGH, 2012). This coefficient works as an external validation criterion in cluster analysis, i.e., to assess the agreement between labels assigned to two different solutions. The Kappa coefficient returns values between zero and one, where values close to one indicate high agreement between the clusters compared; and values close to zero indicate low agreement.

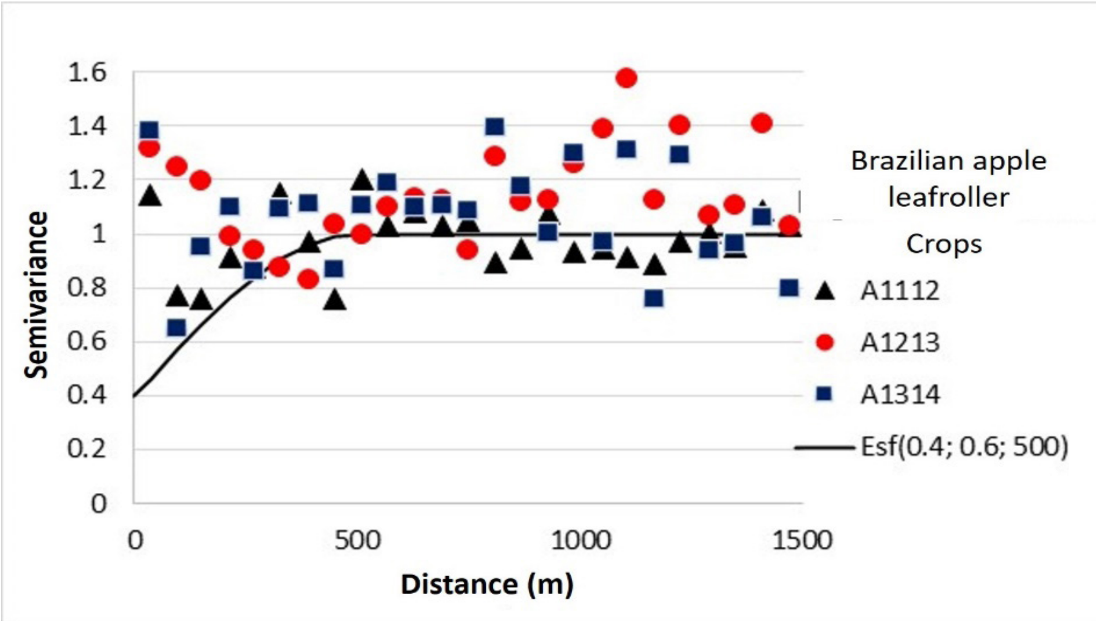
RESULTS AND DISCUSSION

Table 1 presents the minimum, maximum and median incidence values of the three pests in each of the three crop years, as well as their coefficients of variation and standard deviations of the data collected by the traps spread throughout the study area. These data show difficulties in the control of the pest Brazilian apple leafroller and, especially, the pest South American fruit fly, with an increase in incidence over the years. The pest Oriental fruit moth showed a small increase in incidence in the 2012/2013 crop year, which practically did not influence the result of this experiment.

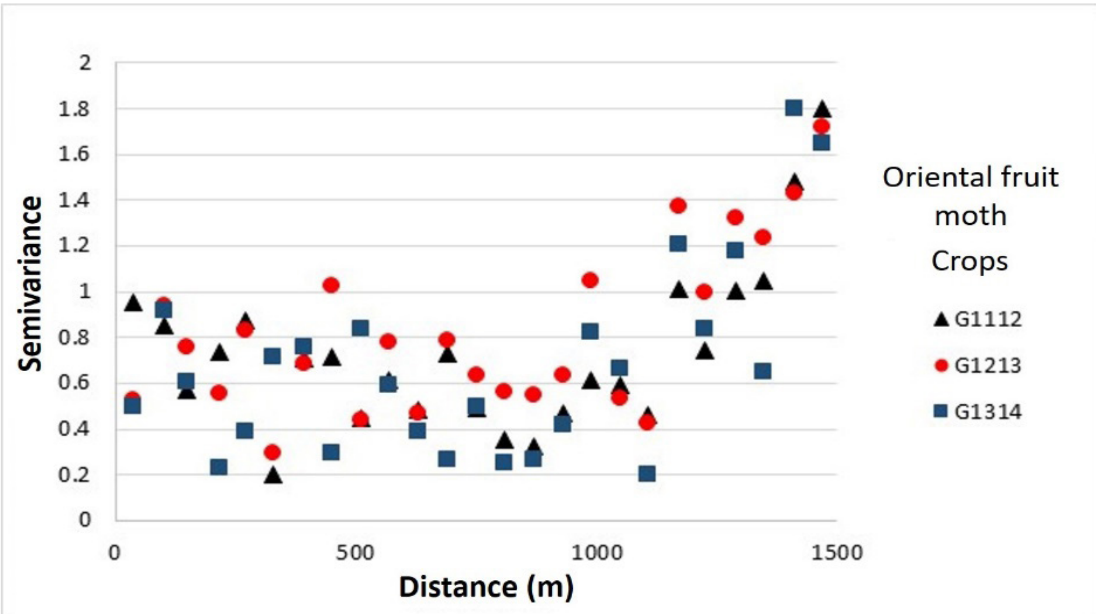
Figure 2 shows the scaled semivariograms generated for Brazilian apple leafroller, Oriental fruit moth and South American fruit fly regarding the three studied crop years. Spatial dependence was identified for Brazilian apple leafroller and South American fruit fly (Figure 2) with spherical adjustments.

Table 1. Minimum (Min), maximum (Max), median (Med), standard deviation (SD) and variation coefficient (VC) values (in %) from incidence of the three pests in the three crop years evaluated

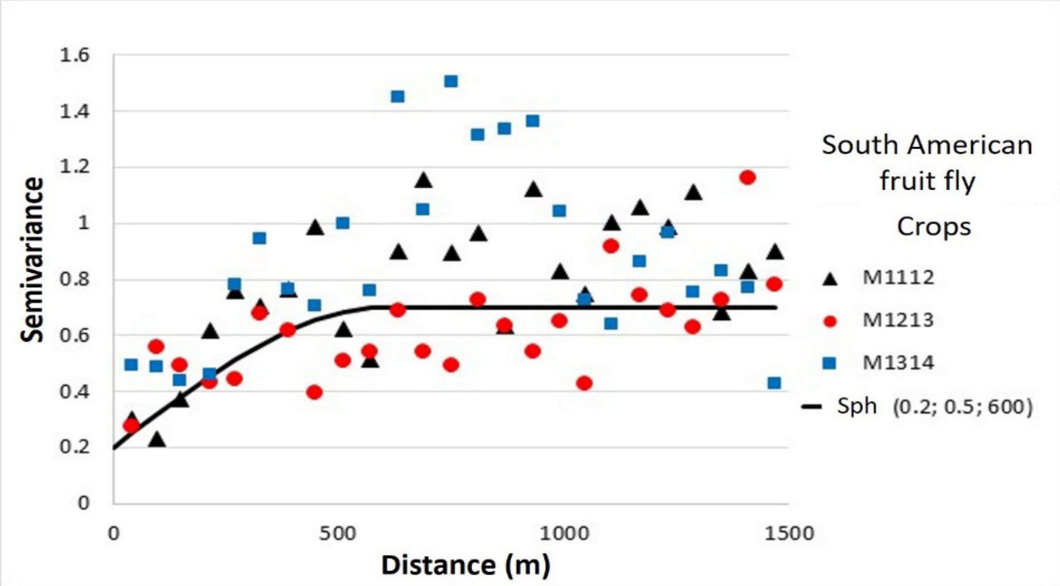
Pest / Crop	2011/2012					2012/2013					2013/2014				
year	*Min*	*Max*	*Med*	*SD*	*VC*	*Min*	*Max*	*Med*	*SD*	*VC*	*Min*	*Max*	*Med*	*SD*	*VC*
Barazilian apple leafroller	0	2	0	0.56	132	0	12	0	2.75	186	0	9	1	2.12	134
Oriental fruit moth	0	2	0	0.34	451	0	6	0	1.05	329	0	1	0	0.02	682
South American fruit fly	2	12	7	141.50	564	5	19	9.5	154.02	513	1	23	6	167.42	592



(a)



(b)



(c)

Figure 2. Scaled semivariograms for pests: a) Brazilian apple leafroller, b) Oriental fruit moth and c) South American fruit fly, for the 2011/2012, 2012/2013 and 2013/2014 crop years. Spherical adjustments for Brazilian apple leafroller and South American fruit fly in the three crop cycles

The spherical adjustment parameters of dependence for Brazilian apple leafroller (Figure 2a) were $Co=0.4$, $C1=0.6$ and $a=500$ *m*; and for South American fruit fly (Figura 2b) were $Co=0.2$, $C1=0.5$ and $a=600$ *m*. No adjustment was performed for Oriental fruit moth since no increase in semivariance was observed with increase in distance (Figure 2b), thus no occurrence of spatial dependence. Figure 3 shows the maps created from values interpolated by kriging for the variables presenting spatial dependence.

A higher incidence of Brazilian apple leafroller was found in the central area of the experimental field, of the mapped crops (Figure 3a, 3b and 3c). The maps of evolution of South American fruit fly showed a predominance of higher incidence in the lower right portion of the field (Figures 3d, 3e and 3f). This indicates that the occurrence of these pests in these areas must be managed or investigated differently from the other areas in the field. Dinardo – Miranda *et al*. (2007) studied the spatial distribution of sugarcane spittlebug (*Mahanarva fimbriolata*) in sugarcane plantations and found aggregated distribution models considering differences in this distribution according to the time of pest development. At the beginning of the period of occurrence, it was not possible to determine the spatial dependence between the samples, which was carried out only from the second generation of the pest.

Working with localized control of stinkbugs in the soybean crop, Roggia *et al*. (2021) used geostatistics to map the uneven distribution of the stinkbug in the crops. This required the localized application of insecticides to avoid waste and reach the target that really need the pest control product, reducing costs and improving the usability of the process. The delineation of the map assumes that the distribution of the pest in the field does not occur randomly, but follows a distribution dependent on the space and time of each evaluation performed in the field. The authors performed mathematical calculations using the georeferenced data, generating interpolated data with better precision for the mapping, which guides decision making for the pest control in the field.

To obtain maps with different areas of pest incidence, data of the three crops for the pests Brazilian apple leafroller and South American fruit fly were clustered using the algorithms of deterministic annealing and simulated annealing, respectively, both based on the negative binomial

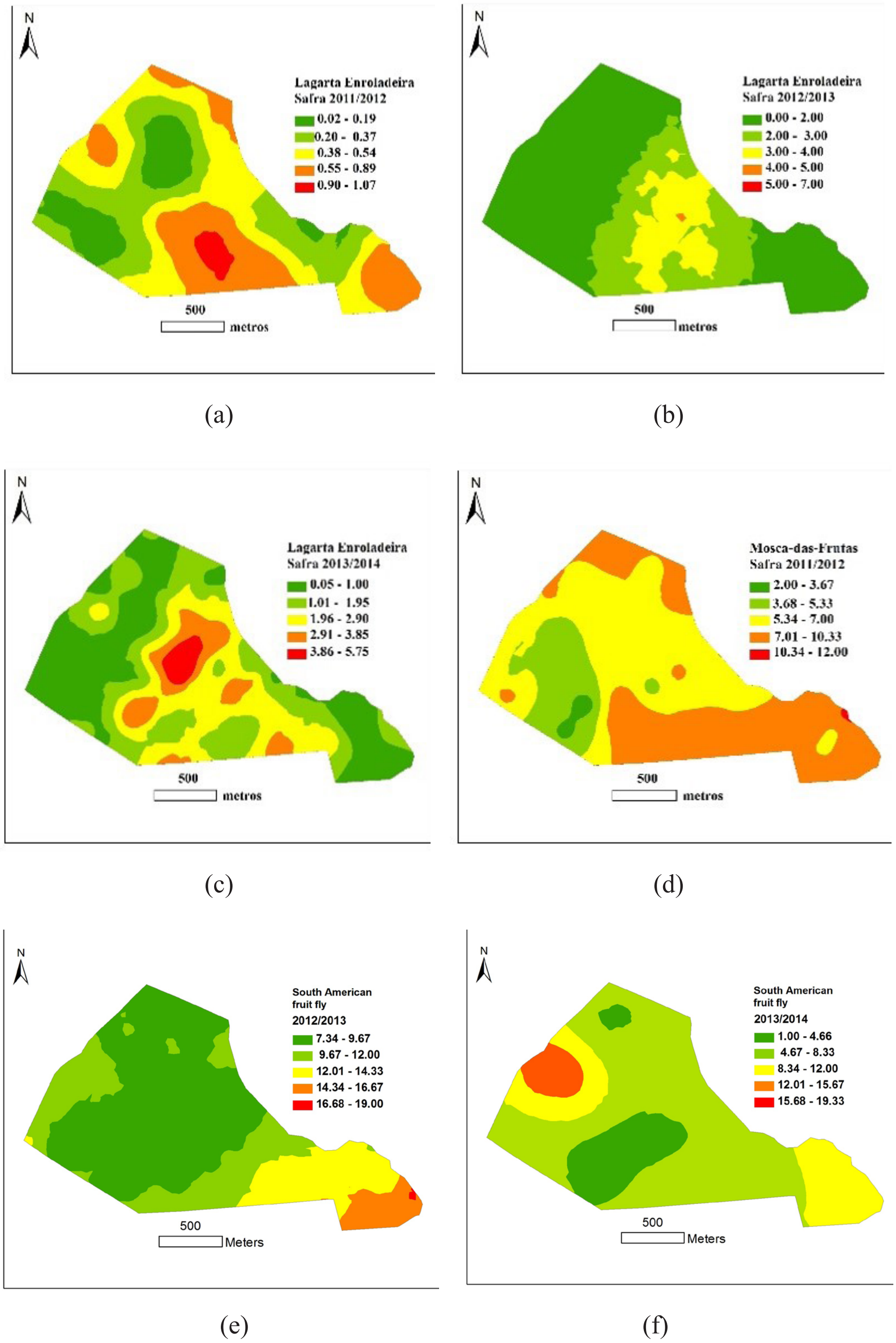


Figure 3. Evolution maps of pest occurrence obtained from interpolation by kriging for Brazilian apple leafroller: (a) 2011/2012 crop year, b) 2012/2013 crop year and c) 2013/2014 crop year; and for South American fruit fly: (d) 2011/2012 crop, e) 2012/2013 crop year and f) 2013/2014 crop year

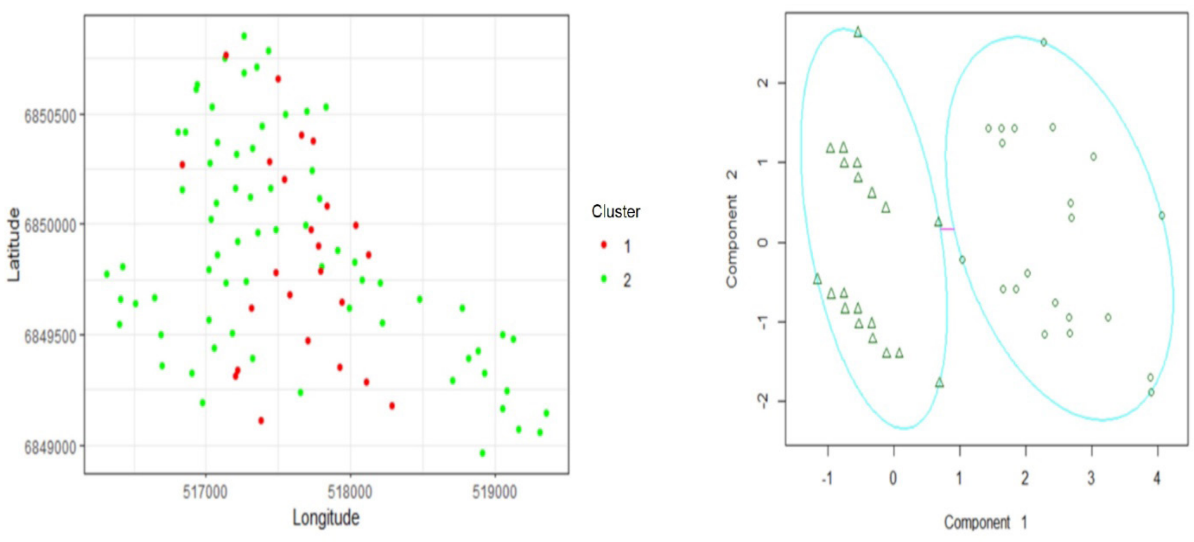
distribution for data counts. In both cases, we achieved the best clusterings regarding the subdivision into two groups. The low incidence of occurrences present in the data in the three cycles prevented the *optCluster* function from finding an ideal algorithm to separate the incidences of the pest Oriental fruit moth into clusters. Thus, we discarded the specific dataset for this pest, as it had already occurred in the geostatistical analysis.

The clustering obtained for the Brazilian apple leafroller in the form of a map (Figure 4a) shows, as already seen in the maps of Figure 3, a smaller central cluster with high incidence of the pest (Cluster 1 - samples in red), reaching a median value of four incidences per crop; and forming a larger cluster in the peripheral regions of the field, with low incidence of the pest (Cluster 2 – samples in green), reaching a median value of only one incidence per crop.

The *clusplot* function, available in the cluster package of the R software, shows the intra-cluster cohesion and inter-cluster separation of a cluster generated from an n-dimensional dataset in a two-dimensional space automatically generated from its two principal components. The clustering generated for Brazilian apple leafroller shows a satisfactory separation of the samples into two clusters (Figure 4b). However, this type of analysis only considers the input parameter pest incidence data from the three crop years analyzed, without taking into account the spatial location of the traps during the clustering process. Thus, when we return to the map in Figure 4a, we observe that some areas with a high incidence of pests appear isolated in the periphery of the field, showing a spatial discontinuity that can hinder the map use as a pest management tool.

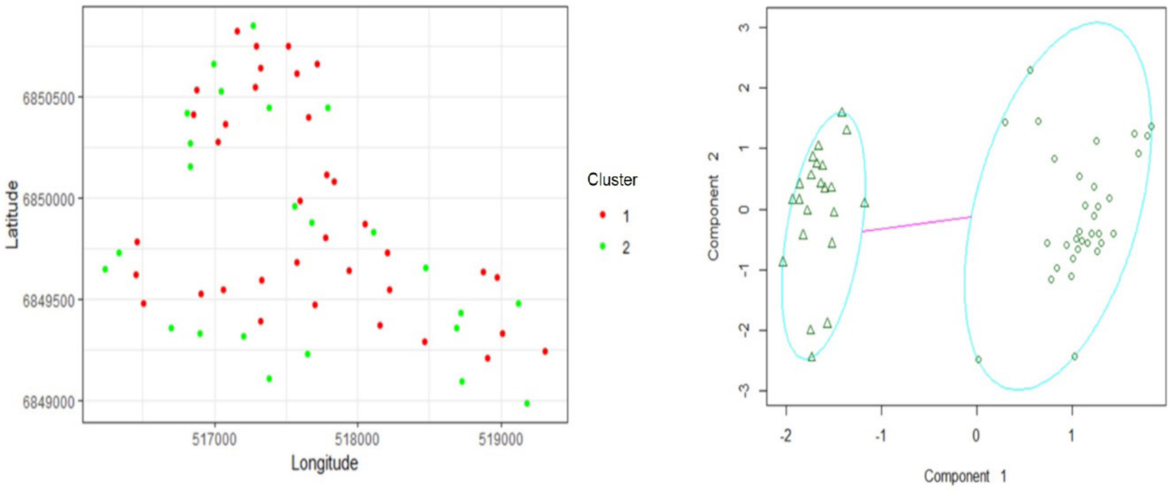
For the clustering of the pest South American fruit fly the sample distribution map (Figure 5a) shows an even more complex spatial subdivision between Cluster 1 (samples in red), with a median incidence of 8 events per crop; and Cluster 2 (green samples), with a median incidence of 12.5 events per crop. However, the visualization of the clusters obtained in the two-dimensional space generated from the two principal incidence components also shows a total separation between them, indicating an adequate subdivision into two clusters (Figure 5b).

Using the results in Figures 4 and 5, and considering the existence of non-coincident points in the location of the traps installed for the two pests, a simple fusion of clusters was performed, joining the samples from the cluster with the lowest incidence of the two pests in a single cluster (median values 1 and 8) and samples with highest incidence (median values 4 and 12.5) in another cluster (Figure 6a). However, these new clusters, despite having very different median values (one



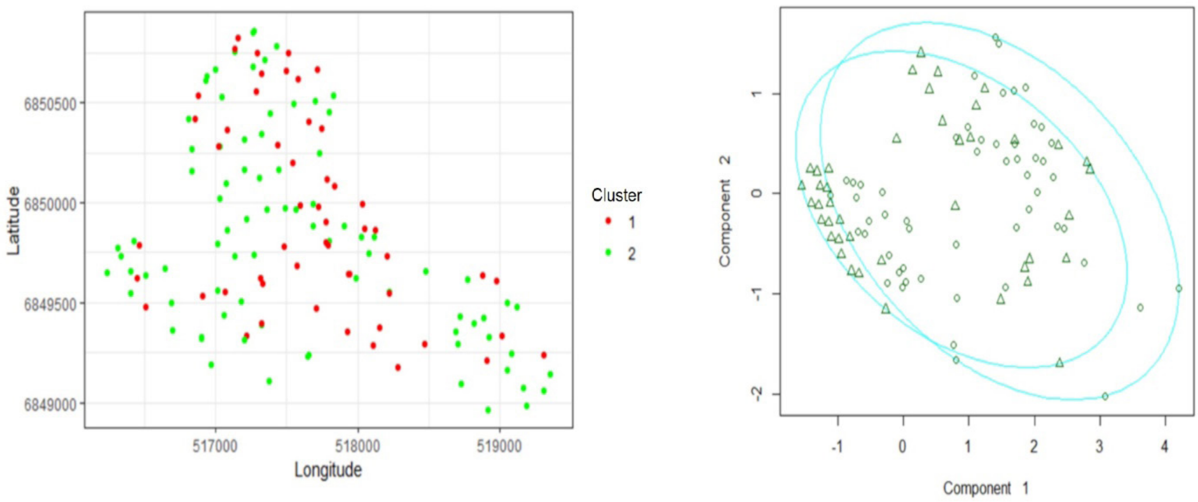
(a) (b)

Figure 4. Clustering of Brazilian apple leafroller using as input data the count of events in 3 crop years: (a) clustering (2 clusters) displayed in map form, red dots represent samples associated with Cluster 1 and green dots represent samples associated with Cluster 2; (b) visualization of clusters in two-dimensional space (samples in different clusters represented by triangles or circles)



(a) (b)

Figure 5. Clustering of South American fruit fly, using as input data the count of events in 3 crop years: (a) clustering (2 clusters) displayed in map form; (b) visualization of clusters in two-dimensional space (samples in different clusters represented by triangles or circles)



(a) (b)

Figure 6. Clustering resulting from the merging of the clusters in Figures 2 and 3: (a) clustering (two clusters) displayed in map form; (b) two-dimensional space clustering visualization (samples in different groups represented by triangles or circles)

and five, respectively), do not appear to be well separated in general as there are many points of intersection between them when viewed from the two principal components (Figure 6b). Although this result indicates certain difficulty in separating the study field into two unique areas with higher and lower overall pest incidence, it is possible to observe a continuous area of above-average incidence for both pests in the central area of the map (Figure 6a). Triangles in the only non-intersecting area between the clusters (Figure 6b) represent these samples. Additionally, in the northern part of the map it is also possible to identify a continuous area of pest incidence, mainly influenced by South American fruit fly.

Using the clustered data in Figure 6a, we created Voronoi diagrams, dissolution and smoothing of polygons in the QGIS software, transforming the sample point map into polygon map. Thus, a continuous general map of the spatial distribution of pests in the study field was created based on the analysis performed from the three crop years (Figure 7).

In Figure 7, areas in red (Group 1) are identified as having a high incidence of pests, with median values of 5 incidences per crop and the areas in green (Group 2) are identified as having a low incidence of pests, with median values of one incidence per crop. As in Figure 6, it is possible to identify a continuous central area and another smaller area located to the north of the map with a high incidence of pests.

The Kappa correlation coefficient was used to verify the agreement of the map obtained from the cluster analysis (Figure 7) with maps obtained from geostatistics (Figure 3). For this purpose, the five classes of each of the six maps in Figure 3 were converted into only two classes (with the highest and the lowest incidence of pests) based on the division of values from the natural breaks algorithm (JENKS, 1967). Table 2 shows the correlation results obtained with the Kappa correlation coefficient.

Following the classification used for the Kappa correlation criterion (VIERA & GARRETT, 2005), for both pests, the correlation with the general map obtained from cluster analysis can be considered fair for the 2011/2012 and 2013/2014 crop years; and moderate, for the 2012/2013 crop year. Whereas no correlations with Kappa values closer to one (substantial or nearly perfect classification) shows that the clustering algorithm used was able to provide a more general solution that fairly considers the temporal dynamics of infestation pests observed on individual maps obtained from geostatistics. Additionally, the non-occurrence of values below 0.20 for the Kappa index also indicates that none of the individual maps showed little correlation with the unified solution.

From new georeferenced analyses of factors

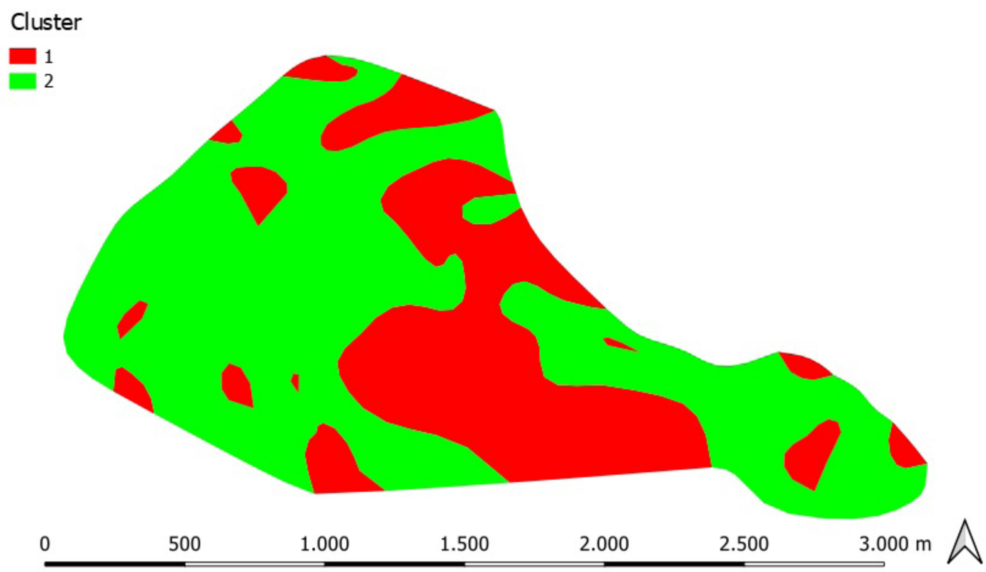


Figure 7. Suggested map for differentiated pest population management based on the incidence found in three crop years for the pests Brazilian apple leafroller and South American fruit fly. Areas with high pest incidence are shown in red (Cluster 1) and areas with low pest incidence are shown in green (Cluster 2)

Table 2. Kappa correlation coefficient for comparison between single pest incidence maps obtained from clustering analysis, with individual maps obtained from geostatistics for each pest in each crop year

Pest / Crop year	2011/2012	2012/2013	2013/2014
Brazilian apple leafroller	0.23	0.48	0.26
South American fruit fly	0.36	0.47	0.35

related to soil, crop and microclimate, it is possible to carry out a better investigation of the causes of the higher incidence of pests in the locations shown in the map of Figure 7, favoring a more specific management to reduce the risks of increased pest population in the following crop years. Additionally, knowledge on the temporal variability of the behavior of the pests analyzed allows the farmer, or the decision maker, to define better pest control strategies, aiming to manage orchards based on the historical risk of the areas, especially in conditions of resource or labor limited. Therefore, it is possible to direct greater efforts towards the areas of greater risk (red zones) at the expense of those of lesser risk (green zones) and, as the database grows, allow more clarity of the causes of spatial and temporal variability in the behavior of insects in field.

Overall, the performed geostatistical analysis allowed the assessment of infestations of Brazilian apple leafroller in the central area of the study field; and of South American fruit fly in the lower right area of the study field. The cluster analysis identified the areas with incidence above the average of both pests at the same time, mainly in the central area of the map, in agreement with part of the result obtained by the geostatistical analysis and verified from the analyzes using the Kappa coefficient. In the approach used in this study, we do not use the continuous maps obtained from geostatistics as inputs for the clustering algorithms. However, we can use this type of approach, widely used in the delineation of management zones, to reduce the appearance of small areas with a high incidence of pests observed in the map of Figure 7.

CONCLUSION

- This study allowed identifying, through a dataset from incidence of the pests Brazilian apple leafroller and South American fruit fly, in three different crop years, areas with greater susceptibility to pests within a commercial apple orchard.
- The geostatistics methodology applied here allowed the generation of continuous individual maps of the incidence of different pests in different crop years, providing the visualization of their temporal evolution. However, some factors involving the construction of these maps such as the adjustment of semivariograms, are highly dependent on visual interpretation by a specialist and spatial interpolation, which predicts mean values regarding the neighborhood, makes its practical use very restricted. On the other hand, the use of specific clustering algorithms in counting data allows for a more generalized identification of infested areas, aggregating in a single map information on the incidence of different pests in different crops, without requiring more specialist-mediated analyses. Since the collection and storage of pest incidence data is mandatory for certification programs, its application as metadata can be used for risk analysis, with management benefits and cost reduction to the farmer.

AUTHORSHIP CONTRIBUTION STATEMENT

SPERANZA, E.A.: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing; **GREGO, C.R.:** Conceptualization, Data curation, Methodology, Writing – original draft, Writing – review & editing; **GEBLER, L.:** Conceptualization, Formal Analysis, Validation, Writing – original draft, Writing – review & editing.

DECLARATION OF INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGEMENTS

The authors thank Agropecuária Schio for performing the collection and availability of data.

REFERENCES

BONIECKI, P.; KOSZELA, K.; PIEKARSKA-BONIECKA, H.; WERES, J.; ZABOROWICZ, M.; KUJAWA, S.; MAJEWSKI, A.; RABA, B. Neural identification of selected apple pests. **Computers and Electronics in Agriculture**, Amsterdam, v. 110, p. 9-16, 2015.

BRUNELLI, D.; ALBANESE, A.; D'ACUNTO, D.; NARDELO, M. Energy neutral machine learning based iot device for pest detection in precision agriculture. **IEEE Internet of Things Magazine**, [s.l.], v. 2, n. 4, p. 10-13, 2019.

CANDEIA, M. G.; GEBLER, L.; BUENO, M. B. Uso da AP no gerenciamento de metadados de longo prazo de monitoramento de pragas da macieira. In: CONGRESSO BRASILEIRO DE AGRICULTURA DE PRECISÃO – CONBAP, 2016, Goiânia. **Anais** [...]. Goiânia: Sociedade Brasileira de Engenharia Agrícola, 2016.

DINARDO-MIRANDA, L. L.; VASCONCELOS, A. C. M.; VEIRA, S. R.; FRACASSO, J. V.; GREGO, C. R. Uso da geoestatística na avaliação da distribuição de Mahanarva fimbriolata em cana-de-açúcar. **Bragantia**, Campinas, v. 66, p. 449-455, 2007.

FIORAVANÇO, J. C.; DOS SANTOS, R. S. S. (Ed.). **Maçã -** o produtor pergunta, a Embrapa responde. Brasília: Embrapa, 2013.

JAIN, A. K.; DUBES, R. C. **Algorithms for clustering data**. Hoboken: Prentice-Hall, 1988.

JENKS, G. F. The data model concept in statistical mapping. **International yearbook of cartography**, [s.l.], v. 7, p. 186-190, 1967.

KOVALESKI, A.; RIBEIRO, L. G. **Manejo de pragas na produção integrada da maçã**. Bento Gonçalves: Embrapa Uva e Vinho, 2002.

LAZZAROTTO, J. J. **Indicadores econômicos e financeiros em sistemas típicos de produção de maçã no Brasil**. Bento Gonçalves: Embrapa Uva e Vinho, 2018.

MCHUGH, M. L. Interrater reliability: the kappa statistic. **Biochemia medica**, Zagreb, v. 22, n. 3, p. 276-282, 2012.

MOON, T. K. The expectation-maximization algorithm. **IEEE Signal processing magazine**, [s.l.], v. 13, n. 6, p. 47-60, 1996.

PROTAS, J. F. S.; SANHUEZA, R. M. V. **Normas Técnicas e Documentos de Acompanhamento da Produção Integrada de Maçãs**. Bento Gonçalves: Embrapa Uva e Vinho, 2002.

QGIS. **A Free and OpenSource Geographic Information System**. Available in: https://qgis.org. Access in: July 26, 2021.

ROGGIA, S.; BUENO, A. de F.; OLIVEIRA, M. C. N. de; GREGO, C. R.; VICENTE, A. K.; VICENTE, L. E. Controle localizado. **Cultivar Grandes Culturas**, v. 21, n. 263, p. 24-28, 2021.

SEKULA, M.; DATTA, S.; DATTA, S. OptCluster - An R package for determining the optimal clustering algorithm. **Bioinformation**, [s.l.], v. 13, n. 3, p. 101, 2017.

SI, Y.; LIU, P.; LI, P.; BRUTNELL, T. P. Model-based clustering for RNA-seq data. **Bioinformatics**, Oxford, v. 30, n. 2, p. 197-205, 2014.

UEDA, N.; NAKANO, R. Deterministic annealing EM algorithm. **Neural networks**, [s.l.], v. 11, n. 2, p. 271-282, 1998.

VAN LAARHOVEN, P. J. M.; AARTS, E. H. L. **Simulated annealing: Theory and applications**. Dordrecht: Springer, 1987.

VIEIRA, S. R.; GARCIA, M. A. G.; GONZÁLEZ, A. P.; SIQUEIRA, G. M. Variabilidade espacial e temporal do teor de água do solo sob duas formas de uso. **Bragantia**, Campinas, v. 69, n. 1, 2010

VIERA, A. J.; GARRETT, J. M. Understanding interobserver agreement - The kappa statistic. **Fam med**, Leawood, v. 37, n. 5, p. 360-363, 2005.

WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data Mining: Practical Machine Learning Tools and Techniques**. 3rd ed. London: Morgan Kaufmann, 2011.

WRZESIEŃ, M.; TREDER, W.; KLAMKOWSKI, K.; RUDNICKI, W. R. Prediction of the apple scab using machine learning and simple weather stations. **Computers and Electronics in Agriculture**, Amsterdan, v. 161, p. 252-259, 2019.