Journal of Intelligent Computing and Health Informatics

Integration of GSTAR-X and Uniform Location Weights for Forecasting Inflation Survey of Living Costs in Central Java

Research Article

Alwan Fadlurrohman

Department of Statistics, Universitas Muhammadiyah Semarang, Semarang 50185, Indonesia

alwanr21@gmail.com (coresponden author)

ABSTRACT

Inflation is a tendency to increase prices of goods and services that take place continuously. Inflation is a monthly time series data that is thought to be influenced by location elements. Modeling for inflation forecasting that involves time and location (spatio temporal) can use the Generalized Space Time Autoregressive (GSTAR) method. To increase accuracy in modeling and forecasting, the GSTAR model was developed into the GSTARX model by involving exogenous variables. Exogenous Variavel used in GSTARX modeling for forecasting Inflation is a variation of the Eid calendar. This GSTARX modeling is applied for inflation forecasting in six cities Cost of Living Survey (SBH) in Central Java, namely Cilacap, Purwokerto, Semarang, Kudus, Magelang and Surakarta. The purpose of this study is to get the best GSTARX model for inflation forecasting for six SBH cities in Central Java. The selection of the best model from the GSTARX method is seen with the smallest RMSE value of each model. Obtained that the GSTARX model with uniform weights is the best model because it has a smaller RMSE compared to the GSTARX model with inverse distance weights, the RMSE values are 0.6122 and 0.6137, respectively. It can be concluded that the GSTARX method with Uniform weighting can provide better performance and can be used to predict the inflation of the six SBH cities in Central Java in the next 12 periods.

Keywords : GSTAR, GSTARX, inflation, central java, cost of living survey

Integration of GSTARX and Uniform location weights for forecasting the Six Cities Inflation Survey of Living Costs in Central Java

pISSN: 2715-6923, eISSN: 2721-9186, Vol. 1, No. 1, March 2020, *pp.23-29*

21

1. INTRODUCTION

Inflation is a tendency to increase prices of goods and services that take place continuously. If the price of goods and services in the country increases, inflation will also increase. Rising prices of goods and services cause a decrease in the value of money. Thus, inflation also has the meaning as a decrease in the value of money on goods and services (BPS, 2008). Inflation in Central Java resulted from a Living Cost Survey (SBH) conducted by the Central Statistics Agency. The purpose of the SBH itself is to record household expenditure, both consumption and non-consumption, which aims to obtain a picture of people's consumption patterns that change over time. The value of household consumption as a result of SBH is used as a basis for selecting commodity packages and weighing diagram in inflation calculation. Out of 82 SBH cities, 6 of them are cities in Central Java. Therefore, inflation in Central Java Province is calculated based on the aggregation of 6 cities, namely Cilacap, Purwokerto, Kudus, Surakarta, Semarang and Tegal. The selection of sample cities is based on the level of development in the economic sector which is relatively fast when compared to other cities (BPS, 2008).

Research on inflation has been done using several methods, including Wavelet Tresholding (Ismiatul & Rosyida, 2019), ARIMA (Wei, 2013), and many more Inflation data is time series data, so that it can be modeled using time series analysis methods. In addition, inflation data for six SBH cities in Central Java also have inter-location links. The Generalized Space Time Autoregressive (GSTAR) model is a model that has a relationship between time and location where the location under study has non-uniform characteristics (heterogeneous) (Nurani, 2002) So that the inflation data of six SBH cities in Central Java can be analyzed using the GSTAR method.

The calendar variation model is a time series model that is used to forecast data based on seasonal patterns with varied periods (Borovkova et al., 2008). In most Islamic countries, monthly time series data in the field of economics and business can be identified by two types of calendar effects, namely the effect of the working day or trading day effect in each month, which is commonly referred to as the day trading effect and holiday effects such as year New China, Christmas, and Eid al-Fitr where the determination of the holiday is different from the calendar AD. Indonesia is one of the countries that feel the variation of the calendar, especially when entering the month of Ramadhan. When Ramadhan the level of consumption increases so the government and companies need to make a policy to keep the stock of goods in order to stay awake. One example of the data is about inflation. Analyzing such data using classical time series methods, such as ARIMA and GSTAR (Borovkova et al., 2008), may fail to capture the effects of these variations. In this

journal, a method is proposed for modeling time series and location data in the presence of calendar variations. The proposed method is a calendar variation model that was developed to obtain additional information, as well as to improve accuracy in modeling and modeling of time series data. So that the GSTAR model was developed by involving exogenous variables known as GSTARX modeling. The GSTARX model is a model that includes exogenous variables (X) in the model, so that it is not only influenced by the variable itself in the time period and also the location factor but is also influenced by the exogenous variable (X) (Borovkova et al., 2008).

The journal is organized as follows: Second part Literature Review, Third Part Research Methodology, Fourth Part Modeling of GSTARX in Forecasting Six Cities Inflation Cost of Living Survey in Central Java, The last part of the conclusion.

2. LITERATURE REVIEW

The GSTAR model was developed by involving exogenous variables known as GSTARX modeling. The GSTARX model is a model that involves exogenous variables (X) in the model, so that they are not only influenced by the variables themselves in a period of time and also by location factors but are also influenced by exogenous variables (X). In matrix notation, the GSTAR-X model (p; λ_1, λ_2, ..., λ_p) can be written as follows (Nurchayani, 2016):

$$Y(t) = \sum_{k=1}^{p} \sum_{l=0}^{\lambda_k} [\Phi_{kl} W^{(l)} Y(t-k)] + \Upsilon_{kl} X(t-s+1) + e(t) \quad (1)$$

The parameters of the GSTAR model can be estimated using OLS, the OLS method approach can also be used in estimating the GSTAR model involving exogenous variables (X) (Ruchjana, 2019).

2.1. Checking stationarity

A data is said to be stationary in time series data if the mean and variance values are constant or do not undergo systematic changes. Markidakis et al (1992) state that the visual form of a time series data plot is often sufficient to ensure that the data is stationary or not stationary. However, formally to identify the stationarity of the data is done by the Augmented Dickey Fuller (ADF) test or by looking at the MACF and MPACF cross correlation matrix schemes. If the MACF and MPACF plots fall slowly then the data is not stationary to the mean so it needs to be done differencing. Conversely, the data is not stationary with respect to variants if the upper and lower limits of the lamda are less than zero, so a Box Cox transformation is needed so that the data is stationary.

2.2. Orde selection on GSTARX model

According to Wutsqa (2010) in (Nurchayani, 2016)

states that the selection of the spatial order of the GSTAR-X model is generally limited to order 1, because a higher order will be difficult to interpret. Whereas the order of time (autoregressive) can be determined by AIC (Akaike Information Criterian). The selection of the best model order on GSTAR-X can be determined by looking at the smallest AIC value.

2.3. Weight selection of location in GSTARX model

In GSTAR-X modeling there is a problem that often occurs is the determination of location weights. Determination of location weights in the GSTAR-X model is divided into 3 main weights, namely uniform location weights, inverse distance weights and cross correlation normalization weights.

- *Uniform location weight*: (Suhartono & Subanar, 2006) defines the selection of uniform location weights as (Equation 2):

$$W_{ij} = \frac{1}{n_i} \qquad (2)$$

where n_i declare the states number of locations adjacent to location i in the spatial lag 1. The weight in this model has the properties as in Equation 3.

$$W_{ij} > 0, W_{ii} = 0, \sum_{j=1}^{N} W_{ij} = 1, \forall_i, \sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij} = N \qquad (3)$$

This location weight gives the same weight value at each location. Therefore, this location weight is often used on uniform data or has the same distance for each location. The weight of W_{ij} in lag 1 is expressed by W in the form of nxn matrix as follows Equation 4:

$$W = \begin{bmatrix} 0 & W_{12} & \dots & W_{1N} \\ W_{21} & 0 & \dots & W_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ W_{N1} & W_{N2} & \dots & 0 \end{bmatrix} \qquad (4)$$

- *Inverse distance weights:* the value of the weight of the inverse location distance is obtained based on the distance between the actual locations. The closer the location, the greater the weight of the location. Weighting with inverse distance refers to the distance between locations, for example the distance between 4 locations is defined:

r_1 = distance between location 1 and location 2
r_2 = distance between location 1 and location 3
r_3 = distance between location 1 and location 4
r_4 = distance between location 2 and location 3
r_5 = distance between location 2 and location 4
r_6 = distance between location 3 and location 4

written in the form of a matrix (see Equation 5):

$$\begin{bmatrix} 0 & \frac{r_2+r_3}{r_1+r_2+r_3} & \frac{r_1+r_3}{r_1+r_2+r_3} & \frac{r_1+r_2}{r_1+r_2+r_3} \\ \frac{r_4+r_5}{r_1+r_4+r_5} & 0 & \frac{r_1+r_5}{r_1+r_4+r_5} & \frac{r_1+r_4}{r_1+r_4+r_5} \\ \frac{r_4+r_6}{r_2+r_4+r_6} & \frac{r_2+r_6}{r_2+r_4+r_6} & 0 & \frac{r_2+r_4}{r_2+r_4+r_6} \\ \frac{r_5+r_6}{r_3+r_5+r_6} & \frac{r_6+r_3}{r_6+r_5+r_3} & \frac{r_5+r_3}{r_3+r_5+r_6} & 0 \end{bmatrix} = \qquad (5)$$

$$\begin{bmatrix} 0 & W_{12} & W_{13} & W_{14} \\ W_{21} & 0 & W_{23} & W_{24} \\ W_{31} & W_{32} & W_{33} & W_{34} \\ W_{41} & W_{42} & W_{43} & 0 \end{bmatrix}$$

2.4. Parameter estimation of GSTARX model

Basically, the GSTAR-X model is an extension of the GSTAR model, so the estimated parameters can be estimated using *Ordinary Least Square* (OLS). This approach can also be used in estimating the parameters of the GSTAR model involving exogenous variables (X) (Ruchjana, 2019). So, the form of the least-squares estimates $\hat{B}_T$ shown in Equation 6.

$$\hat{B}_T = (X'X)^{-1}X'Y \qquad (6)$$

2.5. Choose the best model

RMSE (Root Mean Square Error) is used to determine the best model in each model where the smallest RMSE value represents the best model. The RMSE is formulated as follows (see Equation 7):

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{M}\sum_{t=1}^{M}(Z_t - \hat{Z}_t)^2} \qquad (7)$$

M is the number of forecast data performed. Z_t is the actual data and forecast data. RMSE values range from 0 to ~. The smaller the RMSE value, the better the model used (Suhartono & Subanar, 2006).

3. METHODOLOGY

The data used in this study are secondary data sourced from central Java.bps.go.id, namely inflation data. The inflation data in question is the monthly inflation data in six cities Living Cost Survey in Central Java including Purwokerto, Cilacap, Tegal, Kudus, Surakarta and Semarang from January 2010 to November 2019 with a total of 119 data. Exogenous variables used are calendar variations, i.e. variations of the Eid calendar. Figure 1, shown the research steps.

3.1. Modeling six inflation city survey cost of life in central java using GSTARX method

3.1.1. Descriptive analysis

Integration of GSTARX and Uniform location weights for forecasting the Six Cities Inflation Survey of Living Costs in Central Java

pISSN: 2715-6923, eISSN: 2721-9186, Vol. 1, No. 1, March 2020, *pp.23-29*
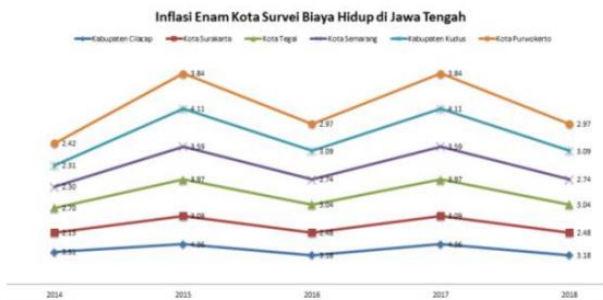
23

Fig. 2. Inflation of six cities in the cost of living survey in central java.

Figure 2 shows the inflation of six cities in the cost of living survey in Central Java is very volatile from 2014 to 2018. Where the average and standard deviation of monthly inflation for six cities living cost survey in Central Java is as follows (see Table 1):
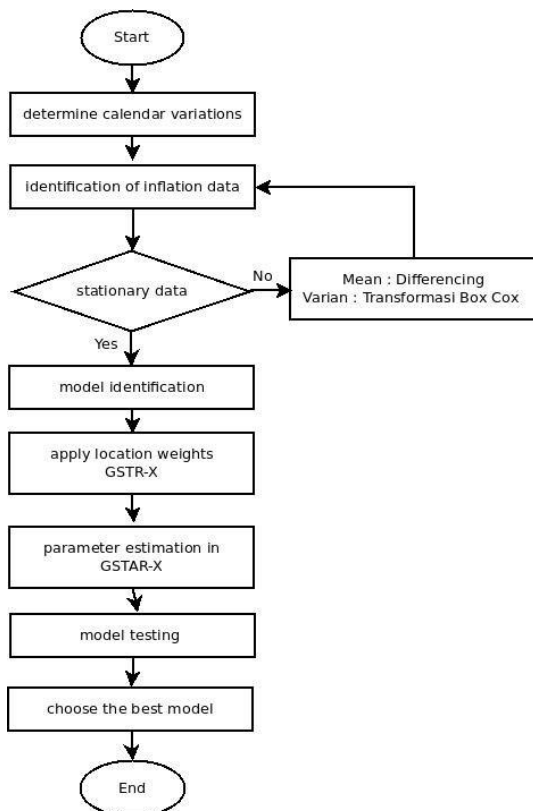

Fig. 1. Flowchart of the research steps.

Table 1. The caption must be shown before the table.

City SBF*	Average	standard deviation
Cilacap	0,50	0,71
Surakarta	0,38	0,74
Tegal	0,48	0,79
Semarang	0,43	0,67
Kudus	0,48	0,75
Purwokerto	0,43	0,63

The highest average inflation per month is in Cilacap and the lowest average inflation per month is Surakarta.

3.1.2. Data identification

Data identification is done to find out the data stationarity. This, to determine the stationarity of the data used the Augmented Dickey Fuller (ADF) test, the results of the ADF test are as follows (see Table 2):

Table 2. The results of ADF test.

City SBH*	ADF**	P-value
Cilacap	-4.3206	0,01
Surakarta	-4.9702	0,01
Tegal	-3.8813	0,02
Semarang	-4.9582	0,01
Kudus	-4.6557	0,01
Purwokerto	-4.6761	0,01

*SBH (survey biaya hidup), **Augmented Dickey Fuller

Based on the table above it can be seen that the six SBH cities have a p-value <0.05, so it can be said that the data does not contain a root unit or the data is stationary. So that it can be continued for the next stage.

3.1.3. Orde determination of GSTARX Model

After the stationary data, the next step is to determine the time P9 and the orde spatial. Where the orde spatial that can be used is orde spatial 1, because if the orde spatial is higher, then the model is difficult to interpret. While for orde time, it is done through the Vector Autoregressive (VAR) approach by looking at the smallest Akaike's Information Criterion (AIC) value.

Table 3. Akaike's information criterion (AIC) value.

Minium Information Criterion					
Lag	MA 0	MA 1	MA 2	MA 3	MA 4	MA
AR 0	-7,532197	-6,903041	-6,737091	6,278608	-5,963418	-5,49322
AR 1	-9,141695	-8,741586	8,507667	8,177402	-7,747075	-7,0439
AR 2	-10,266 78	-9,861451	9,496847	8,774765	-8,250094	7,83849
AR 3	-10,937 78	-10,52448	9,977777	10,00236	-9,499874	8,50642
AR 4	-11,080 58	-10,97784	10,43681	-9,81082	-8,35156	7,10462
AR 5	-10,896 47	-10,90234	10,42551	9,233712	7,322102	4,72319
AR 6	-10,861 48	-9,006936	8,042073	6,187855	3,388905	0,507563
AR 7	-10,215 94	-7,890389	4,971079	2,227964	2,013691	8,205435
AR 8	-9,0208 34	-6,0108	2,097757	3,196360 9	9,983926 1	20,6110 8
AR 9	-6,4557 63	-2,394224	3,144237 4	11,14423 7	23,71566 6	45,2336 6

Based on the smallest AIC value in the Table 3 , it is at AR (4) and MA (0) with previously known $\lambda p = 1$, so that the GSTARX model formed is GSTARX $(1; 4)$.

3.1.4. Weighting

- Uniform Location Weight: The uniform location weights in GSTARX modeling assume that the inflation data of six SBH cities in Central Java at one location has the same effect on inflation at other locations. The uniform weight matrix used in this study is as follows:

$$W_{ij} = \begin{bmatrix} 0 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.2 & 0 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0 & 0.2 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0 \end{bmatrix} \quad (8)$$

- Weight of Inverse Distance Location: The distance inverse location weights in GSTARX modeling assumes that inflation in one location is influenced by the distance or proximity of the distance that is owned by other locations. The distance between two distant locations will have a weight that tends to be lower than the distance between two close locations. Below is the distance (km) between the six SBH cities in Central Java (see Table 4).

Table 4. Distance (km) between six SBH cities in Central Java.

	1	2	3	4	5	6
1	0	264	186	282	302	61
2	264	0	267	102	153	224
3	186	267	0	165	216	114
4	282	102	165	0	51	221
5	302	153	216	51	0	262
6	61	224	114	221	262	0

Note:
1 = Cilacap, 2 = Surakarta, 3 = Tegal, 4 = Semarang, 5 = Kudus, and 6 = Purwokerto

The distance between the six SBH cities in Central Java normalized the inverse distance value, then used as a weighting matrix. The inverse distance weighting matrix based on the above table is as follows:

$$W_{ij} = \begin{bmatrix} 0 & 0.12 & 0.17 & 0.11 & 0.10 & 0.51 \\ 0.13 & 0 & 0.13 & 0.35 & 0.23 & 0.16 \\ 0.19 & 0.13 & 0 & 0.21 & 0.16 & 0.31 \\ 0.08 & 0.23 & 0.14 & 0 & 0.45 & 0.10 \\ 0.09 & 0.17 & 0.12 & 0.52 & 0 & 0.10 \\ 0.43 & 0.12 & 0.23 & 0.12 & 0.10 & 0 \end{bmatrix} \quad (9)$$

3.1.5. GSTARX model estimation

In the discussion above it has been found that the orde time of GSTARX is 4. For orde spatial is limited to only one orde, so the GSTARX model used is the GSTARX model (1; 4). Then parameter estimation is performed in the GSTARX model (1; 4). At the parameter estimation stage, the estimated values of the model parameters are obtained through several estimation methods, including the Ordinary least square method. The estimated results of each weighting are as follows:

- GSTARX Model Estimates of uniform location weights (see Table 5)

Table 5. Distance (km) between six SBH cities in Central Java.

Variable		Estimate	P-value	Information
Intercept		0,181	0,000	
Cilacap	$\emptyset^1_{10}$	0,057	0,781	No*
	$\emptyset^1_{11}$	0,464	0,036	Yes**
Surakarta	$\emptyset^2_{10}$	-0,057	0,756	No*
	$\emptyset^2_{11}$	0.482	0.009	Yes**
Tegal	$\emptyset^3_{10}$	0.681	0,000	Yes**

	$\emptyset^3_{11}$	-0.153	0,361	No*
Semarang	$\emptyset^4_{10}$	0.063	0.831	No*
	$\emptyset^4_{11}$	0.361	0.211	No*
Kudus	$\emptyset^5_{10}$	-0.135	0.635	No*
	$\emptyset^5_{11}$	0.710	0.026	Yes**
Purwokerto	$\emptyset^5_{10}$	-0.075	0.776	No*
	$\emptyset^5_{11}$	0.496	0.041	Yes**
Exogenous Variables		0,564	0,000	Yes**

*No Significant, **Significant

- GSTARX Model Estimation of inverse distance location weights (see Table 6)

Table 6. GSTARX Model Estimation of inverse distance location weights.

Variable		Estimate	P-value	Information
Intercept		0,183	0,000	
Cilacap	$\emptyset^1_{10}$	0,032	0,877	No*
	$\emptyset^1_{11}$	0,504	0,029	Yes**
Surakarta	$\emptyset^2_{10}$	-0,084	0.665	No*
	$\emptyset^2_{11}$	0.508	0.019	Yes**
Tegal	$\emptyset^3_{10}$	0.684	0,000	Yes**
	$\emptyset^3_{11}$	-0.159	0.357	No*
Semarang	$\emptyset^4_{10}$	0.151	0.622	No*
	$\emptyset^4_{11}$	0.265	0.365	No*
Kudus	$\emptyset^5_{10}$	0.075	0.789	No*
	$\emptyset^5_{11}$	0.471	0.136	Yes*
Purwokerto	$\emptyset^5_{10}$	-0.065	0.802	No*
	$\emptyset^5_{11}$	0.481	0.043	Yes**
Exogenous Variables		0.561	0.000	Yes**

*No Significant, **Significant

3.1.6. Choose the best model

After obtaining a model from each weighting, the selection of the best model of the model is done by looking at the value of the RMSE. Where you get the respective RMSE values:

Table 7. Result of model comparison.

Model	RMSE
GSTARX + Uniform location weights	0,6122
GSTARX + Weight of inverse distance location	0,6137

Table 7, shows that the best model of GSTARX is to use uniform location weights with an RMSE value of 0.6122.

3.1.7. Inflation prediction

Table 8, shows the forecasting results of the best models.

Table 8. Forcasting results of the best model..

City	A	B	City	A	B
Cilacap	1	0.68	Semarang	1	1.57
	2	0,65		2	0.97
	3	0,21		3	0,97
	4	0,19		4	1,08

Integration of GSTARX and Uniform location weights for forecasting the Six Cities Inflation Survey of Living Costs in Central Java

pISSN: 2715-6923, eISSN: 2721-9186, Vol. 1, No. 1, March 2020, *pp.23-29*

25

City	A	B	City	A	B
	5	0,33		5	1,48
	6	0,60		6	0,41
	7	0,48		7	0,11
	8	0,53		8	0,31
	9	0,00		9	0,06
	10	0,10		10	0,25
	11	0,56		11	0,37
	12	1,61		12	0,65
Surakarta	1	0.65	Kudus	1	O,63
	2	0,51		2	0,21
	3	0,22		3	0,22
	4	0,28		4	0,39
	5	0,36		5	0,23
	6	0,65		6	0,72
	7	0,34		7	0,42
	8	0,26		8	0,13
	9	0,09		9	0,24
	10	0,29		10	0,43
	11	0,50		11	0,47
	12	0,53		12	0,71
Tegal	1	0,97	Purwokerto	1	0,00
	2	0,27		2	0,26
	3	0,41		3	0,17
	4	0,74		4	0,30
	5	1,08		5	0,48
	6	0,05		6	0,57
	7	-0,18		7	0,36
	8	-0,09		8	0,21
	9	-0,04		9	0,17
	10	0,18		10	0,21
	11	0,45		11	0,46
	12	0,85		12	0,79

4. CONCLUSION

Based on the results of the above analysis it can be concluded that the inflation of six SBH cities in Central Java Province has the GSTARX Model (1; 4) which is built from both location weights which are uniform and inverse distance. The best model is to get a uniform weight.

REFERENCES

Borovkova, S., Lopuhaä, H. P., & Ruchjana, B. N. (2008). Consistency and asymptotic normality of least squares estimators in generalized STAR models. *Statistica Neerlandica*, *62*(4), 482–508. https://doi.org/10.1111/j.1467-9574.2008.00391.x

BPS. (2008). PERKEMBANGAN INDEKS HARGA KONSUMEN/INFLASI. *BPS: Berita Resmi Statistik*, *September*, 1–9.

Ismiatul, S., & Rosyida, I. (2019). Implementasi Model Fuzzy-Wavelet dan FIS Metode Mamdani dalam Prediksi Nilai Tukar EUR / IDR. *PRISMA, Prosiding Seminar Nasional Matematika*, *2*, 313–322.

Nurani, B. (2002). Pemodelan Kurva Produksi Minyak Bumi Menggunakan Model Generalisasi Star. *Forum Statistika Dan Komputasi*, *September*(September), 1–6.

Nurchayani, F. (2016). *Pengelompokan Stasiun Hujan untuk Model Generalized Space Time Autoregressive (GSTAR) pada Peramalan Curah Hujan Kabupaten Jember dengan Tiga Pembobotan*. Universitas Jember.

Ruchjana, B. N. (2019). Pengembangan Model Spatio Temporal. *Prosiding Seminar Nasional Matematika Dan Pendidikan Matematika 2019*, 1–19.

Suhartono, & Subanar. (2006). The Optimal Determination of Space Weight in GSTAR Model Using Cross-Correlation Inference. *Journal of Quantitative Methods*, *2*(2), 45–53.

Wei, W. W. . (2013). Time Series Analysis. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods.Volume 2: Statistical Analysis* (2nd ed., pp. 458–485). Oxford University Press.