Accuration of Time Series and Spatial Interpolation Method for Prediction of Precipitation Distribution on the Geographical Information System

To cite this article: S Y J Prasetyo and K D Hartomo 2018 IOP Conf. Ser.: Mater. Sci. Eng. 288 012140

View the article online for updates and enhancements.
Accuration of Time Series and Spatial Interpolation Method for Prediction of Precipitation Distribution on the Geographical Information System

S Y J Prasetyo* and K D Hartomo
Study Centre SIMITRO - Informatic Engineering
Satya Wacana Christian University, Salatiga, Indonesia

* sri.yulianto@staff.uksw.edu

Abstract. The Spatial Plan of the Province of Central Java 2009-2029 identifies that most regencies or cities in Central Java Province are very vulnerable to landslide disaster. The data are also supported by other data from Indonesian Disaster Risk Index (In Indonesia called Indeks Risiko Bencana Indonesia) 2013 that suggest that some areas in Central Java Province exhibit a high risk of natural disasters. This research aims to develop an application architecture and analysis methodology in GIS to predict and to map rainfall distribution. We propose our GIS architectural application of “Multiplatform Architectural Spatiotemporal” and data analysis methods of “Triple Exponential Smoothing” and “Spatial Interpolation” as our significant scientific contribution. This research consists of 2 (two) parts, namely attribute data prediction using TES method and spatial data prediction using Inverse Distance Weight (IDW) method. We conduct our research in 19 subdistricts in the Boyolali Regency, Central Java Province, Indonesia. Our main research data is the biweekly rainfall data in 2000-2016 Climatology, Meteorology, and Geophysics Agency (In Indonesia called Badan Meteorologi, Klimatologi, dan Geofisika) of Central Java Province and Laboratory of Plant Disease Observations Region V Surakarta, Central Java. The application architecture and analytical methodology of “Multiplatform Architectural Spatiotemporal” and spatial data analysis methodology of “Triple Exponential Smoothing” and “Spatial Interpolation” can be developed as a GIS application framework of rainfall distribution for various applied fields. The comparison between the TES and IDW methods show that relative to time series prediction, spatial interpolation exhibit values that are approaching actual. Spatial interpolation is closer to actual data because computed values are the rainfall data of the nearest location or the neighbour of sample values. However, the IDW’s main weakness is that some area might exhibit the rainfall value of 0. The representation of 0 in the spatial interpolation is mainly caused by the absence of rainfall data in the nearest sample point or too far distance that produces smaller weight.

1. Introduction

It is estimated that the Geographic Information System (GIS) software technology began to develop in 1960 as the result of accumulated research in geography, cartography, and computer science [1]. GIS then develops into a main instrument to generate solutions of problems in mapping, scientific documentation, artworks, thematic mapping, and spatial visualization and analysis based on statistical function [2] [3]. The GIS architectural application itself consists of four layers, namely: (1) data entry and modelling, (2) preparation of data storage media, (3) analysis system and method, and (4)
geographic information representation or output [4][5][6][7][8][9]. One can make use of GIS application for various objectives, such as in agriculture, forestry, epidemiology, and disaster early-warning [10][11][12][13]. There has been also rapid development in the analysis techniques and methods in GIS such as linear regression, spatial interpolation, and time-series and spatial autocorrelation. This paper focuses on the disaster early-warning because Indonesia’s unique geographical position implies a higher vulnerability in the hydrometeorology disaster.

Based on the data from the National Disaster Prevention Agency, the hydrometeorological factors (i.e. typhoon, drought, flood, and landslide) account for 97% of natural disasters in Indonesia in 2013[14]. Landslide is mainly caused by changes in geological characteristics, rain intensity, and excessive land use in hillsides. Every year, Indonesia witnesses increasing landslide disasters intensity and more extensive distribution areas, especially in the provinces of Central Java, West Java, East Java, West Sumatera, and East Kalimantan [15]. For example, the Spatial Plan of the Province of Central Java 2009-2029 identifies several regencies or cities that are vulnerable to the landslide disaster, such as Kebumen, Purworejo, and Karanganyar. Also, according to the Indonesian Disaster Risk Index 2013, these regencies exhibit a high natural disaster risk. There are various methods to manage areas with high risk and vulnerability of landslide disaster, namely: (1) spatial and temporal monitoring of changes of area topography, (2) detection of vulnerability level of areas, (3) land movement simulation and modelling, (4) determination of land movement zonation, (5) early warning system of landslides and rainfall intensity distribution [16]. Rainfall intensity is an important factor in predicting land movement, especially in mountain areas. Efforts to mitigate hydrometeorology disasters, especially landslides, relies heavily on the rainfall intensity data [16]. Spatial and temporal models of the relationship between rainfall intensity and landslide incidents can predict rainfall-driven landslides. At global scale, currently there is no system to identify distribution of rainfall that can trigger landslide because of the limited number of rainfall measurement station networks that can monitor rainfall intensity on a real time based [17]. This research aims to develop an analysis application and methodology architecture on GIS to predict and map rainfall distribution. We propose our GIS architectural application of “Multiplatform Architectural Spatiotemporal” and data analysis methods of “Triple Exponential Smoothing” and “Spatial Interpolation” as our significant scientific contribution. We develop the architectures and methods in this application to handle the complexity of data analysis and to visualize the spatiotemporal information of hydrometeorology disaster mitigation. We subsequently compare the prediction and mapping results between the Triple Exponential Smoothing dan Spatial Interpolation methods.

2. The related works
Time series are a sequence of observation data that are arranged based on certain time variables with discrete interval pattern, in which past occurrences always repeat in the future [18]. The Exponential Smoothing (ES) method is one of the time series methods that are commonly used. ES itself consists of three models, namely Simple Exponential Smoothing (SES) to model stationary data, Double Exponential Smoothing (DES) to model trend data, and Triple Exponential Smoothing (TES) to model seasonal data. This research uses TES method (1-3).

\[ b_t = g(S_t - S_{t-1}) + (1 - g)b_{t-1} \]  
(1)

\[ I = b \cdot \frac{t^x}{tS+(1-b)t-L+m} \]  
(2)

\[ F_t + m = (S_t + b_t m)I_t - L + m \]  
(3)

We use the IDW as spatial interpolation function to perform spatial prediction. The IDW spatial interpolation is a function to compute the distance parameter between known value and unknown value of surrounding objects. The weight of each sample point is inversely proportional to the distance between points. The IDW method is used to predict unknown data based on spatial connectivity between a sample point and its surrounding points. It is also used to predict elevation value, rainfall, temperature, humidity, invasive plants, and other continuous phenomena [19][20]. The following is the IDW equation:
\[ Z(S_0) = \sum_{i=1}^{n} \lambda_i Z(S_i) \]  

(4)

where \( Z(S_0) \) is the unknown point, \( \lambda_i \) is the weight of known point \( Z(S_i) \). The weight value (\( \lambda_i \)) is determined by the following equation:

\[ \lambda_i = \left[ d(s_i, s_0) \right]^p / \sum_{i=1}^{n} \left[ d(s_i, s_0) \right]^p \]  

(5)

where \( d(s_i, s_0) \) is the euclidean distance between \( s_i \) and \( s_0 \) [21].

### 3. Methods

This research consists of two parts, namely data attribute prediction using the TES method and spatial data prediction using the IDW method. Exponential in the TES method computes moving average such that the sum of residual square between the actual data values and predictive data values is minimal. Each processed data value at period \( t \) is weighted \( \beta = (1-\alpha) \) where \( \beta < 1.0 \). Meanwhile, \( \alpha \) itself is the smoothing factor. Every historical data (data closest to the beginning of the process) has a smaller weight than the current data. This computation process produces predicted values for several future prediction periods. The predictive ability of the IDW method is determined by spatial connectivity of known and unknown sample points [21]. A comparison of Euclidean Distance of the equation 

\[ \lambda_i = \left[ d(s_i, s_0) \right]^p / \sum_{i=1}^{n} \left[ d(s_i, s_0) \right]^p \]  

between \( S_i \) and \( S_0 \) will produce specific rainfall prediction on unsampled location. The power value (\( p \)) gives more weight on input data with closer distance than input data with farther distance. As a consequence of more weight, the visualization of spatial information is more detailed.

![Rainfall Forecast using Triple Exponential Smoothing](image1)

![Spatial Interpolation using Inverse Distance Weight](image2)

### 4. Experiment

We conduct our research in 19 sub districts in Boyolali Regency, Central Java Province, Indonesia. We collect the 2000-2016 biweekly rainfall data from the Climatology, Meteorology, and Geophysics Agency of Central Java Province and Laboratory of Plant Disease Observations Region V Surakarta, Central Java. We then store our data attribute in MySQL database and shape files (ESRI) for spatial data. We use the Multiplatform Architecture Spatiotemporal (MAS) application architecture concept to develop the GIS software prototype. MAS is a platform of application architecture that consists of 3 application layers based on its programming language, namely (1) GUI application framework layer using Java programming language, (2) computing application layer with R (http://cran.r-project) programming language and (3) layer of spatial and temporal information layer using PHP programming language and Mapserver framework. Figure 2 below displays the MAS design.
Our research phases basically modify Prasetyo (2015) that consists of 4 stages. The first stage is literature review on our main topic, mainly prediction using Triple Exponential Smoothing and geographic information analysis, especially spatial interpolation. The second stage is the initial process of research data. At this stage we classify data into two groups, namely rainfall data that is predicted using TES and thematically mapped in shapes files, and data analysis using spatial interpolation of IDW mathematical function. We design and develop the prototype using Multiplatform Architecture Spatiotemporal at stage three. At this stage we also make three application modules, namely: (1) Framework module for GUI and Input Output (IO) application using Java programming language, (2) Computing and processing module that consists of three packages, namely ESSA, IDW and PRECI packages, made by R programming language, and (3) Mapserver and APACHE framework module, configured in PHP programming language. The fourth or last stage tests and implements GIS application to predict and map rainfall. At this stage we visualize information in the form of time series chart and spatial interpolation map and map of prediction of rainfall distribution.

5. Results and discussion
The prediction process consists of two stages, namely the initial data processing (data preprocessing) with data smoothing. Each data value is processed at period \( t \) by weighting it of \( \beta = (1 - \alpha) \) where \( \beta < 1.0 \). The \( \alpha \) is the smoothing constant. The smoothing constant of \( \alpha, \beta \) and \( \gamma \) take the value between 0 and 1. When the smoothing constant is equal to 1, there is no curve smoothing. However, when the value of smoothing constant is 0, there is a curve smoothing, no cyclical variation or the straight line [21]. We develop the following algorithm for smoothing processing stage predicts data using the following Holt-Winters additive method:

**Input:**
\[
data = (\text{precipitation}), \text{constanta}_\alpha = 0.95, \text{constanta}_\beta = 0.06, \text{constanta}_\gamma = 0.9, \text{periodic} = 12, \text{constanta}_\text{forecasting} = \text{constanta}_\text{error}, \text{level}, \text{trend}, \text{seasonal}, \text{MAE}
\]

**Function:**
For \( x=1 \) To \( N \) Do
level = constanta_level\(_i\)  
\( \text{trend} = \text{constanta}_\text{trend}\(_i\) \)  
seasonal = constanta_seasonal\(_i\)  
constanta_forecasting = {}  
For \( y=12 \) To \( M \) Do
level_{xy} = constanta_{alpha} \cdot (data_{xy} - seasonal_{xy-periodic}) + (1 - constanta_{alpha}) \\
    \cdot (level_{xy-1} + trend_{xy-1})

trend_{xy} = constanta_{beta} \cdot (level_{xy} - level_{xy-1}) + (1 - constanta_{beta}) \cdot trend_{xy-1}

seasonal_{xy} = constanta_{gamma} \cdot (data_{xy} - level_{xy}) + (1 - constanta_{gamma}) \cdot seasonal_{xy-periodic}

constanta_{forecasting_{xy}} = level_{xy} + trend_{xy} + seasonal_{xy}

End For

constaACP_error = {}

For \ y = 12 \ To \ M \ Do

  constaACP_error_{xy} = abs(data_{xy} - constanta_{forecasting_{xy}})

End For

MAE = \frac{\sum constantaACP_error}{M - 12}

constanta_{forecasting_{M+1}} = level_{M} + (trend_{M} \cdot 1) + seasonal_{M-periodic+1}

End Function

The TES method describes prediction data line that closely resembles actual line for the rainfall parameter in 2001-2003. This method represents predicted line (red) with two components that have to be updated when we make new prediction points, namely constant trend and level from historical data. Level value is estimated data value smoothing for each end of period while trend value is estimated growth average smoothing for each period. These methods analyze seasonal data independent of the average level value. Predicted and actual values show different repeating patterns. The variation of monthly cycle on predicted data (red) tend to exhibit similar patterns from previous years. We interpret the sharp increase of actual data (blue) as an abnormal data pattern or an anomaly and the repeating pattern follow previous data cycle. We visualize the spatial information in the interpolated map using Inverse Distance Weight (IDW). We determine the observation sampling points based on the village data as the smallest area unit (in the form of village polygon). The following algorithm explains the IDW computation process:

**Input:** data = (data_precipitation), location_data = coordinates(data_observation), location_sample = coordinates(observation_sample), data_idw, power = 2, estimated_value, weighted_value, estimated_value, distance

**Function:**

\begin{align*}
  &data_{idw} = \\
  &\text{For } x = 1 \text{ To } N \text{ Do} \\
  &  \text{estimated_value} = \\
  &  \text{For } y = 1 \text{ To } M \text{ Do} \\
  &  \text{distance} = observation_{sample_{x}} - observation_{data_{y}} \\
  &  \text{weighted_value}_{y} = \frac{1}{\text{distance}^{power}} \\
  &  \text{estimated_value}_{y} = \text{weighted_value}_{y} \cdot data_{observation_{y}} \\
  &\text{End For} \\
  & \text{estimated_value} = \frac{\sum \text{estimated_value}}{\sum \text{weighted_value}} \\
  &\text{End For} \\
  &\text{End Function}
\end{align*}

The computation process produces visualized rainfall distribution as can be seen at Figure 3 below. More specifically, Figure 3 displays the visual comparison of spatial information of rainfall using IDW method (a) and using time series or TES method (b). The following are the main differences between these two methods:
1. The spatial rainfall prediction using the IDW method works through the unknown sample point interpolation based on the neighbor’s known distance value.
2. The closer the distance between the unknown sample point and its neighbor’s sample point, the higher the weight point of the sample; implying more detailed spatial information presented.
3. We predict the rainfall data attribute using the time series prediction based on historical data. Prediction results are then the basis for thematic map of rainfall prediction.
4. The thematic map of rainfall distribution prediction uses administrative map as the basis of sample area position and different colors of the areas indicate different rainfall volume.

![Figure 3](image)

Figure 3. Visualization of spatial information of rainfall distribution prediction using IDW method (a) and thematic map of rainfall distribution (b).

Figure 4 show the comparison of rainfall prediction using the time series and IDW methods. The comparison indicates that the spatial interpolation exhibit values that are closer to actual data relative to the time series prediction. The spatial interpolation is closer to actual data because calculated value is the rainfall of the closest location or the neighbour location of sample value. The spatial connectivity between sample points serves as an indicator in determining the rainfall volume. We perform the time series prediction based on the value of historical data on the same points without the effects of neighbour’s value. Black solid line is spatial interpolation data, red dot line is actual data and blue solid line is time series analysis data. In the spatial interpolation analysis, some areas exhibit rainfall of 0 such as the sub districts of Selo, Teras, Simo, Juwangi, Sawit, and Wonosegoro while from the actual data these sub districts are known to exhibit high rainfall volume. The representation of 0 in the spatial interpolation is mainly caused by the unavailability of rainfall data at the closest sample point or by the distance that is too far so that the weight is smaller.
6. Conclusion and future work

The application architecture and analysis methodology “Multiplatform Architectural Spatiotemporal” and spatial data analysis methodology “Triple Exponential Smoothing” and “Spatial Interpolation” can be developed as a GIS application framework of rainfall distribution for various applied fields. The TES methods can analyze data with seasonal pattern that do not depend on the value of average level, and predicted and actual data show different repeating patterns. In the case of predicted data, the variance of monthly cycle tend to exhibit a similar pattern to previous years. Meanwhile, the actual data show an abnormal spike or anomaly and the repeating patterns follow previous data cycle. A comparison between the TES and IDW methods indicate that the spatial interpolation produces values that are closer to actual data compared to the time series prediction. The spatial interpolation produce results that are closer to actual data because calculated value is the rainfall volume of the closest location or of the sample value’s neighbour. The spatial connectivity between sample points is an indicator in determining the rainfall volume. Our time series prediction is based on the historical value data of same points without any impact from neighbour’s point. In IDW, some areas exhibit rainfall volume of 0, such as sub districts of Selo, Teras, Simo, Juwangi, Sawit, and Wonosegoro sub districts although the actual data show that these sub districts exhibit higher rainfall volume. The 0 representation in the spatial interpolation is mainly because of the absence of rainfall data in the nearest sample points or of the distance that is too far that produces too little weight.

Acknowledgements

This research was conducted using Featured Grant College Directorate General of Higher Education, Ministry of National Education. This research was conducted with the support of data from the Laboratory of Plant Disease Observations Region V Surakarta, Central Java BPOPT and Climatology Meteorology and Geophysics Agency of Central Java.
References

[1] Ji W and Cui J 2011 Application of Geographical Information System (GIS) in Agricultural Land classification and Grading International Conference on Agricultural and Natural Resources Engineering Advances in Biomedical Engineering 3 5 201-205

[2] Maracchia G, Perarnaud V and Kleschenko A D 2000 Applications of geographical information systems and remote sensing in agrometeorology Agricultural and Forest Meteorology 103 119–136

[3] Albites F A H 2008 Relationship of Cartography, Geographic Information Systems and The Mexicans Spatial Data Infrastructure The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 37 B4 39-44

[4] Anselin L 1993 Exploratory Spatial Data Analysis and Geographic Information System National Center for Geographic Information and Analysis University of California Santa Barbara

[5] Anselin L 1993 Local Indicator of Spatial Analysis National Center for Geographic Information and Analysis University of California Santa Barbara

[6] Anselin L 1994 Local Indicator of Spatial Analysis, Regional Research Institute West Virginia University Morgantown WV

[7] Anselin L 1995 Local Indicator of Spatial Association Geographical Analysis 27 2 93-115

[8] Anselin L 1996 The Moran Scatterplot as an ESDA Tool to Asses Local Instability in Spatial Association Spatial Analytical Perspective in GIS, Taylor and Francis London 111-125

[9] Anselin L 1998 GIS Research Infrastructure for Spatial Analysis of Real Estate Markets Journal of Housing Research 9 1 113-133

[10] Aziz F A A, Shariff A R M and Mohd M A 2008 GIS based System for Paddy Precision Farming, IAALD Afita WCCA Proc. World Conference on Agricultural Information and IT

[11] Malek I A A 1999 Conceptual Design and Implementation of Geographic Information System (GIS) for Hutang Simpan Ayer Hitam, Selangor, Malaysia Pertanika J.Trap.Agric.Sci. 22 2 171-177 1999

[12] Bindu B and Janak J 2012 GIS In Epidemiology: Application and Services National Journal of Community Medicine 3 2 259-263

[13] Mioc D 2008 Early Warning and Mapping for Flood Disasters The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 37 B4 1507-1509

[14] Qodriyatun S N 2013 Bencana Hidrometeorologi Dan Upaya Adaptasi Perubahan Iklim Info Singkat V 10 9-11

[15] Amri M R 2016 Tanah Longsor, Resiko Bencana Indonesia Jati Raditya 86-89

[16] Liu J K & Shih P T Y 2013 Topographic Correction of Wind-Driven Rainfall for Landslide Analysis in Central Taiwan with Validation from Aerial and Satellite Optical Images Remote Sens. 5 2571-2589

[17] Hong Y and Adler R F 2008 Predicting global landslide spatiotemporal distribution: Integrating landslide susceptibility zoning techniques and real-time satellite rainfall estimates International Journal of Sediment Research 23 249-257

[18] Ramasubramanian 1998 Time Series Analysis IASRI Package Avenue: New Delhi

[19] Cheng L C 2015 Preparing a landslide and shadow inventory map from high-spatial-resolution imagery facilitated by an expert system Journal of Applied Remote Sensing 9 1-13

[20] Roberts E A, Seley R L and Lawrence R L 2004 Using Sampling and Inverse Distance Weighted Modeling for Mapping Invasive Plants Westerns North American Naturalist 64 3 312-323

[21] Janert P K 2006 Exponential Smoothing, The Analysis of Time Series Chapman and Hall 6th ed.