Time series forecasting of rain-induced attenuation using coupled ARIMA-SARIMA models for radio propagation applications over subtropical locations

Adewumi O. Ayo^{1*}, Pius A. Owolawi², Joseph S. Ojo³.

^{1,2}Computer Systems Engineering, Tshwane University of Technology, Pretoria 0152, South Africa; hardewumiayour@gmail.com (A.O.A.) owolawiPA@tut.ac.za (P.A.O.) ³Department of Physics, Federal University of Technology, Akure 034, Nigeria; ojojs_74@futa.edu.ng (J.S.O.).

Abstract: This present research aimed to identify the optimum model for predicting the rain-induced attenuation along the terrestrial and satellite communication networks and employed the time series forecasting models of Autoregressive integrated moving average (ARIMA) and Seasonal Autoregressive integrated moving average (SARIMA) to examine and contrast several techniques for predicting raininduced attenuation across a sub-tropical area of South Africa. In this research, the ARIMA was developed using the Box and Jenkins technique to predict long-term rain-induced attenuation in the following South African provinces: Kwazulu-Natal, Eastern Cape, Gauteng, and Northern Cape. The datasets used for the research were obtained from the South African Weather Station from 1994-2023 were used to build and check the model after generating rain-induced attenuation using synthetic storm techniques (SST). The forecasting performance of the seasonal autoregressive integrated moving average (SARIMA) model and that of autoregressive integrated moving average (ARIMA) were compared with four forecast performance measures: - Mean Squared Error (MSE), Mean Absolute Error (MAE), R- squared or the coefficient of determination (R2) and Root Mean Squared Error (RMSE). The results of the study showed that due to its lower forecast performance error indicators, SARIMA performed better than ARIMA in forecasting rain-induced attenuation in South Africa. There is no statistically significant difference between the two projected values, according to the results of a Ljungbox test for significant difference. The authors draw the inference that the two approaches can effectively be employed in their proper locations. Rain-attenuation prediction data indicates that this long-term projection can help decision makers to improve the performance of microwave networks in the face of random fluctuations and avoiding unexpected signal loss with efficient installation of communications infrastructure. It also has applications in prediction of floods, urban planning, and crop management.

Keywords: ARIMA model, Attenuation, Forecast, Rain-induced, SARIMA model, SST.

1. Introduction

Rain in propagation paths degrades signals' performance in millimetre wave telecommunication systems [1], especially for radio signals above 10 GHz along both terrestrial and satellite paths [1]. In any way, tropospheric collisions can induce signal loss on Earth-space pathways for prolonged periods of time, resulting in a decrease in the quality and availability of actual signals. Precipitation attenuation, depolarization, and tropospheric scintillation are likely the most fundamental tropospheric influences, which are outlined here under: (1) When there is insufficient power for successful communication with the recipient, the link is inaccessible. (2) deteriorated performance, as shown by intervals when service quality falls below the desired level; and (3) Depolarization causes undesirable signal interference from other sources operating at comparable frequencies or cross-polarization blockage (intersystem or intraframework). The goal of this study is to determine the optimum model for predicting signal loss

along terrestrial and satellite communication networks to identify rain-induced attenuation. The study time series attenuation demonstrates the probability of attenuation occurring during precipitation within a specific timeframe, facilitating the comprehension of signal loss during transmission. The primary obstacles are maintaining a high quality of service (Qos) and managing rain-induced attenuation, which has an impact on communication networks. Rain attenuation is the term used to describe the attenuation of a microwave transmission caused by these factors. It is a significant flaw in the transmission of information for both satellite and terrestrial microwave connections. Microwave networks are vulnerable to random oscillations in the wireless channel, which complicates network connection design $\lceil 2 \rceil$. To predict such events, a variety of approaches, including numerical and machine learning procedures based on historical time series and radar data, have been utilized $\lceil 2 \rceil$.

Currently, the most common technique for rain attenuation prediction collects and analyses radar image data from several organizations to forecast rain attenuation. Many quantitative studies, however, are frequently useful in predicting rainfall [3]. The Box and Jenkins model [4], was improved on by Box et al. [5], which is one of the most effective methods for analysing time series data, commonly known as Autoregressive Integrated Moving Average (ARIMA) [6],[7]. ARIMA has been frequently used to forecast rainfall trends throughout the years [3], reservoir and river modelling [8], economics and production [9], and evapotranspiration [10]. The approach has some intriguing qualities that make it appealing to researchers. It facilitates forecasting by enabling researchers to utilize only one variabletime data series for simple scenarios while permitting numerous more complicated cases. To intelligently mark out rain-induced signal attenuation, limit attenuation impacts, and enable network systems to make the right decisions ahead of time, an ARIMA or SARIMA model for time series forecasting must be designed and assessed to help anticipate and enhance microwave network performance [11]. The study aims to compare different models over this region to determine the signal attenuation induced by rain and to perform real-time predictions on rain attenuation data for application on microwave, terrestrial, and satellite communications systems.

For improved performance, there is a need to carry out different forecasting models that are reviewed and compared in different geographical regions. The proposed model can anticipate rain attenuation with high accuracy and assist network systems in making appropriate decisions to prevent rain fade. Time series data analysis may be used to simulate meteorological data such as humidity, temperature, rainfall, and other environmental factors [11]. Forecasting attenuation is critical for making significant decisions and carrying out strategic planning. The capacity to objectively anticipate and forecast attenuation leads to the management of signal and water-related problems such as signal loss, quality of service, floods, and drought, among other concerns [12]. As a result, forecasting attenuation is beneficial for anticipating severe occurrences that may result in the loss of radio transmissions over 10 GHz on both terrestrial and satellite communication channels. By using a set of historical data, such as precipitation over time, the ARIMA model is a variant of time-series data analysis that can be used to predict future events [12]. Hence, to predict future rainfall, the ARIMA model is used in this study. [13], [10, 11] performed research on rainfall forecasting using the ARIMA model. The growth in population coincides with higher frequency demand in terrestrial and satellite communication systems, an increase in demands for domestic water supplies, and at the same time results in higher frequency consumption due to expansion in the telecommunications industry [14]. Mismanagement and a lack of knowledge about existing frequencies and changes in climatic conditions have consequences for an imbalance in signal strength and rain fade. A few pieces of literature exist in the time series analysis of monthly rainfall in some states in Nigeria; they include: Akwa-Ibom State, Uyo [14]. Over the years, SARIMA has frequently been used to predict rainfall trends [2], modelling of streams and lakes [3], manufacturing and finances [15], loss of water [16]. Using the SARIMA and MLP approaches to anticipate prices on the Indian power exchange traded market, the findings show that the SARIMA model had a high prediction accuracy [17] and was further validated by [18] for automatic SARIMA modeling and forecast. Price forecasting has also made extensive use of the SARIMA models. For instance, the SARIMA model outperformed the Back-propagation Neural

Edelweiss Applied Science and Technology ISSN: 2576-8484 Vol. 8, No. 6: 3100-3128, 2024 DOI: 10.55214/25768484.v8i6.2661 © 2024 by the authors; licensee Learning Gate

Network in providing a more generic and accurate forecast for the Korea Stock Prices Index (KOSPI) [19].

Researchers shows interest due to the strategy because of its fascinating features. It makes forecasting easier by allowing researchers to use just one variable-time data series for straightforward scenarios while allowing a variety of more complex cases. A SARIMA model for time series forecasting must be created and evaluated in order to accurately identify rain-induced signal attenuation, restrict attenuation effects, and allow network systems to take the appropriate actions in advance. This will help improve broadband network reliability. The SARIMA model has the advantage of being able to study stochastic, seasonal, and even periodic circumstances in the time series under consideration. Using the ARIMA and SARIMA models, the research time series attenuation illustrates the likelihood that attenuation would happen during precipitation at a certain time frame, which aids in understanding the occurrence of signal loss during transmission. The study analyzed 2023 data and predicted future occurrences using a set of attenuation time series data.

The SARIMA model (0, 1, 0) (0,1,1) $_{(12)}$ was the best model of all stations considered for this study by observing the behaviour of the plot of the autocorrelation function (ACF) and the Partial autocorrelation function (PACF). Moreover, the diagnostic check led to opt for SARIMA (0,1,0) (0,1,1) $_{(12)}$ as the p-value was seen to be statistically significant. The Authors conducted a t- test of significance between the forecast values of the fitted model and thus actually observed. A prediction model can assist microwave networks in improving performance in the face of random fluctuations and avoiding unexpected losses [20]. Wireless technology is critical in next-generation radio access networks, and this paradigm can assist microwave network systems in advancing [21]. Hence, the purpose of this our work.

2. Materials and Methods

2.1. Study Area

Figure 1 depicts a topographic map of South Africa (henceforth referred to as SA), highlighting the most relevant provinces and sites where data for this study was obtained [6]. Since the majority of SA's heartland is higher in altitude and the neighbouring oceans regulate the climate, SA has a wider variety of climates than most other Sub-Saharan African countries and lower average temperatures than other countries in this latitude region, such as Australia.



A map of South Africa with the regions and locations where the data was collected is shown [6].

The freezing Benguela current, which flows north from Antarctica along the Atlantic coast, goes up to the west coast, as does the much warmer Agulhas Current, which flows south from the Indian Ocean and is the principal source of rain for much of the country. The constant evaporation of the eastern oceans produces abundant rainfall, but the Benguela current holds moisture, resulting in arid conditions in the west [6]. Even though South Africa receives about 500 mm of rain annually (as opposed to the 860 mm global average), there are significant and unpredictably fluctuating variances. The landscape of South Africa is characterized by nearly plain and flat highlands, with narrow coastal sections. South Africa's climate spans subtropical in the northeast, temperate in the central plateau, and Mediterranean in the southwest. The centre, surrounded by tall mountains, has a steppe environment, whereas the northwest has a desert climate $\lceil 6 \rceil$. The highlands regularly block moist air from the ocean from reaching the steppe; as a result, the climate most frequently alternates between decades of rain and aridity. The average yearly temperature and precipitation influence the distinction between steppes and deserts. The steppe may easily become a desert with a little less rain. It would be categorized as a prairie if it received more rain [6]. South Africa observes four distinct seasons: spring (May to July), winter (May to July), and summer (mid-October to mid-February), with average highs of 300 degrees Celsius. The majority of the summer months are hot and sunny, and places with significant rainfall will have high attenuation; the region is known for severe convective storms that are accompanied by thunder, lightning, and frequently hail $\lceil 6 \rceil$. With the exception of the Western Cape, which has a Mediterranean climate and gets rain throughout the winter, the majority of attenuation occurs in South Africa during the summer $\lceil 6 \rceil$. Except in the highlands, where orographic effects may generate heavy showers, winter rain is typically prolonged and rarely severe. There is a transitional zone between the winter and summer rainfall areas; autumn and spring, where rain occurs in all seasons, not just winter and summer. A southern coastal strip with an average annual precipitation total of 375-875 mm and a drier interior corridor beyond the coastal hills with an average annual precipitation total of 50-250 mm make up this transition zone [6]. As a result, there is need to forecasting rain-induced attenuation over all the South Africa a using the time series model. The augmented Dickey-Fuller test and Ljung-Box test could not find any evidence to contradict the original date's plot, which indicates that the time series is stationary.

Edelweiss Applied Science and Technology ISSN: 2576-8484 Vol. 8, No. 6: 3100-3128, 2024 DOI: 10.55214/25768484.v8i6.2661 © 2024 by the authors; licensee Learning Gate

2.2. Data Collection

For over ten years, South African Weather Services (SAWS) has been collecting daily rainfall data for all provinces in respect to different locations in South Africa and the neighbouring islands. The Stations are as follows: Bloemfontein, Durban, Nelspruit, Polokwane, Pretoria, and East London. At least one station was assigned to each province to offer a picture of the region's climatic features. The data utilized is the rain rate data gathered by the South African Weather Service in collaboration with Tshwane University of Technology, Pretoria, from 1994 to 2023. However, the data utilized in modelling ranges from January 2023 to December 2023, being compared to the forecasted outcome. The data were collected by inquiring to the relevant authorities on a valid letter of authorization to do research. The rain rate data has been converted in [22] and the rain induced attenuation data was implemented in [23].

2.3. Data Analysis

The data starting October 1994 to December 2023 constituted the training set, while the data from January 2023 served as the test set. The input data is prepared before being supplied to the time series model. Standardization aids in rescaling the input data to a comparable scale. The command standardizes the inputs to fall inside the range (-1,1) [24]. This preprocessing task assists the modelling process by standardizing all inputs, making them easily comparable.

In order to forecast the testing, set of the daily rain attenuation time series, and the training set of data were fitted using the ARIMA and SARIMA models with metrological variables. To achieve the best model's performance, four indices were used. Root mean square error (RMSE) is a performance metric that is used to compare the actual values to the forecast values.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (z_m - \bar{z}_p)^2}{n}}$$
(1)

Furthermore, mean absolute error (MAE), mean square error (MSE), and correlation coefficient (R^2) is used as the performance metric as shown below:

$$MAE = \frac{\sum_{i=1}^{n} |z_m - \bar{z}_p|}{n}$$
(2)
$$MSE = \frac{\sum_{i=1}^{n} (Z_m - \bar{Z}_p)^2}{(3)}$$

$$R^{2} = 1 - \frac{\sum_{i=1}^{n} \left(Z_{m} - \bar{Z}_{p}\right)^{2}}{\sum_{i=1}^{n} \left(Z_{m} - Z_{p}\right)^{2}}$$
(4)

where Z_m and \tilde{Z}_p are measured and predicted values of output rain attenuation, where n is the number of observation and Z_p is the average of estimated data. The Akaike Information Criteria (AIC) is the primary criterion used to choose the optimal SARIMA model. The AIC measures the relative effectiveness of various statistical techniques, and among its many values, the proposed model, which is utilized for forecasting, has the lowest AIC value [25]. The performance of the model is assessed using some statistical evaluation techniques, including the correlation coefficient (R²), mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE).

Kullback-Leibler information was suggested to be used for model selection by Akaike (1974), mentioned by Burnham and Anderson (2001). Another parameter considered is the Akaike Information Criterion (AIC), a commonly deployed evaluation metric for both ARIMA and SARIMA models. It measures the model's simplicity as well as its quality of fit. It needs to be as low as feasible [26].

AIC = -2[log - likelihood] + 2n

where n is the number of estimated parameters (i.e., the number of variables plus the intercept) that are part of the model. The statistical output provides the log-likelihood of the model given in the data, which is a convenient way to assess the model's overall fit (lower numbers imply worse fit). The ARIMA and SARIMA model was subjected to a variety of goodness of fit calculations, including MSE, R², RMSE, AIC, and MAE

2.4. Tests for Stationarity

The term "stationary model" refers to a recently popular type of stochastic model for explaining time series. Stationary models suggest that the process is statistically balanced, with probabilistic properties that remain constant throughout time, implying that it oscillates around a set constant mean level with constant variation. The fundamental idea behind stationarity is that the probability laws that control a process' behaviour do not alter over time. The method is statistically balanced in several aspects. For any time point t_1, t_2, \ldots, t_n , and any time lag k. $\langle Y_t \rangle$ is said to be strictly stationary if the joint distribution of $Y_{t_1}, Y_{t_2}, \ldots, Y_{t_n}$ is the same as the joint distribution of $Y_{t_1} - k, Y_{t_2} - k, \ldots, Y_{t_n} - k$. where Y_t is the observed time series, Y_{t_1} is the observed time at t=1, k is the time lag, t is time and ρ_k is the autocorrelation at lag k [27]. Thus, for every t and k, the (univariate) distribution of $Y_{t_1} - k$ when n =1; in other words, the Y's are marginally distributed. Then a E $(Y_t - k)$ for all t and k so that the mean function is constant across time.

Furthermore, a $Var(Y_t - k) = var(Y_s - k)$ for all t and k, implying that the variance is constant throughout time. With n = 2, we can see that the bivariate distribution of a Y_t and Y_s must be the same as that of $(Y_t - k)$ and $(Y_s - k)$, implying that a $Cov((Y_t, Y_s) = Cov(Y_t - k, Y_s - k))$ for every k, t and s. Putting k = s and k = t yields;

$$\begin{aligned} \gamma_{t,s} &= Cov(Y_{t-s}Y_0) \\ &= Cov(Y_0,Y_{s-t}) \\ &= Cov(Y_0,Y_{|s-t|}) \end{aligned}$$

$$= (\gamma_{0,|t-s|})$$

And it is the covariance between Y_t and Y_s is indeed time dependent via the time difference |t-s| rather than the actual times t and s. As a result, we may simplify our notation for a stationary process and write:

$$\gamma_{k} = Cov(Y_{t}, Y_{t-k}) \text{ and } a \rho_{k} = Corr(Y_{t}, Y_{t-k})$$
(1)
Moreover, note that a $\rho_{k} = \frac{\gamma_{k}}{\gamma_{0}}$
The general properties now become:

The general properties now become:

$$\begin{array}{l} \gamma_0 = Var(Y_t) \quad \rho_0 = 1 \\ \gamma_k = \gamma_{-k} \quad \rho_k = \rho_{-k} \\ |\gamma_k| \le \gamma_0 \quad |\rho_k| \le 1 \end{array}$$

If a phenomenon is entirely stable and possesses finite variance, the covariance function is solely influenced by the time lag [27]. 2.5. Autoregressive Models

The autoregressive model is a stochastic model that may be particularly effective in representing certain realistically occurring time series. In this research, the process' present value is represented as a finite, linear aggregate of its past values together with a random shock, a_(t.) proceeding to use an evenly spaced period to indicate the worth of a procedure $t, t_{-1}, t_{-2} \dots$ by $Z_t, Z_{t-1}, Z_{t-2}, \dots$. Additionally let a $\widetilde{Z_t} = Z_t - \mu$ be the series of deviations from μ . Then a $\widetilde{Z_t} = \emptyset_{1\widetilde{Z_{t-1}}} + \emptyset_2 \widetilde{Z_{t-2}} + \dots + \emptyset_p \widetilde{Z_{t-p}} + a_t$ (3)

is called an autoregressive (AR) process of order p, it is a linear model

Edelweiss Applied Science and Technology

ISSN: 2576-8484

Vol. 8, No. 6: 3100-3128, 2024

DOI: 10.55214/25768484.v8i6.2661

 $[\]ensuremath{\mathbb{C}}$ 2024 by the authors; licensee Learning Gate

$$\tilde{Z} = \phi_1 \tilde{x}_1 + \phi_2 \tilde{x}_2 + \cdots + \phi_p \tilde{x}_p + a$$

Relates to a "dependent" variable z and a set of "independent" variables $x_1, x_2, ..., x_p$. As stated in equation (3), the variable z is regressed on prior iterations to its own value, indicating that the model is autoregressive. We can define an autoregressive operator of order p in terms of the backward shift operator B by

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

The autoregressive model (3) maybe written economically as

$$\emptyset(B)_{\widetilde{z}_t} = a_t$$

A stationary or nonstationary autoregressive process can exist. The \emptyset 's weights must be such $\psi_1, \psi_2,...$ are such that the process is stationary. A convergent series is formed by $\psi(B) = \phi^{-1}(B)$. The autoregressive operator a $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$, when viewed as a polynomial in B of degree of p, must have all roots of $\emptyset(B) = 0$ greater than 1 in absolute value; that is, all roots must lie outside the unit circle. For the first-order AR process a $\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + a_t$, this condition is reduced to the requirement $|\Phi| < 1$, as stated by the preceding argument.

2.6. Moving Average (MA) Models

The autoregressive model in equation (3) describes the process's deviation \tilde{Z}_t as a fixed weighted total of p prior deviations $\tilde{Z}_{t+1}, \tilde{Z}_{t-2}, \dots, \tilde{Z}_{t-p}$, in addition to a random shock a_t . It can be observed that \tilde{Z}_t is expressed as an infinite weighted sum of the a's. The finite moving average process is another important paradigm in the representation of observed time series. In the case we take a \tilde{Z}_t , which linearly dependent on a finite number q of proceeding a's.

Hence, $Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$ (4)

is referred to as a moving average (MA) process of order q. A moving average operator of order q may thus be termed, namely:

$$\theta(B) = 1 - \theta_1 B - \theta_1 B^2 - \dots - \theta_q B^q$$

The moving average model could be expressed as: $\tilde{Z}_t = \theta(B)a_t$

which comprises q+2 unidentified variables μ , $\phi_{1,\dots,}\phi_{p,\theta_{1,\dots,}}a$, $\theta_{q,a}$, $\sigma_{a,}^2$, which must be determined from the data in practice.

2.7. Differencing "I" in ARIMA

The ARIMA and SARIMA model includes the order of differencing, shown by the symbol 'd'. It describes how frequently the unprocessed observations in a time series diverge. Making the time series stable, or ensuring that its statistical characteristics, such as variance and mean, remain constant across time series, is the aim of performing differencing. For the reason that it enables us to model the data using a fixed set of time-independent parameters, stationarity is a crucial presumption in many time series models, including ARIMA and SARIMA. Analysing the partial autocorrelation function (PACF), and autocorrelation function (ACF) plots of the time series data yields the order of differencing, d. The information regarding the correlation between the time series and its lags that is supplied by these graphs may be used to establish the appropriate values of p, d, and q for both ARIMA and SARIMA model. The time series may need first-order differencing (d = 1) to become stationary if the ACF plot reveals a sluggish decrease in the autocorrelation values and the PACF plot reveals a big spike at lag 1 and decays thereafter. The time series may need second-order differencing (d = 2) to become stationary if the ACF plot exhibits a rapid decline while the PACF plot exhibits a notable spike at lag 1 and subsequent decay. The time series may already be stationary if the ACF and PACF plots reveal no significant correlation at any lag, in which case no differencing is needed (d = 0). It's crucial to remember that excessive differencing (i.e., setting d too high) will cause the time series to lose crucial

Edelweiss Applied Science and Technology ISSN: 2576-8484 Vol. 8, No. 6: 3100-3128, 2024 DOI: 10.55214/25768484.v8i6.2661 © 2024 by the authors; licensee Learning Gate

information and result in an erroneous model. To establish the right value of d for your time series data, it is crucial to carefully inspect the PACF and ACF plots.

2.8. Seasonal ARIMA Model.

Time series analyses are frequently used in the study of rainfall data since the data are time structured. Seasonal autoregressive integrated moving average (SARIMA) modeling was applied to the data. An algebraic statement that explains the statistical relationship between a time series and its own past is called an ARIMA model. The seasonal ARIMA model is a multiplicative model that considers both seasonal and non-seasonal elements.

$ARIMA(p,d,q) X(P,D,Q)_{s}(1)$

p is referred to as non-seasonal, AR order, d is non-seasonal difference, q is non-seasonal, MA order, P is seasonal AR order, D is seasonal differencing, Q is seasonal MA order, and S is time span of repeating seasonal pattern. Without differencing operations, the model could be written more formally as:

 $\phi(B^s)\phi(B)(X_t - \mu) = \theta(B^s)\theta(B)w_t (2)$ The non-seasonal component is:

AR:
$$\varphi(B) = 1 - \varphi_i B - \dots, -\varphi_p B^p$$
 (3)
MA: $\theta(B) = 1 + \theta_i B + \dots, +\theta_a B^q$ (4)

The seasonal components are:

Seasonal autoregressive (AR): $\phi(B^s) = 1 - \phi_i B^s -, \dots, -\phi_p B^{ps}$ (4) Seasonal moving average (MA): $\theta(B^s) = 1 + \phi_i B^s +, \dots, +\phi_a B^{qs}$ (5)

2.9. Visualization

A fast exploratory study should be performed before defining the model, the first step should be to verify the quality of the data to see if there is missing data, outliers, check for plausibility, size, type, number of variables, and time stamp intervals. The next step is to determine whether there is a trend, seasonality, or other pattern present to have a better idea of how to build a suitable strategy for determining whether the data is stationary or not [28]. To demonstrate that the data is stable, it is more reassuring to trace the autocorrelation values to zero after a certain delay. The data is not steady if the autocorrelation approaches zero slowly or deviates significantly from zero. The data must be changed to make it stationary if it is not already stationary. One of the most prevalent methods is the differencing technique of mean and modification of variance.

2.10. Correlogram

A correlogram is a stationary method of time series authentication that makes use of the autocorrelation function (ACF) produced by plotting the space between k and k (lag). A population correlogram is formed by graphing k and k. In actuality, we can only rely on the sample autocorrelation function. As k grows, the correlation for stationary data decreases considerably. Meanwhile, the correlogram does not trend to zero for non-stationary data (slow down).

2.11. Pattern Identification

Unless the data is constant in terms of mean and variance, the autocorrelation function (ACF) and partial autocorrelation function (PACF) can be shown below. Some probable acceptable models may be discovered using ACF and PACF. Model identification is used to determine whether autocorrelation and data stationary exist so that the transformation or differencing procedure may be undertaken.

2.12. Estimation of Parameters

The parametric factor level of significance is used to choose the best prediction model. Based on Tables 1–3, there are nine models in total once they have been recognized. Finding out the model's parameter estimation is the next step. According to Table 1–3 findings, the best ARIMA and SARIMA model based on AIC criteria is SARIMA model (0, 1, 0) $(0,1,1)_{(12)}$. It indicates that the SARIMA model with the smallest AIC value is (0, 1, 0) $(0,1,1)_{(12)}$. It is essential to understand whether or not the parameter is significant. A parameter significance test must be carried out as a result. The SARIMA model's parameters are examined using the Ljung box (0, 1, 0) $(0,1,1)_{(12)}$. The important results are listed in Tables 1.

2.13. Residual Assumption Test (Diagnostic Checking)

To determine whether the estimated model is appropriate for the data, diagnostic testing is done. Residual analysis is used for diagnostic purposes. The theoretical basis that the residual is a random variable with a normal distribution and a fixed mean is the basis of the ARIMA and SARIMA models.

2.14. Test for Independence

The Ljung-Box test [29] is used to determine if independent residuals may be used. The first residual is examined using sample autocorrelation from the residual. The following is the hypothesis:

$H_0: \rho_k = 0$ ($H_0:$ nonstationary attenuation data)

$H_1: \rho_k \neq 0H_1:$ stationary attenuation data)

If the p-value < 0.05, reject the attenuation data appears to be steady with respect to its mean and varianceIn the meanwhile, the accuracy is improved by performing a stationary test with the Augmented Dickey-Fuller (ADF) test. [24]. The Criteria for rejecting hypothesis H_0 , if p value is < α , ADF obtains p value = 0.01 < α 0.05, then H_0 is discarded. As a result, the attenuation data may be assumed to be stationary.

2.15. Choosing the Best Model

Akaike's Information Criterion is used to select the optimal model (AIC). The model with the lowest AIC value is selected [24].

3. Results and Discussion

3.1. Preliminary Testing of Attenuation data

Testing the data to see if the mean and variance are stationary is the first stage in time series modelling. If the data is not stationary with either the mean or the variance, differencing is required, and if neither is stationary, a transformation is needed. Since the data is a time series of rain attenuation measurements, Figs. 1–8 show that the series is stationary and has a seasonal cycle. The original data's ACF and PACF, as illustrated in Figs. 9–16, indicate that the attenuation data is stationary. Stationary data in both mean and Variance are required to fit an ARIMA and SARIMA model.



Figure 2. Plot of Bloemfontein Attenuation from 1994 to 2023.



Figure 3. Plot of Durban Attenuation from 1994 to 2023.



Figure 4. Plot of Nelspruit Attenuation from 1994 to 2023.



Figure 5. Plot of Polokwane Attenuation from 1994 to 2023.



Figure 6.

Plot of Pretoria Attenuation from 1994 to 2023.



Figure 7.

Plot of East London Attenuation from 1994 to 2023.

According to Figs. 1-6 there is a shift in attenuation and a decrease in the strength of attenuation. The highest attenuation occurs for about 52 seconds in Nelspruit, which is around 12.4 dB, while the

Edelweiss Applied Science and Technology ISSN: 2576-8484 Vol. 8, No. 6: 3100-3128, 2024 DOI: 10.55214/25768484.v8i6.2661 © 2024 by the authors; licensee Learning Gate

minimum occurs in Bloemfontein (Fig. 1). The average attenuation was 8 dB with a standard deviation of 0.1. A visual examination of Fig. 1 reveals that the attenuation data is stationary with respect to its mean and variance.

3.2. ACF and PACF Identification

Once the data is steady, it is prepared for recognition so that the forecast from the SARIMA model may be obtained (p, d, q) (P, D, Q) (12). To identify it, the attenuation data ACF and PACF are displayed. For ACF and PACF plots in Durban, Pretoria, Polokwane, Bloemfontein, East London and Nelspruit, respectively, rain attenuation data are shown in Figures 8-13.



Figure 8.



Plot of Durban ACF and PACF.



Figure 11. Plot of Polokwane ACF and PACF.



Plot of East London ACF and PACF.

According to the ACF and PACF plots of the time series rain attenuation data, the significant value of the lag is minimal.

3.3. SARIMA and ARIMA Model Prediction

In the context of a few models, we examine ARIMA and SARIMA models. Table 1 displays the combination of the ARIMA and SARIMA model.

ARIMA model	AIC	ADF	P-value	Ljung-Box
(0,1,1)	451.88	2.02	0.01	3.21
(2,1,0)	358.72	2.40	0.02	0.17
(0,1,2)	87.349	0.42	0.02	0.14
(2,1,1)	239.88	0.58	0.01	0.00
(0,1,2)	179.95	0.07	0.01	0.01
(0,1,0)	12.866	1.56	0.04	3.21

Ta	ble 1.							
Pre	diction	Based on A	ARIMA and	I SARIM	A Model	(p, d,	q) (P, 1	$D, Q)_{(12)}$

SARIMA model	AIC	ADF	P-value	Ljung-Box
$(0,1,1)(0,1,1)_{(12)}$	458.32	2.02	0.01	3.31
$(2,1,0)(2,1,0)_{(12)}$	781.15	2.40	0.02	0.32
$\left(0,1,2 ight)\left(0,1,1 ight){}_{\scriptscriptstyle \left(12 ight)}$	532.82	0.42	0.02	0.02
$(2,1,1)(2,1,0)_{(12)}$	577.02	0.58	0.01	0.02
$\left(0,1,2 ight)\left(0,1,2 ight)_{\left(12 ight)}$	370.18	0.07	0.01	0.01
$(0,1,0)(0,1,1)_{(12)}$	188.87	1.56	0.04	1.37

Table	2.
-------	----

Performance metrics for ARIMA Model (p, d, q).

ARIMA model	MSE	RMSE	MAE	\mathbb{R}^2	Std. err
(0,1,1)	1.44	2.10	1.41	0.50	0.093
(2,1,0)	2.45	6.01	2.09	0.40	0.096
(0,1,2)	0.78	0.88	0.12	0.75	0.102
(2,1,1)	1.89	3.60	6.18	0.35	0.289
(0,1,2)	6.36	7.98	8.03	0.50	0.060
(0,1,0)	7.31	2.70	3.72	0.40	0.016

Table	e 3.
Laor	c o.

Performance metrics for SARIMA Model (p, d, q) (P, D, Q) (12)

SARIMA model	MSE	RMSE	MAE	\mathbf{R}^{2}	Std. err
$(0,1,1)(0,1,1)_{(12)}$	86.58	9.31	6.19	0.86	0.090
$(2,1,0)(2,1,0)_{(12)}$	75.18	8.67	5.71	0.52	0.101
$(0,1,2)(0,1,1)_{(12)}$	26.87	5.18	4.04	0.35	0.183
$(2,1,1)(2,1,0)_{(12)}$	28.66	5.35	3.27	0.33	0.086
$\left(0,1,2 ight)\left(0,1,2 ight)_{\left(12 ight)}$	3.54	1.88	1.36	0.30	0.376
$(0,1,0)(0,1,1)_{(12)}$	7.31	2.70	3.71	0.40	0.353

3.4. Forecasting

Time series analysis's principal purpose is to anticipate future values [5]. The model is prepared for prediction after the best one is chosen. The forecasting approach is meant to increase trust in the upcoming data. It was observed that the attenuation for each location under study follows the same trend, indicating that the attenuation will occur in the same period of the transmission of the signal at various locations across South Africa. In order to determine whether non-stationarity, seasonality, and trend were present, the time series plot and correlogram of the series were examined. The time plot of Figures 2 to 7 shows wave-like patterns, which makes it clear that the series has seasonal components. The sinusoidal structure of the ACF plots in Figures 8 to 13 serves as proof of this. Furthermore, it is

evident from Figs. 7-12 that neither the mean nor the variance of the series showed any apparent fluctuations during the investigated period. We applied seasonal differencing to the series, as evidenced by the time series plot, which now shows stationarity in the series. The results of the Augmented Dickey Fuller (ADF) Test, which are displayed in Table 1, further verify that the series became stationary following the application of seasonal differencing. Table 1 -3 is the summary statistics and graphs in Fig. 13 to 18 is attenuation time series and box plot of rain induced attenuation in locations under consideration. The graph shows a gradual decrease in attenuation during the late hours of the day. The best model for both ARIMA and SARIMA model was identified using AIC, ADF and p-value. The PACF plots show notable rises during the first two seasonal lags (Lags 2 to 9, where lags are in months) when looking at Figs. 7-12. The non-seasonal delays show notable increases. ACF plots, depicted in Figures 8 to 13, reveal a single significant peak during the first seasonal lag (Lag 2). SARIMA (0, 1, 0) (0,1,1) (12) appears to be a suitable model based on the behaviors of the correlogram of the seasonally differenced series. The model's coefficients were evaluated using the ADF, and the results were satisfactory for time series attenuation prediction. ARIMA and SARIMA model were evaluated and compared in terms of root mean square errors (RMSE), mean absolute errors (MAE), mean absolute percentage errors (MAPE) and Correlation co-efficient (R^2) . The model that meets the least amount of the specified requirements is the best one.

The time series plot, ACF and PACF plots, histograms, and Normal Q-Q plots of the fitted model's residuals were also examined in order to assess the model's suitability. The Ljung-Box Test was utilized to determine whether or not there was serial correlation among the fitted model's residuals. As expected, Figures 14 to 19 demonstrate explicit that the variance remains constant over time and the residuals snap about zero. Furthermore, every spike in the ACF and PACF plots in Figures 8 to 13 is within boundaries, with the exception of Bloemfontein and East London, indicating that none of the spikes are substantial and are not autocorrelated. The residuals' normalcy is shown in Figures 13 to 18. The Ljung-Box plot Test supports the conclusion that the residuals are normal and do not exhibit autocorrelation, with a test statistics value of 1.37 and a P-value of 0.04. Finally, we use graphs of actual and forecasted performance measures as well as the forecasted performance measure in Figure 14–Fig. 19 and Table 2 and 3, respectively, to show how well the fitted ARIMA and SARIMA model performed in forecasting rain-induced attenuation of the stations under investigation. As can be observed in Tables 2 and 3, the values predicted by the fitted SARIMA model are closer to the actual values at more time points than the ARIMA model. This allows us to evaluate the prediction performance of the two techniques.







Bloemfontein ARIMA, SARIMA attenuation forecast, residual, and fitted model's Q-Q plot.





Figure 15.

Durban ARIMA, SARIMA attenuation Forecast and residual with Q-Q plot fitted for the model.





Figure 16. Nelspruit ARIMA, SARIMA attenuation Forecast, residual, and fitted model's Q-Q plot.







Polokwane ARIMA, SARIMA attenuation Forecast, residual, and fitted model's Q-Q plot.







Pretoria ARIMA, SARIMA attenuation Forecast, residual, and fitted model's Q-Q plot.





Edelweiss Applied Science and Technology ISSN: 2576-8484 Vol. 8, No. 6: 3100-3128, 2024 DOI: 10.55214/25768484.v8i6.2661 © 2024 by the authors; licensee Learning Gate



Figure 19.

East London ARIMA, SARIMA attenuation Forecast, residual, and fitted model's Q-Q plot.

4. Conclusions

Implementing four performance measures, correlation analysis, and a t-test for significant difference, the current research has effectively evaluated the prediction performance of ARIMA and SARIMA in forecasting rain attenuation of the stations under consideration in South Africa. The outcome demonstrates that the two sets of values did not significantly differ from one another. The two models were compared using four forecast error statistics: RMSE, MSE, MAE and R². The SARIMA model was shown to have the least amount of error based on these statistics, making it the most reliable model. The model developed in this study can assist microwave networks in finding an effective strategy to analyze and mitigate rain attenuation to achieve improved wireless transport network control. The correlation index indicates that there is no significant difference in the prediction values obtained from the two approaches, despite SARIMA having the lowest forecast error as indicated by the performance metrics. Thus, they can be successfully used as substitute for loss of signal of forecasted province in South Africa. Having tested the signatures of a group of different models, the SARIMA (0, 1, 0) (0,1,1)(12) model proved to be the best fit for the attenuation time series for the South Africa region based on AIC, with a maximum attenuation of 12 dB and a seasonal shift from the previous period over the studied locations. The SARIMA (0, 1, 0) (0,1,1) (12) method was shown to have the highest accuracy in forecasting future precipitation and the related rain-induced attenuation, with a 98% confidence interval, after the necessary tests and forecast observations were completed. The model could be used to estimate the ideal planning time to minimize the influence of signal attenuation on subtropical terrestrial-tosatellite communication systems. Farmers, construction companies, government agencies working in South Africa are therefore advised to choose SARIMA modelling of rain-induced attenuation over ARIMA. The most significant restriction is the infrastructure used to gather the data. Hybrid models will be created in the future in order to analyze, contrast, and research the problem of signal loss in South Africa.

Funding Statement:

This research has been performed as part of the PhD research under the mentorship of Prof. Pius. A. Owolawi and Prof Joseph. S. Ojo of the Tshwane University of Technology Pretoria, South Africa, and Federal University of Technology Akure Nigeria, respectively. The writing and editing were carried out by the two Professors and has been approved to be published in this innovative journal. We therefore declare no conflict of interest in publishing this paper.

Data Availability:

The data was obtained from the South African Weather Service, Pretoria. https://www.weathersa.co.za/

Conflicts of Interest:

The authors declare that there is no conflict of interest with the publishing of this article for the corresponding author and the co-authors.

Authors Contributions:

This paper was carried out in collaboration between all the authors. Prof Pius A. Owolawi designed the study. Corresponding author Adewumi. O. Ayo performed the statistical analysis, wrote the protocol, wrote the first draft of the manuscript, and managed to the literature searches. Prof. Joseph. S. Ojo managed the analysis of the study. Co-authors read and approved the final manuscript.

Acknowledgments:

The authors gratefully acknowledge the South African Weather Services (SAWS) for the data used in this work and Tshwane University of Technology, Pretoria, South Africa.

Copyright:

 \bigcirc 2024 by the authors. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<u>https://creativecommons.org/licenses/by/4.0/</u>).

References

- [1] Y. P. Pasaribu, H. Fitrianti, and D. R. Suryani, "Rainfall forecast of merauke using autoregressive integrated moving average model," in *E3S Web of Conferences*, vol. 73: EDP Sciences, p. 12010, 2018.
- [2] Y. Lai and D. A. Dzombak, "Use of the autoregressive integrated moving average (ARIMA) model to forecast nearterm regional temperature and precipitation," *Weather and forecasting*, vol. 35, no. 3, pp. 959-976, 2020.
- [3] L. Martínez-Acosta, J. P. Medrano-Barboza, Á. López-Ramos, J. F. Remolina López, and Á. A. López-Lambraño, "SARIMA approach to generating synthetic monthly rainfall in the Sinú river watershed in Colombia," *Atmosphere*, vol. 11, no. 6, p. 602, 2020.
- [4] G. Box, "Box and Jenkins: time series analysis, forecasting and control," in *A Very British Affair: Six Britons and the Development of Time Series Analysis During the 20th Century:* Springer, pp. 161-215, 2013.
- [5] G. T. Wilson, "Time series analysis: forecasting and control, by George EP Box, Gwilym M. Jenkins, Gregory C. Reinsel and Greta M. Ljung, 2015. Published by John Wiley and Sons Inc., Hoboken, New Jersey, pp. 712. ISBN: 978-1-118-67502-1," ed: Wiley Online Library, 2016.
- [6] J. Ojo and P. Owolawi, "Development of one-minute rain-rate and rain-attenuation contour maps for satellite propagation system planning in a subtropical country: South Africa," *Advances in Space Research*, vol. 54, no. 8, pp. 1487-1501, 2014.
- [7] G. P. Zhang and M. Qi, "Neural network forecasting for seasonal and trend time series," *European journal of operational research*, vol. 160, no. 2, pp. 501-514, 2005.
- [8] P. Aghelpour and V. Varshavian, "Evaluation of stochastic and artificial intelligence models in modeling and predicting of river daily flow time series," *Stochastic Environmental Research and Risk Assessment*, vol. 34, no. 1, pp. 33-50, 2020.
- [9] R. Nochai and T. Nochai, "ARIMA model for forecasting oil palm price," in *Proceedings of the 2nd IMT-GT Regional Conference on Mathematics, Statistics and applications*, Academia Penang, pp. 13-15, 2006.

Edelweiss Applied Science and Technology ISSN: 2576-8484 Vol. 8, No. 6: 3100-3128, 2024 DOI: 10.55214/25768484.v8i6.2661 © 2024 by the authors; licensee Learning Gate

- [10] M. Valipour, "Ability of Box-Jenkins models to estimate of reference potential evapotranspiration (A case study: Mehrabad Synoptic Station, Tehran, Iran)," *IOSR Journal of Agriculture and Veterinary Science*, vol. 1, no. 5, pp. 1-11, 2012.
- [11] D. P. Patel, M. M. Patel, and D. R. Patel, "Implementation of ARIMA model to predict Rain Attenuation for KUband 12 Ghz Frequency," *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)*, vol. 9, no. 1, pp. 83-87, 2014.
- [12] N. Alex, F. Nalwanga, I. Mugume, B. A. Ogwang, and P. Waswa, "A community perspective of flood occurrence and weather forecasting over Kampala City," *African Journal of Environmental Science and Technology*, vol. 15, no. 5, pp. 188-201, 2021.
- [13] S. H. Bari, M. Rahman, M. Hussain, and S. Ray, "Forecasting monthly precipitation in Sylhet city using ARIMA model," *Civil and Environmental Research*, vol. 7, no. 1, pp. 69-77, 2015.
- [14] Z. A. Shamsan, "Rainfall and diffraction modeling for millimeter-wave wireless fixed systems," *IEEE Access*, vol. 8, pp. 212961-212978, 2020.
- [15] A. Danandeh Mehr, S. Ghadimi, H. Marttila, and A. Torabi Haghighi, "A new evolutionary time series model for streamflow forecasting in boreal lake-river systems," *Theoretical and Applied Climatology*, vol. 148, no. 1, pp. 255-268, 2022.
- [16] D. S. Kumar, B. Thiruvarangan, A. Vishnu, A. S. Devi, and D. Kavitha, "Analysis and prediction of stock price using hybridization of sarima and xgboost," in 2022 International Conference on Communication, Computing and Internet of Things (IC3IoT), IEEE, pp. 1-4, 2022.
- [17] A. Kożuch, D. Cywicka, and K. Adamowicz, "A comparison of artificial neural network and time series models for timber price forecasting," *Forests*, vol. 14, no. 2, p. 177, 2023.
- [18] J. Arlt and P. Trcka, "Automatic SARIMA modeling and forecast accuracy," *Communications in Statistics-Simulation and Computation*, vol. 50, no. 10, pp. 2949-2970, 2021.
- S. H. Na and S. Y. Sohn, "Forecasting changes in Korea composite stock price index (KOSPI) using association rules," *Expert Systems with Applications*, vol. 38, no. 7, pp. 9046-9049, 2011.
 T. S. Rappaport *et al.*, "Wireless communications and applications above 100 GHz: Opportunities and challenges for
- [20] T. S. Rappaport *et al.*, "Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond," *IEEE access*, vol. 7, pp. 78729-78757, 2019.
- [21] M. Wang, T. Zhu, T. Zhang, J. Zhang, S. Yu, and W. Zhou, "Security and privacy in 6G networks: New areas and new challenges," *Digital Communications and Networks*, vol. 6, no. 3, pp. 281-291, 2020.
- [22] L. J. Mpoporo, P. A. Owolawi, and A. O. Ayo, "Utilization of artificial neural networks for estimation of slant-path rain attenuation," in 2019 international multidisciplinary information technology and engineering conference (imitec), IEEE, pp. 1-7, 2019.
- [23] A. O. Ayo, P. A. Owolawi, and J. S. Ojo, "Performance of ARIMA modelling on sub-tropical rain attenuation at Kaband terrestrial link in Kwazulu-Natal, South Africa," in 2021 IEEE International Conference on Microwaves, Antennas, Communications and Electronic Systems (COMCAS), IEEE, pp. 389-394, 2021.
- [24] O. Al Balasmeh, R. Babbar, and T. Karmaker, "Trend analysis and ARIMA modeling for forecasting precipitation pattern in Wadi Shueib catchment area in Jordan," *Arabian Journal of Geosciences*, vol. 12, pp. 1-19, 2019.
- [25] J. E. Cavanaugh and A. A. Neath, "The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 11, no. 3, p. e1460, 2019.
- [26] J. C. Paul, S. Hoque, and M. M. Rahman, "Selection of best ARIMA model for forecasting average daily share price index of pharmaceutical companies in Bangladesh: A case study on square pharmaceutical ltd," *Global Journal of Management and Business Research Finance*, vol. 13, no. 3, pp. 14-25, 2013.
- [27] T. Škrinjarić, "Mills, TC (2019.) Applied Time Series Analysis-A Practical Guide to Modeling and Forecasting, United Kingdom: Academic Press, 339 str," Notitia-časopis za ekonomske, poslovne i društvene teme, vol. 5, no. 1., pp. 119-122, 2019.
- [28] J. Fattah, L. Ezzine, Z. Aman, H. El Moussami, and A. Lachhab, "Forecasting of demand using ARIMA model," International Journal of Engineering Business Management, vol. 10, p. 1847979018808673, 2018.
- [29] A. Geetha and G. Nasira, "Time-series modelling and forecasting: Modelling of rainfall prediction using ARIMA model," *International Journal of Society Systems Science*, vol. 8, no. 4, pp. 361-372, 2016.