Application of a spatial risk model of the crystalline spider mite (*Oligonychus* sp.) to avocado crop damage using remote sensing

Aplicación de un modelo de riesgo espacial de la araña cristalina (*Oligonychus* sp.) al daño de cultivo de aguacate utilizando sensores remotos

Harry Wilson Báñez-Aldave[1], Ledyz Cuesta-Herrera[2*], Juan López-Hernández[3], Jesús Andrades-Grassi[3]
and Hugo Alexander Torres-Mantilla[4]

ABSTRACT

Keywords:
Avocado
Crystal mite
Geostatistical simulation
Kriging predicted and indicator
Univariate and multivariate geostatistics

The avocado is one of the most consumed foods in the world and it is affected by the mite *Oligonychus* sp., which affects the generation of chlorophyll by the plant, resulting in a decrease in productivity. Given the economic importance of the avocado, a spatial statistical methodology was used to analyze the risk of a pest in its crops. A total of 202 observations of a 1.1 ha avocado farm were used to measure the number of mites per leaf in the area of Barranca, Perú. Predictive geostatistical methods and indicators were applied. A spherical semivariogram was adjusted to estimate a univariate ordinary kriging, covariates such as vegetation indicators and geomorphometric variables were used to improve the spatial resolution of the covariates and geostatistical simulation was used and linear co-regionalization models were adjusted with which pest predictions were made with co-Kriging. Finally, the predictions were transformed into a risk model using kriging indicator. The results obtained show that the mite presents a stationary process in second order with spatial dependence of less than 10 m, in which univariante ordinary kriging was the most efficient. Despite the results, the linear co-regionalization models are consistent, but the geostatistical simulation was not enough to improve the predictions. Covariate data should be incorporated at a higher level of detail and small-scale variations should be analyzed. It is suggested to incorporate covariate data with a higher level of detail and analyze small-scale variations.

RESUMEN

Palabras clave:
Aguacate
Ácaro de cristalino
Simulación geoestadística
Kriging predicto e indicado
Geoestadística univariante y multivariante

El aguacate es uno de los alimentos de mayor consumo a nivel mundial. Sus plantaciones se ven afectadas por el ácaro *Oligonychus* sp., el cual interfiere en la generación de clorofila por parte de la planta, por lo cual la productividad se ve disminuida. Dada la importancia económica de este cultivo se abordó una metodología espacial para el análisis de riesgo de esta plaga. Se dispuso de un total de 202 observaciones de una finca de aguacate de 1,1 ha en las que se midió la cantidad de ácaros por hoja en la zona de Barranca, Perú. Métodos geoestadísticos predictivos e indicadores fueron aplicados. Se ajustó un semivariograma de tipo esférico con lo que se estimó un Kriging ordinario univariante, posteriormente se utilizaron covariables como índices de vegetación y variables geomorfométricas. Para mejorar la resolución espacial de las covariables se utilizó una simulación geoestadística y se ajustaron modelos lineales de corregionalización y co-Kriging, con lo que se realizaron predicciones de la plaga. Finalmente, se transformaron las predicciones a un modelo de riesgo utilizando kriging indicador. Los resultados obtenidos manifiestan que el ácaro presenta un proceso estacionario en segundo orden con una dependencia espacial menor a 10 m, en el que kriging ordinario univariante fue el más eficiente. A pesar de los resultados, los modelos lineales de corregionalización son consistentes, pero la simulación geoestadística no fue suficiente para mejorar las predicciones. Los datos de las covariables deben incorporarse con un mayor nivel de detalle y deben analizarse las variaciones a pequeña escala.

[1] Universidad Nacional de Barranca. Perú. Hbaneza132@unab.edu.pe
[2] Universidad Católica del Maule. Chile. ledyz.cuesta@alu.ucm.cl
[3] Universidad de Los Andes. Mérida, Venezuela. jlopez@ula.ve , andrades@ula.ve
[4] Universidad de Santander. Colombia. h.a.torresmantilla@hotmail.com
* Corresponding author

Avocado (*Persea americana*) consumption has increased worldwide, with a large increase in planted areas in all avocado-producing countries (Hoddle et al. 2002). This food has a wide range of possible uses such as in industrialized products (Cango et al. 2014). The phytophagous mites commonly called "spider mites" are tetranychid, cosmopolitan and highly polyphagous mites, which can develop optimally in more than 150 plant species and affect practically all crops. In avocado, the spider mites that affect this crop belong mainly to the genus *Oligonychus* (Guanilo et al. 2012; Muñoz and Rodriguez 2014). These mites develop mainly on mature leaves near the veins. They feed upon the content of the superficial cells of the leaf during in both immature and adult stages. Usually, infested leaves fall off prematurely due to cell breakdown, removal of chlorophyll and saliva injected by mites leads to leaf malfunction, such as an increased rate of perspiration, resulting in its early wilting and shedding (Cango et al. 2014).

The *Oligonychus* sp. also known as crystal mite or avocado mite, is characterized by forming colonies on the underside of the avocado leaf. The nest is covered by a dense web, under which the different stages of the mite can be observed, forming nests on the underside typically circular, between 1 and 5 mm^2 (Hoddle et al. 2002), also circular necrotic spots can be observed (Hoddle et al. 2002). The damage to the underside of leaves, leading to leaf drop in the EEUU when the percentage of the damaged area is equal to or above 7.5%-10%, corresponds to spider mite populations exceeding 100-500 individuals per leaf (Hoddle et al. 2002). In the EEUU, the presence of high densities of the mite has caused partial or total defoliation of avocado trees (Hoddle et al. 2002; Cango et al. 2014), increasing the risk of sunburn for young fruit and causing premature fruit to drop and size reduction at harvest (Aponte and McMurtry 1997). This mite is one of the most important and frequent that attack the Hass variety, it has a direct consequence on the quality of the fruit and the yields of the mentioned crop (Cango et al. 2014).

By the year 2021 the annual production of avocado in Perú totaled 62,744 metric tons according to data from the National Institute of Statistics and Informatics of Perú, being considered one of the world's leading producers of the fruit (Mendoza 2020). The threat to the economy from *Oligonychus* sp. has been mentioned in several areas of the world, such as México and Spain (Suárez et al. 2023). This pest has been widely studied in Perú and it shows a strong seasonal behavior with minimums during Autumn months and maximums at the beginning of Summer, where the highest reproductions of mite eggs occur, related to dry periods and where there is little rainfall, followed by an abrupt decline in the reproductive ranges, in early Autumn. Another factor that interferes with their reproduction is high temperatures above 27.9 °C and minimum temperatures of 22.8 °C (Chávez 2020).

The study of the population distribution of *Oligonychus* sp., during the development of the avocado in Perú was study by Cango et al. (2014), the population fluctuation of individuals (nymphs and adults) was recorded, which was obtained during the period 2013-2014, where it was found that this developed mainly between February and June, which coincides with the development of favorable climatic and agronomic factors for this pest.

Statistical risk modeling provides a holistic approach to describing and predicting the probabilities of adverse health responses based on individual exposures and other study variables. Rather than attempting to describe or simulate the detailed causal and biological processes that lead from exposure to response probabilities within an individual, statistical risk modeling relies on observed data relating exposures, covariates, and responses at the level of the whole organism or a population of organisms (Cox 2012).

The effective detection and control of diseases in humans and animals needs to consider spatial patterns of the disease's occurrence and any associated risk factors. This includes efficient data collection, management, and analysis. The integration of Geographic Information System (GIS) functionality into most modern disease information systems reflects a recognition of the importance of the spatial dimension of disease control (Pfeiffer et al. 2008). The spatial risk of pests and diseases in agricultural crops has been addressed by several authors such as Jang et al. (2006) and Nelson et al. (1999). *Oligonychus* sp. is considered a key pest, which the Hass variety is affected by fruit quality and

crop yields (Hoddle et al. 2002). Considering the above, the purpose of this work is to determine an efficient method to analyze the spatial risk of *Oligonychus* sp. through spatial geostatistical methods for risk analysis using remote sensing.

MATERIALS AND METHODS
Description of the study area
The Field data from an avocado farm of 1.1 ha of avocado (*Persea americana*) var Hass planted at high density, the distance between trees is 2 m and between rows 5 m was collected (Figure 1). The study area is located in the city of Barranca, a province located 190 km northwest of the city of Lima, Perú, in the coastal desert, in the foothills of the western slope of the central Andes (Lira-Camargo et al. 2020). Given this location, an almost total absence of precipitation was perceived, as well as a high

level of atmospheric humidity and persistent cloudiness. The average annual temperature at the study site is 17.5 to 19 °C, with an annual summer maximum of about 29 °C. Temperatures during the summer months, from December to April, range from 29 to 30 °C during the day and 21 to 22 °C at night (Ruiz et al. 2019).

The avocado trees have a height of 5.20 m, which were collected by the same expert. A total of 202 observations in which the measurement variable is the quantity of the crystalline mite *Oligonychus* sp. per leaf were collected with a Qfield (2019), which allows working on QGIS projects for field work equipment (QField 2019). The projection system with the European Petroleum Survey Group (EPSG) code 32718 - WGS 84/UTM zone 18S was used. Data manipulation and analysis was performed using R statistical software and QGIS software.

Figure 1. Location of the avocado cultivation study area.

Structural analysis of spatial dependence
For the estimation of the risk model, a geostatistical approach was applied; in this, the random variable Z(x)

with spatial index varies continuously across the spatial region D (Cressie 1992)**.** The behavior of the frequency distribution and the possibility of a transformation were

Rev. Fac. Nac. Agron. Medellín 76(2): 10309-10321. 2023

10311

evaluated, as geostatistical analysis and predictions are sensitive to outliers (Giraldo 2002).

Applied to agronomic sciences, geostatistics considers each sample value Z(x) it is associated with a position and uses this same dependence to make inferences about the distribution of the data, in this approximation a model of spatial variation that contains at least three components: first a general structure, which can be defined as a trend; a second, superimposed structure, related to spatial correlation and gradual variation; and finally a third component consisting of random variation caused by sampling errors or spatial variations at different scales sampling errors or spatial variations at scales smaller than the sample network (Cressie 1992; Li and Heap 2008).

In other words, it should be evaluated whether the process generating the data is stationary by mean or formally expressed as follows $E(Z(x_i)) = \mu$ (Giraldo 2002), to test this situation, a first-order polynomial surface was estimated and the significance of the model and the adjusted R^2 were evaluated, since this indicator essentially imposes a penalty for including additional predictors that do not contribute much to explaining the variation observed in the variable response (Diez et al. 2020).

Secondly, it is necessary to test whether the process is stationary in variance $Var(Z(x_i)) = \sigma^2$, i.e., whether the data is isotropic or anisotropic (Giraldo 2002; Plant 2018), this be analyzed by defining the behavior of spatial dependence in cardinal directions (Bivand et al. 2008). Thirdly, the spatial dependence must be modeled using the experimental semivariogram, $\gamma_{(h)} = 1/2N_{(h)} \sum \left[Z_{(x)} - Z_{(x+h)} \right]^2$ where $\Upsilon_{(h)}$ is the semivariance for all samples located in space, separated by the distance interval h; $N_{(h)}$ is the total number of pairs of samples separated by a distance interval h; $Z_{(x)}$ is the value of the sample at a location x; $Z_{(x+h)}$ is the value of the sample at interval distance h from x . Variogram modelling and estimation is important for structural analysis and spatial interpolation (Oliver and Webster 2015; Li and Heap 2008).

For the adjustment of the theoretical semivariogram, the method proposed by Bivand et al. (2008) was used, which utilizes non-linear regression to fit the coefficients. For this, a weighted sum of square errors $\sum_{j=1}^{p} w_j (\gamma(h) - \hat{\gamma}(h))^2$, with $\Upsilon_{(h)}$ the value according to the parametric model is minimized. A theoretical Spherical semivariogram model was adjusted, following Equation (1).

$$\gamma(h) = \begin{cases} 0, h = 0 \\ C_0 + C_1(1.5(h/C_2) - 0.5(h/C_2), 0 < h \le C_2) \\ C_0 + C_1, h > C_2 \end{cases} \quad (1)$$

The parameter C_0 represents the nugget value induced by the spatial error when the distance is smaller than the lag distance. Whereas the variance of the process is denoted by C_1 and a combination between C_0 and C_1, where C_0 is referred to as sill. The parameter C_2 is the range in which it indicates a distance from the origin to the point of sill achieved (Moonchai and Chutsagulprom 2020).

Predictive Geostatistical Interpolation and Validation
The parameters previously analyzed, anisotropy and spatial dependence, allowed us to determine the type of geostatistical predictor interpolator most suitable to be used, since the three large families of geostatistical interpolators differ in three basic conditions: 1) Simple Kriging: the process is stationary by mean and the population mean is known; 2) Ordinary Kriging: the process is stationary by mean and the population mean is unknown; 3) Universal and Residual Kriging: the process is not stationary by mean and the population mean is unknown (Bivand et al. 2008; Li and Heap 2008; Cressie 1992; Oliver and Webster 2015).

The process of validation and quality control of the interpolation is similar to that of traditional statistical models, evaluation of the behavior of the residual semivariogram, homoscedasticity, normality of residuals (Bivand et al. 2008; Cressie 1992; Gujarati and Porter 2010). Also, Root-mean-square deviation (RMSE) and Mean Error (ME) were estimated as a validation process, this is a standardized methodology used by Andrades-Grassi et al. (2020, 2021). In addition, it was estimated the z-score, computed as: $Z_i = \frac{Z(x_i) - \hat{Z}_{[i]}(x_i)}{\sigma_{[i]}(x_i)}$; with $\hat{Z}_{[i]}(x_i)$ the cross validation prediction for x_i, and $\sigma_{[i]}(x_i)$ the corresponding kriging standard error (Bivand et al. 2008).

Generation of Spatial Risk Model and Covariate Analysis

Finally, an indicator function of risk was estimated by *Oligonychus* sp., for this the kriging predictor method selected according to the validation process was used. To convert the prediction into a risk function, the Indicator kriging was used, in this type of kriging we are interested in the event: at a given point x the value Z(x) exceeds the level Z_0. The event can be represented by a binary function, the indicator function, valued 1 if the event is true, and 0 if it is false, whose expected value is the probability of the event Z(x), exceeds Z_0 (Chilès and Delfiner 2009). In this case it must be assumed that the process of the regionalized variable $\{Z(x) : x \in D \subset \mathbb{R}^d\}$ is stationary in which the following transformation is defined:
$$I(x_i, z_i) = \begin{cases} 1 \text{ if } Z_0 \leq Z(x) \\ 0 \text{ otherwise} \end{cases}$$
(Chilès and Delfiner 2009; Giraldo 2002). Values which are much greater than a given cut-off, will receive the same indicator value as those values which are only slightly greater than that cut-off (Giraldo 2002).

Thus, indicator transformation of data is an effective way of limiting the effect of very high values. Simple or Ordinary kriging of a set of indicator-transformed values will provide a resultant value between 0 and 1 for each point estimate. This is in effect an estimate of the proportion of the values in the neighborhood which are greater than the indicator or threshold value (Chilès and Delfiner 2009). In this case, the threshold was set between 50-100 motile forms of the Hass variety of *Oligonychus* sp. per leaf which is equivalent to 40% infestation, since leaf drop occurs when the percentage of damaged area is equal to or above 7.5-10 %, which corresponds to spider populations exceeding 100-500 individuals per leaf (Hoddle et al. 2002). As can be seen, this is a wide range, so the risk function cutoff was set at five thresholds (50, 60, 70, 80 and 100 count/leaf).

As it was evidenced, up to this moment exclusively univariate geostatistical models were evaluated, therefore, it is not known if there are mappable geomorphometric causal factors that have an influence on the growth of *Oligonychus* sp. mite, the Digital Elevation Model (DEM) of ALOS Parsal of 12.5 m was used (Racoviteanu et al. 2007), as this was the free product with the best spatial resolution.

It is known that the spectral response of vegetation changes according to whether it is affected or vigorous. The typical behavior of vigorous vegetation shows reduced reflectivity in the visible spectrum, but high reflectivity in the near infrared, gradually decreasing towards the mid-infrared.

These spectral characteristics are primarily related to the action of photosynthetic pigments and water stored in the leaves. Specifically, the low reflectivity in the visible portion of the spectrum is due to the absorbing effect of the leaf pigments. As for the high reflectivity in the near infrared, it is due to the internal cellular structure of the leaf. In particular, the spongy layer of the mesophyll, with its internal air cavities, plays a leading role by scattering and dispersing most of the incident radiation in this band of the spectrum (Harris 1987). Therefore, the healthy leaf offers high reflectivity in the near infrared (between 0.7 μm and 1.3 μm), in clear contrast to the low reflectivity it offers in the visible spectrum, especially in the red range (Crippen 1990).

There are different vegetation indices that could indicate if there is a potential problem in a crop and use broad information from the electromagnetic spectrum (Crippen 1990). The Normalized Difference Vegetation Index (NDVI) is defined as the parameter calculated from reflectance values at different wavelengths and is particularly sensitive to vegetation cover. The use of this index is based on the particular radiometric behavior of vegetation, which is characterized by the contrast between the red band (between 0.6 μm and 0.7 μm), which is largely absorbed by the leaves, and the near infrared (between 0.7 μm and 1.1 μm), which is mostly reflected. The NDVI is estimated as follows NDVI = (IRC - R) / (IRC + R), where IRC is the reflectivity of the near infrared band and R is the reflectivity of the red band.

The NDVI were generated from a Sentinel-2 image taken on October 10, 2020, at 10 m spatial (Copernicus 2021) resolution, this was preprocessed and transformed to reflectance using the Qgis SPC plug-in (Congedo 2021). It was as close as possible in time to the date of field spatial data collection of *Oligonychus* sp. Information derived from vegetation indices are not causal risk factors, but rather a consequence of the effect of mite damage. Low values of vegetation indices usually indicate low vigorous vegetation, while high values indicate very

Rev. Fac. Nac. Agron. Medellín 76(2): 10309-10321. 2023

10313

vigorous vegetation (Crippen 1990). As this DEM and NDVI do not have a spatial resolution and level of detail that is sufficient for geostatistical analysis, a multi-Gaussian conditional simulation process was performed. One of the main problems of Kriging predictions is that the smoothing of the map of estimated values is smoother than the map of true values (Chilès and Delfiner 2009).

A multi-Gaussian random function $Y(x)$, $\mathbb{R}^d$ of variogram $g(h)$ was simulated at the sites $x_1,... x_n$ in the space. The sequential simulation is performed as follows: 1) simulate a Gaussian value U_1 (mean 0, variance 2) with $Y(x_1)=U_1$. 2) for each $i\in\{2,...n\}$, with: $Y(x_i)=Y(x_i)^{SK} + \sigma_{SK}(x_i)U_i$, where $Y(x_i)^{SK}$ is the simple Kriging of $Y(x_i)$ from the previously simulated values $\{Y(x_1),...Y(x_i\text{-}1)\}$, $\sigma_{SK}(x_i)$ is the Kriging standard deviation U_i is a Gaussian value independent of $U_1,... U_{i\text{-}1}$. At each stage, the value at one site is simulated and added to the conditioning data, the set of simulated values $\{Y(x_1),...Y(x_i\text{-}1)\}$ has a multi-Gaussian distribution, with mean 0 and variogram $g(h)$, a simulation of the random function is obtained at the sites $x_1,...x_n$ (Giraldo 2002).

This procedure, although complex, allows evaluating the behavior of spatial autocorrelation and forces the exploration of the NDVI and DEM covariate data. Upon completion of this procedure, the DEM issued the Topographic Wetness Index (TWI), this indicator allows the identification of potential wetland concentration sites or water accumulation zones and the Topographic Position Index (TPI) (Sörensen et al. 2006). The conditional simulation process has already been used by Andrades-Grassi et al. (2021), for the improvement of continuous random variables of low spatial quality.

The Kriging predictions equations can be simply extended to obtain multivariable prediction equations. The general idea is that multiple variables may be cross correlated, meaning that they exhibit not only autocorrelation but that the partial variability of variable A is correlated with variable B, and can therefore be used for its prediction, and vice versa (Bivand et al. 2008). This statistical technique requires the availability of information of secondary variables at all points being estimated, co-Kriging is proposed to use non-exhaustive secondary information and to explicitly account for the spatial cross correlation between the primary and secondary variables (Goovaerts 1997; Li and Heap 2008).

Co-Kriging requires that both target and co-variable be, individually, spatially autocorrelated, and in addition that they be spatially cross-correlated. To perform co-Kriging, one must first model the spatial structure of a co-variable and its covariance with the target variable, this is called a co-regionalization, and it is an extension of the theory of a single regionalized variable (Bivand et al. 2008).

The easiest way to ensure this is to fit a linear model of co-regionalization: all models (direct and cross) have the same shape and range but may have different partial sills and nuggets. The Fit uses an iterative procedure of two steps: first, each variogram model is fitted to a direct or cross variogram; next each of the partial sill coefficient matrices is approached by least squares (Bivand et al. 2008). As with the model obtained with univariate indicator kriging the RMSE and Mean Error were estimated. It should be noted that co-Kriging was evaluated by introducing a single covariate (NDVI, TWI and TPI). Finally, the best co-Kriging prediction was transformed into an indicator function, using the previously described indicator Kriging.

RESULTS AND DISCUSSION

The results of the univariate count of *Oligonychus* sp.*,* show a double-peaked frequency distribution without the presence of additive outliers, so no transformation of the frequency distribution was required (Giraldo 2002). A stationary process in second order was identified and the condition of stationarity in first order by means was fulfilled, since the estimated parameters of the polynomial trend in first order were not significant (Table 1). The regression which shows a non-significant (R^2 adjusted = -0.008, P=0.87) process and a very low adjusted R^2 (Diez et al. 2020).

Also, the stochastic process is stationary by variance (Figure 2A). According to these two results what is being expressed is the effect of pure spatial dependence (Cressie 1992). A theoretical Spherical semivariogram model was adjusted, the parameter represents the nugget value induced by the spatial error when the distance is smaller than the lag distance. Whereas the variance of the process

Table 1. First-order polynomial model of *Oligonychus* sp. count in avocado trees.

Coefficients	Estimate	Std. Error	t value	*P*
Intercept	1.84×10^5	3.74×10^5	0.49	0.62
East	-8.8×10^{-3}	0.12	-0.06	0.94
North	-2.1×10^{-2}	0.04	-0.47	0.63

is denoted by and a combination between and is referred to as sill. The parameter is the range in which it indicates a distance from the origin to the point of sill achieved, the adjusted semivariogram has the characteristic that the spatial autocorrelation is significant at very short distances (8.75 m), in this case the sill represents the maximum variability in the absence of spatial dependence (Giraldo 2002; Oliver and Webster 2015), as shows in the Figure 2B.
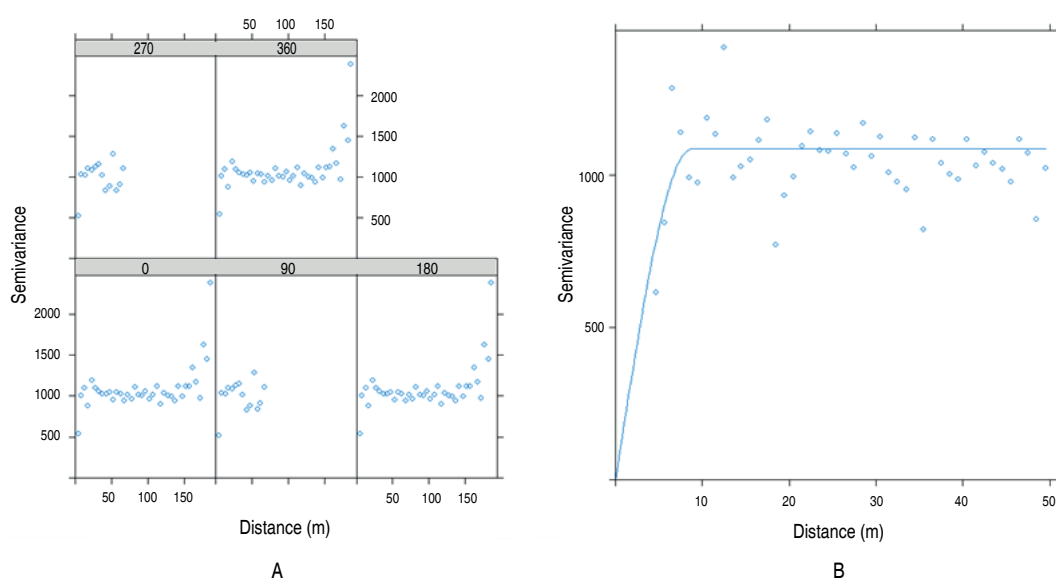

Figure 2. A. Directional semivariogram of *Oligonychus* sp. count in avocado trees; B. Adjusted semivariogram of *Oligonychus* sp. count in avocado trees.

The prediction of ordinary Kriging evidences a RMSE 32.72 counts/leaf and ME -0.16 counts/leaf, with a normal and homoscedastic distribution process of the residuals (*P*>0.05), ME is used for determining the degree of bias in the estimates (Hohn 1991) but it should be used cautiously as an indicator of accuracy because negative and positive estimates counteract each other and resultant ME tends to be lower than actual error (Nalder and Wein 1998), evidencing underestimation in forecasting. RMSE provides a measure of the error size, but it is sensitive to outliers as it places a lot of weight on large errors (Li and Heap 2008).

Based on this prediction, the indicator function is presented, which clearly shows a reduction of the risk zones (yellow) as the critical threshold, the incidence of *Oligonychus* sp. risk, increases. Also, the Probability As>100 mites/Leaf Figure shows the trees with the highest incidence of this mite, located to the north of the study area (Figure 3).

Two questions arise from these results, which are addressed in the following sections: 1) Which indicator function is the closest to reality? and 2) What is the influence of the covariates used in the prediction? To answer these questions, the starting point was the characterization of the spatial statistical process of the NDVI, TWI and TPI and, secondly, the co-regionalization model. The NDVI and DEM show a clearly non-stationary process by means, so it was

necessary to extract the trend for the multigaussian conditional simulation process (Table 2). Additionally, the TPI shows a clear first order gradient, but the TWI does not show a gradient.
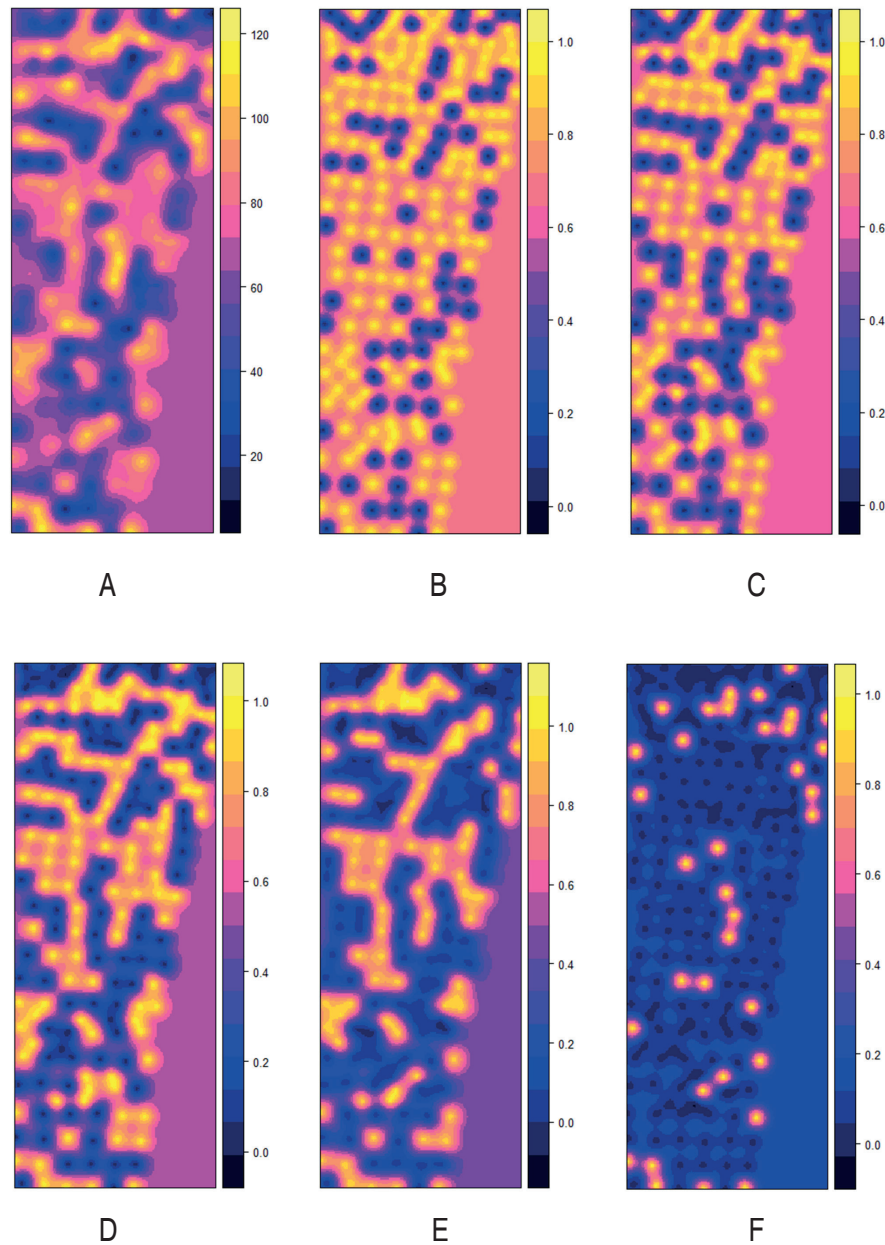

Figure 3. Univariate Predictions and risk indicator functions of *Oligonychus* sp. count in avocado trees. A. Prediction of Ordinary Kriging; B. Kriging Indicator > 50 mites/leaf; C. Kriging Indicator. >60 mites/leaf; D. Kriging Indicator > 70 mites/leaf; E. Kriging Indicator > 80 mites/leaf; F. Kriging Indicator > 100 mites/leaf.

The linear model of co-regionalization indicates that the process of the NDVI (Figure 4A) and TWI (Figure 4B) relationship are very similar in form, because they are possibly providing the same information as it is known the influence of soil moisture on the spectral response in the infrared and red bands (Bivand et al. 2008; Crippen

Table 2. First-order polynomial models of DEM, NDVI, TPI and TWI.

Variable	Coefficients	Estimate	Std. Error	t value	*P*
NDVI	Intercept	4.09×10^{-1}	8.55×10^{-3}	47.89	0***
	East	1.57×10^{-4}	6.31×10^{-5}	2.48	0.0138*
	North	-5.73×10^{-4}	6.31×10^{-5}	-9.07	0***
TPI	Intercept	1.97×10^{-1}	7.56×10^{-2}	2.60	0.00979**
	North	-2.18×10^{-3}	6.45×10^{-4}	3.38	0.00086***
TWI	Intercept	4.75	1.95×10^{-1}	24.32	0***
	East	4.49×10^{-4}	1.44×10^{-3}	0.31	0.755
	North	1.95×10^{-3}	1.44×10^{-3}	1.36	0.177

1990). It is possible to consider that both variables can be multicollinear (Gujarati and Porter 2010) and expressing the same information, both present a negative long-distance experimental covariance with the Count of *Oligonychus* sp.

The opposite occurs with TPI (Figure 4C), which shows a negative spatial correlation but with a short distance

that allowed the inclusion of an effect similar to that found in the univariate analysis, indicating that this morphometric variable has local influence. Finally, the results of the co-regionalization model for the *Oligonychus* sp. count with the NVDI/TWI and with the NDVI/TWI/TPI are presented; they show a similar behavior to the previously described models with an apparent multicollinearity and poor fit (Figure 4D, E).
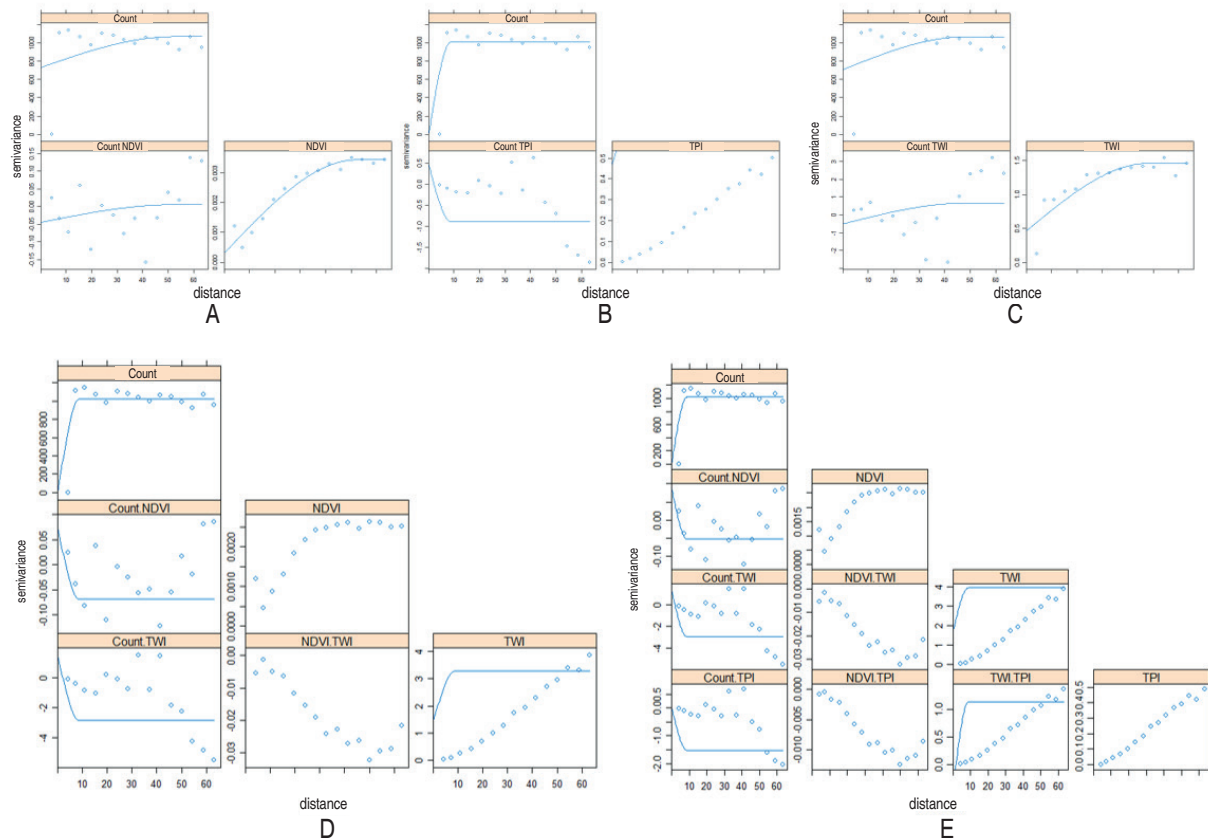

Figure 4. Linear model of co-regionalization models of count *Oligonychus* sp., A. NDVI; B. TWI; C. TPI; D. NDVI/TWI and E. NDVI/TWI/TPI.

Rev. Fac. Nac. Agron. Medellín 76(2): 10309-10321. 2023

10317

All five cases show a poor fit of the linear model of co-regionalization, since the assumptions of the co-regionalization model were not fulfilled. The easiest to fit a linear co-regionalization model is that all models (direct and cross) have the same shape and range but may have different partial sills and nuggets (Bivand et al. 2008). This was not achieved because the ranges for all cases of the covariates are not similar to those of the response variable. The situation of the predictions of the covariates evidences it, since the predictions made by the co-Kriging with the NDVI and TWI show greater distance amplitude, while the TPI shows short distances (Figure 5). It is evident that although the conditional multigaussian simulation was not effective in improving the level of spatial detail, it is convenient to recommend for this study the use of Drone technology to obtain high spatial resolution data. This would homogenize the spatial scales and could improve the prediction.
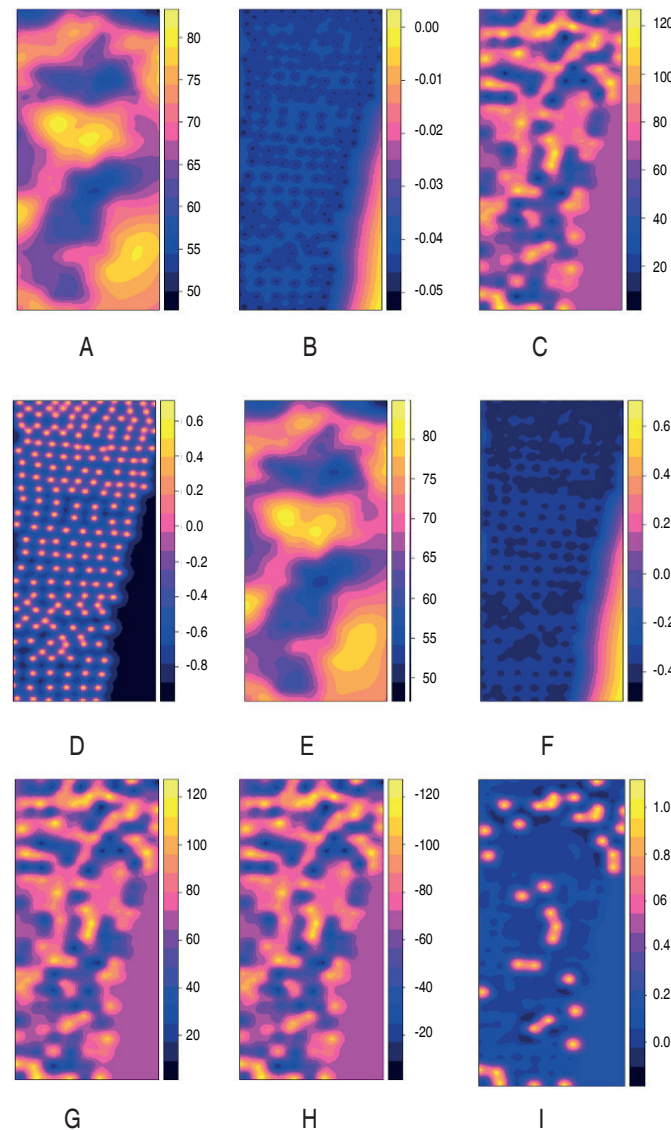

Figure 5. Co-Kriging Prediction, Covariance, and Indicator of count *Oligonychus* sp. A. co-Kriging Prediction with NDVI as covariable; B. co-Kriging Covariance with NDVI as covariable; C. co-Kriging Prediction with TPI as covariable; D. co-Kriging Covariance with TPI as covariable; E. co-Kriging Prediction with TWI as covariable; F. co-Kriging Covariance with TWI as covariable; G. co-Kriging Prediction with NDVI/TWI/TPI as covariable; H. co-Kriging Prediction with NDVI/TWI as covariable; I. co-Kriging Indicator > 100 mites/leaf NDVI/TWI/TPI as covariable.

As a selection rule, the best prediction is the one with the ME closest to zero and the lowest RMSE (Li and Heap 2008). One problem is that RMSE changes with the variability within the distribution of error magnitudes and with the square root of the number of errors (Willmott and Matsuura 2005). The RMSE is more appropriate to represent model performance than other indicators when the error distribution is expected to be Gaussian, the sensitivity of the RMSE to outliers is the most common concern with the use of this metric (Li and Heap 2008).

Considering these criteria, the best performances are those obtained by Univariate Ordinary Kriging, co-Kriging with NDVI as covariate and co-Kriging with TWI as covariate (Table 3), the co-Kriging errors are generally larger, especially near the subset sample points. This is because the co-Kriging predictions are based not only on the target variable, but also on the covariate at these points, as well as its covariance (Bivand et al. 2008). Another possible reason for these results may

be that significant differences in Perú were found in the populations of the pest according to the location in the tree (Cango et al. 2014), as it was not possible to obtain data with higher spatial resolution, therefore the results are not from co-Kriging they are not entirely consistent.

The Ordinary Univariate Kriging validation is superior in all measures given that the overall precision is greater, and the extreme errors are smaller. Since the RMSE of the multivariate predictions are affected by outliers (which generate an increase in the bias of the predictions) and despite the fact that linear co-regionalization models are consistent, there are two possible scenarios for this response: 1) The spatial resolution of the covariables remains spatially insufficient to represent the detail the variability of the mite effect, and the assumptions of the co-Kriging model did not met. 2) The relationship between both the spectral response and the amount of mite per plant is nonlinear, so the multivariate Kriging model is insufficient to satisfactorily represent a spatial risk model (Bivand et al. 2008).

Table 3. Residuals per interpolation method.

Kriging type	RMSE	ME	Min.	Max.
			Error	Error
Ordinary Kriging	32.72	-0.16	-63.33	60.38
Co-Kriging with NDVI	32.76	-0.17	-101.89	102.57
Co-Kriging with TPI	36.52	-0.16	-101.81	102.88
Co-Kriging with TWI	32.76	0.16	-101.69	103.02
Co-Kriging with NDVI/TWI	36.38	-0.19	-102.08	102.92
Co-Kriging with NDVI/TWI/TPI	36.39	-0.19	-102.02	102.84

Also, the sampling density may play a key role in the performance of the spatial interpolation methods. It is often argued that if the sample size is big enough, then the effects of sample size would disappear. Apparently, this assertion is not true, in intensely sampled cases, there is still a clear pattern whereas sample size increases the performance of the spatial interpolation methods continues to improve, because the effects of sampling density is also dominated by the variation in the data (Li and Heap 2008).

In this research, the sampling scheme did not include information about the properties of the site being

sampled. A reasonable size of the lag group is required to compute a good estimate of the semivariogram, so it is evident that a lot of data values are needed to adequately estimate the variogram. Authors such as Plant (2018), and Oliver and Webster (2015), state that between 100 and 200 samples should be used to generate an adequate experimental variogram.

The sampling efficiency is defined as the inverse of the sampling variance, it is desirable, from a statistical standpoint to select a configuration that minimizes the prediction error, the efficiency is calculated for all possible realizations of the variable Z by $\mathrm{Var}_{\xi}\left[Z_D^* - Z_D \right]$

Rev. Fac. Nac. Agron. Medellín 76(2): 10309-10321. 2023

10319

using σ_K^2 , which is the geostatistical prediction error. In terms of the sampling variance, stratified random sampling is at least always equally or more accurate than random sampling; its relative efficiency is a monotone increasing function of sample size. Ideally, the density of sample points should increase in locations exhibiting greater spatial variability. When the autocorrelation is a linear decreasing function of distance, stratified random sampling has a smaller variance than a systematic design (Quenouille 1949).

If the decrease in autocorrelation is not linear, yet concave upwards, systematic sampling is more accurate than stratified random sampling. A centered systematic design, where each point falls exactly in the middle of each interval, is more efficient than a random systematic sampling configuration (Bellhouse 1977).

CONCLUSIONS

A geostatistical method for a spatial risk model of the mite *Oligonychus* sp. was presented. The univariate ordinary kriging model performs better than the multivariate ones because they present minimal error and therefore better performance (despite the linear co-regionalization models consistency). Nevertheless, it is evident that the spatial covariates do not have sufficient spatial resolution to provide enough information for model building. Although there is evidence that they are relevant variables that can be mappable at a high resolution through the use of photogrammetric techniques for the construction of the spatial risk model, a possible way forward is the use of information from Drones or Lidar surveys. It allows obtaining high resolution products at low cost, with which a more efficient model could be built. Another possible solution is the use of nonlinear spatial statistics, which allows the interaction of complex relationships.

REFERENCES

Andrades J, Cuesta L, Camargo C et al (2020) Propuesta metodológica para la construcción y selección de modelos digitales de elevación de alta precisión. Colombia Forestal 23(2):34-46. https://doi.org/10.14483/2256201x.15155

Andrades-Grassi J, Rangel R, López-Hernández J et al (2021) Modelado y simulación del terreno del compartimiento 9, en la Reserva Forestal El Dorado-Tumeremo, Bolívar-Venezuela. Recursos Rurais (17):35-45. https://doi.org/10.15304/rr.id7496

Aponte O and McMurtry JA (1997) Damage on 'Hass' avocado leaves, webbing and nesting behaviour of *Oligonychus perseae* (Acari: Tetranychidae). Experimental & Applied Acarology 21(5): 265-272. https://doi.org/10.1023/A:1018451022553

Bellhouse D (1977) Some optimal designs for sampling in two dimensions. Biometrika 64(3):605-611. https://doi.org/10.1093/biomet/64.3.605

Bivand RS, Pebesma EJ, Gómez-Rubio V and Pebesma EJ (2008) Applied spatial data analysis with R. 10(2): 237-268. New York: Springer. https://doi.org/10.1007/978-1-4614-7618-4

Cango MN, Cabrejo CV, Quispe RQ et al (2014) Distribución poblacional de la arañita roja *Oligonychus* sp.(Acari: Tetranychidae), sobre árboles del palto (*Persea americana* Miller) en Lima, Perú. http://www.avocadosource.com/WAC8/Section_03/NarreaCangoM2015.pdf

Chávez Acosta R (2020) Fluctuación poblacional de *Oligonychus punicae* Hirts (Acari: Tetranychidae), y predatores en *Persea americana* Mill."palto", provincia de Virú, La Libertad, 2016 .(Disertación Ingeniería). Universidad Privada Antenor Orrego. Trujillo, Perú. 48 p.

Chilès JP and Delfiner P (2009) Geostatistics: modeling spatial uncertainty (Vol. 497). John Wiley & Sons. 703 p.

Copernicus (2021) Copernibus Open Access Hub. En, Data and Information Access Services, https://scihub.copernicus.eu/., accessed: July 2021.

Congedo L (2021) Semi-automatic classification plugin: A Python tool for the download and processing of remote sensing images in QGIS. The Journal of Open Source Software 6(64):3172, https://doi.org/10.21105/joss.03172

Cox LA (2012) Statistical risk modeling. In: Risk Analysis Foundations, Models, and Methods. International Series in Operations Research & Management Science, vol 45. Springer, Boston, MA. 556 p. https://doi.org/10.1007/978-1-4615-0847-2

Cressie N (1992) Statistics for spatial data. vol. 4, issue 5. Terra Nova, 613-617 p. https://doi.org/10.1111/j.1365-3121.1992.tb00605.

Crippen RE (1990) Calculating the vegetation index faster. Remote Sensing of Environment 34(1):71-73. https://doi.org/10.1016/0034-4257(90)90085-Z

Diez D, Barr C and Çetinkaya-Rundel M (2020) Introductory statistics for the life and biomedical sciences. Julie Vu David Harrington Derivative of OpenIntro Statistics Third edition. 348 p.

Giraldo R (2002) Introducción a la geoestadística: Teoría y aplicación. Universidad Nacional de Colombia, Bogotá. 94 p.

Goovaerts P (1997) Geostatistics for natural resources evaluation. Oxford University Press, Oxford. 477 p.

Guanilo A, Moraes G, Fletchmann C and Knapp M (2012) Phytophagous and fungivorous mites (Acari: Prostigmata, Astigmata) from Peru. International Journal Acaralogy 38(2):120-134.

Gujarati D and Porter D (2010) Econometría. México D. F.: Mc Graw Hill. 946 p.

Harris R (1987) Satellite remote sensing. An introduction. Geological Magazine, 125(3):314. https://doi.org/10.1017/S0016756800010335

Hoddle MS, Nakahara S and Phillips PA (2002) Foreign exploration for *Scirtothrips perseae* Nakahara (Thysanoptera: Thripidae) and associated natural enemies on avocado (*Persea americana* Miller). Biological Control 24(3):251-265. https://doi.org/10.1016/S1049-9644(02)00037-3

Hohn ME (1991) An Introduction to applied geostatistics: by Edward H. Isaaks and R. Mohan Srivastava, 1989, Oxford University Press, New York, 561 p. Computers Geosciences 17(3):471–473. https://www.usb.ac.ir/FileStaff/3299_2019-6-2-0-56-45.pdf

Jang CS, Liu CW, Lin KH et al (2006) Spatial analysis of potential carcinogenic risks associated with ingesting arsenic in aquacultural tilapia (*Oreochromis mossambicus*) in blackfoot disease hyperendemic areas. Environmental Science & Technology 40(5): 1707-1713. https://doi.org/10.1021/es051875m

Li J and Heap a D (2008) A review of spatial interpolation methods for environmental scientists. Geoscience Australia, Record 2008/23, 137 p.

Lira-Camargo ZR, Alfaro-Cruz SC and Villanueva-Tiburcio JE (2020) Contaminación sonora en la ciudad de Barranca-Lima-Perú. Investigación Valdizana 14(4):213-219. https://doi.org/10.33554/riv.14.4.744

Mendoza García LK (2020) Impacto de la producción de palta en la agroexportación peruana (2009-2018).Trabajo de Investigación. Universidad San Ignacio de Loyola. https://repositorio.usil.edu.pe/handle/usil/11493

Moonchai S and Chutsagulprom N (2020) Semiparametric semivariogram modeling with a scaling criterion for node spacing: A case study of solar radiation distribution in Thailand. Mathematics 8(12):2173. https://doi.org/10.3390/math8122173

Muñoz JL and Rodriguez A (2014) Ácaros asociados al cultivo del aguacate (*Persea americana* Mill) en la costa central de Perú. Agronomía Costarricense 38(1):217-221. https://doi.org/10.15517/rac.v38i1.15206

Nalder IA and Wein RW (1998) Spatial interpolation of climatic normals: test of a new method in the Canadian boreal forest. Agricultural and Forest Meteorology 92(4):211-225. https://doi.org/10.1016/S0168-1923(98)00102-6

Nelson MR, Orum TV, Jaime-Garcia R and Nadeem A (1999) Applications of geographic information systems and geostatistics in plant disease epidemiology and management. Plant Disease 83(4): 08-319. https://doi.org/10.1094/PDIS.1999.83.4.308

Oliver MA and Webster R (2015) Basic steps in geostatistics: the variogram and kriging (Vol. 106). Springer, New York. 100 p. https://doi.org/10.1007/978-3-319-15865-5

Pfeiffer DU, Robinson TP, Stevenson M et al (2008) Chapter 8 - spatial risk assessment and management of disease. pp. 110-119. In spatial analysis in epidemiology. Oxford University Press. 137 p. https://doi.org/10.1093/acprof:oso/9780198509882.001.0001

Plant RE (2018) Spatial data analysis in ecology and agriculture using R. CRC Press, Boca Raton. 684 p. https://doi.org/10.1201/9781351189910

Quenouille MH (1949) Problems in plane sampling. The Annals of Mathematical Statistics 20(3):355–375. http://www.jstor.org/stable/2236533

QField (2019) Legacy documentation. In QField Documentation, https://qfield.org/docs/en/old-doc/ . 2 p.; accessed: November 2021.

Racoviteanu AE, Manley WF, Arnaud Y and Williams MW (2007) Evaluating digital elevation models for glaciologic applications: An example from Nevado Coropuna, Peruvian Andes. Global and Planetary Change 59(1-4):110-125. https://doi.org/10.1016/j.gloplacha.2006.11.036

Ruiz L F, Carcelén F and Sandoval-Monzón R (2019) Evaluación de los indicadores de estrés calórico en las principales localidades de lechería intensiva del departamento de Lima, Perú. Revista de Investigaciones Veterinarias del Perú 30(1):88-98.

Sörensen R, Zinko U and Seibert J (2006) On the calculation of the topographic wetness index: evaluation of different methods based on field observations. Hydrology and Earth System Sciences 10(1):101-112. https://doi.org/10.5194/hess-10-101-2006

Suárez EH, Álvarez C, Álvarez ET et al (2023) Evolución de los problemas fitosanitarios del cultivo del aguacate en Canarias. Phytoma España: La revista profesional de sanidad vegetal (345): 31-37.

Willmott CJ and Matsuura K (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research 30:79-82. http://doi.org/10.3354/cr030079

Rev. Fac. Nac. Agron. Medellín 76(2): 10309-10321. 2023

10321