A Comparison of Spatial Interpolation Methods for Regionalizing Maximum Daily Rainfall Data in South Sulawesi, Indonesia

Wahidah Sanusi¹, Sahlan Sidjara¹, Sudarmin Patahuddin², and Muhammad Danial³

¹Department of Mathematics, Universitas Negeri Makassar, Makassar, South Sulawesi, Indonesia
²Department of Statistics, Universitas Negeri Makassar, Makassar, South Sulawesi, Indonesia
³Department of Chemistry, Universitas Negeri Makassar, Makassar, South Sulawesi, Indonesia

Abstract. The aim of this research is to compare between the Inverse Distance Weighted (IDW) and Ordinary Kriging (OK) interpolation methods for regionalization of areas within the South Sulawesi province based on maximum daily rainfall. The data utilized consists of maximum daily rainfall data from 56 rain stations within the South Sulawesi from 1986 to 2021. The spatial interpolation methods applied include the power 2 IDW, and OK. Various semivariogram models, namely Spherical, Gaussian, and Exponential, are employed within the OK method. The selection of the best method is based on the smallest Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) values. The findings of this research reveal that the optimal method for regionalization of maximum daily rainfall is the OK method with a Gaussian semivariogram model. The RMSE values for this method are 57.45, and the MAE values are 46.49. The results of the spatial interpolation demonstrate that the South Sulawesi is divided into four zones characterized by maximum daily rainfall (in mm) as follows: Zone I: less than 230 mm (Eastern and Southeastern regions), Zone II: 230-260 mm (Northern region), Zone III: 260-280 mm (Western region), and Zone IV: more than 280 mm (Southwestern region).

1 Introduction

One of the common challenges encountered in spatial hydrological data modeling, such as rainfall data, is the absence of rainfall measurement instruments at all locations. This occurs when there are no rain gauge stations available at specific locations. To address this issue, rainfall data from nearby stations can be utilized to estimate the missing data. One of the methods for estimating values at locations lacking data is spatial interpolation. Spatial interpolation assumes that the attributes of the data are continuous in space and exhibit spatial relationships. Several spatial interpolation methods have been widely employed by researchers, including Multiple Linear Regression, Local Polynomial, Inverse Distance Weighted, Ordinary Kriging, Simple Kriging, Universal Kriging, and Empirical Bayesian Kriging (EBK) [1-3].

* Corresponding author: wahidah.sanusi@unm.ac.id

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).
Boumpoulis et al. [1] employed deterministic and geostatistical interpolation methods in their study of sediment data in Western Greece. The deterministic interpolation methods utilized included Radial Basis Function, Local Polynomial, and Inverse Distance Weighted. Meanwhile, the geostatistical interpolation methods employed were Ordinary Kriging, Simple Kriging, Empirical Bayesian Kriging, and Universal Kriging. Caloiero et al. [2] employed the Inverse Distance Weighted, Ordinary Kriging, the Kriging with External Drift, and Ordinary Cokriging interpolation methods to interpolate monthly rainfall data in New Zealand. They conducted a comparative analysis of these four interpolation methods. Similarly, Chen et al. [3] utilized the Principal Component Regression with Residual Correction, Inverse Distance Weighted, and Multiple Linear Regression interpolation methods for annual precipitation data in southeastern China. They also conducted a comparative assessment of these three methods. Hariyanti [4] utilized the Inverse Distance Weighted, Ordinary Kriging, and Natural Neighbour interpolation methods for the regionalization of annual rainfall data in relation to cocoa production levels in the South Sulawesi.

Based on the given information, the objective of this study is to assess and compare the effectiveness of the Inverse Distance Weighted (IDW) and Ordinary Kriging (OK) interpolation techniques in delineating the spatial distribution of maximum daily rainfall within the South Sulawesi.

2 Methods

2.1 Data source

The data utilized in this research consists of maximum daily rainfall data from 56 rain gauge stations located within the South Sulawesi during the period from 1986 to 2021 (Table 1 and Figure 1). These data were obtained from the Water Resources, Human Settlements, Spatial Planning and Development Office of South Sulawesi Province, and the Meteorology, Climatology, and Geophysics Agency (BMKG) of Makassar.

Table 1. Names of rain gauge stations

<table>
<thead>
<tr>
<th>No.</th>
<th>Name of station</th>
<th>Location</th>
<th>No.</th>
<th>Name of station</th>
<th>Location</th>
<th>No.</th>
<th>Name of station</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Moti</td>
<td>Bantaeng</td>
<td>20</td>
<td>Padangloang</td>
<td>Sidrap</td>
<td>39</td>
<td>Lanrae</td>
<td>Barru</td>
</tr>
<tr>
<td>2</td>
<td>Tino Toa</td>
<td>Bantaeng</td>
<td>21</td>
<td>Lamasi</td>
<td>Luwu</td>
<td>40</td>
<td>Leworeng</td>
<td>Soppeng</td>
</tr>
<tr>
<td>3</td>
<td>Onto</td>
<td>Bantaeng</td>
<td>22</td>
<td>PadangSappa</td>
<td>Luwu</td>
<td>41</td>
<td>Lalange Lajoa</td>
<td>Soppeng</td>
</tr>
<tr>
<td>4</td>
<td>Pandang-pandang</td>
<td>Bulukumba</td>
<td>23</td>
<td>Bajo</td>
<td>Luwu</td>
<td>42</td>
<td>Salobunne</td>
<td>Soppeng</td>
</tr>
<tr>
<td>5</td>
<td>Paenrelompoa</td>
<td>Bulukumba</td>
<td>24</td>
<td>Salokarajae</td>
<td>Enrekang</td>
<td>43</td>
<td>Tabo-tabo</td>
<td>Pangkep</td>
</tr>
<tr>
<td>6</td>
<td>Bettu</td>
<td>Bulukumba</td>
<td>25</td>
<td>Maroangin</td>
<td>Enrekang</td>
<td>44</td>
<td>Leanglonrong</td>
<td>Pangkep</td>
</tr>
<tr>
<td>7</td>
<td>Arango</td>
<td>Sinjai</td>
<td>26</td>
<td>Talang Raja</td>
<td>Enrekang</td>
<td>45</td>
<td>Sigeri</td>
<td>Pangkep</td>
</tr>
<tr>
<td>8</td>
<td>Palangka</td>
<td>Sinjai</td>
<td>27</td>
<td>Latapppa</td>
<td>Palopo</td>
<td>46</td>
<td>Camba</td>
<td>Maros</td>
</tr>
<tr>
<td>9</td>
<td>Bikeru</td>
<td>Sinjai</td>
<td>28</td>
<td>Tandung Nanggala</td>
<td>Toraja Utara</td>
<td>47</td>
<td>Panyalingan</td>
<td>Maros</td>
</tr>
<tr>
<td>10</td>
<td>Biru</td>
<td>Bone</td>
<td>29</td>
<td>Kaluku</td>
<td>Luwu utara</td>
<td>48</td>
<td>Maroangin</td>
<td>Maros</td>
</tr>
<tr>
<td>11</td>
<td>Unra</td>
<td>Bone</td>
<td>30</td>
<td>Pamhassian</td>
<td>Luwu utara</td>
<td>49</td>
<td>Paoterere</td>
<td>Makassar</td>
</tr>
<tr>
<td>12</td>
<td>Bengo</td>
<td>Bone</td>
<td>31</td>
<td>Katulungan</td>
<td>Luwu utara</td>
<td>50</td>
<td>Ujuangrandang</td>
<td>Makassar</td>
</tr>
<tr>
<td>13</td>
<td>Palattae</td>
<td>Bone</td>
<td>32</td>
<td>Wonorefo</td>
<td>Luwu Timur</td>
<td>51</td>
<td>Bontosallang</td>
<td>Gowa</td>
</tr>
<tr>
<td>14</td>
<td>Ongkoe</td>
<td>Wajo</td>
<td>33</td>
<td>Margalenmo</td>
<td>Luwu Timur</td>
<td>52</td>
<td>Paku</td>
<td>Gowa</td>
</tr>
<tr>
<td>15</td>
<td>Welle</td>
<td>Wajo</td>
<td>34</td>
<td>Paria</td>
<td>Pinrang</td>
<td>53</td>
<td>Sungaminasa</td>
<td>Gowa</td>
</tr>
<tr>
<td>16</td>
<td>Cannu</td>
<td>Wajo</td>
<td>35</td>
<td>Data</td>
<td>Pinrang</td>
<td>54</td>
<td>Paloko</td>
<td>Takalar</td>
</tr>
<tr>
<td>17</td>
<td>Bulu cemrana</td>
<td>Sidrap</td>
<td>36</td>
<td>Benteng</td>
<td>Pinrang</td>
<td>55</td>
<td>Tanrang</td>
<td>Jeneponto</td>
</tr>
<tr>
<td>18</td>
<td>Talawe</td>
<td>Sidrap</td>
<td>37</td>
<td>Manuba</td>
<td>Barru</td>
<td>56</td>
<td>Bulo-Bulo</td>
<td>Jeneponto</td>
</tr>
<tr>
<td>19</td>
<td>Tanete</td>
<td>Sidrap</td>
<td>38</td>
<td>Mareppang</td>
<td>Barru</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.2 Spatial interpolation

Interpolation is a method used to estimate values at locations where data is not available [4]. There are two interpolation methods that can be employed: deterministic and probabilistic (geostatistical) methods [5]. In this study, the deterministic interpolation method used is the Inverse Distance Weighting (IDW) method, while the geostatistical interpolation method is the Ordinary Kriging (OK) method.

2.2.1 Inverse Distance Weighting (IDW)

This method is an estimation approach, where the average value of an estimation point/target point \( \hat{X}_{IDW} \) is the weighted average of sample values \( (X) \) around the target point. The data points closest to the target point receive greater weights, whereas data points farther away from the target point receive smaller weights. These weights are inversely proportional to the distance between the target point and the sample point. The formula for obtaining the estimation point \( \hat{X}_{IDW} \) using the IDW method is provided in Equation (1).
\[ \hat{X}_{IDW} = \sum_{i=1}^{n} \left( \frac{d_i^p}{\sum_{i=1}^{n} d_i^p} \right) X_i \]  

(1)

Here, \( X_i \) represents the sample data at location \( i \), \( d_i \) represents the Euclidean distance between the target point and the sample data at location \( i \), \( n \) is the total number of sample data points, and \( p \) is a weighting power. Typically, a value of \( p = 2 \) is commonly used [5].

### 2.2.2 Ordinary Kriging (OK)

Ordinary Kriging is a geostatistical method employed to estimate the value at a specific point as a linear combination of sample values from neighboring points. The estimation of the point's value is determined using a weighting function. Kriging weights are derived from the minimization of variance estimation, employing the semivariogram. The accuracy of kriging estimation heavily relies on the selected semivariogram model [4]. In this study, three types of semivariogram models are utilized, namely Spherical, Gaussian, and Exponential. The formulations for these three semivariogram models, \( \gamma(. ,) \), are respectively provided in Equations (2)-(4) [5-7].

- **Spherical model**
  \[
  \gamma(h; \theta) = \begin{cases} 
  C_0 + C_1, & h > R \\
  C_0 + C_1 \left[ \frac{5h}{R} - 0.5 \left( \frac{h}{R} \right)^3 \right], & 0 < h \leq R \\
  0, & h = 0 
  \end{cases} \quad (2)
  \]

- **Gaussian model**
  \[
  \gamma(h; \theta) = \begin{cases} 
  C_0 + C_1 \left[ 1 - \exp \left( -3 \left( \frac{h}{R} \right)^2 \right) \right], & h > 0 \\
  0, & h = 0 
  \end{cases} \quad (3)
  \]

- **Exponential model**
  \[
  \gamma(h; \theta) = \begin{cases} 
  C_0 + C_1 \left[ 1 - \exp \left( -\frac{3h}{R} \right) \right], & h > 0 \\
  0, & h = 0 
  \end{cases} \quad (4)
  \]

Where \( h \) represents the lag or distance, \( \theta = (C_0, C_1, R)' \), where the parameter \( C_0 \) signifies the nugget effect (\( C_0 \geq 0 \)), \( C_1 \) denotes the partial sill (\( C_1 \geq 0 \)), and \( (C_0 + C_1) \) represents the sill, while \( R \) stands for the range (\( R > 0 \)).

The formula for obtaining the Ordinary Kriging estimation point (\( \hat{X}_{OK} \)) using the OK method is given in Equation (5).

\[ \hat{X}_{OK} = \sum_{i=1}^{n} W_i X_i \]  

(5)

Where \( W_i \) represents the weight (\( 0 < W_i < 1 \), \( \sum_{i=1}^{n} W_i = 1 \)), and these weight values are estimated based on the semivariogram estimation [5].

### 2.3 Interpolation method selection

The best interpolation method is selected based on the indicators of the minimum Root Mean Square Error (RMSE) and the minimum Mean Absolute Error (MAE) values [1, 8-10]. The formulations for each indicator are provided in Equations (6) and (7).

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \hat{X}_i)^2}{n}} \quad (6)
\]

\[
MAE = \frac{\sum_{i=1}^{n} |X_i - \hat{X}_i|}{n} \quad (7)
\]
3 Results and discussion

In this study, maximum daily rainfall data from 56 rain gauge stations in South Sulawesi were utilized. As shown in Figure 2, it is evident that at station 49, specifically Maroangin Maros station, the highest recorded maximum daily rainfall reached 400 mm. Conversely, station 1, known as Moti Bantaeng station, registered the lowest maximum daily rainfall, which was 120 mm.

![Fig. 2. Scatter plot of data](image)

Subsequently, this data was used to compare the Inverse Distance Weighted (IDW) method and the Ordinary Kriging (OK) method. However, in the OK method, the semivariogram values for each model were initially estimated. Figure 3 displays the semivariogram graphs for each model, along with their respective nugget, sill, and range values. Based on these three figures, it is evident that the Gaussian model exhibits the smallest sum of squared errors (SSE). Table 2 also illustrates that the OK method with the Gaussian semivariogram model provides the smallest RMSE and MAE values compared to the other two semivariogram models.

<table>
<thead>
<tr>
<th>Indicator</th>
<th>IDW</th>
<th>Ordinary Kriging</th>
</tr>
</thead>
<tbody>
<tr>
<td>p=2</td>
<td>59.95651</td>
<td>58.15767</td>
</tr>
<tr>
<td>RMSE</td>
<td>57.45251</td>
<td>57.84119</td>
</tr>
<tr>
<td>MAE</td>
<td>46.49909</td>
<td>46.75477</td>
</tr>
</tbody>
</table>

Table 2. RMSE and MAE values
Table 3 indicates that the results of maximum daily rainfall estimation from both methods have minimum and maximum values that significantly differ from the observed data. However, they exhibit median and mean values that closely resemble the observed data.

Table 3. Descriptive Statistics of Observed and Estimated Maximum Daily Rainfall Data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>1st Quartile</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Quartile</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation</td>
<td>120.0</td>
<td>205.8</td>
<td>239.5</td>
<td>249.4</td>
<td>282.5</td>
<td>400.0</td>
</tr>
<tr>
<td>IDW p=2</td>
<td>195.3</td>
<td>223.8</td>
<td>237.9</td>
<td>250.7</td>
<td>283.7</td>
<td>361.9</td>
</tr>
<tr>
<td>Spherical</td>
<td>200.8</td>
<td>225.2</td>
<td>234.3</td>
<td>249.4</td>
<td>267.1</td>
<td>338.5</td>
</tr>
<tr>
<td>Gaussian</td>
<td>202.7</td>
<td>227.0</td>
<td>239.1</td>
<td>249.8</td>
<td>271.5</td>
<td>335.9</td>
</tr>
<tr>
<td>Exponential</td>
<td>198.5</td>
<td>226.4</td>
<td>235.2</td>
<td>249.5</td>
<td>266.7</td>
<td>340.7</td>
</tr>
</tbody>
</table>

For the regionalization of South Sulawesi based on maximum daily rainfall, the Ordinary Kriging (OK) interpolation method with a Gaussian semivariogram model was employed. The results are presented in Figure 4, which divides South Sulawesi into the following zones: Zone I: Eastern and Southeastern Regions (less than 230 mm), Zone II: Northern Region (230-260 mm), Zone III: Western Region (260-280 mm), Zone IV: Southwestern Region (more than 280 mm).

Research results related to regional rainfall in South Sulawesi have also been carried out by Yanto [4] who divided South Sulawesi into five regions based on amount annual rainfall (in mm), where the two regions in Yanto's research (region-I and region-II) became one region in our study (Zone I).
4 Conclusion

This research concludes that the Ordinary Kriging interpolation method with a Gaussian semivariogram model outperforms the IDW method in regionalizing South Sulawesi based on maximum daily rainfall. The regionalization results classify South Sulawesi into four zones.

The authors would like to thank the Ministry of Education, Culture, Research, and Technology of Indonesia for providing funding. This research was financed by grant number: 2402/UN36.11/LP2M/2023. The authors also thank the staff of the Meteorological, Climatological, and Geophysical Agency (BMKG) of Makassar, and the Water Resources, Human Settlements, Spatial Planning and Development Office of South Sulawesi Province, Indonesia for supplying rainfall data.

References

