Investigating consequences of choosing IDW interpolation parameters in East Java using raster analyses

A Kurniawan¹*, M Ryan², A M Rafi¹, B E A Haq², Sudirman³ and A Suprayitno¹

1 Malang Climatological Station, Malang, 65152, Indonesia
2 Meteorology Climatology and Geophysics Agency Head Quarter, Angkasa I Street
2, Kemayoran, Jakarta, 10720, Indonesia
3 Rahadi Usman Meteorological Staion, Ketapang, 78851, Indonesia

*E-mail: andang.kurniawan@bmkg.go.id/ andangtedx@gmail.com

Abstract. An observation network will never be enough for creating good information about monthly rainfall. An interpolation method is always needed. For operational purposes, inverse distance weighting (IDW) method is used. In East Java, 197 observation points are involved, then IDW's parameters used are neighbor=12 and power=2. The consequences of this framework are investigated in this study. By reversing IDW's formula, two kinds of raster analyses are developed, distance to neighbor used (DNU) and coefficient from point (CFP). DNU shows how far points are used for doing interpolation in some area by kilometer (km) meanwhile CFP describes an area impacted by a point and value sent to another cell from an observation point. Data used in this study are longitude and latitude of 197 observation points. The scripts are written in R Language. Analysis based on local governmental region shows that Sumenep has very far DNU. In average, the first point used located more than 25 km, and its twelfth is 112 km (average other regions DNU1=7 km and DNU12=35 km). It means there should be a unique interpolation mechanism for Sumenep. CFP confirms that some points give impact in unnatural ways (impacted area=5741 km2). We propose DNU and CFP as alternative quality control parameters for investigating consequences in interpolating rainfall.

Keywords: IDW, raster analysis, DNU, CFP

1. Introduction
There are more than one thousand rainfall observation points in East Java. One thousand nineteen of them have registration identities given by Indonesia's BMKG (Meteorology Climatology and Geophysics Agency). According to a previous study, the quality is below 20 percent [1]. That is why not all can be used to produce operational rainfall information. East Java itself is not only on Java main-island. There are at least three other islands (Madura, Kangean, and Masalembu). Madura is the second largest.
An observation network will never be enough for creating good information about monthly rainfall. However, it is an obligation for Malang Climatological Station to provide information. Interpolation is an essential bridge in climate analysis due to its flexibility. Inverse Distance Weighting (IDW) also can be used as a filling empty data method [2]. Creating gridded rainfall data in Indonesia is a big challenge. Interpolation with more and more data through the nation’s border is done. SACAD is an example [3]. Nevertheless, instead of focusing on big raster thing, this study will pay more attention to more local things. Problems in interpolating rainfall are not that there is no method. There are a lot of interpolation methods, and studies about searching for the best ones [4][5]. Even after the best is known, there is still a problem, consequences of using any interpolation. In this study, we try to do a consequences-based investigation.

With increasing complexity, East Java province has the highest number of lower regional governments in Indonesia. Thirty-eight regional governments consist of twenty-nine regencies (‘Kabupaten’) and nine cities (‘Kota Madya’) [6]. It will be better if the quality of information is in good distribution. Efforts for creating information fairness should see this administrative condition.

The interpolation method is always needed. For operational purposes, IDW method is used. This study will not focus on searching the best interpolation method. Previous studies give underlying this study that in East Java, IDW is better than Spline, Ordinary Kriging, and Regression Kriging in interpolating monthly rainfall [7][8]. In East Java, 197 observation points are involved, then IDW’s parameters used are neighbor=12 and power=2. The consequences of this framework are investigated in this study by developing two parameters. We try to find what still can be wrong even if the best interpolation is used.

2. Data and Method

Data used in this study are coordinates from one hundred and ninety-seven main rainfall observation points (called "Pos Hujan Utama" in Bahasa) in East Java. East Java Province is astronomically bordered by 110.89°E–116.27°E and 8.78°S–5.04°S. All the rainfall observation points used are interpolation material for monthly rainfall information in East Java [9]. In figure 1, the dots from the right ones are the 197 observation points, and the black filled area shows the area position inside the maritime continent and shapefile of the regional administration.

IDW is formulated below [11].

\[
\lambda_i(s_0) = \frac{1}{\sum_{i=0}^n \frac{1}{d^\alpha(s_0,s_i)}} ; \alpha > 1
\]  

(1)

Nearest neighbor (n) is set to 12, and power (\( \alpha \)) is set to 2 at operational level. Using n =12, we can create a threshold for making an area that is not influenced by points be null cell. Then coefficients are counted that will be used on the cell with \( \alpha = 2 \). By reversing IDW’s formula, two kinds of raster analysis
are developed, distance to neighbor used (DNU) and coefficient from point (CFP). DNU shows how far points are used for doing interpolation in some area by kilometer (km). Meanwhile, CFP describes an area impacted by a point and value sent to another cell from an observation point. R Statistics software is used for processing in this study [12]. Package raster is also used to enable raster calculations [13]. The scripts used in this study are written in R Language, available by contact through the first author.

3. Result and Discussion
The first script is used to make a matrix contains center point coordinate both longitude and latitude in resolution 0.01 degree or about one kilometer. Resolution 0.01 degree is justified based on the evaluation of distance between points. From 197 points, the parameter recommends 0.018 degrees [14]. Extent is then set longitude 110.795 until 116.295 and latitude -9.005 until -5.005. The raster will be used as a base in the subsequent analysis.

![Figure 2. Plotted DNU from the first neighbor (left), seventh (middle), and twelfth (right)](image)

The first raster is calculated for distance from a cell to a first nearest point in kilometer. The second raster is for the second nearest up until the twelfth be the farthest observation used in the specific point. Unit of kilometer is used for making the interpreting easier for stakeholder. It can be easier due to its popularity as a unit of distance in Indonesia. Figure 2 shows DNU plotted all over East Java. DNU Distance 1 means the distance between a cell representing a location, with its first point or nearest used as interpolation material for that place. There is some malicious pattern in Probolinggo where id Distance 12 other area shows more than 20 km, Probolinggo shows below 20 km (green). It means there is something not acceptable there.

![Figure 3. Average of DNU based on 38 local administrations](image)
Analysis based on local governmental region shows that Sumenep (showed in black dash line, figure 3) has very far DNU. In average, the first point used located more than 25 km, and its twelfth is 112 km (average other regions DNU1=7 km and DNU12=35 km). It means there should be a unique interpolation mechanism for Sumenep. The second highest is Gresik (showed in solid black line). This fact can be related to its islands separated from Java. Gresik has Bawean islands north of Java, and Sumenep has at least Kangean islands north-east of Java, all separated by sea.

Only after computing DNU, the CFP can be produced. There are two features of CFP. First, CFP can show how far and how significant a point impacts on its relatively nearest cells of a raster. The top two, from an area perspective, are Ponorogo and Tanjung Kamal observation points. Both of them give impact to an area of 5741 km\(^2\). As shown in figure 4, Ponorogo-point has an impact far through Pacitan and Magetan regency in a weird “V”-shape. Meanwhile, Tanjung Kamal-point that is located in Java gives an impact on jumping the sea until Kangean Island. These two conditions are very unnatural.

The second feature of CFP is its ability to describe how high values are added to the raster from a point. Figure 4 shows coefficients that sent from a point (in percent). 100% means that the rainfall of the cell will be the same as the point. For example, if the coefficient displayed is 80%, then the cell will have 0.8 multiplied with rainfall from the source point. By accumulated (sum) all the coefficients, we can know how high is the impact of a point to raster created after interpolation.

| Regency      | Observation Point | Area Covered (km\(^2\)) | Sum Coefficient(unitless) |
|--------------|-------------------|--------------------------|---------------------------|
| Probolinggo  | Asemjajar          | 1636                     | 419                       |
| Probolinggo  | Klampokan          | 2037                     | 410                       |
| Probolinggo  | Batur              | 2130                     | 406                       |
| Probolinggo  | Paiton             | 1358                     | 394                       |
| Probolinggo  | Glagah             | 1760                     | 375                       |

CFP detects there are too many observation points in Probolinggo. The table above shows the five lowest CFP. It all came from Probolinggo. Probolinggo (regency and city combined) only has an area of about 1752 km square (1696+56) [15]. Compared to East Java, it is only 3.6% of the area. If we use simple principle of fair distribution, there should be only seven observation points (3.6% x 197). It confirms that 12 points from Probolinggo bring inequality to the system. Very low CFP and DNU confirmed this condition.

If Table 1 shows five lowest, Table 2 shows five highest. There is a similarity between all these five observation points. They are located on the edge of East Java. Arjasa and Stamet Kalianget are on the edge of east, Stamet Bawean on north, Pacitan on south-west, both Tegaldlimo and Grajagan are on
south-east. It can be related to the DNU plot these cells in the edge of the East Java raster have high DNU. It can be concluded that DNU and CFP can show problem in interpolation of these area.

Table 2. Five highest CFP’s sum coefficient values

| Regency | Observation Point | Area Covered | Coefficient |
|---------|------------------|--------------|-------------|
| Sumenep | Arjasa            | 3118         | 5662        |
| Banyuwangi | Grajagan         | 3025         | 1908        |
| Banyuwangi | Tegaldlimo      | 3056         | 1870        |
| Sumenep | Stamet Kalianget | 5618         | 1802        |
| Pacitan | Pacitan           | 2284         | 1710        |
| Gresik | Stamet Bawean    | 833          | 1594        |

Sumenep is a unique regency compared to the rest. There are nine sub-regencies (‘Kecamatan’) whose 126 islands are separated from Madura Island [16]. Its area is also spread from the east of Madura up until north Bali Island. It is against our common sense if the rainfall in this area is produced using IDW. So, we propose these islands’ rainfall information should be generated using satellite-based with or without correction. For a more straightforward solution, we factually can use the GPM products [17].

Similar to Sumenep’s case, the Gresik regency has islands separated from Java as an island. There are two sub-regencies there, Sangkapura and Tambak [18]. These two are located in the Bawean islands. The distance of its nearest shore to Java’s one is more than 100 km. The same solution should be applied here. For better a view of these problems, we plot it below.

![Figure 5. Three gridding zonation: Java-interpolation (red), Madura-interpolation (orange), and independent-gridding (blue)](image)

The alternatives for this distribution problem are making interpolation in Madura Island and archipelagos near it going to be independent. By using raster manipulation, Java-part and the others can be interpolated in non-single step. It is not the first time that interpolation is done only for Java, excluded Madura [19]. The idea behind this kind of interpolation is that different islands can create different rainfall. Sumenep’s and Gresik’s cases do empower this idea. By using raster manipulation provided by raster packages in R Statistics, it is very possible to do many gridding methods, and then merge it into one raster. The other way is by not using constant neighbor in interpolation. Instead, interpolation can be computed by distance limitation.
4. Conclusion
We propose DNU and CFP as alternative quality control parameters for investigating consequences in rainfall interpolation. These two parameters are easy to be understood and relatively light to be computed. A very straightforward way to use the parameter is by giving attention to its maximum and minimum values. Then check something wrong that could be happened. In the East Java case, malicious DNU and CFP values bring some recommendation that interpolation should not be done in a single gridding system. We believe that climate information should give focus on local stakeholders. This study also proves and gives some examples of how data from Statistics Agency can be used in determining optimal rainfall observation network.

Acknowledgments
This project is one of five projects dedicated to participating in ICTMAS called ASRAB. All authors declare that there is no conflict of interest in doing this project. We would like thanks all entities that give us a chance to produce these kinds of project.

References
[1] Supari 2012 Spatio-temporal Characteristics of Extreme Rainfall Event over Java Island Indonesia Thesis Gajah Mada University
[2] Abbas A A, A A Nashaat, G Abozeid and H I Mohamed 2019 Comparison of Different Methods for Estimating Missing Monthly Rainfall Data 22nd Int. Water Tech. Conf. pp 569-83
[3] van den Besselaar E J M, G van der Schrier, R C Cormes, A S Iqbal and A M G Klein Tank 2017 J. Clim. 30 5151–65
[4] Basconcillo J Q, G A W Duran, A A Francisco, R G Abastillas, F D Hilario, E I Juanillo, A L S Solis, A J Lucero and S A Maratas 2017 SOLA 13 114-119
[5] Nguyen-Xuan T, T Ngo-Duc, H Kamimera, L Trinh-Tuan, J Matsumoto, T Inoue and T Phan-Van 2016 SOLA 12 291-6
[6] Indonesia’s Statistics Agency Head Quarter 2020 Statistics Indonesia (Jakarta: BPS)
[7] Hanif K, E Aprilia, J B Utomo, A Kurniawan and A Safril 2018 Pros. Sem. Nas. Geotik Vol 2, Ed B Muriyasa et al (Surakarta: University Muhammadiyah Surakarta) pp 213-220
[8] A Kurniawan, E Makmur and Supari 2021 Pros. Sem. Nas. Geomatika 2020, R Ati et al (Bogor: Badan Informasi Geospasial) p 263-72
[9] Kurniawan A and D Tarmana 2019 J. Met. Klim. dan Geo. 5 40-48
[10] Mahmudiah R R, A Kurniawan and E S Makmur 2019 IOP Conf. Ser.: Earth Environ. Sci. 303 012023
[11] Giarno, D Didiharyono, A A Fisu and A Mattingaragau 2020 IOP Conf. Ser.: Earth Environ. Sci. 469 012003
[12] R Core Team 2020 R: A language and environment for statistical computing (Vienna: R Foundation)
[13] Hijmams R J et al 2020 Geographic Data Analysis and Modeling (CRAN/raster)
[14] Hengl T 2006 Computers & Geosciences 3 2 1283–98
[15] East Java Province Office Statistics Agency 2020 Jawa Timur Province in Figures (Surabaya: BPS)
[16] Sumenep Regency Office Statistics Agency 2020 Sumenep Regency in Figures (Sumenep: BPS)
[17] Huffman G J, D T Bolvin, D Braithwaite, K Hsu, R Joyce, C Kidd, E J Nelkin, S Sorooshian, J Tan and P Xie 2018 NASA Global Precipitation Measurement (GPM) Integrated Multi-Satellite Retrievals for GPM (IMERG) NASA
[18] Gresik Regency Office Statistics Agency 2020 Gresik Regency in Figures (Sumenep: BPS)
[19] Yanto, B Livneh and B Rajagopalan 2017 Development of a gridded meteorological dataset over Java island, Indonesia 1985–2014 Sci. Data (10.1038/sdata.2017.72)