Comparison of Inverse Distance Weighted and Natural Neighbor Interpolation Method at Air Temperature Data in Malang Region

Jaka Pratama Musashi[1], Henny Pramoedyo, Rahma Fitriani[2]
[1,2]Department of Statistics, Brawijaya University, Malang

Email: jakapratama193@gmail.com, hennyp@ub.ac.id, rahmafitriani@ub.ac.id

ABSTRACT

The purpose of this study was to compare the results of Inverse Distance Weighted (IDW) and Natural Neighbor interpolation methods for spatial data of air temperature in the Malang Region. Interpolation is one way to determine a point of events from several points around the known value. Spatial interpolation can be used to estimate an area that does not have a data record using the value of its known surroundings. 38 points observation air temperature of Malang Region in 2016 is used as a sample point to interpolate the surrounding air temperature. Obtained optimum parameter power value is 2 for IDW interpolation method. The RMSE comparison results show that IDW method is better to be used than the Natural Neighbor Interpolation method with the RMSE values of 1,2292 for the IDW method and 1,6173 for the NN method.

Keywords: inverse distance weighted, natural neighbor, interpolation, air temperature, RMSE

INTRODUCTION

Geostatistical has been developed and able to explain the diversity that occurs due to natural phenomena and human-made on earth. Geostatistics can be used to predict data at unobserved locations. To expect values in unsampled locations, a statistical method called interpolation can be used. The process of estimating data at a location that can not be observed requires a model, but at different cases do not have to get a specific model. Geostatistical have very important role in it, using the estimation method based on the model. One of the geostatistical techniques for estimation is the Interpolation Method.

Interpolation is a method for obtaining data based on known data. In cartography, interpolation is the process of estimating values in areas that are not sampled or measured, to form a map or distribution of values across the entire region. One of the interpolation methods that can be used is Inverse Distance Weighted (IDW). IDW method has a local influence on the distances that give greater weight to the nearest point compared to the farther point [1]. IDW has the advantage which that the interpolation characteristics can be controlled by limiting the input points in the interpolation process. Any input that have small spatial correlations or even no spatial correlation can be removed from the calculation method.

IDW Interpolation has weakness that it can not predict the value above the maximum value and below the minimum value of the sample points or can be called the bullseye effect [2]. Another interpolation method with a weighting method similar to IDW is Natural Neighbor Interpolation. The Natural Neighbor Interpolation is known as Sibson Interpolation or "Area-Stealing" where this method works by looking for points adjacent to the sample point and apply the weights at those points. The basic of this interpolation method is uses only the sample that is around the point to be interpolated and the results obtained are similar to the sample point as the input value of the interpolation process [3].

The logic in the spatial interpolation mentioned in Tobler's Geography Law is that the value of the adjacent observation point will have the same value (close to) as compared to the value at the farther point. Spatial interpolation assumes that attributes are continuous in space and spatial dependencies. Both assumptions indicate that the data attribute estimate can be done

based on the locations at the surrounding points [4]. According to Watson [5], the IDW interpolation technique uses the nearest distance weight, while the Natural Neighbor interpolation technique uses the average weight of the surrounding area. Both of these techniques are local interpolation techniques which of course can be compared.

Global warming is an increasing average temperature in the atmosphere, sea, and land including one of the earth's big damage. Global warming is indicated by rising earth surface temperatures and the process which is called the greenhouse effect. The greenhouse effect is formed because of troposphere gas that exceeds the natural state. Causes of global warming include population growth, industry, refrigeration, transportation, forest burning, land use diversion and greenhouse gas in the atmosphere.

Transportation is a very valuable part of society. The other side of the excessive use of transportation causing a negative effect. Air temperature increase is most prevalent in vehicle dense points or congestion-prone. The number of vehicles also contributed in rising air temperature. The vehicle's exhaust gases cause an increase in the greenhouse effect so that solar radiation to the earth can not bounce back into the atmosphere. The number of transportation continues to increase especially in big cities in Indonesia, one of the big cities in Indonesia is the city of Malang.

The city of Malang is known as a city of tourism and education. The development of Malang city can be seen on the increase of transportation and population which is accompanied by government facility and investor. The area of Malang Raya is located between several mountains so that it has a very wide geographical condition and is widely spread, thus allowing for an unsampled location for air temperature data.

This is an interesting phenomenon, but there has been no research done to interpolate the air temperature data in this city, especially regarding the study of IDW and Natural Neighbor Interpolation Methods. This study uses two interpolation methods by considering the ability of the method to estimate simple but accurate to interpolate air temperature in Malang. Interpolation results are presented in a prediction map to determine the spread of air temperature in Malang. This research is expected to be valuable and useful for various parties, especially for the agricultural sector to provide an overview of the air temperature in the future considering the increase in air temperature will affect the development of plants.

METHODS

Spatial interpolation steps consist of 6 stages: First, we preparing maps with spatial data attributes, then testing spatial autocorrelation, and we do spatial interpolation of IDW and Natural Neighbor. After that we can draw two contour maps from the results of the two interpolation methods. For validation we do cross-validation, and then testing the validation of results research.

First we must fulfilled the assumtion of Spatial Autocorrelation Test. Autocorrelation can be defined as a series of inter elements in the series and elements of the same circuit and separated by an interval. One of the spatial autocorrelation tests is using the Moran I test. Gumprecht [6] describes Moran I statistics as the most widely known and used method for the measurement and testing of spatial dominance. The test measures the intensity of spatial autocorrelation. The autocorrelation coefficient of Moran is denoted by I as follows:

$$\hat{I} = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\left(\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}\right) \sum_{i=1}^{n}(x_i - \bar{x})}$$

where x_i and x_j is a observation value at location i and j, $\bar{x}$ is average of the observed value, w_{ij} is the weight value between location i and j, and n is the sample size. Moran I autocorrelation test has a value between -1 to 1. The value -1 shows a strong negative autocorrelation, while the value of 1 shows a strong positive correlation. The expected value of I under H_0 is that autocorrelation is not expected to be 0 but given by I_0 [7].

In the autocorrelation test should be considered also spatial pattern. In general, spatial patterns can be described as 3 types of clustered, dispersed (chessboard) and random. Spatial

autocorrelation is positive if in an adjacent area has a similar value and if it is poured in the image it forms a hoop pattern. Spatial autocorrelation is negative if in an adjacent area has a much different value and if it is depicted to form a chessboard pattern. Between the two mentioned patterns there is a random pattern showing no spatial autocorrelation.

Spatial interpolation is one way of knowing the value of a point location based on the value of some point around its known value. For example, to make a precipitation map of a country, an available weather stations will never be evenly distributed to provide data on the territory of the country. Spatial interpolation can estimate the value of an area that does not have a data record using the value of its known surroundings [8].

After the Spatial Autocorrelation assumption test has been fulfilled, we can continue to the Interpolation method, which is IDW and Natural Neighbor method. The Inverse Distance Wighted (IDW) method assumes that each input point has a local effect on distance. This method gives higher weight to the closest cell to the data point compared to the further cells. Points on a given radius are used in determining the output value for each location. The general weighting function of IDW is inverse of the squared distance formulated in the following equation:

$$Z^* = \sum_{i=1}^{n} \omega_i Z_i$$

where Z_i represents the value of the data to be interpolated by a number of n points and weights (ω_i) formulated as

$$\omega_i = \frac{h_i^{-p}}{\sum_{j=0}^{n} h_j^{-p}}$$

p is a positive value that can be changed is called the power parameter and h_j is the distance to the point of interpolation which is described as:

$$h_i = \sqrt{(x - x_i)^2 + (y - y_i)^2}$$

(x, y) is the coordinate of the interpolation point and (x_i, y_i) is the coordinate of the default data point.

The Natural Neighbor Interpolation Method was first introduced by Sibson [3] as a weighted average weighted interpolation method such as IDW. The Natural Neighbor interpolation does not use the distance as a weight but instead forms the Delauney Triangulation of the input points and selects the nearest cell that forms the convex hull around the interpolation point and uses the proportional area as weight. This method is used when the sample data points are distributed with uneven density. To interpolate Natural Neighbor, the first process is to build a polygon for all input points in interpolation, then the new voronoi diagram will be created around the interpolation points.
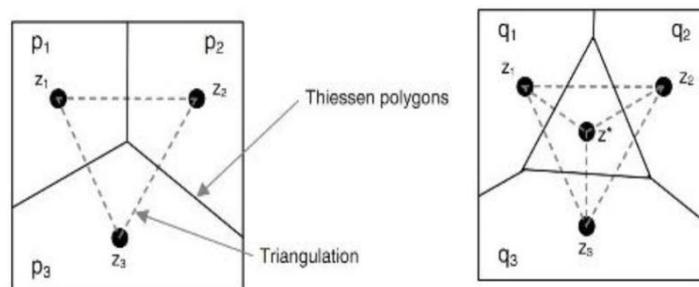

Figure 1. Voronoi Diagram (Thiessen Polygon)

If the point that wanted to be interpolated z is inserted into the data set, the area around the voronoi diagram will decrease. If p_i and q_i are the voronoi diagram areas of the sample points z_i before and after the addition of z^*, the weight for the sample point z_i [9] :

$$\omega_i = \frac{(p_i - q_i)}{pi}$$

According to Armstrong [10], Cross-Validation is a way to know the accuracy of some interpolation methods. Cross validation test using Root Mean Square Error (RMSE) is described as follows :

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}$$

RESULTS AND DISCUSSION

This study uses secondary data of the average monthly air temperature of Malang Raya Area which includes Malang City, Malang Regency, and Batu Town obtained from UPT Water Resources Management of Malang City. Monthly air temperature data for 1 year in °C/year in 2016 from 38 points of observation which located spread in Malang Region.

The autocorrelation assumption test results using Moran's I is shown in Figure 2.
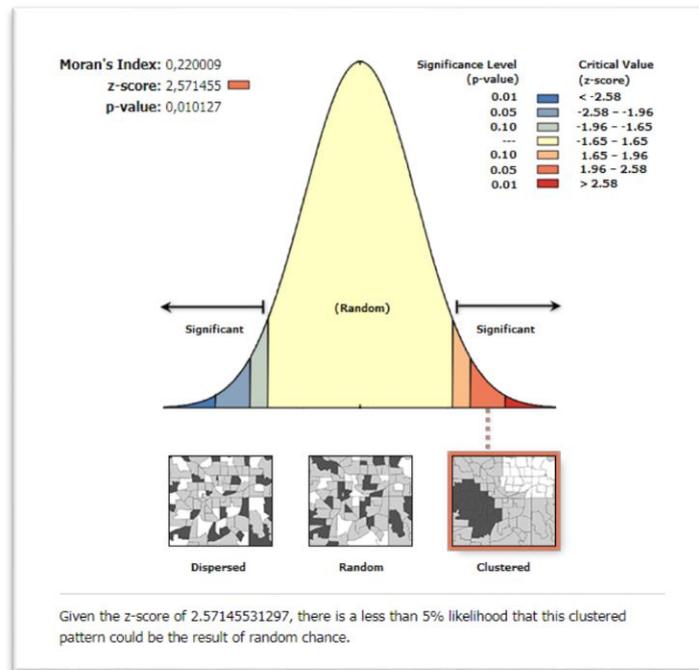

Figure 2. Moran I Test Result from Air Temperature 2016 in Malang Region

Figure 2 shows that with 95% confidence level can be said that formed spatial pattern clustered, which means that the temperature data of Malang Region in 2016 has positive spatial autocorrelation. In addition to using spatial distribution pattern, this assumption test can also be seen from Moran I test statistic. The result of Moran I test resulted Z test statistic (I) of 2.57 with Moran I index of 0.22. Based on test criteria with 95% confidence level, because Z (I)> Z ($\alpha/2$) then H_0 is rejected and concluded that air temperature data of Malang region in 2016 has positive spatial autocorrelation.

Interpolation IDW uses two methods of determining area and a power parameters. Malang Raya area is located between several mountains so it has a geographical condition that is very varied and widely spread. Based on this situation, we use Variable search radius method on IDW Interpolation. In addition to the area determination, also used some weighting for power parameters. In this study we use the power values 1,2,3,4 and 5. Determination of value must be positive and some power values are used only by selecting multiple values to see the resulting difference.

The power parameters to be used are the most optimum, ie by looking at the RMSE value for each power. The RMSE value for each power parameters can be seen in Table 1.

Table 1. RMSE Value for each power

POWER	RMSE
1	1.257122
2	**1.229251**
3	1.261345
4	1.318822
5	1.376546

Table 1 shows that a large RMSE value (1.376546) is obtained from a larger power value (power value 5), because the observation points that affected by this interpolation point are less. Otherwise, if the value of small power (power value 2), resulting RMSE value of 1.229251. From the RMSE value for air temperature in Malang Region, it can be said that it would be better to use IDW Interpolation method with power value 2.

After the most optimum power parameters are obtained, the next step is to interpolate IDW. Interpolation results are used to create a prediction map at all points in Malang Region which presented with raster data format in Figure 3.
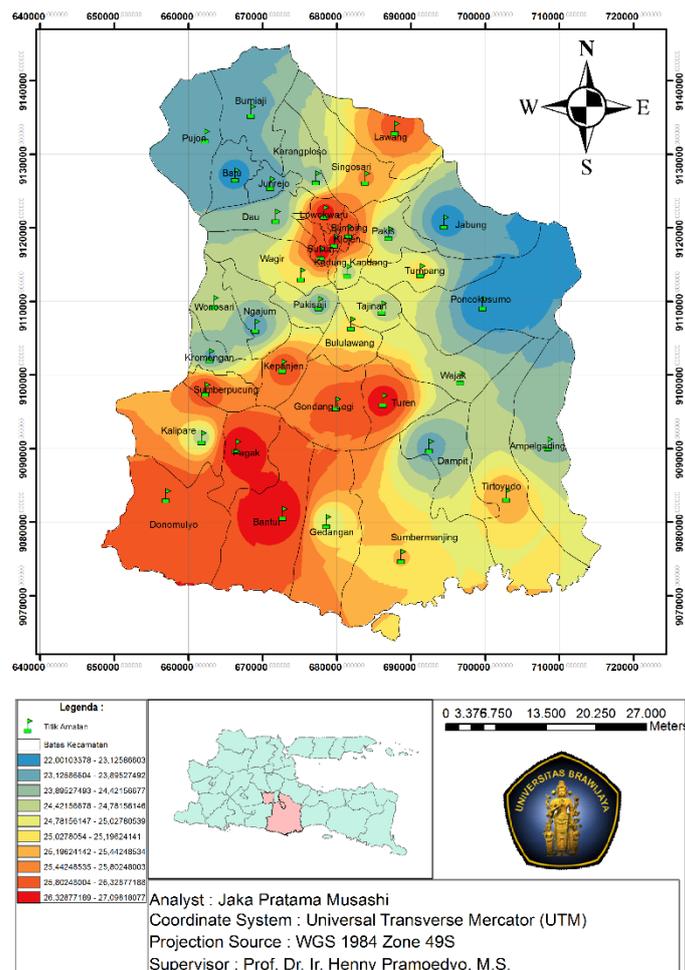

Figure 3. IDW Interpolation Result

Figure 3 is a prediction map of IDW interpolation result using power value 2 using air temperature data of Malang Area. Based on Figure 3 it is observed that the highest interpolation

value of air temperature is in the Malang city and Bantur Subdistrict, while the lowest interpolation air temperature is in the mountains such as Batu City and Poncokusumo Subdistrict.

The Natural Neighbor Interpolation Method is local, where only the samples are located around the points to be interpolated, so that the air temperature data to be interpolated will be similar to the value of the air temperature at the sample point. This interpolation produces a smooth surface which is shown in Fig. 4.
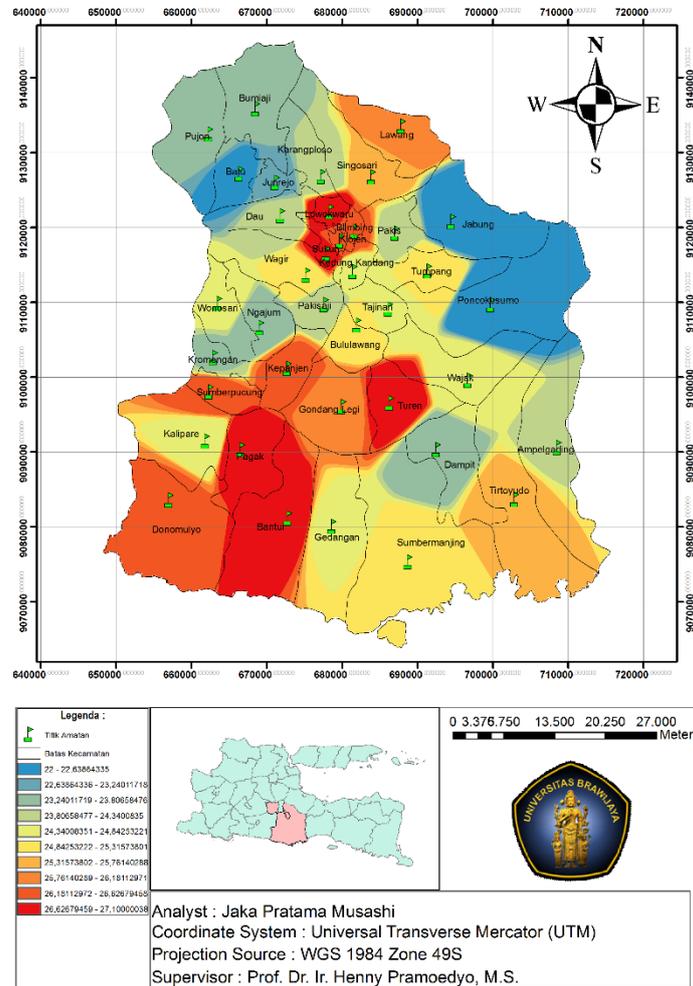

Figure 4. Natural Neighbor Interpolation Result

Using this interpolation method resulting more smoothed topography surface compared to the IDW interpolation method, caused by the theory contained in this method where the Thiessen Polygon system is used to calculate the point interpolation value based on the area that affecting the interpolation point.

Cross Validation test result is RMSE value. The interpolation method with the smallest RMSE value is the best method. The results of RMSE calculations for each interpolation method are presented in Table 2.

Table 2. RMSE Value of each Interpolation Method

Interpolation Method		RMSE
Natural Neighbor		1.617307
IDW	Power = 1	1.257122
	Power = 2	**1.229251**
	Power = 3	1.261345
	Power = 4	1.318822
	Power = 5	1.376546

CONCLUSION

Based on the RMSE value, the IDW Interpolation Method with the power parameter 2 yields much more accurate prediction values than the Natural Neighbor Interpolation Method. Areas with low air temperatures are located around Kota Batu and Kecamatan Poncokusumo, while areas with high temperatures are in the vicinity of Malang and Bantur Subdistricts.

REFERENCES

[1] D. F. Watson dan G. M. Phillip, "A Refinement of Inverse Distance Weighted Interpolation," *Geo-Processing,* vol. 2, pp. 315-327, 1985.

[2] G. H. Pramono, "Akurasi Metode IDW dan Krigging untuk Interpolasi Sebaran Sedimen Tersuspensi," *Forum Geografi,* vol. 22, no. 1, pp. 97-110, 2008.

[3] R. Sibson, "A Brief Description of Natural Neighbor Interpolation (Chapter 2)," dalam *Interpreting Multivariate Data*, Chichester, John Wiley, 1981, pp. 21-36.

[4] S. Anderson, An Evaluation of Spatial Interpolation Methods on Air Temperature in Phoenix, Arizona: Department of Geography, Arizona University, 2001.

[5] D. F. Watson and G. M. Phillip, "Neigborhood Based Interpolation," *Geobyte,* vol. 2, pp. 12-16, 1987.

[6] D. Gumprecht, "Treatment of Far-Off Objects in Moran's I Test," Vienna University of Economics and Business Administration, Vienna, 2007.

[7] M. J. Fortin dan M. R. T. Dale, Spatial Analysis: A Guide for Ecologist, NewYork: Cambridge University Press, 2005.

[8] S. Naoum dan I. K. Tsanis, "Ranking Spatial Interpolation Techniques Using A GIS-Based DSS," *Global Nest Journal,* vol. 6, no. 1, pp. 1-20, 2004.

[9] V. Merwade, D. R. Maidment dan J. A. Goff, "Anisotropic Considerations while Interpolating River Channel Bathymetry," *Journal of Hydrology,* vol. 331, no. 3, pp. 731-741, 2006.

[10] M. Amstrong, Basic Linear Geostatistics, NewYork: Springer, 1998.