Research Article

Data Association Methodology to Improve Spatial Predictions in Alternative Marketing Circuits in Ecuador

Washington R. Padilla ¹ and Jesús García ²

¹Salesian Polytechnic University of Quito-Ecuador Engineer Systems, Research Group Ideia Geoca, Quito, Ecuador

²Carlos III University, Applied Artificial Intelligence Group, Madrid, Spain

Correspondence should be addressed to Washington R. Padilla; wpadillaa@ups.edu.ec

Received 9 May 2018; Revised 29 August 2018; Accepted 24 September 2018; Published 5 November 2018

Academic Editor: Cornelio Yáñez-Márquez

Copyright © 2018 Washington R. Padilla and Jesús García. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This work proposes a methodology that reduces the error of future estimations in commercialization based on multivariate spatial prediction techniques (cokriging) considering the products with strong associations. It is based on the *Apriori* algorithm to find association rules in sales of agricultural products of local markets. Results show the improvement in spatial prediction accuracy after using the best association rules.

1. Introduction

Family farming is an economic and social sector provider of food in the world that guarantees processes of food security within a country. In Latin America and the Caribbean, family farming is the main source of agricultural and rural employment, comprising 80% of farms that represent around 60 million people. This type of productive model combines agriculture, livestock, forestry, fishing, aquaculture, and grazing within the same farm and provides on average between 27 and 67% of the total food production for each country in Latin America [1].

This type of family farming has certain characteristics:

- (i) High presence of labor and family administration
- (ii) It is a diverse agriculture that allows self-sufficiency but also guarantees the feeding of other families through the surplus
- (iii) It is an agriculture that has limited access to productive resources such as land, water, and working capital compared to large-scale operations
- (iv) As an intangible heritage, family farming develops its own social and cultural dimension, which generates intergenerational links for the transfer of knowledge, traditions, and customs

- (v) Generates social and community ties through the generation of cooperatives

Due to its productive and social characteristics, family farming is not a sector solely focused on production but also on the commercialization of products. However, the active participation of these farmers, focused on producing food for consumption in markets, is part of the sustainable and participatory development of the sector. In this sense, there are several limitations such as the geographical dispersion of the different farms of family farming, the production volumes of each family farm, and the limited capacity to meet quality standards established by marketing chains that demand access to markets. As indicated by Contreras et al., a lack of adequate coordination between the consumer and producer in the production and marketing system is not allowing adequately to respond to new demands or dissatisfaction of the consumer [2].

In response to this problem, local initiatives have emerged from family farmers to access markets such as the alternative marketing circuit (CIALCO). These spaces that propose the direct encounter between producers and consumers, in recent years, have acquired a high importance within the agendas of public policy for the development of family farming. This importance is linked to the local assessment of production through the promotion of a more local food consumption focused on the assessment of the

agrobiodiversity of each territory [3]. These short marketing circuits are characterized as follows:

- (i) Low or no presence of intermediation for the commercialization of products
- (ii) Generation of bonds of trust and closeness between producers and consumers
- (iii) Assessment of the temporality of production of each product.

This type of direct marketing can be presented through different strategies such as public purchase, fairs, or local markets for the sale of food, among other modalities. In Ecuador, these initiatives promoted by family farmers are supported by the Ministry of Agriculture and Livestock through the strengthening of local fairs to meet producers and consumers, the baskets of family farming, the export of family farming products, and the public purchase, among other initiatives. At the same time, this promotion has the objective of generating agricultural, environmental, and social policies that improve public policies aimed at responding according to the challenges and needs faced by family farmers, which make public action and its impacts effective, equitable, and sustained development of this sector. So far one of the constraints for the generation of public policies according to the needs of family farming is the scarcity of information on production, but also the data related to the income obtained by the marketing of their products [4].

The research objective is to generate a methodology to improve the prediction of commercialization of different agricultural products using geostationary spatial data mining techniques, using the existing data corresponding to the year 2014 that allows generating future scenarios for the evaluation of public policies that help the development of the family agricultural sector.

The country Ecuador is crossed by the equatorial line, that means its territory is located both north and south of Latitude Zero (Figure 1(a)). At the south-central region are located the provinces of Tungurahua and Chimborazo (Figure 1(b)). From these two provinces, information has been collected on the sale of agricultural products in the so-called alternative marketing circuits (CIALCO).

The paper is organized in six sections, and it begins with a description of alternative marketing circuits, focuses on the main problem that is the lack of historical data that does not allow using statistical techniques, and presents the alternative use of algorithms used in data mining for the generation of future estimates in the commercialization of agricultural products generated by peasant families in specific places in the provinces of Tungurahua and Chimborazo in Ecuador.

In the second section, we present other works that use data mining techniques used in this research oriented to different domains establishing validity and probity of the algorithms used such as association rules, kriging, and cokriging, in addition to the relationship with topics oriented in the same line of research.

In Section 3, the theoretical description of the methods and data mining algorithms are presented, emphasizing their mathematical development, and the description of the

information provided on which the different processes are applied is also presented in this section.

Section 4 describes the methodology proposed to improve the process of future estimates in the commercialization of products, a multivariable function is generated using the products resulting from the association rules, and it is verified if the errors in the prediction tend to decrease.

In Section 5, the proposed methodology is applied and, with the values obtained, the percentage of error reduction in the predicted values is calculated using algorithms for multivariable.

To conclude in the last section, using the percentage of improvement in future predictions, a tomato production scenario is presented graphically in the provinces of Tungurahua and Chimborazo in Ecuador, which allows to establish policies to improve the functionality in this type of circuits.

2. Related Work

This section overviews some relevant previous works related to the developed research, both in theoretical and practical aspects. A series of studies conducted in various fields of science try to use the rules of association as a criterion to establish future estimates, so we can see some works such as [5], whose authors analyze the stock of a supermarket, or [6], to predict admission decisions for students. In works like [7], a relationship between association rules and a fuzzy classification is established.

In [8], an explanation of the mathematical development of kriging and cokriging based on substitution models within the framework of optimization is made, [9] and they propose to improve the construction of the variogram using information of magnitude and direction applied to data of the National Network of the Geomagnetic Observatories of China.

In spite of researching specific works, there is very little documentation on the improvement in future estimation using association rules, as of 2016 the works that focus on this topic in a specific way [10] give the first guidelines in a process to establish the most consumed products and the best association rules that produce the first patterns with the highest consumption of family farming in Ecuador, and using rules of association, here we can see the generation of the first scenarios in the consumption of the products.

In the second work [11], it is the set of products obtained from the application of the Apriori algorithm, inside an improvement that oscillates in a value between twenty and thirty percent, the estimates of future sales using time series considering the half-squared error, and this work establishes scenarios regarding the periodicity of a product.

The third work [12] focuses on the estimation of the commercialization of products using their geographical location and their relationship as an influence in the improvement of consumption predictions based on the set of items resulting from the application of association rules.

The research developed in the area of statistical science has the concepts mature enough to deal with this type of approach with great solvency; however, in this particular case, there is not a sequence of data of several years that allow to use statistical techniques. Available data are limited to

FIGURE 1: Geographic location: (a) the country Ecuador; (b) Tungurahua and Chimborazo.

2014, creating an appropriate scenario to test data mining techniques that allow to establish future estimates of product consumption, with these results it is expected to generate scenarios where the implementation of policies to improve alternative circuits can be evaluated by marketing.

3. Methods and Materials

This section presents in detail the theoretical basis of the data mining processes used, and a detail of the information on which the future estimates are made.

3.1. Association Rules and Apriori Algorithm. The first way to establish a relationship between products in this research is based on the number of times some products appear together in a sale transaction [13], and for this it is necessary to discretize the transactional data file, so that if a product is acquired, it is identified with a value T of true and F if it is not a part. This differentiation of products acquired allows to establish the minimum support that is known as the relationship between the number of times a product appears in a transaction with respect to the total of transactions made, and this process is repeated for a single item. Once the sets that meet for an item are established, we proceed to a similar calculation with two items and so on, identifying all sets that meet a preestablished minimum coverage, looking for the rules that meet a minimum of confidence, i.e., if the product appears in the antecedent of a rule, it has a minimum confidence level of appearing in the consequent of the rule.

The pseudocode is shown below.

3.1.1. Pseudocode Algorithm Apriori [14]

Step 1. Generate all item sets L with a single element; this set is used to form a new set with two, three, or more elements all possible pairs which are taken as Sup equals minsup

Step 2. For every frequent item, set L' is found:

For each subset J , of L'

Determine all association rules of the form:

If $L' - J \rightarrow J$

Select those rules whose confidence is greater or equal than minconf

Repeat Step 1, including next element into L

One of the best known algorithms to search for association rules is the Apriori method [15]. It is based on two parameters support and confidence:

- (i) The *support* of a rule is defined as the number of instances that the rule correctly predicts:

$$\text{sup}_a(x) = |x|,$$

$$\text{sup}_r(x) = \frac{|x|}{|D|}, \text{ being } D \text{ the total set of transactions,}$$

$$\text{sup}_a(A \Rightarrow B) = \text{sup}_a(A \cap B).$$

(1)

- (ii) The *confidence* indicates the percentage of times that a rule is met among the instances selected by the antecedent A:

$$\text{conf}(A \Rightarrow B) = \frac{\text{sup}_a(A \Rightarrow B)}{\text{sup}_a(A)} = \frac{|A \cap B|}{|A|}. \quad (2)$$

3.2. Spatial Estimation. As mentioned in [16], “in the geographical space everything is related to everything, but the closest spaces are more related to each other”. Geostatistics use the concept of a random function to find nondeterministic values on a region D , and if x crosses the region, a series of random variables are obtained, defined as

$$Z = \{Z(x), x \in D\}, \quad (3)$$

which constitutes a random function on domain D .

To simplify the feature of the random function, we consider some descriptive parameters or moments that summarize the information, the expectation, or first-order moment $m(x) = E[Z(x)]$, represent the average around which the values taken by the realizations of the random function are distributed, and the variance is calculated as follows:

$$\begin{aligned} \sigma^2(x) &= \text{var}[Z(x)] = E\{[Z(x) - m(x)]^2\} \\ &= E[Z(x)^2] - m(x)^2, \end{aligned} \quad (4)$$

and the variance and its square root called standard deviation constitute measures of dispersion of $Z(x)$ around its mean value; the covariance centered between two random variables is given by the relationship

$$C(x_1, x_2) = E[Z(x_1)Z(x_2)] - m(x_1)m(x_2), \quad (5)$$

and gives us an elementary vision of the interaction that exists between $Z(x_1)$ and $Z(x_2)$, and the semivariogram, defined between the two random variables, is given by the expression

$$\gamma(x_1, x_2) = \frac{1}{2} \text{var}[Z(x_1) - Z(x_2)], \quad (6)$$

and it reflects the way in which a point has influence on another point at different distances. The variogram is equal to the variance minus the covariance:

$$\gamma(h) = C(0) - C(h). \quad (7)$$

3.2.1. Experimental and Modeling Variogram. If we consider the z regionalized variable known in n sites $\{x_1, \dots, x_n\}$, the estimator of the experimental variogram for a separation vector h , it is defined as follows:

$$\gamma(h) = \frac{1}{2} E\{[Z(x+h) - Z(x)]^2\}. \quad (8)$$

An experimental variogram cannot be used because it is defined only for certain distances and directions, to interpret the spatial continuity of the study variable, and a theoretical model should be adjusted around the experimental variogram.

A variogram, $\gamma(h)$, is isotropic if it is identical in all directions of the space and if it does not depend on the orientation of the vector h but only on its magnitude $|h|$; otherwise, there is anisotropy in its distribution [17].

In general, the modeled variogram grows from the origin and stabilizes at a distance a , around a plateau; the two random variables $Z(x)$ and $Z(x+h)$ are correlated if the length of the separation vector h is less than the distance a , called the reach or zone of influence, beyond $|h| = a$; and the variogram is constant and equal to its plateau.

A spherical variogram of reach a and plateau C is defined as

$$\gamma(h) = \begin{cases} C \left\{ \frac{3}{2} \frac{|h|}{a} - \frac{1}{2} \left(\frac{|h|}{a} \right)^3 \right\}, & |h| \leq a, \\ C, & \text{otherwise.} \end{cases} \quad (9)$$

In processes involving geostatistics, the spatial correlation is modeled by the variogram, and this process is generated by a random function $Z(s)$ composed by the mean (m) and the residue $e(s)$: $Z(s) = m + e(s)$, with an average constant $E(Z(s)) = m$, and the variogram defined as $\gamma(h) = 1/2 E(Z(s) - Z(s+h))^2$. The variance of Z is constant, and the correlation of Z does not depend on the location s but only on the separation distance h . Then, we can form multiple pairs $\{Z(s_i), Z(s_j)\}$, that have identical separation vector $h = s_i - s_j$, and we estimate the correlation between them [18, 19]. An experimental variogram $\gamma(h)$ is isotropic if it is identical in all directions of space; otherwise, there is anisotropy.

If we assume the entropy is in the independent direction of the semivariance, we replace the vector h with the magnitude $\|h\|$. Under this assumption, the variogram can be estimated for $N(h)$ as a simple pair of data $Z(s_i), Z(s_i + h)$. For some distances (intervals), $\bar{h}_j$ is defined as

$$\hat{\gamma}(\bar{h}_j) = \frac{1}{2N(h)} \sum_{i=1}^{N_h} (Z(s_i) - Z(s_i + h))^2, \quad \forall h \in \bar{h}_j, \quad (10)$$

and this estimate is called a simple variogram.

The experimental variogram [20, 21] measures the average dissimilarity between two data as a function of their separation, often presents slope changes, which indicate a change in spatial continuity from certain distances, and the variogram can be modeled as the sum of several elementary models called models nested or nested structures [22]

$$\gamma(h) = \gamma_1(h) + \gamma_2(h) + \dots + \gamma_s(h). \quad (11)$$

The adjustment to a model is not done considering only the experimental variogram, but it must consider all the available information on the regionalized variable, and a more detailed explanation can be found in [23].

3.2.2. Estimation with Kriging. The kriging method in this case is considered as a linear prediction with unbiased linear estimator, and there are some types of kriging depending on the average of the known population. These types are ordinary and simple, and for this study, we are interested in the ordinary type.

The regionalized variable is the obtaining of the stationary Z random function that fulfills

$$\begin{cases} \forall x \in V, & E[Z(x)] = m \text{ unknown,} \\ \forall x, x+h \in V, & \text{cov}[Z(x+h), Z(x)] = C(h), \end{cases} \quad (12)$$

where V is the neighborhood considered in the kriging process. The following conditions are considered:

(i) Linearity:

$$Z^*(x_0) = a + \sum_{\alpha=1}^n \lambda_{\alpha} Z(x_{\alpha}). \quad (13)$$

where x_0 is the place where an estimate is established, $\{x_{\alpha}, \alpha = 1, \dots, n\}$ are the sites with known data, and $\{\lambda_{\alpha}, \alpha = 1, \dots, n\}$ are the weights that together with “a” they are the unknowns.

(ii) *Unbiased estimation constraint*: it is expressed that the expectation of estimation error must be zero:

$$\begin{aligned} E[Z^*(x_0) - Z(x_0)] &= 0 \\ &= a + \sum_{\alpha=1}^n \lambda_{\alpha} E[Z(x_{\alpha})] - E[Z(x_0)] \\ &= a + m \left(\sum_{\alpha=1}^n \lambda_{\alpha} - 1 \right). \end{aligned} \quad (14)$$

(iii) *Minimum variance*: find weights that minimize the variance of the estimation error:

$$\begin{aligned} \text{var}[Z^*(x_0) - Z(x_0)] &\text{it is minimal} \\ &= \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(x_{\alpha} - x_{\beta}) + C(0) - 2 \sum_{\alpha=1}^n \lambda_{\alpha} C(x_{\alpha} - x_0). \end{aligned} \quad (15)$$

Being the variogram, a tool equivalent to the covariance from the relationship,

$$\gamma(h) = C(0) - C(h). \quad (16)$$

The calculation of kriging is done as follows:

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} \gamma(x_{\alpha} - x_{\beta}) - \mu = \gamma(x_{\alpha} - x_0) & \forall \alpha = 1 \dots n, \\ \sum_{\alpha=1}^n \lambda_{\alpha} = 1, \end{cases}$$

$$\begin{pmatrix} \gamma(x_1 - x_1) & \dots & \gamma(x_1 - x_n) & 1 \\ \vdots & & \ddots & \vdots \\ \gamma(x_n - x_1) & \dots & \gamma(x_n - x_n) & 1 \\ 1 & & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ -\mu \end{pmatrix} = \begin{pmatrix} \gamma(x_1 - x_0) \\ \vdots \\ \gamma(x_n - x_0) \\ 1 \end{pmatrix}. \quad (17)$$

3.2.3. Multivariate Prediction: Cokriging. In this case, multiple spatial variables are analyzed together to build the prediction model. The first step is modeling a multivariable variogram, and the main tool for estimating semivariances between different variables is the crossed variogram, defined as follows:

$$\gamma_{ij}(h) = \frac{1}{2} E[(Z_i(s) - Z_i(s+h))(Z_j(s) - Z_j(s+h))]. \quad (18)$$

Two variables can have cross correlation, which means that the variables not only exhibit autocorrelation but that the spatial variability of a variable A is correlated with variable B, and vice versa. This can be extended to multiple variables; the measurements are taken in a limited set of locations, and the interpolation can be made to an unlimited number of locations. The cokriging seeks to estimate the value of a variable considering the data of this variable and other correlated variables, for this uses the following relationships [24, 25].

The crossed variogram between two variables Z_1 and Z_2 is defined as follows:

$$\gamma_{12}(h) = \frac{1}{2} \text{cov}\{Z_1(x+h) - Z_1(x), Z_2(x+h) - Z_2(x)\}, \quad (19)$$

and can be computed from the available data:

$$\hat{\gamma}_{12}(h) = \frac{1}{2|N(h)|} \sum_{N(h)} [Z_1(x_{\alpha}) - Z_1(x_{\beta})][Z_2(x_{\alpha}) - Z_2(x_{\beta})], \quad (20)$$

where $N(h) = \{\alpha, \beta, \text{ such that } x_{\alpha} - x_{\beta} = h\}$, being both variables z_1 and z_2 measure in x_{α} and x_{β} .

3.3. Materials. The analysis is based on information from 2014, provided by the General Coordination Network Marketing Ministry of Agriculture and Livestock of Ecuador. It contains the weekly performance of sales of agricultural products made by small farmers located in Ecuador's central highlands specifically the provinces of Tungurahua and Chimborazo. The available data contains information about the number and volume of sales of products such as vegetables, legumes, meat, dairy, fruits, tubers, and processed products, finding an average of 1,200 items per month divided on a weekly basis.

The elected products that have greater relevance in relation to information in the universe to be part of the research consists of thirty products, indicated in Table 1 (it contains the names in English, scientific name, and Spanish, the original language of the study). Further details of this dataset can be appreciated in the initial part of the investigation [26].

The available record contains the products that are part of the marketing, the value of sales, date, and fair to which each transaction belongs (Table 2).

On the one hand, the first data sheet contains all the recorded transactions, organized in packages named “canastas” (baskets), each one representing a sale of certain products, containing the products present in each purchase, and implicitly also contains the spatial geolocalization of the operation (the location of the fair) and the time stamp (date) of operation. As shown in Table 3, the transactions have the dates and, for each of the 30 products, the label with character “s” denotes it was present in transaction, and otherwise “No”. This table contains 550 transactions recorded along all months in the year 2014. This table

TABLE 1: Set of selected agricultural products.

Spanish name	Scientific name	English name
Acelga	<i>Beta vulgaris</i> var. cicla	Chard
Ajo	<i>Allium sativum</i>	Garlic
Arveja	<i>Pisum sativum</i>	Vetch
Babaco	<i>Carica pentagona</i>	Babaco
Brócoli	<i>Brassica oleracea</i> italica	Broccoli
Cebolla blanca	<i>Allium fistulosum</i>	White onion
Cebolla paitaña	<i>Allium fistulosum</i>	Onion
Choclo	<i>Zea mays</i>	Corn
Col	<i>Brassica oleracea</i>	Cabbage
Col verde	<i>Brassica oleracea</i> var. sabellica	Green cabbage
Coliflor	<i>Brassica oleracea</i> var. botrytis	Cauliflower
Espinaca	<i>Spinacia oleracea</i>	Spinach
Frejol	<i>Phaseolus vulgaris</i>	Frejol
Frutilla	<i>Fragaria</i>	Strawberry
Habas	<i>Vicia faba</i>	Broad beans
Hierbas	<i>Coriandrum sativum</i> , <i>Petroselinum crispum</i>	Weeds
Lechuga	<i>Lactuca sativa</i>	Lettuce
Meloco	<i>Ullucus tuberosus</i>	Meloco
Nabo	<i>Brassica rapa</i>	Turnip
Papas	<i>Solanum tuberosum</i>	Potatoes
Pepinillo	<i>Cucumis sativus</i>	Pickle
Pepino	<i>Cucumis sativus</i>	Cucumber
Pimiento	<i>Capsicum annum</i>	Pepper
Rábano	<i>Raphanus sativus</i>	Radish
Remolacha	<i>Beta vulgaris</i>	Beet
Tomate de árbol	<i>Solanum betaceum</i>	Tamarillo
Tomate riñón	<i>Solanum lycopersicum</i>	Tomato
Vainita	<i>Phaseolus vulgaris</i> L	Vainita
Zanahoria	<i>Daucus carota</i>	Carrot
Zapallo	<i>Cucurbita máxima</i>	Zapallo

containing binary attributes was the base for the association analysis performed in the first place.

On the other hand, Table 3 contains the sales value reported for each product aggregated in weeks and locations. It contains the numerical attributes reflecting the weekly variation of sales, with a blank space when there is no registered value. This second table, containing the sales information of 48 weeks with a total of 1260, was based to carry out the prediction analysis.

4. Proposed Methodology

The proposed methodology to improve the prediction of commercialization of products consists in searching the set of elements with the highest degree of associativity in commercialization. It is used to reduce the error in the spatial estimate of commercialization of agricultural products. It consists in the following steps:

- (i) Establish a baseline with future estimated values for the marketing of agricultural products, using the deterministic method IDW (inverse distance weight)
- (ii) Establish the set of associated products (using the *Apriori* algorithm of association rules)

(iii) For the (unique “u”) selected product,

- (a) Establish the experimental variogram model and the theoretical variogram that best suits the existing data (the adjusted variogram)
- (b) Calculate the estimate of the behavior of a future product based on the prediction (kriging)
- (c) Carry out the cross validation to estimate the error of the waste (CVu residuals)

(iv) For the set of products associated with the highest transaction ratio (multivariable “mv”)

- (a) Verify the correlation between the selected elements
- (b) Repeat steps 3.3.1 to 3.3.2 for the multivariable set
- (c) Carry out cross validation cokriging (CVmv residuals)

(v) Compare the residual values obtained in the two cases (Cvu and CVmv) with the future estimate values of IDW

Detail of implementation and results obtained from the proposed methodology are done in the following section.

5. Results and Discussion

5.1. Experimental Analysis

5.1.1. Data Processing. The proposed methodology has been applied to the marketing information of agricultural products provided by the Ministry of Agriculture of Ecuador, of the result collected from the different fairs located in Tungurahua and Chimborazo, with the data of sale of products of the month of July the year 2014 (Figure 2(a)).

To implement, the proposed methodology, we use the mathematical algorithms found in the R language, and the libraries used are SpatialPoints (sp), ggmap, tmap, ggplot2, GADMTools, rgeos, gdalUtils, gstat, geoR, proj4, crs, raster, maps, readr, in version 1.0.143 [27–30], and Weka 3.7 [31], and the generation of association rules is carried out.

The first activity is centered in the creation of the grid or mesh [32, 33] to determine the prediction area, and a dimension structure is defined with parameters: cellcentre.offset $x = -79.1085$, $y = -2.531218$, cellsize $x = 0.05$, $y = 0.05$, cells. dim $x = 21$; $y = 32$. In the sector of the equatorial line one degree of length equals 111.32 km, the distance occupied in length by the two provinces $x_{\min} = -79.133499$ and $x_{\max} = -78.0834991$ is 1.049 degrees, the equivalent to 116 km, for the conformation of the grid (spgrd1c), and the distance between cells is 5.84 km (Figure 2(b)).

5.1.2. Search for Association Rules. To find association rules, information must be quantized, so you can identify whether an agricultural product is part of the procurement process.

If part of the transaction is the label with the character “T”, otherwise “F” for all months of 2014, to optimize the process of searching for the best value association rules is replaced with “F” by the symbol “?”.

TABLE 2: Transactions with products contained in each sale ("canasta").

	10/5/2014	10/5/2014	10/5/2014	10/5/2014	10/5/2014	10/5/2014	10/5/2014	10/5/2014	10/5/2014	10/5/2014	10/5/2014
Acelga	s	s	No	No	s	No	s	No	s	No	s
Ajo	No	No	No	s	No	s	s	s	No	No	No
Arveja	No	s	No	No	s	s	s	No	No	s	No
Babacos	No	No	No	No	No	No	No	s	No	No	No
Brócoli	s	s	No	s	No	s	s	s	s	No	No
Cebolla blanca	No	No	s	s	s	No	s	s	s	s	s
Cebolla paiteña	s	No	s	No	s	No	No	No	s	s	No
Choclo	s	s	s	No	s	s	s	No	No	s	No
Col	s	No	s	s	s	s	s	No	s	No	s
Coles Verde	No	No	No	No	No	No	No	No	No	No	No
Coliflor	No	No	No	s	No	s	No	No	No	No	No
Espinaca	No	No	No	No	No	No	No	No	No	No	No
Frejol	No	s	No	s	s	s	No	No	No	s	No
Frutilla	No	No	No	s	No	No	s	s	No	s	No
Habas	No	No	No	s	No	s	s	s	No	s	s
Hierbas	s	No	No	No	s	No	No	No	No	s	No
Lechuga	s	No	No	s	s	s	s	No	s	No	No
Melloco	No	s	No	No	No	No	s	No	No	No	No
Nabo	No	No	No	No	No	No	No	No	No	No	No
Papas	No	s	s	s	s	s	s	No	s	No	No
Pepinillo	No	No	s	s	No	No	No	No	s	s	No
Pepino	No	No	No	No	No	No	s	No	No	No	No
Pimiento	No	s	s	s	s	s	s	s	No	s	No
Rabano	No	No	s	s	s	No	No	s	No	No	s
Remolacha	No	No	No	s	No	No	s	No	s	No	No

The Apriori algorithm for association rules is applied to a set of 550 transactions, with minimum support parameters equal to 0.4 (220 occurrences) and a confidence of 0.8. The resulting set is

- (i) Each time a white onion transaction is made, a tomato transaction is performed with a confidence of 87%, tamarillo (86%), carrot (83%), and broccoli (82%) (Figure 3), and each one of these elements generates a rule of association with the tomato, which constitutes the set of multivariable.
- (ii) The product with the highest commercial ratio of the study sample is tomato.
- (iii) The set of greater associativity is structured as $A = \{\text{Tomato, White Onion, Tamarillo, Carrot, Broccoli}\}$ Figure 4.

With the set of best association rules, the source of data is generated on which the different estimation processes are carried out in the future, and this file is called Fjespacial Table 4.

5.1.3. Baseline Analysis (IDW). In order to establish a baseline of analysis, the deterministic method inverse distance weighting (IDW) is used to calculate a first estimate in the future using the set of products with the greatest associativity such as broccoli, white onion, tomato, tamarillo, and carrot established in Section 5.2. The prediction of consumption for a single variable, idw

((TOMATE)~1, Fjespacial, spgridtc), where the variable to predict is the tomato, Fjespacial contains the values of sales, and spgridtc is called the grid or area where the prediction is made (Figure 5(a)).

5.2. Spatial Data Analysis. The data used correspond to the commercialization of the input denominated tomato of the month of July 2014, in the provinces of Tungurahua and Chimborazo, this file is converted to a spatial type, transforming the location data x and y into geographic coordinates [17, 34–36], that represent the latitude and length of each of the alternative circuits of commercialization type fairs that act in the study, and the fourth column corresponds to the values of the behavior of commercialization of tomato.

5.2.1. Theoretical and Experimental Variogram Models. The distance between the points that identify the fairs is expressed in tenths of a degree, and between each jump, there is a distribution of two fairs.

The model variogram (m), of the spherical type with a range of 0.157, is where the spatially correlated points are found, with a plateau equal to 2151 and distance 0,473.

In Figure 3, the adjusted variogram can be observed using the experimental variogram and the model variogram.

TABLE 3: Value sale products.

Fecha	Acéga	Ajo	Arveja	Babacos	Brócoli	Cebolla blanca	Cebolla colorada	Cebolla paitera	Choclo	Col verde	Coliflor	Espinaca	Frejol	Frutilla	Habas	Hierbas	Lechuga	Mellico	Nabo	Papas	Pepinillo	Pepino	Pimiento	Rabano	Remolacha	Tomate de arbol	Tomate riñon	Vainita	Zanahoria	Zapallo	
1/5/2014	13.75	21	82	9	32.8	44.5	26	41	56	22.5	8	11	42.5	5	38	84.5	23.5	18.25	26	3	180	16.5	12	17	6.5	3.5	42	121	2.5	29	8.5
1/12/2014	14.75	21	65	9	30.3	51.25	26	52	56	22.5	8	14	0.25	5	1	97	23.5	30.25	26	3	198	16.5	12	20	6.5	5.5	42	131	2.5	29	8.5
1/19/2014	11.25	21	51	9	25.5	44	16	28	16	30	9	9	0.25	5		85.5	19.5	18.25	16	3	160	12.5	12	16	4.5	5.5	30	139	0.5	24	8.5
1/26/2014	11.25	21.5	52	9	32.8	44.5	26	36	15	5	8	15	42.5	15		89.5	18	29.25	26	3	156	16.5	12	14	4.5	8.5	30	98	2.5	26	8.5
2/2/2014	11.25	3.5	76	9	15.9	22.5	41	75	5	9	17.3	6.5	34	52	92.5	19	29.6	25.6	3	156	48	12	14	5.25	10.5	39	6	6	162	8.5	
2/9/2014	7.5	6.5	28.5	51	7.6	26	6	37.5	40	13	4	3.7	3	37	87	60	12.5	21.4	4.4	4.7	151	47	48	54.5	9.5	6.5	131	6.5	178	5	
2/16/2014	5	3.5	13.3	10	2.5	9.5	5	11	17	4	17.5	7.5	6	12	34	45	5.5	8	1.7	82	24	8	37	6.25	27.5	10	5	5	114	3	
3/2/2014	11.5	35	78	10.5	21	52.75	26	87	36	17.5	20.75	2.5	28	37	92	18	23.5	31	3	78	24	12	15	5.25	5.5	77	113	6	38	8.5	
3/9/2014	5.25	27	20	16	40.2	62.75	30	47.35	24.5	7.2	4.25	5	18.4	50	15	7.5	36.75	10	6	45.5			25.5	11	2	110	133.75		37.5	0.6	
3/16/2014	3.55	37	22	6	28.3	158.25	64.4	50.75	34.5	12.8	4.25	7	10.4	50	22	32.7	29.55	12.5	6	103		5	36.5	46	7.5	151	95.9		20	0.6	
3/23/2014	14	18	13	6	32.4	40	49		107	29.6	7.5	7	13	50	37	15.5	43.5	47	4.75	65	2.5		26	4.5	7.5	151	95.9		48	5	
3/30/2014	14	18	13	6	25.6	49	49		89	23.2	7.5	7	13	50	37	14.5	40.5	34	3.75	65	5		25	8	7.5	149	99.9		58	5	
4/6/2014	2.5		25	30	4			15	20	15.75			40	20	6		15	23	200	5		30			5	32	58		38		
4/13/2014	1.25	2	31.5	21.25	28	22.15	18.8	15	35.6	18.15	4	1.2	29.5	19.5	11	9.25	23.4	25.5	7.7	140		3	23.15	9.25		23	53.6		26.5		
4/20/2014	3.5	2.4	22	10	40.25	79.3	6.8	42	64.2	33.45			9.6	35	16.5	3.25	50.75	24	3.25	319	30		24.9	9.95	1.75	85	138.1	1.5	54.5		
4/27/2014	14	18	13	6	25.6	20	24		71.15	23.2	10	7	13	50	34	14.5	40.5	34	27.3		125		25	8	7.5	149	71.1		67	5	
5/18/2014	12	23	17	36.25	2735	30.6		65.75	93		2.5		38.4	45	46	7.5	57.8	33	6.75	378	91.5		32.5	3.5	8	88	126.65		64.5	0.6	
5/25/2014		14	12	16.5	20.35	23.6		39.8	61.2				13	13	16.5	4.5	36	33.5	4.25	319	75		31.5	11	3	75	67.2	3	34	3	
6/1/2014	12.75	66	39	26.5	32.2	46	100.75		53.5	23.5	5		81.4	31	62	9.5	35	52	6	441	13		34.25	14.75	8	108	165.75	5.75	68.5	4.6	
6/8/2014	12.5	15	14	6	19.65	43	33.4		160	18.15	15.5	28	50	18	15.7	40.5	36.5	3.75	361	70	28		7	2.5	158	106.9		42.5			
6/15/2014	4.5	7	17	36.25	17.25	78.6	47		75	11.6	2.5		30	45	40		36.2	33	0.75	360	9.5		25	3.5	5	63	107.9		52		
6/22/2014	13.25	96	24	15	36.45	197	77.5		52	13.65	9	7.75	39.5	68	93	6.25	52.5	11	3.25	322	16.75		28.5	1	0	82	76.5	10	60.5	5	
6/29/2014	13.25	96	24	15	36.45	190.75	77.5		52	13.65	9	7.75	39.5	68	93	6.25	52.5	11	3.25	358	16.75		28.5	15	0	82	76.5	10	60.5	5

FIGURE 2: Prediction mesh: (a) fairs location; (b) mesh; (c) prediction area.

FIGURE 3: Tomato variogram: (a) variogram model; (b) adjusted variogram

Apriori
=====

Minimum support: 0.4 (220 instances)

Minimum metric <confidence>: 0.8

Number of cycles performed: 12

Generated sets of large itemsets:

Size of set of large itemsets L(1): 14

Size of set of large itemsets L(2): 15

Best rules found:

1. CebollaBlanca=t 338 → Tomate Riñon=t 293 <conf:0.87> lift:(1.1) lev:(0.05) [27] conv(1.58)
2. Tomate de arbol=t 312 → Tomate Riñon=t 268 <conf:0.86> lift:(1.09) lev:(0.04) [23] conv(1.49)
3. Zanahoria=t 374 → Tomate Riñon=t 311 <conf:0.83> lift:(1.06) lev:(0.03) [17] conv(1.26)
4. Brocoli=t 327 → Tomate Riñon=t 269 <conf:0.82> lift:(1.05) lev:(0.02) [12] conv(1.19)

FIGURE 4: Association rules.

TABLE 4: FJespacial data source.

Feria	Brócoli	Cebolla blanca	Tomate	Tomate de arbol	Zanahoria	x	y
Colta	6	60	26	30	30	-78.723827	-1.8880031
Tisaleo	37.9	36	88.25	185	47.5	-78.6922532	-1.4095144
Riobamba1	8	9.6	18	0	0	-78.6736758	-1.6615342
Riobamba2	26	42.5	60	15	17.5	-78.6729892	-1.6608907
Riobamba3	0	10	8	8	5	-78.6686976	-1.6602473
Riobamba4	2.5	74	60	35	29.5	-78.6588056	-1.6762586
Cevallos	9	9.6	51	45	20	-78.6566384	-1.2571434
Riobamba5	7.2	50	30	35	20	-78.655823	-1.6871008
Riobamba6	6	65	42	27	29	-78.6538275	-1.6759905
Riobamba7	2	0	10	0	5	-78.6500294	-1.6780281
Riobamba8	6	26.5	33	0	36	-78.6497031	-1.6746817
Riobamba9	2.5	0	6.9	0	3	-78.6462744	-1.6894172
Chambo	21	50	30	20	20	-78.6076928	-1.7302981
Pillaro	20.1	141.5	92	78	40	-78.551033	-1.3283574

FIGURE 5: Tomato sales estimate: (a) IDW; (b) kriging; (c) cokriging.

5.2.2. *Kriging.* Using the continuous function of the adjusted variogram, the tomato consumption prediction values are obtained based on distance and spatial correlation in the following way:

`krige (TOMATE~1, FJespacial, spgridtc, model = m)`, ~1 defines a single constant predictor.

Based on the ordinary kriging method that is considered the best unbiased linear estimator type, the values found in the interpolation vary especially in two foci on which the predictions are generated.

The values closest to the points of information are more influenced than those that are far away (Figure 5(b)).

5.2.3. *Spatial Prediction Based on Associated Products.* Because of the interrelation of products found with the *Apriori* algorithm, a set of associated products in the commercialization with the highest incidence in the process was identified. The five products resulting from association rules is $A = \{\text{Tomato, Broccoli, White Onion, Tamarillo, Carrot}\}$.

The correlation between the elements of set A was verified, and the model variogram with each element will be generated, as can be seen in Figure 6.

At this point, the linear model of coregionalization is adjusted to a variogram of multivariable samples using the products.

5.2.4. *Cokriging.* In the same way as made for the tomato variable, we proceed to estimate the future sales of the target variable (tomato), with an extended model integrating all the associated products, as shown in Figure 5(c). The variable g represents a function with all the products resulting from the added association rules for which the new variogram is calculated, `vmra <- variogram (g)`.

The adjusted variogram is obtained from the interaction between the variogram of the function g (multivariable) and the model m of a single variable, `vm.fit <- fit.lmc (vmra, g, and m)`.

The multivariate prediction is derived from the relation `xt <- predict (vm.fit and spgridtc)`. A summary of the three cases of prediction of future consumption is presented in Figure 5.

5.3. *Discussion.* To perform the assessment of the prediction model, the cross validation divides the data into two sets: the

FIGURE 6: Multivariate correlation and variogram

modeling subset is used by the model variogram to estimate the coefficients, and then, kriging is applied in the locations of the validation set, so that validation measures are compared with their predictions.

The procedure known as leave-one-out cross validation (LOOCV) was applied, and it performs as many iterations as data (N) has the set, using $N-1$ data to train the model and the data left for testing, being the result the arithmetic mean of the N error results obtained $E = 1/N \sum_{i=1}^n e_i$.

Cross validation usually gives a pessimistic estimate of performance (bias), since most models would improve if the training set would be bigger. For this reason, LOOCV has the lowest bias since the training set contains the whole dataset except one datum. On the other hand, some authors point out that the error estimated by LOOCV may have greater variance than k -fold cross validation, with $k \ll n$, since the size of datasets is higher and estimation smoother. However, this is open to discussion, as indicated [37], since k -fold cross-validation produces dependent test errors, and their correlations cannot be estimated unbiasedly. As indicated, in [38], in learning problems employing models with moderate/low instability (as linear regression problems), LOOCV often has lower variability both in bias and variance.

In any case, for situations with small datasets the variance in fitting the model tends to be higher, implying that k -fold cross-validation is likely to have a high variance (as well as a higher bias) with respect to LOOCV. This is why LOOCV is often the best choice with limited amounts of available data, as the case study in this work, in order to get the maximal use of data to compare the performance of alternative learning structures.

The estimation error (difference between the estimated value and the true value) is calculated in each site with data, and a statistical analysis is made of the errors committed in all data sites.

The results obtained from performing the cross validation for each method chosen for this study indicate that when comparing the residual values of the predictions, the IDW and kriging method have similar prediction values while the cokriging process (multivariate) presents an improvement for its smaller amplitude in the results (Figure 7). As can be seen in the cash flow diagrams, for all the estimation processes of future sales, the values are located in a range between -20 and 30 .

Figure 8 shows the result of subtracting the residual values between the prediction of (1) IDW/kriging method (left frame) and (2) IDW/cokriging (right frame).

FIGURE 7: Comparative cross validation.

FIGURE 8: Cross validation relations.

The first frame represents the set of values obtained from performing the cross validation of estimated values for the methods of future estimation residual between IDW and ordinary kriging (single variable), and the positive values are eight and the negative ones six.

In the second part, the residual difference of the cross validation between the IDW method and the ordinary cokriging (multivariable) is calculated, establishing nine positive and five negative values.

Positive samples indicate that the residual value of the multivariable function is smaller.

6. Conclusions

This research is focused on a target area for analysis located in Tungurahua and Chimborazo provinces of Ecuador, where there are fourteen alternative marketing circuits called fairs. These locations were used to create the grid of future sales estimates, and the analysis of sales transactions containing agricultural products generated the set of strongly associated products, based on the Apriori algorithm. As result, the set of associated products with support parameters = 0.4 and confidence greater than 0.8 are, in order, the following: white onion (0.87), tamarillo (0.86), carrot (0.83), and broccoli (0.82), and each of these products is associated with the sale of tomato.

Using the IDW process as baseline for comparison, the leave-one-out cross validation of predictions was done to compare with geostatistical techniques based on the variogram to generate interpolations of product sales in the target area. Based on the functions of kriging, the sales values of products were established according to their spatial locations and influences of close neighbors. Finally, a multivariable set of products for predictions was established resulting from association rules with greatest associativity (tomato, broccoli, tamarillo, white onion, and carrot). Based on this multivariable set, the prediction values are calculated using the same procedure described in the first and second stages. With this improvement in the sales prediction process, it would be possible to establish scenarios to generate consumption maps (Figure 9) that can be supplied with a better production process and its subsequent commercialization which is reflected in a better level of economic income for the farmer families.

Finally, the products resulting from applying the Apriori algorithm with the greatest associativity are tomato, broccoli, tamarillo, white onion, and carrot.

Based on this multivariable set, the prediction values are calculated using the same procedure described in the first and second stages.

The residual value of the IDW prediction minus kriging prediction delivers eight positive values and six negative

FIGURE 9: Results for scenarios.

values. In the same way, it is calculated for the IDW minus cokriging process, and in this case, nine positive and five negative values are obtained.

In the process of cokriging (multivariable), there are a greater number of cases with positive differences that shows that this process using the set of association rules as multivariable has a 16% improvement when establishing future sales estimate.

The proposed methodology is to find a set based on association rules to establish the multivariable process, and this research has shown acceptable improvement in prediction values. With this improvement in the sales prediction process, it is possible to establish scenarios to generate consumption maps (Figure 9) that can be supplied with a better production process and its subsequent commercialization which is reflected in a better level of economic income for the family farmer.

Finally, it should be emphasized that a methodology is established based on the use of association rules that allow future estimates to be improved using cokriging (multivariable) processes.

Taking into account that in order to use a conventional statistical process, there is not enough data available to establish a distribution and an estimate of future prediction, and it is considered that the proposed technique and methodology are useful for initial cases of study, have a limited amount of data, especially to create a baseline, as more annual data series are obtained which can be compared

between the values obtained using data mining techniques contrasted with values of traditional statistical techniques.

Data Availability

The data used are provided by the General Coordination of Marketing Networks of the Ministry of Agriculture and Livestock of Ecuador, within the framework of an interinstitutional agreement with the Salesian Polytechnic University.

Disclosure

This research is part of the Doctorate in Computer Science Program that is being studied by Washington R. Padilla.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by Project MINECO TEC2014-57022-C2-2-R, Salesian Polytechnic University of Quito-Ecuador and by Commercial Coordination Network, Ministry of Agriculture and Livestock Ecuador. This research has the partial economic support of the Salesian Polytechnic University of Ecuador.

References

- [1] M. Leporati, S. Salcedo, B. Jara, V. Boero, and M. Muñoz, *La Agricultura Familiar en Cifras, en Recomendaciones de Política*, FAO, Rome, Italy, 2014.
- [2] R. Contreras, E. Krivonos, and L. Sáez, *Mercados Locales y Ferias libres: El Caso de Chile, en Recomendaciones de Política*, FAO, Rome, Italy, 2014.
- [3] C. R. Pablo Díaz and M. Arosio, *Circuitos Cortos de Comercialización, Un Resumen Ejecutivo*, Centro Latinoamericano para el desarrollo Rural RIMISP, Centro Latinoamericano para el desarrollo Rural RIMISP, Metropolitana, Chile.
- [4] S. Salcedo, A. Paula, and L. Guzmán, *El Concepto de Agricultura Familiar en América Latina y el Caribe, en Recomendaciones de Política*, FAO, Rome, Italy, 2014.
- [5] S. Asadifar and M. Kahani, "Semantic association rule mining: a new approach for stock market prediction," in *Proceedings of 2nd Conference on Swarm Intelligence and Evolutionary Computation (CSIEC)*, pp. 106–111, Kerman, Iran, March 2017.
- [6] R. V. Mane and V. R. Ghorpade, "Predicting student admission decisions by association rule mining with pattern growth approach," in *Proceedings of International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT)*, pp. 202–207, Mysore, Karnataka, December 2016.
- [7] P. S. V. V. S. R. Kumar, L. R. D. P. Maddireddi, V. A. Lakshmi, and J. N. K. Dirisala, "Novel fuzzy classification approaches based on optimisation of association rules," in *Proceedings of 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, pp. 1–5, Bengaluru, India, April 2016.
- [8] K. S. Won and T. Ray, "Performance of kriging and cokriging based surrogate models within the unified framework for surrogate assisted optimization," in *Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No.04TH8753)*, vol. 2, pp. 1577–1585, Portland, OA, USA, June 2004.
- [9] D. Chen, D. Liu, Y. Li, L. Meng, and X. Yang, "Improve spatiotemporal kriging with magnitude and direction information in variogram construction," *Chinese Journal of Electronics*, vol. 25, no. 3, pp. 527–532, 2016.
- [10] W. R. Padilla and H. J. García, "CIALCO: alternative marketing channels," in *Proceedings of International Conference on Practical Applications of Agents and Multi-Agent Systems Highlights of Practical Applications of Scalable Multi-Agent Systems. The PAAMS Collection*, pp. 313–321, Seville, Spain, June 2016.
- [11] W. R. Padilla, J. García, and J. M. Molina, "Improving forecasting using information fusion in local agricultural markets," in *Proceedings International Conference on Hybrid Artificial Intelligence Systems Hybrid Artificial Intelligent Systems*, pp. 479–489, Oviedo, Spain, June 2018.
- [12] W. R. Padilla, J. García, and J. M. Molina, "Information fusion and machine learning in spatial prediction for local agricultural markets," in *Proceedings of International Conference on Practical Applications of Agents and Multi-Agent Systems Highlights of Practical Applications of Agents, Multi-Agent Systems, and Complexity: The PAAMS Collection*, pp. 235–246, Toledo, Spain, June 2018.
- [13] W. B. Zulfikar, A. Wahana, W. Uriawan, and N. Lukman, "Implementation of association rules with Apriori algorithm for increasing the quality of promotion," in *Proceedings of 4th International Conference on Cyber and IT Service Management*, pp. 1–5, Bandung, Indonesia, April 2016.
- [14] J. Hernandez-Orallo, *Introducción a la Minería de Datos*, ene, Madrid Pearson Prentice Hall, Upper Saddle River, NJ, USA, 2004.
- [15] S. D. Patil, R. R. Deshmukh, and D. K. Kirange, "Adaptive Apriori algorithm for frequent itemset mining," in *Proceedings of International Conference System Modeling Advancement in Research Trends (SMART)*, pp. 7–13, Moradabad, India, November 2016.
- [16] J. P. Celemin, "Autocorrelación espacial e indicadores locales de asociación espacial: importancia, estructura y aplicación," *La Revista Universitaria de Geografía*, vol. 18, no. 1, pp. 11–31, 2009.
- [17] A. Uversky, D. Ramljak, V. Radosavljević, K. Ristovski, and Z. Obradović, "Which links should I use? A variogram-based selection of relationship measures for prediction of node attributes in temporal multigraphs," in *Proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013)*, pp. 676–683, Niagara Falls, Canada, August 2013.
- [18] V. Pinel, J. P. Gastellu-Etchegorry, and V. Demarez, "Retrieval of forest biophysical parameters from remote sensing images with the DART model," in *Proceedings of International Geoscience and Remote Sensing Symposium, IGARSS'96, Remote Sensing for a Sustainable Future*, vol. 3, pp. 1660–1662, Lincoln, NE, USA, May 1996.
- [19] Y. Chen, X. Du, and L. Zhou, "Transformer defect correlation analysis based on Apriori algorithm," in *Proceedings of IEEE International Conference on High Voltage Engineering and Application (ICHVE)*, pp. 1–4, Chengdu, China, September 2016.
- [20] E. Tonye, J. Fotsing, B. E. Zobo, N. T. Tankam, T. F. N. Kanaa, and J. P. Rudant, "Contribution of variogram and feature vector of texture for the classification of big size SAR images," in *Proceedings Seventh International Conference on Signal Image Technology Internet-Based Systems*, pp. 382–389, Dijon, France, November 2011.
- [21] *How Do I Generate a Variogram for Spatial Data in Stata? | Stata FAQ*, IDRE Stats, <https://stats.idre.ucla.edu/stata/faq/how-do-i-generate-a-variogram-for-spatial-data-in-stata/>.
- [22] C. Ma, "Linear combinations of space-time covariance functions and variograms," *IEEE Transactions on Signal Processing*, vol. 53, no. 3, pp. 857–864, 2005.
- [23] R. S. Bivand, E. Pebesma, and V. Gómez-Rubio, *Applied Spatial Data Analysis with R*, Springer, New York, NY, USA, 2013.
- [24] H. Liu, B. Yang, and E. Kang, "Cokriging method for spatio-temporal assimilation of multi-scale satellite data," in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 3314–3316, Milan, Italy, July 2015.
- [25] N. W. Park and H. Y. Yoo, "The effects of spatial prediction of grain size fractions on intertidal surface sediments classification," in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium-IGARSS*, pp. 3876–3878, Melbourne, Australia, July 2013.
- [26] W. R. Padilla, G. H. Jesus, and J. M. Molina, "Model learning and spatial data fusion for predicting sales in local agricultural markets," in *Proceedings of 21st International Conference on Information Fusion (FUSION)*, pp. 1–5, Cambridge, UK, July 2018.
- [27] RStudio—open source and enterprise-ready professional software for R," December 2017, <https://www.rstudio.com/>.
- [28] R for Spatial Scientists," July 2017, http://gsp.humboldt.edu/OLM/R/01_03_GettingStarted.html.

- [29] beckmw, Breaking the rules with spatial correlation R is my friend, <https://beckmw.wordpress.com/2013/01/07/breaking-the-rules-with-spatial-correlation/>.
- [30] Fields: on-line manual," July 2017, <http://www.image.ucar.edu/GSP/Software/Fields/fields.manual.printable.shtml#Intro>.
- [31] Weka 3—data mining with open source machine learning software in Java," September 2017, <http://www.cs.waikato.ac.nz/ml/weka/>.
- [32] M. H. Merrikhpour and M. Rahimzadegan, "Improving the algorithm of extracting regional total precipitable water vapor over land from MODIS images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 10, pp. 5889–5898, 2017.
- [33] A. Boucher, K. C. Seto, and A. G. Journel, "A novel method for mapping land cover changes: incorporating time and space with geostatistics," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 11, pp. 3427–3435, 2006.
- [34] E. Sertel, S. Kaya, and P. J. Curran, "The use of geostatistical methods to identify severe earthquake damage in an urban area," in *Proceedings of 2007 Urban Remote Sensing Joint Event*, pp. 1–5, Xuzhou, China, April 2007.
- [35] H. Zhang, C. Wang, Y. Shanzhen, and y Q. Jiang, "Geostatistical analysis of spatial and temporal variations of groundwater depth in Shule River," *WASE International Conference on Information Engineering*, vol. 2, pp. 453–457, 2009.
- [36] H. Zhang, C. Wang, Y. Shanzhen, and Q. Jiang, "Geostatistical analysis of spatial and temporal variations of groundwater depth in Shule River," in *Proceedings of WASE International Conference on Information Engineering*, vol. 2, pp. 453–457, Taiyuan, China, July 2009.
- [37] Y. Bengio and Y. Grandvalet, "No unbiased estimator of the variance of K-fold cross-validation," *Journal of Machine Learning Research*, vol. 5, pp. 1089–1105, 2004.
- [38] Y. Zhang and Y. Yang, "Cross-validation for selecting a model selection procedure," *Journal of Econometrics*, vol. 187, no. 1, pp. 95–112, 2015.