Spatial Variation Prediction and Mapping of Soil Temperature

Damian C Martinez Martinez
Pontificia Universidad Javeriana
Bogotá, Colombia
damian.martinez@javeriana.edu.co

Margarita S. Narducci
Pontificia Universidad Javeriana
Bogotá, Colombia
m.narducci@javeriana.edu.co

***Abstract*—Spatial variation of soil temperature is a measurement of great help to analysis in precision agriculture. These measurements are used to create yield prediction models, diagnosis of diseases and reveal high or low concentrations of temperature. The present paper carries out a study of different methods to predict the variation of temperature in soil using geostatistical tools based on georeferenced data in a crop. Three interpolation techniques were tested to map soil properties and thus compare the accuracy of the prediction. Given the data generated, heat maps were created referring to the variables calculated by the different methods. The methods were compared using a performance criterion that included root mean square error (RMSE) and mean absolute error (ME). This comparison showed that the kriging method was the most accurate for the interpolated soil temperature, the RMSE of the method was the lowest among the others, and ME was very close to 0, suggesting that the predictions are unbiased.**

I. Introduction

Currently we are facing a food crisis due to the rapid and constant increase in the population in many areas of the world, as a consequence, mechanisms must be sought to significantly increase production to achieve food security that allows access, timely and permanent consumption of food in quantity, quality by all people [1].

Soil temperature is one of the parameters that directly influence the growth of crops, the germination of many types of seed depends on soil temperature layers [2], absorption of nutrients is also affected especially Phosphorus in cold soils.

In this sense, monitoring soil temperature in a crop can help farmers to make accurate decisions that could allow them to increase productivity and reduce losses. As a possible solution to this problem a set of technologies and estimation techniques are combined. A network of sensors measures temperature at various points of the corp then this data is processed by the application of methods of interpolation to determine the value and variation of temperature at the points not measured according to the collected data. Methods such as IDW, Natural Neighbor and Kriging are compared in this paper.

II. Methods of Interpolation

A. Inverse Distance Weighted (IDW) Technique

Inverse Distance Weighted (IDW) interpolation is a deterministic method which estimates unobserved location variables in a particular area based on the sampled values of the observed locations [3]. In this method the distance is a weighting parameter and thus, the unsampled points are determined by a linear combination of values at known sample points. The technique is based on the inverse weight of distance between a location to be estimated and all the observed points [4]. The surface to be interpolated must be of a location-dependent variable. This method assumes that the variable being assigned decreases influence with distance from its sampled location, therefore, nearby observations will have a heavier weight. The distance between the sampled point and the interpolation point is calculated with the Euclidian distance formula.

Fig. 1 shows a yellow circle which is called the search radius and inside are the points that will be used to make the interpolation.

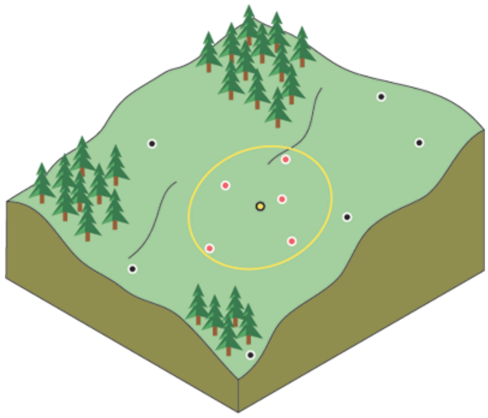


Fig. 1. IDW to a selected point Source: Adapted from [5]

Equation 1 relates the calculated value to the measured data, where $Z(x_0)$ is the interpolated value, n represents the total number of samples, x_i is the data value, h_{ij} is the data distance between the interpolated value and the sample value and y β determines the power weight [6].

$$Z(x_0) = \frac{\sum_{i=1}^{n} \frac{x_i}{h_{ij}^{\beta}}}{\sum_{i=1}^{n} \frac{1}{h_{ij}^{\beta}}} \quad (1)$$

El β value allows to control the importance of the known points, as β is increased the interpolated values begin to approach the value of the closest sample point.

B. Natural Neighbor Technique

This algorithm searches for the closest subset of input samples to a query point and applies weights based on provided areas to interpolate a value for which a grid is defined

Equation 2 represent the natural neighbor interpolation.

$$G(x,y) = \sum_{i=1}^{n} w_i f(x_i, y_i) \quad (2)$$

Where G(x,y) is interpolated value, n the number of natural neighbors, $f(x_i, y_i)$ value of the sample and w_i weight associated with the sample. The number of natural

978-1-6654-3331-0/20/$31.00 ©2020 IEEE

neighbors is determined by constructing neighboring natural circles. Two points are natural neighbors if they are within the same circle.

This technique will not produce peaks, pits, ridges, or valleys that are not yet represented by the input samples. The surface passes through the input samples and is smoothed everywhere except at the input sample locations

C. *Kriging Technique*

The kriging method is an advanced geostatistical procedure that generates an estimated surface from a dispersed set of points with values, incorporates statistical properties of the measured data (spatial autocorrelation) that are used to explain the variation in the surfaces.

The kriging approach uses the semivariogram to express spatial continuity (autocorrelation). The semivariogram measures the strength of the statistical correlation as a function of distance. The range is the distance at which the spatial correlation disappears

Fig. 2 shows the relationship of the red dot to all other measured locations. This process continues for each measured point.

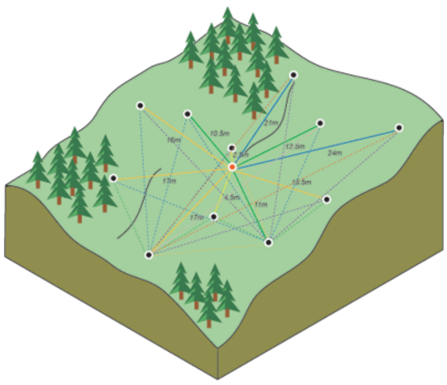


Fig. 2. Kriging Method: Relationship between points. Source: Adapted from [7]

Kriging is like IDW in that it weights the surrounding measured values to derive a prediction for an unmeasured location [8]. Equation (3) represents the weighted sum of the data performed for this model:

$$Z(s_0) = \sum_{i=1}^{N} \lambda_i \, Z(s_i) \qquad (3)$$

Where $Z(s_0)$ is the measured value at the location i, λ_i an unknown weight for the measured value at the i, s_0 location prediction and N the number of measured values, for technique of kriging the λ_i weight depends on a fitted model to the measured points, the distance to the prediction location, and the spatial relationships among the measured values around the prediction location [7].

D. *Cross – validation*

The cross-validation technique was used to evaluate and compare the performance of the different interpolation methods. The sample points were divided into two data sets, one for training and the other for validation. To reduce variability, the training and validation sets should be crossed in successive rounds, so that each data point can be validated. Mean Error (ME) mean relative error (MRE) and mean square error (RMSE) were calculated to assess the accuracy of the interpolation methods.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(O_i - S_1)^2}{N}}$$

$$ME = \frac{\sum_{i=1}^{0} O_i - S_1}{N}$$

$$MRE = \frac{1}{N}\sum_{i=1}^{N} \frac{(O_i - S_1)}{O_i}$$

Where O_i is observed value, S_1 the calculated value, N number of samples.

III. SIMULATION

A. *Field Site Details*

Interpolation techniques were applied on a data set collected by a network of sensors found at the 37-hectare Cook Agronomy Farm (CAF) in the northwestern United States operated by Washington State University, located near Pullman. , WA (46-47'N, 117-5'W) and whose data is public on the website[9]. Soil temperature measurements were made at five depths (0.3 m, 0.6 m, 0.9m, 1.2m and 1.5m), recording every hour. Fig. 3 shows the geographic distribution of the 47 sensor groups on the farm.

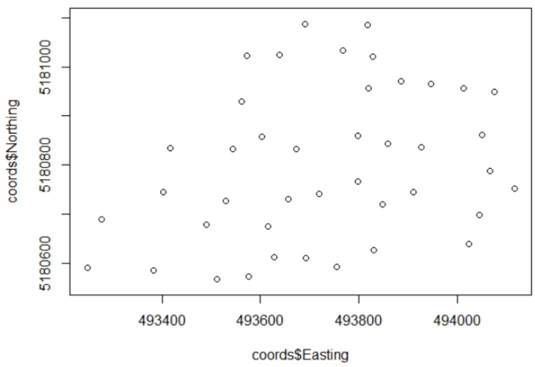


Fig. 3. Sensor locations on the farm

The farm is divided into three experimental subfields, each divided into 30 m wide segments, which serve as the basis for crop rotation. In general, wheat (Triticum aestivum L.) are planted two out of every 3 years, with the remaining year planted with an alternative crop (canola (Brassica napus annua Koch), chickpeas (Cicer arietinum L.), pea (Pisum arvense L.), barley (Hordeum vulgare L.), or triticale(xTriticosecale Wittm.)). CAF is center to multiple agricultural research projects carried out by interdisciplinary teams from multiple institutions. Research focuses include climate change, sustainable agriculture, and precision

978-1-6654-3331-0/20/$31.00 ©2020 IEEE

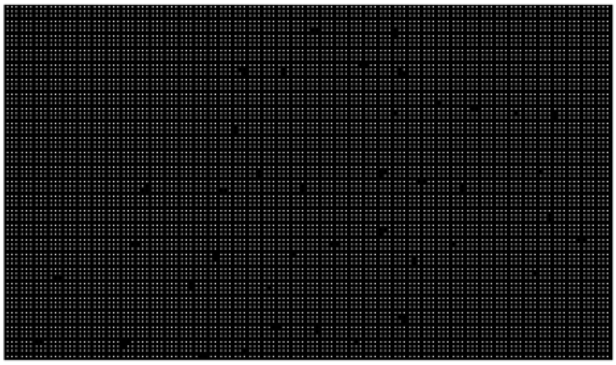
agriculture, with data collected on soil biogeochemical processes [6].

B. *Results of interpolation.*

Using the R program, the set of data related to temperature measurement was grouped, for this analysis the average values of all years were taken for the temperature records at a depth of 30 cm and thus validate the behavior of each method.

To apply the IDW method, a grid was built which represents the size of each area to be estimated, the greater the number of grids, the better the image resolution is obtained, but a higher computational cost since it is required to calculate more points.

With a number of grids of 50,000 over the total area that frames the farm, a representation is obtained as shown in Fig. 4, each black point represents the location of the sensors.



*F*ig. 4. Grid with n=50000

To apply the IDW method, the gstat package in R software was used applying the idw function with a power of two, this data is saved in a data frame to later apply a mask with the outline of the farm and each point is painted with a different color according to the value obtained from white being the lowest value, going through red and yellow and green colors for the highest (Fig. *6*5).

The natural neighbor method was used in the same package as the previous gstat method, applying the interpolate function, applying the 50,000 grid, the value was calculated for each of the fields, obtaining the result that is presented in Fig. 6.

For the Kriging technique, semivariogram modeling has been applied to show the spatial relationship. The experimental semivariogram is calculated by averaging the squared difference of the values in all pairs of observations with the distance and direction of separation (Fig. *6*7).

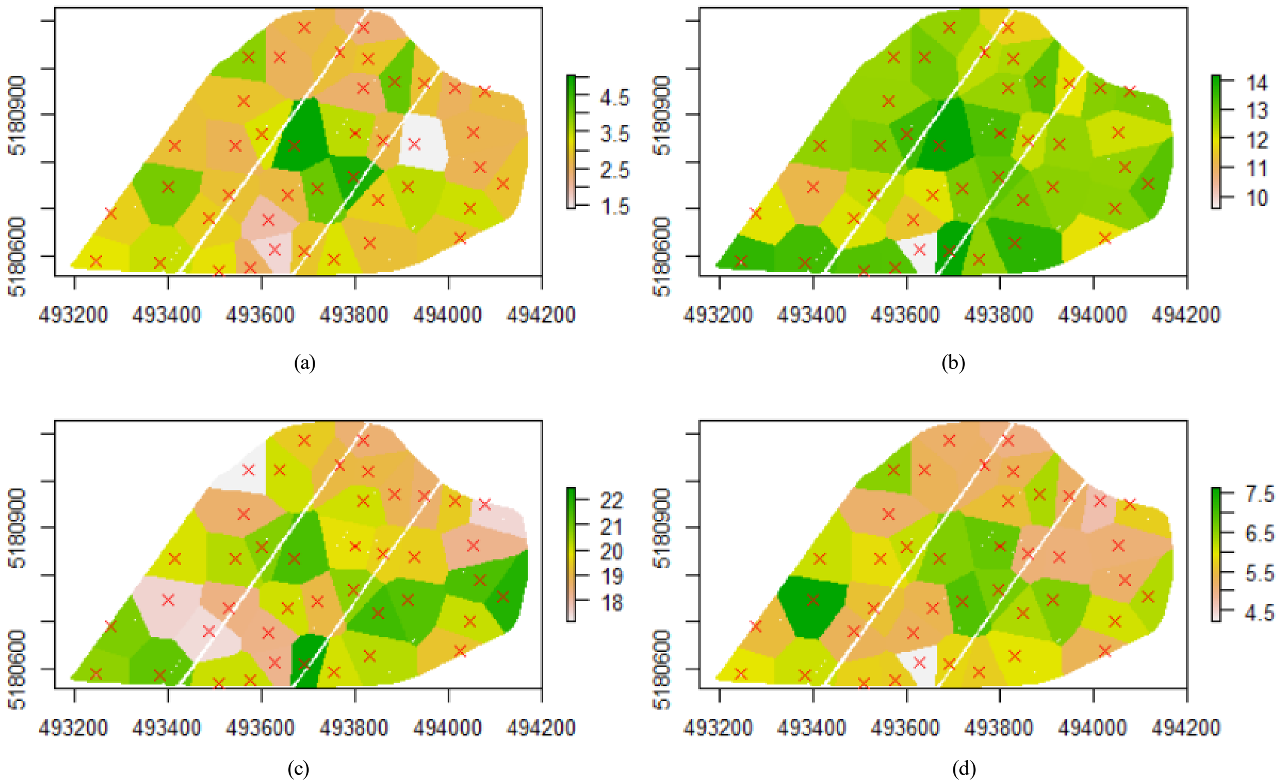


Fig. 5 Result applying IDW: a) Average January - March b) Average April - June c) Average July - September d) Average October - December

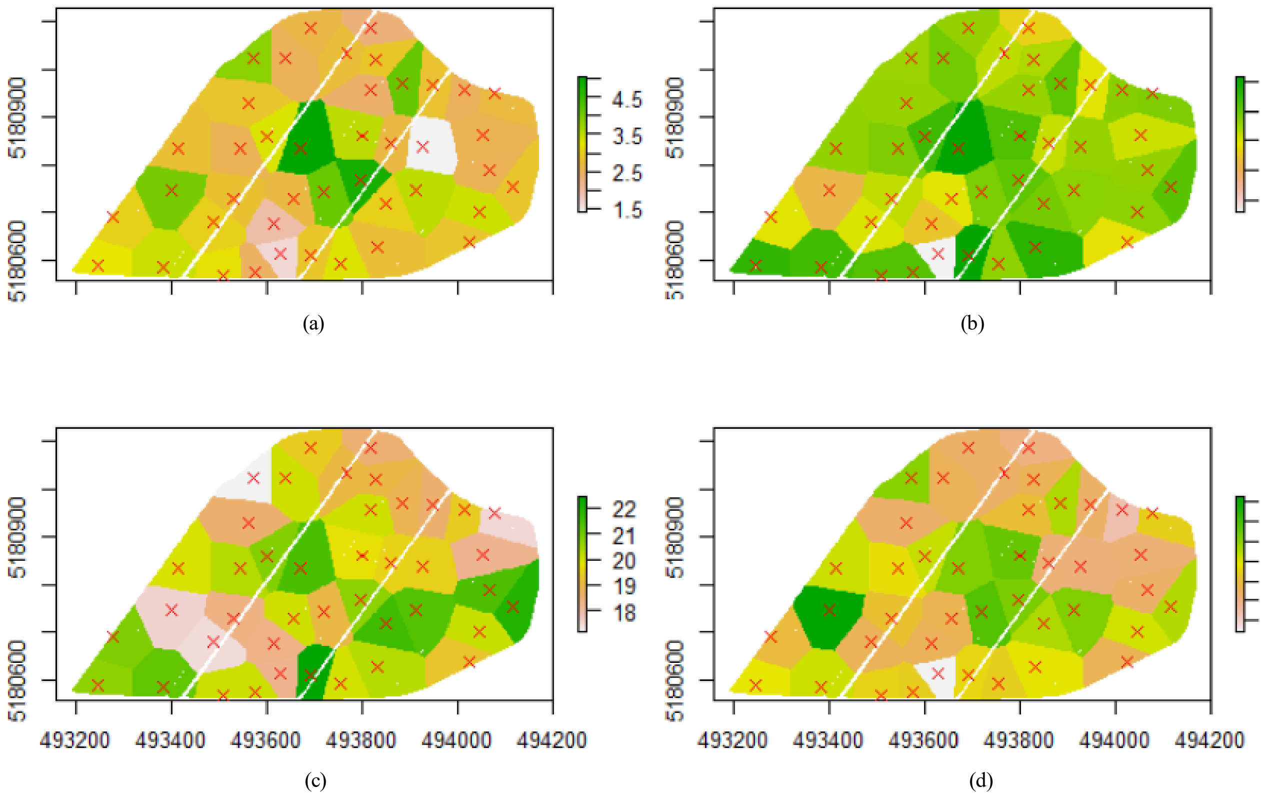


Fig. 6. Result applying Natural Neighbor: a) Average January - March b) Average April - June c) Average July - September d) Average October - December

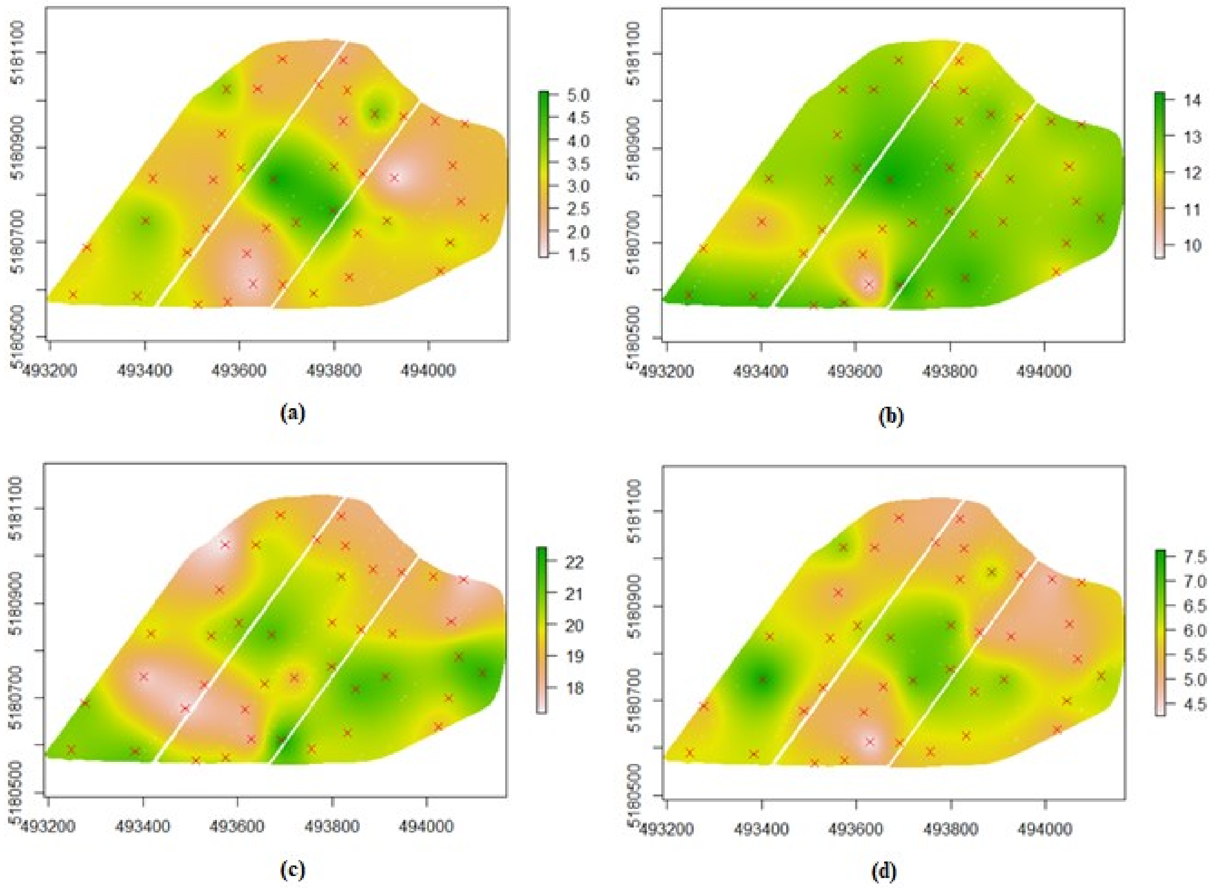


Fig. 7. Result applying Kriging: a) Average January - March b) Average April - June c) Average July - September d) Average October - December

The Fig. 8 illustrated the model that represents the Semivariance used to interpolate soil temperature with the kriging method obtained from for the measurements taken in the first three months of the year.

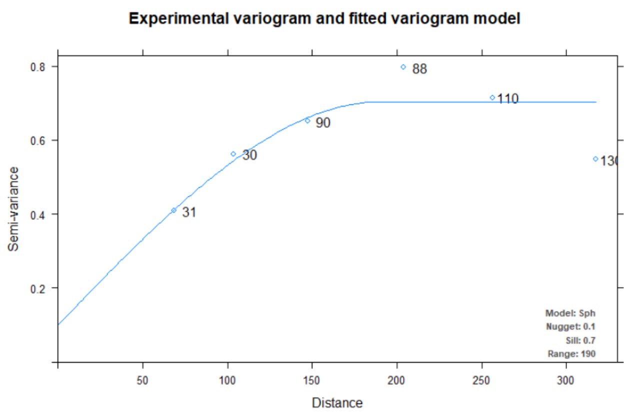


Fig. 8. Semivariance generated from the average data for the month of January – March

IV. EVALUATION

For evaluation, we use the cross-validation procedure to compare the performance of the different interpolation's method based on the data set. For all points, cross-validation sequentially skips a point, predicts its value using the rest of the data, and then compares the measured and predicted values.

For the model to provide accurate predictions, the standardized mean error must be close to zero, and the root mean-square error (RMSE) and the mean standard error must be as small as possible. The error in the standardized error of the root mean should be close to 1.

TABLE 1. COMPARISON OF RESULTS FOR JANUARY - MARCH

Month	Models	RMSE	MRE	ME
January - March	IDW	0.7003	0.18357	-0.01442
	Natural Neighbor	0.8396	0.2287	0.0110
	Kriging	0.6749	0.1745	0.00618

Table 2. COMPARISON OF RESULTS FOR APRIL - JUNE

Month	Models	RMSE	MRE	ME
April - June	IDW	0.8877	0.04128	-0.01762
	Natural Neighbor	0.8931	0.04193	-0.02591
	Kriging	0.8290	0.04017	-0.01082

Table 3. COMPARISON OF RESULTS FOR JULY - SEPTEMBER

Month	Models	RMSE	MRE	ME
July - September	IDW	1.2433	0.0509	0.00353
	Natural Neighbor	1.2681	0.0532	0.02823
	Kriging	1.2107	0.0522	-0.01130

Table 4 COMPARISON OF RESULTS FOR OCTOBER - DECEMBER

Month	Models	RMSE	MRE	ME
October - December	IDW	0.6935	0.1000	-0.0251
	Natural Neighbor	0.6778	0.0949	-0.0008
	Kriging	0.6587	0.0929	-0.0170

V. CONCLUSIONS

According to the collected data, we can observe how the temperature varies depending on the season of the year, obtaining the highest temperature between the months of July to September and the lowest between January and March. Evaluating each method according to the value of RMSE and ME, a lower value is obtained for the Kriging inteporlation between the measured values and the calculated values, likewise for this method there is a ME value close to zero suggesting that the predictions are unbiased. As a future work we are proposing to validate the interpolation method with own temperature measurements in a rice field in Colombia.

ACKNOWLEDGEMENT

This work was funded by the OMICAS program: Optimización Multiescala In-silico de Cultivos Agrícolas Sostenibles (Infraestructura y validación en Arroz y Caña de Azúcar), anchored at the Pontificia Universidad Javeriana in Cali and funded within the Colombian Scientific Ecosystem by The World Bank, the Colombian Ministry of Science, Technology and Innovation, the Colombian Ministry of Education, the Colombian Ministry of Industry and Tourism, and ICETEX, under grant FP44842-217-2018 and OMICAS Award ID: 792-61187.

REFERENCES

[1] Ministerio de ambiente, "Seguridad Alimentaria y Nutricional." [Online]. Available: https://www.minambiente.gov.co/index.php/asuntos-ambientales-sectorial-y-urbana/sostenibilidad-sectores-productivos/seguridad-alimentaria-y-nutricional.

[2] M. Zeynoddin, I. Ebtehaj, and H. Bonakdari, "Development of a linear based stochastic model for daily soil temperature prediction: One step forward to sustainable agriculture," *Comput. Electron. Agric.*, vol. 176, p. 105636, Sep. 2020.

[3] A. F. O. Gaffar, R. Malani, A. B. W. Putra, Mulyanto, and Ibayasid, "Prediction of the Topographic Shape of the Ground Surface Using IDW Method through

the Rectangular-Neighborhood Approach," in *Proceeding - 2018 International Symposium on Advanced Intelligent Informatics: Revolutionize Intelligent Informatics Spectrum for Humanity, SAIN 2018*, 2019, pp. 107–111.

[4] E. Oktavia, Widyawan, and I. W. Mustika, "Inverse distance weighting and kriging spatial interpolation for data center thermal monitoring," in *Proceedings - 2016 1st International Conference on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2016*, 2016, pp. 69–74.

[5] "How IDW works—Help | ArcGIS for Desktop." [Online]. Available: https://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/how-idw-works.htm. [Accessed: 18-Nov-2020].

[6] G. S. Bhunia, P. K. Shit, and R. Maiti, "Comparison of GIS-based interpolation methods for spatial distribution of soil organic carbon (SOC)," *J. Saudi Soc. Agric. Sci.*, vol. 17, no. 2, pp. 114–126, Apr. 2018.

[7] ArcGis, "How Kriging works—Help | ArcGIS for Desktop." 2014.

[8] L. Belkhiri, A. Tiri, and L. Mouni, "Spatial distribution of the groundwater quality using kriging and Co-kriging interpolations," *Groundw. Sustain. Dev.*, vol. 11, p. 100473, Oct. 2020.

[9] "Data from: A field-scale sensor network data set for monitoring and modeling the spatial and temporal variation of soil moisture in a dryland agricultural field | Ag Data Commons." [Online]. Available: https://data.nal.usda.gov/dataset/data-field-scale-sensor-network-data-set-monitoring-and-modeling-spatial-and-temporal-variation-soil-moisture-dryland-agricultural-field. [Accessed: 08-Oct-2020].

978-1-6654-3331-0/20/$31.00 ©2020 IEEE