Research article

Thrips incidence prediction in organic banana crop with Machine learning

Jose Manrique-Silupu *, Jean C. Campos, Ernesto Paiva, William Ipanaqué

Automatic Control Systems Laboratory, Universidad de Piura, Av. Ramón Maguña 131, Piura, 20009, Piura, Peru

HIGHLIGHTS

- Precision agriculture for pest control in organic banana.
- Prediction of Pest incidence with machine learning techniques.
- Machine learning modeling applied to organic banana cultivation.
- IOT sensor network and cloud-based data logging system in banana plantation.

ARTICLE INFO

Keywords:
Multi-class classification
Machine learning
Organic banana pest
Support vector machine
Red rust thrips
Twin Support Vector Machine

ABSTRACT

The organic banana is one of the most popular products worldwide and its popularity is mainly due to its excellent nutritional properties and tasty flavor. Peru is considered one of the major producers and exporters of this product, being the city of Piura the main region with most of the national agro-producers. It is also considered a key factor in the development of the economy of this region as it creates job opportunities because of the productive chain required in the process (harvest, post-harvest, and export). The main problem faced by producers is the existence of pests such as Red spot thrips, Black Sigatoka, and others, which affect the production and the quality of the final product. Therefore, this article aims to propose an alternative solution, using the 4.0 Industry technology as well as the installation of an IoT sensor network in banana plantations in order to develop a model which estimates the classification of the pest incidence level based on Machine learning techniques, making use of the atmospheric variables measured with the IoT sensor network as input data. In the research, we have used The Support Vector Machine techniques, which have successfully achieved models with a high level of accuracy. The implementation of this system aims to help producers improve the management of pest control by scheduling spraying dates more effectively, optimizing not only the quality of the product but also reducing costs.

1. Introduction

The organic banana is a product in great demand worldwide for its nutritional properties and exquisite flavor [1]. This product is considered the fifth most important crop worldwide, representing a calorific source for more than 500 million people in tropical countries [2], its main producers are located in Asia and America [3].

Peru has an area of 197,837 hectares which are used for agriculture [4], and is one of the main suppliers of this product in the American market [5], accounting for approximately 3% of the world's organic banana production. As with other organic products, one of the main problems to be faced is the poor management of pests and diseases associated with climatic conditions [6, 7], which impair the productivity and quality of the product.

To face this type of problem precision agriculture has appeared as an alternative that promises to improve the efficiency and quality of agricultural production [8], transforming manual work into a technological one [9].

Precision agriculture can be defined as the combination of information and communication technologies, also known as 4.0 technologies, which allows obtaining information from numerous sources, allowing more efficient decision-making based on the analysis of this information [10, 11].

* Corresponding author.
E-mail address: jose.manrique@udep.edu.pe (J. Manrique-Silupu).

https://doi.org/10.1016/j.heliyon.2021.e08575
Received 30 April 2021; Received in revised form 25 August 2021; Accepted 6 December 2021

2405-8440/ © 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
The 4.0 technologies applied in the agricultural sector are Big Data [12, 13], Cloud Computing [14, 15], Cyber-Physical System [16, 17], Internet of Things (IoT) [18, 19], and Artificial Intelligence (AI) [20, 21]. The interpretation of information with Machine Learning (ML) techniques allows the development of mathematical models with high estimation accuracy [22].

In [23], it’s been used a multi-sensor network system and the Support Vector Machine technique to work on an early warning model of pests in vegetables based on the information collected in a multidimensional way: the number of pests, soil, environment, climate, and meteorological factors in real-time, to have a Big Data system.

In [24], we introduce a description of current trends for crop pest prediction using machine learning techniques such as Support Vector Machine. In [25], a high-resolution spatial image is used to determine the extent of Yellow Sigatoka infection in a banana crop being able to predict phenotypic phenomena from a Support Vector Machine model. In [26], a system for early prediction of Black Sigatoka outbreak in banana plantations is developed through the use of techniques such as linear regression, multivariate regression, regression tree, and additive regression.

In [27], a multiple logistic regression model was used to find the relationship between the level of infection in the fungus commonly found in the fly larvae of the Mediterranean fruit. In [28], we have detailed the multiple regression model for crop pest risk prediction as well as we have evaluated the performances of multiple regression models and neural networks. In [29], the cereal loss rate in the harvesting phase has been estimated by designing a model based on logistic regression. In [30], the model based on the multiple logistic regression shows the factors that determine the level at which the IPM (Integrated Pest Management) system is adopted by a producer. In [31], it’s been introduced a model based on the Decision Trees model to be able to predict the behavior of nutrients in the soil of banana plantations, facilitating decision-making and optimizing resources. In [32], it’s been developed a pest prediction model for banana crops using linear regression and Support Vector Machine techniques, evaluating and comparing both alternatives.

Thus, this work proposes a mathematical model to predict the level of pest incidence in banana crops, using Support Vector Machine (SVM). By comparing its performance and accuracy, this new tool would facilitate the decision-making for farmers in the region of Piura, Peru.

The methodology to develop these models includes data collection in the field with a data recording system, description of the binary classification for pest incidence, protocol and parameters of the experiment management, selection of the input variable, design of the SVM prediction models, comparison of the metrics and discussion of the results.

2. Information gathering

A data logging system has been installed in a banana plantation located in the district of Buenos Aires, province of Morropón, department of Piura, Peru, with geographical coordinates 5° 16’ 13.4” S 79° 57’ 10.1” W, this plot has an area of 1 Ha.

2.1. Data logging system

The installed system consists of a network of IoT (Internet of Things) sensors to measure climatic and soil variables. The sensors were installed at 3 levels: the soil (subway level), below the banana leaves (microclimate level), and above the banana leaves (climate level).

The sensor network is composed of two nodes and a weather station as shown in Fig. 1, due to the seasonal behavior of the pest these variables are strongly related to the development of pests in banana crops as detailed in [33, 34].

Table 1 shows the variables measured by the data recording system at all levels. It is important to note that not all variables were relevant to the mathematical model design.

3. Pest control

The procedure followed to determine the incidence of the pest in the plot starts by choosing 25 plants at random; afterward, a detailed inspection is carried out for which it is necessary to cut leaves from the plant and perform a visual inspection to count the number of insects found. This information is recorded because it will allow us to decide when to carry out the pest control process. The diagnosis of the crop plot is carried out using the pest incidence level “a”, which is defined as:

\[
\alpha = \frac{\text{Total insect}}{\text{Total of plants inspected}} \tag{1}
\]

The application of the control action will depend on the classification of the level of incidence of thrips, as indicated in Table 3. So, considering the level of pest incidence, farmers apply products such as Sulphocalic Broth, which is a combination of Sulphocalic Broth with agricultural oil or Entrust SC. Crop rotation is relevant in the use of these products to avoid creating resistance in the plants to be sprayed. In addition, one of the main challenges is to avoid the excessive use of these products in the plantations.

4. Experiment management

Data has been systematically stored since the installation of the statistics collection system in the organic banana plantation in November 2019. Likewise, 112 samples were taken for the development of this research until March 2021.

4.1. Limitations

Considering the geographic location detailed in section 2, the results of this research were developed in a region with a desert climate, low rainfall during the year, an annual temperature ranging between 20 °C and 31 °C, low wind speed, and high solar radiation.

It’s also been modeled the behavior of the growing population of the red spot Thrips (Chaetanaphothrips signipennis), which affects plants of the Musaceae family.
4.2. **Thrips incidence behavior analysis**

The incidence prediction model was designed for the pest called Red spot thrips (*Chaetanaphothrips signipennis*) since it has the most severe impact on organic banana crops in the region of Piura, Peru. The behavior of this pest is seasonal and its increase is due to the high temperatures during the summer months, from December to March.

By analyzing the level of incidence, it was possible to identify that the behavior of the frequency of the Red spot thrips resembles an exponential function since it can be used to describe the growth of the population of this pest. (See Fig. 2.) Thus, the following can be written:

\[ I_T(t) = I_{T,0} e^{rt} \]  \hspace{1cm} (2)

Considering that \( I_T \) represents the incidence of Red rust thrips, \( I_{T,0} \) is the initial population, \( r \) is the effective population growth rate and \( t \) is the time represented in days, the population can be calculated for time \( t_2 \) from a previous population at time \( t_1 \).

\[ I_T(t_2) = I_T(t_1) e^{r(t_2-t_1)} \]  \hspace{1cm} (3)

It has been determined that the most relevant meteorological variables measured by data logging systems are micro-climate temperature (°C), rain rate (mm), and wind speed (m/s) [26, 34, 35]. These variables vary their behavior cyclically throughout the year. In order to explain the relationship between pest and atmospheric variables, it was decided to write \( r \) as follows:

\[ r = a_0 + a_1 T + a_2 S_w + a_3 R + a_4 CEF_1 + a_5 CEF_2 + a_6 CEC \]  \hspace{1cm} (4)

\( T \) is representing the average temperature, \( S_w \) is the average wind speed and \( R \) is the average rainfall incidence. The meteorological variables are averaged over the time interval \( [t_1, t_2] \). \( CEF_1 \) is the Cumulative Effect of Fumigation for the red rust thrips, \( CEF_2 \) is the Cumulative Effect of Fumigation for other pests, \( CEC \) is the Cumulative Effect for the Cleaning and the following sections are showing how they are calculated considering the fumigation and cleaning dates. The coefficients \( a_0, a_1, a_2, a_3, a_4, a_5, a_6 \) are expected to be estimated with mathematical modeling.

It is possible to linearize (3) by applying a natural logarithm to the expression. By linearizing the function, we can apply a regression method on the incidence of Red rust thrips.

\[ \ln I_T(t_2) = \ln I_T(t_1) + (a_0 + a_1 T + a_2 S_w + a_3 R + a_4 CEF_1 + a_5 CEF_2 + a_6 CEC)(t_2 - t_1) \]  \hspace{1cm} (5)

**Table 2. Features variables.**

| Nomenclature | Expression | Description |
|--------------|------------|-------------|
| \( x_0 \)   | \( \Delta t \) | Days since the previous Thrips incidence |
| \( x_1 \)   | \( T \Delta t \) | Average temperature by time delta |
| \( x_2 \)   | \( S_w \Delta t \) | Average Wind speed by time delta |
| \( x_3 \)   | \( R \Delta t \) | Average rain incidence by time delta |
| \( x_4 \)   | \( CEF_1 \Delta t \) | Cumulative Effect Fumigation for Red rust thrips by time delta |
| \( x_5 \)   | \( CEF_2 \Delta t \) | Cumulative Effect Fumigation for other pest by time delta |
| \( x_6 \)   | \( CEC \Delta t \) | Cumulative Effect Cleaning for other pest by time delta |
| \( x_7 \)   | \( \ln I_T(t_1) \) | Natural logarithm of the previous thrips incidence multiplied by time delta |

**Table 3. Thrips incidence level classification.**

| Incidence level | Control action | Cost |
|-----------------|----------------|------|
| Low Incidence   | Leaf fumigation | 18.64 $/ha |
| Medium Incidence| Stem fumigation | 35.13 $/ha |
| High Incidence  | Stem fumigation and cleaning | 54.04 $/ha |

It is possible to replace \( t_2 - t_1 \) with \( \Delta t \).

\[ \ln I_T(t_2) = \ln I_T(t_1) + a_0 \Delta t + a_1 T \Delta t + a_2 S_w \Delta t + a_3 R \Delta t + a_4 CEF_1 \Delta t + a_5 CEF_2 \Delta t + a_6 CEC \Delta t \]  \hspace{1cm} (6)

It is possible to determine the variables for machine learning algorithms by applying equation (6).

An analysis has been performed in section 5.2 in order to select the variables to be used in the mathematical model training.

**4.2.1. Thrips incidence level classification**

The incidence in the red rust thrips can be determined by using the values shown in Table 3. So, taking these levels into account, the control strategies can be applied as recommended in [36].

Ranking the values of thrips incidence together with the criteria described above gives the following results, see Fig. 3.

At present, farmers from the region of Morropón spend $861/ha, without including labor and scheduled plot cleaning which is done every 20 days. This is because 12 leaf fumigations, 12 stem fumigations,
and 4 stem fumigations with cleaning must be applied during the year considering the costs in Table 4.

5. Prediction model design

The objective of the mathematical model designed in this research is to predict the incidence level classification based on input variables relevant to Thrips reproduction and death behavior.

5.1. Cumulative effect of fumigation and cleaning

It is important to consider the causes of mortality of the species in order to calculate the incidence of red rust thrips, including strategies such as spraying and plot cleaning applied to control the pest.

These variables have a behavior that can be described as an exponential function with a negative rate, which is called degradation rate, and describes the speed with which the chemical used or the cleaning in the plot decreases until it converges to a value close to 0.

\[ F_i(t) = e^{-r_F t} \]  

In equation (7), \( t_F \) represents the time in days since the last fumigation, this variable can be calculated from the recorded dates of fumigation. But \( F_i(t) \) only describes the effect on a day, so in order to describe the Cumulative Effect of Fumigation \((CEF)\) between the inspection days, it is necessary an integral one.

\[ CEF_i(t) = \int_{t_F-2}^{t_F} F_i(t) dt = e^{-r_F t_F} - e^{-r_F (t_F - 2)} \]  

\( t_F \) represents the days since the last fumigation according to the last inspection and \( t_F - 2 \) represents the day since the last fumigation was made. This can be applied to red rust thrips fumigations, other pest fumigations as well as to the cleaning (weeding) of the plot.

5.2. Features selection

It is possible to determine the mathematical importance through numerical weighting by applying sensitivity analysis of the characteristics in the variables described in Table 2.

Previously, all variables were multiplied by Delta T between plot inspections, in order to resemble the mathematical model to the linearization shown in equation (6).

The results of the analysis have been shown in Fig. 4, specifically in the graphs showing sensitivity to shuffled values (“shuffle”) and sensitivity to the omitted values (“missing”), the variable with the highest sensitivity for predicting thrips incidence is the logarithm of the previous incidence, then the cumulative effect of spraying for red spot thrips \((CEF)\), followed by mean microclimate temperature and mean wind speed.

The variables with lower sensitivity are rain rate, the cumulative effect of plot cleaning, \( \Delta t \), and the cumulative effect of spraying for other pests.

The variables \( \Delta t \) and rain rate have been eliminated since they did not have a significant effect on the output variable; however, it was decided to leave the cumulative effects of spraying other pests and cleaning as they are the cause of death.

5.3. Data pre-processing

Since the characteristics are worked with different units and scales, we have considered applying a scaling method, so for this research, we have chosen the standard normalization described in equation (9).

\[ z = \frac{x - \mu}{\sigma} \]  

5.4. Support Vector Machine

The Support Vector Machine is a supervised machine learning algorithm that allows solving classification problems as in [23, 24, 25].

5.4.1. Support vector regression

Support vector regression (SVR) allows predefining the acceptable error for the model and, consequently, to find a hyperplane that fits the data.

Unlike linear regression, the objective function of SVR is to minimize the norm of the coefficient vector and not the squared error. This is handled in the constraints of the problem, where the absolute error can be less than or equal to a specified margin, also called the maximum error \((\epsilon)\), \( \epsilon \) can be adjusted to the desired accuracy of the model. There are scenarios in which the model can’t fit all points, for this reason, it is necessary to include slack variables, which are defined as the deviation of the data that are not within the margin. These deviations are assumed to exist, but the objective is to minimize them as much as possible. So the objective function is modified to:

\[ \min_{\omega} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{n} \xi_i \]  

subject to \( |y_i - \omega^T x_i| \leq \epsilon + \xi_i \)

\( x_i \) represents the vector containing the input data of the mathematical model, \( \omega \) is the vector containing the weighting corresponding to each input variable and the value of the bias. \( C \) can be defined as the tolerance of the points outside the margin. The tolerance increases with \( C \), so \( C \) should be as close to 0 as possible. (See Fig. 5.)

The Kernel function is an additional tool that allows fitting the mathematical model to data with non-linear behavior by adding a higher dimension to the space generated by the input variables.

\[ K(x, x_i) = \phi(x)^T \cdot \phi(x_i) \]  

The equation describing the hyperplane is \( \omega^T \phi(x_i) \), the objective function results in:

\[ \min_{\omega} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{n} \xi_i \]  

subject to \( |y_i - \omega^T \phi(x_i)| \leq \epsilon + \xi_i \)

5.4.2. Support vector classifier

The Support vector classifier (SVC) tries to separate different classes by optimizing the location of a plane and maximizing the margin. But it is also necessary that the hyperplane has the maximum possible margin concerning the data it is separating; the margin is the distance between the points closest to the hyperplane (which are known as support vectors). Thus, the condition expressed in equation (13) must be fulfilled.

\[ y_i(\omega^T x_i) = \pm 1 \]  

Thus, the distance or margin \((d)\) is related to the vector modulus \( \omega \) and can be calculated with equation (14).

\[ d = \frac{2}{\|\omega\|} \]  

An example of the optimal placement of a hyperplane is illustrated in Fig. 6.

The objective of the SVC consists of finding the highest margin which satisfies the separation between two classes, so it becomes an optimization problem described in equation (15).

\[ \min_{\omega} \frac{1}{2} \|\omega\|^2 \]  

subject to \( y_i(\omega^T x_i) \geq 1 \)

It is not always possible to find a hyperplane which satisfies the conditions posed by this problem, so it is necessary to include a slack variant
whose value must satisfy the condition of always being positive. Then the new function to be optimized is:

$$\min \frac{1}{2} \|\omega\|^2 + \lambda \sum_{i=1}^{n} \eta_i$$

subject to $y_i(\omega^T \phi(x_i)) \geq 1$

$$\eta_i \geq 0$$

(16)

It is possible to solve this problem if the points can be separated linearly, however, it is not necessarily feasible, so it is recommended to apply a transformation (Kernel function) [37] to the input data to create a higher dimension so that a linear separation can be applied in this new division.

$$\min \frac{1}{2} \|\omega\|^2 + \lambda \sum_{i=1}^{n} \eta_i$$

subject to $y_i(\omega^T \phi(x_i)) \geq 1$

$$\eta_i \geq 0$$

(17)

5.4.3. Twin support vector classifier

The Twin support vector machine (TSVM) differs from the previous approach because it uses two non-parallel hyperplanes to separate classes. Therefore, it tries to solve a pair of small problems instead of a much more complex problem as it would be the conventional SVC. (See Figs. 7 and 8.)
Having two hyperplanes, the objective function of the Twin Support Vector Machine is composed of two equations:

\[
\min \frac{1}{2} \|w^T x_i\|^2 + \lambda \sum_{i=1}^{n} \eta_i \\
\text{subject to } -w^T x_i + \eta_i \geq 1 \\
\eta_i \geq 0
\]

(18)

\[
\min \frac{1}{2} \|w^T x_i\|^2 + \lambda \sum_{i=1}^{n} \xi_i \\
\text{subject to } w^T x_i + \xi_i \geq 1 \\
\xi_i \geq 0
\]

(19)

Adding the Kernel function, the result is:

\[
\min \frac{1}{2} \|\phi(x_i)\|^2 + \lambda \sum_{i=1}^{n} \eta_i \\
\text{subject to } -\phi(x_i) + \eta_i \geq 1 \\
\eta_i \geq 0
\]

(20)

\[
\min \frac{1}{2} \|\phi(x_i)\|^2 + \lambda \sum_{i=1}^{n} \xi_i \\
\text{subject to } \phi(x_i) + \xi_i \geq 1 \\
\xi_i \geq 0
\]

(21)

6. Results and discussion

For the results, we have used the GridCVSearch algorithm of SciKit Learn for the different combinations of SVM parameter values, with 20-fold cross-validations.

6.1. SVR results

Several models were run with SVR to numerically estimate the incidence of red spot thrips. The metrics of the models in which the cumulative effects of spraying and cleaning have not been considered are at least 20% lower than those shown below, Table 5 shows the results of the models that include the cumulative effects of spraying and cleaning.

We can observe that the results with RBF were better even by dividing the preciseness by seasons, so the accuracy of the SVR Linear Kernel does not exceed the accuracy of the RBF SVR Kernel. When analyzing the yield by season, it is found that in summer a more accurate prediction is possible, since a more differentiated behavior of the pest is noticed, and this is aligned with the information collected in the field, since the agro-producers informed us that the pest has a higher incidence in summer months as there is a higher average daily temperature and a greater amount of rainfall per week.

For the autumn and winter seasons, low goodness-of-fit values were obtained, the SVR RBF kernel manages to follow a trend in the real behavior but does not manage to adjust itself, this may be because these are the seasons when the population reaches the lowest values. (See Fig. 9.)

None of the SVR alternatives can predict the spring season. At the end of spring, there is an increase in the thrips population, which prepares the conditions for much more rapid development of the pest in the summer season. Therefore, it is not easy to adjust the model to the behavior during this season.

Finally, the numerical value obtained from the SVR model prediction was converted into classes as shown in Table 2. We were also able to obtain metrics of 82% in F1-score as well as in its precision with the SVR RBF kernel transformation. The rest of the models did not exceed 73% in any of the aforementioned metrics.

6.2. SVC results

The confusion matrix was used to analyze the results of the classification, the correct as well as the incorrect classifications are accounted [38].

As it can be seen in Table 6, better results were obtained taking into account the total points of the year, however, in the analysis done considering the season, it is possible to observe that during summer the SVC Linear Kernel has the highest metrics. On the other hand, in winter, the highest metrics were obtained by the SVC RBF Kernel and during the other two seasons, the SVC Sigmoid Kernel excels.
Table 6. SVC Results.

|                | F1-Score | Accuracy |
|----------------|----------|----------|
| SVC Linear Kernel | 73%      | 73%      |
| SVC RBF Kernel   | 70%      | 72%      |
| SVC Sigmoid Kernel | 74%  | 75%      |

These models did not improve the accuracy and F1-score metrics obtained by converting the SVR RBF Kernel SVR model results into classes. (See Fig. 10.)

Considering the classifiers these were able to improve the predictions during the fall and winter seasons, which was not possible to achieve with the SVR. Spring is the season with the worst accuracy performance, which is reflected in the behavior observed in the SVR results.

6.3. Twin SVC results

Finally, we trained a One-versus-One multiclass Twin SVC model. The results were as follows, see Fig. 11 and Table 8.

When comparing TSCV and SVC with Linear Kernel, we realize that both have similar performance if metrics for the whole year are taken into account. However, if the analysis is seasonal we find that the values of the TSCV metric fall in summer while it increases during the rest of the seasons, so it is required a model which allows better monitoring of the population growth.

The highest metrics are those of the TSCV RBF kernel, which outperforms the ranking of the SVR model prediction. In addition, it has a very stable behavior in summer, autumn, and spring, with metrics above 87%, that is, it is superior to all the proposed models in all the circumstances under which they have been evaluated.

6.4. Discussion

It can be observed from the implemented models that the behavior during summer differs from the rest of the seasons, so it is possible to model it using linear regressions with a high level of adjustment; during the rest of the seasons the adjustment drops considerably, this may be because, during the time interval it has been sampled, two summers
correspond to the years 2020 and 2021, which allows identifying more precisely the behavior of the pest during this season.

Any alternative studied is not able to model the behavior of population growth during spring. Spring can be considered as the transition season since during the winter and autumn months the average temperature is lower than the rest of the year, however, during spring the temperature begins to increase preparing the circumstances for the considerable increase of the population that occurs in summer taking into account that during the first months there are still no adequate temperature values for an uncontrolled growth of the pest, which are found at the end of December due to the fact that in this month there are values that promote considerable increases.

It is important to point out that the seasonal performance of TSVG RBF Kernel significantly outperforms the other alternatives analyzed for the fall and winter seasons. From a practical point of view, it can be used for 9 months of the year as it has the potential to be implemented as a monitoring system that can not only be adapted with little work but also does not require much time.

7. Conclusions

This research has an important value for the development of precision agriculture in Peru, considering that the agricultural sector has had a growth trend in the last 10 years, due to the development of tools such as the one based on industry 4.0. It would open possibilities to apply new technologies to this sector, and thus allow small and medium agro-producers to be more competitive nationally and internationally.

The possibility of developing a seasonally differentiated alarm system has been determined taking the seasonal results into account. It can be affirmed that the predictions in the winter season are less reliable if we consider the results shown in Tables 7 and 9, so it would be convenient to resume specific modeling of this season as soon as enough points have been collected. This could take approximately 4 years.

Determining that the growing population of pests has an exponential behavior has permitted us to propose mechanistic mathematical modeling alternatives, with the possibility of developing new alternatives and with greater precision, based on a deeper understanding of the principles of the process. Presently, the inclusion of the logarithm of the previously measured incidence allows the model to linearize the exponential behavior of reproduction in pest species.

If we consider the costs shown in Table 4, it is possible to analyze the economic impact of implementing a pest early warning system with a high level of reliability. It is estimated that low-class spraying would no longer be necessary and stem spraying with cleaning would be reduced by half its costs. We came to this conclusion as a result of the interviews we had with the farmers who reported that they could only perform one pest inspection per plot, that most of the spraying they do is not properly programmed and in many opportunities, they must spray stems during the month; considering that the tool can be identified at intervals decided by the user, it would be possible to program the spraying when the incidence level is projected to increase. This would result in an annual cost of $529.73/ha, i.e. a saving of 39% of the total cost.

Finally, it is proposed to develop new pest alert systems focused on red spot thrips, the use of tools to optimize the decision making as well as include optimization for the scheduling of cleaning and fumigation taking into account the quality of the agro-product, the costs of production, fumigation, and cleaning, thus achieving greater efficiency in the process and the competitiveness of small farmers in the international market.

Declarations

Author contribution statement

Jose Manrique-Silupu: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.
Jean C. Campos: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.
Ernesto Paiva: Conceived and designed the experiments; Analyzed and interpreted the data. William Ipanaque: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Funding statement

The authors acknowledge the financial support of the Project Concytec-Banco Mundial “Mejoramiento y ampliación de los Servicios del Sistema Nacional de Ciencia Tecnología e Innovación Tecnológica” 8682-PE, through its executing unit Proiciencia [contrato número 165-2018-FONDECYT-BM-IADTAV – Project: “Transformación Digital del sector Agro-Industrial aplicado al Banano Orgánico”].
