Journal of Quality Measurement and Analysis *JQMA* 19(3) 2023, 143-153
e-ISSN: 2600-8602 *http://www.ukm.my/jqma*

GENERALIZED SPACE-TIME AUTOREGRESSIVE (GSTAR) FOR FORECASTING AIR POLLUTANT INDEX IN SELANGOR

(*Autoregressive Ruang Masa Teritlak (GSTAR) untuk Meramal Indeks Pencemaran Udara di Selangor*)

NUR MAISARA MOHAMED, NUR HAIZUM ABD RAHMAN* & HANI SYAHIDA ZULKAFLI

ABSTRACT

This study presents the Generalized Space-Time Autoregressive (GSTAR) model, a multivariate time series approach that integrates spatial and temporal observations for data forecasting. This study's primary objective is to develop and apply the GSTAR model to forecast the Air Pollutant Index (API), which exhibits spatial-temporal dependencies between locations and time. Three areas in Selangor have been used in this study: Banting, Petaling, and Shah Alam. The model employs uniform and inverse distance weights to consider spatial relationships. The forecasting performance is assessed using Root Mean Square Error (RMSE). Although both weight methods yield comparable results, the GSTAR model with inverse distance weight is promising for API data forecasting with consistently low RMSE values. The result of this study emphasises the significance of location-based information in generating more efficient and informed solutions.

Keywords: GSTAR; forecasting; uniform weight; inverse distance weight; Air Pollutant Index

ABSTRAK

Kajian ini membentangkan model Generalized Space-Time Autoregressive (GSTAR), pendekatan siri masa berbilang yang mengintegrasikan pemerhatian spatial dan temporal untuk peramalan data. Objektif utama kajian ini adalah untuk membangunkan dan menggunakan model GSTAR untuk meramalkan Indeks Pencemaran Udara (IPU), yang mempamerkan kebergantungan spatial-temporal antara lokasi dan masa. Tiga lokasi di Selangor telah digunakan dalam kajian ini: Banting, Petaling, dan Shah Alam. Model ini menggunakan pemberat seragam dan jarak songsang untuk mempertimbangkan perhubungan spatial. Prestasi ramalan dinilai menggunakan Ralat Min Kuasa Dua (RMSE). Walaupun kedua-dua kaedah berat menghasilkan hasil yang setanding, model GSTAR dengan berat jarak songsang menjanjikan peramalan data IPU dengan nilai RMSE rendah secara konsisten. Hasil kajian ini menekankan kepentingan maklumat berasaskan lokasi dalam menjana penyelesaian yang lebih cekap dan bermaklumat.

Kata kunci: GSTAR; ramalan; pemberat seragam; pemberat jarak songsang; Indeks Pencemaran Udara (IPU)

1. Introduction

Time series data find applications in various disciplines, such as meteorology for weather forecasting, medicine for monitoring patient progress over time, and finance for studying fluctuations. The increasing complexities and challenges related to spatial-time issues have led to various observations and statistical concepts applied as solutions. Continuous advancements in computing technology and the availability of massive databases have facilitated extensive research in developing and improving time series models driven by the escalating volume of time series data.

Notably, it was only in the mid-1970s that researchers began examining statistical and economic models that traced the spatial-temporal development of single or multi-variable relationships through time. Over the next two decades, this field experienced a significant surge in development and interest (Kamarianakis & Prastacos 2006). Time series data modelling comprises both univariate and multivariate data approaches. The Autoregressive Integrated Moving Average (ARIMA) model is a common univariate time series modelling example. On the other hand, multivariate time series modelling involves techniques such as the Vector Autoregressive Integrated Moving Average (VARIMA) model.

The space-time model is a functional multivariate time series approach that simultaneously combines time and spatial observations (Atluri *et al.* 2018; Koutsaki *et al.* 2023). In spatial-temporal problems, the data is multidimensional and correlated with past events and specific locations or regions (Akbar *et al.* 2020). The Space-Time Autoregressive (STAR) and Space-Time Autoregressive Moving Average (STARMA) models, initially introduced by Cliff and Ord (1975), serve as prominent examples of models used for modelling and forecasting spatial-temporal time series data. STARMA represents an improvement over the Vector Autoregressive Moving Average (VARMA) model by reducing the required parameters (Suhartono *et al.* 2016). This model includes a spatial lag operator that signifies how neighbouring locations influence a specific spatial point through weightings (Munandar *et al.* 2023).

While STARIMA models are adequate for large-scale geographical applications, they may need to be more complex for small-scale spatial time series research (Kamarianakis & Prastacos 2006). The STAR model demonstrates effectiveness in various fields; however, its lack of flexibility becomes apparent when encountering locations with distinct characteristics (Monika *et al.* 2023). Additionally, the assumption of constant autoregressive parameters across all locations becomes impractical since different locations typically yield different parameters. To address these limitations, Borovkova *et al.* (2008) and Ruchjana *et al.* (2012) conducted further studies and proposed an enhanced model called the Generalized Space-Time Autoregressive (GSTAR) model. The GSTAR model is a natural generalisation of STAR models, allowing its application in samples with varying characteristics, as it accommodates varying autoregressive parameters for each location (Ruchjana *et al.* 2012).

Numerous research studies have been conducted concerning GSTAR modelling since its introduction. Anggraeni *et al.* (2018) utilised the GSTAR model in their research to explore spatial effects between stations and their potential for forecasting future rainfall. Their study revealed that the aggregate stacking models of Autoregressive Moving Average (ARMA) and GSTAR outperformed individual models and ensemble averaging of ARMA and GSTAR in all clusters. However, the non-seasonal ARMA model outperformed GSTAR in most clusters, suggesting that seasonal models might not be relevant in the context of recent unprecedented climatic shifts.

Zewdie *et al.* (2018) focused on forecasting temperatures in Northern Ethiopia using the GSTAR model in a separate study. They demonstrated that due to parameter variations within the area, the GSTAR techniques play a crucial role in enhancing prediction reliability, as measured by the root-mean-squared error function (RMSEF). On the other hand, Vector Autoregressive (VAR) models, which assume no restrictions on parameters and do not utilise weighted matrices to calculate location impact, fall short compared to the GSTAR model. While STAR models outperform VAR, they make the incorrect assumption of equal parameters across all locations, leading to a higher RMSEF and less reliable forecasts. It is worth noting that one weakness of their study is the lack of consideration for seasonal impacts or informal statistical tests.

Air quality data exhibit spatial-temporal dependencies between locations and time that rapidly change. The parameter used to describe air quality status is the Air Pollutant Index

(API). In Malaysia, the API framework draws inspiration from the one introduced by the United States Environmental Protection Agency (USEPA). This index is computed through the aggregation of sub-index values corresponding to five primary pollutants: ozone (O3), carbon dioxide (CO2), particulate matter (PM10), sulfur dioxide (SO2), and nitrogen dioxide (NO2). The API values are categorised into distinct ranges, namely ‘‘Good (0-50), Moderate (51-100), Unhealthy (101-200), Very Unhealthy (201-300), and Hazardous (301 and above)” (Leh *et al.* 2012). Severe deterioration of air quality globally has been caused by rapid urbanisation, leading to increased hospitalisation and premature deaths. Therefore, daily API and forecasting air quality can help the general public to take action to protect themselves from polluted air and advance air pollution management by the government.

Air pollution can have adverse effects on various population groups, particularly those who are more susceptible due to age, pre-existing health conditions, or socioeconomic factors. Vulnerable populations include children since their developing respiratory systems are more sensitive to pollutants, and exposure can lead to long-term health problems. Meanwhile, elderly individuals or older adults may have weaker immune systems and pre-existing health conditions that can be exacerbated by exposure to air pollutants. The case study used in this research focused on forecasting the API data using GSTAR in three districts in Selangor, Malaysia: Banting, Petaling, and Shah Alam. Air is necessary for all life on Earth to survive; hence, studying air quality is crucial as it significantly impacts the country's economic growth and the population's health (Manisalidis *et al.* 2020).

2. Methodology

2.1 *Data description*

The Air Pollutant Index (API) data used in this study was obtained from the Department of Environment (DOE) Malaysia. Collected hourly from January 1st, 2018, to December 31st, 2018, the data examined the relative changes in air quality status. To address missing data, a simple imputation technique was applied, replacing the gaps with the average value of the dataset. The data was then divided into two parts: in-sample data, spanning from January 1st, 2018, to December 17th, 2018, and out-sample data, covering the period from December 18th, 2018, to December 31st, 2018. For this study, three air quality monitoring stations were selected in Selangor, Malaysia, specifically in Banting, Petaling, and Shah Alam districts.

2.2 *Stationary test*

Many time series approaches assume that the data is stationary. Time series analysis needs stationarity because forecasting is made possible by a predictable distribution. The nonstationary process can be made stationary by detrending or differencing if the stationarity assumption deviates (Ng’ang’a & Oleche 2022). David A. Dickey and Wayne A. Fuller introduced the DF test in 1979, and it was extended to the Augmented Dickey-Fuller (ADF) test in 1984 by Said and Dickey. The error term in the DF test is assumed to be uncorrelated, while the ADF test allows for correlated error terms as more lagged terms will be added until the error terms are uncorrelated (Gujarati & Porter 2009). Since the ADF test can examine a more extensive and complex set of time series models, this study implements the ADF test to check for the stationarity of data rather than the DF test. The ADF test is expressed as in Eq. (1).

$$\Delta Y_t = \delta_0 + \delta_1 t + \lambda Y_{t-1} + \sum_{j=1}^{p-1} \rho_j \Delta Y_{t-j} + u_t \tag{1}$$

The hypothesis is given by, $H_0: \lambda = 0$ (The series is non-stationary) meanwhile $H_1: \lambda < 0$ (The series is stationary). The test statistics are expressed as follows:

$$\text{Test statistics} = \frac{\hat{\lambda}}{se(\hat{\lambda})} \tag{2}$$

According to Marsani and Shabri (2020), the ADF statistic is a negative value at which a more significant negative value than the critical value indicates a more decisive rejection of the null hypothesis. In other words, if the p-value is lesser than the level of significance, we reject the null hypothesis and can conclude that the series is stationary. Further examination of the data can be carried out.

2.3 *Generalized Space-Time Autoregressive (GSTAR) model*

The STAR models can be naturally generalized into what is known as a Generalized STAR (GSTAR) model, which allows the autoregressive parameters to vary depending on the location. Let $Y_i(t) = (Y_1(t), Y_2(t), \dots, Y_N(t))'$ at location $i = 1,2,\dots,N$ and time $t = 1,2,\dots,T$ follows GSTAR $(p;\ \lambda_1, \lambda_2, \dots, \lambda_p)$ model with time order p and spatial $\lambda_1, \lambda_2, \dots, \lambda_p$ that can be written as follows:

$$Y(t) = \sum_{s=1}^{p} \left(\phi_{s0} Y(t-s) + \sum_{k=1}^{\lambda s} \phi_{sk} W_{ij}^{k} Y(t-s) \right) + \varepsilon(t) \tag{3}$$

where $Y(t-s)$ is the observed value at time lag s, s is time autoregressive order, k is spatial autoregressive order, p is the time order of $p{-}th$ autoregressive term, λs is the spatial order of $s{-}th$ autoregressive term, $W_{ij}(k)$ represents the weight of $k{-}th$ order spatial, ϕ_{s0} is the diagonal matrices with the diagonal elements as autoregressive of lag time for each location, and ϕ_{sk} is the diagonal matrices. The diagonal elements are space-time parameters in lag spatial and time lag, and $\varepsilon(t)$ is the white noise.

The process of identifying the optimal GSTAR model involved the utilisation of the Box-Jenkins methodology. This methodology follows a three-step iterative procedure for constructing the ARIMA model, encompassing the stages of model identification, parameter estimation, and diagnostic evaluation (Adhikari & Agrawal 2013). When working with a specific time series dataset, developing a well-suited model that aligns with the principle of parsimony becomes crucial. This principle emphasises the creation of a model with the fewest necessary parameters to effectively capture and represent the time series data with accuracy and sufficiency. With that, the three steps are necessary in order to select a satisfactory model that can be utilised to forecast future values.

2.4 *Spatial weights*

The number of surrounding observed sites in spatial order influences spatial weight. This study utilised two types of spatial weight: uniform weight and inverse distance weight. Uniform weight assigns an equal weight value to each site and can be calculated using the following formula:

$$W_{ij} = \frac{1}{N_i^{(s)}} \tag{4}$$

where W_{ij} is the weight between i and j. N_i is the number of neighbours site with a site. The second spatial weight, inverse distance weight, calculates the distance between locations. The locations used in this research are latitude and longitude. The distance between locations is defined as follows:

$$W_{ij} = \frac{1/d_{ij}}{\sum_{i \neq j} 1/d_{ij}} \tag{5}$$

where d_{ij} is calculated using the Euclidean distance between the location i and j.

$$d_{ij} = \sqrt{(u_i - u_j)^2 - (v_i - v_j)^2} \tag{6}$$

where u and v represent the latitude and longitude coordinate location, respectively.

2.6 *Accuracy measurement*

The accuracy measure of forecasting models serves as a basis for identifying the most suitable models for making forecasts. By employing the Root Mean Square Error (RMSE), the best model could be identified with the lowest RMSE. To calculate RMSE, the following formula is used:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{7}$$

where i is the period of time, n is the total number of observations, y_i is the actual value and $\hat{y}_i$ is the forecast value at time i.

3. Result and Discussion

The daily Air Pollutant Index (API) data in three monitoring stations, Banting, Petaling and Shah Alam, were plotted in Figure 1. Based on Figure 1, the API in Banting experiences rapid fluctuations, with a notable increase from January 2018. The increase rate slows between March 2018 and September 2018, with a peak API reading exceeding 100 during this period. Subsequently, there was a sharp decline in API until December 2018. Shah Alam's API gradually increased from January 2018, followed by fluctuations until May 2018. From May to July 2018, there was a decline in API, which then rose again from August 2018 to October 2018 before decreasing until December 2018. API in Petaling is more stable and does not show much variation compared to Banting and Shah Alam. Interestingly, Figure 1 illustrates that all three locations show a similar increasing pattern in API around October 2018, followed by a subsequent decrease.

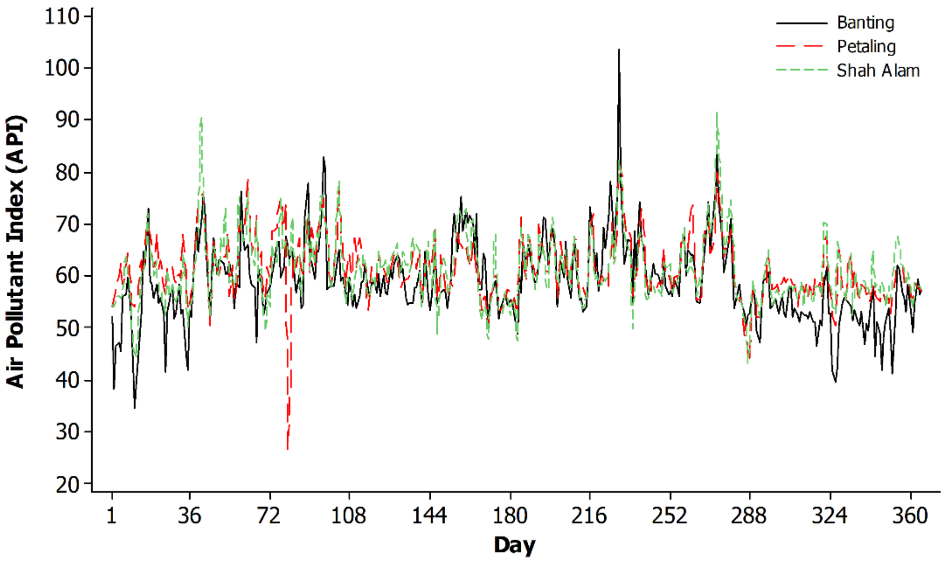


Figure 1: Time series plot of air pollutant data

Table 1: Descriptive statistic for Air Pollutant Index in Banting, Petaling, and Shah Alam

Location	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Banting	59.004	7.878	34.708	54.250	58.042	63.208	103.500
Petaling	61.372	6.254	26.286	57.333	60.417	65.229	81.833
Shah Alam	61.268	6.548	43.333	56.813	60.417	64.958	91.458

Table 1 presents a comprehensive overview of the descriptive statistics for the Air Pollutant Index (API) data in all research locations, encompassing minimum, first quartile (Q1), median, mean, standard deviation (StDev), third quartile (Q3), and maximum values. The API varied from 34.71 (minimum) to 103.5 (maximum) in the Banting area. For Petaling, the API ranged from a minimum of 26.29 to a maximum of 81.83. Similarly, Shah Alam exhibited API data with a minimum reading of 43.33 and a maximum reading of 91.46. Notably, the mean API value in Banting was the lowest, measuring 59, whereas the mean values for the other locations exceeded 60. These findings indicate that the API readings at all three locations fall within the Moderate level. Within this range, vulnerable individuals appear to be most at risk, as studies have linked air pollution to reduced cognitive performance among the elderly, while others suggest that poor air quality poses particular dangers to children.

To begin model building, the API data need to be in a stationary form. The stationarity of the data can be confirmed by examining the consistency of their variance and mean to determine if they significantly affect the series' behaviour. In Table 2, the lambda values of the API data are summarised both before and after transformation using the Box-Cox transformation. Before the transformation, the lambda value indicates that only Petaling data exhibits stationarity in variance, while the data for Banting and Shah Alam do not. Stationarity in variance ensures that the statistical properties of the data remain relatively constant over time. If the variance is not constant, the model parameters may be unstable, making it challenging to estimate and interpret the model accurately. This can lead to unreliable forecasts. Consequently, the data for Banting and Shah Alam were transformed using the natural logarithm (ln) and reciprocal (1/y) transformations, respectively. Subsequently, the stationarity of the transformed series was reevaluated, and now both series show stationarity in variance as the lambda values are equal to 1. This confirms the successful transformation of the data, rendering them suitable for further analysis and model building.

Table 2: The results of lambda value before and after transformation

Location	Before	After	
	λ	Transformation	λ
Banting	0.0	ln y	1.0
Petaling	1.0	No transformation	1.0
Shah Alam	-1.0	1/y	1.0

Once the data has been transformed, it is necessary to assess its stationarity in terms of the mean. A formal method to test for data stationarity in the mean is through the Augmented Dickey-Fuller (ADF) test. Based on the ADF test results in Table 3, the *p*-values for all locations are greater than α=0.01. Consequently, H_0 (null hypothesis) cannot be rejected. This leads to the conclusion that the time series data is non-stationary, indicating the presence of a unit root in the data. Therefore, differencing is necessary since the data is non-stationary in the mean.

Table 3: The results of the Augmented Dickey-Fuller test for transformed API data

Location	ADF	*p*-value
Banting	-3.5517	0.03791
Petaling	-5.6219	0.0100
Shah Alam	-5.1642	0.0100

The Box-Jenkins methodology uses stationary data to initiate model identification. In the GSTAR model, time and spatial lag orders are identified using Space-time autocorrelation functions (STACF) and Space-time partial autocorrelation functions (STPACF), as illustrated in Figure 2, depicting a "dies down" pattern. These autocorrelation functions are instrumental in determining the order of the GSTAR model, which is a particular case of the STAR model (Nurhayati *et al*. 2012). Therefore, the model shown in this context is identified as GSTAR(1,1) and the parameter estimations are given in Table 4.

Table 4: The parameter estimation of GSTAR(1,1) model

Location	Parameter	Coefficients	
		Uniform	Inverse
Banting	${\emptyset_{10}}^1$	-0.1802	-0.1802
	${\emptyset_{11}}^1$	0.0012	0.0021
Petaling	${\emptyset_{10}}^2$	-0.1648	-0.1641
	${\emptyset_{11}}^2$	14.2979	15.8354
Shah Alam	${\emptyset_{10}}^3$	-0.1535	-0.1530
	${\emptyset_{11}}^3$	4.4212×10−6	3.3838×10−6

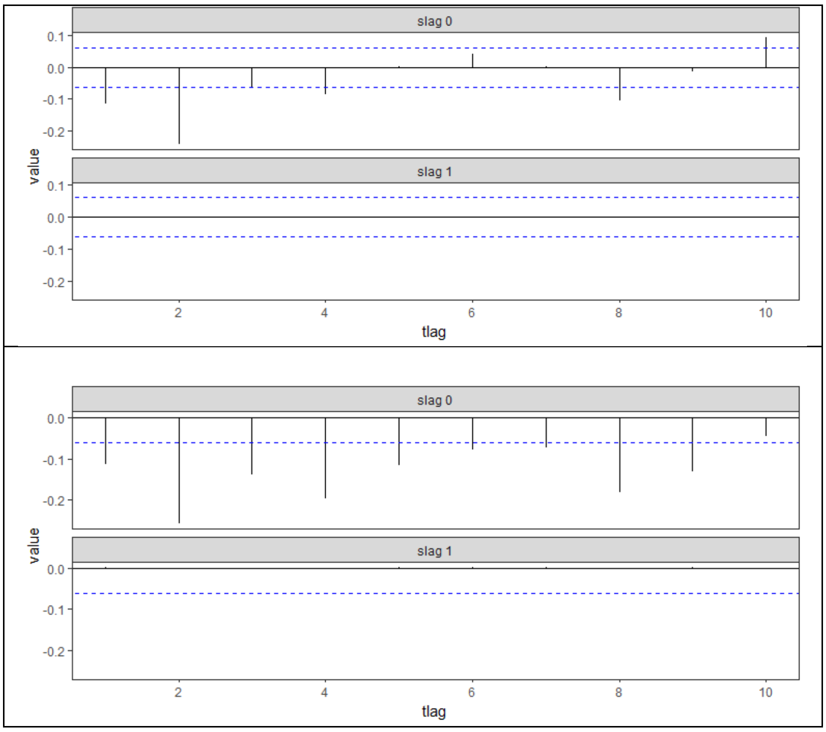


Figure 2: STACF and STPACF for stationary data

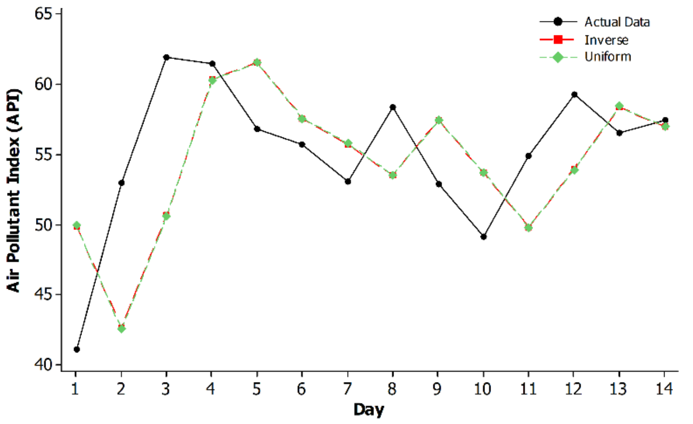
Before the forecasting stage, the developed model needs to be validated using the Ljung-Box test, and the results are presented in Table 5. Based on the Ljung-Box test results, all the *p*-values of the variables are more than a significant level of 0.01. As a result, we can infer that the errors are random and exhibit white noise. This confirms that the residuals are uncorrelated.

Table 5: The model adequacy test using the Ljung-Box test

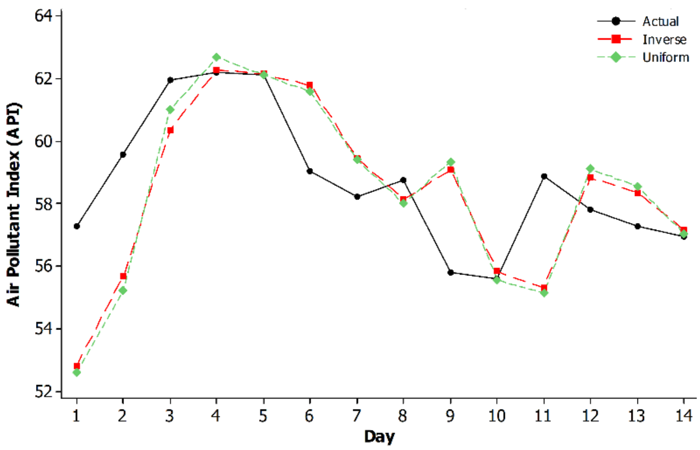
Models	Uniform		Inverse	
Location	χ^2	*p*-value	χ^2	*p*-value
Banting	0.0147	0.9037	0.0142	0.9050
Petaling	1.1718	0.2790	1.6031	0.2055
Shah Alam	1.4145	0.2343	1.4111	0.2349

Table 6: The forecast performance by using RMSE

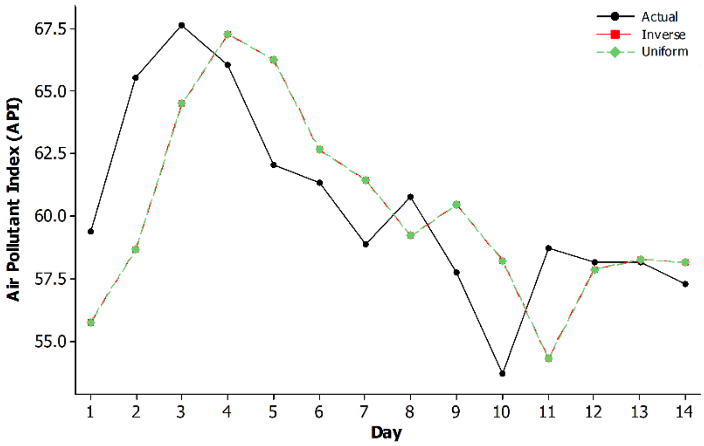
Location	Uniform	Inverse
Banting	5.7988	5.7705
Petaling	2.3916	2.2772
Shah Alam	3.2469	3.2467



(a) Banting



(b) Petaling



(c) Shah Alam

Figure 3: Time series plot of actual and forecast air pollutant data (a) Banting, (b) Petaling, and (c) Shah Alam

Multiple-step ahead forecasting, from December 18th, 2018, to December 31st, 2018, was evaluated by assessing the forecast performance using Root Mean Square Error (RMSE) with uniform and inverse distance weights. The computed RMSE values are presented in Table 6. The data in Table 6 shows that the GSTAR(1,1) model with inverse distance weight yields the lowest average RMSE when compared to the uniform weight model. Inverse distance weights are inversely proportional to the distance between units. Closer locations receive higher weights. Thus, inverse distance weight is better for understanding the spatial effect between the datasets compared to uniform weight. The actual and forecasted values for each location are given in Figure 3.

4. Conclusion

The Generalized Space-Time Autoregressive (GSTAR) model is a commonly used spatial-temporal model in time series analysis, particularly for forecasting. The data visualisation analysis in this study indicates significant correlations between the Air Pollutant Index (API) at three locations: Banting, Petaling, and Shah Alam. This suggests that a spatial-temporal model can effectively forecast the API for these locations. The stationary test reveals that the API data for 2018 are nonstationary in mean and variance. The Box-Cox transformation method and differencing are applied to obtain stationary data to address this. Ensuring stationarity is crucial for precise forecasts using the GSTAR model.

The Root Mean Square Error (RMSE) reveals that the GSTAR model with uniform and inverse distance weights yields nearly identical results. Nevertheless, the GSTAR model with inverse distance weight is favoured due to its lower RMSE value. Specifically, using the inverse distance weight in Petaling leads to a 4.78% improvement in forecast accuracy. These findings emphasise the importance of identifying the spatial relationships between locations to achieve more accurate forecasts.

Air quality significantly affects both the population's health and the country's economic prosperity. Malaysia's current decline in air quality is attributed to industrialisation, an increase in private vehicles, and fossil fuel combustion. Accurate predictions using a spatial-temporal model like GSTAR can aid the government in monitoring and regulating air pollution. In conclusion, the GSTAR model is suitable for predicting API in Selangor. Further development could involve incorporating covariates, using different spatial weights, and comparing them with Maximum Likelihood Estimation (MLE) to enhance forecasting accuracy.

References

Adhikari R. & Agrawal R. 2013. *An Introductory Study on Time series Modeling and Forecasting*. ArXiv, abs/1302.6613

Akbar M.S., Setiawan, Suhartono, Ruchjana B.N., Prastyo D.D., Muhaimin A. & Setyowati E. 2020. A Generalized Space-Time Autoregressive Moving Average (GSTARMA) model for forecasting air pollutant in Surabaya. *Journal of Physics: Conference Series* **1490**: 012022.

Anggraeni D., Kurnia I.F. & Hadi A.F. 2018. Ensemble averaging and stacking of ARIMA and GSTAR model for rainfall forecasting. *Journal of Physics: Conference Series* **1008**: 012019.

Atluri G., Karpatne A. & Kumar V. 2018. Spatio-temporal data mining: A survey of problems and methods. *ACM Computing Surveys* **51**(4): 1-41.

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.

Cliff A.D. & Ord, J.K. 1975. Space-time modelling with an application to regional forecasting. *Transactions of the Institute of British Geographers* **64**: 119–128.

Dickey D.A. & Fuller W.A. 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* **74**(366): 427–431.

Gujarati D.N. & Porter D.C. 2009. *Basic Econometrics*. 5th Edition. New York, US: McGraw Hill Inc.

Kamarianakis Y. & Prastacos P. 2006. Spatial Time-Series Modeling: A review of the proposed methodologies. *University of Crete, Department of Economics, Working Papers*.

Koutsaki E., Vardakis G. & Papadakis N. 2023. Spatiotemporal data mining problems and methods. *Analytics* **2**(2): 485–508.
Leh O.L.H., Ahmad S., Aiyub K., Jani Y.M. & Hwa T.K. 2012. Urban air environmental health indicators for Kuala Lumpur city. *Sains Malaysiana* **41**(2): 179–191.
Manisalidis I., Stavropoulou E., Stavropoulos A. & Bezirtzoglou E. 2020. Environmental and Health impacts of air pollution: A review. *Frontiers in Public Health* **8**: 14.
Marsani M.F. & Shabri A. 2020. Non-stationary in extreme share return: World indices application. *ASM Science Journal* **13**: 1–9.
Monika P., Ruchjana B.N., Abdullah A.S. & Budiarto R. 2023. Systematic literature review on an integrated Generalized Space Time Autoregressive Integrated Moving Average (GSTARIMA) Model with heteroscedastic error and Kriging method for forecasting climate. *Preprints* **2023:** 2023081651.
Munandar D., Ruchjana B.N., Abdullah A.S. & Pardede H.F. 2023. Literature review on integrating Generalized Space-Time Autoregressive Integrated Moving Average (GSTARIMA) and deep neural networks in machine learning for climate forecasting. *Mathematics* **11**(13): 2975.
Ng'ang'a F.W. & Oleche M. 2022. Modelling and forecasting of crude oil price volatility comparative analysis of volatility models. *Journal of Financial Risk Management* **11**(1): 154–187.
Nurhayati N., Pasaribu U.S. & Neswan O. 2012. Application of generalized space-time autoregressive model on GDP data in West European Countries. *Journal of Probability and Statistics* **2012**: 867056.
Ruchjana B.N., Borovkova S.A. & Lopuhaä H.P. 2012. Least Squares estimation of Generalized Space Time AutoRegressive (GSTAR) model and its properties. *AIP Conference Proceedings* **1450**(1): 61–64.
Said S.E. & Dickey D.A. 1984. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* **71**(3): 599–607.
Suhartono, Wahyuningrum S.R., Setiawan, & Akbar M.S. 2016. GSTARX-GLS model for spatio-temporal data forecasting. *Malaysian Journal of Mathematical Sciences* **10**(S): 91-103.
Zewdie M.A., Wubit G.G. & Ayele A.W. 2018. G-STAR model for forecasting space-time variation of temperature in Northern Ethiopia. *Turkish Journal of Forecasting* **02**(1): 9–19.

Department of Mathematics and Statistics
Faculty of Science
Universiti Putra Malaysia
43400 UPM Serdang
Selangor DE, MALAYSIA
E-mail: 201112@student.upm.edu.my, nurhaizum_ar@upm.edu.my, hsyahida@upm.edu.my*

Received: 7 August 2023
Accepted: 4 September 2023

*Corresponding author