Method Article

Singular Spectrum Analysis to Identify Excessive Rainfall

ABSTRACT

Indonesia is known for its excessive rainfall. Rainfall trends in an area have different characteristics. Differences in latitude, apparent motion of the sun, geographical position, topography, and the interaction of many forms of air circulation all contribute to this. Rainfall time series is essential for engineering planning, particularly for water infrastructure like irrigation, dams, urban drainage, ports, and wharves. Although meteorological technologies provide short-term rainfall predictions, long-term rainfall prediction is difficult and fraught with uncertainty. Unpredictability and seasonality can cause complex behavior in rainfall time series. This research utilizes the Singular Spectrum Analysis approach to extract trends; seasonality, cyclists, and noise can all be identified with potentially high accuracy.

Keywords: Singular Spectrum Analysis, Excessive Rainfall

1. INTRODUCTION

Medan is the capital and largest city of the Indonesian province of North Sumatra. North Sumatra is a region of the Indonesian Maritime Continent with substantial rainfall variability. Medan is highly strategic based on its geographical conditions, as indicated by the coordinates 1°–4° N and 98°–100° E. It is placed near the equatorial line, passed by the Bukit Barisanmountains, and flanked by the Malacca Strait and the Indian Ocean in [4]. This allows the rainfall environment to be influenced by global climate phenomena, and regional-scale climatic influences such as monsoons, tropical disturbances, and convergence zones are considered.

Medan City has recently suffered exceptional rain intensity, and a number of regions in the city are river basins that discharge into the sea. Excessive rainfall can cause flooding on many roadways and in residential areas. Rainfall patterns are classified in [15] as low (0–100 mm), medium (100–300 mm), high (300–500 mm), and extremely high (>500 mm). Excessive rainfall has the potential to generate floods, flash floods, landslides, and strong winds, especially in places prone to hydrometeorological disasters.

Identification is based on prediction results. Estimating the future value of past data requires high sensitivity and the ability to understand the characteristics of the data so that future probabilities can be estimated. Singular spectrum analysis is one method for predicting, belonging to the area of non-parametric methods that do not require initial data assumptions. It is possible to find and extract trends, seasonality, cyclists, and noise, which are potentially powerful forecasting techniques.

2. MATERIAL AND METHODS

2.1 Singular Spectrum Analysis

The basics of SSA have been explained in detail in [7]. The SSA approach has two complementary stages, which are the decomposition and reconstruction stages. The decomposition stage has two essential steps which are embedding and singular value decomposition. Embedding is the process of forming the original series into the path matrix; SVD decomposes the track matrix and divides the data into trend, seasonality, monthly components, and noise based on their single values. Then the reconstruction stage has two critical steps: creating groups, which involves decomposition of the path matrix, and diagonal averages to reconstruct a new time series.

Step 1. Decomposition
The concept of embedding is the path matrix, which transfers the one-dimensional time series of length N, into a sequence of L-dimensional vectors ($i = 1, K = N - L + 1$). The window length L, ($1 < L < N$) is the single parameter in this delay technique. The columns of the (LxK) trajectory matrix will be formed of the Kvectors Xi. The trajectory matrix X is a Hankel matrix with equal elements on the diagonals ($i + j = const$).

$$\mathrm{X} = \begin{bmatrix} x_1 & x_2 & x_3 & . & x_K \\ x_2 & x_3 & x_4 & . & x_{K+1} \\ . & . & . & . & . \\ x_L & x_{L+1} & x_{L+2} & . & x_N \end{bmatrix} \quad (2.1)$$

Singular Value Decomposition transforms the trajectory matrix into a sum of rank-one biorthogonal elementary matrices. From the $S = XXT$ matrix, the eigenvalues (λ_k) of S taken in decreasing order of magnitude $\lambda_k (1 \le k \le L)$and eigenvector ($U_k$)can be calculated and then sorted by the ortho-normal system of the eigenvectors of the matrix S corresponding to these eigenvalues. Then calculate the i-th element of the k principal components that will be formed $V_i = \mathrm{X}^T U_i / \sqrt{\lambda_i}$. Then the SVD of the path matrix can be calculated by $\mathrm{X}_i = \sqrt{\lambda_i} U_i V_i^T$. SVD can be represented as

$$\mathrm{X} = \mathrm{X}_1 + \mathrm{X}_2 + \cdots + \mathrm{X}_d \quad (2.2)$$

Step 2. Reconstruction
The concept of grouping in the reconstruction stage involves taking the results of path matrix decomposition and uniting those that are considered similar to form several groups. Then diagonal averaging is used to reconstruct a new time series by transforming each elementary matrix of the grouped decomposition. The diagonal averaging algorithm transforms Y into the reconstructed time series $y_1, \ldots, y_N$using the formula:

$$y_k = \begin{cases} \frac{1}{k} \sum_{m=1}^{k} a^*{}_{m,k-m+1}, 1 \le k < O \\ \frac{1}{O-1} \sum_{m=1}^{O-1} a^*{}_{m,k-m+1}, O \le k < P+1 \\ \frac{1}{N-k+1} \sum_{m=k-P+1}^{N-P+1} a^*{}_{m,k-m+1}, P+1 \le k < N \end{cases} \quad (2.3)$$

Another expansion by applying the Hankelization procedure to all matrix components can be written as:

$$\tilde{Y}^{(k)} = \left(\tilde{y}_1^{(k)}, \ldots, \tilde{y}_N^{(k)}\right) \quad (2.4)$$

2.2 SSA Forecasting

In this research using SSA recurrent forecasting. Thus M new data points will be determined to be forecasted, can be written as :$G_{N+M} = (g_1, g_2, \dots, g_{N+M})$

$$g_i \begin{cases} \widetilde{y_i} \ , & i = 0,1, \dots, N \\ \sum_{j=1}^{L-1} r_j \, g_{i-j} \ , & i = N+1, \dots, N+M \end{cases} \quad (2.5)$$

The concept of SSA recurrent forecasting is to create a model

$$y_{i+d} = \sum_{k=1}^{d} r_k \, y_{i+d-k} \, , 1 \le i \le N-d \quad (2.6)$$

To obtain a model for SSA recurrent forecasting, the LRF coefficient$(r_1, \dots, r_d)$ can be estimated from the eigenvector at the SVD.With $\mathrm{P} = (\mathrm{p}_1, \mathrm{p}_2, \dots, \mathrm{p}_{\mathrm{L}-1}, \mathrm{p}_{\mathrm{L}})$ and$\mathrm{P}^{\nabla} = (\mathrm{p}_1, \mathrm{p}_2, \dots, \mathrm{p}_{\mathrm{L}-1})$using the formula:

$$(r_{L-1}, \dots, r_1)^T = \frac{1}{1-v^2} \sum_{i=1}^{L-1} \pi_i \, \mathrm{P}_i^{\nabla}, v^2 = \sum_{i=1}^{L-1} \pi_i^2 \quad (2.7)$$

2.3 Forecasting Accuracy

2.3.1 *Mean Absolute Percentage Error (MAPE)*

Forecasting accuracy using MAPE to measure the accuracy of prediction results with actual data in the form of an average absolute error proportion and MAPE criteria explained in [5].

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \frac{|Y_t - \hat{Y}_t|}{Y_t} \times 100\% \quad (2.8)$$

With MAPE criteria If < 10% = highly accurate forecasting, 10–20% = good forecasting, 20–50% = reasonable forecasting, and > 50% = weak and inaccurate forecasting.

2.3.2 *Tracking Signal*

The tracking signal is used to determine the possibility of using the forecasting results if the pattern changes. If the tracking signal value is outside the acceptable limit (+/- 5), then the forecasting model must be reviewed, as explained in [2].

$$Tracking\ Signal = \frac{\sum_1^n e_n}{\sum_1^n \frac{|e_n|}{n}} \quad (2.9)$$

3. RESULTS AND DISCUSSION

3.1Singular Spectrum Analysis

The data utilized in this research are the rainfall time series of Medan City for the observation period of January 2017–December 2022 from [15].The plot shows that the data pattern identifies the influence of trend and seasonality over several time.

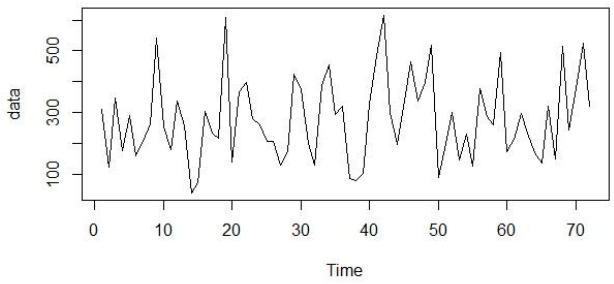

Fig 1. Plot of the Rainfall Time Series of Medan City

Step 1. Decomposition

First step in decomposition process is embedding by determining the window length. The data in this case is 60. The optimum L is determined by conducting experiments up to $L = 30$, then choosing the MAPE with the minimum value. It is obtained at a minimum of L = 13.

Table 1. Decomposition

L	13	16	20	22	30
MAPE	**17,33**	25,10	31,26	29,02	33,85

With L = 13 and K = 48,the Hankel matrix can be organized as shown below :

$$\mathrm{X} = \left(x_{ij}\right)_{i,j=1}^{13;48} = \begin{bmatrix} 312 & 124 & 347 & . & 397 \\ 124 & 347 & 175 & . & 519 \\ . & . & . & . & . \\ 258 & 40 & 74 & . & 174 \end{bmatrix}$$

From the Hankel matrix, the SVD step produces 13 eigentriples. Eigentriples consist of singular value ($\boldsymbol{\lambda_i}$),eigenvector(U_i), and principal component (V_i).

Table 2. Singular value ($\boldsymbol{\lambda_i}$) from X matrix

NO	$\boldsymbol{\lambda_i}$	$\sqrt{\boldsymbol{\lambda_i}}$
1	49480263.5	7034.22
2	1607298.0	1267.79
⋮	⋮	⋮
13	288936.9	537.53

Table 3. Eigenvector (U_i)from X matrix

NO	$\boldsymbol{U_1}$	$\boldsymbol{U_2}$	...	$\boldsymbol{U_{13}}$
1	-0.277	0.338	...	0.205
2	-0.281	0.236	...	-0.083
⋮	⋮	⋮	⋮	⋮
13	-0.278	0.283	...	-0.275

Table 4. Principal Component (V_i) fromX matrix

NO	V_1	V_2	...	V_{13}
1	-0.136	0.039	...	-0.044
2	-0.694	0.012	...	0.066
⋮	⋮	⋮	⋮	⋮
47	-1.244	0.067	...	0.043
48	-1.285	0.076	...	-0.016

Step 2. Reconstruction

The grouping step with a scree plot shows the estimation of the number of groups that will be formed.

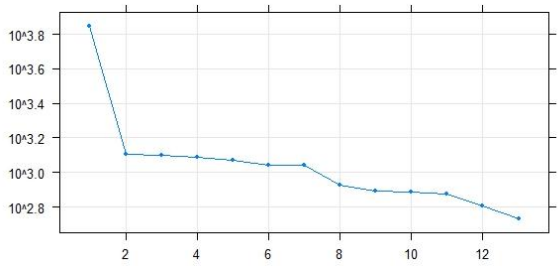

Fig 2. Scree Plot

The plot shows that there will probably be around 2 groups formed. To facilitate grouping by calculating the period value of each eigenvector. Adjacent periodicities are indicated to be in the same group. Several possible groups can be formed along with their respective MAPE values, as follows :

Table 5.MAPE values

No	Model	MAPE
1	Trend (1,2,11) Seasonal (4,6,13,8-10)	18,35
2	Trend (1,2) Seasonal (4,6,13,8-11)	18,58
3	Trend (1) Seasonal (2,4,6,13,8-11)	17,33

The diagonal averaging step in the reconstruction is to create a new series from groups that have been formed as below :

Table 6.Diagonal averaging step to create a new series

No	Actual	Reconstruction		Diagonal Averaging	Residual
		Trend	Seasonal		
1	312	265.33	8.81	274.14	37.86
2	124	256.62	-137.49	119.13	4.87
⋮	⋮	⋮	⋮	⋮	⋮
59	494	285.67	134.78	420,45	73.55
60	174	278.06	5.55	283.61	-109.61

3.2 SSA Forecasting

Forecasting model-based SSA recurrent forecasting using the LRF coefficient as follows:

$$g_i = \sum_{j=1}^{L-1} r_j g_{i-j} = 12{,}1135(g_{i-1}) + (-1{,}0817)(g_{i-2}) + \cdots + (-2{,}3728)(g_{i-12})$$

3.3 Forecasting Accuracy

Forecasting accuracy using MAPE and the tracking signal, it is known that the tracking signal values for the 12 forecasted time periods are within acceptable tolerance limits.

Table 7.Forecasting Accuracy

No	Outsample D_t	Forecast F_t	Tracking Signal	MAPE
61	213	379,88	-1,00	17,33%
62	296	382,19	-1,21	
63	232	256,50	-1,27	
64	169	366,41	-1,78	
65	135	217,03	-1,96	
66	319	259,62	-1,69	
67	150	249,78	-1,94	
68	514	177,40	-0,74	
69	243	275,75	-0,83	
70	376	226,24	-0,39	
71	526	257,47	0,32	
72	321	328,28	0,30	

The results of forecastingare as follows:

Table 8. Forecasting the rainfall time series of Medan City for January - December 2023

January	264.22	July	319.05
February	386.44	August	280.11
March	318.19	September	232.63
April	347.12	October	283.74
May	352.77	November	197.81
June	298.99	December	276.37

4. CONCLUSION

The results of the Medan City rainfall forecast for January–December 2023 show that almost average rainfall is in the high categorybecause it is in the range of 300–500 mm each month. This can be a reference in many matters, such as engineering planning, especially for water infrastructure such as irrigation, dams, and urban drainage. Besides that, this information can be an illustration for local governments in making policies.

REFERENCES

[1] A Shlemov,Golyandina N,D Holloway and, A Spirov. Shaped 3D Singular Spectrum Analysis for Quantifying Gene Expression, with Application to the Early ZebrafishEmbryo :Biomed Research International. 2015 : 2015 : 1-18.

[2] Abraham B, Ledolter J. Statistical Methods for Forecasting:John Willey and Sons; 1983.

[3] Bougas C. Forecasting air passenger traffic flows in Canada:An evaluation of time series models and combination methods. Quebec : Laval University; 2013.

[4] Budi Prasetyo, HendiIrwandi, Nikita Pusparini, Variable Topography-Based Rainfall Characteristic in North Sumatera, JurnalSains&TeknologiModifikasiCuaca.2018 :19(1) : 11 – 20.

[5] C. D. Lewis, Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting. Butterworth-Heinemann; 1982.

[6] Darmawan G, Rosadi D, Nurani B. Performance of Bootstrap Method in Singular Spectrum Analysis on Forecasting Rainfall Data, Proceedings oh The 2nd International Conference On Advance And Scientific Innovation (ICASI) ; 2019.

[7] Golyandina N and AZhigljavsky. Singular Spectrum Analysis for Time Series. SpringerHeidelberg; 2013.

[8] Golyandina N, V Nekritkuin, AZhigljavsky. Analysis of Time Series Structure,SSA and Related Techniques. Chapman & HALL /CRC; 2001.

[9] Hassani H, Heravi S, Brown G. Forecasting before, during, and after recession with singular spectrum analysis. Journal of Applied Statistics.2013 : 40(10):2290-2302

[10] Hassani H and Mahmoudvand R. Singular Spectrum Analysis Using R Macmillan Publishers Ltd. London;2018.

[11] Purnama, Eka. AplikasiMetode Singular Spectrum Analysis (SSA) PadaPeramalanCurahHujan Di ProvinsiGorontalo, (JAMBURA) Journal of Probability and Statistics. 2022 : 3(2).

[12] Sisti Nadia Amalia, ZulAmry, Forecasting of Air Passengers using Singular Spectrum Analysis, Scirea of Journal Mathematics. 2023 : 8(2) : 51-61.

[13] SM. Shaharudin, N. Ahmad, NH. Zainuddin, NS. Mohamed, Identification of rainfall patterns on hydrological simulation using robust principal component analysis, Indonesian Journal of Electrical Engineering and Computer Science (IJEECS). 2017 : 5(3) : 401-408.

[14] SM. Shaharudin, N. Ahmad, NH. Zainuddin, Modified singular spectrum analysis in identifying rainfall trend over peninsular Malaysia, Indonesian Journal of Electrical Engineering and Computer Science (IJEECS). 2019 : 15(1) : 283 - 293.

[15] www.bmkg.go.id